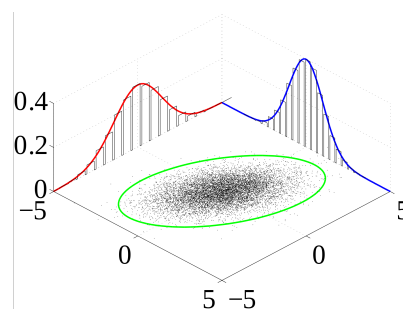
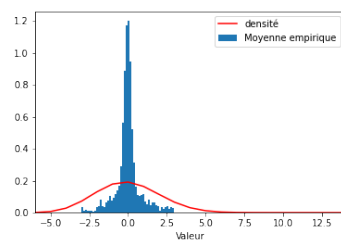
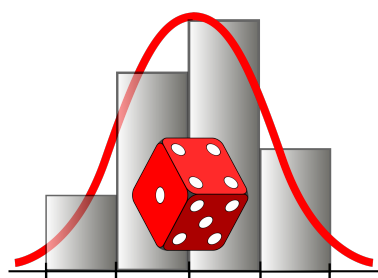


Probabilités

Notions de base d'analyse probabiliste



Paul FRAUX

Version du 11 février 2025

Avant-Propos

Ce polycopié présente les fondamentaux de la théorie des probabilités, utile dans tout cursus scientifique. Il suit le même plan que le cours, en approfondissant certains points, et parfois en allant plus loin (parties marquées d'un H.P.).

Ce cours vise à faire de l'analyse chiffrée pour décrire un phénomène qui dépend du hasard, comprendre et modéliser des phénomènes complexes, aider à la prise de décision et faire des prévisions.

Pour cela, les objectifs de ce cours sont :

- savoir calculer une probabilité,
- connaître les principaux modèles du hasard,
- s'initier aux techniques statistiques.

On commencera par rappeler la théorie générale. Puis nous présenterons les principaux outils en théorie des probabilités, d'abord sur une seule dimension avant d'aborder les vecteurs aléatoires. La fin du cours présentera deux des théorèmes fondamentaux sur les limites d'expériences aléatoires.

Ce polycopié est nouveau, et contient presque sûrement des erreurs, incohérence et fautes d'orthographe. Merci de nous les signaler pour contribuer à son amélioration.

Notations

Ensembles classiques :

- Pour n un entier naturel, on note $\llbracket 1, n \rrbracket := \{1, \dots, n\}$.
- On note \mathbb{N} l'ensemble des entiers naturels, et \mathbb{N}^* l'ensemble des entiers supérieurs à 1.
- On note \mathbb{R} l'ensemble des nombres réels.
- Pour E un ensemble, on note $\mathcal{P}(E)$ l'ensemble contenant tous les sous-ensembles de E .

Fonctions utiles :

- Pour (E, F) deux espaces topologiques, on note $C(E, F)$ l'ensemble des fonctions continues de E vers F . En particulier $C(\mathbb{R}, \mathbb{R})$ est l'ensemble des fonctions réelles continues.
- Pour E un espace vectoriel normé, on note $C^k(\mathbb{R}, E)$ l'ensemble des fonctions k fois dérivable, de $k^{\text{ème}}$ dérivée continue.
- Pour A un sous-ensemble d'un ensemble E , on note $\mathbb{1}_A$ la fonction

$$\begin{aligned} \mathbb{1}_A : E &\rightarrow \mathbb{R}, \\ x &\mapsto \mathbb{1}_A(x) := \begin{cases} 1, & \text{si } x \in A, \\ 0, & \text{si } x \notin A. \end{cases} \end{aligned}$$

Concepts probabilistes :

- Ω est un univers, \mathcal{T} l'ensemble des événements et \mathbb{P} une mesure de probabilité.
- A^c ou \overline{A} est l'événement complémentaire de A , *i.e.* celui où A ne se produit pas.
- \mathbb{P}_A ou $\mathbb{P}(\cdot|A)$ est la probabilité conditionnellement à l'événement A .
- X est une variable ou un vecteur aléatoire réel;
- $\mathbb{E}(X)$ est l'espérance de X , $\text{Var}(X)$ sa variance et σ son écart-type. Si X est vectoriel, Σ est sa matrice de covariance et ρ sa matrice de corrélation.
- F_X ou F est la fonction de répartition de X , f_x ou f est sa densité s'il en a une.
- M_X est la fonction génératrice des moments de X .
- φ_X est la fonction caractéristique de X .

Table des matières

1	Espaces de probabilité	6
1.1	Événements et univers	6
1.1.1	Espace probabilisable	6
1.1.2	Espace probabilisé	8
1.2	Premiers exemples et la formule du savoir	9
1.2.1	Probabilité uniforme	9
1.2.2	Espaces probabilisés au plus dénombrable	10
1.2.3	Espace de probabilité non-discret	10
1.2.4	Conditionnement et formule du savoir	11
1.3	Indépendances d'évènements	13
1.3.1	Définition fondamentale	13
1.3.2	Un exemple d'utilisation : le lemme de Borel-Canteli	14
1.4	Exercices	15
2	Probabilité univarié : outils et exemples classiques	19
2.1	Trousse à outils du probabiliste	19
2.1.1	Variables aléatoires	19
2.1.2	Premier outil : la fonction de répartition	21
2.1.3	Deuxième outil : l'espérance et les moments	22
2.1.4	Un troisième outil : la fonction génératrice des moments (H.P.)	25
2.1.5	Théorème de transfert et autres outils	27
2.1.6	Exercices	27
2.2	Les variables aléatoires discrètes	28
2.2.1	Définition et forme générale	28
2.2.2	La loi uniforme	28
2.2.3	La loi de Bernoulli et les binomiales	30
2.2.4	La loi géométrique	31
2.2.5	La loi de Poisson	31
2.2.6	Exercices	33
2.3	Les variables aléatoires absolument continue	34
2.3.1	Définition et forme générale	34
2.3.2	La loi uniforme	35
2.3.3	La loi exponentielle et la loi Gamma	36
2.3.4	Le changement de variable pour les lois absolument continu	39

2.3.5	La loi normale et la loi log-normale	39
2.3.6	La loi de Cauchy	41
2.3.7	Exercices	41
3	Analyse multivariée des multiples variables	43
3.1	Vecteurs aléatoires et généralisations des outils univariés	43
3.1.1	Tribu pour les vecteurs aléatoires (H.P.)	43
3.1.2	Loi de vecteurs aléatoires et fonction de répartition	44
3.1.3	Espérance, matrice de covariance et de corrélations	46
3.1.4	Deux exemples	48
3.1.4.15	Exercice	50
3.2	Calculer avec plusieurs aléas	51
3.2.1	Marginales	51
3.2.2	Indépendance	52
3.2.3	Changements de variable dans les hyper-espaces	54
3.2.4	Exercice	55
4	Théorèmes limites de l'aléatoire	57
4.1	Fonction caractéristique	57
4.1.1	Définitions de la fonction caractéristique	57
4.1.2	Calculs type de fonctions caractéristiques	59
4.1.3	La fonction caractéristique caractérise la loi	61
4.1.4	Exercice	62
4.2	Les divers modes de convergences des variables aléatoires	64
4.2.1	Convergence en loi	64
4.2.2	Convergence en norme L^p	67
4.2.3	Convergence en probabilité et presque sûr	69
4.2.4	Résumé des relations entre les divers modes de convergence	74
4.2.5	Exercice	74
4.3	Théorèmes limites	75
4.3.1	Loi des grands nombres	75
4.3.2	Théorème central limite de la limite centrale	78
5	Solution ou indications des exercices	80
5.1	Section 1 : espace de probabilité	80
5.2	Section 2 : probabilité univarié	87
5.3	Section 3 : analyse multivariée	94
5.4	Section 4 : théorèmes limites de l'aléatoire	97
A	Loi forte des grands nombres	99
B	Convergence et convolution	104
	Bibliographie	105

Chapitre 1

Espaces de probabilité

La théorie des probabilités a la prétention d'étudier le hasard dans toutes ses formes, et en particulier chercher les lois qui le régissent. Elle cherche à décrire les régularités dans le hasard. Créée au 17^{me} siècle, entre autres par Pascal et Fermat, elle a de nombreuses applications actuelles en sciences théoriques (principe d'incertitude d'Heisenberg en mécanique quantique, théorie de l'évolution de Darwin en biologie...) comme en pratiques (évaluations de sécurité de fonctionnements de systèmes, évaluations de paramètres de matériaux inconnus...)

1.1 Événements et univers

1.1.1 Espace probabilisable

Nous nous intéresserons à une expérience aléatoire et aux événements que l'on peut y associer, afin de pouvoir les mesurer.

Considérons par exemple l'expérience qui consiste à lancer un dé à 6 faces et à noter le résultat de ce lancé X .

Des événements possibles sont alors les ensembles $\{X = 3\}$, $\{X \leq 4\}$ et $\{X \text{ est pair}\}$. Si on rajoute l'évènement *impossible* : $\{X = 7\} = \{X \in \emptyset\} = \{X \neq 1, \dots, 6\} = \{\text{dé en équilibre sur un coin}\} = \emptyset$, alors d'un point de vue ensembliste l'ensemble des événements est exactement $\mathcal{P}(\{1, \dots, 6\})$.

On peut alors naturellement définir des opérations correspondantes au langage naturel et les divers cas possibles.

Si $A \in \mathcal{P}(\{1, \dots, 6\})$ est un événement, alors on peut considérer l'évènement A^c "A n'arrive pas".

Si A et B sont deux événements, on peut considérer l'évènement " A ou B advient" noté $A \cup B$, ou encore l'évènement "A et B arrivent", noté $A \cap B$.

On dira que deux événements sont *incompatibles* ou *disjoint* s'ils ne peuvent pas arriver simultanément, i.e. en termes ensemblistes, si l'évènement $A \cap B$ est l'évènement impossible \emptyset .

On s'attend donc à ce que l'ensemble des événements soit au moins une algèbre de Boole

Définition 1 : H.P.

Soit E un ensemble.

On appelle algèbre de Boole sur E une collection d'ensemble $\mathcal{E} \subset \mathcal{P}(E)$, i.e. $\mathcal{E} \in \mathcal{P}(\mathcal{P}(E))$ qui vérifie :

1. \mathcal{E} est non vide.
2. $\forall A \in \mathcal{E}, A^c \in \mathcal{E}$, c'est-à-dire que pour tout élément de \mathcal{E} , son complémentaire est aussi un élément de \mathcal{E} .
3. $\forall (A, B) \in \mathcal{E}^2, A \cup B \in \mathcal{E}$, c'est-à-dire que pour toute paire d'éléments de \mathcal{E} , leur conjonction est aussi un élément de \mathcal{E} .

Tant que nous étudions des espaces aléatoires avec un nombre fini d'événements, ce cadre conviendrait. Pour étudier des situations plus compliquées, ce cadre pourrait être insuffisant.

Considérons par exemple un joueur qui lance une pièce de monnaie jusqu'à obtenir face :

$$E_n := \{\text{le premier tirage d'un face arrive à la } n^{\text{ieme}} \text{ étape}\}$$

L'événement "le jeu se termine en un nombre fini de coups" est alors $E := \bigcup_{n \in \mathbb{N}^*} E_n$. Bien que $E^c = \bigcap_{n \in \mathbb{N}^*} E_n^c = \{\text{face ne sort jamais}\}$ a bien peu de chance d'arriver, il est raisonnable de pouvoir le considérer dans une théorie générique des probabilités.

De nombreuses questions analogues conduisent à demander :

Définition 2 :

Un ensemble probabilisable est un couple (Ω, \mathcal{T}) tel que Ω soit un ensemble et $\mathcal{T} \in \mathcal{P}(\mathcal{P}(\Omega))$ vérifie :

1. $\Omega \in \mathcal{T}$.
2. $\forall A \in \mathcal{T}, A^c \in \mathcal{T}$.
3. $\forall (A_n)_{n \in \mathbb{N}} \in \mathcal{T}^{\mathbb{N}}, \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$

On dira alors que \mathcal{T} est une σ -algèbre, ou encore une tribu.

On dit que Ω est l'univers.

Exemples : Soit Ω un ensemble, alors $\{\emptyset, \Omega\}$ et $\mathcal{P}(\Omega)$ sont deux tribus.

Proposition 3 :

Soit (Ω, \mathcal{T}) un espace probabilisable, A et B deux événements et $(A_n)_{n \in \mathbb{N}}$ une suite d'événement. Alors :

1. $\emptyset \in \mathcal{T}$
2. $A \cup B \in \mathcal{T}$
3. $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{T}$

Démonstration.

Soit (Ω, \mathcal{T}) , A , B et $(A_n)_{n \in \mathbb{N}}$ comme dans l'énoncé

1. $\emptyset = \Omega^c$ est dans \mathcal{T} d'après les points 1 et 2 de la définition.
2. $A \cup B = A \cup B \cup \bigcup_{n > 1} \emptyset \in \mathcal{T}$ d'après la preuve précédente et le point 3 de la définition
3. $\bigcap_{n \in \mathbb{N}} A_n = (\bigcup_{n \in \mathbb{N}} A_n^c)^c \in \mathcal{T}$ d'après les points 2 et 3 de la définition.

□

1.1.2 Espace probabilisé

Quand on demande la probabilité d'un événement, la réponse est un réel compris entre 0 et 1.

Une probabilité est donc une application de l'ensemble des événements dans $[0,1]$.

L'intuition suggère que la probabilité d'un événement est liée à la fréquence empirique d'occurrence de l'événement. Par exemple, pour un dé à 6 face équilibré, le 3 sort environ dans $\frac{1}{6}$ des cas, on s'attend alors à ce que $\mathbb{P}(X = 3) = \frac{1}{6}$.

Si l'on veut conserver cette intuition, alors on doit prendre comme probabilité une fonction qui vaut 1 pour l'univers, et additive sur les événements disjoints. On pose donc :

Définition 4 :

Un ensemble probabilisé est un triplet $(\Omega, \mathcal{T}, \mathbb{P})$ tel que (Ω, \mathcal{T}) soit un espace probabilisable et \mathbb{P} une probabilité, c'est-à-dire une application $\mathbb{P} : \mathcal{T} \rightarrow [0,1]$ vérifiant :

1. $\mathbb{P}(\Omega) = 1$
2. $\forall (A_n)_{n \in \mathbb{N}}$ suite d'événements deux à deux disjoints, $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$

Remarque : Certains analystes jaloux diront que la théorie des probabilités n'est que l'étude de mesures positive de masse totale 1 (voir un cours d'intégration de Lebesgue).

Proposition 5 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, A et B deux événements et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements. Alors :

1. $\mathbb{P}(\emptyset) = 0$
2. $A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. Cette relation s'étend à n'importe quelle union finie d'événements disjoints.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
4. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
5. (*Continuité croissante*) Si $\forall n \in \mathbb{N}, A_n \subset A_{n+1}$ (i.e. (A_n) est croissante), alors $\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$
6. (*Continuité décroissante*) Si $\forall n \in \mathbb{N}, A_{n+1} \subset A_n$ (i.e. (A_n) est décroissante), alors $\mathbb{P}(\bigcap_{n \in \mathbb{N}} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$
7. (*Croissance*) Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$

Démonstration.

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, A et B deux événements et $(A_n)_{n \in \mathbb{N}}$ une suite d'événements.

1. Comme la suite des événements vides est une suite d'événements deux à deux disjoints,

$$\sum_{n \in \mathbb{N}} \mathbb{P}(\emptyset) = \mathbb{P}(\emptyset) < +\infty.$$

La seule possibilité est alors que $\mathbb{P}(\emptyset) = 0$.

2. Soient deux événements A et B disjoints ($A \cap B = \emptyset$).

Considérons la suite A_n définie par $A_0 = A$, $A_1 = B$ et $\forall n > 1, A_n = \emptyset$.

Il s'agit d'une suite d'événements deux à deux disjoints, donc $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) + \sum_{n>1} \mathbb{P}(\emptyset) = \mathbb{P}(A) + \mathbb{P}(B)$

3. On utilise les décompositions en événements disjoints $A = A \setminus B \cup (A \cap B)$, $B = B \setminus A \cup (A \cap B)$ et $A \cup B = A \setminus B \cup (A \cap B) \cup B \setminus A$. Le point précédent montre alors que $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$, mais aussi que $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$ et que $\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A)$.

En réarrangeant ces trois égalités, on trouve le résultat attendu.

4. On obtient que $A \cup A^c = \Omega$, et ces deux événements sont disjoints. Par le point 2 prouvé précédemment, $\mathbb{P}(A^c) + \mathbb{P}(A) = 1$

5. Supposons A_n croissante et posons $B_n := A_n \setminus \left(\bigcup_{k < n} A_k \right)$.

Alors pour tout $n \in \mathbb{N}$, $B_n \subset A_n$, les B_n sont deux à deux disjoints et $A_n = \bigcup_{k \leq n} A_k = \bigcup_{k \leq n} B_k$.

Alors :

$$\mathbb{P}\left(\bigcup_{k \in \mathbb{N}} A_k\right) = \mathbb{P}\left(\bigcup_{k \in \mathbb{N}} B_k\right) = \sum_{k \in \mathbb{N}} \mathbb{P}(B_k) = \lim_{n \rightarrow \infty} \sum_{k \leq n} \mathbb{P}(B_k) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n),$$

qui est le résultat souhaité.

6. En passant au complémentaire, ce point découle des deux résultats précédents.
7. Supposons $A \subset B$, alors on peut écrire $B = A \cup (B \cap A^c)$ comme réunion disjointe de A et d'un autre événement, donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A)$

□

Définition 6 : Terminologie

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

On appelle événement certain l'événement Ω .

On appelle événement impossible l'événement \emptyset .

On dit qu'un événement A est négligeable si $\mathbb{P}(A) = 0$.

On dit qu'un événement A est presque sûr (ou p.s.) si $\mathbb{P}(A) = 1$.

1.2 Premiers exemples et la formule du savoir

1.2.1 Probabilité uniforme

Si on se donne un ensemble fini E , la probabilité la plus simple que l'on peut se donner consiste à donner exactement le même poids à chaque point de l'ensemble (pensez à un lancer de dé équilibré).

C'est-à-dire demander que

$$\forall x \in E, \mathbb{P}(\{x\}) = \frac{1}{|E|}.$$

Alors, on doit avoir

$$\forall A \subset E, \mathbb{P}(A) = \frac{|A|}{|E|}.$$

Ainsi, faire des calculs avec la probabilité uniforme revient à faire des calculs de cardinaux, *i.e.* de la combinatoire.

1.2.2 Espaces probabilisés au plus dénombrable

Cas fini :

Considérons le cas où l'univers Ω est fini, c'est-à-dire que $\Omega := \{x_1, \dots, x_n\}$. La tribu naturelle sur un tel univers est $\mathcal{P}(\Omega)$, car il s'agit de la seule tribu contenant tous les singletons.

Soit alors \mathbb{P} une probabilité sur Ω , et posons pour tout $i \in \llbracket 1, n \rrbracket, p_i := \mathbb{P}(\{x_i\})$.

On notera que les nombres p_i vérifient :

$$\forall i \in \llbracket 1, n \rrbracket, 0 \leq p_i \leq 1, \quad \text{et} \quad \sum_{i=1}^n p_i = 1. \quad (\text{A})$$

Alors pour tout $A \subset \Omega$, on obtient que

$$\mathbb{P}(A) = \sum_{i: x_i \in A} p_i \quad (\text{o})$$

En particulier, la mesure de probabilité est uniquement déterminée par la donnée des réels p_i .

Réciproquement, étant donné une famille p_i vérifiant (A), on peut définir une probabilité sur Ω via la formule (o).

Cas infini : Considérons maintenant le cas où l'univers Ω est dénombrable, c'est-à-dire que $\Omega := \{x_n, n \in \mathbb{N}\}$, avec

comme tribu $\mathcal{P}(\Omega)$.

Soit alors \mathbb{P} une probabilité sur Ω , et posons pour tout $i \in \mathbb{N}, p_i := \mathbb{P}(\{x_i\})$.

On a encore que :

$$\forall i \in \mathbb{N}, 0 \leq p_i \leq 1, \quad \text{et} \quad \sum_{i \in \mathbb{N}} p_i = 1. \quad (\text{B})$$

Et pour tout $A \subset \Omega$, que

$$\mathbb{P}(A) = \sum_{i \in \mathbb{N}: x_i \in A} p_i \quad (\text{o})$$

Donc la mesure de probabilité est encore uniquement déterminée par la donnée des réels p_i .

Réciproquement, étant donné une famille p_i vérifiant (B), on peut définir une probabilité sur Ω via la formule (o).

Exemple : Soit $(u_n)_{n \in \mathbb{N}}$ une série à terme positif convergente, et on peut définir une probabilité sur \mathbb{N} par la relation :

$$p_n := \frac{u_n}{\sum_{n \in \mathbb{N}} u_n}.$$

Question : Peut-on construire une probabilité uniforme sur \mathbb{N} ? Malheureusement non... Expliquez pourquoi.

1.2.3 Espace de probabilité non-discret

Les espaces finis ou dénombrables sont dits discrets, puisqu'on se ramène toujours à la probabilité des singletons $\mathbb{P}(\{x\})$. Si on veut construire des probabilités plus complexes, on peut par exemple utiliser la mesure de Lebesgue :

Théoreme 7 :

Il existe une unique mesure positive λ sur $\mathcal{B}([0, 1])$ appelée mesure uniforme telle que

$$\forall a < b \in [0, 1]^2, \lambda([a, b]) = b - a$$

Elle permet de construire un modèle non discret uniforme, qui permet de modéliser "tirer uniformément un nombre aléatoire entre 0 et 1".

1.2.4 Conditionnement et formule du savoir

Supposons maintenant qu'on dispose d'une information partielle sur la nature du résultat, notée B. Pour un événement A on ne va donc retenir que ce qui est commun à A et à B en considérant que le nouvel univers des possibles se résume à B. En suivant cette intuition, on définit :

Définition 8 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{T}$ un événement de probabilité non nulle ($\mathbb{P}(B) \neq 0$).

On appelle probabilité conditionnelle de A sachant B, noté $\mathbb{P}(A|B)$ ou $\mathbb{P}_B(A)$ le réel

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Proposition 9 :

L'application $\mathbb{P}(\cdot|B) : A \in \mathcal{T} \mapsto \mathbb{P}(A|B)$ est une mesure de probabilité sur (Ω, \mathcal{T}) .

Démonstration.

On a bien que $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.

Soit $(A_n)_{n \in \mathbb{N}} \in \mathcal{T}^{\mathbb{N}}$ d'événements deux à deux disjoints, alors $(A_n \cap B)$ est une suite d'événements deux à deux disjoints. Alors :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n | B\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n \cap B\right)}{\mathbb{P}(B)} = \frac{\sum_{n \in \mathbb{N}} \mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n | B).$$

□

Remarque : On aura en particulier toutes les propriétés élémentaires des probabilités, par exemple que $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$.

Remarque 2 : On a que $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

On a les résultats élémentaires suivants, conséquences immédiates de la proposition et des remarques :

Proposition 10 : Probabilités composées

Soient A_1, \dots, A_n n événements tels que $\mathbb{P}(\cap_{k=1}^{n-1} A_k) \neq 0$, on a :

$$\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2 \cap A_1) \cdots \mathbb{P}(A_n|\bigcap_{k=1}^{n-1} A_k)$$

Proposition 11 : Probabilités totales

Soit $(A_k)_{k \in \llbracket 1;n \rrbracket}$ un système complet d'événements, c'est-à-dire si $i \neq j$, alors $A_i \cap A_j = \emptyset$ et $\bigcup_{k \in \mathbb{N}} A_k = \Omega$.
On suppose de plus que pour tout $k \in \llbracket 1;n \rrbracket$, $\mathbb{P}(A_k) > 0$. On a alors pour tout $B \in \mathcal{T}$:

$$\mathbb{P}(B) = \sum_{k \in \llbracket 1;n \rrbracket} \mathbb{P}(B|A_k)\mathbb{P}(A_k).$$

Nous pouvons utiliser cette formule pour compléter la formule de Bayes suivante. La formule de Bayes est essentielle, puisqu'elle nous dit comment il faudra ajuster *a posteriori* notre connaissance *a priori* de la probabilité de l'événement A en fonction de la probabilité de l'événement B bien connu qui vient de se produire.

Théoreme 12 : Formules de Bayes, ou la formule du savoir

Soit A et B deux événements de probabilité non nulle, alors

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

On en déduit que si $(A_k)_{k \in \llbracket 1;n \rrbracket}$ un système complet d'événements avec $\forall k \in \llbracket 1;n \rrbracket, \mathbb{P}(A_k) > 0$, alors

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{j=1}^n \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

1.3 Indépendances d'évènements

1.3.1 Définition fondamentale

Définition 13 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et (A, B) un couple d'évènements.

On dit que A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Il ne faut pas confondre événements incompatibles, qui veut dire que $A \cap B = \emptyset$, avec événements indépendants. En particulier, deux événements incompatibles, de probabilités non nulles, sont dépendants : si l'un est réalisé, l'autre ne l'est pas.

Si $\mathbb{P}(A)$ ou $\mathbb{P}(B)$ est nul, alors A et B sont indépendants.

Remarque : Si $\mathbb{P}(B) > 0$, on a l'équivalence A et B indépendants $\iff \mathbb{P}(A|B) = \mathbb{P}(A)$, c'est-à-dire que la connaissance de la réalisation de B n'informe en rien sur la réalisation de A .

Exercice : Montrer l'équivalence A et B indépendant $\iff A$ et B^c indépendant $\iff A^c$ et B^c indépendant.

Et maintenant, que faire si l'on considère plus que 2 événements ? On utilise alors la définition suivante :

Définition 14 :

Soient $(A_i)_{i \in I}$ une famille d'évènements. On dira que ces événements sont mutuellement indépendants (ou indépendant dans leur ensemble) si pour tout sous-ensemble $J \subset I$ fini, on a :

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

On voit facilement que l'indépendance mutuelle entraîne l'indépendance deux à deux. La réciproque est fautive comme le montre l'exemple suivant :

considérons le jet de deux dés différenciés (l'un rouge et l'autre jaune). On note :

- A l'évènement « le dé rouge donne un nombre pair » ;
- B l'évènement « le dé jaune donne un nombre pair » ;
- et enfin C l'évènement « la somme des deux dés est paire ».

On vérifie facilement que A , B et C sont deux à deux indépendants, mais non mutuellement indépendants.

Remarque : La notion d'indépendance des variables aléatoires que nous verrons dans la partie 2 est très similaire.

1.3.2 Un exemple d'utilisation : le lemme de Borel-Canteli

Voyons une première application de l'indépendance pour déterminer la plausibilité de certains événements asymptotiques.

Lemme : (Borel-Cantelli)

1. Si la série $\sum_n \mathbb{P}(A_n)$ converge, alors $\mathbb{P}(A_n \text{ arrive une infinité de fois}) = 0$
2. Si la série $\sum_n \mathbb{P}(A_n)$ diverge et que **la famille (A_n) est indépendante dans son ensemble**, alors $\mathbb{P}(A_n \text{ arrive une infinité de fois}) = 1$

Démonstration.

1) Écrivons

$$\begin{aligned} \mathbb{P}(A_n \text{ arrive une infinité de fois}) &= \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k\right) = \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \\ &\leq \lim_{n \rightarrow \infty} \sum_{k \geq n} \mathbb{P}(A_k) = 0 \end{aligned}$$

2) Commençons par remarquer que

$$\mathbb{P}(\{A_n \text{ arrive une infinité de fois}\}^c) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k^c\right) \leq \sum_{n \in \mathbb{N}} \mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right)$$

Il suffit donc de montrer que $\mathbb{P}(\bigcap_{k \geq n} A_k^c)$ est nulle pour tout $n \in \mathbb{N}$ pour conclure. C'est donc ce à quoi nous allons nous atteler.

Par continuité décroissante puis indépendance, on a que

$$\mathbb{P}\left(\bigcap_{k \geq n} A_k^c\right) = \lim_{N \rightarrow +\infty} \mathbb{P}\left(\bigcap_{N \geq k \geq n} A_k^c\right) = \lim_{N \rightarrow +\infty} \prod_{k=n}^N \mathbb{P}(A_k^c).$$

Or, on a que

$$\ln\left(\prod_{k=n}^N \mathbb{P}(A_k^c)\right) = \sum_{k=n}^N \ln(1 - \mathbb{P}(A_k)) \leq - \sum_{n \leq k \leq N} \mathbb{P}(A_k) \xrightarrow{n \rightarrow \infty} -\infty,$$

dont on peut déduire que $\lim_{N \rightarrow +\infty} \prod_{k=n}^N \mathbb{P}(A_k^c) = 0$.

On peut donc conclure que sous les hypothèses, $\mathbb{P}(A_n \text{ arrive une infinité de fois}) = 1$. □

Ce lemme permet par exemple d'étudier la récurrence ou la transience d'une marche aléatoire, et donc de répondre à cette question essentielle : Pourquoi un marcheur ivre finira-t-il toujours par retrouver sa maison, alors qu'un oiseau désorienté a des chances d'être perdu à jamais ?

1.4 Exercices

Exercice 1 :

On considère un ensemble fini à 4 éléments $E = \{a_1, a_2, a_3, a_4\}$ muni de tous les événements ($\mathcal{P}(E)$) muni d'une loi de probabilité \mathbb{P} . Calculer la probabilité de l'événement élémentaire $\{a_1\}$ dans les cas suivant :

1. $\mathbb{P}(a_2) = \frac{1}{4} \quad \mathbb{P}(a_3) = \frac{1}{6} \quad \mathbb{P}(a_4) = \frac{1}{5};$
2. $\mathbb{P}(a_1) = 3\mathbb{P}(a_2) \quad \mathbb{P}(a_3) = \frac{1}{4} = \mathbb{P}(a_4);$
3. $\mathbb{P}(\{a_2, a_3, a_4\}) = 2\mathbb{P}(a_1);$
4. $\mathbb{P}(a_1) = \mathbb{P}(a_2) \quad \mathbb{P}(a_3) = 2\mathbb{P}(a_2) \quad \mathbb{P}(a_4) = 3\mathbb{P}(a_3).$

Exercice 2 :

On considère un ensemble muni d'une loi de probabilité \mathbb{P} et deux événements A et B tels que

$$\mathbb{P}(A) = \frac{1}{2} \quad \mathbb{P}(A \cup B) = \frac{3}{4} \quad \mathbb{P}({}^c B) = \frac{5}{8}.$$

Calculer les probabilités $\mathbb{P}(A \cap B)$, $\mathbb{P}({}^c A \cap {}^c B)$, $\mathbb{P}({}^c A \cup {}^c B)$ et $\mathbb{P}(A \cap {}^c B)$.

Exercice 3 :

Au bridge, une main est constituée de $n = 13$ cartes prises dans un jeu de $N = 52$ cartes. On suppose que toutes les mains sont équiprobables. Quelle est la probabilité qu'un joueur de bridge ait une main contenant :

1. un as exactement ?
2. Au moins un as ?
3. Un as et un roi exactement ?
4. Au moins un as et au moins un roi ?
5. Une carte de chaque valeur ?

Exercice 4 : *Statistique de Maxwell-Boltzmann*

On modélise la statistique de Maxwell-Boltzmann de la manière suivante :

On répartit n objets différents dans N boîtes, et on suppose que toutes ces répartitions sont équiprobables.

