Année universitaire 2002-2003

UNIVERSITÉ D'ORLÉANS

Olivier GARET

MA6.06 : Mesure et Probabilités

Table des matières

Chapitre 1

Un peu de théorie de la mesure

La théorie des probabilités décrit les événements comme des sous-ensembles d'un ensemble Ω représentant tous les résultats possibles *a priori* – même s'il peut s'avérer ensuite que certains n'arrivent jamais. Remarquons bien qu'il n'est pas possible de modéliser un phénomène aléatoire quelconque si l'on ne connaît pas les résultats possibles *a priori*.

Soit donc Ω un ensemble. Pour tout $A \subset \Omega$, on note A^c le complémentaire de A dans Ω :

$$A^c = \{x \in \Omega; x \notin A\}.$$

1.1 Tribus

1.1.1 Axiomes de base

On dit qu'une partie $\mathcal{F} \subset \mathcal{P}(\Omega)$ est une tribu si elle vérifie les propriétés suivantes :

1. $\varnothing \in \mathcal{F}$.
2. $\forall A \in \mathcal{F} \quad A^c \in \mathcal{F}$.
3. Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{F}$, $\bigcup\limits_{i=1}^{+\infty} A_i \in \mathcal{F}$.

1.1.2 Propriétés

Les propositions suivantes sont alors des conséquences relativement faciles des axiomes de base :

- $\Omega \in \mathcal{F}$.
- Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{F}$, $\bigcap\limits_{i=1}^{+\infty} A_i \in \mathcal{F}$.

- Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{F}$, $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$.
- Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{F}$, $\bigcap_{i=1}^{n} A_i \in \mathcal{F}$.

Une fois que Ω et $\mathcal{F}$ sont fixés, on appelle événement tout élément de $\mathcal{F}$.

Exercice :

Montrer qu'une partie $\mathcal{F} \subset \mathcal{P}(\Omega)$ est une tribu si elle vérifie les propriétés suivantes :

1. $\Omega \in \mathcal{F}$.
2. $\forall (A,B) \in \mathcal{F} \times \mathcal{F} \quad (A \subset B) \Longrightarrow (B \backslash A \in \mathcal{F})$.
3. $\forall (A,B) \in \mathcal{F}^2 \quad A \cup B \in \mathcal{F}$.
4. Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{F}$.

1.1.3 Sous-tribus

Si $\mathcal{A}$ est une tribu et que la partie $\mathcal{B} \subset \mathcal{A}$ est une tribu, alors on dit que $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$.[1]

1.1.4 Opérations sur les tribus

Intersection de tribus

Soit Ω un ensemble et T un ensemble de tribus sur Ω. T est supposé non vide.[2] Il peut être fini ou infini, voire même infini non dénombrable. Alors $\mathcal{F} = \bigcap_{\mathcal{A}\in T} \mathcal{A}$ est une tribu.

Démonstration. Il suffit de vérifier les 3 axiomes de base des tribus.

- $\forall \mathcal{A} \in T \quad \varnothing \in \mathcal{A}$. Donc $\varnothing \in \bigcap_{\mathcal{A}\in T} \mathcal{A} = \mathcal{F}$.
- Soit $A \in \mathcal{F}$. On doit montrer que $A^c \in \mathcal{F}$. Soit $\mathcal{A} \in T$. Comme les $A \in \mathcal{A}$ et que $\mathcal{A}$ est une tribu, $A^c \in \mathcal{A}$. Comme ceci est vrai pour tout $\mathcal{A} \in T$, on a $A^c \in \bigcap_{\mathcal{A}\in T} \mathcal{A} = \mathcal{F}$.
- Soit $(A_i)_{i\in I}$ une famille dénombrable d'éléments de $\mathcal{F}$. On doit montrer que $\bigcup_{i\in I} A_i \in \mathcal{F}$. Soit $\mathcal{A} \in T$. Comme les A_i sont dans $\mathcal{A}$ et que $\mathcal{A}$ est

[1] Une erreur classique à ne pas commettre : si $\mathcal{B}$ est une sous-tribu de $\mathcal{A}$, que $B \subset A$ avec $A \in \mathcal{A}$, alors rien ne permet d'affirmer que $B \in \mathcal{B}$.

[2] T est donc un ensemble d'ensembles d'ensembles, lol.

une tribu, $\bigcup_{i \in I} A_i \in \mathcal{A}$. Comme ceci est vrai pour tout $\mathcal{A} \in T$, on a $\bigcup_{i \in I} A_i \in \bigcap_{\mathcal{A} \in T} \mathcal{A} = \mathcal{F}$.

□

Tribu engendrée par une famille de tribus

Soit $(\mathcal{F}_i)_{i \in I}$ une famille de tribus sur Ω. L'ensemble des tribus contenant des tribus contenant toutes les $\mathcal{F}_i$ est non vide, puisque $\mathcal{P}(\Omega)$ est une telle tribu. D'après le résultat énoncé ci-dessus, l'intersection de toutes ces tribus est une tribu. Par construction, cette tribu est la plus petite tribu contenant toutes les $\mathcal{F}_i$. On la note

$$\bigwedge_{i \in I} \mathcal{F}_i.$$

Tribu engendrée par une famille d'ensembles

Soit $(A_i)_{i \in I}$ une famille de parties de Ω.

Pour tout i, la plus petite tribu contenant A_i est la tribu $\mathcal{F}_i = (\varnothing, A_i, A_i^c, \Omega)$. Ainsi, la plus petite tribu contenant les ensembles A_i est

$$\bigwedge_{i \in I} \mathcal{F}_i.$$

On note cette tribu $\sigma(A_i; i \in I)$.

Tribu produit

Soit $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces mesurés. On appelle tribu produit sur $\Omega \times \Omega'$ la tribu engendrée par les ensembles $A \times B \in \mathcal{A} \times \mathcal{B}$. On note $\mathcal{A} \otimes \mathcal{B}$ cette tribu.

Voyons maintenant la définition d'une probabilité sur $(\Omega, \mathcal{F})$.

1.2 Espace probabilisé

On appelle
- probabilité
- ou mesure de probabilité
- ou loi

sur $(\Omega, \mathcal{F})$ toute application

$$P : \mathcal{F} \to [0,1]$$

vérifiant les propriétés suivantes :

1. $P(\Omega) = 1$.
2. Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{F}$ deux à deux disjoints,
   $P(\underset{i=1}{\overset{+\infty}{\cup}} A_i) = \sum_{i=1}^{+\infty} P(A_i)$.

Alors, le triplet $(\Omega, \mathcal{F}, P)$ est appelé espace probabilisé.

On remarque qu'un espace probabilisé est très exactement un espace mesuré associée à une mesure positive de masse totale 1.

Remarque sur le vocabulaire : l'image $P(A)$ d'un événemement A par l'application P est appelée probabilité de cet événement. Ainsi le mot "probabilité" peut-il désigner à la fois une application et la valeur de cette application en un point. Le contexte doit permettre de lever toute ambiguité.

Les propositions suivantes sont alors des conséquences relativement faciles des définitions :

1. Pour toute suite $(A_i)_{1\leq i\leq n}$ d'éléments de $\mathcal{F}$ deux à deux disjoints,
   $P(\underset{i=1}{\overset{n}{\cup}} A_i) = \sum_{i=1}^{n} P(A_i)$.
2. $\forall A, B \in \mathcal{F} \quad (A \cap B = \varnothing) \Longrightarrow (P(A \cup B) = P(A) + P(B))$
3. $\forall A \in \mathcal{F} \quad P(A^c) = 1 - P(A)$
4. $P(\varnothing) = 0$
5. $\forall A, B \in \mathcal{F} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
6. $\forall A, B \in \mathcal{F} \quad P(A \cup B) \leq P(A) + P(B)$
7. $\forall A, B \in \mathcal{F} \quad (A \subset B) \Longrightarrow (P(A) \leq P(B))$
8. $\forall A, B \in \mathcal{F} \quad P(A \cap B) \leq \min(P(A), P(B))$
9. $\forall A, B \in \mathcal{F} \quad P(A \cup B) \geq \max(P(A), P(B))$
10. Pour toute suite $(A_i)_{i\geq 1}$ d'éléments de $\mathcal{F}$,
    $P(\underset{i=1}{\overset{+\infty}{\cup}} A_i) \leq \sum_{i=1}^{+\infty} P(A_i)$.
11. Si $(A_i)_{i\in\mathbb{N}}$ est une suite croissante d'événements
    (c'est à dire que $\forall n \in \mathbb{N} \quad A_n \subset A_{n+1}$))
    et que l'on pose $A = \underset{i=1}{\overset{+\infty}{\cup}} A_i$, alors la suite $(P(A_n))_{n\in\mathbb{N}}$ est monotone, croissante, et converge vers $P(A)$.
12. Si $(A_i)_{i\in\mathbb{N}}$ est une suite décroissante d'événements
    (c'est à dire que $\forall n \in \mathbb{N} \quad A_{n+1} \subset A_n$))
    et que l'on pose $A = \underset{i=1}{\overset{+\infty}{\cap}} A_i$, alors la suite $(P(A_n))_{n\in\mathbb{N}}$ est monotone, décroissante, et converge vers $P(A)$.

Démonstration. 1. Il suffit de poser $A_i = \varnothing$ pour $i \geq n+1$ et d'appliquer l'axiome 2.

2. Il suffit d'appliquer la propriété 1 avec $n = 2$, $A_1 = A$ et $A_2 = B$.
3. Il suffit d'appliquer la propriété 2 avec $B = A^c$: A et B sont disjoints donc $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$ d'après l'axiome 1.
4. On applique la propriété 3 avec $A = \Omega$: $P(\varnothing) = P(\Omega^c) = 1 - P(\Omega) = 1 - 1 = 0$.
5. Les ensembles $A \backslash B$, $B \backslash A$ et $A \cap B$ sont disjoints et leur réunion est $A \cup B$, donc l'après la propriété 1, on a

$$\begin{aligned}
P(A \cup B) &= P(A\backslash B) + P(B\backslash A) + P(A \cap B) \\
P(A \cup B) &= (P(A\backslash B) + P(A \cap B) + (P(B\backslash A) + P(A \cap B)) - P(A \cap B) \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}$$

car $A\backslash B$ $A \cap B$ sont disjoints, de réunion A, tandis que $B\backslash A$ et $A \cap B$ sont disjoints, de réunion B.
6. Il suffit d'appliquer la relation 5 en remarquant que $P(A \cap B) \geq 0$
7. Si $A \subset B$, on a B est la réunion disjointe de A et de $B\backslash A$. Donc $P(B) = P(A) + P(B\backslash A) \geq P(A)$.
8. $A \subset (A \cup B)$, donc d'après la propriété 6 $P(A) \leq P(A \cup B)$. De même $P(B) \leq P(A \cup B)$. Finalement $\max(P(A), P(B)) \leq P(A \cup B)$.
9. Posons $B_1 = A_1$ et, pour tout $n \geq 2$ $B_n = A_n \backslash (\bigcup_{i=1}^{n-1} B_i)$. Par construction, les $(B_k)_{k \geq 1}$ sont deux à deux disjoints. De plus, on peut montrer par récurrence sur n que

$$\forall n \geq 1 \quad \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i.$$

On en déduit

$$\bigcup_{i=1}^{+\infty} A_i = \bigcup_{i=1}^{+\infty} B_i.$$

Donc

$$\begin{aligned}
P(\bigcup_{i=1}^{+\infty} A_i) &= P(\bigcup_{i=1}^{+\infty} B_i) \\
&= \bigcup_{i=1}^{+\infty} P(B_i) \\
&\leq \bigcup_{i=1}^{+\infty} P(A_i).
\end{aligned}$$

10. Comme $A_n \subset A_{n+1}$, on a $P(A_n) \leq P(A_{n+1})$, donc la suite est croissante. Comme on a pour tout n : $A_n \subset A$, la suite $(P(A_n))_{n\geq 1}$ est majorée par $P(A)$. Posons $B_1 = A_1$ et, pour tout $n \geq 2$ $B_n = A_n \backslash A_{n-1}$. On a :
$$\forall n \geq 1 \quad A_n = \overset{n}{\underset{i=1}{\cup}} B_i$$
et
$$A = \overset{+\infty}{\underset{i=1}{\cup}} A_i = \overset{+\infty}{\underset{i=1}{\cup}} B_i$$
Ainsi,
$$\begin{aligned} P(A) &= P(\overset{+\infty}{\underset{i=1}{\cup}} B_i) \\ &= \overset{+\infty}{\underset{i=1}{\cup}} P(B_i) \\ &= \lim_{n\to+\infty} \sum_{i=1}^{n} P(B_i) \\ &= \lim_{n\to+\infty} P(\overset{n}{\underset{i=1}{\cup}} B_i) \\ &= \lim_{n\to+\infty} P(A_n) \end{aligned}$$

11. On applique le résultat précédent à la suite croissante $(A'_n)_{n\geq 1}$ définie par $A'_n = A_n^c$.

□

1.3 Partitions et probabilités

Le théorème très simple qui suit est très fréquemment utilisé. Il traduit le fait que pour calculer une probabilité, il faut parfois diviser les cas.

Théorème 1. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω. Alors on a*
$$\forall A \in \mathcal{F} \quad P(A) = \sum_{i\in I} P(A \cap \Omega_i).$$

Démonstration. Comme la famille $(\Omega_i)_{i\in I}$ une partition de Ω, la famille $(A \cap \Omega_i)_{i\in I}$ est une partition de A. A est donc réunion disjointe des $(A \cap \Omega_i)_{i\in I}$, donc $P(A) = \sum_{i\in I} P(A \cap \Omega_i)$. □

1.4 Probabilité conditionnelle

Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité et B un événement observable de probabilité non nulle. On appelle probabilité conditionnelle sachant B l'application

$$\begin{aligned} P(.|B) : \mathcal{F} &\rightarrow \mathbb{R} \\ A &\mapsto P(A|B) = \frac{P(A \cap B)}{P(B)} \end{aligned}$$

$P(A|B)$ se lit "Probabilité de A sachant B".
On a évidemment

$$P(A \cap B) = P(B)P(A|B). \tag{1.1}$$

Remarque : L'application "probabilité conditionnelle" est une probabilité. Elle vérifie donc toutes les propriétés énoncées précédemment.

1.4.1 Conditionnements en chaîne

Si A, B sont deux événements observables avec $A \subset B$ et $P(B) \neq 0$, la formule (1.1) devient

$$P(A) = P(B)P(A|B). \tag{1.2}$$

On a la généralisation suivante :

Théorème 2. *Soient $n \geq 2$ et $E_1, \dots, E_n$ des événements observables vérifiant*

$$E_n \subset E_{n-1} \subset \cdots \subset E_1$$

et $P(E_{n-1}) > 0$. Alors on a

$$P(E_n) = P(E_n|E_{n-1})P(E_{n-1}|E_{n-2}) \dots P(E_2|E_1)P(E_1)$$

Démonstration. La formule se montre par récurrence sur n. Pour $n = 2$, c'est une conséquence immédiate de (1.2). Pour $n > 2$, on applique d'abord la formule pour $n = 2$ aux événements E_n et E_{n-1} :

$$P(E_n) = P(E_n|E_{n-1})P(E_{n-1}),$$

puis on applique la propriété de récurrence au rang $n - 1$. □

Exemple : (d'après André Franquin) Chez les papous, il y a les papous à poux et les papous pas à poux. La probabilité pour qu'un papou ait des poux vaut 0.1. De plus, chez les papous, il y a les papous papas et les papous pas papas. La probabilité pour qu'un papou à poux soit papa vaut 0.6. Or, chez les poux, il y a les poux papas et les poux pas papas : la probabilité pour qu'un papou à poux possède au moins un pou papa est de 0.8.

Question : on tire au hasard un papou. Quelle est la probabilité pour que ce soit un papa papou à poux papa ? Réponse : $0.8 \times 0.6 \times 0.1 = 0.048$.

Ce théorème est parfois énoncé sous la forme plus compliquée – mais équivalente – suivante.

Théorème 3. *Soient $n \geq 2$ et $A_1, \dots, A_n$ des événements observables avec $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$. Alors*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = \left(\prod_{k=1}^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_{k+1} | A_1 \cap A_2 \cap \dots \cap A_k)\right) P(A_1)$$

Démonstration. Il suffit de poser, pour $1 \leq i \leq n$, $E_i = \bigcap_{k=1}^{i} A_k$ et d'appliquer le théorème précédent. □

1.4.2 Conditionnement par tous les cas possibles

Ceci est l'expression en termes de probabilités conditionnelles du principe de partition.

Théorème 4. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i \in I}$ une partition de Ω. Alors on a*

$$\forall A \in \mathcal{F} \quad P(A) = \sum_{i \in J} P(A|\Omega_i) P(\Omega_i),$$

où $J = \{i \in I; P(\Omega_i) > 0\}$.

Démonstration. D'après le théorème 1, on a

$$\begin{aligned} P(A) &= \sum_{i \in I} P(A \cap \Omega_i) \\ &= \sum_{i \in J} P(A \cap \Omega_i) + \sum_{i \in I \setminus J} P(A \cap \Omega_i) \\ &= \sum_{i \in J} P(A \cap \Omega_i) \\ &= \sum_{i \in J} P(A|\Omega_i) P(\Omega_i) \end{aligned}$$

□

1.4.3 Formule de Bayes

Théorème 5. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Soit I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω telle que pour tout $i \in I$, $P(\Omega_i)$ soit non nul. Soit A un élément de probabilité non nulle.*

Alors on a, pour tout $j \in I$, la formule

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

Démonstration.

$$\begin{aligned} P(\Omega_j|A) &= \frac{P(A\cap\Omega_j)}{P(A)} \\ &= \frac{P(A|\Omega_j)P(\Omega_j)}{P(A)} \end{aligned}$$

et on applique le théorème précédent.

□

Exemple :

- 60% des étudiants qui vont en T.D. obtiennent l'examen.
- 10% des étudiants qui ne vont pas en T.D. obtiennent l'examen.
- 70% des étudiants vont en T.D.

Quelle proportion des lauréats a séché les cours ? On note A l'événement "être assidu en cours". On a $P(A) = 0.7$, et donc $P(A^c) = 0.3$. On note L l'événement "obtenir l'examen" : on a $P(L|A^c) = 0.1$ et $P(L|A) = 0.6$. On a alors

$$P(A^c|L) = \frac{P(L|A^c)P(A^c)}{P(L|A^c)P(A^c) + P(L|A)P(A)} = \frac{0.1\times 0.3}{0.1\times 0.3 + 0.6\times 0.7} = \frac{3}{45} = \frac{1}{15}.$$

1.5 Indépendance

1.5.1 Événements indépendants

On dit que deux événements observables A et B sont indépendants si on a

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

Soit $(A_i)_{i\in G}$ une partie d'éléments de $\mathcal{F}$ indexés par un ensemble G. On dit que les événements constituant la famille $(A_i)_{i\in G}$ sont globalement indépendants si l'on a pour tout ensemble fini $I \subset G$:

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

1.5.2 Tribus indépendantes

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé ; $\mathcal{A}$ et $\mathcal{B}$ deux sous-tribus de $\mathcal{F}$. On dit que les tribus $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes sous P si

$$\forall A \in \mathcal{A} \quad \forall B \in \mathcal{B} \quad P(A\cap B) = P(A)P(B).$$

Plus généralement, si $(\mathcal{A}_i)_{i\in I}$ est une famille de sous-tribus de $\mathcal{F}$, on dit que cette famille est indépendante sous P si pour tout ensemble fini $J \subset I$, on a

$$\forall (A_i)_{i\in J} \in \prod_{i\in J} \mathcal{A}_i \quad P(\prod_{i\in J} A_i) = \prod_{i\in J} P(A_i).$$

Remarque : Si I est fini et que $(\mathcal{A}_i)_{i\in I}$ est une famille de sous-tribus de $\mathcal{F}$, cette famille est indépendante sous P si et seulement si on a

$$\forall (A_i)_{i\in I} \in \prod_{i\in I} \mathcal{A}_i \quad P(\prod_{i\in I} A_i) = \prod_{i\in I} P(A_i).$$

Il suffit en effet de poser $A_i = \Omega$ pour $i \in I\backslash I$ pour exprimer une intersection indexée par J en une intersection indexée par I.

Exercice : Soient $A, B \in \mathcal{F}$. Montrer que A est indépendant de B si et seulement si la tribu $\sigma(A)$ est indépendante de la tribu $\sigma(B)$.

Remarque utile : Si les tribus $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes sous P, que $\mathcal{A}'$ est une sous-tribu de $\mathcal{A}$ et $\mathcal{B}'$ est une sous-tribu de $\mathcal{B}$, alors les tribus $\mathcal{A}'$ et $\mathcal{B}$' sont indépendantes sous P.

1.5.3 Indépendance et tribus engendrées

Définition On dit qu'une famille $\mathcal{C}$ de parties de Ω est un π-système si

$$\forall (A,B) \in \mathcal{C} \quad A\cap B \in \mathcal{C}.$$

On donne maintenant un résultat général de théorie de la mesure très utile. Sa preuve, basée sur le théorème $\lambda - \pi$ de Dynkin, est admise ici.[3]

[3]Le lecteur intéressé pourra se référer la section 3.3 de l'ouvrage de Patrick Billingsley : *Probability and measure*, précisément aux théorèmes 3.2 et 3.3.

Proposition 1 (Critère d'identitification d'une probabilité). *Soit P et Q deux probabilités sur l'expace mesuré $(\Omega, \mathcal{F})$. On suppose qu'il existe un π-système $\mathcal{C}$ qui engendre $\mathcal{F}$ ($\sigma(\mathcal{C}) = \mathcal{F}$) et sur lequels P et Q coïncident, c'est à dire que*

$$\forall A \in \mathcal{C} \quad P(A) = Q(A).$$

Alors $P = Q$.

Théorème 6. *Soit $\mathcal{C}$ et $\mathcal{D}$ deux familles de parties mesurables de $(\Omega, \mathcal{F})$. On suppose que $\mathcal{C}$ et $\mathcal{D}$ sont des π-systèmes et que pour tout $(A, B) \in \mathcal{C} \times \mathcal{D}$, on a $P(A \cap B) = P(A)P(B)$. Alors, les tribus $\mathcal{A} = \sigma(\mathcal{C})$ et $\mathcal{B} = \sigma(\mathcal{D})$ sont indépendantes.*

Démonstration. Pour $A \in \mathcal{A}$, on pose $\mathcal{T}_A = \{B \in \mathcal{B}, P(A \cap B) = P(A)P(B)\}$.