Plus formellement, on considère l'univers \mathcal{E}_{MB} des suites ordonnées de n éléments (**avec répétition possible**), à valeurs dans l'ensemble $\{1, \dots, N\}$, munie de la tribu maximale et de la loi uniforme. Une telle suite correspond à noter successivement les numéros de la boîte dans lesquelles vont les objets. Par exemple, pour $n=2$ et $N=3$, on a comme univers :

$$\mathcal{E}_{MB} = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)\},$$

et comme probabilité :

$$\forall (i, j) \in \{1, 2, 3\}^2, \mathbb{P}(\{(i, j)\}) = \frac{1}{9}.$$

Déterminer la probabilité d'avoir exactement k objet dans une boîte fixée, en fonction de n , N et k .

Exercice 5 : *Statistique de Fermi-Dirac*

On modélise la statistique de Fermi-Dirac de la manière suivante :

On répartit n objets différents dans N boîtes, et on suppose que toutes les boîtes contiennent au plus une boîte. On suppose alors que toutes les répartitions valides sont équiprobables.

Plus formellement, on considère l'univers \mathcal{E}_{FD} des suites ordonnées de n éléments (**sans répétition** de valeur possible), à valeurs dans l'ensemble $\{1, \dots, N\}$, munie de la tribu maximale et de la loi uniforme. Une telle suite correspond à noter dans la position k le numéro de la boîte dans laquelle va le $k^{ième}$ objet. Par exemple, pour $n=2$ et $N=3$, on a comme univers :

$$\mathcal{E}_{FD} = \{(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)\},$$

munie de la loi uniforme.

Déterminer la probabilité d'avoir un objet dans une boîte fixée, en fonction de n et N .

Exercice 6 : *Statistique de Bose-Einstein*

On modélise la statistique de Bose-Einstein de la manière suivante :

On répartit n objets différents dans N boîtes, et on suppose alors que toutes les répartitions de nombre de boules par boîtes valides sont équiprobables.

Plus formellement, on considère l'univers \mathcal{E}_{BE} des suites ordonnées de N éléments (**avec répétition** de valeur possible) de somme n , munie de la tribu maximale et de la loi uniforme. Une telle suite correspond à noter dans la position k le nombre d'objets dans la boîte en position k . Par exemple, pour $n=2$ et $N=3$, on a comme univers :

$$\mathcal{E}_{BE} = \{(0, 0, 2), (0, 1, 1), (0, 2, 0), (1, 0, 1), (1, 1, 0), (2, 0, 0)\},$$

munie de la loi uniforme

Exprimer la probabilité d'avoir $R = (r_1, \dots, r_N)$ objets respectivement dans les boîtes $(1, \dots, N)$, en fonction de n , N et R .

Exercice 7 : *Probabilité géométrique*

On considère le polynôme de degré 2 défini par $P(X) = X^2 + aX + b$ où a et b sont tirées suivant des lois uniformes et indépendantes sur $[-1; 1]$ (i.e. la probabilité d'un événement est égale à son aire divisée par 4).

- Quelle est la probabilité que P n'ait pas de racine réelle ?
- Quelle est la probabilité que P n'ait pas de racine réelle et que $a > 0$?

Ces questions peuvent être reliées aux questions de savoir quand est-ce que les solutions d'une équation différentielle de degré 2 soient oscillantes ou oscillantes amorties.

Exercice 8 : *Probabilité géométrique-2*

On considère que x et y sont tirés suivant des lois uniformes et indépendantes sur $[0; 1]$.

Quelle est la probabilité que la somme de x et y ne dépasse pas 1 et que leur produit fasse au plus $\frac{2}{9}$?

Exercice 9 : *Probabilité géométrique-3*

On considère que x et y sont tirés suivant des lois uniformes et indépendantes sur $[-n; n]$.

On note \underline{x} et \underline{y} les parties fractionnaires de x et y .

- Quelle est la probabilité que $\underline{x} \leq 2$?
- Quelle est la probabilité que $\underline{x} \leq 2$ et $\underline{y} \leq 2$?
- Que deviennent ces probabilités quand n tend vers l'infini ?

Exercice 10 : *Une indépendance théorique*

Montrer que deux événements A et B sont indépendants si et seulement si :

$$\mathbb{P}(A \cap B)\mathbb{P}(\overline{A} \cap \overline{B}) = \mathbb{P}(A \cap \overline{B})\mathbb{P}(\overline{A} \cap B)$$

Exercice 11 : *Un peu d'indépendance*

On jette deux dés à six faces bien équilibrés (comme les dés sont différents, les événements dépendant uniquement du premier dé sont indépendants de ceux qui résultent du deuxième dé). On appelle d_i le nombre de points apparaissant sur le $i^{\text{ème}}$ dé. Soient E_1 , E_2 et E_3 les événements :

- $E_1 = "d_1 + d_2 \text{ est pair} "$;
- $E_2 = "d_1 = 6 \text{ et } d_2 \text{ est pair} "$;
- $E_3 = "d_1 + d_2 \text{ est divisible par } 4."$

Quels couples d'événements parmi E_1 , E_2 et E_3 sont indépendants ?

Exercice 12 : *Un peu plus d'indépendance*

On considère un groupe de n personnes. On suppose que ce groupe est suffisamment hétéroclite pour que la probabilité qu'une personne dans le groupe soit de sexe masculin soit de $\frac{1}{2}$.

- On considère $A = " \text{le groupe comporte au plus un garçon} "$ et $B = " \text{le groupe comporte au moins un garçon et au moins une fille} "$. Pour quel(les) valeur(s) de n les événements A et B sont-ils indépendants ?
- Si l'on pose $C = " \text{le groupe contient au plus deux garçons} "$. Pour quelles valeurs de n a-t-on $\mathbb{P}(C) \geq 0.9$?

Exercice 13 : *Conditionnement*

On jette deux dés à 6 faces équilibrés. Trouver la probabilité que

- la somme des nombres obtenus sur les deux dés soit supérieur à 9 sachant que l'on a obtenu au moins un 6.
- l'on tire 4 sur l'un des dés sachant que l'on a tiré au moins un 2.
- la somme des nombres obtenus sur les deux dés vaille 5 sachant que la différence entre la plus grande et la plus petite de ces deux valeurs vaut 4.

Exercice 14 :

Soient trois portes, l'une cachant un trésor, les deux autres une chèvre, réparti de manière uniforme en probabilité. Vous choisissez une porte à ouvrir, quand un étranger connaissant la répartition ouvre devant vous une des deux autres portes, en choisissant sciemment de révéler une chèvre (en tirant une pièce pour choisir si besoin). Quelle est alors la probabilité de trouver le trésor en changeant de porte ?

Il s'agit d'une reformulation du problème de Monty Hall. L'autrice du problème et de sa résolution affirme avoir reçu plus de 10 000 lettres traitant du problème, dont plusieurs provenant d'universitaires remettant en question la pertinence de la démonstration.

Exercice 15 : *Anniversaire*

On considère un groupe de n personnes et on suppose que la date d'anniversaire d'une personne suit une loi uniforme sur 365 jours (personne n'est né le 29 février). Exprimer la probabilité qu'il y ait deux personnes du groupe qui a la même date d'anniversaire.

Avec un calculateur, trouver à partir de quelle valeur de n cette probabilité dépasse $\frac{1}{2}$.

Solution et indication page 80

Chapitre 2

Probabilité univarié : outils et exemples classiques

2.1 Trousse à outils du probabiliste

Lorsque dans la vie réelle, nous allons considérer des variables simples plutôt que l'ensembles des faits qui ont lieu dans l'univers.

Parmi les questions qui peuvent nous intéresser, on retrouve la taille d'individu, leur préférence de vote, la concentration chimique d'un mélange, le temps de fonctionnement d'un appareil électrique, le nombre d'appels reçu par le service client au cours d'un intervalle de temps, le débit de liquide en un point d'espace et de temps.

Ces questions reviennent à prendre le résultat d'une expérience aléatoire (développement physique ou politique d'une personne, expérience subit d'un appareil, évènements extérieurs conditionnant les appels, ensembles des variables d'états du liquide) et ne conserver que les paramètres qui nous intéressent.

2.1.1 Variables aléatoires

Une *variable aléatoire* est une fonction de l'expérience aléatoire réalisée. Sa valeur n'est connue qu'après réalisation de l'expérience, ce qui explique son nom. En revanche, la fonction correspondante est bien déterminée.

Une *variable aléatoire réelle* est une variable aléatoire à valeur réelle, c'est-à-dire une fonction de l'univers Ω dans \mathbb{R} , ou encore $f : \Omega \rightarrow \mathbb{R}$.

Il est naturel de vouloir répondre à des questions sur les probabilités qu'une variable aléatoire soit dans des domaines particuliers (par exemples des intervalles ou une valeur donnée). Cela conduit alors à la définition suivante :

Définition 15 :

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

*On appelle variable aléatoire réelle toute application $X : \Omega \rightarrow \mathbb{R}$ **mesurable** de (Ω, \mathcal{T}) dans $(\mathbb{R}, \mathcal{B}(\mathbb{R})) :=$ tribu engendrée par les ouverts), c'est-à-dire :*

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \subset \mathcal{T}$$

Remarque : En pratique, on ne vérifie que rarement la condition de mesurabilité. Si on doit le faire, il suffit de vérifier la propriété sur une sous-partie des boréliens $\mathcal{B}(\mathbb{R})$ qui les engendrent, comme les intervalles, ou encore les ensembles de la forme $] - \infty, a]$.

Convention : On notera les variables aléatoires avec des lettres majuscules (X, Y, Z, U, V, \dots) et leurs valeurs par des lettres minuscules correspondantes (e.g. $X = x$).

Si $B \in \mathcal{B}(\mathbb{R})$ est un borélien, on notera $\{X \in B\}$ l'événement $\{\omega \in \Omega, X(\omega) \in B\}$.

Proposition 16 :

L'ensemble des variables aléatoires réelles est stable par somme, produit, multiplication par un réel, par passage à la limite et composition par une fonction continue (i.e. si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est continue et si X_1, \dots, X_n sont des variables aléatoires réelles, alors $f(X_1, \dots, X_n)$ est aussi une variable aléatoire réelle).

Pour étudier une variable aléatoire, il est fondamental d'étudier sa loi :

Définition 17 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. On appelle loi de X que l'on note \mathbb{P}_X la mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par

$$\forall B \in \mathcal{B}(\mathbb{R}), \quad \mathbb{P}_X(B) := \mathbb{P}(\{X \in B\})$$

Exercice : Montrer qu'il s'agit bien d'une loi de probabilité.

Il est important de remarquer que la loi d'une variable aléatoire X est une mesure sur l'ensemble des valeurs de la variable aléatoire.

L'objet d'intérêt principal lorsqu'on étudie des variables aléatoires est sa loi.

En effet, elle permet de calculer directement les probabilités d'événement exprimé en fonction de la variable aléatoire.

La loi, c'est fondamental !

Exemple : Prenons l'exemple du lancé de deux dés de couleur différente, et X la variable aléatoire correspondant à faire la somme du résultat des deux dés :

$$X : \llbracket 1, 6 \rrbracket \times \llbracket 1, 6 \rrbracket \rightarrow \llbracket 2, 12 \rrbracket$$

En notant $p_n = \mathbb{P}_X(\{n\}) = \mathbb{P}(\{X = n\}) = \mathbb{P}(\{(i, j) \in \llbracket 1, 6 \rrbracket^2, i + j = n\})$ on trouve :

n	2	3	4	5	6	7	8	9	10	11	12
p_n	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Définition 18 :

On dira que des variables aléatoires $(X_i)_{i \in I}$ sont indépendantes si pour tout n , tous sous-ensembles finis de taille n $(X_{i_1}, \dots, X_{i_n})$ et pour tout boréliens $(B_1, \dots, B_n) \in \mathcal{B}(\mathbb{R})^n$,

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n)$$

Remarque : Cette notion sera manipulée dans les exercices.

2.1.2 Premier outil : la fonction de répartition

Dans sa besace des outils probabiliste, il y a de nombreux instruments. Un premier instrument est la fonction de répartition :

Définition 19 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire.

On appelle fonction de répartition de X , noté F ou F_X , l'application :

$$F : \begin{cases} \mathbb{R} & \rightarrow & [0, 1] \\ t & \mapsto & \mathbb{P}_X(]-\infty, t]) = \mathbb{P}(X \leq t). \end{cases}$$

Cette fonction de répartition a des propriétés bien utiles :

Proposition 20 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. On note F la fonction de répartition de X , alors :

- Si $t \leq t'$, alors $F(t) \leq F(t')$;
- F est continu à droite en tout point ;
- F admet une limite finie à gauche en tout point ;
- $\lim_{t \rightarrow -\infty} F(t) = 0$ et $\lim_{t \rightarrow +\infty} F(t) = 1$.

Démonstration.

- Soit $t \leq t'$, alors comme $]-\infty, t] \subset]-\infty, t']$, on a par croissance que $\mathbb{P}(]-\infty, t]) \leq \mathbb{P}(]-\infty, t'])$ et donc que $F(t) \leq F(t')$.
- Soit $t \in \mathbb{R}$ et $(t_n)_{n \in \mathbb{N}}$ une suite décroissante de réel strictement supérieur à t , et de limite t . Alors $(]-\infty, t_n])_{n \in \mathbb{N}}$ est une suite décroissante d'ensemble, et donc par continuité décroissante des mesures de probabilité, $F(t) = \mathbb{P}(X \leq t) = \mathbb{P}(\bigcap_{n \in \mathbb{N}} \{X \leq t_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(\{X \leq t_n\}) = \lim_{n \rightarrow \infty} F(t_n)$.

- Soit $t \in \mathbb{R}$ et $(t_n)_{n \in \mathbb{N}}$ une suite croissante de réel strictement inférieur à t , et de limite t . Alors $(]-\infty, t_n])_{n \in \mathbb{N}}$ est une suite croissante d'ensemble, et donc par continuité croissante des mesures de probabilité, $\mathbb{P}(X < t) = \mathbb{P}(\bigcup_{n \in \mathbb{N}} \{X \leq t_n\}) = \lim_{n \rightarrow \infty} \mathbb{P}(\{X \leq t_n\}) = \lim_{n \rightarrow \infty} F(t_n)$.
- Comme $0 = \mathbb{P}(\emptyset) = \lim_{n \in \mathbb{N}, n \rightarrow +\infty} \mathbb{P}(]-\infty, -n]) = \lim_{n \in \mathbb{N}, n \rightarrow +\infty} F(-n)$, ce qui permet de conclure avec le premier point que $\lim_{t \rightarrow -\infty} F(t) = 0$

Avec un raisonnement similaire avec $\mathbb{R} = \bigcup_{n \in \mathbb{N}}]-\infty, n]$, on trouve que $\lim_{t \rightarrow +\infty} F(t) = 1$

□

Culture générale : Une question naturelle est celle de la réciproque : si on se donne une fonction $F : \mathbb{R} \rightarrow [0, 1]$, peut-on trouver une variable aléatoire réelle pour laquelle F est exactement la fonction de répartition ? La réponse est oui, et repose sur la construction de mesures via les mesures extérieures, et permet de donner des exemples explicites de mesure de probabilité diffuse (sans points de poids non nul) mais qui n'a pas de partie absolument continue (*i.e.* ne ressemble pas à une mesure à densité par rapport à la mesure de Lebesgue). Un tel exemple peut ainsi être trouvé en prenant comme fonction de répartition l'escalier du diable, une fonction continue, dérivable Lebesgue-presque partout, de dérivée nulle, et de limite 0 et 1 respectivement en 0 et 1.

La fonction de répartition permet de retrouver la loi :

Proposition 21 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle. On note F la fonction de répartition de X , alors pour $(a, b, x_0) \in \mathbb{R}^3$, on a :

$$\mathbb{P}(X \in]a, b]) = F(b) - F(a) \quad \mathbb{P}(X = x_0) = F(x_0^+) - F(x_0^-)$$

La loi se répartit comme prescrit par la fonction de répartition !

2.1.3 Deuxième outil : l'espérance et les moments

Un deuxième outil que le probabiliste doit être prêt à sortir, même s'il faut le manipuler avec prudence, est l'espérance et sa généralisation, les moments.

Définition 22 :

On dit qu'une variable aléatoire réelle X sur $(\Omega, \mathcal{T}, \mathbb{P})$ est intégrable, ou admet un moment d'ordre 1 si $X \in \mathcal{L}^1(\Omega, \mathcal{T}, \mathbb{P})$ est intégrable au sens de Lebesgue.

On définit alors son espérance, que l'on note $\mathbb{E}(X)$ ou $\mathbb{E}_{\mathbb{P}}(X)$ comme l'intégrale au sens de Lebesgue de X :

$$\mathbb{E}(X) := \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} x d\mathbb{P}_X(x)$$

L'espérance ne dépend que de la loi !

Remarque : Pour les allergiques à l'intégrale de Lebesgue, on peut se ramener à une intégrale généralisée de Riemann. En effet, quitte à considérer les parties positives et négatives de la variable, on peut supposer qu'elle est à valeur positive. Alors, avec le lemme ci-après :

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}_+} \mathbb{P}(X > t) d\lambda(t) = \int_{\mathbb{R}_+} (1 - F_X(t)) d\lambda(t)$$

Or comme $1 - F_X(t)$ est continu à droite et admet une limite à gauche en tout point, elle est Riemann-intégrable sur les compacts, et on peut donc voir le terme de droite comme une intégrale de Riemann.

Lemme : Soit (E, \mathcal{E}, μ) un espace mesuré (σ -fini) et $f : E \rightarrow [0, +\infty]$ mesurable, Alors

$$\int_E f d\mu = \int_{\mathbb{R}_+} \mu(\{f > x\}) d\lambda(x)$$

Démonstration. Appliquer le théorème de Tonelli à l'indicatrice 1_C où $C = \{(x, a) \in E \times \mathbb{R}_+, f(x) > a\}$ sur $(E \times \mathbb{R}_+, \mathcal{E} \otimes \mathcal{B}(\mathbb{R}), \mu \times \lambda_{Lebesgue})$. \square

Maintenant, il s'agit de savoir manipuler cette espérance. Comme il s'agit d'une intégrale de masse totale de la mesure finie, on a immédiatement les propriétés suivantes :

Proposition 23 :

L'espérance vérifie que :

- Si X est constante ou constante presque sûrement à un réel b (*i.e.* $\mathbb{P}(X = b) = 1$), alors $\mathbb{E}(X) = \mathbb{E}(b) = b$;
- Pour A événement, $X = 1_A$, on a $\mathbb{E}(1_A) = \mathbb{P}(A)$;
- Pour (X, Y) deux variables aléatoires réelles intégrables et (a, b) deux réels, on a que

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Pour avoir plus d'information sur une variable aléatoire, il est courant de regarder les quantités suivantes :

Définition 24 :

Plus généralement, on dira que X admet un moment d'ordre p si X^p est intégrable.

On appelle alors $\mathbb{E}(X^p)$ le moment d'ordre p de X .

Pour X intégrable, on dira que X est centré si $\mathbb{E}(X) = 0$.

Si X a un moment d'ordre p , on appellera $\mathbb{E}[(X - \mathbb{E}(X))^p]$ le moment centré d'ordre p de X .

Un cas particulièrement important est celui des moments d'ordre 2 :

Définition 25 :

Pour X de carré intégrable, on appelle variance (de X) le moment centré d'ordre 2 noté V , et écart-type la racine carrée de la variance, noté σ .

$$V(X) = \sigma^2 = \mathbb{E} \left[(X - \mathbb{E}(X))^2 \right]$$

On dira que la variable est réduite si $\sigma = 1$.

Proposition 26 :

Soient X et Y deux variables aléatoires réelles de carré intégrable, a et b deux réels. Les propriétés suivantes sont vérifiées :

- $V(aX + b) = a^2 V(X)$.
- (Formule de König-Huygens) $V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
- (Résultat admis) Si X et Y sont indépendantes, $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
- Si X et Y sont indépendantes, on a : $V(X + Y) = V(X) + V(Y)$
- Si l'espérance de X vaut m et son écart type σ , la variable $\frac{X-m}{\sigma}$ est centré-réduite

Démonstration.

- Par linéarité de l'espérance et la valeur de l'espérance des variables constantes,

$$\begin{aligned} V(aX + b) &= \mathbb{E} \left[(aX + b - \mathbb{E}(aX + b))^2 \right] \\ &= \mathbb{E} \left[(aX - \mathbb{E}(aX) + b - b)^2 \right] \\ &= \mathbb{E} \left[a^2 (X - \mathbb{E}(X))^2 \right] = a^2 V(X). \end{aligned}$$

•

$$\begin{aligned} V(X) &= \mathbb{E} \left[(X - \mathbb{E}(X))^2 \right] = \mathbb{E} \left[X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2 \right] \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \end{aligned}$$

- Soit X et Y indépendantes (hypothèse utilisée de la ligne 4 à la ligne 5), on a :

$$\begin{aligned}
 V(X + Y) &= \mathbb{E} \left[(X + Y - \mathbb{E}(X + Y))^2 \right], \\
 &= \mathbb{E} \left[(X - \mathbb{E}(X) + Y - \mathbb{E}(Y))^2 \right], \\
 &= \mathbb{E} \left[(X - \mathbb{E}(X))^2 + 2(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) + (Y - \mathbb{E}(Y))^2 \right], \\
 &= V(X) + 2\mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] + V(Y), \\
 &= V(X) + 2\mathbb{E}[X - \mathbb{E}(X)] \mathbb{E}[Y - \mathbb{E}(Y)] + V(Y), \\
 &= V(X) + 0 + V(Y).
 \end{aligned}$$

- On applique la linéarité de l'espérance pour le caractère centré et le premier point de la présente proposition pour le côté réduit.

□

2.1.4 Un troisième outil : la fonction génératrice des moments (H.P.)

Pour réduire tous ces calculs coûteux, il est parfois possible de s'appuyer sur la fonction génératrice des moments :

Définition 27 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle.

La fonction génératrice des moments d'une variable aléatoire X est la fonction M_X définie par :

$$M_X(t) = \mathbb{E}(e^{tX}),$$

pour tout réel t tel que cette espérance existe.

Cette fonction, comme son nom l'indique, est utilisée afin de calculer les moments associés à la loi de X . Pour cela, nous utiliserons le :

Théoreme 28 :

Soit X une variable aléatoire réelle et M_X sa fonction génératrice des moments. Les deux assertions suivantes sont équivalentes :

1. Il existe $\delta > 0$ tel que $M_X(t) < +\infty$ pour tout $t \in]-\delta, \delta[$.
2. La variable X admet des moments de tout ordre finis et la série $\sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$ a un rayon de convergence non nul $R > 0$.

De plus, si l'une des deux assertions ci-dessus est vérifiée, alors

- Pour tout $t \in]-R, R[$ on a $M_X(t) = \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$.
- Pour tout $k \geq 0$, M_X est k fois dérivable en 0 et $M_X^{(k)}(0) = \mathbb{E}(X^k)$.

Démonstration.

Montrons que 1 implique 2 :

En remarquant que pour $0 \leq t < \delta$ et pour tout réel x , on a que $e^{t|x|} \leq e^{tx} + e^{-tx}$, on en déduit en intégrant que

$$\mathbb{E}(e^{t|X|}) \leq M_X(t) + M_X(-t) < +\infty$$

On en déduit avec le théorème de Tonelli que la série $\sum_{k=0}^n \frac{\mathbb{E}(|X|^k)}{k!} t^k = \mathbb{E}(e^{t|X|}) < +\infty$.

On en conclut que la série entière $\sum_{k=0}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k$ est absolument convergente sur $[0, \delta[$ donc possède un rayon de convergence $R \geq \delta$.

Montrons maintenant que 2 implique 1 :

Pour cela, montrons que la série entière $\sum_{k=0}^{\infty} \frac{\mathbb{E}[|X|^k]}{k!} t^k$ a un rayon de convergence supérieur à $\min(R, 1)$.

Remarquons que pour tout entier k ,

$$\mathbb{E}[|X|^{2k+1}] = \mathbb{E}[|X|^{2k+1} \mathbf{1}_{X < 1}] + \mathbb{E}[|X|^{2k+1} \mathbf{1}_{X > 1}] \leq 1 + \mathbb{E}[|X|^{2k+2}]$$

Alors, pour $0 \leq t < \min(R, 1)$,

$$\sum_{k=0}^{\infty} \frac{\mathbb{E}[|X|^k]}{k!} t^k \leq e^t + 2 \sum_{k=0}^{\infty} \frac{\mathbb{E}[X^{2k}]}{(2k)!} t^{2k} \leq e^t + 2 \sum_{k=0}^{\infty} \frac{|\mathbb{E}[X^k]|}{k!} t^k < +\infty$$

où la dernière somme converge, car qu'une série entière est absolument convergente dans l'intérieur de son disque de convergence.

Ainsi en utilisant le théorème de Fubini-Tonelli, on obtient que $M_X(t) < +\infty$ et que $M_X(-t) < +\infty$, ce qui permet de conclure.

Les deux points suivants sont des résultats classiques de la théorie des séries entières. □

Remarque supplémentaire : $M_X(-t)$ est une extension de la transformée de Laplace.

Exercice : Sous les hypothèses du théorème, donner l'expression de la variance en fonction de la fonction génératrice et de ses dérivées.

Exercice : Soit (X_1, \dots, X_N) des variables aléatoires réelles indépendantes dans leur ensemble ayant chacune une fonction génératrice de moments globalement définie. Posons $S = \sum_{i=1}^N X_i$. Montrer que la fonction génératrice de moments de S est le produit des fonctions génératrices des X_i .

2.1.5 Théorème de transfert et autres outils

Le théorème de transfert permet d'exprimer les diverses quantités d'intérêt d'une variable aléatoire introduites précédemment uniquement en fonction de la loi :

Théorème 29 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire de loi \mathbb{P}_X , et $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ une application réelle mesurable.

Alors Φ est \mathbb{P}_X intégrable si et seulement si $\Phi \circ X$ est \mathbb{P} intégrable. En cas d'intégrabilité, on a

$$\int_{\Omega} \Phi(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}} \phi(t) d\mathbb{P}_X(t)$$

Il reste encore de nombreux autres outils dans la besace du probabiliste. Parmi les plus célèbres, on peut citer la fonction caractéristique, qui sera développé dans la section 4.1.1, à la page 57.

2.1.6 Exercices

Exercice 1 :

Soit X la variable aléatoire de loi définie par :

$$\begin{aligned} \mathbb{P}(X = k) &= ak(8 - k) & k \in \{0, 1, \dots, 8\} \\ \mathbb{P}(X = k) &= 0 & k \notin \{0, 1, \dots, 8\} \end{aligned}$$

Calculer a , la moyenne, la variance et l'écart type de X .

Indication : si on note $S_k(n) = \sum_{p=0}^n p^k$, alors on a que :

$$S_1(n) = \frac{n(n+1)}{2}, \quad S_2(n) = \frac{n(n+1)(2n+1)}{6}, \quad S_3(n) = S_1(n)^2, \quad S_4(n) = \frac{n(n+1)(2n+1)(3n^2+3n-1)}{30}$$

Exercice 2 : *Loi de Rayleigh*

Soit $a > 0$, et soit une variable aléatoire de fonction de répartition

$$\begin{aligned} F(x) &= 1 - \exp^{-\frac{x^2}{2a^2}} & \text{pour } x \geq 0 \\ &= 0 & x < 0. \end{aligned}$$

On appelle la loi correspondante la loi de Rayleigh de paramètre a . Calculer (si elles existent) sa densité, sa moyenne, sa variance, sa médiane (point λ tel que $\mathbb{P}(X < \lambda) = \frac{1}{2}$) et son mode (maximum de la densité).

Exercice 3 :

Utiliser la fonction génératrice, vu comme un polynôme en e^t , pour montrer qu'il est impossible de truquer deux dés 6 (potentiellement de manières différentes) de telle sorte que la variable aléatoire qui renvoie la somme des résultats des deux dés suive une loi uniforme sur $\llbracket 2, 12 \rrbracket$.

Solution et indication page 87

2.2 Les variables aléatoires discrètes

2.2.1 Définition et forme générale

Définition 30 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire

Une variable aléatoire réelle est discrète si et seulement si $X(\Omega)$, l'ensemble des valeurs prises par X , est (presque sûrement) fini ou dénombrable.

Exemple : Le résultat de lancer de deux dés, le nombre de composants qui a mal fonctionné sur un intervalle de temps, ou encore le nombre d'essais nécessaire à faire démarrer un processus.

Proposition 31 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et soit X une variable aléatoire discrète, alors sa loi est entièrement déterminée par :

$$\{(x_k, p_k)_{k \in J} | J \subset \mathbb{N}, \text{ et } \mathbb{P}(X = x_k) = p_k\}$$

Démonstration. Voir la section 1.2.2

□

Remarque : Pour faire des calculs explicites, on notera qu'avec les notations précédentes, on a :

$$\mathbb{E}(X) := \sum_{k \in J} x_k p_k, \quad \mathbb{E}(X^r) := \sum_{k \in J} x_k^r p_k$$

2.2.2 La loi uniforme

Reprenons la première probabilité que nous avons vu : la probabilité *uniforme*. Rappelons que :

Définition 32 :

Soit $(a, b) \in \mathbb{N}^2$ avec $a \leq b$. On pose $n = b - a$.

On dit que la variable aléatoire X suit la loi uniforme, que l'on note $X \sim \mathcal{U}([1, n])$ si X est (presque sûrement) à valeur dans $\{1, \dots, n\}$, et que pour tout $k \in \{1, \dots, n\}$,

$$\mathbb{P}(X = k) = \frac{1}{b - a} = \frac{1}{n}$$

Il s'agit d'un modèle associé au jet d'un dé bien équilibré à n faces. Il sera utilisé à utiliser pour modéliser un tirage au hasard sur un ensemble de cardinal fini n , lorsqu'on n'a pas de raison de préférer un résultat, ou en première approximation s'il n'y a aucune information supplémentaire.

Proposition 33 :

Soit $X \sim \mathcal{U}([1, n])$, alors

$$\mathbb{E}(X) = \frac{n+1}{2}, \quad \text{et} \quad V(X) = \frac{n^2-1}{12}$$

Démonstration. Comme il s'agit la première preuve de ce type, nous allons plus détailler que dans la suite.

Par définition, on a que :

$$\mathbb{E}(X) = \sum_{k=1}^n k \mathbb{P}(X = k) = \sum_{k=1}^n k \frac{1}{n}$$

Maintenant, on peut remarquer

$$\begin{aligned} \sum_{k=1}^n k &= \sum_{k=1}^n \frac{2}{2} k = \frac{1}{2} \sum_{k=1}^n (k + k) = \frac{1}{2} \left(\sum_{k=1}^n k + \sum_{k=1}^n k \right) \\ &= \frac{1}{2} \left(\sum_{k=1}^n k + \sum_{k=1}^n (n - k + 1) \right) = \frac{1}{2} \left(\sum_{k=1}^n k + n - k + 1 \right) \\ &= \frac{1}{2} \left(\sum_{k=1}^n n + 1 \right) = \frac{n(n+1)}{2}. \end{aligned}$$

On peut alors conclure que $\mathbb{E}(X) = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$

Reformulation De manière plus visuelle, nous avons fait le calcul suivant :

$$\begin{array}{cccccccc|c} 1 & + & 2 & + & \cdots & + & n-1 & + & n & | & \sum_{k=1}^n k \\ n & + & n-1 & + & \cdots & + & 2 & + & 1 & | & \sum_{k=1}^n k \\ \hline & & & & \cdots & & & & & & \\ n+1 & + & n+1 & + & \cdots & + & n+1 & + & n+1 & | & n(n+1) = 2 \sum_{k=1}^n k \end{array}$$

Calculons alors la variance. Pour cela, nous aurons besoin de calculer la somme des carrés des n premiers entiers, ce que l'on peut faire en faisant apparaître un télescopage, et en réutilisant le résultat précédent :

$$\begin{aligned} \sum_{k=1}^n k^2 &= \sum_{k=1}^n \frac{(k+1)^3 - k^3 - 3k - 1}{3} = \frac{1}{3} \left((n+1)^3 - 1 - \frac{3n(n+1)}{2} - n \right) \\ &= \frac{n(n+1)(2n+1)}{6}. \end{aligned}$$

Alors :

$$\mathbb{E}(X^2) = \sum_{k=1}^n k^2 \mathbb{P}(X = k) = \sum_{k=1}^n k^2 \frac{1}{n} = \frac{n(n+1)(2n+1)}{6} \frac{1}{n} = \frac{(n+1)(2n+1)}{6}$$

Ce qui nous permet de conclure avec la formule de Koëning-Huygens :

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2} \right)^2 = \frac{n^2 - 1}{12}$$

□

Proposition 34 :

Soit $X \sim \mathcal{U}([1, n])$, alors la fonction génératrice de moments pour $t \neq 0$ vaut $M_X(t) = \frac{e^t}{n} \frac{e^{tn} - 1}{e^t - 1}$

Démonstration. Par définition, pour $t \in \mathbb{R} \setminus \{0\}$, $M_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=1}^n e^{tk} \frac{1}{n} = \frac{e^t}{n} \frac{e^{tn} - 1}{e^t - 1}$

□

2.2.3 La loi de Bernoulli et les binomiales

La loi de Bernoulli est un modèle qui s'applique à l'expérience aléatoire qui consiste à jeter une pièce de monnaie et à regarder si le résultat est pile ou face. Elle modélisera des expériences avec uniquement deux issues possibles : un succès et un échec, et on pose alors X (succès) = 1 et X (échec) = 0.