Regardons d'abord le cas où $A \in \mathcal{C}$. Si $P(A) = 0$, alors A est indépendant de tout, donc $\mathcal{T}_A = \mathcal{B}$. Si $P(A) \neq 0$, on peut définir sur $\mathcal{B}$ la probabilité conditionnelle P_A par

$$\forall B \in \mathcal{B} \quad P_A(B) = \frac{P(A \cap B)}{P(A)}.$$

Les probabilités P et P_A coïncident sur $\mathcal{D}$. Comme $\mathcal{D}$ est un π-système qui engendre $\mathcal{B}$, P et P_A coïncident sur $\mathcal{B}$. On en déduit que lorsque $A \in \mathcal{C}$, on a $\mathcal{T}_A = \mathcal{B}$.

On a donc montré que si $\mathcal{C}$ et $\mathcal{D}$ sont des π-systèmes, alors

$$\forall (A, B) \in \mathcal{C} \times \mathcal{D} \quad P(A \cap B) = P(A)P(B) \Longrightarrow \forall (A, B) \in \mathcal{C} \times \sigma(\mathcal{D}) \quad P(A \cap B) = P(A)P(B).$$

Mais, $\mathcal{B} = \sigma(\mathcal{D})$ est lui-même un π-système. Le résultat que l'on vient de démontrer s'applique cette fois avec $(\mathcal{B}, \mathcal{C})$ à la place de $(\mathcal{C}, \mathcal{D})$, et on obtient que

$$\forall (A, B) \in \mathcal{C} \times \mathcal{D} \quad P(A \cap B) = P(A)P(B) \Longrightarrow \forall (A, B) \in \sigma(\mathcal{C}) \times \sigma(\mathcal{D}) \quad P(A \cap B) = P(A)P(B),$$

ce qui était notre objectif

□

Théorème 7. *Les deux propriétés suivantes sont équivalentes :*

- *Les tribus $(\mathcal{A}_i)_{i \in I}$ sont indépendantes*
- *Pour tout $j \in I$, la tribu $\mathcal{A}_j$ est indépendante de la tribu $\bigwedge_{i \in I \backslash \{j\}} \mathcal{A}_i$.*

Démonstration. – Preuve de $1 \Longrightarrow 2$: Soit $j \in I$. On considère le π-système $\mathcal{C}$ défini par

$$\mathcal{C} = \bigcup_{F \subseteq I \backslash \{j\}} \{ \bigcap_{x \in F} A_x ; \forall x \in F \quad A_x \in \mathcal{A}_x \}.$$

Il est facile de voir que $\mathcal{C}$ est un π-système qui engendre $\underset{i\in I\backslash\{j\}}{\wedge} \mathcal{A}_i$ et que $\forall (A,B)\in \mathcal{A}_j\times\mathcal{C} \quad P(A\cap B)=P(A)P(B)$. Le théorème 6 permet de conclure.

– Preuve de $2 \Longrightarrow 1$:

On montre par récurrence sur n la proposition $\mathcal{P}_n$

$$\mathcal{P}_n : \ |I|=n \Longrightarrow \forall \prod_{x\in I} A_x \in \prod_{x\in I}\mathcal{A}_x \quad P(\underset{x\in I}{\cap} A_x)=\prod_{x\in I} P(A_x).$$

Il est clair que $\mathcal{P}_0$ et $\mathcal{P}_1$ sont vraies. Montrons $\mathcal{P}_n \Longrightarrow \mathcal{P}_{n+1}$. Soit I un ensemble de cardinal $n+1$. On peut écrire $I=J\cup\{x_0\}$ avec $|J|=n$. Soit $E_1=\prod_{x\in I} A_x \in \prod_{x\in I}$. On a $E_1 = A_{x_0}\cap E_2$, avec $E_1=\underset{x\in J}{\cap} A_x$. Comme $A_{x_0}\in \mathcal{A}_{x_0}$ et $E_2\in \underset{i\in I\backslash\{x_0\}}{\wedge}\mathcal{A}_i$, l'hypothèse 2 implique $P(E_1)=P(A_{x_0})P(E_2)$. Mais d'après $\mathcal{P}_n$, on a

$$P(E_2)=P(\underset{x\in J}{\cap} A_x)=\prod_{x\in J} P(A_x),$$

d'où

$$P(E_1)=P(A_{x_0})P(E_2)=\prod_{x\in I} P(A_x),$$

ce qui achève la preuve. □

1.6 Exercices de théorie de la mesure

1. Pour n entier strictement positif, on note $A_n=n\mathbb{N}^*$. Notons $\mathcal{P}$ l'ensemble des nombres premiers positifs et $\mathcal{T}$ la sous-tribu de $(\mathbb{N},\mathcal{B}(\mathbb{N}))$ engendrée par les $(A_p)_{p\in\mathcal{P}}$.
   (a) Montrer que l'ensemble C des entiers qui sont premiers avec 2000 est $\mathcal{T}$-mesurable.
   (b) Montrer que l'ensemble $B=\{2^k ; k\in\mathbb{N}^*\}$ des puissances de deux est $\mathcal{T}$-mesurable.
2. Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants, tous de probabilité non nulle. On pose

$$A=\overline{\lim_{n\to+\infty}}\, A_n=\bigcap_{n=1}^{+\infty} A_n$$

et

$$B = \underline{\lim}_{n\to+\infty} A_n = \bigcup_{n=1} \bigcap_{k=n}^{+\infty} A_k.$$

Montrer que $P(A) = 0$ si et seulement si $P(B) = 0$

3. Le but de cette exercice est de montrer qu'il est impossible de construire un espace probabilisé $(\Omega, \mathcal{F}, P)$ et une variable aléatoire X sur cet espace tels que

$$\forall n \geq 1 \quad P(n \text{ divise } X) = \frac{1}{n}.$$

On note $p_1, \dots, p_n, \dots$ la suite des nombres premiers.

(a) Posons $E = \underline{\lim}_{n\to+\infty} \{p_n \text{ ne divise pas } X\}$. Montrer $E = \Omega$.

(b) On pose $D_n = \bigcap_{k=1}^{n} \{p_k \text{ ne divise pas } X\}$. Montrer

$$\frac{1}{P(D_n)} \geq \sum_{i=1}^{p_n} \frac{1}{i}$$

(c) En déduire que $P(D) = 0$, où on a posé $D = \bigcap_{k=1}^{+\infty} \{p_k \text{ ne divise pas } X\}$.

(d) Conclure.

4. Une enquête effectuée parmi les nouveaux adhérents du parti socialiste français a montré que les femmes représentaient $40,55\%$ des nouveaux adhérents. $20,4\%$ des nouvelles militantes socialistes sont enseignantes, tandis que seulement $12,81\%$ des nouveaux militants de sexe masculin sont enseignants. Parmi les enseignants qui militent nouvellement au parti socialiste, quelle est la proportion de femmes ?

Chapitre 2

Lois des variables et des vecteurs aléatoires

Rappel : si X est un espace topologique (par exemple un espace métrique), on appelle *tribu borélienne* de X et on note $\mathcal{B}(X)$ la tribu engendrée par la famille des ouverts de X.

2.1 Définition

Si $(\Omega, \mathcal{F}, P)$ est un espace probabilisé, on appelle variable aléatoire toute application mesurable de $(\Omega, \mathcal{F}, P)$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, où $\mathcal{B}(\mathbb{R})$ est la tribu borélienne de $\mathbb{R}$. De même, on appelle vecteur aléatoire toute application mesurable de $(\Omega, \mathcal{F}, P)$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, où $\mathcal{B}(\mathbb{R}^d)$ est la tribu borélienne de $\mathbb{R}^d$.

On appelle loi d'une variable aléatoire X definie sur $(\Omega, \mathcal{F}, P)$ la loi image de P par X. Cette loi est notée P_X Rappelons que cette loi image est une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par

$$\forall A \in \mathcal{B}(\mathbb{R}) \quad P_X(A) = P(X^{-1}(A))$$

Par définition, $X^{-1}(A) = \{\omega \in \Omega; X(\omega) \in A\}$. Afin de simplifier les écritures, on écrit toujours $\{X \in A\}$ à la place de $X^{-1}(A)$. Ainsi, on écrira le plus souvent $P(\{X \in A\})$ et même $P(X \in A)$ pour désigner $P_X(A)$.

Exemple : Soit P la mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $P = \frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_1$. P est une mesure positive, de masse totale 1 : c'est donc une probabilité. Considérons l'application $X : \mathbb{R} \to \mathbb{R}$ définie par $\forall \omega \in \mathbb{R} \quad X(\omega) = |\omega|$. Comme X est une application mesurable, X est une variable aléatoire. Pour

P-presque tout $\omega, X(\omega) \in \{0,1\}$. Ainsi, la loi de X sous P est

$$\begin{aligned} P_X &= P(X=0)\delta_0 + P(X=1)\delta \\ &= P(\{0\})\delta_0 + P(\{-1,1\})\delta_1 \\ &= \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1 \end{aligned}$$

Exemple : L'exemple qui suit ne paie pas de mine mais est cependant très instructif. Soit P la mesure sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie par $P = \frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_1$. On a vu que P était une probabilité. Considérons l'application $Y : \mathbb{R} \to \mathbb{R}$ définie par $\forall \omega \in \mathbb{R} \quad Y(\omega) = \omega$. Comme Y est une application mesurable (!), Y est une variable aléatoire. Il est facile de voir que la loi de Y sous P est tout simplement P. Ainsi, on voit que le problème de l'existence d'une variable aléatoire suivant une certaine loi se ramène à celui de l'existence de cette loi et relève donc des théories de la mesure.

2.1.1 Fonction de répartition

Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. On appelle fonction de répartition de X et on note F_X la fonction définie sur $\mathbb{R}^d$ par

$$\begin{aligned} \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d \quad F_X(t) &= P_X(]-\infty, t_1] \times]-\infty, t_2] \times]-\infty, t_d]) \\ &= P(X_1 \le t_1, X_2 \le t_2 \dots, X_d \le t_d) \end{aligned}$$

Théorème 8. *Si deux variables (ou vecteurs) aléatoires ont la même fonction de répartition,alors elles ont même loi.*

Démonstration. Si X et Y sont tels que $F_X = F_Y$, cela veut dire que P_X et P_Y coincident sur les ensembles de la forme $]-\infty, t_1] \times]-\infty, t_2] \times]-\infty, t_d]$. Or ces ensembles forment un π-système qui engendre $\mathcal{B}(\mathbb{R}^d)$, donc P_X et P_Y sont égales. □

C'est surtout en dimension 1 que la fonction de répartition est utile, car en dimension supérieure, ses propriétés sont plus difficiles à exprimer et les calculs sont souvent compliqués, voire infaisables. Nous allons juste nous contenter de donner quelques propriétés de la fonction caractéristique d'une variable aléatoire.

Propriétés de la fonction de répartition d'une variable aléatoire réelle

Théorème 9. *La fonction de répartition F_X d'une variable aléatoire vérifie les propriétés suivantes*

- *F_X est à valeurs dans* $[0, 1]$
- *F_X est croissante sur* $\mathbb{R}$.
- $\lim_{t \to -\infty} F_X(t) = 0$.
- $\lim_{t \to +\infty} F_X(t) = 1$.
- *En tout point, F_X est continue à droite.*
- *En tout point, F_X admet une limite à gauche.*

Démonstration. – Le premier point découle du fait que $F_X(t)$ est la probabilité d'un événement.

- Si $s \leq t$, on a $]-\infty, s] \subset]-\infty, s]$, d'où

$$F_X(s) = P(]-\infty, s]) \leq P(]-\infty, t]) = F_X(t).$$

- Posons pour $n \geq 1$, $A_n =]-\infty, -n]$, on a $A_{n+1} \subset A_n$ et $\bigcap_{n \geq 1} A_n = \varnothing$, d'où $\lim_{n \to +\infty} P_X(A_n) = P_X(\varnothing) = 0$. Soit $\varepsilon > 0$: d'après ce qui précède, il existe n tel que $P(A_n) < \varepsilon$. Comme F_X est croissante et positive, on a

$$t \leq -n \quad \Longrightarrow 0 \leq F_X(t) \leq F_X(-n) \leq \varepsilon,$$

ce qui prouve que $\lim_{t \to -\infty} F_X(t) = 0$.

- Posons pour $n \geq 1$, $A_n =]-\infty, n]$, on a $A_n \subset A_{n+1}$ et $\bigcup_{n \geq 1} A_n = \mathbb{R}$, d'où $\lim_{n \to +\infty} P_X(A_n) = P_X(\mathbb{R}) = 1$. Soit $\varepsilon > 0$: d'après ce qui précède, il existe n tel que $P(A_n) \geq < 1 - \varepsilon$. Comme F_X est croissante et majorée par 1, on a

$$t \geq n \quad \Longrightarrow 1 \geq F_X(t) \geq F_X(n) \geq 1 - \varepsilon,$$

ce qui prouve que $\lim_{t \to +\infty} F_X(t) = 1$.

- Soit $t \in \mathbb{R}$, $(t_n)_{n \geq 1}$ une suite décroissante convergeant vers t. Posons pour $n \geq 1$, $A_n =]-\infty, t_n]$, on a $A_{n+1} \subset A_n$ et $\bigcap_{n \geq 1} A_n =]-\infty, t]$, d'où $\lim_{n \to +\infty} F_X(t_n) = \lim_{n \to +\infty} P_X(A_n) = P_X(]-\infty, t]) = F_X(t)$. Comme cette égalité est obtenue pour toute suite décroissante convergeant vers t, ceci prouve que la limite à droite de F_X au point t est $F_X(t)$ (critère de

continuité séquentiel). Remarquons qu'on aurait pu également utiliser des critères analogues pour les preuves des deux propriétés précédentes et éviter ainsi l'emploi de ε.
– Toute fonction croissante admet une limite à gauche en tout point de l'intérieur de l'ensemble de définition.

□

2.1.2 Tribu engendrée par une ou plusieurs variables aléatoires

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé, X une variable aléatoire (ou un vecteur aléatoire) sur cette espace. On note

$$\sigma(X) = \{X^{-1}(A); A \in \mathcal{B}(\mathbb{R})\}.$$

Cette famille est une tribu. On dit que c'est la tribu engendrée par une variable aléatoire X.

De la même manière, on appelle tribu engendrée par une famille de variables $(X_i)_{i\in I}$ et on note $\sigma((X_i)_{i\in I})$ la tribu

$$\bigwedge_{i\in I} \sigma(X_I).$$

Exemple : Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. Alors, l'événement $\{X = Y\}$ est $\sigma(X, Y)$-mesurable.

Démonstration. On a

$$\begin{aligned} \{X = Y\} &= \bigcup_{k\in\mathbb{N}} \{X = Y\} \cap \{X = k\} \\ &= \bigcup_{k\in\mathbb{N}} \{X = k\} \cap \{Y = k\} \end{aligned}$$

Par définition de $\sigma(X)$, l'événement $\{X = k\}$ est $\sigma(X)$-mesurable. Comme $\sigma(X, Y)$ contient $\sigma(X)$, l'événement $\{X = k\}$ est $\sigma(X, Y)$-mesurable. De même, l'événement $\{Y = k\}$ est $\sigma(X, Y)$-mesurable. Comme $\sigma(X, Y)$ est une tribu, on en conclut que pour tout k, l'événement $\{X = k\} \cap \{Y = k\}$ est $\sigma(X, Y)$-mesurable, puis que l'événement $\{X = Y\}$ est $\sigma(X, Y)$-mesurable.

□

2.2 Indépendance des variables aléatoires

Définition : On dit que des variables aléatoires $(X_i)_{i\in I}$ sont indépendantes si les tribus $(\sigma(X_i))_{i\in I}$qu'elles engendrent sont indépendantes.

Exemple : Si X et Y sont deux variables aléatoires indépendantes, alors pour tout couple de boréliens A et B, on a

$$P(\{X \in A\} \cup \{Y \in B\} = P(X \in A)P(Y \in B).$$

Théorème 10. *Soient $(X_i)_{i\in I}$ une collection de vecteurs aléatoires aléatoires indépendants. On suppose que X_i est à valeurs dans $\mathbb{R}^{n_i}$. Soient $(f_i)_{i\in I}$ une famille d'applications telles que pour tout i, f_i soit une application mesurable de $\mathbb{R}^{n_i}$ dans $\mathbb{R}^{p_i}$. Alors, si on pose $Y_i = f_i(X_i)$ les variables aléatoires $(Y_i)_{i\in I}$ sont indépendantes.*

Démonstration. L'indépendance des variables aléatoires est en fait l'indépendance des tribus engendrées. Soit $B \in \mathcal{B}(\mathbb{R}^{p_i})$ un borélien. On a $\{Y_i \in B\} = \{X_i \in f_i^{-1}(B)\}$. Comme f_i est borélienne, $f_i^{-1}(B) \in \mathcal{B}(\mathbb{R}^{n_i})$, et donc $\{Y_i \in B\}$ est $\sigma(X_i)$-mesurable. Ceci prouve que $\sigma(Y_i)$ est une sous-tribu de $\sigma(X_i)$. Comme les tribus $(\sigma(X_i))_{i\in I}$ sont indépendantes, leur sous-tribus $(\sigma(Y_i))_{i\in I}$ le sont aussi. □

Exemple : Si X, Y et Z sont indépendantes, alors $\operatorname{ch} X, Y^2$ et Z^3 sont indépendantes.

Là, nous restons un peu sur notre faim. En effet, nous voudrions pouvoir dire aussi que $\operatorname{ch} X + Y^2$ est indépendante de Z^3. Pour cela, il faudrait que nous sachions que (X, Y) est indépendant de Z, auquel cas nous pourrions appliquer les fonctions $f(x, y) = \operatorname{ch} x + y^2$ et $g(z) = z^3$.

Par chance (!), ceci est vrai. En effet, on a le résultat suivant :

Théorème 11. *Soient $(\mathcal{A}_i)_{i\in I}$ une famille de sous-tribus de $(\Omega, \mathcal{F})$ indépendantes sous P. Soient $J \subset I$ et $K \subset I$ disjoints.*

Alors les tribus $\underset{j\in J}{\wedge} \mathcal{A}_j$ et $\underset{k\in K}{\wedge} \mathcal{A}_k$ sont indépendantes.

Démonstration. On considère le π-système $\mathcal{C}$ défini par

$$\mathcal{C} = \bigcup_{F\subseteq J} \{ \bigcap_{x\in F} A_x ; \forall x \in F \quad A_x \in \mathcal{A}_x\}.$$

ainsi que le π-système $\mathcal{D}$ défini par

$$\mathcal{C} = \bigcup_{F\subseteq K} \{ \bigcap_{x\in F} A_x ; \forall x \in F \quad A_x \in \mathcal{A}_x\}.$$

Si $B_1 \in \mathcal{C}$, B_2 peut s'écrire sous la forme $B_1 = \underset{x\in F}{\cap} A_x$ où $F \subseteq J$ et où $\forall x \in F \quad A_x \in \mathcal{A}_x$. De même, si $B_2 \in \mathcal{D}$, B_2 peut s'écrire sous la forme $B_2 = \underset{x\in F'}{\cap} A_x$ où $F' \subseteq J$ et où $\forall x \in F \quad A_x \in \mathcal{A}_x$. Ainsi

$$\begin{aligned} P(B_1 \cap B_2) &= P(\underset{x\in(F\cup F')}{\cap} A_x) \\ &= \prod_{x\in(F\cup F')} P(A_x) \\ &= \big(\prod_{x\in F} P(A_x)\big)\big(\prod_{x\in F'} P(A_x)\big) \\ &= P(B_1)P(B_2) \end{aligned}$$

Comme $\mathcal{C}$ est un π-système qui engendre $\underset{j\in J}{\wedge} \mathcal{A}_j$ et $\mathcal{D}$ un π-système qui engendre $\underset{j\in J}{\wedge} \mathcal{A}_j$, le théorème 6 permet de conclure.

□

Théorème 12. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $X_1, \dots, X_n$ n vecteurs aléatoires. Les deux propositions suivantes sont équivalentes*

1. *$X_1, \dots, X_n$ sont indépendantes.*
2. *$P_{X_1,\dots,X_n} = P_{X_1} \otimes \dots \otimes P_{X_n}$*

Démonstration. – Preuve de 1 $\Longrightarrow$ 2. On pose $\mathcal{C} = \prod_{i=1}^{n} \mathcal{B}(\mathbb{R}^{n_i})$. Soit $A = A_1 \times \dots \times A_n \in \mathcal{C}$. On a

$$\begin{aligned} P_{X_1,\dots,X_n}(A) &= P((X_1, \dots, X_n) \in A) \\ &= P(\overset{n}{\underset{i=1}{\cap}} \{X_i \in A_i\}) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \\ &= \prod_{i=1}^{n} P_{X_i}(A_i) \\ &= (P_{X_1} \otimes \dots \otimes P_{X_n})(A) \end{aligned}$$

Ainsi $P_{X_1,\dots,X_n}$ et $P_{X_1} \otimes \dots \otimes P_{X_n}$ coïncident sur un π-système qui engendre $\overset{n}{\underset{i=1}{\wedge}} \mathcal{B}(\mathbb{R}^{n_i})$. Il s'ensuit que ces deux mesures sont égales.

– Preuve de 1 $\Longrightarrow$ 2 Soient $B_1, \dots B_n$ quelconques tels que pour tout i B_i soit $\sigma(X_i)$-mesurable : alors, pour tout i, il existe un borélien A_i tel que $B_i = \{X_i \in A_i\}$. On pose $A = A_1 \times \cdots \times A_n \in \mathcal{C}$.

$$\begin{aligned} P(\overset{n}{\underset{i=1}{\cap}} B_i) &= P(\overset{n}{\underset{i=1}{\cap}} \{X_i \in A_i\}) \\ &= P_{X_1,\dots,X_n}(A) \\ &= (P_{X_1} \otimes \cdots \otimes P_{X_n})(A) \\ &= \prod_{i=1}^{n} P_{X_i}(A_i) \\ &= \prod_{i=1}^{n} P(X_i \in A_i) \\ &= \prod_{i=1}^{n} P(B_i), \end{aligned}$$

ce qui prouve l'indépendance des tribus.

□

Corollaire 1. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $X_1, \dots, X_n$ n vecteurs aléatoires. On suppose qu'il existe des mesures de probabilités $\mu_1, \dots, \mu_n$ telles que $P_{X_1,\dots,X_n} = \mu_1 \otimes \cdots \otimes \mu_n$. Alors*

1. *$X_1, \dots, X_n$ sont indépendantes.*
2. *Pour tout i, la loi de X_i sous P est μ_i. ($P_{X_i} = \mu_i$)*

Démonstration. Soit B un borélien.

$$\begin{aligned} P_{X_i}(B) &= P(X_i \in B) \\ &= P((X_1, \dots, X_n) \in \Omega \times \cdots \times \Omega \times B \times \Omega \dots \Omega) \\ &= P_{(X_1,\dots,X_n)}(\Omega \times \cdots \times \Omega \times B \times \Omega \dots \Omega) \\ &= (\mu_1 \otimes \cdots \otimes \mu_n)(\Omega \times \cdots \times \Omega \times B \times \Omega \dots \Omega) \\ &= \mu_1(\Omega) \times \cdots \times \mu_{i-1}(\Omega) \times \mu_i(B) \times \mu_{i+1}(\Omega) \dots \mu_n(\Omega) \\ &= \mu_i(B) \end{aligned}$$

Ainsi $P_{X_i} = \mu_i$. L'identité $P_{X_1,\dots,X_n} = \mu_1 \otimes \cdots \otimes \mu_n$. peut se réécrire $P_{X_1,\dots,X_n} = P_{X_1} \otimes \cdots \otimes P_{X_n}$ et il suffit alors d'appliquer le théorème précédent. □

2.2.1 Application : loi $0 - 1$ de Kolmogorov

Théorème 13. *Soit S un ensemble infini et $(\mathcal{A}_i)_{i \in S}$ une famille de tribus indépendantes sous la loi P. On pose*

$$\mathcal{T} = \underset{\Lambda \subseteq S}{\cap} \ \underset{k \in S \backslash \Lambda}{\wedge} \mathcal{A}_k.$$

$\mathcal{T}$ est appelée tribu de queue de la famille $(\mathcal{A}_i)_{i\in\Lambda}$. Alors

$$\forall A \in \mathcal{T} \quad P(A) \in \{0,1\}.$$

Démonstration. Posons $I = S \cup \{\infty\}$ et $\mathcal{A}_\infty = \mathcal{T}$. Montrons que les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes. Soit $J \subseteq I$, et pour tout $i \in J$, $A_i \in \mathcal{A}_i$. Si $\{\infty\}$, l'indépendance des tribus $(\mathcal{A}_i)_{i\in S}$ assure que

$$P(\bigcap_{i\in J} A_i) = \prod_{i\in J} P(A_i).$$

Sinon, par définition de la tribu $\mathcal{T}$, on a $A_\infty \in \bigwedge_{k\in S\setminus J} \mathcal{A}_k$.

Comme $\bigcap_{i\in J\setminus\{\infty\}} \in \bigwedge_{k\in J\setminus\{\infty\}} \mathcal{A}_k$, le théorème 11 implique que A_∞ et $\bigcap_{k\in J\setminus\{\infty\}} \mathcal{A}_k$ sont indépendants. Il s'ensuit que

$$\begin{aligned} P(\bigcap_{i\in J} A_i) &= P(A_\infty \cap \bigcap_{i\in J\setminus\{\infty\}} A_i) \\ &= P(A_\infty)P(\bigcap_{i\in J\setminus\{\infty\}} A_i) \\ &= P(A_\infty) \prod_{i\in J\setminus\{\infty\}} P(A_i) \\ &= \prod_{i\in J} P(A_i). \end{aligned}$$

Ainsi, les tribus $(\mathcal{A}_i)_{i\in I}$ sont indépendantes. En particulier, la tribu $\mathcal{A}_\infty = \mathcal{T}$ est indépendante de la tribu $\bigwedge_{k\in S} \mathcal{A}_k$. Mais $\mathcal{T}$ est une sous-tribu de $\bigwedge_{k\in S} \mathcal{A}_k$, donc $\mathcal{T}$ est indépendante d'elle même. Soit donc $A \in \mathcal{T}$: on a

$$0 = P(\varnothing) = P(A \cap A^c) = P(A)P(A^c) = P(A)(1 - P(A)),$$

donc $P(A) \in \{0,1\}$. □

2.3 Variables aléatoires discrètes

On dit qu'une loi μ est discrète s'il existe un ensemble D fini ou dénombrable inclus dans $\mathbb{R}^d$ tel que $\mu(D) = 1$.

De même, on dit qu'une variable aléatoire X définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ est discrète si sa loi P_X est discrète.

Ainsi, si D est un ensemble dénombrable tel que $P(X \in D) = P_X(D) = 1$ et l'on pose

$$\forall i \in D \quad p_i = P(X = i)$$

La famille $(p_i)_{i \in D}$ est une famille de réels positifs vérifiant

$$\sum_{i \in D} p_i = 1.$$

La connaissance de D et des p_i permet de reconstituer la loi de X. En effet, on a le théorème suivant :

Théorème 14. *Soit X est une variable aléatoire discrète, et D un ensemble D fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Pour $i \in D$, on pose $p_i = P(X = i)$. Alors,*

1. *pour tout $A \in \mathcal{B}(\mathbb{R})$, on a*

$$P_X(A) = \sum_{i \in D \cap A} p_i.$$

2. $P_X = \sum_{i \in D} p_i \delta_i$.
3. *P_X admet comme densité par rapport à la mesure de comptage sur D la fonction $f(x)$ définie par*

$$f(x) = \begin{cases} p_x & si\ x \in D \\ 0\ si\ x \notin D \end{cases}$$

Démonstration. On note m la mesure de comptage sur D Soit A un borélien. $\{X \in A \cap D\}$ est réunion dénombrable disjointe des événements $\{X = i\}$, où i décrit $A \cap D$. On peut donc écrire

$$\begin{aligned} P_X(A) &= P(X \in A) \\ &= P(X \in A \backslash D) + P(X \in A \cap D) \\ &= 0 + P(X \in A \cap D) \\ &= \sum_{i \in A \cap D} P(X = i) \\ &= \sum_{i \in A \cap D} p_i \end{aligned}$$

Posons $\mu = \sum_{i \in D} p_i \delta_i$.

$$\begin{aligned} \mu(A) &= \sum_{i \in D} p_i \delta_i(A) \\ &= \sum_{i \in D} p_i \mathbb{1}_A(i) \end{aligned}$$

On a d'une part

$$\begin{aligned} \mu(A) &= \sum_{i \in D} p_i \mathbb{1}_A(i) \\ &= \sum_{i \in D \cap A} p_i \mathbb{1}_A(i) + \sum_{i \in D \setminus A} p_i \mathbb{1}_A(i) \\ &= \sum_{i \in D \cap A} p_i + 0 \end{aligned}$$

Comme A est quelconque, on en déduit que $\mu = P_X$. D'autre part

$$\begin{aligned} \mu(A) &= \sum_{i \in D} p_i \mathbb{1}_A(i) \\ &= \int_D p_i \mathbb{1}_A(i)\ dm(i) \\ &= \int_{\mathbb{R}} f(i) \mathbb{1}_A(i)\ dm(i) \\ &= \int_A f(x)\ dm(x), \end{aligned}$$

ce qui signifie que μ (c'est à dire P_X) admet f comme densité par rapport à la mesure de comptage.

□

On a la réciproque suivante :

Théorème 15. *Soit D un ensemble fini ou dénombrable, $(p_i)_{i \in D}$ une famille de réels positifs vérifiant*

$$\sum_{i \in D} p_i = 1.$$

Alors, on peut construire un espace probabilisé $(\Omega, \mathcal{F}, P)$ *est une variable aléatoire* X *sur cet espace telle que*

$$\forall i \in D \quad p_i = P(X = i)$$

Démonstration. Comme on l'a déjà remarqué, le problème d'existence d'une variable aléatoire se ramène souvent à l'existence d'une loi. Ici,on peut prendre $\Omega = D$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $X(\omega) = \omega$ avec

$$P = \sum_{i \in D \cap A} p_i \delta_i.$$

□

2.3.1 Fonction d'une variable aléatoire discrète

Théorème 16. *La loi image* μ_f *d'une loi discrète* μ *par une application mesurable* f *est une loi discrète.*

Démonstration. Soit D un ensemble fini ou dénombrable tel que $\mu(D) = 1$. $f(D)$ est un ensemble fini ou dénombrable. Il est donc mesurable par rapport à la mesure de Lebesgue. On a $\mu_f(f(D)) = \mu(f^{-1}(f(D)) \geq \mu(D) = 1$ car $f^{-1}(f(D) \supset D$. Il s'ensuit que $f(D)$ est un ensemble fini ou dénombrable dont la mesure sous μ_f est 1. □

Corollaire 2. *Soit* X *une variable aléatoire discrète définie sur un espace probabilisé* $(\Omega, \mathcal{F}, P)$ *et* f *une fonction quelconque définie sur* $X(\Omega)$*. Alors, la fonction* Y *définie par*

$$\forall \omega \in \Omega \quad Y = f(X(\omega))$$

est une variable aléatoire discrète sur $(\Omega, \mathcal{F}, P)$*.*

De manière plus concise, on écrit $Y = f(X)$.

Exemple : Soit X une variable aléatoire vérifiant $X(\Omega) = \{-1; 0; 1\}$, avec $P(X = -1) = P(X = 0) = P(X = 1) = \frac{1}{3}$.

On pose $Y = X^2$.

On a $Y(\Omega) = \{0; 1\}$, avec

$$\{Y = 0\} = \{X = 0\}$$

et

$$\{Y = 1\} = \{X = 1\} \cup \{X = -1\},$$

d'où

$$P(Y = 0) = P(X = 0) = \frac{1}{3}$$

et

$$P(Y = 1) = P(X = 1) + P(X = -1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

2.4 Variables et vecteurs aléatoires à densité

On dit qu'une loi μ est à densité (sous-entendu par rapport à la mesure de Lebesgue) s'il existe une fonction mesurable f qui soit une densité de μ par rapport à la mesure de Lebesgue.

Ainsi, si f est une densité de la loi μ, on a pour tout borélien A

$$\mu(A) = \int_A f(\omega)d\lambda(\omega).$$

Exercice : Montrer que si f est la densité d'une loi, alors $\lambda(f < 0) = 0$.

Évidemment, si f est la densité d'une loi, on a

$$1 = \mu(\mathbb{R}^d) = \int_{\mathbb{R}^d} f(\omega)d\lambda(\omega).$$

Réciproquement, si f est une fonction mesurable, positive λ presque partout et d'intégrale 1, $\mu = f.\lambda$ est une mesure de probabilité admettant f pour densité.

Ainsi, on dit qu'une variable (ou un vecteur) aléatoire X est à densité si sa loi P_X est à densité.

2.4.1 Premières propriétés

Soit X une variable aléatoire de densité f. f a les propriétés suivantes :

- $\forall a, b \in \mathbb{R} \quad a \leq b \Longrightarrow \quad P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x)\ dx$.
- $\forall a \in \mathbb{R} \quad P(a \leq X) = P(a < X) = \int_a^{+\infty} f(x)\ dx$.
- $\forall a \in \mathbb{R} \quad P(a \geq X) = P(a > X) = \int_{-\infty}^a f(x)\ dx$.
- $\int_{-\infty}^{+\infty} f(x)\ dx = 1$.

2.4.2 Densités et lois marginales

Théorème 17. *Soit $(\Omega, \mathcal{F}, \mu)$ et $(\Omega', \mathcal{F}', \mu')$ deux espaces mesurés. On suppose que la loi ν sur $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ admet une densité h par rapport à $\mu \otimes \mu'$. Alors la loi image de ν par l'application*

$$\begin{array}{rcl} \pi : \Omega \times \Omega' & \rightarrow & \Omega \\ (x, x') & \mapsto & x' \end{array}$$

admet comme densité par rapport à μ la fonction f définie par $f(x) = \int_{\Omega'} h(x, x')d\mu'(x')$.

Démonstration. Soit $B \in \mathcal{F}$

$$\begin{aligned}
\pi\nu(B) &= \nu(\pi^{-1}(B)) \\
&= \nu(B \times \Omega') \\
&= \int_{B\times\Omega'} d\nu(x,x') \\
&= \int_{B\times\Omega'} h(x,x')d(\mu\otimes\mu')(x,x') \\
&= \int_B (\int_{\Omega'} h(x,x')d\mu(x'))d\mu(x) \\
&= \int_B f(x)d\mu(x),
\end{aligned}$$

ce qui prouve le résultat. □

Théorème 18. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. Si $h(x,y)$ est une densité de (X,Y) par rapport à la mesure de Lebesgue sur $\mathbb{R}^{n+p}$, alors X admet la densité $f(x) = \int_{\mathbb{R}^p} h(x,y)d\lambda^p(y)$, tandis que*
Y admet la densité $g(x) = \int_{\mathbb{R}^n} f(x,y)d\lambda^n(x)$.

Démonstration. Le diagramme commutatif ci-dessous traduit que
$X = \pi \circ (X,Y)$.

$$\begin{array}{ccc}
(\Omega,\mathcal{F},P) & \overset{(X,Y)}{\longrightarrow} & (\mathbb{R}^n\times\mathbb{R}^p, \mathcal{B}(\mathbb{R}^n)\otimes\mathcal{B}(\mathbb{R}^p), P_{(X,Y)}) \\
& X \searrow & \downarrow \pi \\
& & (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_X)
\end{array}$$

Il s'ensuit que P_X est la mesure image de $P_{(X,Y)}$ par π. Comme $h(x,y)$ est la densité de (X,Y) par rapport à $\lambda^n \otimes \lambda^p = \lambda^{n+p}$, la densité de X est bien $f(x) = \int_{\mathbb{R}^p} h(x,y)d\lambda^p(y)$, On procède de même pour calculer la densité de Y. □

2.4.3 Indépendance et densités

Théorème 19. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On suppose que X admet une densité f et Y une densité g. Alors, le vecteur aléatoire (X,Y) admet la fonction $h(x,y) = f(x)g(y)$ comme densité.*

Démonstration. Comme X et Y sont indépendantes, sous P, on a $P_{X,Y} = P_X \otimes P_Y$. Ainsi

$$\begin{aligned} P_{X,Y} &= P_X \otimes P_Y \\ &= f\lambda^n \otimes g\lambda^p \\ &= ((x,y) \mapsto f(x)g(y))\lambda^n \otimes \lambda^p \\ &= h\lambda^{n+p}, \end{aligned}$$

ce qui montre bien que (la loi de) (X,Y) admet h comme densité. □

Théorème 20. *Soit X un vecteur aléatoire à valeurs dans $\mathbb{R}^n$ et Y un vecteur aléatoire à valeurs dans $\mathbb{R}^p$ définis sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On suppose que (X,Y) admet une densité h_1 qui s'écrive sous la forme $h_1(x,y) = f_1(x)g_1(y)$, où f et g sont des fonctions positives. Alors, X et Y sont indépendantes ; X admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ la fonction*

$$f(x) = \frac{f_1(x)}{\int_{\mathbb{R}^n} f_1(x')\ d\lambda^n(x')}$$

et Y admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^n$ la fonction

$$g(y) = \frac{g_1(y)}{\int_{\mathbb{R}^p} g_1(y')\ d\lambda^p(y')}$$

Démonstration. Posons $A = \int_{\mathbb{R}^n} f_1(x)d\lambda^n(x)$ et $B = \int_{\mathbb{R}^p} g_1(y)d(\lambda^p)(y)$. Comme f_1 et f_2 sont positives, on a $A \geq 0$ et $B \geq 0$. Comme h est une densité, on a

$$\begin{aligned} 1 &= \int_{\mathbb{R}^n \times \mathbb{R}^p} h(x,y)d(\lambda^n \otimes \lambda^p)(x,y) \\ &= \int_{\mathbb{R}^n \times \mathbb{R}^p} f_1(x)g_1(y)d(\lambda^n \otimes \lambda^p)(x,y) \\ &= (\int_{\mathbb{R}^n} f_1(x)d\lambda^n(x))(\int_{\mathbb{R}^p} g_1(y)d(\lambda^p)(y)) \\ &= AB \end{aligned}$$

On en peut donc dire que A et B sont tous deux strictement positifs. Ainsi les fonctions f et g sont bien définies. Elles sont positives, et, par construction, chacune d'entre elle admet 1 comme intégrale par rapport à la mesure de Lebesgue. Ainsi $\mu = f\lambda^n$ et $\nu = g\lambda^p$ sont des mesures de probabilité.

Comme dans le théorème précédent, la densité de $\mu \otimes \nu =$ est $h(x,y) = f(x)g(y)$. Donc $h(x,y) = f(x)g(y) = \frac{f_1(x)}{A}\frac{g_1(y)}{B} = \frac{f_1(x)g_1(y)}{AB} = h_1(x,y)$.

On en déduit $P_{(X,Y)} = \mu \otimes \nu$.

Il suffit d'appliquer le corollaire 1 pour conclure.

□

2.5 Variables et lois discrètes classiques

2.5.1 Indicatrice d'un évévenement

On rappelle que pour $A \subset \Omega$, l'application $1\!\!1_A$ (appelée indicatrice de A) est définie sur Ω par

$$1\!\!1_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

1_A est une variable aléatoire à valeurs dans $\{0;1\}$.

2.5.2 Masse de Dirac

On appelle masse de Dirac en un point $x \in \Omega$ la mesure δ_x définie par

$$\delta_x(A) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

C'est bien une loi car elle est positive et $\delta_x(\Omega) = 1$.

Remarque : Si Ω est un groupe abélien $\delta_{x+y} = \delta_x * \delta_y = \delta_y * \delta_x$.

2.5.3 Loi de Bernoulli

On appelle loi de Bernoulli de paramètre p la loi $\mu = (1-p)\delta_0$.

Ainsi, on dit qu'une variable aléatoire X suit la loi de Bernoulli de paramètre p si on a $P(X=1) = p$ et $P(X=0) = 1-p$.

Remarques importantes :

- Pour tout évènement A, $1\!\!1_A$ suit la loi de Bernoulli de paramètre $P(A)$.
- Réciproquement, si une variable aléatoire X suit une loi de Bernoulli, elle vérifie $X = 1\!\!1_{X=1}$.(Réfléchir un peu...)

Ainsi les variables aléatoires qui suivent des lois de Bernoulli sont exactement les indicatrices d'événements.

2.5.4 Loi uniforme sur un ensemble

Soit $E \subset \Omega$ un ensemble fini. On appelle loi uniforme sur E la loi définie sur $\mathcal{P}(\Omega)$ par

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to & [0,1] \\ A \mapsto \dfrac{|A \cap E|}{|E|} & & \end{array}$$

Ainsi, une variable aléatoire X suit la loi uniforme sur E si l'on a

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

Exemple : La variable aléatoire X représentant le résultat du lancer d'un dé non truqué suit la loi uniforme sur l'ensemble $\{1, 2, 3, 4, 5, 6\}$.

2.5.5 Loi binomiale

On appelle loi binomiale de paramètres n et p et on note $\mathcal{B}(n, p)$ la loi de la somme de n variables de Bernoulli indépendantes de même probabilité p.

Ainsi $\mathcal{B}(n, p) = (\mathrm{Ber}(p))^{*n}$.

Théorème 21. *$\mathcal{B}(n, p)$ charge les entiers $\{0, \dots, n\}$. Plus précisément, on a*

$$\forall k \in \{0, \dots, n\} \quad \mathcal{B}(n, p)(\{k\}) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Démonstration. Posons $\mu = \mathcal{B}(n, p)$. On a

$$\begin{array}{rcl} \mu & = & (\mathrm{Ber}(p))^{*n} \\ & = & ((1 - p)\delta_0 + (1 - p)\delta_1)^{*n} \end{array}$$

Comme l'ensemble des mesures positives munie de (+,*) est une algèbre commutative, la formule du binôme de Newton s'applique et l'on a

$$\begin{aligned}
\mu &= ((1-p)\delta_0 + p\delta_1)^{*n} \\
&= \sum_{k=0}^{n} \binom{n}{k} ((1-p)\delta_0)^{*(n-k)} * (p\delta_1)^{*k} \\
&= \sum_{k=0}^{n} \binom{n}{k} ((1-p)^{n-k}(\delta_0)^{*(n-k)}) * (p^k(\delta_1)^{*k}) \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_0^{*(n-k)} * \delta_1^{*k} \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_0 * \delta_k \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k(1-p)^{n-k}\delta_k
\end{aligned}$$

□

Corollaire 3. *Soit $A_1, \dots, A_n$ n événements indépendants de même probabilité p. On pose*

$$X = \sum_{k=1}^{n} \mathbb{1}_{A_k}.$$

Alors X suit la loi binomiale de paramètres n et p .

Remarque : X est le nombre de A_k qui sont réalisés.

Exemple : On lance n fois une pièce de monnaie équilibrée. Le nombre X de "pile" obtenus suit une loi binomiale de paramètres n et $\frac{1}{2}$.