Définition 35 :

On dit que la variable aléatoire réelle X suit la loi de Bernoulli de paramètre p , noté $X \sim \mathcal{B}(p)$, si X ne prend que deux valeurs 1 ou 0 avec les probabilités respectives p et $q=1-p$.

Proposition 36 :

Si $X \sim \mathcal{B}(p)$, alors

$$\mathbb{E}(X) = p, \quad V(X) = p(1-p), \quad \text{et} \quad \forall t \in \mathbb{R}, M_X(t) = 1 + p(e^t - 1)$$

Démonstration. Il est immédiat que $\mathbb{E}(X) = p$ et que $\mathbb{E}(X^2) = p$. Donc $V(X) = p - p^2 = p(1-p)$.

Enfin, pour $t \in \mathbb{R}$, $M_X(t) = pe^t + (1-p) = 1 - p + pe^t$. □

Supposons que l'on répète maintenant n fois l'expérience pile ou face. On note X le nombre de piles obtenues. Alors, on peut modéliser X par une loi binomiale. Une variable aléatoire binomiale représente le nombre de succès dans une suite de n expériences identiques et indépendantes.

Définition 37 :

On dit que la variable aléatoire réelle X suit la loi Binomiale de paramètre n et p , noté $X \sim \text{Bin}(n, p)$, si X a pour valeur dans $\{0, \dots, n\}$ et

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

En particulier, X suit la même loi qu'une somme de n Bernoulli indépendantes de paramètre p .

Proposition 38 :

Si $X \sim \text{Bin}(n, p)$, alors

$$\mathbb{E}(X) = np, \quad V(X) = np(1-p), \quad \text{et} \quad \forall t \in \mathbb{R}, M_X(t) = (1 + p(e^t - 1))^n$$

Démonstration. Soit $(Y_i)_{i \in \{1, \dots, n\}}$ des expériences de Bernoulli indépendantes, alors $X \sim \sum Y_i$.

Par linéarité de l'espérance, $\mathbb{E}(X) = np$ et comme les Y_i sont indépendantes, $V(X) = np(1-p)$, et pour $t \in \mathbb{R}$, $M_X(t) = (1 - p + pe^t)^n$ (en utilisant le résultat de l'exercice sur les fonctions génératrices). □

2.2.4 La loi géométrique

Pour modéliser le nombre de répétitions d'une expérience nécessaire à obtenir un succès, on pourra utiliser la loi géométrique. Elle consiste à répéter de manière indépendante une épreuve de Bernoulli, dont la probabilité de succès est p et à noter X le rang du premier succès. Alors X est à valeurs dans $\{1, 2, \dots\} = \mathbb{N}^*$.

Définition 39 :

Soit $p \in]0, 1]$ On dit que la variable aléatoire réelle X suit la loi géométrique de paramètre p , noté $X \sim \mathcal{G}(p)$, si X est à valeur dans $\{1, 2, \dots\} = \mathbb{N}^*$ et

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

En particulier, X suit la même loi que le rang de premier succès dans une suite de Bernoulli indépendantes de paramètre p .

Proposition 40 :

Si $X \sim \mathcal{G}(p)$, alors

$$\mathbb{E}(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2}, \quad \text{et} \quad \forall t < -\ln(1-p), \quad M_X(t) = \frac{pe^t}{1 - (1-p)e^t}$$

Démonstration.

On calcule directement : $\mathbb{E}(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = -p \frac{d}{dp} \sum_{k=0}^{\infty} (1-p)^k = -p \frac{d}{dp} \frac{1}{p} = \frac{1}{p}$ Et avec le même type de manipulation, $\mathbb{E}(X^2) = \frac{2}{p^2} - \frac{1}{p}$, ce qui permet de conclure pour la variance.

Enfin, pour $t < -\ln(1-p)$, on a que $M_X(t) = \sum_{k=1}^{\infty} e^{tk} p(1-p)^{k-1} = pe^t \sum_{k=1}^{\infty} (e^t(1-p))^{k-1} = \frac{pe^t}{1 - (1-p)e^t}$ \square

Nous verrons en exercice que la loi géométrique est sans mémoire, et n'est donc pas pertinente dans une expérience aléatoire mettant en jeu de l'usure des pièces.

2.2.5 La loi de Poisson

On souhaite modéliser le nombre d'événements se produisant durant un temps T , avec une probabilité d'obtenir de grand nombre d'événements qui décroît rapidement. Ce peut être :

- le nombre d'arrivées à un péage d'autoroute ;
- le nombre de pannes d'un système ;
- le nombre d'appels à un standard téléphonique.

La loi que nous retenons est la loi de Poisson, aussi appelée « loi des événements rares »

Définition 41 :

Soit $\lambda > 0$.

On dit que la variable aléatoire réelle X suit la loi de Poisson de paramètre λ , noté $X \sim \mathcal{P}(\lambda)$, si X a pour valeur dans \mathbb{N} et

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

Proposition 42 :

Si $X \sim \mathcal{P}(\lambda)$, alors

$$\mathbb{E}(X) = \lambda, \quad V(X) = \lambda, \quad \text{et} \quad \forall t \in \mathbb{R}, \quad M_X(t) = \exp(\lambda(e^t - 1))$$

Démonstration.

On calcule directement :

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \\ \mathbb{E}(X^2) &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} (k(k-1) + k) e^{-\lambda} \frac{\lambda^k}{k!} = \lambda^2 + \lambda \end{aligned}$$

ce qui permet encore de conclure pour la variance.

Enfin, pour $t \in \mathbb{R}$, on a que $M_X(t) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \exp(\lambda e^t) = \exp(\lambda(e^t - 1))$

□

La propriété suivante permet de comprendre l'appellation de loi des événements rares : si on compte le nombre d'événements ayant une probabilité très faible d'arriver, on va se rapprocher d'une loi de Poisson (ce qui permet d'approximer un tel comportement). La preuve sera laissée en exercice

Proposition 43 : Approximation d'une loi binomiale par une loi de Poisson

Soient $\lambda > 0$ un réel, $Y \sim \mathcal{P}(\lambda)$ et $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles suivant la loi $\mathcal{Bin}(n, p_n)$ telles que $\lim_{n \rightarrow +\infty} np_n = \lambda$. Alors pour $k \in \mathbb{N}$ entier quelconque, on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(Y = k).$$

On dit alors que X_n converge en loi vers Y .

2.2.6 Exercices

Exercice 4 :

Recopiez le tableau suivant où k est entier, fermez le cours et essayer de le remplir (ce n'est pas grave si vous ne réussissez pas au premier essai).

Nom	Paramètres notation usuelle	$\mathbb{P}(X = k)$	$\mathbb{E}(X)$	$V(X)$	Fonction génératrice	Hypothèses Cas d'utilisation
$X \sim \text{uniforme}$						
$X \sim \text{Bernoulli}$						
$X \sim \text{Binomiale}$						
$X \sim \text{Géométrique}$						
$X \sim \text{Poisson}$						

Exercice 5 :

Soit m et n deux entiers, $\Omega_1 = \{1, m\}$ et $\Omega_2 = \{1, n\}$, et $\Omega = \Omega_1 \times \Omega_2$. On munit Ω de la loi uniforme, et on considère les variables aléatoires X_1 et X_2 définies pour tout couple $(\omega_1, \omega_2) \in \Omega$ par

$$X_1(\omega_1, \omega_2) = \omega_1,$$

$$X_2(\omega_1, \omega_2) = \omega_2.$$

Montrer que les variables aléatoires X_1 et X_2 sont indépendantes.

Exercice 6 :

Pour une valeur donnée de n , déterminez la valeur de p telle que la variance d'une loi binomiale $\text{Bin}(n, p)$ est maximale.

Exercice 7 :

Une usine produit des composants électroniques en grande quantité, avant de les regrouper dans des boîtes de mille composants. On observe que la probabilité qu'une boîte contienne au moins un composant défectueux est de 0.52. On relève aléatoirement 12 boîtes, et l'on pourra supposer en première approximation que l'état des composants sont indépendants (hypothèse fortement discutable en pratique). Soit X la variable aléatoire égale au nombre de composants défectueux dans la première boîte, et Y le nombre de boîtes contenant au moins un composant défectueux parmi les douze.

- Quelle est la loi suivie par les variables aléatoires X et Y ?
- Calculer la probabilité que 5 boîtes exactement parmi ces 12 contiennent un composant défectueux.
- Calculer l'espérance mathématique de la variable aléatoire Y , arrondi à l'entier le plus proche.
- Calculer la variance et l'écart-type de la variable aléatoire X .

Exercice 8 :

Soient $p \in]0, 1[$, et $X \sim \mathcal{G}(p)$. Montrer que X est sans mémoire, c'est-à-dire que pour $k \geq m$ deux entiers naturels,

$$\mathbb{P}(X > k | X > m) = \mathbb{P}(X > k - m).$$

En déduire, si on a déjà lancé une pièce équilibrée 18 fois et on a obtenu uniquement le côté face, la probabilité que l'on obtienne pile au bout de 24 lancers.

Exercice 9 :

Soit $X \sim \mathcal{P}(\lambda)$ une variable aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$.

Calculer

$$\mathbb{E}\left(\frac{1}{X+1}\right)$$

Exercice 10 :

Soit $X \sim \mathcal{G}(p)$ et $Y \sim \mathcal{P}(\lambda)$. On suppose que X et Y soient indépendantes. Calculer quand elle existe l'espérance de la variable aléatoire $Z = X^Y$

Exercice 11 :

Montrer la proposition 43

Solution et indication page 87

2.3 Les variables aléatoires absolument continue

2.3.1 Définition et forme générale

Définition 44 :

On dit que la loi de X (ou abusivement que X) est absolument continue s'il existe une fonction $f > 0$ intégrable pour la mesure de Lebesgue dx telle que :

$$\forall A \in \mathcal{B}(\mathbb{R}), \mathbb{P}_X(A) = \int_A f(x)dx = \int_{\mathbb{R}} f(x)\mathbb{1}_A(x)dx$$

On notera que cela implique que $\int_{\mathbb{R}} f(x)dx = 1$.

Dans ce cas, on dit que f est la densité de probabilité de X .

Pour les fonctions absolument continues, il est plus simple de calculer les quantités d'intérêt :

Proposition 45 :

Soit X une variable aléatoire absolument continue, de densité de probabilité f . Soient $(t, a, b) \in \mathbb{R}^3$ et $g \in C(\mathbb{R}, \mathbb{R})$

- *Fonction de répartition* : elle vérifie $F(t) = \int_{-\infty}^t f(x)dx$

En particulier, F est continu et $\mathbb{P}(X \in [a, b]) = \int_a^b f(t)dt$

Si en plus $f \in C(\mathbb{R}, \mathbb{R})$, alors F est dérivable et $F' = f$.

- *Calculs* : On a également $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x)dx$.

En particulier :

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx,$$

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x)dx,$$

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f(x)dx.$$

Remarque : On retrouve l'affirmation sur le fait que la fonction génératrice de moments est une extension de la transformée de Laplace. En effet, pour une fonction à densité, elle vaut exactement ladite transformée.

Culture générale : On a vu que la fonction de répartition est continue pour une variable *absolument continue*, mais est-ce que toutes les variables aléatoires ayant une fonction de répartition continue sont à densité? La réponse est non, ce qui peut par exemple se voir en prenant la fonction escalier du diable (une fonction continue, dérivable presque partout et de dérivée nulle, valant 0 en 0 et 1 en 1) comme fonction de répartition, et la mesure associée comme loi de variable aléatoire.

2.3.2 La loi uniforme

Si l'on veut maintenant modéliser un phénomène continu, avec l'idée de généraliser un résultat équiprobable, on choisira la loi uniforme. Elle correspond également au choix par défaut en l'absence d'informations supplémentaires.

Définition 46 :

On dit que la variable aléatoire réelle X suit la loi uniforme entre a et b , noté $X \sim \mathcal{U}([a, b])$, si la loi de X est absolument continue, de densité :

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

Proposition 47 :

Si $X \sim \mathcal{U}([a, b])$, alors

$$\mathbb{E}(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}, \quad \text{et} \quad \forall t \in \mathbb{R}, M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$$

Démonstration. On a comme X est à densité que :

$$\mathbb{E}(X) = \int_a^b \frac{1}{b-a} x dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{a+b}{2},$$

et

$$\mathbb{E}(X^2) = \int_a^b \frac{1}{b-a} x^2 dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{a^2 + ab + b^2}{3},$$

Ce qui permet également de conclure que $V(X) = \frac{(b-a)^2}{12}$.

Enfin, pour $t \in \mathbb{R}^*$,

$$M_X(t) = \int_a^b e^{tx} \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{e^{tx}}{t} \right]_a^b = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

□

2.3.3 La loi exponentielle et la loi Gamma

Pour décrire la fiabilité d'un système, on pourra utiliser une loi exponentielle. Elle sert à modéliser une durée de vie ou un temps de bon fonctionnement entre deux pannes (ou encore le taux d'émission d'un matériel radioactif).

Définition 48 :

Soit $\lambda > 0$.

On dit que la variable aléatoire réelle X suit la loi exponentielle de paramètre λ , noté $X \sim \mathcal{E}(\lambda)$, si la loi de X est absolument continue, de densité :

$$x \mapsto \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Proposition 49 :

Si $X \sim \mathcal{E}(\lambda)$, alors

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}, \quad \text{et} \quad \forall |t| < \lambda, M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

Démonstration. On a comme X est à densité, on a par intégration par partie (deux des termes sont bien convergents) que :

$$\mathbb{E}(X) = \int_0^{+\infty} x \lambda e^{-\lambda x} dx = [-x e^{-\lambda x}]_0^{+\infty} - \int_0^{+\infty} \lambda - e^{-\lambda x} dx = \frac{1}{\lambda},$$

et de même en intégrant par partie deux fois :

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 \lambda e^{-\lambda x} dx = \int_0^{+\infty} 2x - e^{-\lambda x} dx = \int_0^{+\infty} \frac{2}{\lambda} e^{-\lambda x} dx = \frac{2}{\lambda^2},$$

Ce qui permet également de conclure que $V(X) = \frac{1}{\lambda^2}$.

Enfin, pour $|t| < \lambda$,

$$M_X(t) = \int_0^{+\infty} \lambda e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}.$$

□

Comme pour la loi géométrique, la loi exponentielle est sans mémoire (on peut en fait montrer que c'est la seule loi à densité sans mémoire).

Proposition 50 :

Soient $\lambda > 0$, $X \sim \mathcal{E}(\lambda)$, et $(t_1, t_2) \in \mathbb{R}_+^2$ deux réels positifs alors

$$\mathbb{P}(X > t_1 + t_2 | X > t_1) = \mathbb{P}(X > t_2)$$

Dit autrement, la probabilité d'atteindre le temps $t_1 + t_2$ sachant qu'on a atteint le temps t_1 est la même que celle au départ d'atteindre le temps t_2 .

Démonstration. On commence par calculer $\mathbb{P}(X > t)$ pour $t > 0$:

$$\mathbb{P}(X > t) = \int_t^{+\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_t^{+\infty} = e^{-\lambda t}.$$

Maintenant, en utilisant la définition des probabilités conditionnelle et le fait que $\{X > t_1 + t_2\} \cap \{X > t_1\} = \{X > t_1 + t_2\}$, on trouve :

$$\mathbb{P}(X > t_1 + t_2 | X > t_1) = \frac{\mathbb{P}(X > t_1 + t_2)}{\mathbb{P}(X > t_1)} = \frac{e^{-\lambda(t_1+t_2)}}{e^{-\lambda t_1}} = e^{-\lambda t_2} = \mathbb{P}(X > t_2)$$

□

L'inconvénient de la loi exponentielle dans les modélisations est qu'elle suppose une absence de vieillissement du dispositif. Nous allons donc chercher un meilleur modèle lorsque cela importe. Pour cela, nous avons besoin d'introduire la fonction suivante :

Définition 51 : Fonction Gamma

Pour $x > 0$, l'intégrale

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

est bien définie. On l'appelle fonction Gamma ou intégrale eulérienne de première espèce. Elle vérifie :

- $\Gamma(1) = 1$ et $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
- $\Gamma(x+1) = x\Gamma(x)$ En particulier, $\Gamma(n+1) = n!$.

...

Démonstration. On a $\Gamma(1) = \int_0^{+\infty} e^{-u} du = [-e^{-u}]_0^{+\infty} = 1$.

De plus, avec l'intégrale de Gauss,

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} u^{\frac{1}{2}-1} e^{-u} du = 2 \int_0^{+\infty} e^{-u^2} du = \int_{-\infty}^{+\infty} e^{-u^2} du = \sqrt{\pi}.$$

Enfin, pour $x > 0$, en intégrant par partie, on trouve :

$$\Gamma(x+1) = \int_0^{+\infty} u^x e^{-u} du = [-u^x e^{-u}]_0^{+\infty} + \int_0^{+\infty} x u^{x-1} e^{-u} du = 0 + x \Gamma(x)$$

□

On va maintenant pouvoir introduire une extension de la loi exponentielle :

Définition 52 :

Soient $\lambda > 0$ et $r > 0$.

On dit que la variable aléatoire réelle X suit la loi gamma de paramètre r, λ , noté $X \sim \gamma(r, \lambda)$, si la loi de X est absolument continue, de densité :

$$x \mapsto \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Proposition 53 :

Si $X \sim \gamma(r, \lambda)$, alors

$$\mathbb{E}(X) = \frac{r}{\lambda}, \quad V(X) = \frac{r}{\lambda^2}, \quad \text{et} \quad \forall |t| < \lambda, M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}$$

Démonstration. On a comme X est à densité, on a par changement de variable que :

$$\mathbb{E}(X) = \int_0^{+\infty} x \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^{+\infty} \left(\frac{x}{\lambda}\right)^r e^{-x} \frac{dx}{\lambda} = \frac{\Gamma(r+1)}{\lambda \Gamma(r)} = \frac{r}{\lambda},$$

et de même :

$$\mathbb{E}(X^2) = \int_0^{+\infty} x^2 \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^{+\infty} \left(\frac{x}{\lambda}\right)^{r+1} e^{-x} \frac{dx}{\lambda} = \frac{\Gamma(r+2)}{\lambda^2 \Gamma(r)} = \frac{r(r+1)}{\lambda^2},$$

Ce qui permet de dire que $V(X) = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}$.

Enfin, pour $|t| < \lambda$,

$$M_X(t) = \int_0^{+\infty} e^{tx} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx = \int_0^{+\infty} \frac{\lambda^r}{\Gamma(r)} \left(\frac{x}{\lambda-t}\right)^{r-1} e^{-x} \frac{dx}{\lambda-t} = \frac{\lambda^r}{(\lambda-t)^r}.$$

□

Remarque : On a $\gamma(1, \lambda) \stackrel{loi}{=} \mathcal{E}(\lambda)$. Pour $r > 1$, on prend en compte le vieillissement, et en particulier qu'un appareil sorti d'usine a moins de chance d'être défectueux qu'après un long usage.

Remarque 2 : Les lois $\gamma(\frac{k}{2}, \frac{1}{2})$ pour k entier ont le droit à un nom particulier : on parle de loi du χ^2 (lire ki-deux). Cela est dû au fait qu'il s'agit de la loi de la somme des carrés de k variables indépendantes dans leur ensemble suivant des lois normales standards (*i.e.* centré réduit). Elle sera particulièrement intéressante dans des tests d'indépendances.

2.3.4 Le changement de variable pour les lois absolument continues

Faisons un petit détour de notre catalogue du bon probabiliste du continu pour voir comment transformer la densité d'une loi lorsqu'on transforme la variable aléatoire correspondante. Pour cela, nous pourrions utiliser le théorème suivant (où les analystes crieront à l'arnaque, qu'il ne s'agit que du changement de variable des intégrales... et ils n'auront pour une fois pas tout à fait tort) :

Théorème 54 :

Soit $D \subset \mathbb{R}$ et, soit X une variable aléatoire de densité $f_X \mathbb{1}_D$.

Soit $\Psi : D \rightarrow \Delta \subset \mathbb{R}$ qui, soit continue, strictement monotone et de dérivée Ψ' continue et non nulle pour tout x .

Alors la variable aléatoire $Y = \Psi(X)$ est uniformément continue, de densité :

$$f_Y(y) = f_X(\Psi^{-1}(y)) \cdot |(\Psi^{-1})'(y)| \mathbb{1}_\Delta(y).$$

Notation : Il peut être plus commode d'écrire Ψ^{-1} sous la forme $x(y)$. La densité de Y s'écrit alors :

$$f_Y(y) = f_X(x(y)) \cdot \left| \frac{dx}{dy} \right| \mathbb{1}_\Delta(y).$$

2.3.5 La loi normale et la loi log-normale

La loi normale est la loi à densité la plus connue et l'une des plus importantes. Elle apparaîtra naturellement lorsqu'on fait une moyenne de nombreuses petites variations, comme nous le verrons formellement dans 4.3.2.

Définition 55 :

Soient $m \in \mathbb{R}$ et $\sigma > 0$.

On dit que la variable aléatoire réelle X suit la loi normale de paramètre m, σ , noté $X \sim \mathcal{N}(m, \sigma)$, si la loi de X est absolument continue, de densité :

$$x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Proposition 56 :

Si $X \sim \mathcal{N}(m, \sigma)$, alors

$$\mathbb{E}(X) = m, \quad V(X) = \sigma^2, \quad \text{et} \quad \forall |t| < \lambda, M_X(t) = e^{mt + \frac{\sigma^2 t^2}{2}}$$

Démonstration. Par changement de variables,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} (u+m) \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = m + \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \right]_{-\infty}^{+\infty} = m,$$

et de même :

$$\begin{aligned} \text{Var}(X) &= \sigma \text{Var}\left(\frac{X-m}{\sigma}\right) = \sigma \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x.x.e^{-\frac{x^2}{2}} dx \\ &= \sigma \frac{1}{\sqrt{2\pi}} \left([xe^{-\frac{x^2}{2}}]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} -e^{-\frac{x^2}{2}} dx \right) \\ &= \sigma, \end{aligned}$$

Enfin, pour $|t| < \lambda$,

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m-t\sigma^2)^2}{2\sigma^2} + tm + \frac{\sigma^2 t^2}{2}} dx = e^{tm + \frac{\sigma^2 t^2}{2}}.$$

□

Comme cette loi est essentielle en probabilité, il est essentiel de savoir faire des calculs avec. Malheureusement, la fonction $\exp(-x^2)$ n'a pas de primitive connue, il est donc vain de chercher à calculer directement sa fonction de répartition. Pour $X \sim \mathcal{N}(m, \sigma)$, la solution est alors de se ramener à une loi normale centrée réduite avec le changement de variable $u = \frac{X-m}{\sigma}$, puis d'aller chercher les probabilités d'intérêt pour cette loi dans une table.

Pour la loi Normale, je reste normal et je regarde sur la table !

Une autre loi, très similaire puisqu'elle correspond à l'effet multiplicatif cumulé de nombreuses variations, est la loi *log – normale*. Elle est très utilisée pour modéliser des revenus, comme le salaire mensuel, ou encore des durées de vie. Il a trouvé également un domaine d'application dans la linguistique : la longueur d'une phrase, mesurée en nombre de mots par phrase, suit approximativement une loi log-normale. Cette loi ne charge que les réels positifs et offre une distribution étalée à droite.

Définition 57 :

Soient $m \in \mathbb{R}$ et $\sigma > 0$.

On dit que la variable aléatoire réelle X suit la loi log-Normale de paramètre m, σ , si la loi de X est absolument continue, de densité :

$$x \in \mathbb{R} \mapsto \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\ln(x)-m)^2}{2\sigma^2}} \mathbb{1}_{\mathbb{R}_+^*}.$$

Proposition 58 : Admise

Si X suit une loi log-Normale de paramètre m, σ , alors $Y = \ln(X)$ suit une loi Normale, et

$$\mathbb{E}(X) = e^{m + \frac{\sigma^2}{2}}, \quad V(X) = e^{2m + \sigma^2} (e^{\sigma^2} - 1).$$

2.3.6 La loi de Cauchy

Lorsqu'on cherche à modéliser un comportement extrêmement étalé dans l'espace, on peut penser à la loi de Cauchy. Elle peut aussi servir si l'on fait des manipulations de variables aléatoires, puisque le quotient de deux variables aléatoires réelles indépendantes suivant des lois normales standards suit une loi de Cauchy.

Définition 59 :

On dit que la variable aléatoire réelle X suit la loi de Cauchy si la loi de X est absolument continue, de densité :

$$x \in \mathbb{R} \mapsto \frac{1}{\pi(1+x^2)}.$$

Proposition 60 :

Si X suit la loi de Cauchy, sa fonction de répartition est

$$F_X(t) = \frac{1}{\pi} \left(\arctan(t) + \frac{\pi}{2} \right)$$

X n'a pas d'espérance.

c'est fichu, la loi de Cauchy n'a pas d'espérance !

2.3.7 Exercices

Exercice 12 :

Recopiez le tableau suivant où k est entier, fermez le cours et essayer de le remplir (ce n'est pas grave si vous ne réussissez pas au premier essai).

Nom	Paramètres notation usuelle	densité	$\mathbb{E}(X)$	$V(X)$	Fonction génératrice	Hypothèses Cas d'utilisation
$X \sim \text{uniforme}$						
$X \sim \text{exponentielle}$						
$X \sim \text{Gamma}$						
$X \sim \text{Normale}$						
$X \sim \text{log-Normale}$						
$X \sim \text{Cauchy}$						

Exercice 13 :

Soit $X \sim \mathcal{N}(0, 1)$, donnez la loi de $\sigma X + m$.

Soit $X \sim \mathcal{N}(m, \sigma)$, donnez la loi de $\frac{X-m}{\sigma}$.

Soit $X \sim \mathcal{N}(0, 1)$, donnez la densité de la loi de X^2 .

Exercice 14 :

Soit X une variable aléatoire réelle absolument continue de densité f et de fonction de répartition F strictement croissante. On pose la variable aléatoire $Y = F(X)$

- Montrer que pour $t \in [0, 1]$, on a l'égalité $\mathbb{P}(Y \leq t) = t$;
- Donner la loi de Y .

Exercice 15 :

On considère $[0, 1]$ muni de la mesure uniforme. Soit $b \geq 2$ un entier, et pour $x \in [0, 1]$ on note $x \rightarrow (\epsilon_{i,b}(x))_{i \in \mathbb{N}^*}$ l'unique fonction à valeur dans les suites d'entiers entre 0 et $b-1$ et non constamment égale à $b-1$ à partir d'un certain rang, telle que $x = \sum_{i=1}^{+\infty} \frac{\epsilon_{i,b}(x)}{b^i}$

On appelle *nombre univers en base b* tout nombre réel x tel que toute suite de nombre entre 0 et $b-1$ fini se retrouve dans les chiffres de manière successives : $\forall (y_1, \dots, y_n) \in \llbracket 0, b-1 \rrbracket^n, \exists N \in \mathbb{N}, \epsilon_{N+1,b}(x) \cdots \epsilon_{N+n,b}(x) = y_1 \cdots y_n$

1. Pour $i \in \mathbb{N}$, donner la loi de la variable aléatoire $\epsilon_{i,b}(x)$
2. Montrer que la suite $(\epsilon_{i,b}(x))_{i \in \mathbb{N}^*}$ est indépendante dans son ensemble.
3. Montrer que la probabilité d'une suite finie donnée n'apparaisse pas dans la suite des $(\epsilon_{i,b}(x))_{i \in \mathbb{N}^*}$ est nulle
4. Montrer qu'un réel dans $[0, 1]$ est presque sûrement un nombre univers en base b .
5. Montrer qu'un réel dans $[0, 1]$ est presque sûrement un nombre univers dans toutes les bases.

On vient de voir que les nombres univers dans toutes les bases remplissent $[0, 1]$, mais il est étonnant de noter que les mathématiciens sont encore incapables d'exhiber explicitement un tel nombre.

Solution et indication page 87

Chapitre 3

Analyse multivariée des multiples variables

3.1 Vecteurs aléatoires et généralisations des outils univariés

Une fois que l'on maîtrise le comportement d'une variable aléatoire réel, il est naturel de s'intéresser aux interactions potentielles, c'est-à-dire étudier des variables aléatoires de la forme $X = (X_1, \dots, X_d)$ avec chaque X_k une variable aléatoire réelle. Pourquoi cherche-t-on à privilégier une étude multivariée ? Pour pouvoir prendre en compte la structure d'association qui lie les variables aléatoires X_1, \dots, X_d entre elles. Lorsqu'une propriété ou une définition est une généralisation naturelle des notions univariés qui n'ajoute aucun nouveau comportement, nous l'énoncerons sans forcément la commenter en longueur.

3.1.1 Tribu pour les vecteurs aléatoires (H.P.)

Pour pouvoir se munir d'une probabilité, on va devoir se munir d'une tribu d'évènements. Il est naturel (si on veut pouvoir manipuler les fonctions continues sans se poser de questions) de demander que celle-ci contienne les pavés d'intervalles ouverts, c'est-à-dire les ensembles de \mathbb{R}^d de la forme

$$]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_d, b_d[$$

On a alors deux possibilités. Prendre directement la tribu engendrée par ces pavés $\mathcal{B}(\mathbb{R}^d)$, ou prendre la tribu engendrée par les pavés $A_1 \times \dots \times A_d$ où $A_i \in \mathcal{B}(\mathbb{R})$, noté $\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$. Heureusement, ces deux définitions coïncident :

Proposition 61 :

On a $\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$

Démonstration. Comme $]a_1, b_1[\times]a_2, b_2[\times \dots \times]a_d, b_d[\in \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$, par définition de la tribu engendrée comme plus petite tribu contenant tous les éléments, on a l'inclusion $\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R}) \subset \mathcal{B}(\mathbb{R}^d)$

La réciproque est plus compliquée, et nous allons utiliser un argument classique des tribus.