2.5.6 Loi géométrique

On dit qu'une variable aléatoire X suit la loi géométrique de paramètre p si l'on a

$$\forall k \in \mathbb{N}^* \quad P(X = k) = p(1-p)^{n-1}.$$

Théorème 22. *Soit $(A_i)_{i\in\mathbb{N}^*}$ une suite d'événements indépendants de même probabilité p. On pose*

$$X(\omega) = \inf\{k \in \mathbb{N}^* \quad \omega \in A_k\}.$$

Alors X suit la loi géométrique de paramètre p . De plus

$$\forall k \in \mathbb{N} \qquad F_X(k) = 1 - (1-p)^k.$$

Démonstration.

$$\{X > k\} = \bigcap_{i \in \{1,\dots,k\}} A_i^c.$$

Comme les A_i sont dans $\mathcal{F}$, les A_i^c le sont aussi, et donc comme on peut l'écrire comme intersection finie d'éléments de $\mathcal{F}$,$\{X > k\}$ est dans $\mathcal{F}$, et donc, par passage au complémentaire

$$\{X \leq k\} \in \mathcal{F}.$$

En utilisant l'indépendance des (A_i), on obtient

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

Comme

$$\{X = +\infty\} = \bigcap_{k \geq 1} \{X > k\},$$

on obtient, par continuité séquentielle décroissante

$$P(X = +\infty) = \lim_{k \to +\infty} P(\{X > k\}) = \lim_{k \to +\infty} (1-p)^k = 0.$$

Ainsi X est bien une variable aléatoire, et l'on a

$$\forall k \in \mathbb{N}^* \quad P(X = k) = P(X > k-1) - P(X > k) = (1-p)^{k-1} - (1-p)^k = (1-p)^{k-1}p$$

De plus

$$\forall k \in \mathbb{N} \quad F_X(k) = P(X \leq k) = 1 - P(X > k) = 1 - (1-p)^k.$$

$\square$

Exemple : On lance une pièce de monnaie équilibrée jusqu'à obtention de "pile". Le nombre de lancers effectués suit une loi géométrique de paramètre $\frac{1}{2}$.

2.5.7 Loi de Poisson

On dit qu'une variable aléatoire X suit la loi de Poisson de paramètre λ si l'on a

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

La construction de telles variables est bien possible car $e^{-\lambda} \frac{\lambda^k}{k!} \geq 0$ et

$$\sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

2.5.8 Loi hypergéométrique

La loi hypergéométique $\mathcal{H}(N, n, k)$ modélise le phénomène suivant : on tire au hasard k individus dans une population de N individus, et l'on compte le nombre d'individus possédant une certaine particularité, sachant qu'il y a exactement n personnes dans la population totale qui possédaient cette particularité.

De manière théorique, la loi hypergéométrique est la loi image de la loi uniforme sur $\Omega = \mathcal{B}(N, k)$ par l'application

$$\begin{array}{rcl} X : \mathcal{B}(N,k) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1,\dots,n\}) \cap \omega| \end{array}$$

Ainsi pour $i \in \{0, \dots, \min(n, k)\}$, on a

$$\mathcal{H}(N,n,k)(i) = \frac{\binom{n}{i}\binom{N-n}{k-i}}{\binom{N}{k}}$$

Démonstration. Notons P la loi uniforme sur Ω. On a

$$\mathcal{H}(N,n,k)(i) = P(\omega \in S),$$

où $S = \{\omega \in \mathcal{B}(N, k); |\{1, \dots, n\} \cap \omega| = i$. L'application

$$\begin{array}{rcl} \mathcal{B}(\{1,\dots,n\}N,i) \times \mathcal{B}(\{n+1,\dots,N\},k-i) & \to & S \\ (A,B) \mapsto A \cup B & & \end{array}$$

est une bijection, donc

$$|S| = |\mathcal{B}(\{1,\dots,n\}N,i) \times \mathcal{B}(\{n+1,\dots,N\},k-i)| = \binom{n}{i}\binom{N-n}{k-i}$$

Comme P est la loi uniforme sur Ω, et que $|\Omega| = \binom{N}{k}$, le résultat s'ensuit. □

2.6 Lois à densité usuelles

2.6.1 Loi uniforme sur un compact de $\mathbb{R}^d$

On dit qu'une variable aléatoire X suit la loi uniforme sur un compact K de $\mathbb{R}^d$ $[a, b]$ si elle admet la densité

$$x \mapsto \frac{1}{\lambda(K)} \mathbb{1}_K(x)$$

2.6.2 Loi uniforme sur un intervalle

On dit qu'une variable aléatoire X suit la loi uniforme sur l'intervalle $[a, b]$ si elle admet la densité

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

2.6.3 Loi gaussienne de paramètres m et σ^2

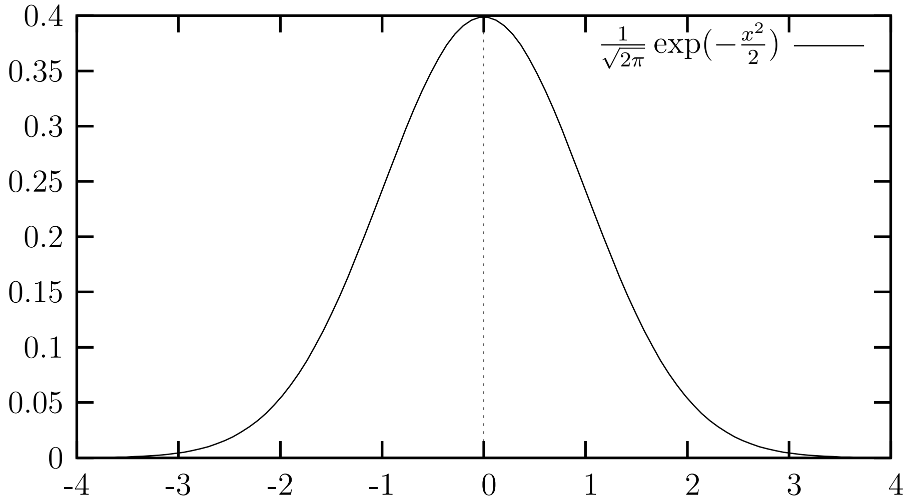
Soit $m \in \mathbb{R}$ et $\sigma^2 > 0$. On dit qu'une variable aléatoire X suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ^2 si elle admet la densité

$$x \mapsto \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(x-m)^2}{2\sigma^2}).$$

On emploie également parfois le mot "normale" à la place de "gaussienne" : ces deux mots signifient exactement la même chose. gamma On dit qu'une variable gaussienne est centrée lorsque $m = 0$.
On dit qu'une variable gaussienne est centrée lorsque $\sigma^2 = 1$.

Quelques résultats qui seront prouvés ultérieurement : si $X \sim \mathcal{N}(m, \sigma^2)$, alors $aX + b \sim \mathcal{N}(m + b, a^2\sigma^2)$. En particulier, si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$; et si $X \sim \mathcal{N}(0, 1)$, alors $\sigma X + m \sim \mathcal{N}(m, \sigma^2)$.

Densité de la loi normale $\mathcal{N}(0, 1)$

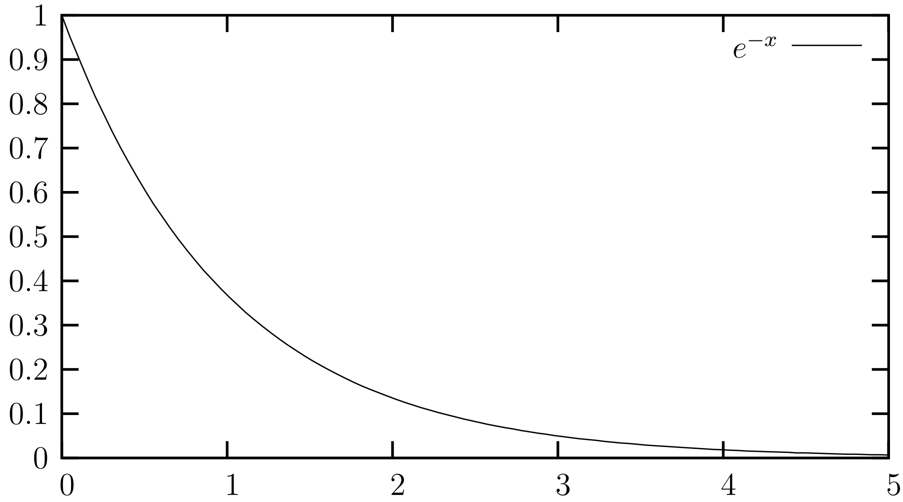


2.6.4 Loi exponentielle de paramètres a

Soit $a > 0$. On dit qu'une variable aléatoire X suit la loi exponentielle de paramètres a si elle admet la densité

$$x \mapsto a \exp(-ax) \mathbb{1}_{\mathbb{R}^+}(x).$$

Densité de la loi exponentielle de paramètre 1

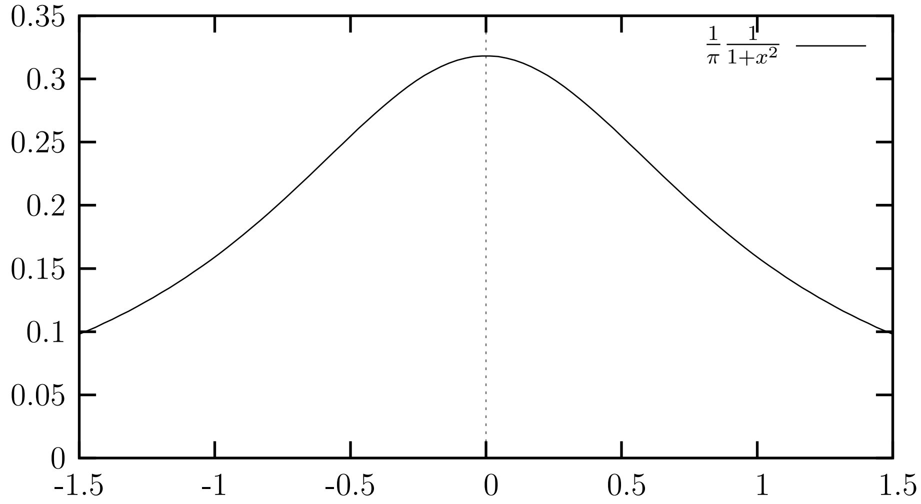


2.6.5 Lois de Cauchy

Soient $a \in \mathbb{R}, b > 0$. La loi de Cauchy $\mathcal{C}(a,b)$ admet comme densité par rapport à la mesure de Lebesgue :

$$x \mapsto \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}.$$

Densité de la loi de Cauchy $\mathcal{C}(0,1)$



2.6.6 Lois Gamma

Soient a et λ des réels strictement positifs. On appelle loi Gamma $\Gamma(a,\lambda)$ la loi dont la densité par rapport à la mesure de Lebesgue est

$$x \mapsto \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{]0,+\infty[}(x),$$

où $\Gamma(a)$ est la valeur au point a de la fonction Γ, définie par

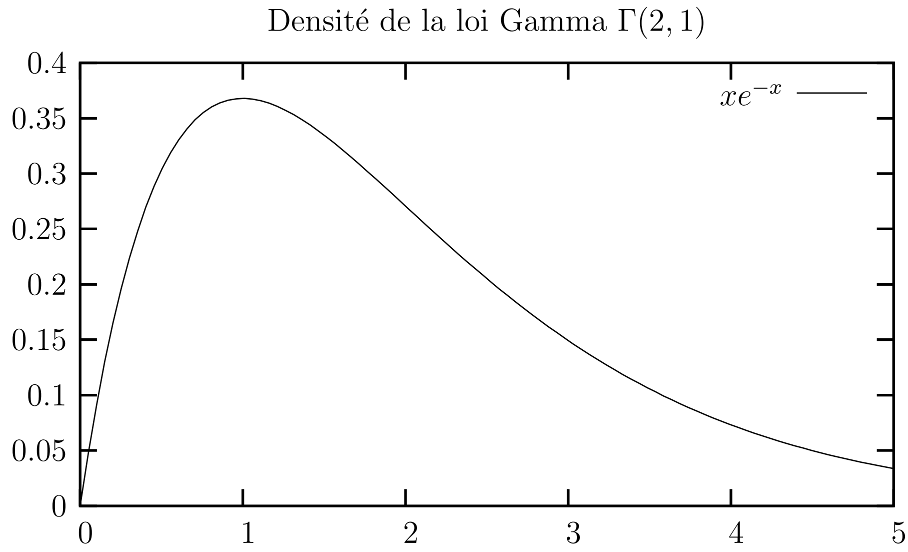
$$\Gamma(a) = \int_{\mathbb{R}+} x^{a-1} e^{-x}\, dx.$$

On dit parfois que a est le paramètre de forme et λ le paramètre d'échelle de la loi. En effet, on montrera plus loin que si $X \sim \Gamma(a,\lambda)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a, \lambda\mu)$.

On rappelle quelques propriétés classiques de la fonction Γ qui seront utiles dans la suite :

- $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
- $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

La preuve de ces deux propriétés sera vue en exercice.



2.7 Exercices sur les lois

1. Un jeu consiste à effectuer une mise en choisissant un nombre entre 1 et 6, puis à lancer simultanément trois dés. Si le numéro choisi sort une fois, le joueur récupère sa mise plus une somme égale à sa mise. Si le numéro choisi sort deux fois, le joueur récupère sa mise plus une somme égale à deux fois sa mise. Enfin, si le numéro choisi sort trois fois, le joueur récupère sa mise plus une somme égale à trois fois sa mise. Quelle est l'espérance de gain à ce jeu ?
2. Soit $s > 1$. On dit que X suit une loi ζ de paramètre si l'on a
$$\forall n \in \mathbb{N}^* \quad P(X = n) = \frac{1}{\zeta(s)} \frac{1}{n^s},$$
où l'on a posé
$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}.$$
Soit donc X suivant une loi ζ de paramètre s. On tire Y au hasard – c'est à dire avec équiprobabilité – entre 1 et X.
   (a) Pour $n, k \in \mathbb{N}^*$, calculer $P(Y = k|X = n)$.
   (b) On pose $Z = \frac{Y}{X}$. Montrer que la fonction de répartition F_Z est strictement croissante sur $[0, 1]$.
   (c) Soient p, q deux entiers positifs premiers entre eux, avec $p \leq q$. Calculer $P(Z = \frac{p}{q})$.

(d) On rappelle que $\varphi(n)$ désigne le nombre d'entiers entre 1 et n qui sont premiers avec n. Déduire de ce qui précède une preuve probabiliste de l'identité

$$\zeta(s+1)\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^{s+1}} = \zeta(s).$$

(e) Montrer

$$\Big(\sum_{n=1}^{+\infty}\frac{1}{n^2}\Big)\Big(\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^3}\Big)\Big(\sum_{n=1}^{+\infty}\frac{\varphi(n)}{n^5}\Big) = \int_2^3 x\ dx.$$

(On pourra admettre que $5\zeta(4) = 2\zeta(2)^2$.)

3. Donner un exemple de familles d'événements $\mathcal{C}$ et $\mathcal{D}$ telles que
   – $\forall A \in \mathcal{C} \quad \forall B \in \mathcal{D} \quad P(A \cap B) = P(A)P(B)$.
   – les tribus $\sigma(\mathcal{C})$ et $\sigma(\mathcal{D})$ ne sont pas indépendantes.
4. Donner un exemple de deux lois distinctes sur $(\Omega, \mathcal{F})$ coïncidant sur un système $\mathcal{C}$ engendrant $\mathcal{F}$.
5. On choisit de manière uniforme sur $[0,1]$ un réel Y. Quelle est la probabilité pour que le polynome

$$p(x) = x^2 + x + Y$$

   ait des racines complexes ? des racines distinctes ?
6. Dans le segment $[AB]$ de longueur 1, on choisit au hasard un point M. Quelle est la probabilité pour que l'on ait $AM.MB \geq \frac{2}{9}$?
7. Soient $X_1, \dots, X_n, \dots$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1, \dots, X_n)$. Déterminer la fonction de répartition de M_n. Montrer que M_n admet une densité que l'on déterminera.
8. Soient $X_1, \dots, X_n, \dots$ des variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $M_n = \max(X_1, \dots, X_n)$ et $m_n = \min(X_1, \dots, X_n)$. Montrer que M_n et $1 - m_n$ ont même loi.
9. Soit X une variable aléatoire suivant la loi $\mathcal{B}(2n, 1/2)$. On pose $Y = |X - n|$. Déterminer la loi de Y. À k fixé, il faut déterminer les valeurs de X qui sont telles que $Y = k$.
10. Soit (X,Y) un vecteur aléatoire suivant la loi uniforme sur le rectangle $[-1,2] \times [-1,1]$. Montrer que

$$P(1 - Y \geq 2|X|) = \frac{1}{3}.$$

11. Pour n entier strictement positif, on note $A_n = n\mathbb{N}^*$. Notons $\mathcal{P}$ l'ensemble des nombres premiers positifs et $\mathcal{T}$ la sous-tribu de $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ engendrée par les $(A_p)_{p\in\mathcal{P}}$. Pour $\omega \in \mathbb{N}^*$, on pose

$$F(\omega) = \prod_{p\in\mathcal{P};d \text{ divise } \omega} p.$$

    (a) Montrer que $\mathcal{T} = \sigma(X)$.
    (b) Déterminer le plus petit ensemble $\mathcal{T}$-mesurable contenant 1980.
    (c) Montrer que A_n est $\mathcal{T}$-mesurable si et seulement si n n'est divisible par aucun carré.
    (d) On munit $(\mathbb{N}^*, \mathcal{B}(\mathbb{N}^*))$ de la mesure de probabilité ζ de paramètre s, c'est à dire que

$$\forall n \in \mathbb{N}^* \quad P(\{n\}) = \frac{1}{\zeta(s)}\frac{1}{n^s},$$

    où l'on a posé

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}.$$

    Montrer que $P(A_n) = \frac{1}{n^s}$. À quelle condition les événements A_n et A_m sont-ils indépendants sous la loi P ?
    (e) Soit $\mathcal{N} = \{\omega \in \mathbb{N}^*; \quad A_\omega \text{est } \mathcal{T} - \text{mesurable}\}$ Montrer que $\mathcal{N} = \underset{p\in\mathcal{P}}{\cap} A^c_{p^2}$, puis que $0 < P(\mathcal{N})$.

12. (a) Soit $(\mathcal{A}_n)_{n\geq 1}$ une suite de tribus indépendantes. Montrer que la tribu $\mathcal{Q} = \overset{+\infty}{\underset{k=1}{\cap}} \underset{i\geq k}{\wedge} \mathcal{A}_i$ est triviale.
    (b) Soit $(A_n)_{n\geq 1}$ une suite d'événements indépendants. Soit A l'événement " une infinité de A_i se produisent. Montrer que $P(A)$ ne peut valoir que 0 ou 1

13. Démontrer les propriétés de la fonction Γ laissées en exercice :
    - $\forall a > 0 \quad \Gamma(a+1) = a\Gamma(a)$.
    - $\forall n \in \mathbb{N} \quad \Gamma(n+1) = n!$

14. *Queue de la gaussienne*
    Soit X une variable aléatoire suivant la loi normale $\mathcal{N}(0,1)$. On pose $\psi(t) = P(X > t) = 1 - F_X(t)$. Montrer

$$\Psi(t) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{x}e^{-\frac{x^2}{2}}.$$

Chapitre 3

Espérances et calculs

3.1 Quelques rappels sur la construction de l'espérance

Définition Si X est une variable aléatoire intégrable définie sur $(\Omega, \mathcal{F}, P)$, on appelle espérance de X et on note $\mathbb{E}X$ le réel défini par

$$\mathbb{E}X = \int_\Omega X(\omega)\ dP(\omega).$$

Remarque : En toute rigueur, il faudrait écrire $\mathbb{E}_P X$.

Définition Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes sont intégrables, on note $\mathbb{E}X$ le vecteur $(\mathbb{E}X_1, \dots, \mathbb{E}X_n)$.

3.2 Quelques propriétés

- L^1 est un espace vectoriel.
- $\forall X, Y \in L^1 \quad \mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$.
- $\forall X \in L^1, \forall a \in \mathbb{R} \quad \mathbb{E}aX = a\mathbb{E}X$.
- $\forall A \in \mathcal{F} \quad \mathbb{E}(\mathbb{1}_A) = P(A)$.
- La variable aléatoire X est intégrable si et seulement si $|X|$ est intégrable.
- $\forall X \in L^1 \quad P(X \ge 0) = 1 \Longrightarrow \mathbb{E}X \ge 0$
- $\forall X \in L^1 \quad P(X \le a) = 1 \Longrightarrow \mathbb{E}X \le a$.
- $\forall X \in L^1 \quad P(X \ge b) = 1 \Longrightarrow \mathbb{E}X \ge b$.
- $\forall X \in L^1 \quad P(X = a) = 1 \Longrightarrow \mathbb{E}X = a$.
- $\forall X \in L^1 \quad P(|X| \le a) = 1 \Longrightarrow \mathbb{E}|X| \le a$.
- Soient X, Y deux variables aléatoires vérifiant $0 \le X \le Y$. Si Y est intégrable, alors X est intégrable.

– $\forall X, Y \in L^1 \quad P(X \leq Y) = 1 \Longrightarrow \mathbb{E}X \leq \mathbb{E}Y$.

Vocabulaire : on dit qu'une variable aléatoire X est centrée si $\mathbb{E}X = 0$. On définit de même ce qu'est un vecteur aléatoire centré.

3.3 Application : Formule de Poincaré et inégalités de Bonferroni

La formule de Poincaré est l'analogue de la formule du même nom du cours de dénombrement. On peut considérer que c'en est une généralisation.

Théorème 23 (Formule de Poincaré). *Pour tous événements $A_1, A_2, ..., A_n$ sous la probabilité P*

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B \in \mathcal{P}(\{1,\dots,n\}) \setminus \varnothing} (-1)^{1+|B|} P(\cap_{j \in B} A_j) &(3.1)\\ &= \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i_1 < i_2 \leq n} P(A_{i_1} \cap A_{i_2}) + \cdots &(3.2)\\ &\cdots + (-1)^{k+1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) + \cdots &(3.3)\\ &\cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n). &(3.4) \end{aligned}$$

Exemple : Pour $n = 3$, on a

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_1 \cap A_3) + P(A_1 \cap A_2 \cap A_3$$

Pour prouver la formule de Poincaré, on va utiliser un lemme qui va nous permettre d'obtenir des encadrements de la probabilité d'une réunion.

Lemme 1. *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé ; $(A_x)_{x \in I}$ des événements.*

Pour $n \geq 1$, on pose

$$V_n = P_n\Big(\sum_{x \in I} \mathbb{1}_{A_x} - 1\Big) \mathbb{1}_{\cup_{x \in I} A_x},$$

où $(P_k)_{k \geq 0}$ est la suile de polynômes définie par

$$\begin{cases} P_0 = 1 \\ P_1 = X \\ P_2 = \frac{X(X-1)}{2} \\ \dots \\ P_k = \frac{X(X-1)\dots(X-k+1)}{k!} \end{cases}$$

Ainsi pour $n \geq k$ $\quad P_k(n) = \binom{n}{k}$ tandis que $P_k(n) = 0$ pour $0 \leq n < k$.
Alors

$$\forall n \in \mathbb{N}^* \quad P(\bigcup_{x\in I} A_x) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\bigcap_{x\in J} A_x) + (-1)^n \mathbb{E}V_n \quad (3.5)$$

Démonstration. Il suffit de montrer

$$\forall n \in \mathbb{N}^* \quad \mathbb{1}_{\bigcup_{x\in I} A_x} = \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} \prod_{x\in J} \mathbb{1}_{A_x} + (-1)^n V_n \qquad (3.6)$$

Soit $\omega \in \Omega$. Si $\omega \notin \cup_{x\in I} A_x$, les deux membres de l'égalité sont nuls. Sinon, posons $N = \sum_{x\in I} \mathbb{1}_{A_x}(\omega) = |\{x \in I ; \omega \in A_x\}|$.

On a $\prod_{x\in J} \mathbb{1}_{A_x} = 1$ si et seulement si $J \subset \{x \in I ; \omega \in A_x\}$.

Ainsi $\sum_{J\in\mathcal{B}_k(I)} \prod_{x\in J} \mathbb{1}_{A_x} = P_k(N)$
On doit donc montrer

$$1 = \sum_{k=1}^{n} (-1)^{k+1} P_k(N) + (-1)^n P_n(N-1),$$

ce qui est équivalent à

$$\sum_{k=0}^{n} (-1)^k P_k(N) = (-1)^n P_n(N-1).$$

On va montrer par récurrence sur n :

$$\sum_{k=0}^{n} (-1)^k P_k(X) = (-1)^n P_n(X-1)$$

Pour $n = 0$, c'est vérifié. Pour passer de n à $n+1$, on a

$$\begin{aligned}
\sum_{k=0}^{n+1}(-1)^k P_k(X) &= \sum_{k=0}^{n}(-1)^k P_k(X) + (-1)^{n+1}P_{n+1}(X) \\
&= (-1)^n P_n(X-1) + (-1)^{n+1}P_{n+1}(X) \\
&= (-1)^{n+1}(P_{n+1}(X) - P_n(X-1)) \\
&= (-1)^{n+1}(\frac{1}{n+1}XP_n(X-1) - P_n(X-1)) \\
&= (-1)^{n+1}\frac{(X-(n+1))P_n(X-1)}{n+1} \\
&= (-1)^{n+1}P_{n+1}(X-1)
\end{aligned}$$

□

Ensuite, il suffit d'intégrer (3.6) pour obtenir (3.5).

Démonstration. En prenant $I = J = \{1, \dots, n\}$ dans le lemme précédent, on obtient

$$P(\underset{x\in I}{\cup} A_x) = \sum_{k=1}^{|I|} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\underset{x\in J}{\cap} A_x),$$

car V_n est identiquement nulle. Cela démontre la formule de Poincaré. □

Théorème 24 (Inégalités de Bonferroni). *Soit* $(\Omega, \mathcal{F}, P)$ *un espace probabilisé ;* $(A_x)_{x\in I}$ *des événements. Soit* $n \in \mathbb{N}^*$.

Alors,

– *Si n est impair, on a*

$$P(\underset{x\in I}{\cup} A_x) \leq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\underset{x\in J}{\cap} A_x) \tag{3.7}$$

– *Si n est pair, on a*

$$P(\underset{x\in I}{\cup} A_x) \geq \sum_{k=1}^{n} (-1)^{k+1} \sum_{J\in\mathcal{B}_k(I)} P(\underset{x\in J}{\cap} A_x) \tag{3.8}$$

Démonstration. Il suffit d'appliquer le lemme 1 en remarquant que V_n est une variable aléatoire positive, et que donc son espérance l'est aussi. □

3.4 Intégrale et queue de distribution

Théorème 25. *Soit X une variable aléatoire positive. On a*

$$\mathbb{E}X = \int_{\mathbb{R}_+} P(X > t)\ d\lambda(t).$$

Démonstration.

$$\begin{aligned}
\int_{\mathbb{R}_+} P(X > t)\ d\lambda(t) &= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} 1\!\!1_{]t,+\infty[}(s)\ dP_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} 1\!\!1_{[0,s]}(t)\ dP_X(s)\ d\lambda(t) \\
&= \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} 1\!\!1_{[0,s]}(t)\ d\lambda(t)\ dP_X(s) \\
&= \int_{\mathbb{R}_+} s\ dP_X(s) \\
&= \mathbb{E}X
\end{aligned}$$

□

Corollaire 4. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. On a*

$$\mathbb{E}X = \sum_{k=0}^{+\infty} P(X > k).$$

Démonstration. D'après le théorème précédent, $\mathbb{E}X = \int_{\mathbb{R}_+} P(X > t)\ d\lambda(t)$. Comme $t \mapsto P(X > t)$ est une fonction positive, on a

$$\int_{\mathbb{R}_+} P(X > t)\ d\lambda(t) = \sum_{k=0}^{+\infty} \int_{[k,k+1[} P(X > t)\ d\lambda(t).$$

Mais comme X est à valeurs entières, on a

$$\forall t \in [k, k+1[\quad P(X > t) = P(X > k),$$

d'où le résultat. □

3.5 Théorèmes de transfert

Théorème 26. *Soit X est une variable aléatoire dont la loi P_X admet une densité f par rapport à la mesure m. Soit g une fonction mesurable.*

Alors, $g(X)$ est intégrable si et seulement si

$$\int_{\mathbb{R}} |g(x)|d(x)\underset{\cdot}{m}(x) < +\infty.$$

Si cette intégrale est finie, on a alors

$$\mathbb{E}g(X) = \int_{\mathbb{R}} |g(x)|f(x)\underset{\cdot}{m}(x) < +\infty.$$

Démonstration. D'après le théorème de transfert

$$\begin{aligned}\int_{\Omega} |g(X(\omega))|dP(\omega) &= \int_{\mathbb{R}} |g(x)|dP_X(x) \\ &= \int_{\mathbb{R}} |g(x)|f(x)\ dm(x)\end{aligned}$$

De même, si cette quantité est finie, le théorème de transfert nous dit encore que

$$\begin{aligned}\int_{\Omega} g(X(\omega))dP(\omega) &= \int_{\mathbb{R}} g(x)dP_X(x) \\ &= \int_{\mathbb{R}} g(x)f(x)\ dm(x)\end{aligned}$$

□

3.5.1 Calcul de l'espérance d'une variable aléatoire discrète

Théorème 27. *Soit X est une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Soit g une fonction quelconque de D dans $\mathbb{R}$. Alors, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si*

$$\sum_{i \in D} |g(i)|p_i < +\infty,$$

où l'on a posé $p_i = P(X = i)$. Si cette somme est finie, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \sum_{i \in D} g(i)p_i.$$

Démonstration. D'après nos hypothèses, P_X admet une densité par rapport à la mesure de comptage de support D : c'est la fonction $i \mapsto p_i$. Il suffit donc d'appliquer le théorème 26 en prenant pour m la mesure de comptage sur D et pour f :

$$f(x) = \begin{cases} p_x & \text{si } x \in D \\ 0 \text{ si } x \notin D \end{cases}$$

On a alors

$$\int_{\mathbb{R}} |g(x)|f(x)\ dm(x) = \sum_{x \in D} |g(x)|f(x),$$

et, si cette somme est finie :

$$\int_{\mathbb{R}} g(x)f(x)\ dm(x) = \sum_{x \in D} g(x)f(x).$$

□

Corollaire 5. *Soit X est une variable aléatoire discrète, et D un ensemble fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$. Alors, X est intégrable si et seulement si*

$$\sum_{i \in D} |i|p_i < +\infty,$$

où l'on a posé $p_i = P(X = i)$. Si cette somme est finie, on a alors

$$\mathbb{E}X = \sum_{i \in D} ip_i.$$

Démonstration. Il suffit d'appliquer le thèorème précédent avec $g(x) = x$. □

3.5.2 Calcul de l'espérance d'une variable aléatoire à densité

Voici maintenant le théorème de transfert pour les variables à densité

Théorème 28. *Soit X est une variable aléatoire admettant la fonction f comme densité, et g une fonction continue par morceaux définie sur $X(\Omega)$. Alors, la variable aléatoire $Y = g(X)$ est intégrable si et seulement si*

$$\int_{\mathbb{R}} |g(x)|f(x)\ dx < +\infty.$$

Si cette intégrale est convergente, on a alors

$$\mathbb{E}Y = \mathbb{E}g(X) = \int_{\mathbb{R}} g(x)f(x)\ dx.$$

Démonstration. Il suffit d'appliquer le théorème 26 avec pour m la mesure de Lebesgue. □

Corollaire 6. *Soit X est une variable aléatoire admettant la fonction f comme densité. Alors, X est intégrable si et seulement si*

$$\int_{\mathbb{R}} |x|f(x)\ dx < +\infty.$$

Si cette intégrale est convergente, on a alors

$$\mathbb{E}X = \int_{\mathbb{R}} xf(x)\ dx.$$

Démonstration. Il suffit d'appliquer le thèorème précédent avec $g(x) = x$. □

Remarque importante : Si la densité de X est paire et que X est intégrable, alors $\mathbb{E}X = 0$.

Démonstration. On a

$$\begin{aligned}
\mathbb{E}X &= \int_{-\infty}^{0} xf(x)\ dx + \int_{0}^{+\infty} xf(x)\ dx \\
&= \int_{0}^{+\infty} -xf(-x)\ dx + \int_{0}^{+\infty} xf(x)\ dx \\
&= \int_{0}^{+\infty} x(f(x) - f(-x))\ dx \\
&= 0
\end{aligned}$$

car $f(x) = f(-x)$. □

3.6 Moments d'ordre 2

On dit qu'une variable aléatoire X admet un moment d'ordre 2 si elle est de carré intégrable, c'est à dire si $X^2 \in L^1$.

On note $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ (ou encore L^2) l'ensemble des variables aléatoires de carré intégrable.

Lemme 2. *Soient $X, Y \in L^2$. Alors la variable aléatoire XY est intégrable.*

Démonstration. Pour tous les réels a, b, on a $|ab| \le \frac{1}{2}(a^2 + b^2)$.
(En effet, $a^2+b^2+2ab = (a+b)^2 \ge 0$, d'où $(a^2+b^2)/2 \ge -ab$ et $a^2+b^2-2ab = (a-b)^2 \ge 0$, d'où $(a^2+b^2)/2 \ge ab$.)
On a donc

$$0 \le |XY| \le \frac{1}{2}(X^2 + Y^2),$$

Comme $X^2 + Y^2$ est intégrable, on en déduit que $|XY|$ est intégrable, ce qui est ce que l'on voulait montrer □

Corollaire 7. $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ *est un espace vectoriel.*

Démonstration. La stabilité par multiplication ne pose pas de problème. Pour la stabilité par addition, il faut remarquer que

$$(X + Y)^2 = X^2 + Y^2 + 2XY,$$

puis utiliser le lemme précédent est le fait que L^1 est un espace vectoriel. □

3.6.1 Covariance et variance

Soient X et Y deux variables aléatoires admettant chacune un moment d'ordre 2. On appelle covariance du couple (X, Y) le nombre

$$\mathrm{Covar}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$$

On appelle variance de X le nombre

$$\mathrm{Var}\, X = \mathrm{Covar}(X, X) = \mathbb{E}(X - \mathbb{E}X)^2.$$

On appelle écart-type de X le nombre

$$\sigma(X) = (\mathrm{Var}\, X)^{1/2}.$$

Vocabulaire : On dit qu'une variable aléatoire est réduite si on a $\mathrm{Var}\, X = 1$ (ou de manière équivalente si $\sigma(X) = 1$).

On a les propriétés suivantes

1. $(X, Y) \mapsto \mathrm{Covar}(X, Y)$ est une forme bilinéaire symétrique positive.
2. $\forall a, b \in \mathbb{R} \quad \mathrm{Covar}(X - a, Y - b) = \mathrm{Covar}(X, Y)$.
3. $\mathrm{Var}(X + Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\,\mathrm{Covar}(X, Y)$.
4. $\mathrm{Covar}(X, Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
5. $\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

6. $|\mathbb{E}XY|^2 \leq \mathbb{E}X^2\mathbb{E}Y^2$ (inégalité de Cauchy-Schwarz)
7. $|\operatorname{Covar}(X,Y)| \leq \sigma(X)\sigma(Y)$.

Démonstration. 1. Notons $\langle X, Y\rangle = \mathbb{E}X, Y$. Il est facile de voir que $(X,Y) \mapsto \langle X, Y\rangle$ est une forme bilinéaire symétrique positive. Posons $L(X) = X - \mathbb{E}X$. $X \mapsto L(X)$ est une application linéaire de L^2 dans lui-même. On a $\operatorname{Covar}(X,Y) = \langle L(X), L(Y)\rangle$. Les deux observations faites ci-dessus permettent de dire que $(X,Y) \mapsto \operatorname{Covar}(X,Y)$ est une forme bilinéaire symétrique positive.

2.

$$\begin{aligned}
\operatorname{Covar}(X-a, Y-b) &= \langle L(X-a), L(Y-b)\langle \\
&= \langle L(X) - L(a), L(Y) - L(b)\langle \\
&= \langle L(X) - 0, L(Y) - 0\langle \\
&= \operatorname{Covar}(X-a, Y-b)
\end{aligned}$$

3.

$$\begin{aligned}
\operatorname{Var}(X+Y) &= \operatorname{Covar}(X+Y, X+Y) \\
&= \operatorname{Covar}(X,X) + 2\operatorname{Covar}(X,Y) + \operatorname{Covar}(Y,Y) \\
&= \operatorname{Var} X + 2\operatorname{Covar}(X,Y) + \operatorname{Var} Y
\end{aligned}$$

Pour passer de la ligne 1 à la ligne 2, on utilise le fait que la covariance est bilinéaire symétrique.

4. $(X - \mathbb{E}X)(Y - \mathbb{E}Y) = XY + \mathbb{E}X\mathbb{E}Y - (\mathbb{E}X)Y - (\mathbb{E}Y)X$.
D'où

$$\begin{aligned}
\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) &= \mathbb{E}XY + \mathbb{E}(EX\mathbb{E}Y) - \mathbb{E}X\mathbb{E}Y - \mathbb{E}Y\mathbb{E}X \\
&= \mathbb{E}XY + \mathbb{E}X\mathbb{E}Y - 2\mathbb{E}X\mathbb{E}Y \\
&= \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y).
\end{aligned}$$

5. Il suffit d'appliquer la formule précédente avec $X = Y$.
6. Comme $\langle ., .\rangle$ est une forme bilinéaire symétrique positive, l'inégalité de Cauchy-Schwarz s'applique.
7. Il suffit d'appliquer l'inégalité de Cauchy-Schwarz à la forme bilinéaire symétrique positive Covar , puis de prendre la racine carrée.

$\square$

Lorsque $\sigma(X)$ et $\sigma(Y)$ sont non nuls, on définit le coefficient de corrélation de X et Y par

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Covar}(X,Y)}{\sigma(X)\sigma(Y)}.$$

D'après ce qui précède, $\mathrm{Covar}(X,Y) \in [-1;1]$. Lorsque $\mathrm{Covar}(X,Y) = 0$ (ce qui implique $\mathrm{Corr}(X,Y) = 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X ne sont pas corrélées.
Lorsque $\mathrm{Covar}(X,Y) \geq 0$ (ce qui implique $\mathrm{Corr}(X,Y) \geq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont positivement corrélées.
Lorsque $\mathrm{Covar}(X,Y) \leq 0$ (ce qui implique $\mathrm{Corr}(X,Y) \leq 0$ si $\sigma(X)$ et $\sigma(Y)$ sont non nuls), on dit que X et Y sont négativement corrélées.

3.6.2 Matrice de covariance

Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire dont toutes les composantes admettent un moment d'ordre deux, on convient de dire que le vecteur a un moment d'ordre deux et on appelle matrice de covariance de X la matrice $n \times n$ dont les coefficients sont $(\mathrm{Covar}(X_i, X_j))_{1 \leq i,j \leq n}$.

Théorème 29. *Si $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux , la matrice de covariance de X est la matrice dans la base canonique de l'application bilinéaire positive*

$$\begin{array}{ccc} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ (a,b) & \mapsto & \mathrm{Covar}(\langle X, a\rangle, \langle X, b\rangle) \end{array}$$

C'est une matrice symétrique positive.

Démonstration. À X fixé, l'application $X \mapsto \langle X, a\rangle$ est une application linéaire. Comme on a déjà montré que Covar était une forme bilinéaire symétrique positive, il s'ensuit que l'application considérée ici est une forme bilinéaire symétrique positive. Cette application envoie le couple (e_i, e_j) sur $\mathrm{Covar}(\langle X, e_i\rangle, \langle X, e_j\rangle) = \mathrm{Covar}(X_i, X_j)$. La matrice d'une forme bilinéaire symétrique positive est une matrice symétrique positive. $\square$

Théorème 30. *Soit $X = (X_1, \dots, X_n)$ est un vecteur aléatoire admettant un moment d'ordre deux et de matrice de covariance C_X et d'espérance m_X. Soit A une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^p$, et b un vecteur de $\mathbb{R}^p$. Alors $Y = AX + b$ admet $C_Y = AC_XA^*$ comme matrice de covariance et l'espérance de Y vaut $Am_X + b$.*

Démonstration.

$$\begin{aligned}
\mathbb{E}Y_i &= \mathbb{E}(\sum_{k=1}^{n} a_{i,k}X_k + b_i) \\
&= \sum_{k=1}^{n} a_{i,k}\mathbb{E}X_k + b_i \\
&= \sum_{k=1}^{n} a_{i,k}m_k + b_i \\
&= (Am+b)_i
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\,\mathrm{Covar}(\langle Y,a\rangle,\langle Y,b\rangle) &= \mathbb{E}\,\mathrm{Covar}(\langle AX+c,a\rangle,\langle AX+c,b\rangle) \\
&= \mathbb{E}\,\mathrm{Covar}(\langle AX,a\rangle,\langle AX,b\rangle) \\
&= \mathbb{E}\,\mathrm{Covar}(\langle X,A^*a\rangle,\langle X,A^*b\rangle) \\
&= \langle C_X A^*a, A^*b\rangle \\
&= \langle AC_X A^*a, b\rangle
\end{aligned}$$

□

3.6.3 Espérance et indépendance

Le théorème suivant est très important :

Théorème 31. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors, leur produit XY est une variable aléatoire intégrable et l'on a*

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y.$$

Démonstration. D'après le théorème de transfert, on a

$$\mathbb{E}|XY| = \int_{\mathbb{R}^2} |xy|\; dP_{(X,Y)}$$

Il vient

$$\begin{aligned}
\mathbb{E}|XY| &= \int_{\mathbb{R}^2} |xy|\; dP_{(X,Y)} \\
&= \int_{\mathbb{R}^2} |x|.|y|\; d(P_X\otimes P_Y)(x,y) \\
&= \int_{\mathbb{R}} |x| dP_X(x). \int_{\mathbb{R}} |y|\; dP_Y(y) \\
&= \mathbb{E}|X|.\mathbb{E}|Y| \\
&< +\infty
\end{aligned}$$

Ainsi, le théorème de Fubini nous a permis de montrer que XY était intégrable. Maintenant, et on a

$$\begin{aligned}
\mathbb{E}XY &= \int_{\mathbb{R}^2} xy \; dP_{(X,Y)} \\
&= \int_{\mathbb{R}^2} xy \; d(P_X \otimes P_Y)(x,y) \\
&= \int_{\mathbb{R}} x dP_X(x). \int_{\mathbb{R}} y \; dP_Y(y) \\
&= \mathbb{E}X\mathbb{E}Y
\end{aligned}$$

□

Corollaire 8. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors X et Y ne sont pas corrélées.*

Démonstration. On a $\mathrm{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$. □

Remarque importante : Des variables aléatoires peuvent être non corrélées sans être indépendantes.
Exemple : soient deux variables aléatoires vérifiant

$P(\{X=1\} \cap \{Y=1\}) = P(\{X=1\} \cap \{Y=-1\}) = P(\{X=-1\} \cap \{Y=0\}) = 1/3.$

La matrice M associée à la loi du couple est

$$\begin{pmatrix} 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 \end{pmatrix}$$

La loi de Y s'obtient en faisant la somme des lignes : on obtient

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \end{pmatrix}$$

On a donc $\mathbb{E}Y = \frac{1}{3} \times (-1) + \frac{1}{3} \times (0) + \frac{1}{3} \times (1) = 0$.
D'autre part $\mathbb{E}XY = \sum_{i \in \{-1;1\}} \sum_{j \in \{-1;0;1\}} ijP(\{X=i\} \cap \{Y=j\}) = 1/3 - 1/3 = 0$.
On a donc $\mathrm{Covar}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0$.
Cependant

$$0 = P(\{X=1\} \cap \{Y=0\}) \neq P(X=1)P(Y=0) = \frac{2}{3} \times \frac{1}{3}.$$

Corollaire 9. *Soient X, Y deux variables aléatoires indépendantes de carré intégrable. Alors on a*

$$\mathrm{Var}(X+Y) = \mathrm{Var}\, X + \mathrm{Var}\, Y.$$

Démonstration. On a toujours $\mathrm{Var}\, X+Y = \mathrm{Var}\, X + \mathrm{Var}\, Y + 2\,\mathrm{Covar}(X,Y)$.Comme X et Y sont indépendantes, elles ne sont pas corrélées, d'où le résultat. □

3.7 Calculs de lois images

On rappelle un théorème de théorie de la mesure de Lebesgue très important :

Proposition 2. *Soit O_1 et O_2 deux ouverts de $\mathbb{R}^d, d \geq 1$. On suppose que T est un C^1-difféomorphisme de O_1 dans O_2. Soit maintenant μ_1 une mesure positive sur $\mathbb{R}^d$ telle que $\mu_1(\mathbb{R}^d \backslash O_1) = 0$ et admettant une densité f_1 par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, la mesure image de μ_1 par T admet comme densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$ la fonction f_2 définie par*

$$f_2(y) = \begin{cases} f_1(T^{-1}(y))|\det DT_y^{-1}| \ si\ y \in O_2 \\ 0 \ si\ y \notin O_2 \end{cases}$$

3.7.1 Exemple fondamental

Théorème 32. *Soit $A \in Gl_d(\mathbb{R})$ et $b \in \mathbb{R}^d$. On suppose que le vecteur aléatoire X admet la densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Alors, le vecteur aléatoire $Y = AX + b$ admet la densité*

$$g(y) = \frac{1}{|\det A|} f(A^{-1}(y-b)).$$

Démonstration. Ici $O_1 = O_2 = \mathbb{R}^d$ et $T^{-1}(y) = A^{-1}(y - b)$. La différentielle en un point d'une transformation affine se confond avec l'application linéaire associée, d'où le résultat. □

Applications :

1. Si X suit la loi uniforme sur un compact K de $\mathbb{R}^d$, alors $Y = AX + b$ suit la loi uniforme sur l'image de K par $x \mapsto Ax+b$. Cette application est particulièrement intéressante en dimension 1.
2. Si $X \sim \Gamma(a, \lambda)$, alors pour tout $\mu > 0$, on a $\frac{1}{\mu}X \sim \Gamma(a, \lambda\mu)$.
3. Si X suit la loi exponentielle de paramètre e et que $a > 0$ alors $\frac{1}{\mu}X$ suit la loi exponentielle de paramètre μa. (Remarquer que ceci constitue un cas particulier de la remarque prédente.)
4. Pour $a \in \mathbb{R}$ et $b > 0$, X suit la loi de Cauchy $\mathcal{C}(0, 1)$ si et seulement si $Y = bX + a$ suit la loi de Cauchy $\mathcal{C}(a, b)$.
5. Soit $\sigma > 0$, $m \in \mathbb{R}$. On a $X \sim \mathcal{N}(m, \sigma^2) \iff \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$.