On va montrer par récurrence finie que $H(n)$: "pour tout $0 \leq n \leq d$, pour tout $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})^n$ et

$(a_{n+1}, b_{n+1}, \dots, a_d, b_d) \in \mathbb{R}^{d-n}$, on a que

$$A_1 \times \dots \times A_n \times]a_{n+1}, b_{n+1}[\times \dots]a_d, b_d[\in \mathcal{B}(\mathbb{R}^d) \text{ " .}$$

L'initialisation est évidente, puisque pour $n=0$, elle revient à demander que les pavés d'intervalles ouverts soient dans $\mathcal{B}(\mathbb{R}^d)$.

Pour l'hérédité, supposons $H(n)$ pour $n < d$ et posons

$$\mathcal{A} := \left\{ X \subset \mathbb{R} \mid \forall (A_j) \in \mathcal{B}(\mathbb{R})^n, \forall (a_i, b_i) \in \mathbb{R}^{2(d-n-1)}, A_1 \times \dots \times A_n \times X \times]a_{n+2}, b_{n+2}[\times \dots]a_d, b_d[\in \mathcal{B}(\mathbb{R}^d) \right\}$$

Alors $\emptyset \in \mathcal{A}$ par hypothèse de récurrence (en fait tous les intervalles ouverts).

Si $A \in \mathcal{A}$, alors pour tout $(A_j) \in \mathcal{B}(\mathbb{R})^n$ et $(a_i, b_i) \in \mathbb{R}^{2(d-n-1)}$, comme $\prod_{j=1}^n A_j \times \mathbb{R} \times \prod_{i=n+1}^d]a_i, b_i[\in \mathcal{B}(\mathbb{R}^d)$, on a par stabilité d'une tribu par différence finie que $\prod_{j=1}^n A_j \times A^c \times \prod_{i=n+1}^d]a_i, b_i[\in \mathcal{B}(\mathbb{R}^d)$ car $A^c =]-\infty, +\infty[\setminus A$.

Enfin, par stabilité de $\mathcal{B}(\mathbb{R}^d)$ par union dénombrable, si $(A_i) \in \mathcal{A}$, alors $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

Donc \mathcal{A} est une tribu contenant les pavés ouverts, donc $\mathcal{B}(\mathbb{R}) \subset \mathcal{A}$, c'est-à-dire que $H(n+1)$ est vrai.

D'après le principe de récurrence, on a donc $H(d)$ qui est vrai, c'est-à-dire que $\mathcal{B}(\mathbb{R}) \times \dots \times \mathcal{B}(\mathbb{R}) \subset \mathcal{B}(\mathbb{R}^d)$

Donc par définition de la tribu engendrée, on a l'inclusion $\mathcal{B}(\mathbb{R}^d) \subset \mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})$.

□

Culture générale : Ce résultat n'est pas généralisable en dimension infini, on peut prouver que $\bigotimes_{\mathbb{R}} \mathcal{B}(\mathbb{R}) \neq \mathcal{B}(\mathbb{R}^{\mathbb{R}})$ où l'on a muni $\mathbb{R}^{\mathbb{R}}$ de la topologie de la convergence simple.

3.1.2 Loi de vecteurs aléatoires et fonction de répartition

Maintenant que nous pouvons mettre de côté les discussions techniques des évènements considérés, on va pouvoir parler de vecteurs aléatoires réels :

Définition 62 :

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

Soit $d \in \mathbb{N}^*$ un entier, on dira que X est un vecteur aléatoire (de dimension d) si

$$X : \begin{cases} \Omega & \rightarrow & \mathbb{R}^d \\ \omega & \mapsto & (X_1(\omega), \dots, X_d(\omega)) \end{cases}$$

est une application mesurable de (Ω, \mathcal{T}) vers $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Lorsque c'est le cas, les X_i sont des variables aléatoires réelles.

Pour un vecteur aléatoire, on définit sa loi comme la mesure de probabilité \mathbb{P}_X sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ définie par :

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X(A) := \mathbb{P}[X^{-1}(A)] = \mathbb{P}[X \in A]$$

Remarque : Si $A = A_1 \times A_2 \times \dots \times A_d$, comme $X^{-1}(A) = \bigcap_{k=1}^d X_k^{-1}(A_k)$, on a que

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}\left(\bigcap_{k=1}^d \{X_k \in A_k\}\right)$$

Proposition 63 :

(Admise) Pour toute loi de probabilité Q sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, on peut trouver un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et un vecteur aléatoire X tel que $\mathbb{P}_X = Q$.

Maintenant que nous avons nos vecteurs aléatoires, peut-on généraliser les outils vus précédemment pour les variables aléatoires ? La réponse est oui, même si parfois la situation multivariée est plus complexe que l'univariée. Commençons donc par la fonction de répartition :

Définition 64 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X un vecteur aléatoire dessus.

On appelle fonction de répartition (de la loi) de X la fonction :

$$F : \begin{cases} \mathbb{R}^d & \rightarrow [0, 1] \\ (t_1, \dots, t_d) & \mapsto \mathbb{P}_X([-\infty, t_1] \times \dots \times [-\infty, t_d]) = \mathbb{P}\left(\bigcap_{k=1}^d \{X_k \leq t_k\}\right) \end{cases}$$

On retrouve dans la version multivariée de la fonction de répartition de nombreux comportements déjà vu pour les variables aléatoires, mais il y a néanmoins quelques subtilités d'utilisation qu'il faut bien maîtriser. Les plus basiques sont énoncées dans la proposition suivante :

Proposition 65 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X un vecteur aléatoire de fonction de répartition F . Alors :

1. $0 \leq F \leq 1$;
2. Si pour tout $k \in \llbracket 1, d \rrbracket$, $t_k \leq t'_k$, alors

$$F(t_1, \dots, t_d) \leq F(t'_1, \dots, t'_d);$$

3. F est continue à droite et admet une limite à gauche en chacune de ses variables ;
4. S'il existe k tel que $\lim_{n \rightarrow +\infty} t_{k,n} = -\infty$, et tous les autres t_i sont fixés alors $F(t_1, t_2, \dots, t_{k,n}, \dots, t_d) \rightarrow 0$;
si pour tout $k \in \llbracket 1, d \rrbracket$, $\lim_{n \rightarrow +\infty} t_{k,n} = +\infty$ alors $F(t_{1,n}, t_{2,n}, \dots, t_{d,n}) \rightarrow 1$.

Remarque : Il faut faire attention puisqu'on ne peut rien dire *a priori* sur la continuité globale, ni l'existence de limite dans d'autres directions que les axes.

Démonstration. La preuve est la même qu'en univariée, il s'agit juste d'être méticuleux dans la manipulation des inclusions ensemblistes. □

On retrouvera encore la distinction entre les lois discrètes et uniformément continue :

Définition 66 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X un vecteur aléatoire .

On dit que X est discrète si l'image de X est \mathbb{P}_X -presque sûrement dénombrable (ou fini)

On dit que X est uniformément continue s'il existe une fonction (appelé la densité) $f : \mathbb{R}^d \rightarrow \mathbb{R}_+$, intégrable telle que

$$\forall (A_i) \in \mathcal{B}(\mathbb{R})^d, \quad \mathbb{P}_X \left(\prod_{i=1}^d A_i \right) = \int_{A_1} \int_{A_2} \cdots \int_{A_d} f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

Heureusement, comme pour les variables aléatoires absolument continus, on peut pour les vecteurs aléatoires absolument continus passer de la densité de la loi à la fonction de répartition, et réciproquement :

Proposition 67 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X un vecteur aléatoire de dimension n uniformément continue de fonction de répartition F et de densité f .

Alors

$$F(t_1, \dots, t_n) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \cdots \int_{-\infty}^{t_d} f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

et si f est continue, alors

$$\frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F = f.$$

Remarque : Attention cependant, il est fort possible d'avoir un vecteur aléatoire avec une composante qui est une variable aléatoire discrète et une deuxième composante qui est une variable aléatoire absolument continue.

3.1.3 Espérance, matrice de covariance et de corrélations

À présent, nous allons chercher à généraliser les moments. Les seuls ayant une intuition facilement généralisable sont les moments d'ordre 1 et 2.

Définition 68 :

Pour $\mathbf{X} = (X_1, \dots, X_d)$ est un vecteur aléatoire de \mathbb{R}^d , on dit que \mathbf{X} est de puissance r -ième intégrable si chacune de ses composantes l'est.

Lorsque $r = 1$, on parle d'espérance mathématique et l'espérance du vecteur est alors le vecteur des espérances, c'est-à-dire que

$$\mathbb{E}(\mathbf{X}) := \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \vdots \\ \mathbb{E}(X_d) \end{pmatrix}$$

Pour l'espérance, rien ne change à part la dimension de l'espace ambiant :

Proposition 69 :

L'espérance vérifie encore que :

- Si \mathbf{X} est presque sûrement égal à un vecteur b , *i.e.* $\mathbb{P}(X = b) = 1$, alors $\mathbb{E}(X) = \mathbb{E}(b) = b$.
- Pour (X, Y) deux vecteurs aléatoires réels et (a, b) deux réels, on a que

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

En revanche, le moment d'ordre 2 quand il existe permet d'apercevoir de nouveau comportement. Il permet d'apporter un début de réponses à la question naturelle lorsqu'on relève plusieurs résultats : sont-ils tous pertinents (ou est-ce que l'un permettrait de calculer l'autre) ? Il permet même de s'intéresser à la puissance explicative de l'un des paramètres pour l'autre (quel part de ce premier phénomène s'explique par ce second phénomène).

Définition 70 :

Soit $\mathbf{X} = (X_1, \dots, X_d)$ est un vecteur aléatoire de \mathbb{R}^d quadratiquement intégrable.

Comme pour tous couples de réels l'inégalité $|X_i X_j| \leq \frac{1}{2}(X_i^2 + X_j^2)$ est vérifiée, on a l'existence pour tout couple (i, j) de $\mathbb{E}(X_i X_j)$.

On appelle matrice des variances covariances ou encore matrice de covariance, la matrice (où $m_i := \mathbb{E}(X_i)$) : $\Lambda = (\sigma_{ij})_{1 \leq i, j \leq d}$ définie par :

$$\sigma_{ij} = \text{cov}(X_i, X_j) = \mathbb{E}[(X_i - m_i)(X_j - m_j)],$$

On dira que le paramètre σ_{ij} est la covariance entre X_i et X_j .

Proposition 71 :

Sous les hypothèses de la définition précédente, on a que :

1. $\sigma_{ii} = V(X_i)$;
2. $\sigma_{ij} = \sigma_{ji} = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)$;
3. $\text{Cov}(\sum_{k=1}^n X_k, \sum_{j=1}^m X_j) = \sum_{k=1}^n \sum_{j=1}^m \text{Cov}(X_k, X_j)$,

La covariance est une forme bilinéaire et la matrice Λ est symétrique semi-définie positive ;

4. $V(a_1 X_1 + a_2 X_2) = a_1^2 V(X_1) + a_2^2 V(X_2) + 2a_1 a_2 \text{cov}(X_1, X_2)$;
5. Si les X_i sont deux à deux indépendantes, alors Λ est diagonale.

Remarque : Attention, pour le dernier point, la réciproque est fautive, sauf si \mathbf{X} est un vecteur Gaussien (voir 3.1.4 page 49).

Maintenant, lorsqu'il s'agit d'interpréter les résultats pour décrire une interaction, un problème se pose : la covariance dépend de l'ordre de grandeur des variables X_i et X_j . Ceci est facile à résoudre, il suffit de standardiser les variables,

de les renormaliser. Les tailles auxquelles on peut naturellement comparer la covariance sont les variances des X_i et X_j , de taille caractéristique $\sigma(X_i)\sigma(X_j)$.

Définition 72 :

Soient (X_1, X_2) un vecteur aléatoire, de moment d'ordre 2 fini, et on suppose que $V(X_k) \neq 0$ pour $k = 1, 2$ (i.e. aucune des composantes n'est presque sûrement constante).

Rappelons que $\sigma(X_j) = V^{\frac{1}{2}}(X_j)$. On appelle coefficient de corrélation linéaire entre X_1 et X_2 , noté $\rho(X_1, X_2)$, le réel :

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}.$$

Un premier résultat dans la direction que j'annonce est la proposition suivante. Il existe des résultats plus avancés de décomposition des lois, qui généralisent le dernier point en particulier lorsque l'égalité n'est pas vérifiée, mais sort largement de la visée de ce cours.

Proposition 73 :

Soient (X_1, X_2) un vecteur aléatoire, de moment d'ordre 2 fini, tels que $V(X_k) \neq 0$ pour $k = 1, 2$.

1. $-1 \leq \rho(X_1, X_2) \leq 1$,
2. Si X_1 et X_2 sont indépendantes, alors $\rho(X_1, X_2) = 0$ (la réciproque est fausse).
3. Si $|\rho(X_1, X_2)| = 1$, alors il existe trois réels $(a, b, c) \in \mathbb{R}^3$ avec a et b non nuls tels que $aX_1 + bX_2 + c = 0$ presque sûrement (donc une des variables explique totalement l'autre).

3.1.4 Deux exemples

Exemple de loi discrète : la loi multinomiale

La loi multinomiale servira à modéliser des comparaisons de proportions entre de multiples résultats possibles (comme par exemple les résultats d'un sondage sur la présidentielle avec remise).

Schématiquement, il s'agit dans un tirage de boules, contenu dans une urne avec d couleurs de boules distinctes (où pour $k \in \llbracket 1, d \rrbracket$, on note p_k la proportion de boules de la k -ième couleur), répété r fois au hasard et avec remise, de s'intéresser au vecteur aléatoire $X = (X_1, \dots, X_d)$ avec X_k qui représente le nombre de boules obtenues de la k -ième couleur.

Définition 74 :

Soit $0 < p_1, \dots, p_d < 1$ tels que $\sum_{i=1}^d p_i = 1$

On dit que $X = (X_1, \dots, X_d)$ suit une loi multinomiale de paramètres r, p_1, \dots, p_d noté $X \sim \mathcal{M}(r; p_1, \dots, p_d)$ si

$$\forall (r_i) \in \mathcal{R}, \mathbb{P}(X_1 = r_1, \dots, X_d = r_d) = \frac{r!}{r_1! r_2! \dots r_d!} p_1^{r_1} p_2^{r_2} \dots p_d^{r_d}$$

où $\mathcal{R} := \{(r_1, \dots, r_d) | 0 \leq r_1, \dots, r_d \leq r, \sum_{i=1}^d r_i = r\}$

Intuition derrière la loi : il y a deux termes principaux à comprendre. Le produit $p_1^{r_1} p_2^{r_2} \dots p_d^{r_d}$ qui, compte tenu de

l'indépendance des tirages successifs, correspond à la probabilité d'une série de r tirages durant lesquels on a obtenu r_1 boules de la couleur 1, ..., et r_d boules de la couleur d . Ensuite, le terme combinatoire qui précède correspond au nombre de séries distinctes aboutissant à une telle composition finale (c'est-à-dire le nombre de manières de réarranger l'ordre dans lequel on tire ces boules).

Origine du nom : La loi multinomiale tire son nom du développement du multinôme (généralisation du binôme)

$$(p_1 + \dots + p_n)^r = \sum_{r_1, \dots, r_n : \sum r_i = r} \frac{r!}{r_1! r_2! \dots r_d!} p_1^{r_1} p_2^{r_2} \dots p_d^{r_d}$$

Liaison : La contrainte qui pèse sur les r_k se retrouve sur les X_k sous la forme :

$$\sum_{i=1}^d X_i = r$$

En particulier, les composantes X_k sont linéairement dépendantes et donc ne peuvent pas être indépendantes.

Dernière remarque : Pour $d = 2$, on retrouve les probabilités d'une loi binomiale. Pour $d \geq 2$, on montre que chaque composante X_k suit une loi $\text{Bin}(r, p_k)$.

Exemple de loi à densité : la loi normale

Tout comme précédemment, la loi normale suit un rôle central en probabilité.

Définition 75 :

On dit qu'un vecteur aléatoire $X = (X_1, \dots, X_n)$ est un vecteur gaussien si pour toute collection de réels $(\lambda_1, \dots, \lambda_n)$, la variable aléatoire $\sum_{i=1}^n \lambda_i X_i$ suit une loi normale.

Proposition 76 : Admis

Pour X_1, \dots, X_n une suite de variable aléatoire gaussienne **indépendante**, le vecteur aléatoire $X = (X_1, \dots, X_n)$ est gaussien.

Si en plus les X_i sont centrés et réduits, on parlera de *vecteurs gaussiens standard*.

Dans les faits, nous utiliserons plutôt la caractérisation suivante :

Théoreme 77 :

Soit $m \in \mathbb{R}^d$, et Λ une matrice $d \times d$ symétrique *définie positive*. On considère X un vecteur aléatoire, alors on a l'équivalence entre les points

- X est un vecteur gaussien de moyenne m et de matrice de covariance Λ ;
- X admet une densité par rapport à Lebesgue de la forme

$$\frac{1}{(\sqrt{2\pi})^d \sqrt{\det(\Lambda)}} \exp\left(-\frac{(x-m)^t \Lambda^{-1} (x-m)}{2}\right)$$

Lorsque c'est le cas, la variable aléatoire X_j est alors de loi $\mathcal{N}(m_j, \Lambda_{jj})$.

La démonstration est laissée en exercice en 4.2.5, page 74.

3.1.415 Exercice**Exercice 1 :**

Soit X, Y un couple de variable aléatoire dont la loi est donnée par le tableau suivant, contenant les probabilités de tirer un couple de valeurs :

$X \backslash Y$	1	3	5	8
-3	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
5	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{3}{16}$

Calculer la moyenne du vecteur (X, Y) , ainsi que la matrice de covariance et le coefficient de corrélation entre X et Y .

Exercice 2 :

Soit $X \sim \mathcal{M}(r; p_1, \dots, p_d)$, montrer que chaque composante X_k suit une loi $\text{Bin}(r, p_k)$.

Exercice 3 :

Soit $X \sim \mathcal{N}(0, 1)$, et ϵ une variable aléatoire indépendante de X qui vaut ± 1 avec probabilité $\frac{1}{2}$.

Montrer que $(X, \epsilon X)$ n'est pas un vecteur gaussien. Est-ce en contradiction avec la proposition 76 ? Pourquoi ?

Solution ou indication page 94

3.2 Calculer avec plusieurs aléas

3.2.1 Marginales

À partir de la loi d'un vecteur aléatoire, il est possible de retrouver la loi de ses composantes :

Définition 78 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $X = (X_1, \dots, X_d)$ un vecteur aléatoire réel.

On appelle loi conjointe (de X) la loi du vecteur aléatoire.

On appelle loi marginale de X_1 (par rapport à X) la loi de la composante X_1 .

Proposition 79 :

La loi conjointe détermine, de façon unique, les lois marginales.

La réciproque est fautive (i.e. les lois marginales ne déterminent pas de façon unique la loi conjointe).

Démonstration. Considérons un vecteur aléatoire $X = (X_1, \dots, X_d)$ de loi \mathbb{P}_X , et cherchons la loi de X_1 .

Soit $A \in \mathcal{T}$, alors on a :

$$\mathbb{P}_{X_1}(A) = \mathbb{P}(X_1 \in A) = \mathbb{P}(X_1 \in A, \forall i > 1, X_i \in \mathbb{R}) = \mathbb{P}_X(A \times \mathbb{R}^{d-1})$$

Pour la réciproque, on peut par exemple considérer l'exercice de la section précédente fournissant un vecteur aléatoire ayant des marginales normales, mais qui n'est pas un vecteur gaussien. \square

Exemple : Prenons par exemple le vecteur aléatoire comptant le nombre de pièces (X) et le nombre de personnes (Y) vivants dans un appartement, dans un groupe de $N = 270$ personnes, et présentons les effectifs (qui permettent de trouver la loi conjointe en divisant par l'effectif total) dans un tableau à double entrée :

$X \setminus Y$	1	2	3	≥ 4	
1	35	25	10	3	
2	20	23	25	10	
3	2	15	35	20	
≥ 4	1	3	18	28	
					N=270

Maintenant, supposons que l'on se demande la répartition du nombre de personnes par appartement (loi de X), ou encore celle du nombre de pièces par appartement, on peut la trouver à partir de ce tableau. L'idée générale est de « sommer » la loi conjointe par rapport aux autres variables. On trouve alors les lois marginales :

$X \setminus Y$	1	2	3	≥ 4	Loi marginale de X
1	35	25	10	3	70
2	20	23	25	10	78
3	2	15	35	20	72
≥ 4	1	3	18	28	50
Loi marginale de Y	55	66	88	61	N=270

Vecteurs aléatoire discret : Voyons en pratique quels calculs sont nécessaires pour trouver les lois marginales pour un vecteur aléatoire discret

On suppose que I et J sont deux ensembles finis de la forme $I = \llbracket 1; l \rrbracket$ et $J = \llbracket 1; l \rrbracket$. On s'intéresse à un couple (X, Y) à valeurs dans $\{(x_i, y_j)_{i \in I, j \in J}\}$ et tels que

$$\mathbb{P}[X = x_i, Y = y_j] = p_{ij}.$$

En utilisant la formule des probabilités totales, on écrit :

$$\mathbb{P}[X = x_i] = \mathbb{P}\left[X = x_i \cap \left(\bigcup_{j \in J} Y = y_j\right)\right] = \sum_{j \in J} \mathbb{P}[X = x_i, Y = y_j] = \sum_{j \in J} p_{ij} := p_{i.}$$

Avec le même calcul, on trouve que Y est presque sûrement à valeur dans $\{y_j | j \in J\}$ et :

$$\mathbb{P}[Y = y_j] = \sum_{i \in I} p_{ij} := p_{.j}$$

On retrouve bien cette intuition de sommation.

Et pour les vecteurs aléatoires absolument continus ? On connaît la loi conjointe d'un couple (X_1, X_2) définie par une densité $(x_1, x_2) \mapsto f(x_1, x_2)$. Il est naturel de se demander comment trouver les lois marginales à partir de cette densité si elles sont à densité, et si oui, comment calculer cette densité. La réponse à ces questions est oui, comme on peut par exemple le voir en calculant avec le théorème de Fubini-Tonelli la fonction de répartition d'une marginale :

$$\begin{aligned} F_{X_1}(t) &= \mathbb{P}(X_1 \leq t) = \mathbb{P}(X_1 \leq t, X_2 \in \mathbb{R}) = \mathbb{P}((X_1, X_2) \in]-\infty; t[\times \mathbb{R}) \\ &= \int \int_{]-\infty; t[\times \mathbb{R}} f(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^t \left(\int_{\mathbb{R}} f(x_1, x_2) dx_2 \right) dx_1 \end{aligned}$$

On conclut de ce calcul que X_1 est à densité, de densité égale à :

$$x_1 \mapsto f_1(x_1) := \int_{\mathbb{R}} f(x_1, x_2) dx_2$$

On retrouve encore une fois cette intuition de sommation, puisque l'intégration est la généralisation de la somme à des ensembles indénombrables.

Pour obtenir la densité marginale, on somme (intègre) sur tous les possibles la densité conjointe par rapport aux variables qui ne nous intéressent pas.

3.2.2 Indépendance

L'indépendance des variables aléatoire est une des notions les plus importantes de la théorie des probabilités, permettant la simplification de nombreux calculs. Rappelons la définition de l'indépendance d'une famille finie, un concept bien pratique pour de nombreux calculs :

Définition 80 :

Soient X_1, \dots, X_d une famille de d variables aléatoires, on dit que ces variables aléatoires sont mutuellement indépendantes si pour toute suite d'événements A_1, \dots, A_d , on a :

$$\mathbb{P}(X_1 \in A_1, \dots, X_d \in A_d) = \prod_{j=1}^d \mathbb{P}(A_j),$$

Pour montrer qu'une famille de variable aléatoire est indépendante, ou encore pour utiliser concrètement cette notion, on peut se servir des deux caractérisations suivantes :

Théoreme 81 :

Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire réel.

On appelle F_k la fonction de répartition de X_k et F celle de X . Alors l'indépendance de X_1, X_2, \dots, X_d est équivalente à :

$$\forall (x_1, \dots, x_d) \in \mathbb{R}^d, \quad F(x_1, \dots, x_d) = \prod_{i=1}^d F_k(x_i)$$

Démonstration. Il s'agit d'appliquer la définition de l'indépendance aux événements de la forme $] - \infty, x_1] \times] - \infty, x_2] \times \dots \times] - \infty, x_d]$. \square

Théoreme 82 :

Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire réel, et on suppose que pour tout k , X_k admet la densité f_k . Alors

- Si les X_k sont indépendantes, X admet la densité :

$$f(x_1, \dots, x_d) = \prod_{k=1}^d f_k(x_k)$$

- Inversement, si X admet une densité f et que $f(x_1, \dots, x_d) = \prod_{k=1}^d f_k(x_k)$, alors les X_k sont indépendants.

Démonstration. Pour le premier point, il s'agit de remarquer que pour toute suite d'événements A_1, \dots, A_d ,

$$\prod_{j=1}^d \mathbb{P}(X_j \in A_j) = \prod_{j=1}^d \int_{A_j} f_j(x_j) dx_j = \int_{A_1} \int_{A_1} \dots \int_{A_d} \prod_{j=1}^d (f_j(x_j)) dx_1 \dots dx_d$$

Pour le deuxième point, on remonte le calcul. \square

Une des propriétés des variables aléatoires indépendantes la plus utilisée en pratique est la suivante :

Proposition 83 :

On suppose que les $(X_k)_{k=1,\dots,d}$ sont des variables aléatoires réelles indépendantes.

- Soient $(g_k)_{k=1,\dots,d}$ des fonctions de \mathbb{R} dans \mathbb{R} mesurables. Alors les variables aléatoires réelles $g_k(X_k)$ sont indépendantes. *En général, la réciproque est fausse.*

En particulier :

- Le produit $X_1 X_2 \cdots X_d$ est intégrable si et seulement si pour tout k , X_k est intégrable, et on a :

$$\mathbb{E}\left(\prod_{i=1}^d X_i\right) = \prod_{i=1}^d \mathbb{E}(X_i).$$

Notons qu'encore une fois cette dernière égalité ne suffit pas à garantir l'indépendance.

3.2.3 Changements de variable dans les hyper-espaces

Mais maintenant, comment faire pour faire les calculs lorsqu'on n'a pas d'hypothèses d'indépendance ? Regardons un cas particulier fondamental :

On dispose d'un vecteur aléatoire X dont on connaît la loi. Soit T une application suffisamment régulière de \mathbb{R}^d dans \mathbb{R}^d . On veut connaître la loi du vecteur aléatoire $T(X)$.

Déjà, pour tous $A \in \mathcal{B}(\mathbb{R}^d)$ on a :

$$\mathbb{P}[T(X) \in A] = \mathbb{P}[X \in T^{-1}(A)],$$

ce qui permet d'obtenir la loi de $T(X)$ à partir de celle de X .

En pratique, on s'en servira surtout pour les variables aléatoires à densité :

Théoreme 84 :

Soit $X = (X_1, \dots, X_d)$, un vecteur aléatoire de \mathbb{R}^d admettant la densité :

$$(x_1, x_2, \dots, x_d) \mapsto f(x_1, x_2, \dots, x_d) \mathbb{1}_D(x_1, x_2, \dots, x_d),$$

où D est un ouvert de \mathbb{R}^d .

Soit une application $T : D \rightarrow \Delta \subset \mathbb{R}^d$ bijective, de classe C^1 , telle que T^{-1} soit aussi de classe C^1 .

On note pour $k = 1, \dots, d$, $x_k = g_k(y_1, \dots, y_d)$ les formules définissant T^{-1} , c'est-à-dire $T^{-1} = (g_1, g_2, \dots, g_d)$

Alors le vecteur $T(X)$ admet la densité $h(y_1, \dots, y_d)$ qui s'écrit :

$$f \circ T^{-1} |J_{T^{-1}}| \mathbb{1}_\Delta = f(g_1(y_1, \dots, y_d), g_2(y_1, \dots, y_d), \dots, g_d(y_1, \dots, y_d)) |J_{T^{-1}}| \mathbb{1}_\Delta(y_1, \dots, y_d)$$

Le terme $|J_{T^{-1}}|$ dans le théorème représente la valeur absolue du jacobien $J_{T^{-1}}$, qui est le déterminant, supposé non nul en tout point, de la matrice jacobienne :

$$JacT^{-1}(y) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(y_1, \dots, y_d) & \cdots & \frac{\partial g_1}{\partial y_d}(y_1, \dots, y_d) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_d}{\partial y_1}(y_1, \dots, y_d) & \cdots & \frac{\partial g_d}{\partial y_d}(y_1, \dots, y_d) \end{pmatrix}$$

Et dans la pratique ? Soient deux variables aléatoires réelles X et Y dont on connaît la loi conjointe de densité $(x, y) \mapsto f(x, y)$. Supposons que l'on veuille calculer la loi de la somme, du produit ou du quotient de X et Y , c'est-à-dire connaître la loi de $Z_1 = X + Y$, $Z_2 = XY$ ou encore $Z_3 = \frac{X}{Y}$ en supposant $Y(\omega) \neq 0$ p.s.

On opère dans \mathbb{R}^2 avec un des deux changements de variables bijectifs suivants :

$$\begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} X \\ Z_i \end{pmatrix} \quad \text{ou} \quad \begin{pmatrix} X \\ Y \end{pmatrix} \mapsto \begin{pmatrix} Z_i \\ Y \end{pmatrix}$$

Par le théorème du changement de variables, on obtient la loi conjointe du couple $\begin{pmatrix} X \\ Z_i \end{pmatrix}$ ou du couple $\begin{pmatrix} Z_i \\ Y \end{pmatrix}$, puis on récupère la loi de Z_i comme loi marginale.

En particulier, si X et Y sont indépendantes de densités respectives f et g , la densité h de $Z_1 = X + Y$ s'écrit alors comme le produit de convolution $f * g$ de f et g , c'est-à-dire :

$$h(z) = \int_{\mathbb{R}} f(z-x)g(x)dx = \int_{\mathbb{R}} g(z-x)f(x)dx.$$

Comme pour les variables aléatoires, la loi est fondamentale pour les vecteurs aléatoires, et elle permet de retrouver les quantités d'intérêt :

Proposition 85 :

Soit X un vecteur aléatoire de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ de loi Q_X . Soit g une application mesurable de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors $g(X)$ est intégrable sur $(\Omega, \mathcal{T}, \mathbb{P})$ ssi g est intégrable sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), Q_X)$ et l'on a :

$$\int_{\Omega} g(X(\omega))D\mathbb{P}(\omega) = \int_{\mathbb{R}^d} g(x_1, \dots, x_d)dQ_X(x_1, \dots, x_d).$$

Corollaire 3.2.1. *Supposons que $X = (X_1, X_2)$ soit un vecteur aléatoire de \mathbb{R}^2 admettant la densité f . Soit g une application de \mathbb{R}^2 dans \mathbb{R} telle que $Y = g(X)$ soit une variable aléatoire de loi Q_Y . Alors, on a :*

$$\mathbb{E}(Y) = \int_{\mathbb{R}} ydQ_Y(y) = \int_{\mathbb{R}^2} g(x_1, x_2)f((x_1, x_2))dx_1dx_2$$

Cette dernière formule nous permet, en particulier, de calculer des covariances.

3.2.4 Exercice

Exercice 4 :

Montrer que si $X \sim \mathcal{M}(r; p_1, \dots, p_d)$, alors ses composantes ne sont pas indépendantes dans leur ensemble.

Exercice 5 :

Soient X_1 , X_2 et X_3 des variables aléatoires mutuellement indépendante de loi de Rademacher, c'est-à-dire

$$\forall i \in \{1, 2, 3\}, \quad \mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = \frac{1}{2}$$

On pose alors $Y_1 = X_2X_3$, $Y_2 = X_1X_3$ et $Y_3 = X_1X_2$.