3.7.2 Application aux lois gaussiennes

Lemme 3. *Soient $X = (X_1, X_2)$ un vecteur aléatoire formé de deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0,1)$. On pose $Y_1 = \cos\theta X_1 + \sin\theta X_2$ et $Y_2 = -\sin\theta X_1 + \cos\theta X_2$. Alors Y_1 et Y_2 sont deux variables aléatoires indépendantes suivant la loi normale $\mathcal{N}(0,1)$.*

Démonstration. Si l'on note, pour $x \in \mathbb{R}^2$, $x = (x_1, x_2)$, la densité de X est

$$\frac{1}{\sqrt{2\pi}}\exp(-\frac{x_1^2}{2}) \times \frac{1}{\sqrt{2\pi}}\exp(-\frac{x_2^2}{2}) = \frac{1}{2\pi}\exp(-\frac{x_1^2+x_2^2}{2}) = \frac{1}{2\pi}\exp(-\frac{\|x\|_2^2}{2}).$$

On a donc $Y = MX$, avec

$$M = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Ainsi, le vecteur $Y = MX$ admet pour densité

$$y \mapsto \frac{1}{\det M}\frac{1}{2\pi}\exp(-\frac{\|M^{-1}y\|_2^2}{2}).$$

M est une matrice de rotation, donc son déterminant vaut 1 et c'est une isométrie pour la norme euclidienne, ce qui implique que pour tout $x \in \mathbb{R}^2$ on a $\|M^{-1}y\|_2 = \|y\|_2$: la densité de Y est donc

$$y \mapsto \frac{1}{2\pi}\exp(-\frac{\|y\|_2^2}{2}),$$

ce qui est précisément la densité de X : Y a donc même loi que X, donc ses composantes Y_1 et Y_2 sont indépendantes et suivent la loi normale $\mathcal{N}(0,1)$. □

Théorème 33. *Soient U_1 et U_2 deux variables aléatoires indépendantes, avec $U_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $U_2 \sim \mathcal{N}(m_2, \sigma_2^2)$. Alors $U_1 + U_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Démonstration. Si $\sigma_1 = 0$ ou $\sigma_2 = 0$, la variable aléatoire associée est constante est donc le résultat provient de la remarque faite plus haut – l'application affine est une translation.

Supposons donc $\sigma_1 > 0$ et $\sigma_2 > 0$. On pose $X_1 = \frac{U_1 - m_1}{\sigma_1}$ et $X_2 = \frac{U_2 - m_2}{\sigma_2}$. On peut trouver θ tel que $\cos\theta = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$ et $\sin\theta = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}}$. Alors, si on pose $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$, on a

$$U_1 + U_2 = m_1 + m_2 + \sigma(\cos\theta X_1 + \sin\theta X_2).$$

D'après le lemme, $\cos\theta X_1 + \sin\theta X_2 \sim \mathcal{N}(0,1)$, donc $U_1 + U_2 \sim \mathcal{N}(m_1 + m_2, \sigma^2)$. □

3.7.3 Application : convolution de deux lois à densité

Théorème 34. *Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, P)$, de densité f et g. Alors $Z = X + Y$ admet comme densité la fonction*

$$x \mapsto \int_{-\infty}^{+\infty} f(x-t)g(t)\ dt.$$

Si, de plus, X et Y sont à valeurs positives, alors la densité est simplement

$$x \mapsto 1\!\!1_{\mathbb{R}_+}(x) \int_0^x f(x-t)g(t)\ dt.$$

Démonstration. On pose

$$\begin{pmatrix} Z \\ T \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix}, \text{ avec } A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

La densité de (X, Y) est $h(x, y) = f(x)g(y)$. D'après l'exemple fondamental, la densité de (Z, T) est

$$g(z,t) = \frac{1}{|\det A|} f(A^{-1} \begin{pmatrix} z \\ t \end{pmatrix}).$$

On a

$$\det A = 1 \text{ et } A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

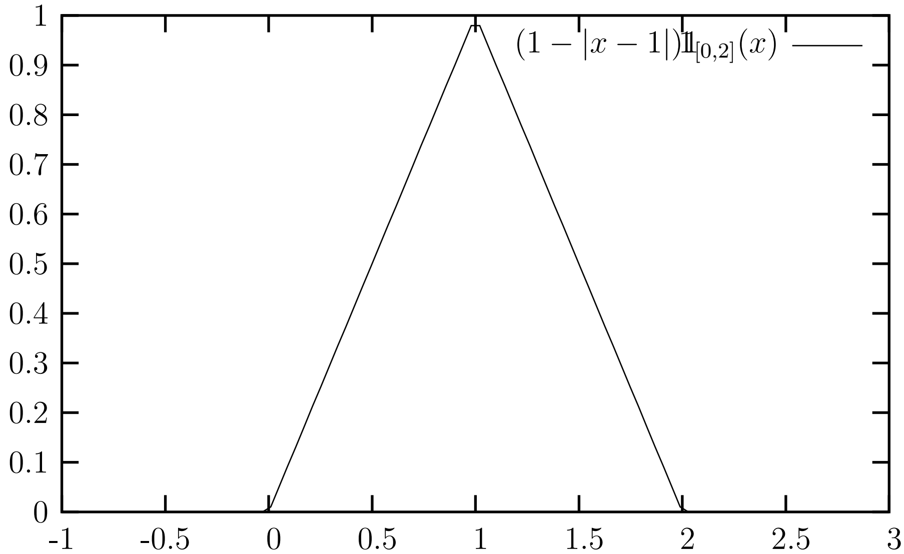
donc $g(z, t) = h(z - t, t) = f(z - t)g(t)$.

D'après le théorème 17, Z admet comme densité par rapport à la mesure de Lebesgue :

$$z \mapsto \int_{\mathbb{R}} g(z,t)\ d\lambda(t) = \int_{\mathbb{R}} f(z-t)g(t)\ d\lambda(t).$$

Dans le cas où X et Y sont à valeurs positives, il suffit de remarquer que $f(z - t)$ est nul si z dépasse t et que $g(t)$ est nul si t est négatif. Ainsi, $f(z - t)g(t)$ ne peut être non nul que pour z vérifiant $0 \le t \le z$, ce qui n'est évidemment jamais vérifié si z est négatif. □

Exemple : ci-dessous, le graphe de la densité de Z lorsque X et Y suivent toutes les deux la loi uniforme sur $[0, 1]$.



Application : $\Gamma(a,\lambda)*\Gamma(b,\lambda)=\Gamma(a+b,\lambda)$

Théorème 35. *Soit a,b,λ strictement positifs, X et Y deux variables aléatoires indépendantes, X suivant la loi $\Gamma(a,\lambda)$ et Y la loi $\Gamma(b,\lambda)$. Alors $Z=X+Y$ suit la loi $\Gamma(a+b,\lambda)$.*

Démonstration. Pour tous a et λ strictement positifs, on note $f_{a,\lambda}$ la densité de la loi $\Gamma(a,\Lambda)$, soit (rappel)

$$f_{a,\lambda}(x)=1\!\!1_{\mathbb{R}_+}(x)\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}.$$

D'après le théorème précédent, Z admet une densité f_Z. Cette densité est nulle sur $\mathbb{R}_-$ tandis que pour x positif, on a

$$\begin{aligned}
f_Z(x) &= \int_0^x f_{a,\lambda}(x-t)f_{b,\lambda}(t)\ dt\\
&= \int_0^x \frac{\lambda^a}{\Gamma(a)}t^{a-1}e^{-\lambda t}\frac{\lambda^b}{\Gamma(b)}(x-t)^{b-1}e^{-\lambda(x-t)}\ dt\\
&= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)}\int_0^x t^{a-1}(x-t)^{b-1}\ dt
\end{aligned}$$

On fait le changement de variable $t = \theta x$. On obtient

$$\begin{aligned} f_Z(x) &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} x^{b-1} x \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta \\ &= \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a+b-1} \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta \\ &= K_{a,b} f_{a+b,\lambda}(x), \end{aligned}$$

où $K_{a,b} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta$. Evidemment f_Z et $x \mapsto K_{a,b} f_{a+b,\lambda}(x)$ coincident également sur $\mathbb{R}_-$ où elles sont nulles. On a donc

$$\int_{\mathbb{R}} f_Z(x)\, d\lambda(x) = \int_{\mathbb{R}} K_{a,b} f_{a+b,\lambda}(x)\, d\lambda(x) = K_{a,b} \int_{\mathbb{R}} f_{a+b,\lambda}(x).$$

Mais f_Z et $f_{a+b,\lambda}$ sont des densités donc leur intégrale sur $\mathbb{R}$ vaut un. On en déduit $K_{a,b} = 1$, d'où

$$f_Z(x) = f_{a+b,\lambda}(x).$$

La densité de Z est la densité de $\Gamma(a+b,\Lambda)$, donc Z suit la loi $\Gamma(a+b,\Lambda)$.

Remarque : comme sous-produit de cette démonstration, on a obtenu le résultat non trivial suivant :

$$\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 \theta^{a-1}(1-\theta)^{b-1}\, d\theta.$$

□

3.8 Calcul des premiers moments des lois discrètes usuelles

3.8.1 Indicatrice d'un événement

On rappelle que pour $A \subset \Omega$, l'application $1\!\!1_A$ (appelée indicatrice de A) est définie sur Ω par

$$1\!\!1_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

1_A est une variable aléatoire à valeurs dans $\{0;1\}$. Il est important de remarquer que, comme $\forall x \in \{0;1\} \quad x^2 = x$, on a $1\!\!1_A^2 = 1\!\!1_A$. Maintenant, on a

– $\mathbb{E} 1\!\!1_A = P(A)$.

–

$$\begin{aligned}\mathrm{Var}\,\mathbb{1}_A &= \mathbb{E}\mathbb{1}_A^2 - (\mathbb{E}\mathbb{1}_A)^2\\ &= \mathbb{E}\mathbb{1}_A - (\mathbb{E}\mathbb{1}_A)^2\\ &= P(A) - P(A)^2\\ &= P(A)(1-P(A)).\end{aligned}$$

3.8.2 Loi binomiale

On a vu que la loi binomiale de paramètres n et p était la loi de

$$X = \sum_{k=1}^{n} \mathbb{1}_{A_k},$$

où $A_1, \ldots, A_n$ sont n événements indépendants de même probabilité p. On a donc

$$\mathbb{E}X = \sum_{k=1}^{n} \mathbb{E}\mathbb{1}_{A_k} = \sum_{k=1}^{n} P(A_k) = np,$$

et comme les variables aléatoires sont indépendantes

$$\mathrm{Var}\,X = \sum_{k=1}^{n} \mathrm{Var}\,\mathbb{1}_{A_k} = \sum_{k=1}^{n} P(A_k)(1-P(A_k)) = np(1-p).$$

3.8.3 Loi géométrique

Soit X une variable aléatoire suivant une loi géométrique de paramètre $p \in]0,1]$. On a

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{+\infty} kP(X=k) \\
&= \sum_{k=1}^{+\infty} kP(X=k) \\
&= \sum_{k=1}^{+\infty} kp(1-p)^{k-1} \\
&= p\sum_{k=1}^{+\infty} k(1-p)^{k-1} \\
&= p\frac{1}{(1-(1-p))^2} \\
&= \frac{1}{p}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}X(X-1) &= \sum_{k=0}^{+\infty} k(k-1)P(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)P(X=k) \\
&= \sum_{k=2}^{+\infty} k(k-1)p(1-p)^{k-1} \\
&= p(1-p)\sum_{k=2}^{+\infty} k(k-1)(1-p)^{k-2} \\
&= p(1-p)\frac{2}{(1-(1-p))^3} \\
&= \frac{2(1-p)}{2p^2}.
\end{aligned}$$

On a alors $\mathbb{E}X^2 = \mathbb{E}X(X-1) + \mathbb{E}X = \frac{1}{p} + \frac{2(1-p)}{p^2}$ et
$\mathrm{Var}\, X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

3.8.4 Loi de Poisson

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

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

On a alors $\mathbb{E}X^2 = \mathbb{E}X(X-1) + \mathbb{E}X = \lambda^2 + \lambda$ et
$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

3.8.5 Loi hypergéométrique

On rappelle que la loi hypergéométrique $\mathcal{H}(N,n,k)$ est la loi image de la loi uniforme sur $\Omega = \mathcal{B}(N,k)$ par l'application

$$\begin{array}{rcl} X : \mathcal{B}(N,k) & \to & \mathbb{N} \\ \omega & \mapsto & X(\omega) = |\{1,\dots,n\}) \cap \omega| \end{array}$$

On va montrer que $\mathbb{E}X = k\frac{n}{N}$ et $\operatorname{Var} X = k\frac{n}{N}(1-\frac{n}{N})\frac{N-1}{N}$.

Démonstration. Notons P la loi uniforme sur Ω. Par souci de lisibilité, on définit l'ensemble aléatoire A par $A(\omega) = \omega$. Ainsi

$$\begin{array}{rcl} X & = & |\{1,\dots,n\}) \cap A| \\ & = & \sum_{i=1}^{n} \mathbb{1}_{\{i \in A\}} \end{array}$$

Pour tout $k \in \{1,\dots,n\}$, on a

$$\mathbb{E}\mathbb{1}_{\{i\in A\}} = P(i \in A) = 1 - \frac{\binom{N-1}{k}}{\binom{N}{k}} = 1 - \frac{(N-1)!(N-k)!}{N!(N-k-1)!} = 1 - \frac{N-k}{N} = \frac{k}{N}.$$

Ainsi

$$\mathbb{E}X = \sum_{i=1}^{n} \mathbb{1}_{\{i\in A\}} = \frac{nk}{N} = k\frac{n}{N}.$$

Maintenant, on a

$$\operatorname{Var} X = \sum_{i=1}^{n} \sum_{j=1}^{n} \operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}).$$

Pour $i = j$, on a

$$\operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}) = \operatorname{Var}\mathbb{1}_{\{i\in A\}} = P(j \in A)(1 - P(j\in A)) = \frac{k}{N}(1-\frac{k}{N}).$$

Pour $i \neq j$, on a

$$\begin{array}{rcl} \operatorname{Covar}(\mathbb{1}_{\{i\in A\}}, \mathbb{1}_{\{j\in A\}}) & = & \operatorname{Covar}(1-\mathbb{1}_{\{i\in A\}}, 1-\mathbb{1}_{\{j\in A\}}) \\ & = & \operatorname{Covar}(\mathbb{1}_{\{i\notin A\}}, \mathbb{1}_{\{j\notin A\}}) \\ & = & P(i\notin A, j \notin A) - P(i\notin A)P(j\notin A) \\ & = & \frac{\binom{N-2}{k}}{\binom{N}{k}} - (\frac{N-k}{N})^2 \\ & = & \frac{(N-k)(N-k-1)}{N(N-1)} - (\frac{N-k}{N})^2 \\ & = & -\frac{1}{N-1}\frac{k}{N}(1-\frac{k}{N}) \end{array}$$

On en déduit

$$\begin{aligned}
\operatorname{Var} X &= n \times \frac{k}{N}(1-\frac{k}{N}) + n(n-1)\times(-\frac{1}{N-1}\frac{k}{N}(1-\frac{k}{N})) \\
&= n\frac{k}{N}(1-\frac{k}{N})(1-\frac{n-1}{N-1}) \\
&= n\frac{k}{N}(1-\frac{k}{N})\frac{N-n}{N-1} \\
&= k\frac{n}{N}(1-\frac{n}{N})\frac{N-1}{N}
\end{aligned}$$

On remarque qu'une loi hypergéométrique a la même espérance qu'une loi binomiale $\mathcal{B}(k, \frac{n}{N})$ et que sa variance ne diffère de celle de cette binomiale que d'un facteur $\frac{N-1}{N}$.

□

3.9 Calcul des premiers moments des lois à densité usuelles

3.9.1 Loi uniforme sur un segment

Soit X une variable aléatoire suivant la loi uniforme sur $[-1, 1]$. La densité de X est donc

$$x \mapsto \frac{1}{2}\mathbb{1}_{[-1,1]}(x).$$

On a donc

$$\mathbb{E}X = \int_{-1}^{1} \frac{1}{2}x\ dx = 0$$

et

$$\mathbb{E}X^2 = \int_{-1}^{1} \frac{1}{2}x^2\ dx = \frac{1}{3}.$$

Comme X est centrée, on a $\operatorname{Var} X = \mathbb{E}X^2$.

Passons au cas général : on pose $Y = \frac{a+b}{2} + \frac{b-a}{2}X$. X suit la loi uniforme sur $[-1, 1]$ si et seulement Y suit la loi uniforme sur $[a, b]$. On a alors

- $\mathbb{E}Y = \mathbb{E}\frac{a+b}{2} + \frac{b-a}{2}\mathbb{E}X = \frac{a+b}{2}$.
- $\operatorname{Var} Y = (\frac{b-a}{2})^2 \operatorname{Var} X = \frac{(b-a)^2}{12}$

3.9.2 Loi gaussienne

Soit X une variable aléatoire suivant la loi $\mathcal{N}(0,1)$. On rappelle que la densité de X est

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

Il est facile de vérifier que

$$\frac{\mathrm{d}}{\mathrm{dx}}(xf(x)) = (1-x^2)f(x).$$

On a donc

$$\forall a,b \in \mathbb{R} \quad bf(b) - af(a) = \int_a^b f(x)\ dx - \int_a^b x^2 f(x)\ dx$$

Comme $\lim\limits_{a\to-\infty} af(a) = \lim\limits_{b\to+\infty} bf(b) = 0$ et que $\int_{-\infty}^{+\infty} f(x)\ dx = 1$, on en déduit que

$$\int_{-\infty}^{+\infty} x^2 f(x)\ dx = 1.$$

Autrement dit, X admet un moment d'ordre 2 : $\mathbb{E}X^2 = 1$.

D'autre part, l'existence d'un moment d'ordre 2 implique celle d'un moment d'ordre 1. Comme la densité de X est paire, on en déduit que $\mathbb{E}X = 0$. On a donc $\operatorname{Var} X = \mathbb{E}X^2 = 1$.

Passons au cas général. Si l'on a $Y = m + \sigma X$, on sait que Y suit la loi $\mathcal{N}(m, \sigma^2)$. On a alors $\mathbb{E}Y = m + \sigma\mathbb{E}X = m$ et $\operatorname{Var} Y = \sigma^2 \operatorname{Var} X = \sigma^2$.

3.9.3 Lois Gamma

Soit X une variable aléatoire suivant la loi $\Gamma(a,\lambda)$. Alors, X admet des moments de tout ordre, avec pour tout $\alpha \ge 0$, on a

$$\mathbb{E}X^\alpha$$

. En particulier $\mathbb{E}X = \frac{a}{\lambda}$ et $\operatorname{Var} X = \frac{a}{\lambda^2}$.

Démonstration. Pour tous a et λ strictement positifs, on note $f_{a,\lambda}$ la densité de la loi $\Gamma(a,\Lambda)$, soit (rappel)

$$f_{a,\lambda}(x) = \mathbb{1}_{\mathbb{R}_+}(x) \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}.$$

D'après le théorème de transfert,

$$\begin{aligned}\mathbb{E}X^\alpha &= \int_{\mathbb{R}_+} x^\alpha f_{a,\lambda}(x)\ dx \\ &= \int_{\mathbb{R}_+} \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)} f_{a+\alpha,\lambda}(x)\ dx \\ &= \lambda^{-\alpha}\frac{\Gamma(a+\alpha)}{\Gamma(a)}\end{aligned}$$

Ainsi $\mathbb{E}X = \lambda^{-1}\frac{\Gamma(a+1)}{\Gamma(a)} = \frac{a}{\lambda}$, $\mathbb{E}X^2 = \lambda^{-2}\frac{\Gamma(a+2)}{\Gamma(a)} = \frac{a(a+1)}{\lambda^2}$ $\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{a(a+1)}{\lambda^2} - \frac{a^2}{\lambda^2} = \frac{a}{\lambda^2}$. □

3.9.4 Lois exponentielles

Soit X suivant une loi exponentielle de paramètre $\lambda > 0$. La loi exponentielle est un cas particulier de la loi Gamma : on a $\mathcal{E}(\lambda) = \Gamma(1, \lambda)$. On déduit du calcul précédent que $\mathbb{E}X = \frac{1}{\lambda}$ et $\operatorname{Var} X = \frac{1}{\lambda^2}$.