- Donner les lois marginales des Y_i .
- Calculer les lois des couples (Y_i, Y_j) pour $i \neq j$. Les variables Y_1 , Y_2 et Y_3 sont-elles deux à deux indépendantes ?
- Calculer la loi du triplet. Les variables Y_1 , Y_2 et Y_3 sont-elles mutuellement indépendantes ?

Solution ou indication page 94

Chapitre 4

Théorèmes limites de l'aléatoire

4.1 Fonction caractéristique

4.1.1 Définitions de la fonction caractéristique

On va dans cette section introduire la fonction caractéristique, qui va permettre de caractériser la loi, mais aussi de donner des critères de convergence. On peut aussi la voir comme une généralisation de la transformée de Fourier appliqué en une loi de probabilité.

Soit X une variable aléatoire réelle de loi \mathbb{P}_X . On peut alors remarquer que l'application $\omega \mapsto e^{itX(\omega)}$ est, pour tout $t \in \mathbb{R}$, une variable aléatoire à valeurs complexes, et bornée. Elle est donc intégrable puisqu'une mesure de probabilité est de masse finie, et on a :

$$\mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x).$$

Rappelons qu'une telle intégrale peut se calculer en l'écrivant comme somme de sa partie réelle et de sa partie imaginaire et en intégrant séparément chacune d'elles.

Cette remarque permet de justifier le bien fondé de la définition suivante :

Définition 86 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle.

On appelle fonction caractéristique de (la loi de) X la fonction à valeur complexe :

$$t \mapsto \varphi_X(t) := \mathbb{E}(e^{itX})$$

Exemple : On considère X une variable aléatoire réelle de loi uniforme sur $\{1, \dots, n\}$, et on veut calculer sa fonction caractéristique. Par définition, pour $t \in \mathbb{R}$ on a :

$$\begin{aligned} \varphi_X(t) &= \sum_{k=1}^n e^{tk} p_k = \sum_{k=1}^n e^{tk} \frac{1}{n} \\ &= \frac{e^{it}}{n} \sum_{k=0}^{n-1} e^{tk} = \begin{cases} \frac{e^{it}}{n} \frac{1-e^{itn}}{1-e^{it}} & \text{si } t \neq 0[2\pi] \\ e^{i0} = 1 & \text{sinon} \end{cases} \end{aligned}$$

Exemple : On considère X une variable aléatoire réelle à densité, de densité f . Si on note

$$\mathcal{F} : f \mapsto [t \mapsto \int_{\mathbb{R}} f(x) e^{-2i\pi t x} dx]$$

la transformée de Fourier usuelle (et \mathcal{F}^{-1} son inverse), alors on a que :

$$\forall t \in \mathbb{R}, \varphi_X(t) = \mathcal{F}(f) \left[-\frac{t}{2\pi} \right] = \mathcal{F}^{-1}(f) \left[\frac{t}{2\pi} \right].$$

Proposition 87 :

Soit X une variable aléatoire réelle, de loi \mathbb{P}_X . On a alors, pour $t \in \mathbb{R}$:

1. $|\varphi_X(t)| \leq 1$ et $\varphi_X(0) = 1$
2. La fonction φ_X est continue sur \mathbb{R}
3. Si a et b sont deux réels et $Y = aX + b$, alors on a $\varphi_Y(t) = e^{itb} \varphi_X(at)$
4. Si X et Y sont indépendantes, alors $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$

Démonstration. En suivant les notations de l'énoncé :

1. Par inégalité triangulaire de l'intégrale, $|\varphi_X(t)| = |\mathbb{E}(e^{itX})| \leq \mathbb{E}(|e^{itX}|) = \mathbb{E}(1) = 1$.
De plus, $\varphi_X(0) = \mathbb{E}(e^0) = 1$.
2. Pour tout ω , la fonction $t \mapsto e^{itX(\omega)}$ est continue sur \mathbb{R} , et est dominée par la fonction \mathbb{P} -intégrable constante égale à 1.
Donc par continuité par le symbole intégral, φ_X est continue sur \mathbb{R} .
3. Soient a et b sont deux réels et $Y = aX + b$. alors on a $\varphi_Y(t) = \mathbb{E}(e^{itY}) = \mathbb{E}(e^{it(aX+b)}) = \mathbb{E}(e^{itb} e^{itaX}) = e^{itb} \varphi_X(at)$ où l'on a utilisé la linéarité de l'espérance pour la dernière égalité.
4. Si X et Y sont indépendantes alors $\varphi_{X+Y}(t) = \mathbb{E}(e^{it(X+Y)}) = \mathbb{E}(e^{itX} e^{itY}) = \mathbb{E}(e^{itX}) \mathbb{E}(e^{itY}) = \varphi_X(t) \cdot \varphi_Y(t)$

□

Proposition 88 :

Soit X est une variable aléatoire réelle telle que pour tout $k \leq K$ entier, $\mathbb{E}(|X|^k) < +\infty$, alors φ_X est dérivable au moins jusqu'à l'ordre K et l'on a :

$$\forall k \leq K, \varphi_X^{(k)}(t) = i^k \mathbb{E}(X^k e^{itX})$$

En particulier, cela permet de retrouver les moments de X , puisque :

$$\forall k \leq K, \varphi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$$

Démonstration. En suivant les notations de l'énoncé :

Pour tout $k \leq K$ et tout $\omega \in \Omega$, la fonction $t \mapsto e^{itX(\omega)}$ est C^k sur \mathbb{R} . De plus, sa k^{ieme} dérivée est dominée sur \mathbb{R} par la fonction $\omega \mapsto |X(\omega)|^k$, qui est \mathbb{P} intégrable par hypothèse.

Donc par dérivation sous le signe intégral, on a le résultat annoncé. \square

Définition 89 :

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire, on généralise la fonction caractéristique en l'application :

$$\varphi_X : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{C} \\ \bar{t} = (t_1, \dots, t_n) & \mapsto \mathbb{E}(e^{i\langle \bar{t}, X \rangle}) = \mathbb{E}(e^{i \sum_p t_p X_p}) \end{cases}$$

Les deux propriétés précédentes sont encore vérifiées en remplaçant la dérivation par une dérivée partielle.

On admettra dans la suite que tout les énoncé à propos de la fonction caractéristique de variables aléatoires se généralisent aux vecteurs aléatoires.

4.1.2 Calculs type de fonctions caractéristiques

On a les fonctions caractéristiques suivantes :

Proposition 90 :

Soit $t \in \mathbb{R}$,

- Si $X \sim \mathcal{B}(p)$, alors $\varphi_X(t) = pe^{it} + q$;
- Si $X \sim \mathcal{G}(p)$, alors $\varphi_X(t) = \frac{pe^{it}}{1 - qe^{it}}$;
- Si $X \sim \mathcal{P}(\lambda)$, alors $\varphi_X(t) = e^{\lambda(e^{it} - 1)}$;
- Si $X \sim \gamma(r, \lambda)$, alors $\varphi_X(t) = \left(\frac{\lambda}{\lambda - it} \right)^r$;
- Si $X \sim \text{Cauchy}$, alors $\varphi_X(t) = e^{-|t|}$;
- Si $X \sim \mathcal{N}(0, 1)$, alors $\varphi_X(t) = e^{-\frac{t^2}{2}}$.

Démonstration. Soit $t \in \mathbb{R}$,

- Si $X \sim \mathcal{B}(p)$, alors par définition de l'espérance, $\varphi_X(t) = pe^{it} + q$;

- Si $X \sim \mathcal{G}(p)$, alors :

$$\begin{aligned}\varphi_X(t) &= \sum_{k=1}^{+\infty} pq^{k-1} e^{itk}, \\ &= pe^{it} \sum_{k=0}^{+\infty} (qe^{it})^k, \\ &= \frac{pe^{it}}{1 - qe^{it}};\end{aligned}$$

- Si $X \sim \mathcal{P}(\lambda)$, alors :

$$\begin{aligned}\varphi_X(t) &= \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk}, \\ &= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(\lambda e^{it})^k}{k!}, \\ &= e^{\lambda(e^{it}-1)};\end{aligned}$$

- Si $X \sim \gamma(r, \lambda)$, alors :

$$\begin{aligned}\varphi_X(t) &= \int_{\mathbb{R}_+} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} e^{itx} dx, \\ &= \frac{\lambda^r}{\Gamma(r)} \frac{1}{(\lambda - it)^r} \int_{\mathbb{R}_+} x^{r-1} e^{-x} dx, \\ &= \frac{\lambda^r}{(\lambda - it)^r};\end{aligned}$$

- Soit $X \sim \text{Cauchy}$, on note f la densité de la loi de Cauchy, c'est-à-dire $f : x \mapsto \frac{1}{\pi(1+x^2)}$. Considérons la fonction L^1 définie par $\psi : t \mapsto e^{-|t|}$ et calculons son image par la transformation de Fourier :

$$\begin{aligned}\mathcal{F}(\psi)(t) &= \int_{\mathbb{R}} e^{-|t|} e^{-2i\pi tx} dx \\ &= \int_{-\infty}^0 e^t e^{-2i\pi tx} dx + \int_0^{+\infty} e^{-t} e^{-2i\pi tx} dx \\ &= \left[\frac{e^{(1-2i\pi x)t}}{1-2i\pi x} \right]_{-\infty}^0 + \left[-\frac{e^{-(1+2i\pi x)t}}{1+2i\pi x} \right]_0^{+\infty} \\ &= \frac{1}{1-2i\pi x} + \frac{1}{1+2i\pi x} \\ &= \frac{2}{1+(2\pi x)^2}\end{aligned}$$

Alors par changement d'échelle, on a que

$$\begin{aligned}\varphi_X(t) &= \mathcal{F}^{-1}(f)\left(\frac{t}{2\pi}\right) \\ &= \mathcal{F}^{-1}\left(\frac{1}{2\pi} \mathcal{F}(\psi)\left[\frac{\cdot}{2\pi}\right]\right)\left(\frac{t}{2\pi}\right) \\ &= \mathcal{F}^{-1}(\mathcal{F}(\psi))(t) \\ &= \psi(t) \quad dx\text{-presque partout}\end{aligned}$$

- Soit $X \sim \mathcal{N}(0, 1)$, on va calculer $\varphi_X(t)$ en remarquant qu'il s'agit d'une fonction dérivable, qui vérifie une équation différentielle :

$$\begin{aligned}\varphi'_X(t) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} ixe^{-\frac{x^2}{2}} e^{itx} dx \\ &= \left[-\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}} e^{itx} \right]_{-\infty}^{+\infty} + i^2 t \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}} e^{itx} dx \\ &= -t\varphi_X(t).\end{aligned}$$

Donc il existe une constante C telle que $\varphi_X(t) = Ce^{-\frac{t^2}{2}}$. Comme $\varphi_X(0) = 1$, on a alors $C = 1$, ce qui conclut la preuve. □

On trouvera en exercices le calcul des fonctions caractéristiques de diverses autres lois.

4.1.3 La fonction caractéristique caractérise la loi

Cette section cherche à démontrer l'injectivité de l'application qui à une loi associe sa fonction caractéristique. À première lecture, on pourra aller directement à la partie application.

Proposition 91 :

L'application $X \mapsto \varphi_X$ est injective.

Démonstration.

Soit X une variable aléatoire réelle, de loi \mathbb{P}_X .

On va poser pour $\sigma > 0$ la fonction g_σ correspondante à la densité d'une loi normale centrée d'écart-type σ :

$$g_\sigma : x \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

Comme on connaît sa transformée de Fourier, on a pour tout $u \in \mathbb{R}$ que :

$$g_\sigma(u) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{iut} g_{1/\sigma}(t) dt.$$

L'idée de la preuve est de se servir de g_σ pour approcher par convolution la loi de probabilité.

On définit la fonction réelle suivante :

$$g_\sigma * \mathbb{P}_X : x \mapsto \int_{\mathbb{R}} g_\sigma(x - y) d\mathbb{P}_X(y).$$

Avec le théorème de Tonelli-Fubini, on peut déduire du fait que g_σ est une densité de probabilité que

$$\int_{\mathbb{R}} g_\sigma * \mathbb{P}_X(x) dx = \int \int g_\sigma(x - y) d\mu dx = 1.$$

En tant qu'intégrale d'une fonction positive, $g_\sigma * \mathbb{P}_X$ est positive. On peut donc en déduire que $g_\sigma * \mathbb{P}_X$ est une densité de probabilité.

Notons $V_\sigma := g_\sigma * \mathbb{P}_X$ la loi de probabilité associée.

On va montrer que V_σ converge vers \mathbb{P}_X dans un certain sens faible. Plus précisément, on va montrer que pour toute fonction continue bornée f ,

$$\lim_{\sigma \rightarrow 0} \int_{\mathbb{R}} f(x) dv_\sigma(x) = \int_{\mathbb{R}} f(x) d\mathbb{P}_X(x)$$

En effet, si on se donne f continue et bornée, on a bien par convergence dominée que :

$$\int_{\mathbb{R}} f(x) dv_\sigma(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x) g_\sigma(x-y) d\mathbb{P}_X(y) dx = \int_{\mathbb{R}} (g_\sigma * f) d\mathbb{P}_X \xrightarrow{\sigma \rightarrow 0} \int_{\mathbb{R}} f d\mathbb{P}_X,$$

où l'on a utilisé que pour une fonction continue f bornée, et x un réel, $g_\sigma * f(x) \xrightarrow{\sigma \rightarrow 0} f(x)$ démontré dans l'annexe B

Le dernier argument nécessaire à la preuve est alors que la densité $g_\sigma * \mathbb{P}_X(x)$ ne dépend en fait que de la fonction caractéristique :

$$g_\sigma * \mathbb{P}_X(x) = \int_{\mathbb{R}} g_\sigma(x-y) d\mathbb{P}_X(y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma}} g_{1/\sigma}(t) e^{it(x-y)} dt \right) d\mathbb{P}_X(y) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma}} e^{itx} g_{1/\sigma}(t) \varphi_X(-t) dt$$

De tout ce qui précède, on peut conclure que si X et Y sont deux variables aléatoires ayant la même fonction caractéristique, on a pour toute fonction continue et bornée que

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

En passant à la limite avec des fonctions tendant vers des indicatrices de fermées, on trouve alors que $\mathbb{P}_X = \mathbb{P}_Y$. □

La fonction caractéristique caractérise la loi !

Applications : Somme de deux variables aléatoires de loi de Poisson

On suppose X_1 et X_2 indépendantes et $X_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 \sim \mathcal{P}(\lambda_2)$. On veut connaître la loi de $X_1 + X_2$. Alors, en utilisant les propriétés des fonctions caractéristiques, on obtient que :

$$\varpi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t) = e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{it}-1)} = e^{(\lambda_1+\lambda_2)(e^{it}-1)}$$

qui est la loi caractéristique d'une loi de Poisson $\mathcal{P}(\lambda_1 + \lambda_2)$. Donc $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

4.1.4 Exercice

Pour les exercices qui suivent, on pensera à utiliser la proposition 90

Exercice 1 :

Montrer que si $X \sim \mathcal{Bin}(n, p)$, alors $\varphi_X(t) = (pe^{it} + q)^n$.

Exercice 2 :

Montrer que si $X \sim \mathcal{U}(a, b)$, alors $\varphi_X(t) = \frac{\sin\left(t \frac{b-a}{2}\right)}{t \frac{b-a}{2}} e^{it \frac{a+b}{2}}$.

Exercice 3 :

Montrer que si $X \sim \mathcal{E}(\lambda)$, alors $\varphi_X(t) = \frac{\lambda}{\lambda - it}$.

Exercice 4 :

Montrer que si $X \sim \mathcal{N}(m, \sigma)$, alors $\varphi_X(t) = e^{imt} e^{-\frac{t^2 \sigma^2}{2}}$.

Exercice 5 :

Soient X_1 et X_2 deux variables aléatoires indépendantes

- Montrer que si $X_1 \sim \mathcal{Bin}(n_1, p)$ et $X_2 \sim \mathcal{Bin}(n_2, p)$, alors $X_1 + X_2 \sim \mathcal{Bin}(n_1 + n_2, p)$. Est ce surprenant ?
- Montrer que si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$, alors $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.
- Montrer que si $X_1 \sim \mathcal{E}(\lambda)$ et $X_2 \sim \mathcal{E}(\lambda)$, alors $X_1 + X_2 \sim \gamma(2, \lambda)$.
- Plus généralement, montrer que si $X_1 \sim \gamma(r_1, \lambda)$ et $X_2 \sim \gamma(r_2, \lambda)$, alors $X_1 + X_2 \sim \gamma(r_1 + r_2, \lambda)$.

4.2 Les divers modes de convergences des variables aléatoires

Le concept de variables aléatoires est vaste, et très riche. Il permet de nombreux modes de convergence, qui permettent d'étudier plus profondément leurs relations.

4.2.1 Convergence en loi

On va commencer par définir la convergence en loi, qui est la convergence la moins restrictive, et donc la plus simple à obtenir.

Définition 92 : (Convergence en loi)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On note F_n (respectivement F) la fonction de répartition de X_n (resp. X).

On dit que la suite (X_n) converge en loi vers X , que l'on note $X_n \xrightarrow{\text{loi}} X$ ou $\lim_{n \rightarrow +\infty} X_n \stackrel{\text{loi}}{=} X$, si :

$$\lim_{n \rightarrow +\infty} F_n(t) = F(t), \quad \text{en tout point } t \in \mathbb{R} \text{ où } F \text{ est continue.}$$

Remarque : Comme la convergence n'utilise en fait que la loi de X , on peut parler de convergence en loi même si les variables aléatoires sont définies sur des univers différents. C'est le seul mode de convergence ayant cette possibilité parmi celles que nous allons présenter.

Cas particulier : Variables à valeurs entières

Supposons que toutes les variables aléatoires X_n sont à valeurs dans \mathbb{N} . Alors la suite $(X_n)_{n \geq 1}$ converge en loi vers X si pour tout $k \in \mathbb{N}$ on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P}[X_n = k] = \mathbb{P}[X = k]$$

Cas particulier : Limite à densité

Si la suite (X_n) converge en loi vers X et si X admet une densité, alors pour tous $a < b$ on a :

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_n \in]a, b]) = \mathbb{P}(X \in]a, b])$$

La suite de cette section est technique, et servira à démontrer des théorèmes fondamentaux de convergence en loi.

Lemme : Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé, $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles et X une autre variable aléatoire réelle.

Alors X_n converge en loi vers X si et seulement si pour toute fonction f uniformément continue bornée,

$$\lim_n \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$$

Démonstration.

(\Leftarrow) Soit x_0 un point de continuité de F_X , on peut trouver pour $\epsilon > 0$ des fonctions uniformément continue et bornée f et g telles que :

$$\mathbb{1}_{]-\infty; x_0 - \epsilon]} f \leq \mathbb{1}_{]-\infty; x_0]} \leq g \leq \mathbb{1}_{]-\infty; x_0 + \epsilon]}$$

(prendre par exemple $f=1$ jusqu'à $x_0 - \epsilon$, puis $x \mapsto \frac{1}{\epsilon}(x - x_0)$ sur $[x_0 - \epsilon; x_0]$, puis 0, et la translatée de cette fonction par ϵ pour g).

Alors en intégrant la troisième inégalité en X_n , on obtient avec l'hypothèse que :

$$F_{X_n}(x_0) = \mathbb{E}[\mathbb{1}_{]-\infty; x_0]}] \leq \mathbb{E}[g(X_n)] \xrightarrow{n \rightarrow \infty} \mathbb{E}[g(X)]$$

Avec la quatrième en X , on obtient que $\mathbb{E}[g(X)] \leq F_X(x_0 + \epsilon)$

D'où en passant à la limite supérieure,

$$\limsup_{n \rightarrow +\infty} F_{X_n}(x_0) \leq F_X(x_0 + \epsilon)$$

Avec le même type de manipulation avec la première et deuxième inégalité, on trouve que $\liminf_{n \rightarrow +\infty} F_{X_n}(x_0) \geq F_X(x_0 - \epsilon)$. Comme F_X est continu en x_0 , on a donc par passage à la limite en ϵ que $F_X(x_0) = \liminf_{n \rightarrow +\infty} F_{X_n}(x_0) = \limsup_{n \rightarrow +\infty} F_{X_n}(x_0)$, c'est-à-dire que $\lim_{n \rightarrow +\infty} F_{X_n}(x_0) = F_X(x_0)$.

(\Rightarrow) On suppose que (X_n) converge en loi vers X , et on se donne f une fonction uniformément continue et bornée. Quitte à prendre les parties positive et négative de f , on peut supposer f positive.

Idée de la preuve : On va pour cette preuve choisir une subdivision adaptée à la fonction (en ses points de discontinuité, et considérer les sommes de Darboux, en espérant avoir une convergence.

Par uniforme continuité, il existe $u_k > 0$ tel que si $|x - y| < u_k$, alors $|f(x) - f(y)| < \frac{1}{k}$. Comme l'ensemble des points de discontinuité des fonctions de répartition, on peut trouver une suite de subdivision $(x_{k,p})_{k \in \mathbb{N}^*, p \in \mathbb{Z}}$ de points de continuités de F_X tels que

$$x_{k,p} < x_{k,p+1}, \quad |x_{k,p} - x_{k,p+1}| \leq u_k \quad \text{et} \quad \mathbb{R} = \bigcup_{p \in \mathbb{Z}}]x_{k,p}; x_{k,p+1}].$$

Remarquons qu'alors

$$f - \frac{1}{k} \leq \sum_{p \in \mathbb{Z}} \inf_{t \in]x_{k,p}; x_{k,p+1}]} [f(t)] \mathbb{1}_{]x_{k,p}; x_{k,p+1}]} \leq f \leq \sum_{p \in \mathbb{Z}} \sup_{t \in]x_{k,p}; x_{k,p+1}]} [f(t)] \mathbb{1}_{]x_{k,p}; x_{k,p+1}]} \leq f + \frac{1}{k} \quad (\text{A})$$

Quitte à se restreindre à un compact où la somme est une somme finie puis à appliquer le théorème de convergence dominée, on peut intervertir l'espérance et la somme dans l'équation (A) évalué en X_n , et on obtient :

$$\sum_{p \in \mathbb{Z}} \inf_{t \in]x_{k,p}; x_{k,p+1}]} [f(t)] [F_{X_n}(x_{k,p+1}) - F_{X_n}(x_{k,p})] \leq \mathbb{E}[f(X_n)] \leq \sum_{p \in \mathbb{Z}} \sup_{t \in]x_{k,p}; x_{k,p+1}]} [f(t)] [F_{X_n}(x_{k,p+1}) - F_{X_n}(x_{k,p})]$$

Soit en passant à la limite inférieure/supérieure en n (théorème de domination sous l'intégrale, où l'on majore f par sa borne supérieure) :

$$\begin{aligned}
 \sum_{p \in \mathbb{Z}} \inf_{t \in [x_{k,p}; x_{k,p+1}]} [f(t)] [F_X(x_{k,p}) - F_X(x_{k,p+1})] &\leq \liminf \mathbb{E}[f(X_n)], \\
 &\leq \limsup \mathbb{E}[f(X_n)], \\
 &\leq \sum_{p \in \mathbb{Z}} \sup_{t \in [x_{k,p}; x_{k,p+1}]} [f(t)] [F_X(x_{k,p}) - F_X(x_{k,p+1})].
 \end{aligned}$$

Donc en intégrant une dernière fois l'équation (A) évalué en X , on a que

$$\mathbb{E}[f(X)] - \frac{1}{k} \leq \sum_{p \in \mathbb{Z}} \inf_{t \in [x_{k,p}; x_{k,p+1}]} [f(t)] [F_X(x_{k,p}) - F_X(x_{k,p+1})]$$

et aussi que

$$\sum_{p \in \mathbb{Z}} \sup_{t \in [x_{k,p}; x_{k,p+1}]} [f(t)] [F_X(x_{k,p}) - F_X(x_{k,p+1})] \leq \mathbb{E}[f(X)] - \frac{1}{k}$$

Donc

$$\mathbb{E}[f(X)] - \frac{1}{k} \leq \liminf \mathbb{E}[f(X_n)] \leq \limsup \mathbb{E}[f(X_n)] \leq \mathbb{E}[f(X)] + \frac{1}{k}$$

Et passant alors à la limite en k , on a alors bien que $\mathbb{E}[f(X_n)]$ converge et est de limite $\mathbb{E}[f(X)]$. □

Cette caractérisation permet de prouver le théorème suivant :

Théorème 93 : Paul Lévy

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On note φ_n (respectivement φ) la fonction caractéristique de X_n (resp. X). Alors on a l'équivalence suivante :

$$\lim_{n \rightarrow \infty} X_n \stackrel{loi}{=} X \quad \Leftrightarrow \quad \forall t \in \mathbb{R}, \lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$$

Démonstration.

On va utiliser caractérisation 92 pour montrer cette équivalence :

(\Rightarrow) Le sens direct est immédiat puisque la fonction $x \mapsto e^{itx}$ est continue et bornée.

(\Leftarrow) Considérons f une fonction uniformément continue et bornée, et supposons que pour tous $u \in \mathbb{R}$, $\varphi_{X_n}(u) \rightarrow \varphi_X(u)$.

On reprend les notations de la proposition 91, avec

$$g_\sigma : x \mapsto \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}.$$

On a vu qu'alors

$$g_\sigma(u) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} e^{iut} g_{1/\sigma}(t) dt,$$

et que pour Y une variable aléatoire,

$$\int_{\mathbb{R}} (g_\sigma * f) d\mathbb{P}_Y \xrightarrow{\sigma \rightarrow 0} \int_{\mathbb{R}} f d\mathbb{P}_Y.$$

Soit $\epsilon > 0$

On découpe alors l'espérance

$$\mathbb{E}[f(X_n) - f(X)] = \mathbb{E}[f(X_n) - g_\sigma * f(X_n)] + \mathbb{E}[g_\sigma * f(X_n) - g_\sigma * f(X)] + \mathbb{E}[g_\sigma * f(X) - f(X)]$$

Comme f est uniformément continue (voir l'annexe B), il existe $\sigma > 0$ tel que $\|f - g_\sigma * f\|_\infty \leq \epsilon$.

Écrivons pour $\sigma > 0$ fixé que

$$\begin{aligned} \mathbb{E}[g_\sigma * f(X_n)] &= \int_{\mathbb{R}} g_\sigma * f(u) d\mathbb{P}_{X_n} = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x) g_\sigma(u-x) dx d\mathbb{P}_{X_n}(u), \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f(x)}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{i(u-x)t} g_{1/\sigma}(t) dt dx d\mathbb{P}_{X_n}(u), \\ &= \int_{\mathbb{R}} \frac{f(x)}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{-ixt} g_{1/\sigma}(t) e^{-itx} g_{1/\sigma}(t) \varphi_{X_n}(t) dt dx, \\ &\xrightarrow{n \rightarrow +\infty} \int_{\mathbb{R}} \frac{f(x)}{\sqrt{2\pi\sigma}} \int_{\mathbb{R}} e^{-ixt} g_{1/\sigma}(t) e^{-itx} g_{1/\sigma}(t) \varphi_X(t) dt dx \quad \text{par convergence dominée,} \\ &\xrightarrow{n \rightarrow +\infty} \mathbb{E}[g_\sigma * f(X)] \end{aligned}$$

Soit alors N tel que pour $n \geq N$, $|\mathbb{E}[g_\sigma * f(X_n) - g_\sigma * f(X)]| \leq \epsilon$. Alors pour un tel $n \geq N$, par inégalité triangulaire, et notre choix de σ on a que :

$$\begin{aligned} |\mathbb{E}[f(X_n) - f(X)]| &\leq \|f - g_\sigma * f\|_\infty + |\mathbb{E}[g_\sigma * f(X_n) - g_\sigma * f(X)]| + \|f - g_\sigma * f\|_\infty \\ &\leq 3\epsilon \end{aligned}$$

Donc $|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \rightarrow 0$, ce qui nous donne par la caractérisation que $X_n \xrightarrow{\text{loi}} X$. □

La convergence en loi des X_n vers X est équivalente à la convergence simple de la suite (φ_n) des fonctions caractéristiques vers celle de X .

La fonction caractéristique remplace convergence en loi par de la convergence simple, c'est plus simple !

Exemple d'application :

Pour $m \geq 2$, considérons X_m une variable aléatoire de loi uniforme sur $\{0, \frac{1}{m}, \dots, \frac{m-1}{m}\}$.

Soit X une variable aléatoire de loi uniforme sur $[0, 1]$. Alors suite $(X_m)_{m \geq 2}$ converge en loi vers X quand $m \rightarrow +\infty$.

En effet, on a pour $t \in \mathbb{R}$, en tant que limite de somme de Riemann, que :

$$\varphi_n(t) = \sum_{k=0}^{m-1} e^{it \frac{k}{m}} \frac{1}{m} \xrightarrow{n \rightarrow +\infty} \int_0^1 e^{itx} dx = \varphi(t).$$

4.2.2 Convergence en norme L^p

On peut aussi voir certaines des variables aléatoires comme des fonctions intégrables, ce qui permet de considérer un autre mode de convergence. Pour les variables aléatoires ayant des moments d'ordre p , on introduit la convergence en norme L^p . L'on se rappelle pour cela que les espaces L^p constituent des espaces vectoriels normés complets pour la norme $X \mapsto \mathbb{E}(|X|^p)$.

Définition 94 : (Convergence en norme L^p)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On suppose de plus que les variables aléatoires X_n et X_0 ont toutes des moments d'ordres p (i.e. sont de puissance $p^{\text{ième}}$ intégrable (noté $X \in L^p$)).

Alors, on dira que $(X_n)_{n \geq 1}$ converge dans L^p , ou encore converge en moyenne d'ordre p vers X_0 , que l'on note $X_n \xrightarrow{L^p} X_0$ si

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X_0|^p) = 0.$$

Remarque : Attention, il n'est pas suffisant d'avoir $\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n|^p) = \mathbb{E}(|X_0|^p)$ pour avoir une convergence en norme L^p .

Les cas particuliers importants sont $p = 1$ et $p = 2$ et nous les traitons séparément.

Convergence dans L^2 (ou convergence en moyenne quadratique) :

Soient X_n et X des variables aléatoires de carré intégrable. Rappelons que la suite $(X_n)_{n \geq 1}$ converge vers X dans L^2 ssi :

$$\lim_{n \rightarrow +\infty} \mathbb{E}(|X_n - X|^2) = 0.$$

Par exemple, soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires telles que pour tout n , $X_n \sim \text{Bin}(n, p)$. Pour $n \geq 1$, on pose $Y_n = \frac{X_n}{n}$. Ici, Y_n représente la proportion de succès lors des expériences de Bernoulli associées à X_n . On peut donc conjecturer que sa limite dans L^2 va être p . Calculons alors :

$$\mathbb{E}(|Y_n - p|^2) = \mathbb{E}\left(\frac{1}{n^2} |X_n - np|^2\right) = \frac{1}{n} V(X_n) \rightarrow 0.$$

On obtient donc que la suite (Y_n) converge bien vers p dans L^2 .

Conséquence de la convergence dans L^1 :

La suite $(X_n)_{n \geq 1}$ converge vers X dans L^1 ssi :

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

Cela implique en particulier que : $\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

Proposition 95 : (Admise)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire réelle. Si X a un moment d'ordre $r \geq 1$, alors pour tout $1 \leq s \leq r$, X a un moment d'ordre s . De plus, si $(X_n)_{n \geq 1}$ converge dans L^r vers X , alors $(X_n)_{n \geq 1}$ converge dans L^s vers X .

Remarque : En particulier, une convergence quadratique implique une convergence dans L^1 .

4.2.3 Convergence en probabilité et presque sûr

Un autre moyen de demander un rapprochement des variables aléatoires consiste à demander que pour tout écart, la probabilité d'être plus éloigné que cet écart tende vers 0.