3.9.5 Lois de Cauchy

Soient $a \in \mathbb{R}, b > 0$. La loi de Cauchy $\mathcal{C}(a, b)$ admet comme densité par rapport à la mesure de Lebesgue :

$$x \mapsto \frac{1}{\pi}\frac{b}{(x-a)^2 + b^2},$$

donc pour $k \geq 1$, on a

$$\mathbb{E}|X|^k = \frac{1}{\pi}\frac{b|x|^k}{(x-a)^2 + b^2} = +\infty,$$

donc les lois de Cauchy n'admettent pas de moment d'ordre 1, ni , a fortiori, d'ordre supérieur.

3.10 Exercice sur les espérances

1. Soient A, B deux éléments observables. On note

$$A\Delta B = \{x \in A; x \notin B\} \cup \{x \in B; x \notin A\}.$$

Ce sont donc les éléments qui sont dans A ou dans B, mais pas dans les deux. Montrer $\mathbb{1}_{A\Delta B} = (\mathbb{1}_A - \mathbb{1}_B)^2$. En déduire

$$P(A\Delta B) = P(A) + P(B) - 2P(A \cap B).$$

2. Soient A, B deux éléments observables. Montrer que

$$|P(A \cap B) - P(A)P(B)| \leq \sqrt{P(A)P(B)}.$$

3. Soit X une variable aléatoire à valeurs dans $\mathbb{N}^*$ telle la suite $(p_n)_{n \geq 1}$ définie par $p_n = P(X = n)$ soit décroissante. Montrer que pour toute injection σ de $\mathbb{N}^*$ dans lui-même, on a

$$\mathbb{E}\sigma(X) \geq \mathbb{E}X.$$

4. On suppose que $Y = \ln X$ vérifie $X \sim \mathcal{N}(m, \sigma^2)$ (on dit alors que Y est log-normale). Calculer $\mathbb{E}X$ et $\operatorname{Var} X$.
5. Calculer $\mathbb{E} \sin X$, où $P(X = \frac{\pi}{2}) = \frac{1}{6}, P(X = \frac{\pi}{3}) = \frac{1}{3}, P(X = \frac{\pi}{6}) = \frac{1}{2}$.
6. Soient X, Y deux variables aléatoires suivant chacune une loi uniforme sur $[a, b]$. Montrer que $\mathbb{E}|X - Y| \leq \frac{b-a}{2}$. Que vaut $\mathbb{E}|X - Y|$ lorsque X et Y sont indépendantes ?
7. Soit X une variable aléatoire suivant la loi uniforme sur $[0, 1]$. Déterminer la loi de $Z = -\ln(1 - X)$.
8. Soit X une variable aléatoire de densité f. Montrer que la variable aléatoire $|X|$ admet comme densité

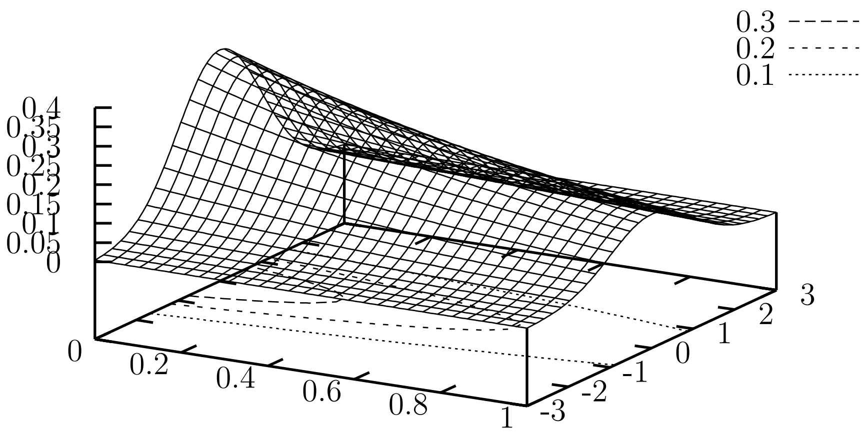
$$x \mapsto \mathbb{1}_{R_+}(x)(f(x) + f(-x)).$$

9. Soit X une variable aléatoire positive de densité f. Montrer que la variable aléatoire $X^{1/2}$ admet comme densité

$$x \mapsto \mathbb{1}_{R_+}(x)2x(f(x^2)).$$

10. Soit X une variable aléatoire normale centrée réduite. Montrer que la variable aléatoire X^2 est à densité et la déterminer.
11. La figure ci-dessous représente la densité $f(x, y)$ d'un couple de variables aléatoires indépendantes X et Y. X suit une loi exponentielle de paramètre 1 et Y une loi normale centrée réduite.

    On a tracé quelques isoclines, c'est à dire des courbes reliant des points de même densité : $f(x, y) = $ constante. Quelle est la nature géométrique de ces isoclines ?



12. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On pose $S = X + Y$ et $P = XY$.

    Déterminer la loi de (S,P).

13. Soit f une fonction réelle continue sur l'intervalle fermé $[0,1]$. Pour $n \in \mathbb{N}^*$, on note B_n le polynôme de Bernstein

$$B_n(x) = \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k} \ .$$

    Pour tout $x \in]0,1[$ on se donne une suite (X_k) de variables de Bernoulli indépendantes de même paramètre x. On note $S_n = \sum_{k=0}^{n} X_k$.

    (a) Déterminer la moyenne $\mathbb{E}[f(\frac{S_n}{n})]$.

    (b) Soit pour tout $\varepsilon > 0$, le réel $\delta(\varepsilon)$ défini par

$$\delta(\varepsilon) = \sup\{|f(x) - f(y)| : x, y \in [0,1] \text{ et } |x - y| \leq \varepsilon\} \ .$$

       i. Démontrer que $\delta(\varepsilon)$ tend vers 0 avec ε.

       ii. Démontrer que

$$\sup_{x\in[0,1]} |B_n(x) - f(x)| \leq \delta(\varepsilon) + \frac{2\|f\|_\infty}{n\varepsilon^2} \ .$$

       En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0,1]$.

14. On place 7 dames sur un échiquier torique 41×41 de telle manière qu'aucune dame ne puisse en prendre une autre. Montrer qu'il est possible de placer sur l'échiquier deux cavaliers pouvant se prendre mutuellement tels que chacun des cavaliers puisse être pris par au moins une des dames.
15. Soient n, r deux entiers tels que $1 \le r \le n$. On prend r nombres au hasard dans $\{1, \dots, n\}$ et on note X le plus petit de ces r nombres.
    (a) Quelles valeurs peut prendre X ? Montrer que pour $k \in \{0, n-r\}$, on a
    $$P(X > k) = \frac{\binom{n-k}{r}}{\binom{n}{r}}.$$
    (b) En déduire que
    $$\mathbb{E}X = \frac{\binom{n+1}{r+1}}{\binom{n}{r}} = \frac{n+1}{r+1}.$$
16. Soient X et Y deux variables aléatoires indépendantes suivant la loi uniforme sur $[0, 1]$.
    (a) Soit r la rotation dans $\mathbb{R}^2$ de centre $(0, 0)$ et d'angle $-\frac{\pi}{4}$. On pose $(U, V) = r(X, Y)$. Montrer que la loi du vecteur (U, V) est la loi uniforme sur un ensemble que l'on déterminera.
    (b) Pour quelles valeurs de α la variable aléatoire $\frac{1}{|X-Y|^\alpha}$ est-elle intégrable ? Lorsqu'elle l'est, calculer sa valeur.

Chapitre 4

Fonctions génératrices et fonctions caractéristiques

4.1 Fonctions génératrices des variables aléatoires à valeurs dans $\mathbb{N}$

Définition : On appelle fonction génératrice d'une variable aléatoire X à valeurs dans $\mathbb{N}$ la fonction

$$z \mapsto G_X(z) = \mathbb{E}z^X = \sum_{k=0}^{+\infty} P(X=k)z^k.$$

Usuellement, on définit cette fonction sur l'intervalle réel $[0,1]$, mais elle est en fait toujours définie sur la boule unité complexe fermée.

4.1.1 Fonction génératrice et indépendance

Théorème 36. *Si deux variables aléatoires X et Y sont indépendantes, on a $G_{X+Y} = G_X G_Y$.*

Démonstration. Soit $z \in B(0,1)$. On a

$$G_{X+Y}(z) = \mathbb{E}z^{X+Y} = \mathbb{E}z^X z^Y = \mathbb{E}z^X \mathbb{E}z^Y = G_X(z)G_Y(z).$$

□

4.1.2 Calculs de fonctions génératrices

Loi de Bernoulli

La fonction génératrice d'une loi de Bernoulli de paramètre p est $z \mapsto (1-p)+pz$, car si X suit une telle loi, on a

$$\mathbb{E}z^X = P(X=0)z^0 + P(X=1)z^1 = (1-p)+pz.$$

Loi binomiale

Si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes suivant la loi de Bernoulli de paramètre p, alors $S_n = X_1 + \dots + X_n$ suit la loi loi binomiale de paramètres n et p. Ainsi, on déduit du théorème 36 que la fonction génératrice d'une loi binomiale de paramètres n et p est

$$\varphi_{S_n}(z) = \varphi_{X_1} \times \dots \varphi_{X_n}(z) = ((1-p)+pz)^n.$$

Loi géométrique de paramètre $p \in]0,1[$

Soit $X\mathcal{G}(p)$ et $z \in B(0,1)$ On a

$$\begin{aligned}
\varphi_X(z) &= \mathbb{E}z^X \\
&= \sum_{n=1}^{+\infty} p(1-p)^{n-1}z^n \\
&= \sum_{n=0}^{+\infty} p(1-p)^n z^{n+1} \\
&= pz \sum_{n=0}^{+\infty} ((1-p)z)^n \\
&= \frac{pz}{1-(1-p)z}
\end{aligned}$$

Loi de Poisson

Soit X une variables aléatoire suivant une loi de Poisson de paramètre λ. On a

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

4.1.3 Fonction génératrice et loi

Théorème 37. *Soit X une variable aléatoire de loi ν sur $\mathbb{N}$. Sur $[0,1[$, la fonction $x \mapsto G_X(x)$ est infiniment dérivable et ces dérivées sont toutes strictement positives, avec*

$$G_X^{(n)}(s) = \mathbb{E}X(X-1)\dots X(X-n+1)s^{X-n}$$

En particulier

$$P(X=n) = \frac{G_X^{(n)}(0)}{n!},$$

ce qui montre que la fonction génératrice caractérise la loi.

Démonstration. La fonction $z \mapsto G_X(z)$ est la somme d'une série entière de rayon de convergence au moins égal à 1. Ainsi $z \mapsto G_X(z)$ est holomorphe sur le disque unité ouverte et y est infiniment dérivable, avec pour tout z dans le disque ouvert unité :

$$G_X^{(n)}(z) = \sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1)P(X=k)z^k$$

Il suffit maintenant d'appliquer le théorème de transfert pour constater que le membre de droite est l'espérance de $X(X-1)\dots X(X-n+1)z^{X-n}$.

En prenant $z=0$, on obtient

$$\begin{aligned} G_X^{(n)}(0) &= \mathbb{E}X(X-1)\dots(X-n+1)\mathbb{1}_{\{X-n=0\}} \\ &= \mathbb{E}n(n-1)\dots(n-n+1)\mathbb{1}_{\{X-n=0\}} \\ &= n!P(X=n) \end{aligned}$$

La restriction à un intervalle de $\mathbb{R}$ d'une fonction holomorphe est évidemment une fonction infiniment dérivable et la notion de dérivée coincide. Lorsque $s \in [0,1[$, on a pour tout $\omega \in \Omega$:

$$X(\omega)(X(\omega)-1)\dots X(\omega)(X(\omega)-n+1)s^{X(\omega)-n} \geq 0.$$

Comme l'espérance d'une variable aléatoire positive est positive, le résultat s'ensuit.

□

4.1.4 Application : convolution des lois de Poisson

Théorème 38. *Soit X et Y deux variables aléatoires indépendantes, X suivant une loi de Poisson de paramètre λ et Y une loi de Poisson de paramètre μ. Alors $X+Y$ suit une loi de Poisson de paramètre $\lambda+\mu$.*

Démonstration. On a vu qu'une v.a.r. suivant une loi de Poisson de paramètre λ a comme fonction génératrice $s \mapsto e^{-\lambda(1-s)}$, ceci quelque soit $\lambda > 0$. En particulier, il s'ensuit que $G_X(s) = e^{-\lambda(1-s)}$ et $G_Y(s) = e^{-\mu(1-s)}$ Maintenant $G_{X+Y}(s) = G_X(s)G_Y(s) = e^{-\lambda(1-s)}e^{-\mu(1-s)} = e^{-(\lambda+\mu)(1-s)}$. Ainsi $X+Y$ a la même fonction génératrice qu'une loi de Poisson de paramètre $\lambda+\mu$. Mais d'après le théorème 37, la fonction génératrice détermine la loi, donc $X+Y$ suit la loi de Poisson de paramètre $\lambda+\mu$. □

4.1.5 Fonction génératrice et espérance

Théorème 39. *Soit X une variable aléatoire à valeurs dans $\mathbb{N}$.*
Alors $\mathbb{E}X < +\infty$ si et seulement si G_X admet une dérivée à gauche en 1. Dans ce cas $G'_X(1) = \mathbb{E}X$.

Démonstration. On note ν la loi de X. Pour $x \in [0,1[$,

$$\frac{G_X(1)-G_X(x)}{1-x} = \sum_{n=0}^{+\infty} \frac{1-x^n}{1-x}\nu(n)$$

Pour tout n, on a $\frac{1-x^n}{1-x} = 1 + x + \dots + x^{n-1}$: c'est donc une fonction croissante de x. De plus $\lim_{x\to 1^-} \frac{1-x^n}{1-x} = n$. D'après le théorème de convergence monotone (on intègre sur $\mathbb{N}$ par rapport à la mesure de comptage), on a donc on a

$$\lim_{x\to+\infty} \frac{G_X(1)-G_X(x)}{1-x} = \sum_{n=0}^{+\infty} n\nu(n) = \int x \, d\nu(x) = \mathbb{E}X.$$

□

4.2 Fonctions caractéristiques

4.2.1 Motivations

La fonction caractéristique est un outil analogue à la fonction génératrice, qui permet de généraliser aux variables aléatoires à valeurs réelles et même aux vecteurs aléatoires les techniques des fonctions génératrices.

Définition: On appelle fonction caractéristique d'une mesure de probabilité sur $\mathbb{R}^d$ la fonction complexe définie en tout point de $\mathbb{R}^d$ par

$$\forall t \in \mathbb{R}^d \quad \varphi_\mu(t_1, \ldots, t_n) = \varphi_{P_X}(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\, d\mu(x) = \int_{\mathbb{R}^d} \exp(i(t_1x_1+t_2x_2+\cdots+t_nx_n))\, d\mu(x)$$

Par extension, on appelle fonction génératrice d'un vecteur aléatoire X et on note φ_X la fonction génératrice de sa loi. Ainsi

$$\forall t \in \mathbb{R}^d \quad \varphi_X(t) = \int_{\mathbb{R}^d} \exp(i\langle t, x\rangle)\, dP_X(x) = \mathbb{E}e^{i\langle X,t\rangle}.$$

On va admettre ici un important résultat d'analyse, qui justifie la dénomination de fonction caractéristique et rend cette outil pertinent pour nos besoins.

Proposition 3. *Soit μ et ν deux mesures de probabilité sur $\mathbb{R}^d$. On a*

$$\forall t \in \mathbb{R}^d \quad \varphi_\mu(t) = \varphi_\nu(t) \iff \mu = \nu.$$

Donnons une conséquence frappante de cette proposition qui nous sera utile dans l'étude des vecteurs gaussiens.

Théorème 40. *Soit X et Y deux vecteurs aléatoires sur $\mathbb{R}^d$ tels que pour tout $a \in \mathbb{R}^d$, $\langle X, a\rangle$ et $\langle Y, a\rangle$ ont même loi. Alors X et Y ont même loi.*

Démonstration. On va montrer que que X et Y ont même fonction caractéristique, ce qui assurera qu'ils ont même loi. Soit $a \in \mathbb{R}^d$ quelconque. On pose $Z = \langle X, a\rangle$ et $T = \langle Y, a\rangle$. Z et T ont même loi, donc $\mathbb{E}e^{iZ} = \mathbb{E}e^{iT}$. Mais $\mathbb{E}e^{iZ} = \mathbb{E}e^{i\langle X,a\rangle} = \varphi_X(a)$ et $\mathbb{E}e^{iT} = \mathbb{E}e^{i\langle Y,a\rangle} = \varphi_Y(a)$, donc $\varphi_X(a) = \varphi_Y(a)$. Ainsi $\varphi_X = \varphi_Y$, donc X et Y ont même loi. □

Le théorème très simple ci-après est d'usage courant.

Théorème 41. *Soit X une vecteur aléatoire de $\mathbb{R}^d$, A une application linéaire de $\mathbb{R}^d$ dans $\mathbb{R}^n$ et $b \in \mathbb{R}^n$. On pose $Y = AX + b$. Alors*

$$\forall t \in \mathbb{R}^n \quad \varphi_Y(t) = e^{i\langle b,t\rangle}\varphi_X(A^*t).$$

Démonstration.

$$\varphi_Y(t) = \mathbb{E}e^{i\langle Y,t\rangle} = \mathbb{E}e^{i\langle AX+b,t\rangle} = \mathbb{E}e^{i\langle AX,t\rangle}e^{i\langle b,t\rangle}\mathbb{E}e^{i\langle X,A^*t\rangle}e^{i\langle b,t\rangle} = e^{i\langle b,t\rangle}\varphi_X(A^*t)$$

□

4.2.2 Propriétés des fonctions caractéristiques

Théorème 42. *Soit μ une mesure de probabilité sur $\mathbb{R}^d$. On a les propriétés suivantes :*
- $\varphi_\mu(0) = 1$
- $|\varphi_\mu| \leq 1$.
- *φ est uniformément continue sur $\mathbb{R}^d$.*
- *Si $e_1, \ldots, e_p$ sont des éléments de $\mathbb{R}^n$, la matrice A de taille $p \times p$ définie par $a_{k,l} = \varphi_\mu(e_k - e_l)$ est hermitienne positive.*

Démonstration. – $\varphi_\mu(0) = \int_{\mathbb{R}^d} e^{i\langle x,0\rangle} d\mu(x) = \int_{\mathbb{R}^d} 1 \; d\mu = 1$.
- $|\varphi_\mu(t)| = |\int_{\mathbb{R}^d} e^{i\langle x,t\rangle} d\mu(x)| \leq \int_{\mathbb{R}^d} |e^{i\langle x,t\rangle}| d\mu(x) = \int_{\mathbb{R}^d} 1 \; d\mu = 1$.
- Soient $t, t' \in \mathbb{R}^d$

$$\begin{aligned} |\varphi_\mu(t) - \varphi_\mu(t')| &= |\int_{\mathbb{R}^d} e^{i\langle t,x\rangle} - e^{i\langle t',x\rangle} \; d\mu(x)| \\ &= |\int_{\mathbb{R}^d} e^{i\langle t,x\rangle}(1 - e^{i\langle (t'-t),x\rangle}) \; d\mu(x)| \\ &\leq \int_{\mathbb{R}^d} |1 - e^{i\langle (t'-t),x\rangle}| \; d\mu(x) \end{aligned}$$

  Il suffit maintenant de voir que la fonction $u \mapsto \int_{\mathbb{R}^d} |1 - e^{i\langle u,x\rangle}| \; d\mu(x)$ admet une limite nulle en 0 pour conclure, or ce dernier point est assuré par le théorème de convergence dominée de Lebesgue.
- On remarque que

$$\begin{aligned} \int_E |\sum_{k=1}^p x_k \exp(i\langle e_k, x\rangle)|^2 \; d\mu(x) &= \int_E \sum_{k=1}^p \sum_{l=1}^p x_k x_l \exp(i\langle e_k - e_l, x\rangle) \; d\mu(x) \\ &= \sum_{k=1}^p \sum_{l=1}^p x_k x_l \int_E \exp(i\langle e_k - e_l, x\rangle) \; d\mu(x) \\ &= \sum_{k=1}^p \sum_{l=1}^p x_k x_l \varphi_\mu(e_k - e_l) \end{aligned}$$

□

Exemple : pour $n = 1$, si f est une fonction positive 2π périodique intégrable sur $[-\pi, \pi[$, on peut prendre pour μ la mesure

$$d\mu(x) = \frac{f(x)\mathbb{1}_{[-\pi,\pi[}(x)}{2\pi} \; d\lambda(x)$$

$\varphi_{\mu(k)}$ est alors le $-k$ ième coefficient de Fourier de F. On prend alors classiquement $e_k = k$.

Ainsi, si f est non nulle, la matrice A est en fait définie positive car

$$\begin{aligned}\sum_{k=1}^{p}\sum_{l=1}^{p} x_k x_l \varphi_\mu(k-l) = 0 &\implies \int_E |\sum_{k=1}^{p} x_k \exp(ikx)|^2\, d\mu(x) = 0\\ &\implies |\sum_{k=1}^{p} x_k \exp(ikx)|^2 f(x) = 0 \text{ p.p. sur}[-\pi,\pi],\end{aligned}$$

Comme un polynôme trigonométrique non nul n'a qu'un nombre fini de zéro (ça se prolonge en un polynôme sur $\mathbb{C}$), on en déduit que f est presque partout nul, ce qu'on avait exclu.

En fait, un difficile théorème nous dit que les propriétés énoncées ci-dessus sont largement suffisantes pour permettre d'affirmer qu'une fonction donnée est une fonction caractéristique.

Il s'agit du théorème de Bochner, que nous admettrons ici.

Proposition 4 (Bochner). *Soit $\varphi : \mathbb{R}^n \to \mathbb{C}$ une fonction continue en 0, vérifiant $\varphi(0) = 1$ et de type positif, c'est à dire que si $e_1, \dots, e_p$ sont des éléments quelconques de $\mathbb{R}^n$, la matrice A de taille $p \times p$ définie par $a_{k,l} = \varphi(e_k - e_l)$ est hermitienne positive.*

Alors il existe une mesure de probabilité μ sur $\mathbb{R}^n$ telle que $\varphi = \varphi_\mu$.