Définition 96 : (Convergence en probabilité)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On dira que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , que l'on note $X_n \xrightarrow{\text{proba}} X$, si

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

Exemple : On considère la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$, tels que la loi de X_n admettent la densité

$$f_n = \frac{ne^{-nx}}{(1 + e^{-nx})^2}$$

Pour $x \neq 0$ fixé, on s'aperçoit que $f_n(x)$ tend vers 0 quand n tend vers l'infini et que le pic de la distribution en 0 tend vers l'infini.

Ceci nous amène à penser que la limite en loi serait la mesure de Dirac en 0, ou encore qu'il y aurait une convergence vers la variable aléatoire nulle. En effet, si on calcule pour $\epsilon > 0$ fixé :

$$\mathbb{P}(|X_n| \leq \epsilon) = \int_{-\epsilon}^{\epsilon} f_n(x) dx = \left[-\frac{1}{1 + e^{-nx}} \right]_{-\epsilon}^{\epsilon} = \frac{1 - e^{n\epsilon}}{1 + e^{n\epsilon}} \xrightarrow{n \rightarrow +\infty} 1$$

On en déduit que la probabilité du complémentaire tend vers 0, ce qui assure la convergence en probabilité de la suite (X_n) vers la variable aléatoire 0.

Proposition 97 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On suppose que (X_n) converge dans L^1 vers une variable X ,

Alors (X_n) converge en probabilité vers X

Pour démontrer cela, nous allons utiliser le lemme suivant :

Lemme : (Markov)

Si Y est une variable aléatoire réelle ayant un moment d'ordre 1, alors pour $\epsilon > 0$, on a que

$$\mathbb{P}(|Y| \geq \epsilon) \leq \frac{\mathbb{E}(|Y|)}{\epsilon}$$

Démonstration.

Soit $\epsilon > 0$.

Preuve du lemme : On a l'inégalité de fonction $\epsilon \mathbf{1}_{\{\epsilon \leq |Y|\}} \leq |Y|$

Donc en intégrant, on a bien que $\epsilon \mathbb{P}(|Y| \geq \epsilon) \leq \mathbb{E}(|Y|)$.

Prouvons à présent la proposition. D'après le lemme, pour n donné, on obtient que

$$\mathbb{P}(|X_n - X| \geq \epsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\epsilon}$$

Or par convergence L^1 , $\mathbb{E}(|X_n - X|) \rightarrow 0$.

Donc $\mathbb{P}(|X_n - X| \geq \epsilon) \xrightarrow{n \rightarrow \infty} 0$. □

Attention cependant à ne pas confondre la convergence en probabilité avec la notion suivante de convergence :

Définition 98 : (Convergence presque sûre)

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On dira que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X , que l'on note $X_n \xrightarrow{p.s.} X$ si l'ensemble

$$A = \{\omega \in \Omega, \lim_{n \rightarrow +\infty} X_n = X\} \text{ est de probabilité } 1 : \mathbb{P}(A) = 1$$

Proposition 99 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

La convergence presque sûre des X_n vers X est équivalente à la condition

$$\forall \epsilon > 0, \mathbb{P}[\sup_{k \geq n} |X_k - X| > \epsilon] \xrightarrow{n \rightarrow +\infty} 0.$$

C'est-à-dire que la variable aléatoire $\sup_{k \geq n} |X_k - X|$ converge en probabilité vers 0.

Démonstration.

Supposons que X_n converge presque sûrement vers X et soit $\epsilon > 0$.

Alors, vu que

$$\bigcap_{n \in \mathbb{N}} \{\omega, \sup_{k \geq n} |X_k - X|(\omega) > \epsilon\} \subset \{\omega, \lim_{n \rightarrow +\infty} X_n(\omega) \text{ n'existe pas ou n'est pas } X(\omega)\} = {}^c A$$

on a par continuité monotone de \mathbb{P} que

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\sup_{k \geq n} |X_k - X| > \epsilon] = 0$$

et ceci est vrai pour tout ϵ .

Réciproquement, supposons que pour tout ϵ ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}[\sup_{k \geq n} |X_k - X| > \epsilon] = 0$$

En étant plus précis que précédemment (et en remplaçant dans la définition de la limite le quantificateur sur tous les réels positifs par un quantificateur sur tous les inverses d'entiers), on obtient que

$$\bigcup_{p \in \mathbb{N}} \bigcap_{n \in \mathbb{N}} \left\{ \sup_{k \geq n} |X_k - X| > \frac{1}{p+1} \right\} = \left\{ \lim_{n \rightarrow +\infty} X_n \text{ n'existe pas ou n'est pas } X \right\} = {}^c A$$

Or par hypothèse et continuité monotone,

$$\mathbb{P} \left[\bigcap_{n \in \mathbb{N}} \left\{ \sup_{k \geq n} |X_k - X| > \frac{1}{p+1} \right\} \right] = \lim_{n \rightarrow +\infty} \mathbb{P} \left[\sup_{k \geq n} |X_k - X| > \frac{1}{p+1} \right] = 0$$

Donc en particulier,

$$\mathbb{P}({}^c A) \leq \sum_{p \in \mathbb{N}} \mathbb{P} \left[\bigcap_{n \in \mathbb{N}} \left\{ \sup_{k \geq n} |X_k - X| > \frac{1}{p+1} \right\} \right] = 0$$

□

Corollaire 4.2.1. Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles. Si X_n converge presque sûrement vers une variable aléatoire X , alors X_n converge en probabilité vers X .

Démonstration. On a pour $\epsilon > 0$ que $0 \leq \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(\sup_{k \geq n} |X_k - X| > \epsilon) \xrightarrow{n \rightarrow +\infty} 0$

□

Remarque : La réciproque n'est pas complètement vraie, mais l'on peut se consoler avec le fait qu'elle est vraie pour une suite extraite. Si l'on suppose que X_n converge en probabilité vers X , alors il existe une extraction (fonction strictement croissante sur les entiers positifs) Ψ telle que $X_{\psi(n)}$ converge presque sûrement vers X .

En effet, prenons X_n converge en probabilité vers X . Pour $k \in \mathbb{N}^*$ avec $\epsilon = \frac{1}{k}$, par définition de la limite en 0 avec l'intervalle $] -\frac{1}{k^2}; \frac{1}{k^2} [$, on peut trouver $N(k)$ tel que :

$$\forall n \geq N(k), \mathbb{P}[|X_n - X| > \frac{1}{k}] \leq \frac{1}{k^2}.$$

Alors, on peut construire Ψ extraction par récurrence comme $\Psi(k+1) = \max[N(k+1), \Psi(k) + 1]$.

Si on pose $A_k := \{|X_{\psi(k)} - X| > \frac{1}{k}\}$, alors la série $\sum_{k \in \mathbb{N}^*} \mathbb{P}(A_k) = \sum_{k \in \mathbb{N}^*} \frac{1}{k^2}$ converge.

Donc par le premier point du lemme de Borel-Canteli (proposition 14), $\mathbb{P}(A_k \text{ arrive infiniment souvent}) = 0$, c'est-à-dire que $\mathbb{P}({}^c A_k \text{ arrive un nombre fini de fois}) = 1$

Donc

$$\exists M(\omega) \forall n \geq M, |X_{\psi(n)}(\omega) - X(\omega)| \leq \frac{1}{n} \quad \text{est presque sûr en } \omega$$

C'est-à-dire que $X_{\psi(n)} \xrightarrow{p.s.} X$.

Proposition 100 :

Soit $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé et $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles.

On suppose que (X_n) converge en probabilité vers une variable X ,

Alors (X_n) converge en loi vers X

Démonstration.

On va utiliser la caractérisation donnée dans la proposition 92.

Soit donc f une fonction continue à support compact.

Soit $\epsilon > 0$, f est continue à support compact, donc uniformément continue, et donc :

$$\exists \delta > 0, |x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon.$$

De plus, par convergence en probabilité, il existe n_0 tel que pour $n \geq n_0$,

$$\mathbb{P}(|X_n - X| < \delta) \leq \frac{\epsilon}{1 + 2\|f\|_\infty}$$

De plus, on peut découper l'intégrale au sens de Lebesgue comme :

$$\begin{aligned} \mathbb{E}[f(X_n)] - \mathbb{E}[f(X)] &= \mathbb{E}[f(X_n) - f(X)] \\ &= \int_{|X_n - X| < \delta} f(X_n(\omega)) - f(X(\omega)) d\mathbb{P}(\omega) + \int_{|X_n - X| \geq \delta} f(X_n(\omega)) - f(X(\omega)) d\mathbb{P}(\omega) \end{aligned}$$

Donc en regroupant ces arguments, pour $n \geq n_0$:

$$\begin{aligned} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq \int_{|X_n - X| < \delta} |f(X_n(\omega)) - f(X(\omega))| d\mathbb{P}(\omega) + \int_{|X_n - X| \geq \delta} |f(X_n(\omega)) - f(X(\omega))| d\mathbb{P}(\omega) \\ &\quad (\text{Par inégalité triangulaire}) \\ &\leq \epsilon \times \mathbb{P}(\Omega) + 2\|f\|_\infty \mathbb{P}(|X_n - X| \geq \delta) \\ &\leq \epsilon + \epsilon = 2\epsilon \\ &\quad (\text{Car } n \geq n_0) \end{aligned}$$

Donc $\lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$, ce qui nous donne exactement la caractérisation.

Donc $X_n \xrightarrow{\text{Loi}} X$. □

Remarque : Et dans l'autre sens ? On n'a que la réciproque partielle suivante : Si $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles converge en loi vers une variable aléatoire constante $b \in \mathbb{R}$, alors elle converge en probabilité.

En effet, on commence par noter F_n la fonction de répartition de X_n et F_a la fonction de répartition de la variable constante égale à a ($F_a = \mathbb{1}_{x \geq a}$).

Pour $\epsilon > 0$, comme les fonctions F_n et F sont toutes des fonctions croissantes, les ensembles de leurs points de discontinuité sont tous dénombrables, donc l'union de ces ensembles est dénombrable. En particulier, il existe un réel $\epsilon > \eta > 0$ tels que toutes les fonctions de répartitions soient continues en $a - \epsilon - \eta$. Avec un tel η , on a donc

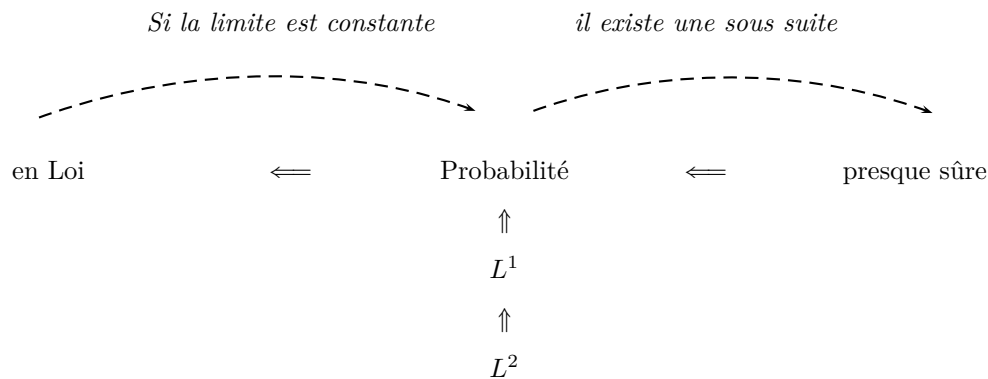
que :

$$\begin{aligned}\mathbb{P}[|X_n - a| > \epsilon] &\leq \mathbb{P}[|X_n - a| > \epsilon - \eta] \\&= \mathbb{P}[X_n > a + \epsilon - \eta] + \mathbb{P}[X_n \leq a - \epsilon - \eta] - \mathbb{P}[X_n = a - \epsilon - \eta] \\&= 1 - F_n(a + \epsilon - \eta) + F_n(a - \epsilon + \eta) - 0 \\&\xrightarrow{n \rightarrow +\infty} 1 - F_a(a + \epsilon - \eta) + F_a(a - \epsilon + \eta) \\&= 1 - \mathbb{1}_{x \geq a}(a + \epsilon - \eta) + \mathbb{1}_{x \geq a}(a - \epsilon + \eta) \\&= 1 - 1 + 0 = 0\end{aligned}$$

Comme ceci est vrai pour tout ϵ , X_n converge en probabilité vers a

4.2.4 Résumé des relations entre les divers modes de convergence

Résumons les diverses implications entre les modes de convergences dans un diagramme dans lequel une flèche (mode 1 \Rightarrow mode 2) veut dire qu'une suite convergente pour le mode 1 est également convergente pour le mode 2 et de même limite. On obtient le graphique suivant :



4.2.5 Exercice

Exercice 6 :

Rappeler la fonction caractéristique d'un variable aléatoire de loi normale et montrer la proposition 76.

Exercice 7 :

On va ici chercher à montrer le théorème 77. Soit $Y = (Y_1, \dots, Y_n)$ un vecteur gaussien centré de covariance K symétrique définie positive.

1. Rappeler la fonction caractéristique d'une variable aléatoire de loi normale ;
2. Montrer que la fonction caractéristique de Y vérifie $\Phi_Y(\bar{t}) = \mathbb{E}(e^{i\langle \bar{t}, Y \rangle}) = e^{-\frac{\langle \bar{t}, \text{Var}(Y) \bar{t} \rangle}{2}}$;
3. Soit $Y = (Y_1, \dots, Y_n)$ un vecteur gaussien centré de covariance K et soit A une matrice symétrique. Montrer que AY est un vecteur gaussien ;
4. Montrer que la matrice de covariance de AY est AKA^t ;
5. Montrer qu'il existe une matrice B tel que BY soit un vecteur gaussien centré de matrice de covariance Id_n ;
6. Montrer le théorème.

4.3 Théorèmes limites

4.3.1 Loi des grands nombres

Nous avons introduit les probabilités par l'intuition que la probabilité d'un événement est liée à la fréquence empirique d'occurrence de l'événement. Nous allons essayer de voir si cette intuition peut être justifiée. L'expérience appartenir à un événement correspond à une variable aléatoire de loi de Bernoulli de paramètre la probabilité de l'événement, et mesurer la fréquence empirique correspond à répéter cette expérience de manière indépendante un grand nombre de fois avant de faire la moyenne des résultats. On voudrait donc un résultat qui, pour une suite (X_i) de variable indépendante identiquement distribuée (*i.e. de même loi*), de loi de Bernoulli de paramètre p , nous dirait pour un mode de convergence

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{mode}} p$$

On a en fait bien plus général que cela, en remarquant que $p = \mathbb{E}(X_1)$, avec le théorème suivant :

Théorème 101 : Loi faible des grands nombres

Soit μ une loi de probabilité sur \mathbb{R} qui admette un moment d'ordre 1. On pose $m = \int_{\mathbb{R}} x d\mu(x)$ sa moyenne. On considère $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi μ .

Alors $\frac{X_1 + \cdots + X_n}{n} \xrightarrow{\text{Proba}} m$, c'est-à-dire

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{X_1 + \cdots + X_n}{n} - m \right| > \epsilon \right] = 0.$$

Démonstration.

On ne va ici prouver qu'une version moins générale, où l'on suppose en plus que les variables aléatoires X_i admettent un moment d'ordre 2 (mais nous ne perdrons pas de généralité, puisque le résultat est impliqué par la loi forte des grands nombres, énoncé après cette preuve et prouvé en annexe A), afin de pouvoir utiliser l'inégalité de Bienaymé-Tchebychev :

Lemme : (*inégalité de Bienaymé-Tchebychev*) Soit X une variable aléatoire qui admet un moment d'ordre 2 ($\mathbb{E}(X^2) \leq +\infty$). Si l'on pose $m = \mathbb{E}(X)$ et $\sigma^2 = \text{Var}(X)$, alors

$$\forall t > 0, \mathbb{P}[|X - m| \geq t] \leq \frac{\sigma^2}{t^2}$$

Démonstration. Pour montrer ce lemme, il suffit de remarquer que $|X - m|^2$ est intégrable, donc par l'inégalité de Markov (lemme 97), on a bien pour $t > 0$ que :

$$\mathbb{P}[|X - m| \geq t] = \mathbb{P}[|X - m|^2 \geq t^2] \leq \frac{\mathbb{E}(|X - m|^2)}{t^2} = \frac{\sigma^2}{t^2}$$

□

Soit alors $(X_i)_{i \in \mathbb{N}}$ comme dans l'énoncé, et qui admettent un moment d'ordre 2. Soit $n \geq 2$ et $\epsilon > 0$, alors

par l'inégalité de Bienaymé-Tchebychev, on obtient que :

$$\mathbb{P} \left[\left| \frac{X_1 + \dots + X_n}{n} - m \right| > \epsilon \right] \leq \frac{1}{\epsilon^2} \text{Var} \left(\frac{X_1 + \dots + X_n}{n} - m \right)$$

ce qui donne par indépendance que

$$\mathbb{P} \left[\left| \frac{X_1 + \dots + X_n}{n} - m \right| > \epsilon \right] \leq \frac{1}{n^2 \epsilon^2} \left(\sum_{i=1}^n \text{Var}(X_i) \right)$$

et comme les variables ont la même loi,

$$\mathbb{P} \left[\left| \frac{X_1 + \dots + X_n}{n} - m \right| > \epsilon \right] \leq \frac{\text{Var}(X_1)}{n \epsilon^2} \xrightarrow{n \rightarrow +\infty} 0$$

□

On a en fait une version encore plus générale :

Théoreme 102 : Loi forte des grands nombres

Soit μ une loi de probabilité sur \mathbb{R} qui admette un moment d'ordre 1. On pose $m = \int_{\mathbb{R}} x d\mu(x)$ sa moyenne. On considère $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi μ .

Alors

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p.s.} m.$$

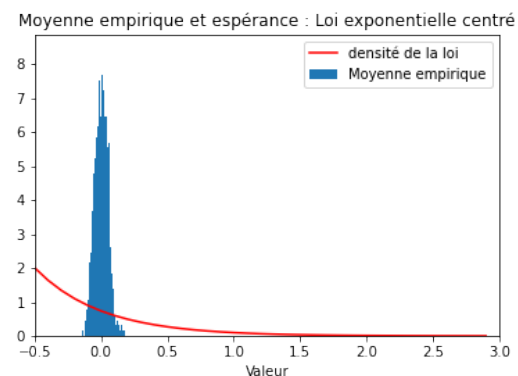
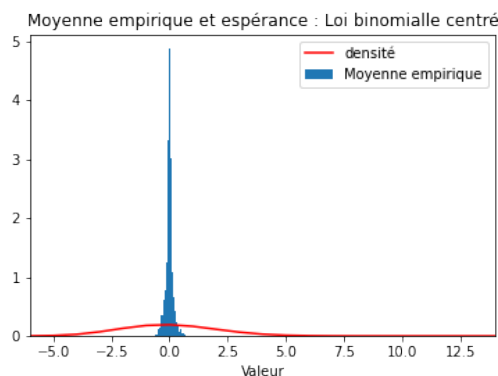
Démonstration. Voir l'annexe A

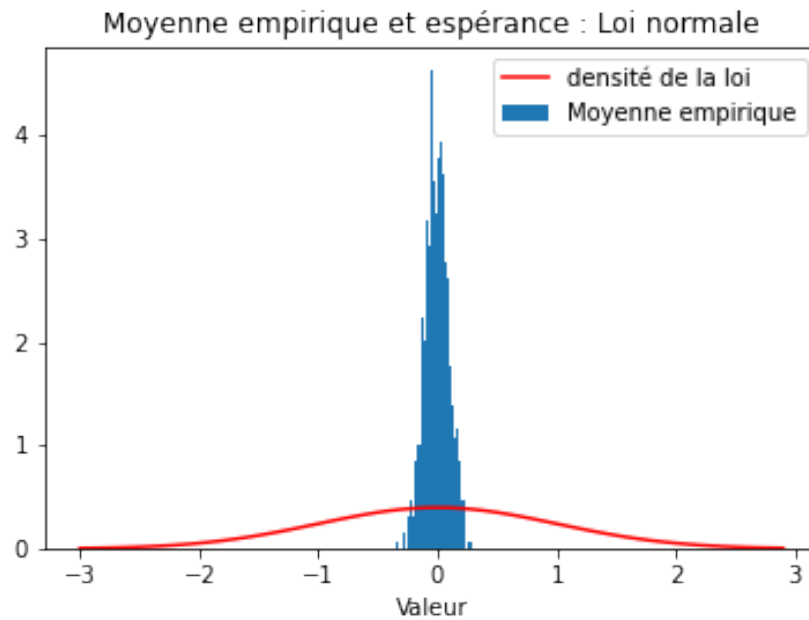
□

Exemple de simulation : On va simuler à 1000 reprises un échantillon de taille 100, noté $\{x_1, \dots, x_{100}\}$ de diverse loi, puis calculer la moyenne empirique des échantillons :

$$\overline{m}(x_1, \dots, x_{100}) = \frac{x_1 + \dots + x_{100}}{100}$$

Puis on cherche à tracer l'histogramme des cinq cents moyennes ainsi obtenues.

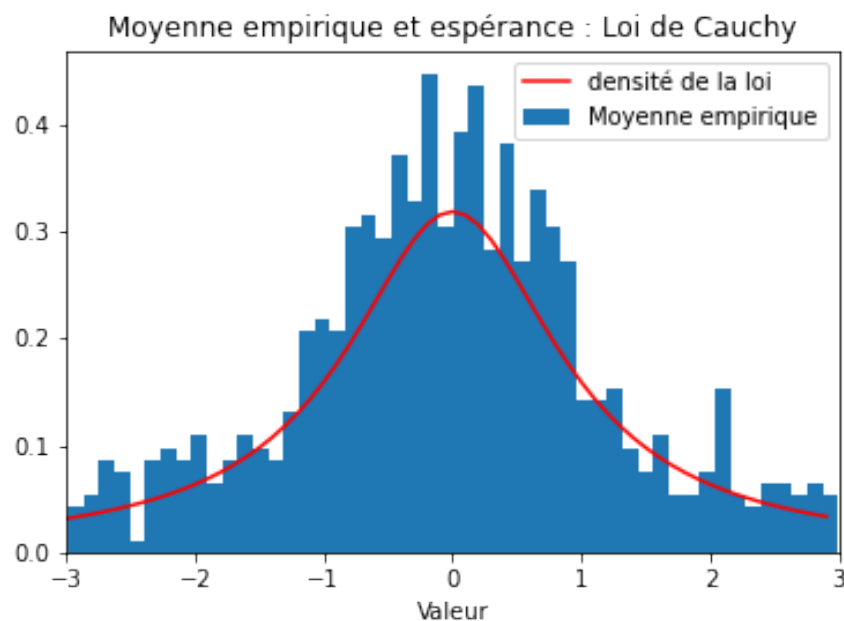




On retrouve bien une moyenne empirique qui se concentre autour de la moyenne de la loi, comme prédit par la loi forte des grands nombres. On semble même décerner une certaine régularité sur la répartition de la moyenne empirique autour de l'espérance.

Avec la loi des grands nombres, la moyenne empirique converge vers l'espérance, c'est trop fort !

Pour la loi de Cauchy, on trouve un résultat bien différent :



La loi de Cauchy n'ayant pas d'espérance, le théorème ne s'applique plus !

4.3.2 Théorème central limite de la limite centrale

Une fois que nous avons établi la loi forte des grands nombres, une question naturelle se pose : pourrions-nous contrôler la vitesse de convergence ? Et encore une fois, la magie s'opère, et la réponse est positive. On va pouvoir dire qu'asymptotiquement, la différence entre la moyenne empirique de n variables indépendantes et leur moyenne se comporte comme une loi normale divisée par la racine carrée de n .

$$\frac{\sum_{i=1}^n X_i}{n} - \mathbb{E}(X_1) \sim \frac{1}{\sqrt{n}} \mathcal{N}(0; 1)$$

Plus précisément, on a le théorème suivant :

Théorème 103 : Théorème central limite

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes, de même loi. On suppose que X_1 admet un moment d'ordre 2, et on note m sa moyenne et σ^2 sa variance. Alors, on a la convergence en loi suivante :

$$\sqrt{n} \cdot \frac{X_1 + \dots + X_n - nm}{\sigma n} = \frac{X_1 + \dots + X_n - nm}{\sigma \sqrt{n}} \xrightarrow[n \rightarrow +\infty]{Loi} \mathcal{N}(0, 1).$$

Démonstration.

Soit $X \sim \mathcal{N}(0, 1)$.

On va pour ce théorème utiliser la caractérisation de Paul Lévy (théorème 93) :

$$\lim_{n \rightarrow \infty} X_n \stackrel{loi}{=} X \quad \Leftrightarrow \quad \forall t \in \mathbb{R}, \lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$$

Remarquons tout d'abord que quitte à soustraire l'espérance, on peut supposer X_1 centré. On pose $Y_n = \frac{X_1 + \dots + X_n}{\sigma \sqrt{n}}$. Commençons donc par calculer sa fonction caractéristique, soit donc $t \in \mathbb{R}$:

$$\varphi_{Y_n}(t) = \mathbb{E}[e^{itY_n}] = \mathbb{E}[e^{i \frac{t}{\sigma \sqrt{n}} \sum_{i=1}^n X_i}] = \mathbb{E}\left[\prod_{i=1}^n e^{i \frac{t}{\sigma \sqrt{n}} X_i}\right]$$

Donc par indépendance,

$$\varphi_{Y_n}(t) = \prod_{i=1}^n \varphi_{X_i}\left(\frac{t}{\sigma \sqrt{n}}\right) = \varphi_{X_1}\left(\frac{t}{\sigma \sqrt{n}}\right)^n$$

Remarquons ensuite que comme X_1 admet un moment d'ordre deux, la fonction caractéristique est deux fois dérivable (proposition 88), et donc admet un développement limité en 0 d'ordre deux de la forme :

$$\varphi_{X_1}(u) = \varphi_{X_1}(0) + \varphi'_{X_1}(0)u + \varphi^{(2)}_{X_1}(0)\frac{u^2}{2} + o(u^2) = 1 - \sigma^2 \frac{u^2}{2} + o(u^2),$$

puisque l'on a supposé X_1 centrée.

On a alors le développement limité suivant :

$$\begin{aligned} \varphi_{Y_n}(t) &= \left(1 - \sigma^2 \frac{t^2}{2\sigma^2 n} + o_n\left(\frac{t^2}{n}\right)\right)^n = e^{n \ln(1 - \frac{t^2}{2n} + o_n(\frac{t^2}{n}))}, \\ &= e^{n(-\frac{t^2}{2n} + o_n(\frac{t^2}{n}))}, \\ &= e^{-\frac{t^2}{2} + o(1)} = e^{-\frac{t^2}{2}} + o_n(1). \end{aligned}$$

D'où on obtient que $\varphi_{Y_n}(t) \rightarrow e^{-\frac{t^2}{2}}$, qui est la fonction caractéristique d'une loi normale centrée et réduite.

Donc d'après le théorème de Paul Lévy (proposition 93), on a bien que la variable aléatoire Y_n converge en loi vers une loi normale centrée et réduite \square

La superposition de causes uniformément petites et indépendantes converge en temps normal vers la loi normale!

En pratique : Lorsque l'on veut utiliser ce théorème pour approcher une somme de variables indépendantes d'écart type connu, la règle empirique est de demander d'avoir *au moins 30 échantillons* ($n \geq 30$). Il existe des lois avec lesquelles la convergence est plus rapide (comme les lois normales avec $n=1$, ou des lois uniformes avec $n=4$), mais en l'absence d'information le critère fonctionne souvent correctement.

Application : On s'intéresse à la loi binomiale $\mathcal{B}in(n, p)$. Si Y_n est une variable aléatoire suivant une telle loi, on sait que Y_n est de même loi $X_1 + \dots + X_n$ où les X_k sont des variables aléatoires indépendantes de même loi de Bernoulli $\mathcal{B}(p)$. On peut donc appliquer le théorème central limite, et on obtient :

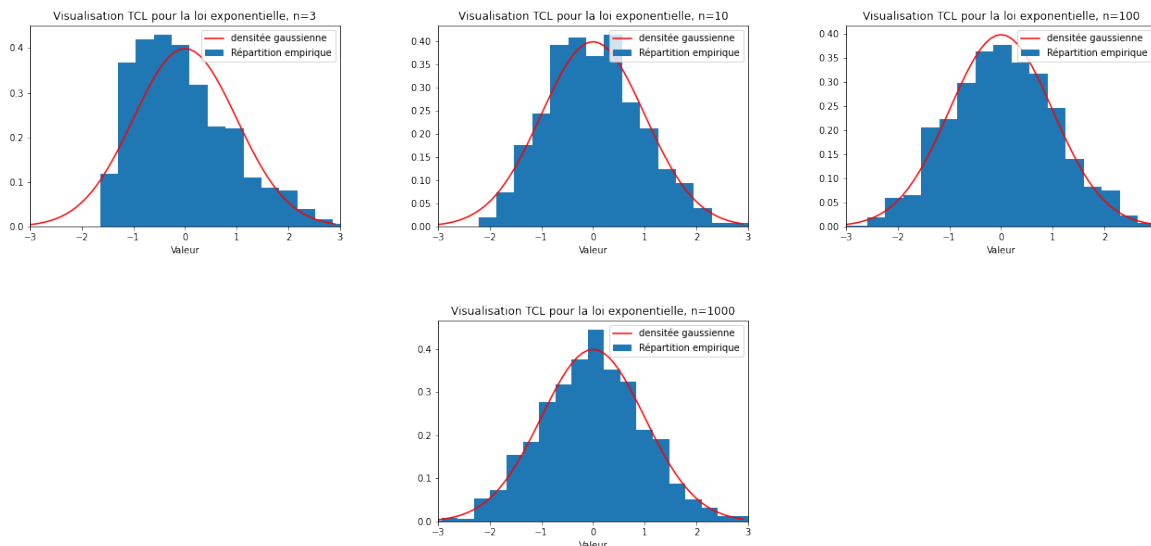
$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0; 1).$$

Ce résultat fournit une approximation de la loi binomiale, dite approximation normale, extrêmement utile en pratique pour simplifier des calculs.

Exemple de simulation : On va simuler à 1000 reprises un échantillon de taille n varié, noté $\{x_1, \dots, x_n\}$ de loi fixée, puis calculer l'écart à de la moyenne empirique à l'espérance renormalisée par la racine de la taille de l'échantillon :

$$\bar{Y}_n(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n - n\mathbb{E}(X_1)}{\sqrt{n}\sigma}$$

Puis on cherche à tracer l'histogramme des cinq cents écarts ainsi obtenu.



Chapitre 5

Solution ou indications des exercices

Vous trouverez dans cette partie des solutions ou des indications pour les exercices dans les sections exercice de ce document. Les solutions seront écrites sans mise en forme, là où *les indications seront mises en avant*.

5.1 Section 1 : espace de probabilité

1) On considère un ensemble fini à 4 éléments $E = \{a_1, a_2, a_3, a_4\}$ munis de tous les événements ($\mathcal{P}(E)$) muni d'une loi de probabilité \mathbb{P} . On veut calculer la probabilité de l'événement élémentaire $\{a_1\}$ dans divers cas. Pour cela, on utilisera que $\mathbb{P}(E) = \sum_{i=1}^4 \mathbb{P}(a_i) = 1$ et que pour tous événements disjoints, $\mathbb{P}(\cup_j A_j) = \sum_j \mathbb{P}(A_j)$.

1. Si $\mathbb{P}(a_2) = \frac{1}{4}$ $\mathbb{P}(a_3) = \frac{1}{6}$ $\mathbb{P}(a_4) = \frac{1}{5}$;
 $\mathbb{P}(a_1) = 1 - \mathbb{P}(a_2) - \mathbb{P}(a_3) - \mathbb{P}(a_4) = \frac{23}{60}$
2. Si $\mathbb{P}(a_1) = 3\mathbb{P}(a_2)$ $\mathbb{P}(a_3) = \frac{1}{4} = \mathbb{P}(a_4)$;
Alors, $\mathbb{P}(a_1) = 1 - \frac{1}{3}\mathbb{P}(a_2) - 2 \cdot \frac{1}{4}$, et donc $\mathbb{P}(a_1) = \frac{3}{8}$
3. Avec $\mathbb{P}(\{a_2, a_3, a_4\}) = 2\mathbb{P}(a_1)$, on trouve que $\mathbb{P}(a_1) = \frac{1}{3}$.
4. Si $\mathbb{P}(a_1) = \mathbb{P}(a_2)$ $\mathbb{P}(a_3) = 2\mathbb{P}(a_2)$ $\mathbb{P}(a_4) = 3\mathbb{P}(a_3)$, alors

$$1 = \mathbb{P}(a_1) + \mathbb{P}(a_1) + 2\mathbb{P}(a_1) + 3 \cdot 2\mathbb{P}(a_1),$$

Ce qui donne $\mathbb{P}(a_1) = \frac{1}{10}$.

2) Utiliser que $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, ainsi que $1 = \mathbb{P}(A) + \mathbb{P}(^c A)$. Ne pas hésiter à faire des dessins des situations.

3) Au bridge, une main est constituée de $n = 13$ cartes prises dans un jeu de $N = 52$ cartes. On suppose que toutes les mains sont équiprobables. Comme l'ordre du tirage est indifférent, on considère comme univers E toutes les suites non ordonnées de 13 cartes sans répétitions, pioché dans les 52 du paquet. On a que :

$$|E| = C_N^n$$

1. Pour piocher exactement un as, on commence par choisir un as parmi les quatre, puis on choisit les 12 cartes restantes de la main parmi toutes les cartes qui ne sont pas des as. On trouve alors

$$\mathbb{P}(\text{"un as exactement"}) = \frac{C_4^1 C_{48}^{12}}{C_{52}^{13}} \simeq 0.439$$

2. Pour piocher au moins un as, on remarque qu'il s'agit du complémentaire de choisir aucun as, qui est de cardinal C_{48}^{13} . On trouve alors

$$\mathbb{P}(\text{"au moins un as"}) = 1 - \frac{C_{48}^{13}}{C_{52}^{13}} \simeq 0.696$$

3. Pour piocher exactement un as et exactement un roi, il s'agit de choisir l'as, puis choisir le roi, et enfin choisir 11 cartes qui ne sont ni un as ni un roi, soit :

$$\mathbb{P}(\text{"un as et un roi exactement"}) = \frac{C_4^1 C_4^1 C_{44}^{11}}{C_{44}^{11}} \simeq 0.193$$

4. Notons E_1 et E_2 les événements "avoir au moins un as" et "avoir au moins un roi". On cherche la probabilité de $E_1 \cap E_2$. Par symétrie du problème, $\mathbb{P}(E_1) = \text{Prob}(E_2)$. Calculons $\mathbb{P}(\overline{E_1 \cup E_2})$. Ceci revient à choisir 13 cartes distinctes qui ne soient ni un roi ni un as, c'est-à-dire parmi 44 cartes. Donc

$$\mathbb{P}(\overline{E_1 \cup E_2}) = \frac{C_{44}^{13}}{C_{48}^{13}}.$$

On trouve alors directement

$$\begin{aligned} \mathbb{P}(E_1 \cap E_2) &= \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cup E_2) \\ &= 2(1 - \frac{C_{48}^{13}}{C_{52}^{13}}) - (1 - \frac{C_{44}^{13}}{C_{48}^{13}}) \\ &= 1 - 2\frac{C_{48}^{13}}{C_{52}^{13}} + \frac{C_{44}^{13}}{C_{48}^{13}} \simeq 0.474 \end{aligned}$$

5. Pour obtenir une main de 13 valeurs différentes, on ne peut choisir que la couleur de chacune des valeurs, ce qui correspond à $(C_4^1)^{13}$ choix distincts si l'on ne compte pas l'ordre. C'est-à-dire :

$$\mathbb{P}(\text{une carte de chaque valeur}) = \frac{(C_4^1)^{13}}{C_{52}^{13}}$$

Statistiques de Maxwell-Boltzmann, Fermi-Dirac et Bose-Einstein :

4) On considère l'univers \mathcal{E}_{MB} des suites ordonnées de n éléments (**avec répétition possible**), à valeurs dans l'ensemble $\{1, \dots, N\}$, munie de la tribu maximale et de la loi uniforme, et on veut déterminer la probabilité d'avoir exactement k éléments de la suite égal à 1, en fonction de n , N et k .

Nombre de cas possibles : Par une récurrence immédiate sur le nombre d'objet, comme chaque objet a N possibilité d'occupation, on a que

$$|\mathcal{E}_{MB}| = N^n.$$

Nombre de cas favorables : Les k positions d'éléments valant 1 sont à choisir parmi n . Il y a C_n^k manière de choisir ces positions.

Pour chacune de ces positions, il faut choisir les valeurs dans l'ensemble $\{2, \dots, N\}$ qui sont dans les $n - k$ positions restantes. Il y a $(N - 1)^{n-k}$ possibilité, et ce pour chaque choix des positions contenant un 1. Donc le nombre de cas favorable N_f vaut

$$N_f = C_n^k (N - 1)^{n-k}$$

Probabilité : La probabilité P_{MB} d'avoir exactement k fois la valeur 1 dans la suite (k objet dans la boîte fixée) est alors

$$P_{MB} = \frac{N_f}{|\mathcal{E}_{MB}|} = \frac{C_n^k (N-1)^{n-k}}{N^n}.$$

Ce qui se simplifie en :

$$P_{MB} = C_n^k \frac{1}{N} \left(1 - \frac{1}{N}\right)^{n-k}$$

5) On considère l'univers \mathcal{E}_{FD} des suites ordonnées de n éléments (**sans répétition** possible), à valeurs dans l'ensemble $\{1, \dots, N\}$, munie de la tribu maximale et de la loi uniforme, et on veut déterminer la probabilité d'avoir une valeur de la suite égal à élément fixé de $\{1, \dots, N\}$, en fonction de n et N .

Nombre de cas possibles : Le nombre de telle suite se calcule en remarquant qu'il s'agit de considérer les suites de n éléments parmi N , sans répétitions

$$|\mathcal{E}_{FD}| = A_N^n = \frac{N!}{(N-n)!}.$$

Nombre de cas favorables : On commence par choisir la position de la suite contenant le numéro choisi, il y a n choix possibles. Pour chacun de ces positions, il s'agit ensuite de répartir dans les $n-1$ positions restantes des nombres différant du numéro choisi, sans répétition, qui comporte A_{N-1}^{n-1} possibilités. En tout, on a donc que le nombre de cas favorable N_f vaut

$$N_f = n A_{N-1}^{n-1}$$

Probabilité : La probabilité P_{FD} vaut :

$$P_{FD} = \frac{N_f}{|\mathcal{E}_{FD}|} = \frac{n A_{N-1}^{n-1}}{A_N^n} = \frac{n}{N}.$$

6) On considère l'univers \mathcal{E}_{BE} des suites ordonnées de N éléments (**avec répétition** de valeur possible) dans $\{0, \dots, n\}$ et de somme n , munie de la tribu maximale et de la loi uniforme, et on veut déterminer la probabilité d'avoir $R = (r_1, \dots, r_N)$ objets respectivement dans les boîtes $(1, \dots, N)$, en fonction de n , N et R . Nombre de cas possibles : Choisir une telle suite revient à choisir $N-1$ éléments a_i parmi $N+n-1$, et à poser r_i le nombre d'éléments strictement compris entre a_{i-1} (où par convention $a_0=0$) et a_i . Donc :

$$|\mathcal{E}_{BE}| = C_{N+n-1}^{N-1}.$$

Nombre de cas favorables : Il n'y a qu'un seul cas favorable.

Probabilité : La probabilité P_{BE} d'avoir R est alors :

$$P_{MB} = \frac{N_f}{|\mathcal{E}_{MB}|} = \frac{1}{C_{N+n-1}^{N-1}}.$$

Probabilité géométrique

7) On considère le polynôme de degré 2, de coefficients $P(X) = X^2 + aX + b$ où a et b sont tirés suivant des lois uniformes et indépendantes sur $[0; 1]$.

- P n'a pas de racine réelle si $\Delta = a^2 - 4b < 0$ (dessiner l'ensemble correspondant dans le carré unité). La probabilité recherchée est alors

$$\mathbb{P}_1 = \frac{2 - \int_{-1}^1 \frac{x^2}{4} dx}{4} = \frac{11}{24}.$$

- La surface sur laquelle P n'a pas de racine réelle et $a > 0$ est alors la moitié de la surface précédente, donc d'aire divisée par deux, donc :

$$\mathbb{P}_2 = \frac{11}{48}.$$

8) Cette probabilité correspond à l'aire sous deux courbes, lesquelles ? Puis découper $[0; 1]$ en trois tiers égaux.

9) On considère que x et y sont tirées suivant des lois uniformes et indépendantes sur $[-n; n]$ et note \underline{x} et \underline{y} les parties fractionnaires de x et y .

- On peut découper l'ensemble en :

$$\{\underline{x} \leq 2\} = \left(\bigcup_{0 \leq k \leq n-1} [k; k + \frac{1}{2}[\cup \bigcup_{0 \leq k \leq n-1} [-k - \frac{1}{2}; k[\right) \times [-n; n],$$

Faites le dessin correspondant.

Comme l'ensemble de droite est d'aire la moitié de l'aire du carré $[-n; n]^2$, sa probabilité vaut $\frac{1}{2}$.

- Par indépendance du tirage de x et y , la probabilité de la conjonction des événements est le produit des probabilités, soit $\frac{1}{4}$.
- Cette probabilité ne dépend pas de n .

Indépendances

10) Montrons que deux événements A et B sont indépendants si et seulement si :

$$\mathbb{P}(A \cap B)\mathbb{P}(\overline{A} \cap \overline{B}) = \mathbb{P}(A \cap \overline{B})\mathbb{P}(\overline{A} \cap B)$$

(\Rightarrow) Supposons A et B indépendants, c'est-à-dire $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

D'après un exercice de cours, A et \overline{B} sont indépendants, ainsi que \overline{A} et B , mais aussi \overline{A} et \overline{B} . On reconnaît alors des deux coté de l'inégalité la quantité

$$\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(\overline{A})\mathbb{P}(\overline{B}).$$

(\Leftarrow) Supposons que l'égalité est vérifiée, et écrivons que

$$\mathbb{P}(\overline{A} \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B);$$

$$\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B);$$

$$\begin{aligned} \mathbb{P}(\overline{A} \cap \overline{B}) &= \mathbb{P}(\overline{A}) + \mathbb{P}(\overline{B}) - \mathbb{P}(\overline{A} \cap \overline{B}); \\ &= 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B) \end{aligned}$$

En remplaçant toutes ces égalités dans l'hypothèse, on a que :

$$\begin{aligned} \mathbb{P}(A \cup B) [1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B)] &= \\ [\mathbb{P}(A) - \mathbb{P}(A \cap B)] [\mathbb{P}(B) - \mathbb{P}(A \cap B)]. \end{aligned}$$

Ce qui se simplifie après développement en

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

Donc A et B sont deux événements indépendants.

11) On jette deux dés à six faces bien équilibrées et on appelle d_i le nombre de points apparaissant sur le i^{eme} dé. Soient E_1 , E_2 et E_3 les événements :

- $E_1 = "d_1 + d_2 \text{ est pair} "$;
- $E_2 = "d_1 = 6 \text{ et } d_2 \text{ est pair} "$;
- $E_3 = "d_1 + d_2 \text{ est divisible par } 4."$

Calculons les probabilités des événements.

- E_1 s'écrit comme réunion des événements disjoints $E_{11} = "d_1 \text{ et } d_2 \text{ sont pairs} "$ et $E_{12} = "d_1 \text{ et } d_2 \text{ sont impairs} "$. Donc $\mathbb{P}(E_1) = \mathbb{P}(E_{11}) + \mathbb{P}(E_{12})$. Or par indépendance, puis par dénombrement,

$$\mathbb{P}(E_{11}) = \mathbb{P}(E_{12}) = \frac{3}{6} \cdot \frac{3}{6} = \frac{1}{4}.$$

$$\text{Donc } \mathbb{P}(E_1) = \frac{1}{2}.$$

- Il n'y a que trois couples de résultats sur les 36 possibles qui mènent à E_2 , donc

$$\mathbb{P}(E_2) = \frac{3}{36}.$$

- En écrivant tous les couples de résultats possibles pour (d_1, d_2) , on trouve 9 éléments dans E_3 , ce qui prouve que :

$$\mathbb{P}(E_3) = \frac{9}{36} = \frac{1}{4}.$$

En remarquant que $E_1 \cap E_2 = E_2$, $E_1 \cap E_3 = E_3$ et que $E_2 \cap E_3 = \{(6, 2), (6, 6)\}$, on a que :

$$\mathbb{P}(E_1 \cap E_2) = \frac{1}{12} \neq \mathbb{P}(E_1)\mathbb{P}(E_2);$$

$$\mathbb{P}(E_1 \cap E_3) = \frac{1}{4} \neq \mathbb{P}(E_1)\mathbb{P}(E_3);$$

$$\mathbb{P}(E_2 \cap E_3) = \frac{2}{36} \neq \mathbb{P}(E_2)\mathbb{P}(E_3).$$

Donc aucunes paires d'événements ne sont indépendantes.

12) Pour ce groupe, les événements A et B sont indépendants seulement pour $n=3$.

De plus

$$\mathbb{P}(C) \geq 0.9 \iff n \geq 9.$$

Conditionnement

13) On jette deux dés à 6 faces équilibrés. Pour trouver les probabilités conditionnelles, on revient à la définition,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ pour } \mathbb{P}(B) \neq 0.$$

On note S la somme des résultats de dés, et D la différence entre le plus grand résultat avec le plus petit résultat

- Il y a onze lancers de dés contenant un 6 (les lancers de résultats (6,1), (6,2), (6,3), (6,4), (6,5), (6,6), (1,6), (2,6), (3,6), (4,6), (5,6)), et parmi ceux-là, seulement 7 sont de somme supérieure à 9. On a donc :

$$\mathbb{P}(S \geq 9 | \text{"au moins un 6"}) = \frac{7/36}{11/36} = \frac{7}{11}.$$

- Il y a aussi onze lancers contenant un 2, et parmi ceux-là seulement deux contiennent aussi un 4. Donc :

$$\mathbb{P}(\text{"au moins un 4"} | \text{"au moins un 2"}) = \frac{2/36}{11/36} = \frac{2}{11}.$$

- Les seuls lancers qui vérifient $D = 4$ sont (5,1), (1,5), (2,6), (6,2).
Parmi ceux-là, aucun n'est de somme 5, donc

$$\mathbb{P}(S = 5 | D = 4) = 0.$$

14) Soient trois portes, l'une cachant un trésor, les deux autres une chèvre, répartis de manière uniforme en probabilité. Vous choisissez une porte à ouvrir, quand un étranger connaissant la répartition ouvre devant vous une des deux autres portes, en choisissant sciemment de révéler une chèvre (en tirant une pièce pour choisir si besoin).

On suppose avoir ouvert la porte 1.

Notons F_i l'événement "le trésor est derrière la porte i ", et O_i "l'étranger ouvre la porte i ". Alors :

$$\mathbb{P}(F_1) = \mathbb{P}(F_2) = \mathbb{P}(F_3) = \frac{1}{3}$$

et comme l'étranger n'ouvre que sur une chèvre,

$$\mathbb{P}(O_2|F_2) = 0;$$

$$\mathbb{P}(O_2|F_3) = 1;$$

et comme il tire une pièce s'il a le choix,

$$\mathbb{P}(O_2|F_1) = \frac{1}{2}.$$

D'après la formule de Bayes,

$$\mathbb{P}(F_3|O_2) = \frac{\mathbb{P}(O_2|F_3)\mathbb{P}(F_3)}{\sum_{i=1}^3 \mathbb{P}(O_2|F_i)\mathbb{P}(F_i)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 + 1 \cdot \frac{1}{3}} = \frac{2}{3}$$

Il y a donc tout intérêt à toujours changer de portes !

15) On considère un groupe de n personnes et on suppose que la date d'anniversaire d'une personne suit une loi uniforme sur 365 jours. En itérant la formule des probabilités totales pour trouver la probabilité du complémentaire, on a que :

$$\mathbb{P}(\text{"il y a deux anniversaires en commun"}) = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{365-i}{365}\right)$$

Surprenamment, on a alors que $n = 23$ suffit à avoir une probabilité supérieure à $\frac{1}{2}$.

5.2 Section 2 : probabilité univariée

Outils du probabiliste

1) Soit X la variable aléatoire de loi définie par :

$$\begin{aligned}\mathbb{P}(X = k) &= ak(8 - k) & k \in \{0, 1, \dots, 8\} \\ \mathbb{P}(X = k) &= 0 & k \notin \{0, 1, \dots, 8\}\end{aligned}$$

On note $S_i = \sum_{p=0}^8 p^i$.

Cette loi doit vérifier que

$$\sum_{p=0}^8 \mathbb{P}(X = k) = 1.$$

Or

$$\begin{aligned}\sum_{p=0}^8 \mathbb{P}(X = k) &= \sum_{p=0}^8 ak(8 - k) = \sum_{p=0}^8 a(8k - k^2) \\ &= a(8S_1 - S_2) = 84a\end{aligned}$$

Donc

$$a = \frac{1}{84}.$$

Calculons la moyenne de X :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{p=0}^8 k\mathbb{P}(X = k) \\ &= a \sum_{p=0}^8 k^2(8 - k) \\ &= a(8S_2 - S_3) = \frac{336}{84} = 4\end{aligned}$$

Calculons de même le moment d'ordre 2 :

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_{p=0}^8 k^2\mathbb{P}(X = k) \\ &= a \sum_{p=0}^8 k^3(8 - k) \\ &= a(8S_3 - S_4) = \frac{1596}{84} = 19\end{aligned}$$

On en déduit immédiatement que

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 19 - 4^2 = 3$$

et donc que l'écart-type vaut $\sigma = \sqrt{3}$.

2) Soit $a > 0$, et soit une variable aléatoire de loi de Rayleigh de paramètre a , c'est-à-dire de fonction de répartition

$$F(x) = \begin{cases} 1 - \exp^{-\frac{x^2}{2a^2}} & \text{pour } x \geq 0 \\ 0 & x < 0. \end{cases}$$

On calcule sa densité :

$$f(x) = \frac{d}{dx}F(x) = \begin{cases} \frac{x}{a^2} \exp^{-\frac{x^2}{2a^2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

On en déduit la moyenne :

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{+\infty} \frac{x^2}{a^2} \exp^{-\frac{x^2}{2a^2}}, \\ &= \sqrt{\frac{\pi}{2}} \int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi}a^2} \exp^{-\frac{x^2}{2a^2}} = \sqrt{\frac{\pi}{2}} a \int_{-\infty}^{+\infty} \frac{x^2}{\sqrt{2\pi}} \exp^{-\frac{x^2}{2}}, \\ &= \sqrt{\frac{\pi}{2}} a, \end{aligned}$$

en reconnaissant dans la dernière intégrale le moment d'ordre 2 d'une loi normale centrée réduite. De même, on en déduit le moment d'ordre 2 comme en intégrant par partie,

$$\begin{aligned} \mathbb{E}(X^2) &= \int_0^{+\infty} x^2 \frac{x}{a^2} \exp^{-\frac{x^2}{2a^2}}, \\ &= \left[-x^2 \exp^{-\frac{x^2}{2a^2}} \right]_0^{+\infty} + 2 \int_0^{+\infty} x \exp^{-\frac{x^2}{2a^2}} \\ &= 2a \end{aligned}$$

Et on en déduit que la variance $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = (2 - \frac{\pi}{2})a^2$.

Pour le calcul de la médiane, comme la loi est à densité, on a que

$$\mathbb{P}(X < \lambda) - \frac{1}{2} = F(\lambda) - \frac{1}{2} = \frac{1}{2} - \exp^{-\frac{\lambda^2}{2a^2}}$$

qui nous permet de directement la trouver : $x_{1/2} = a\sqrt{2\ln(2)}$.

Pour trouver le maximum de la densité, on peut la dériver sur \mathbb{R}_+ , où l'on trouve

$$f'(x) = \frac{1}{a^2} \exp^{-\frac{x^2}{2a^2}} \left(1 - \frac{x^2}{a^2}\right)$$

et donc trouver que l'unique maximum de f est en $x_0 = a$.

, sa moyenne, sa variance, sa médiane (point λ tel que $\mathbb{P}(X < \lambda) = \frac{1}{2}$) et son mode (maximum de la densité).

3) Nous allons montrer qu'il est impossible de truquer deux dés 6 (potentiellement de manières différentes) de telle sorte que la variable aléatoire qui renvoie la somme des résultats des deux dés suive une loi uniforme sur $\llbracket 2, 12 \rrbracket$. Pour cela, raisonnons par l'absurde et supposons que cela soit possible.

Posons $P_{i,j}$ la probabilité que le dé numéro j tombe sur la face i , et X et Y les variables aléatoires qui renvoient le résultat du premier et deuxième dé respectivement.

On considère alors la variable aléatoire $X + Y$, et sa fonction génératrice G_{X+Y} prise en $\ln(t)$. Soit donc $t > 0$. Par hypothèse,

$$G_{X+Y}(\ln t) = \mathbb{E}(t^{X+Y}) = t^2 (1 + t + t^2 + \dots + t^{10}).$$

On connaît les racines de $1 + t + t^2 + \dots + t^{10} = \frac{1-t^{11}}{1-t}$, il s'agit de l'ensemble $S = \{e^{\frac{2i\pi k}{11}}, k \in \llbracket 1, 10 \rrbracket\} \subset \mathbb{C} \setminus \mathbb{R}$.

Mais les résultats de dé distincts sont indépendants, donc la fonction génératrice de la somme est le produit des fonctions génératrices (par un exercice de cours) :

$$G_{X+Y}(\ln t) = G_X(\ln t)G_Y(\ln t) = (p_{1,1}t + p_{2,1}t^2 + \dots + p_{6,1}t^6) (p_{1,2}t + p_{2,2}t^2 + \dots + p_{6,2}t^6)$$

Donc il existe 5 racines (a_i) de S telles que

$$p_{6,1} \prod_{i=1}^5 (t - a_i) = p_{1,1} + p_{2,1}t + \dots + p_{6,1}t^5 \in \mathbb{R}[t]$$

Ce qui est impossible puisque la somme des multiplicités des racines complexes non réelle d'un polynôme réel est forcément paire. Nous avons donc une contradiction, donc un tel truquage est impossible.

Variables discrètes

4) Cet exercice est un exercice d'apprentissage de cours.

5) Soit m et n deux entiers, $\Omega_1 = \{1, m\}$ et $\Omega_2 = \{1, n\}$, et $\Omega = \Omega_1 \times \Omega_2$. On munit Ω de la loi uniforme, et on considère les variables aléatoires X_1 et X_2 définies pour tout couple $(\omega_1, \omega_2) \in \Omega$ par

$$X_1(\omega_1, \omega_2) = \omega_1,$$

$$X_2(\omega_1, \omega_2) = \omega_2.$$

Alors, pour $\omega_1 \in \Omega_1$, on a

$$\mathbb{P}(X_1 = \omega_1) = \sum_{y \in \Omega: X_1(y) = \omega_1} \frac{1}{|\Omega|} = \sum_{(\omega_1, \omega_2, \omega_2 \in \Omega_2)} \frac{1}{|\Omega|} = \frac{|\Omega_2|}{|\Omega|}.$$

De même, pour $\omega_2 \in \Omega_2$,

$$\mathbb{P}(X_2 = \omega_2) = \frac{|\Omega_1|}{|\Omega|}$$

Alors, pour $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$,

$$\mathbb{P}(X_1 = \omega_1, X_2 = \omega_2) = \frac{1}{|\Omega|} = \frac{|\Omega_2|}{|\Omega|} \times \frac{|\Omega_1|}{|\Omega|} = \mathbb{P}(X_1 = \omega_1)\mathbb{P}(X_2 = \omega_2).$$

6) Soit $X \sim \mathcal{Bin}(n, p)$, alors $V(X) = np(1-p)$. Cette quantité est maximale pour n fixé en $p = \frac{1}{2}$.

7) Une usine produit des composants électroniques en grande quantité, avant de les regrouper dans des boîtes de mille composants. On observe que la probabilité qu'une boîte contienne au moins un composant défectueux est de 0.52.

On relève aléatoirement 12 boîtes, et l'on pourra supposer en première approximation que l'état des composants sont indépendants (hypothèse fortement discutable en pratique). Soit X la variable aléatoire égale au nombre de composants défectueux dans la première boîte, et Y le nombre de boîtes contenant au moins un composant défectueux parmi les douze.

- X et Y sont la somme de binomiales indépendante de même paramètre, donc suivent une loi binomiale. On a directement $Y \sim \mathcal{Bin}(12, 0.52)$.

Pour trouver p tel que $X \sim \mathcal{Bin}(1000, p)$, on sait que $0.52 = \mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - 1 \cdot (1-p)^{1000}$.

Donc $p = 1 - 0.48^{\frac{1}{1000}} \simeq 0.00073$.

- On a que

$$\mathbb{P}(Y = 5) = C_{12}^5 \cdot 0.52^5 \cdot 0.48^7 \simeq 0.176$$

- On connaît l'espérance d'une loi binomiale, donc

$$\mathbb{E}(Y) = 12 \times 0.52 = 6.24 \simeq 6$$

- Avec la formule de la variance d'une loi binomiale,

$$V(X) = np(1-p) \simeq 0.733 \quad \sigma_X = \sqrt{np(1-p)} \simeq 0.856$$

8) Soient $p \in]0, 1[$, et $X \sim \mathcal{G}(p)$. Montrons que X est sans mémoire, c'est-à-dire que pour $k \geq m$ deux entiers naturels,

$$\mathbb{P}(X > k | X > m) = \mathbb{P}(X > k - m).$$

On calcule

$$\begin{aligned} \mathbb{P}(X > n) &= \sum_{r=n+1}^{+\infty} p(1-p)^{r-1}, \\ &= p \frac{(1-p)^{n-1}}{1 - (1-p)} = (1-p)^n \end{aligned}$$

Donc on vérifie bien que la loi géométrique est sans mémoire :

$$\mathbb{P}(X > k | X > m) = \frac{\mathbb{P}(X > k, X > m)}{\mathbb{P}(X > m)} = \frac{p^k}{p^m} = p^{k-m} = \mathbb{P}(X > k - m)$$

Posons $X \sim \mathcal{G}(\frac{1}{2})$.

Si l'on a déjà lancé une pièce équilibrée 18 fois et on a obtenu uniquement le côté face, la probabilité que l'on obtienne pile au bout de 24 lancers est alors :

$$\begin{aligned}\mathbb{P}(X = 24|X > 18) &= \mathbb{P}(X > 23|X > 18) - \mathbb{P}(X > 24|X > 18) = \mathbb{P}(X > 5) - \mathbb{P}(X > 6) \\ &= \frac{1}{2}(1 - \frac{1}{2})^5 = \frac{1}{64}.\end{aligned}$$

9) Soit $X \sim \mathcal{P}(\lambda)$ une variable aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$.

$$\begin{aligned}\mathbb{E}\left(\frac{1}{X+1}\right) &= \sum_{k=0}^{+\infty} \frac{1}{k+1} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{(k+1)!} \\ &= e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{k!} \\ &= \frac{e^{-\lambda}}{\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{k!} \\ &= \frac{e^{-\lambda}}{\lambda} \left(\sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} - 1 \right) \\ &= \frac{e^{-\lambda}}{\lambda} (e^{\lambda} - 1) \\ &= \frac{1 - e^{-\lambda}}{\lambda}\end{aligned}$$

10) Soit $X \sim \mathcal{G}(p)$ et $Y \sim \mathcal{P}(\lambda)$ indépendantes. Quitte à avoir un terme infini, on a :

$$\begin{aligned}\mathbb{E}X^Y &= \sum_{(m,n) \in \mathbb{N}^2} m^n \mathbb{P}(X = m, Y = n) = \sum_{(m,n) \in \mathbb{N}^2} m^n \mathbb{P}(X = m) \mathbb{P}(Y = n) \quad \text{par indépendance} \\ &= \sum_{(m,n) \in \mathbb{N}^2} m^n (1-p)^m p \frac{\lambda^n}{n!} e^{-\lambda} \\ &= pe^{-\lambda} \sum_{m \in \mathbb{N}} (1-p)^m \sum_{n \in \mathbb{N}} m^n \frac{\lambda^n}{n!} \\ &= pe^{-\lambda} \sum_{m \in \mathbb{N}} (1-p)^m e^{\lambda m} \\ &= pe^{-\lambda} \frac{1}{1 - (1-p)e^{\lambda}}\end{aligned}$$

Donc X^Y a une espérance que si $(1-p)e^{\lambda} < 1$, d'espérance $pe^{-\lambda} \frac{1}{1 - (1-p)e^{\lambda}}$.

Avec le même type de calcul, on arrive à montrer que $\mathbb{E}(X^Y)^2 = pe^{-\lambda} \sum_{m \in \mathbb{N}} (1-p)^m e^{\lambda m^2}$ qui est

supérieur pour $p \neq 1$ et $\lambda \neq 0$ à $pe^{-\lambda} \sum_{m \geq \frac{\ln(1-p)}{\lambda}} (1-p)^m e^{\lambda m^2} \geq +\infty$. Donc X^Y n'a un moment d'ordre 2 que si $p = 0$ ou $\lambda = 0$.

11) Commencer par montrer que $(1 - p_n)^{n-k} \rightarrow e^{-\lambda}$ et que $\frac{n!}{n^k(n-k)!} \rightarrow 1$, avant de conclure.

Variables à densité

12) Cet exercice est un exercice d'apprentissage de cours.

13) Chercher à appliquer le changement de variable pour les lois absolument continues.

14) On remarque que comme X est absolument continue, F est une fonction continue. Comme de plus F est strictement croissante, c'est une bijection. Notons F^{-1} l'application inverse, qui est également croissante.

Alors, pour $t \in [0; 1]$, on a l'égalité d'événements

$$\{Y \leq t\} = \{X \leq F^{-1}(t)\}.$$

Donc

$$\mathbb{P}(Y \leq t) = \mathbb{P}[X \leq F^{-1}(t)] = F[F^{-1}(t)] = t$$

Il s'agit exactement de la fonction de répartition d'une loi car par croissance des fonctions de répartition, $F_Y(u) = 0$ pour u négatif et $F_Y(u) = 1$ pour $u \geq 1$. Donc Y suit une loi uniforme.

15) On considère $[0, 1]$ muni de la mesure uniforme. Soit $b \geq 2$ un entier, et pour $x \in [0, 1]$ on note $x \rightarrow (\epsilon_{i,b}(x))_{i \in \mathbb{N}^*}$ l'unique fonction à valeur dans les suites d'entiers entre 0 et $b-1$ et non constamment égale à $b-1$ à partir d'un certain rang, telle que $x = \sum_{i=1}^{+\infty} \frac{\epsilon_{i,b}(x)}{b^i}$

1. Soit $i \in \mathbb{N}$, montrons que $\epsilon_{i,b}$ suit une loi uniforme sur $\llbracket 0; b-1 \rrbracket$.

Pour $k \in \llbracket 0; b-1 \rrbracket$, on a que

$$\{\epsilon_{i,b} = k\} = \bigcup_{(a_j)_{1 \leq j \leq i-1} \in \llbracket 0; b-1 \rrbracket^{i-1}} \left[\sum_{j=1}^{i-1} \frac{a_j}{b^j} + \frac{k}{b^i}; \sum_{j=1}^{i-1} \frac{a_j}{b^j} + \frac{k+1}{b^i} \right[.$$

Comme l'union est disjointe, on a donc que

$$\mathbb{P}(\epsilon_{i,b} = k) = \sum_{(a_j)_{1 \leq j \leq i-1} \in \llbracket 0; b-1 \rrbracket^{i-1}} \frac{1}{b^i} = \frac{b^{i-1}}{b^i} = \frac{1}{b}.$$

Qui nous donne bien que $\epsilon_{i,b}$ suit une loi uniforme.

2. Pour montrer que la suite des variables $(\epsilon_{i,b}(x))_{i \in \mathbb{N}^*}$ est indépendante dans son ensemble, il faut montrer que c'est le cas pour toute sous suite finies. Mais pour cela, il suffit de montrer

que pour tout $n \in \mathbb{N}$, le début de la suite jusqu'au rang n est indépendante dans son ensemble. Or pour $(a_j)_{1 \leq j \leq n} \in \llbracket 0; b-1 \rrbracket^n$, on a que

$$\mathbb{P}[(\epsilon_{j,b})_{1 \leq j \leq n} = (a_j)_{1 \leq j \leq n}] = \mathbb{P}\left[\sum_{j=1}^n \frac{a_j}{b^j}, \sum_{j=1}^n \frac{a_j}{b^j} + \frac{1}{b^n}\right] = \frac{1}{b^n}$$

On retrouve bien que

$$\mathbb{P}[(\epsilon_{j,b})_{1 \leq j \leq n} = (a_j)_{1 \leq j \leq n}] = \prod_{i=1}^n \mathbb{P}[\epsilon_{i,b} = a_i].$$

3. On se donne n un entier, et y_1, \dots, y_n une suite finie de chiffre.

$$\mathbb{P}[\forall i \in \llbracket kn+1; (k+1)n \rrbracket, \epsilon_{i,b} = y_i] = \frac{1}{b^n}.$$

Notons, pour un entier naturel k donné, " $A_k(y_1, \dots, y_n)$ " l'événement la suite y_1, \dots, y_n n'apparaît pas dans les chiffres en position $kn+1$ à $(k+1)n$. On remarque que

$$"(y_1, \dots, y_n) \text{ n'apparaît pas dans les chiffres}" \subset \bigcap_{k=1}^{+\infty} A_k(y_1, \dots, y_n).$$

Donc pour p un entier, on a par indépendance que :

$$\begin{aligned} \mathbb{P}[y_1, \dots, y_n \text{ n'apparaissent jamais}] &\leq \mathbb{P}\left[\bigcap_{k=1}^p A_k(y_1, \dots, y_n)\right] \\ &\leq \prod_{k=1}^p \mathbb{P}[A_k(y_1, \dots, y_n)] \\ &\leq \prod_{k=1}^p \left(1 - \frac{1}{b^n}\right) \\ &= \left(1 - \frac{1}{b^n}\right)^p \xrightarrow{p \rightarrow +\infty} 0 \end{aligned}$$

On en déduit bien que presque sûrement, la suite y_1, \dots, y_n apparaît consécutivement dans les chiffres de x .

4. Comme l'ensemble des suites finies est dénombrable, par σ -sous additivité des mesures, on a bien que presque sûrement tous les réels sont des nombres univers en base b .
5. Comme l'ensemble des bases et dénombrables, on a même que presque tous les réels sont univers dans toutes les bases. En particulier, cet ensemble est non vide.

5.3 Section 3 : analyse multivariée

1) Soit X, Y un couple de variable aléatoire dont la loi est donnée par le tableau suivant :

$X \backslash Y$	1	3	5	8
-3	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$
5	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{16}$

Posons tout d'abord $I = \{-3; 5\}$ et $J = \{1; 3; 5; 8\}$ On calcule la moyenne du vecteur (X, Y) :

$$\begin{aligned} \mathbb{E} \begin{bmatrix} X \\ Y \end{bmatrix} &= \begin{pmatrix} \sum_{i \in I} i \mathbb{P}(X = i) \\ \sum_{j \in J} j \mathbb{P}(Y = j) \end{pmatrix}, \\ &= \begin{pmatrix} -3(\frac{3}{16} + \frac{1}{16} + \frac{1}{8} + \frac{1}{8}) + 5(\frac{1}{16} + \frac{1}{8} + \frac{1}{8} + \frac{3}{16}) \\ 1(\frac{3}{16} + \frac{1}{16}) + 3(\frac{1}{16} + \frac{1}{8}) + 5(\frac{1}{8} + \frac{1}{8}) + 8(\frac{1}{8} + \frac{3}{16}) \end{pmatrix}, \\ &= \begin{pmatrix} 1 \\ \frac{73}{16} \end{pmatrix}. \end{aligned}$$

La covariance du couple est alors $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, avec

$$\mathbb{E}[XY] = \sum_{(i,j) \in I \times J} ij \mathbb{P}(X = i, Y = j) = \frac{109}{16}.$$

D'où

$$\text{Cov}(X, Y) = \frac{9}{4}$$

La variance de X est alors :

$$V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 17 - 1 = 16 = 4^2.$$

Et la variance de Y est :

$$V(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{505}{16} - \frac{5329}{256} = \frac{2751}{256}$$

Donc le coefficient de corrélation vérifie

$$\rho(X, Y) = \frac{9/4}{4 \times \sqrt{2751/16}} \simeq 0.0107$$

2) Soit $(X_1, \dots, X_n) \sim \mathcal{M}(r, p_1, \dots, p_d)$, montrons que X_1 suit une loi binomiale de paramètre r et p_1 . D'après le développement du multinôme, on a que

$$(p_2 + \dots + p_d)^r = \sum_{r_2, \dots, r_d: \sum r_i = r} \frac{r!}{r_2! \dots r_d!} p_2^{r_2} \dots p_d^{r_d}$$

Or avec la formule des probabilités totale, pour $(X_1, \dots, X_n) \sim \mathcal{M}(r, p_1, \dots, p_d)$ et tout $0 \leq k \leq r$, on a que :

$$\begin{aligned} \mathbb{P}(X_1 = k) &= \sum_{r_2, \dots, r_n: \sum r_i = r-k} \frac{r!}{r_1! r_2! \dots r_d!} p_1^k p_2^{r_2} \dots p_d^{r_d}; \\ &= \frac{r!}{k!(r-k)!} p_1^k \sum_{r_2, \dots, r_n: \sum r_i = r-k} \frac{(r-k)!}{r_2! \dots r_d!} p_2^{r_2} \dots p_d^{r_d}; \\ &= \frac{r!}{k!(r-k)!} p_1^k (p_2 + \dots + p_d)^{r-k}; \\ &= \frac{r!}{k!(r-k)!} p_1^k (1 - p_1)^{r-k}; \end{aligned}$$

ce qui prouve bien que X_1 suit une loi binomiale

3) Soit $X \sim \mathcal{N}(0, 1)$, et ϵ une variable aléatoire indépendante de X qui vaut ± 1 avec probabilité $\frac{1}{2}$.

Montrer que $(X, \epsilon X)$ n'est pas un vecteur gaussien.

Si l'on considère la combinaison linéaire $Y = X + \epsilon X$, on a que

$$\mathbb{P}[Y = 0] = \mathbb{P}[\epsilon = -1] = \frac{1}{2} \neq 0.$$

Donc Y a au moins un atome, donc ne peut pas suivre une loi continue, et donc en particulier une loi normale.

4) Utiliser le fait que pour une loi multinomiale, les composantes sont linéairement liées.

5) Soient X_1, X_2 et X_3 des variables aléatoires mutuellement indépendante de loi de Rademacher, et on pose alors $Y_1 = X_2 X_3, Y_2 = X_1 X_3$ et $Y_3 = X_1 X_2$.

- Pour calculer les lois marginales des Y_i , par symétrie par permutation circulaire, il suffit de calculer la loi de Y_1 . Or on a que :

$$\begin{aligned} \mathbb{P}[Y_1 = 1] &= \mathbb{P}[X_2 X_3 = 1], \\ &= \mathbb{P}[(X_2 = 1, X_3 = 1) \cup (X_2 = -1, X_3 = -1)], \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}, \\ &= \frac{1}{2} \end{aligned}$$

De même, $\mathbb{P}[Y_1 = -1] = \frac{1}{2}$.

- De nouveau, par symétrie, les lois des couples (Y_i, Y_j) pour $i \neq j$ sont toutes égales à celle du couple (Y_1, Y_2) .

Comme le problème est invariant par multiplication de l'un des X_i par -1 , on a que le couple (Y_1, Y_2) est réparti uniformément sur $\{(1, 1); (1, -1); (-1, 1); (-1, -1)\}$. On aurait aussi pu le

vérifier par des calculs similaires à :

$$\begin{aligned}
 \mathbb{P}[(Y_1, Y_2) = (1, 1)] &= \mathbb{P}[(X_2 X_3 = 1) \cap (X_1 X_3 = 1)], \\
 &= \mathbb{P}[(X_1, X_2, X_3) \in \{(1, 1, 1); (-1, -1, -1)\}], \\
 &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}, \\
 &= \frac{1}{4}
 \end{aligned}$$

En particulier, les variables Y_1 , Y_2 et Y_3 sont donc deux à deux indépendantes en revenant à la définition.

- Soit $(\epsilon_1, \epsilon_2, \epsilon_3) \in \{\pm 1\}^3$ un résultat possible pour le triplet. Alors on a

$$\begin{aligned}
 \mathbb{P}[(Y_1, Y_2, Y_3) = (\epsilon_1, \epsilon_2, \epsilon_3)] &= \mathbb{P}[(X_2 X_3 = \epsilon_1) \cap (X_1 X_3 = \epsilon_2) \cap (X_1 X_2 = \epsilon_3)], \\
 &= \mathbb{P}[(X_1, X_2, X_3) \in \{(\epsilon_1, \epsilon_2, \epsilon_3); (-\epsilon_1, -\epsilon_2, -\epsilon_3)\}], \\
 &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}, \\
 &= \frac{1}{4}
 \end{aligned}$$

Les variables Y_1 , Y_2 et Y_3 ne sont donc pas mutuellement indépendantes, puisque par exemple

$$\mathbb{P}[(Y_1, Y_2, Y_3) = (1, 1, 1)] \neq \mathbb{P}[Y_1 = 1]\mathbb{P}[Y_2 = 1]\mathbb{P}[Y_3 = 1]$$

5.4 Section 4 : théorèmes limites de l'aléatoire

1) On se servira de la fonction caractéristique d'une loi de Bernouilli.

2) Il s'agit d'un calcul explicite d'intégrale.

3) On pourra se rappeler du lien entre loi exponentielle et loi de gamma.

4) Il s'agit de poser le changement de variable adéquat

5) Pour chacune des questions, il s'agit d'utiliser que la fonction caractéristique d'une somme de variables aléatoire est le produit des fonctions caractéristique, puis de reconnaître la fonction caractéristique et d'utiliser qu'elle caractérise la loi.

6) Pour montrer la proposition 76, calculer la fonction caractéristique des combinaisons linéaires, et reconnaître la fonction caractéristique de la loi normale.

7) On va ici chercher à montrer le théorème 77. Soit $Y = (Y_1, \dots, Y_n)$ un vecteur gaussien centré de covariance K .

1. Si $Z \sim \mathcal{N}(m; \sigma^2)$, alors $\forall t \in \mathbb{R}, \varphi_Z(t) = e^{imt - \frac{\sigma^2 t^2}{2}}$.

2. Soit $\bar{t} = (t_1, \dots, t_n)^t \in \mathbb{R}^n$, par définition d'un vecteur gaussien, $\langle \bar{t}; Y \rangle = \sum_{p=1}^n t_p Y_p$ suit une loi normale.

Comme l'on a supposé Y centré, cette variable aléatoire est centrée. De plus, on peut calculer sa variance comme

$$\begin{aligned} V\left(\sum_{p=1}^n t_p Y_p\right) &= \sum_{p=1}^n t_p^2 V(Y_p) + \sum_{\substack{1 \leq p, q \leq n \\ p \neq q}} t_p t_q \text{Cov}(Y_p, Y_q), \\ &= \langle \bar{t}; \Sigma \bar{t} \rangle. \end{aligned}$$

Donc avec le point précédent,

$$\forall \bar{t} \in \mathbb{R}^d, \varphi_Y(\bar{t}) = \varphi_{\langle \bar{t}; Y \rangle}(1) = e^{-\frac{\langle \bar{t}; \Sigma \bar{t} \rangle}{2}}.$$

3. Comme toute combinaison linéaire de combinaison linéaire peut se réécrire comme une seule combinaison linéaire, on a que pour toute matrice A et tout $\bar{t} \in \mathbb{R}^n$, $\bar{t}^t A Y$ est une combinaison linéaire des Y_i , et est donc une variable suivant une loi normale. Donc $A Y$ est aussi un vecteur gaussien, et le résultat en découle si l'on impose que A soit symétrique.

4. Comme $\langle \bar{t}; A Y \rangle = \langle A^t \bar{t}; Y \rangle$, qui est de variance $\bar{t}^t A K A^t \bar{t}$ avec ce qui précède. En faisant varier \bar{t} , on en déduit la matrice de covariances.

5. On prend B une racine carrée dans les matrices symétriques définies positives de K et qui commute avec K . Alors $B Y$ est un vecteur gaussien de covariance I_d .

6. La densité est évidente pour les vecteurs gaussiens standard, comme densité produit.

Pour le sens direct du théorème, on remarque que $\varphi_{BY}(\bar{t})$ est égal à la fonction caractéristique d'un vecteur $X = (X_1, \dots, X_n)$ composé de n vecteurs gaussiens centrés réduits indépendants. Donc comme la fonction caractéristique caractérise la loi, BY admet la densité suivante suivant Lebesgue :

$$\frac{1}{\sqrt{2\pi}^d} e^{-\frac{x^t x}{2}}.$$

On peut alors conclure en faisant le changement de variable linéaire $Z \mapsto B^{-1}Z$, de déterminant $\sqrt{\det(K)}^{-1}$.

Annexe A

Loi forte des grands nombres

On démontre dans cette annexe la loi forte des grands nombres

Théoreme 104 : Loi forte des grands nombres

Soit μ une loi de probabilité sur \mathbb{R} qui admette un moment d'ordre 1. On pose $m = \int_{\mathbb{R}} x d\mu(x)$ sa moyenne. On considère $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi μ .

Alors

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.} m.$$

Nous allons commencer par une série de réductions.

Première réduction : commençons par montrer que si le théorème est vrai pour $X_1 \geq 0$ presque sûrement, alors il est vrai en toute généralité.

Notons pour X une variable aléatoire $X_+ = \max(0, X)$ et $X_- = \max(0, -X)$. On remarque alors que $X_+ \geq 0$, $X_- \geq 0$ et que $X = X_+ - X_-$.

Alors, par l'hypothèse que le théorème est vrai pour les variables positives, comme les $(X_i)_+$ et $(X_i)_-$ sont indépendantes identiquement distribuées,

$$\begin{aligned} \frac{(X_1)_+ + \cdots + (X_n)_+}{n} &\xrightarrow{p.s.} \mathbb{E}[(X_1)_+], \\ \frac{(X_1)_- + \cdots + (X_n)_-}{n} &\xrightarrow{p.s.} \mathbb{E}[(X_1)_-]. \end{aligned}$$

Par linéarité des limites réelles, on a pour tout couple de réels (a, b) et toutes suites de variables aléatoires (U_k) et (T_k) , $\{U_k \rightarrow a\} \cap \{T_k \rightarrow b\} \subset \{U_k + T_k \rightarrow a + b\}$.

En particulier, si $U_k \xrightarrow{p.s.} a$ et $T_k \xrightarrow{p.s.} b$, alors $U_k + T_k \xrightarrow{p.s.} a + b$.

Comme $X_i = (X_i)_+ - (X_i)_-$ et que $\mathbb{E}[X_1] = \mathbb{E}[(X_1)_+] - \mathbb{E}[(X_1)_-]$, on peut donc faire la première réduction.

À partir de maintenant, on suppose donc que les variables aléatoires X_k sont positives.

Deuxième réduction : Montrons à présent que nous pouvons utiliser un argument de troncature.

On pose $Y_k := X_k \mathbb{1}_{X_k \leq k}$ et $S_k := Y_1 + \dots + Y_k$.

On veut montrer dans cette réduction qu'il suffit de montrer que

$$\frac{S_n}{n} \xrightarrow{p.s.} \mathbb{E}(X_1).$$

Pour cela, montrons que presque sûrement, $\frac{X_1 + \dots + X_n}{n} - \frac{S_n}{n} \rightarrow 0$.

D'après le théorème de Cesàro, il suffit pour cela de vérifier que $X_k \mathbb{1}_{X_k > k} = X_k - Y_k \xrightarrow{p.s.} 0$. Pour cela, on peut noter que

$$\{\omega, X_k \mathbb{1}_{X_k > k} \text{ ne tends pas vers } 0\} \subset \bigcap_{k \in \mathbb{N}} \{X_k > k\}.$$

Donc

$$\begin{aligned} \mathbb{P}(X_k - Y_k \text{ ne tends pas vers } 0) &\leq \mathbb{P} \left[\bigcap_{k \in \mathbb{N}} \{X_k > k\} \right], \\ &\leq \inf_{k \in \mathbb{N}} \mathbb{P}[X_k > k], \\ &= \inf_{k \in \mathbb{N}} \mathbb{P}[X_1 > k] = \mathbb{P}[X_1 = +\infty] = 0. \end{aligned}$$

Car comme X_1 est intégrable, il est fini presque partout.

On cherche donc à partir de maintenant à montrer que $\frac{S_n}{n} \xrightarrow{p.s.} \mathbb{E}(X_1)$.

Troisième réduction : Soit $\alpha > 1$, et on pose $k_n = \lfloor \alpha^n \rfloor$. On va chercher à prouver qu'il suffit de prouver que

$$\frac{S_{k_n} - \mathbb{E}(S_{k_n})}{k_n} \xrightarrow{p.s.} 0.$$

Supposons cette convergence presque sûre, et montrons que le théorème est vrai.

Remarquons tout d'abord que comme $|X_1 \mathbb{1}_{X_1 \leq n}| \leq X_1$ qui est intégrable et que $X_1 \mathbb{1}_{X_1 > n} \rightarrow 0$ presque sûrement, on a par égalité des lois et convergence dominée que $\mathbb{E}(Y_n) = \mathbb{E}(X_1 \mathbb{1}_{X_1 \leq n}) \rightarrow \mathbb{E}(X_1)$.

Encore une fois, avec le théorème de Cesàro, on a alors que $\frac{\mathbb{E}(S_n)}{n} \rightarrow \mathbb{E}(X_1)$, et donc en tant que suite extraite que $\frac{\mathbb{E}(S_{k_n})}{k_n} \rightarrow \mathbb{E}(X_1)$.

Maintenant, pour $\alpha > 1$ et $n \in \mathbb{N}$, on peut trouver un p tel que $k_p \leq n \leq k_{p+1}$.

On a que

$$\frac{(\alpha^{p+1} - 1)}{\alpha^p + 1} \leq \frac{\lfloor \alpha^{p+1} \rfloor}{\lfloor \alpha^p \rfloor} \leq \frac{\alpha^{p+1} + 1}{\alpha^p - 1},$$

et donc que $\frac{k_{p+1}}{k_p} \rightarrow \alpha$.

De plus, comme les Y_n sont positifs, (S_n) est une suite croissante, donc

$$\frac{k_p}{k_{p+1}} \frac{S_{k_p}}{k_p} = \frac{S_{k_p}}{k_{p+1}} \leq \frac{S_n}{n} \leq \frac{S_{k_{p+1}}}{k_p} = \frac{S_{k_{p+1}}}{k_{p+1}} \frac{k_{p+1}}{k_p}.$$

En passant à la limite inférieure dans la première inégalité et à la limite supérieure dans la deuxième, on trouve donc :

$$\frac{1}{\alpha} \mathbb{E}(X_1) = \lim \frac{k_p}{k_{p+1}} \frac{S_{k_p}}{k_p} \leq \liminf \frac{S_n}{n} \leq \limsup \frac{S_n}{n} \leq \lim \frac{S_{k_{p+1}}}{k_{p+1}} \frac{k_{p+1}}{k_p} = \alpha \mathbb{E}(X_1) \text{ presque sûrement.}$$

Donc en prenant la limite pour $\alpha \rightarrow 1$ le long d'une suite, on a que la limite de $\frac{S_n}{n}$ existe et vaut $\mathbb{E}(X_1)$ presque sûrement, qui nous donne la troisième réduction.

Critère de convergence : Montrons que pour (U_n) une suite de variables aléatoires, il suffit de montrer que pour tout $\epsilon > 0$, la série $\sum_{n \in \mathbb{N}} \mathbb{P}[|U_n| > \epsilon] < +\infty$ converge pour avoir que $U_n \xrightarrow{p.s.} 0$.

Supposons que pour tout $\epsilon > 0$, la série $\sum_{n \in \mathbb{N}} \mathbb{P}[|U_n| > \epsilon] < +\infty$ converge.

D'après le lemme de Borel-Canteli (lemme 14), on a alors que $\mathbb{P}[|U_n| > \epsilon \text{ infiniment souvent}] = 0$.

Donc

$$\begin{aligned} \mathbb{P}[U_n \text{ ne converge pas vers } 0] &= \mathbb{P}\left[\bigcup_{k \in \mathbb{N}^*} |U_n| > \frac{1}{k} \text{ infiniment souvent}\right] \\ &\leq \sum_{k \in \mathbb{N}^*} \mathbb{P}\left[|U_n| > \frac{1}{k} \text{ infiniment souvent}\right] \\ &= 0. \end{aligned}$$

Qui est bien la condition énoncée.

Nous avons à présent tout ce qui nous faut pour la démonstration, pour laquelle nous allons vouloir utiliser l'inégalité de Bienaymé-Tchebychev. Nous allons donc avoir besoin de majorer les variances

Première estimation :

Soit $k \geq 1$, alors comme Y_k est borné, il admet un moment d'ordre 2, et avec la formule de König-Huygens,

$$V(Y_k) \leq \mathbb{E}(Y_k^2).$$

Maintenant, avec le théorème de Fubini-Tonelli,

$$\begin{aligned} \mathbb{E}(Y_k^2) &= \int_{\Omega} Y_k(\omega)^2 d\mathbb{P}(\omega), \\ &= \int_{\Omega} \int_0^{+\infty} \mathbb{1}_{Y_k(\omega)^2 \geq t}(x) dx d\mathbb{P}(\omega) = \int_0^{+\infty} \int_{\Omega} \mathbb{1}_{Y_k(\omega)^2 \geq t}(x) d\mathbb{P}(\omega) dx, \\ &= \int_0^{+\infty} \mathbb{P}(Y_k^2 \geq x) dx, \\ &\text{et donc par changement de variable } y^2 = x, \\ &= \int_0^{+\infty} 2y \mathbb{P}(Y_k^2 \geq y^2) dy = \int_0^k 2y \mathbb{P}(Y_k \geq y) dy, \\ &= \int_0^k 2y \mathbb{P}(X_1 \geq y) dy \end{aligned}$$

C'est-à-dire que

$$V(Y_k) \leq \int_0^k 2y \mathbb{P}(X_1 \geq y) dy.$$

Deuxième estimation :

De plus, d'après l'inégalité de Bienaymé-Tchebychev, pour $\epsilon > 0$, on a l'inégalité la série

$$\sum_{n \geq 1} \mathbb{P} \left[\left| \frac{S_{k_n} - \mathbb{E}(S_{k_n})}{k_n} \right| > \epsilon \right] \leq \sum_{n \geq 1} \frac{V(S_{k_n})}{\epsilon^2 k_n^2},$$

Puis par indépendance des Y_k ,

$$\begin{aligned} &\leq \sum_{n \geq 1} \sum_{i=1}^{k_n} \frac{V(Y_i)}{\epsilon^2 k_n^2} \\ &\leq \sum_{i \geq 1} \frac{V(Y_i)}{\epsilon^2} \sum_{n \text{ tq } k_n \geq i} \frac{1}{k_n^2}, \\ &\leq \sum_{i \geq 1} d \frac{V(Y_i)}{\epsilon^2 i^2} \quad \text{Pour d une constante.} \end{aligned}$$

Pour la dernière ligne, nous avons utilisé que si on note $n_0 = \min\{n, i \leq k_n\}$, alors :

$$i^2 \sum_{n \text{ tq } k_n \geq i} \frac{1}{k_n^2} = i^2 \sum_{n \geq n_0} \frac{1}{[\alpha^n]^2},$$

$$\text{Et comme } \frac{1}{[\alpha^n]^2} \underset{n \rightarrow +\infty}{\sim} \frac{1}{\alpha^{2n}},$$

$$\underset{n \rightarrow +\infty}{\sim} i^2 \sum_{n \geq n_0} \frac{1}{\alpha^{2n}}, \underset{n \rightarrow +\infty}{\sim} i^2 \frac{\alpha^{-2n_0}}{1 - \frac{1}{\alpha^2}},$$

$$\underset{n \rightarrow +\infty}{\sim} \left(\frac{i}{k_{n_0}} \right)^2 \frac{1}{1 - \frac{1}{\alpha^2}},$$

et donc que $i^2 \sum_{n \text{ tq } k_n \geq i} \frac{1}{k_n^2}$ est borné car $\left(\frac{i}{k_{n_0}} \right)^2 \frac{1}{1 - \frac{1}{\alpha^2}} \leq \frac{1}{1 - \frac{1}{\alpha^2}}$.

Dernière estimation :

Nous pouvons alors conclure, car

$$\begin{aligned}
\frac{\epsilon^2}{d} \sum_{n \geq 1} \mathbb{P} \left[\left| \frac{S_{k_n} - \mathbb{E}(S_{k_n})}{k_n} \right| > \epsilon \right] &\leq \sum_{k \geq 1} \frac{V(Y_k)}{k^2} \quad \text{d'après la deuxième estimation,} \\
&\leq \sum_{k \geq 1} \frac{1}{k^2} \int_0^k 2y \mathbb{P}[X_1 > y] dy \quad \text{d'après la première estimation,} \\
&\leq \sum_{k \geq 1} \frac{1}{k^2} \left(\sum_{n=1}^k \int_{n-1}^n 2y \mathbb{P}[X_1 > y] dy \right), \\
&\leq \sum_{k \geq 1} \frac{1}{k^2} \sum_{n=1}^k \int_{n-1}^n 2n \mathbb{P}[X_1 > y] dy,
\end{aligned}$$

Et d'après le théorème de Fubini-Tonelli :

$$\begin{aligned}
&\leq 2 \sum_{n=1}^{+\infty} n \int_{n-1}^n \mathbb{P}[X_1 > y] dy \sum_{k=n}^{+\infty} \frac{1}{k^2}, \\
\text{Et comme } \frac{1}{k^2} &\leq \frac{2}{k(k+1)} = \frac{2}{k} - \frac{2}{k+1}, \\
&\leq 2 \sum_{n=1}^{+\infty} n \int_{n-1}^n \mathbb{P}[X_1 > y] dy \sum_{k=n}^{+\infty} 2 \left(\frac{1}{k} - \frac{1}{k+1} \right), \\
&\leq 4 \sum_{n=1}^{+\infty} n \int_{n-1}^n \mathbb{P}[X_1 > y] dy \frac{1}{n}, \\
&\leq 4 \sum_{n=1}^{+\infty} \int_{n-1}^n \mathbb{P}[X_1 > y] dy = 4 \int_0^{+\infty} \mathbb{P}[X_1 > y] dy \\
&= 4 \int_0^{+\infty} \int_{\Omega} \mathbb{1}_{[X_1 > y]} d\mathbb{P} dy = 4\mathbb{E}(X_1). \text{ par Fubini Tonelli.} \\
&< +\infty.
\end{aligned}$$

Donc d'après le critère de convergence,

$$\left| \frac{S_{k_n} - \mathbb{E}(S_{k_n})}{k_n} \right| \xrightarrow{p.s.} 0.$$

Donc avec la troisième réduction,

$$\frac{S_n}{n} \xrightarrow{p.s.} \mathbb{E}(X_1).$$

Donc avec la deuxième réduction, la loi forte des grands nombres est vérifiée pour les variables aléatoires positive.

Donc avec la première réduction, la loi forte des grands nombres est vérifiée.

Annexe B

Convergence et convolution

Rappelons que l'on a défini g_σ comme :

$$g_\sigma : x \mapsto \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}.$$

Proposition 105 :

Soit f une fonction intégrable. Alors :

- Si f est continu et borné, $f * g_\sigma \xrightarrow[\sigma \rightarrow 0]{CVS} f$
- Si f est uniformément continue et bornée, $f * g_\sigma \xrightarrow[\sigma \rightarrow 0]{CVU} f$

Démonstration.

Soit f continue et bornée, et x_0 un point, on veut montrer que

$$f * g_\sigma(x_0) - f(x_0) = \int_{\mathbb{R}} [f(x_0 - y) - f(x_0)] g_\sigma(y) dy \rightarrow 0.$$

Soit donc $\epsilon > 0$, alors par continuité de f en x_0 , il existe $\alpha > 0$ tel que si $|x_0 - y| \leq \alpha$, alors $|f(x_0) - f(y)| \leq \epsilon$.

On découpe alors l'intégrale de la convolution en deux morceaux :

$$\begin{aligned} |f * g_\sigma(x_0) - f(x_0)| &= \left| \int_{|y| > \alpha} [f(x_0 - y) - f(x_0)] g_\sigma(y) dy + \int_{|y| \leq \alpha} [f(x_0 - y) - f(x_0)] g_\sigma(y) dy \right| \\ &\leq 2\|f\|_\infty \int_{|y| > \alpha} g_\sigma(y) dy + \epsilon \end{aligned}$$

Or par changement de variable, puis convergence dominée par g_1 ,

$$\begin{aligned}\int_{|y|>\alpha} g_\sigma(y)dy &= \int_{|y|>\frac{\alpha}{\sigma}} g_1(y)dy \\ &= \int_{\mathbb{R}} g_1(y)\mathbf{1}_{|y|>\frac{\alpha}{\sigma}} dy \\ &\xrightarrow{\sigma \rightarrow 0} 0\end{aligned}$$

Donc pour σ proche de 0, $|\int_{|y|>\alpha} g_\sigma(y)dy| \leq \epsilon$.

En particulier,

$$|f * g_\sigma(x_0) - f(x_0)| \leq (2\|f\|_\infty + 1)\epsilon$$

De plus, si f est uniformément continue, toutes ces inégalités sont vérifiées pour tout x_0 pour le même σ , donc la convergence est uniforme. \square

Bibliographie

- [BMMT97] B.LACAZE, C. MAILHES, M.MAUBOURGUET et J.-Y. TURNERET : *Probabilités et statistiques appliquées*. Cépaduès, 1997.
- [Cer16] Raphael CERF : Cours : Intégration et probabilités, 2016.
- [Gar02] B. GAREL : *Modélisation probabiliste et statistique*. Cépaduès, 2002.
- [Ipi] IPIPIPOURAX : Logo de la théorie des probabilités. <https://commons.wikimedia.org/w/index.php?curid=18155165>, dernier accès le 1/10/2023.
- [Ist99] Jacques ISTAS : *Probabilités et statistiques*. Ellipses, 1999.