4.2.3 Fonction caractéristique et indépendance

Théorème 43. *Soient X et Y deux vecteurs aléatoires indépendants à valeurs dans $\mathbb{R}^d$. Alors*

$$\forall t \in \mathbb{R}^d \quad \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

Démonstration.

$$\varphi_{X+Y}(t) = \mathbb{E}e^{i\langle t, X+Y\rangle} = \mathbb{E}e^{i\langle t,X\rangle}e^{i\langle t,Y\rangle} = \mathbb{E}e^{i\langle t,X\rangle}\mathbb{E}e^{i\langle t,Y\rangle} = \varphi_X(t)\varphi_Y(t).$$

□

Théorème 44. *Soient X et Y deux vecteurs aléatoires indépendants, X étant à valeurs dans $\mathbb{R}^n$ et Y à valeurs dans $\mathbb{R}^p$. Alors le vecteur (X, Y) de dimension $n + p$ admet comme fonction caractéristique la fonction*

$$\forall (s,t) \quad \varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t).$$

Démonstration.

$$\varphi_{(X,Y)}(s,t) = \mathbb{E}e^{i(\langle s,X\rangle + \langle t,Y\rangle)} = \mathbb{E}e^{i\langle s,X\rangle}e^{i\langle t,Y\rangle} = \mathbb{E}e^{i\langle s,X\rangle}\mathbb{E}e^{i\langle t,Y\rangle} = \varphi_X(s)\varphi_Y(t).$$

□

Corollaire 10. *Si μ et ν sont des mesures de probabilité respectivement définies sur $\mathbb{R}^n$ et $\mathbb{R}^p$, alors*

$$\forall (s,t) \quad \varphi_{\mu\otimes\nu}(s,t) = \varphi_\mu(s)\varphi_\nu(t).$$

Démonstration. Il suffit de considérer un couple (X, Y) de variables aléatoires de loi $\mu \otimes \nu$. X et Y sont indépendantes, donc

$$\varphi_{\mu\otimes\nu}(s,t) = \varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t) = \varphi_\mu(s)\varphi_\nu(t)$$

□

Le théorème 44 admet une réciproque

Théorème 45. *Soient X et Y deux vecteurs aléatoires, X étant à valeurs dans $\mathbb{R}^n$ et Y à valeurs dans $\mathbb{R}^p$. Si*

$$\forall (s,t) \quad \varphi_{(X,Y)}(s,t) = \varphi_X(s)\varphi_Y(t),$$

alors X et Y sont indépendants.

Démonstration. Ainsi $\varphi_{P_{(X,Y)}}(s,t) = \varphi_{P_X}(s)\varphi_{P_Y}(t)$. Mais d'après le corollaire précédent, $\varphi_{P_X}(s)\varphi_{P_Y}(t) = \varphi_{P_X\otimes P_Y}(s,t)$. Ainsi $\varphi_{P_{(X,Y)}}(s,t) = \varphi_{P_X\otimes P_Y}(s,t)$. Comme la fonction caractéristique caractérise la loi, on a $P_{(X,Y)} = P_X \otimes P_Y$, ce qui signifie que X et Y sont indépendants. □

En revanche, le théorème 43 n'admet pas de réciproque : on verra en exercice des variables aléatoires X et Y non indépendantes telles que $\varphi_{X+Y} = \varphi_X\varphi_Y$.

4.2.4 Fonction caractéristique et moments

Théorème 46. *Si X est une variable aléatoire sur $\mathbb{R}^d$ telle que $\|X\|$ admette un moment d'ordre N, alors la fonction caractéristique de X est de classe C^N.*

Si $k_1, \dots, k_d$ sont des entiers naturels dont la somme $k = k_1 + \dots + k_d$ ne dépasse pas N, on a

$$\forall u \in \mathbb{R}^d \quad \frac{\partial^k \varphi_X}{\partial_1^{k_1} \dots \partial_d^{k_d}}(u) = i^k \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} \prod_{j=1}^d x_j^{k_j} \, d\mu(x) \tag{4.1}$$

Ainsi

$$\frac{\partial^k \varphi_X}{\partial_1^{k_1} \dots \partial_d^{k_d}}(0) = i^k \mathbb{E} \prod_{j=1}^d X_j^{k_j}. \tag{4.2}$$

Démonstration. La première formule se prouve par récurrence sur k en utilisant le théorème de domination de Lebesgue pour la dérivation sous le signe somme. Pour la deuxième formule, il suffit de prendre $u = 0$. □

Dans le cas des variables aléatoires réelles, la forme du théorème est évidemment plus simple :

Corollaire 11. *Si X est une variable aléatoire réelle telle admettant un moment d'ordre N, alors la fonction caractéristique de X est de classe C^N et on a*

$$\forall k \in \{1, \dots, N\} \quad \varphi_X^{(k)}(0) = i^k \mathbb{E}X^k \tag{4.3}$$

En particulier, si X admet un moment d'ordre 2 et est centrée avec une variance σ^2 , on a le développement limité en 0 :

$$\varphi_X(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2)$$

Démonstration. La reformulation ne pose pas de mystère. Si X admet un moment d'ordre 2, φ_X est de classe C^2, donc

$$\varphi_X(t) = \varphi_X(0) + \varphi_X'(0)t + \frac{\varphi_X''(0)}{2}t^2 + o(t^2)$$

Maintenant $\varphi_X(0) = 1$, $\varphi_X'(0) = i\mathbb{E}X = 0$, $\varphi_X''(0) = -\mathbb{E}X^2 = -\operatorname{Var} X$ car $\mathbb{E}X = 0$. Il suffit de substituer pour conclure. □

4.2.5 Fonctions caractéristiques des variables aléatoires à valeurs dans $\mathbb{N}$

Pour une variable aléatoire à valeur dans $\mathbb{N}$, le calcul de la fonction caractéristique est équivalent à celui de la fonction génératrice.

En effet, on a la formule :

$$\forall t \in \mathbb{R} \varphi_X(t) = \mathbb{E}e^{itX} = \mathbb{E}(e^{it})^X = G_X(e^{it}).$$

On laisse ainsi au lecteur le soin de calculer la fonction caractéristique de la loi de Bernoulli, de la loi binomiale, de la loi de Poisson, de la loi géométrique.

4.2.6 Quelques fonctions caractéristiques de mesures à densité

Pour des mesures à densités, le calcul de la fonction caractéristique est en fait (au signe près) le calcul de la transformée de Fourier de la densité. Lorsque l'on sort des cas simples où une primitive peut facilement être trouvée, ces intégrales peuvent souvent être calculées en utilisant des techniques issues de la théorie de la variable complexe, par exemple en utilisant une méthode des résidus ou encore en appliquant un théorème de prolongement analytique. C'est ce que nous allons voir dans les exemples qui vont suivre.

Il y a bien sûr d'autres méthodes pouvant être utiles : reconnaitre la transformée de Fourier d'une fonction connue et utiliser un théorème d'inversion ou expliciter une équation différentielle vérifiée par la fonction caractéristique, puis la résoudre. Nous ne traiterons pas cela ici et vous renvoyons à un cours d'analyse.

Loi uniforme sur $[a, b]$

On commence par calculer la fonction caractéristique de la loi uniforme sur $[-1, 1]$: on a

Pour $t \neq 0$, on a

$$\varphi_X(t) = \mathbb{E}e^{itx} = \int_{-1}^{1} \frac{1}{2} e^{itx}\ dx = [\frac{e^{itx}}{2it}]_{-1}^{1} = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}$$

La formule se prolonge par continuité pour $t = 0$.

Maintenant, si on pose $Y = \frac{a+b}{2} + (b-a)X$, Y suit la loi uniforme sur $[-a, a]$ et on a

$$\varphi_Y(t) = e^{i\frac{a+b}{2}} \varphi_X((b-a)t) = e^{i\frac{a+b}{2}} \frac{\sin(b-a)t}{(b-a)t}.$$

Loi exponentielle de paramètre λ

On commence par calculer la fonction caractéristique de la loi uniforme sur 1 : on a

Pour $t \neq 0$, on a

$$\varphi_X(t) = \mathbb{E}e^{itx} = \int_{0}^{+\infty} e^{-x} e^{itx}\ dx = [\frac{e^{(-1+it)x}}{-1+it}]_{0}^{+\infty} = \frac{0-1}{-1+it} = \frac{1}{1-it}$$

La formule se prolonge par continuité pour $t = 0$.

Maintenant, si on pose $Y = \frac{1}{\lambda}X$, Y suit la loi $\mathcal{E}(\lambda)$ et on a

$$\varphi_Y(t) = \varphi_X(\frac{1}{\lambda}t) = \frac{1}{1 - i\frac{t}{\lambda}} = \frac{\lambda}{\lambda - it}.$$

Variables aléatoires gaussiennes

Théorème 47. *La fonction caractéristique de la loi normale $\mathcal{N}(m, \sigma^2)$ est*

$$t \mapsto \exp(imt)\exp(-\frac{1}{2}\sigma^2 t^2).$$

Démonstration. On va d'abord calculer la fonction caratéristique de $X \sim \mathcal{N}(0, 1)$. Nous devons calculer

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2)\exp(itx)\ dx.$$

Pour $z \in \mathbb{C}$, on pose

$$f_z(x) = \exp(-x^2/2)\exp(xz)$$

Pour tout $x \in \mathbb{R}$, la fonction $z \mapsto f_z(x)$ est holomorphe sur $\mathbb{C}$. D'autre part, pour $z \in D(0, R)$ la fonction $x \mapsto f_z(x)$ est dominée par la fonction intégrable $x \mapsto \exp(-x^2/2)\exp(Rx)$, car

$$\begin{array}{rcl} \forall x \in \mathbb{R} \quad \forall z \in D(0, R) & & \\ |\exp(-x^2/2)\exp(xz)| & = & \exp(-x^2/2)\exp(x\mathrm{Re}\ z) \\ & \leq & \exp(-x^2/2)\exp(Rx) \end{array}$$

Il s'ensuit que $z \mapsto \int_{\mathbb{R}} f_z(x)d\lambda(x)$ est holomorphe.

Pour z réel, on a

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2)\exp(zx)\ d\lambda(x) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-((x-z)^2 - z^2)/2)\ d\lambda(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-(x-z)^2/2)\ d\lambda(x)\exp(\frac{z^2}{2}) \\ &= \exp(\frac{z^2}{2}), \end{aligned}$$

car l'expression intégrée n'est autre que la densité de la loi $\mathcal{N}(z, 1)$. Mais si deux fonctions holomorphes coincident sur $\mathbb{R}$, elles coïncident sur $\mathbb{C}$. On a donc

$$\forall z \in \mathbb{C} \quad \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2)\exp(zx)\ d\lambda(x) = \exp(\frac{z^2}{2}).$$

On particularise alors z en it, t étant réel, et on obtient

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-x^2/2) \exp(itx)\ dx = \exp(-\frac{t^2}{2}).$$

Pour passer au cas général, on pose $Y = \sigma X + m$; on a $Y \sim \mathcal{N}(m, \sigma^2)$, et alors $\varphi_Y(t) = \mathbb{E}e^{itY} = e^{it(\sigma X+m)} = e^{im}\mathbb{E}e^{it\sigma X} = e^{im}\varphi_X(\sigma t) = e^{im}e^{-\frac{\sigma^2 t^2}{2}}$.

□

Lois de Cauchy

Théorème 48. *La fonction caractéristique de la loi de Cauchy $\mathcal{C}(a, b)$ est*

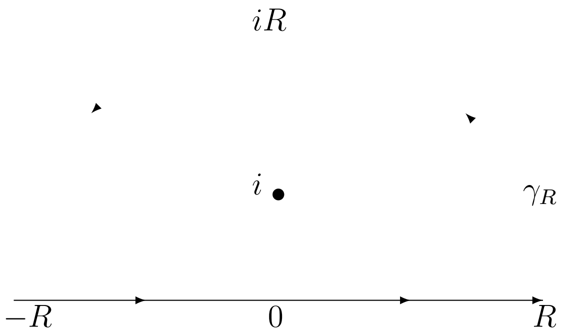
$$\varphi(t) = e^{iat}e^{-b|t|}.$$

Rappel : la loi de Cauchy $\mathcal{C}(a, b)$ admet comme densité par rapport à la mesure de Lebesgue :

$$x \mapsto \frac{1}{\pi} \frac{b}{(x-a)^2 + b^2}.$$

Démonstration. On va d'abord calculer la fonction caractéristique de $\mathcal{C}(0, 1)$ en un point $t > 0$

Pour $R > 1$, on intègre la forme différentielle $f(z) = \frac{e^{itz}}{1+z^2}$ sur le contour γ_R.



Le seul pôle de f à l'intérieur de la courbe est en i, donc pour $R > 1$, on a

$$\frac{1}{2i\pi} \int_{\gamma_R} f(z)\ dz = \mathbf{Res}_i f(z)$$

Or $\frac{1}{1+(i+h)^2} = \frac{1}{h}\frac{1}{2i+h}$, donc $\mathrm{Res}_i f(z) = \frac{e^{-t}}{2i}$.
Ainsi

$$\forall R > 1 \quad \frac{1}{\pi} \int_{\gamma_R} f(z)\ dz = e^{-t}.$$

Or
$$\int_{\gamma_R} f(z)\ dz = \int_{-R}^{R} f(t)\ dt + iR\int_0^\pi f(Re^{i\theta})e^{i\theta}\ d\theta.$$
Mais, lorsque $z = Re^{i\theta}$, on a
$$\begin{aligned} |f(Re^{i\theta})e^{i\theta}| &= \frac{e^{-Rt\sin\theta}}{|1+z^2|} \\ &\leq \frac{1}{R^2-1} \end{aligned}$$
(C'est ici qu'on utilise l'hypothèse $t > 0$) Ainsi,
$$\lim_{R\to+\infty} iR\int_0^\pi f(Re^{i\theta})e^{i\theta}\ d\theta = 0.$$
On en déduit
$$e^{-t} = \frac{1}{\pi}\int_{-\infty}^{+\infty} f(t)\ dt$$

Ainsi si $X \sim \mathcal{C}(0,1)$, on a
$$\forall t > 0 \quad \varphi_X(t) = e^{-t}.$$
Comme la loi de X est symétrique ($P_X = P_{-X}$) et que $\varphi(0) = 1$, on a
$$\forall t \in \mathbb{R} \quad \varphi_X(t) = e^{-|t|}.$$

Si on pose $Y = bX + a$, on a $Y \sim \mathcal{C}(a,b)$, et alors $\varphi_Y(t) = \mathbb{E}e^{itY} = e^{it(bX+a)} = e^{ia}\mathbb{E}e^{itbX} = e^{ia}\varphi_X(bt) = e^{ia}e^{-b|t|}$.

□

4.3 Exercices sur les fonctions génératrices et les fonctions caractéristiques

1. (a) On suppose que X et Y sont des variables aléatoires à valeurs entières, de fonctions génératrices f et g. Soit A un événement indépendant de X et de Y, avec $P(A) = p$. On note Z la variable aléatoire définie par
$$Z(\omega) = \begin{cases} X(\omega) \text{ si } \omega \in A \\ Y(\omega) \text{ si } \omega \notin A \end{cases}$$
Montrer que la fonction génératrice de Z est $pf + (1-p)g$.

(b) On lance 3 fois de suite un dé à 6 faces. À chaque série de 6 lancers est associé un score. Le score se calcule ainsi Si le troisième lancer est un “1”, le score est le nombre de nombre pairs apparus dans les trois premiers lancers. Sinon, le score est le nombre de “6” apparus dans les trois premiers lancers.

Exemples :
- $2-4-1$ rapporte 2 points
- $6-1-2$ rapporte 1 point
- $1-4-2$ rapporte 0 point
- $5-2-3$ rapporte 0 point.

On demande de calculer la loi du score.

2. Soient (X_n) une suite de variables aléatoires indépendantes de même loi non dégénérée à valeurs dans $\mathbb{N}$ et T une variable aléatoire à valeurs dans $\mathbb{N}^*$ indépendante des précédentes. On définit pour tout $n \in \mathbb{N}^*$ la variable $S_n = \sum_{j=1}^{n} X_j$, puis $S(\omega) = S_{T(\omega)}(\omega)$ pour tout $\omega \in \Omega$.
   (a) Si G_T et G_X désignent les fonctions génératrices de T et X_1, montrer que la fonction génératrice de S est donnée par $G_S = G_T \circ G_X$.
   (b) *Formule de Wald*
   Si X_1 et T admettent les moyennes (espérances) m et t, montrer que $\mathbb{E}[S] = mt$.
3. Soit N une variable aléatoire suivant une loi de Poisson de paramètre λ et $X_1, \ldots, X_n \ldots$ une suite de v.a.r.i.i.d de lois de Bernoulli de paramètre p, cette suite étant indépendante de N. Montrer que $S = X_1 + X_2 \cdots + X_N$ suit loi de Poisson de paramètre λp.
4. (a) Soit K un vecteur aléatoire n-dimensionnel dont les composantes sont indépendantes et suivent la loi de Poisson de paramètre λ. Soit L un vecteur aléatoire n-dimensionnel dont les composantes sont indépendantes et suivent la loi Bernoulli de paramètre p. On suppose que K et L sont indépendants. Déterminer la fonction génératrice de $\langle K, L \rangle$.
   (b) On suppose que pour tout $k \geq 0$, X_k suit une loi de Poisson de paramètre $n\lambda$. Soit T une variable aléatoire indépendante de la suite $(X_k)_{k \geq 0}$ et suivant la loi binomiale $\mathcal{B}(n, p)$. Déterminer la fonction génératrice de X_T.
   (c) En déduire que $\langle K, L \rangle$ suit la même loi que X_T.
5. Un joueur joue à pile ou face avec $n \geq 2$ pièces équilibrées de la manière suivante : il lance simultanément ces n pièces ; s'il n'obtient aucun pile, son gain est nul et la partie s'arrête. S'il obtient au moins un pile, il relance la première pièce autant de fois qu'il a obtenu de piles à la

première phase du jeu et gagne autant d'unités que le nombre de piles obtenu lors de cette deuxième série de jets. On note X_1 le nombre de piles obtenu à la première étape, et X_2 le gain du joueur.

(a) Déterminer la fonction génératrice de X_2.

(b) En déduire

$$\forall k \in \{0, \dots, n\} \quad P(X_2 = k) = \binom{n}{k} \frac{1}{3^k} (\frac{3}{4})^n$$

6. Montrer que la convolée de deux lois de Cauchy est une loi de Cauchy.
7. Donner un exemple de variable aléatoire telle que $(\varphi_X)^2 = \varphi_{2X}$. En déduire que la propriété

$$\forall t \in \mathbb{R} \quad \varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$$

n'implique PAS que X et Y sont indépendantes.

8. Soit $X \sim \mathcal{N}(0, 1)$. Montrer que $\forall n \geq 0 \quad \mathbb{E}X^{2n} = \frac{(2n)!}{n!2^n}$.
9. Soit X et Y deux variables aléatoires admettant chacune un moment d'ordre 3. On pose $\varphi(x, y) = \mathbb{E}\exp(i(xX + yY))$. Montrer que X^2Y est intégrable. Régularité de φ ? Exprimer $\mathbb{E}X^2Y$ en fonction de φ.
10. Soit X une variable atéatoire suivant la loi exponentielle de paramètre 1 et ε une variable aléatoire prenant la valeur 1 avec probabilité 1/2 et la valeur -1 avec probabilité 1/2. On appelle *Loi de Laplace* la loi de εX.

    (a) Montrer que la loi de Laplace est une loi à densité.

    (b) Calculer la fonction caractéristique de la loi de Laplace.
11. On considère quatre variables aléatoires indépendantes $X_{1,1}$, $X_{1,2}$, $X_{2,1}$, $X_{2,2}$ de même loi $\mathcal{N}(0, 1)$.

    (a) On note $U = X_{1,1}X_{2,2}$ et $V = X_{1,2}X_{2,1}$. Déterminer la fontion caractéristique de U et de V.

    (b) Montrer que le déterminant $\begin{vmatrix} X_{1,1} & X_{1,2} \\ X_{2,1} & X_{2,2} \end{vmatrix}$. suit la loi de Laplace.
12. (a) Montrer qu'il existe une constante K telle que

$$f(x) = \frac{K}{\|x\|_2^{1/2}} \mathbb{1}_{B(0,1)}(x)$$

soit la densité d'une loi de probabilité sur $\mathbb{R}^3$.

    (b) Soit μ cette loi. Montrer qu'il existe une constante L non nulle telle que

$$\varphi_\mu(t) = L\frac{\cos \|t\|}{\|t\|^2} + O(\frac{1}{\|t\|^{2,5}})$$

Chapitre 5

Lois des grands nombres

5.1 Inégalités classiques

5.1.1 Inégalité de Markov

Théorème 49. *Soit X une variable aléatoire positive, intégrable. Alors, on a*

$$\forall a > 0 \quad P(X \geq a) \leq \frac{\mathbb{E}X}{a}.$$

Démonstration. Comme X est positive, on a

$$\mathbb{E}X = \int_{\Omega} x \, dP(\omega) \geq \int_{\{X \geq a\}} x \, dP(\omega) \geq \int_{\{X \geq a\}} a \, dP(\omega) = aP(X \geq a).$$

□

5.1.2 Inégalité de Tchebytchef

Théorème 50. *Soit X une variable aléatoire admettant un moment d'ordre 2. Alors, on a*

$$\forall a > 0 \quad P(|X - \mathbb{E}X| \geq a) \leq \frac{\operatorname{Var} X}{a^2}.$$

Démonstration.

$$P(|X - \mathbb{E}X| \geq a) = P(|X - \mathbb{E}X|^2 \geq a^2)$$

Il suffit alors d'appliquer l'inégalité de Markov à la variable aléatoire $Y = |X - \mathbb{E}X|^2$. Comme $\mathbb{E}Y = \operatorname{Var} X$, l'inégalité s'ensuit. □

5.2 Convergence presque sûre

Définition : on dit qu'une suite de variables (ou de vecteurs) aléatoires $(X_n)_{n\geq 0}$ converge presque sûrement vers une variable (ou un vecteur) aléatoire X lorsqu'il existe un ensemble mesurable $\Omega' \subset \Omega$ tel que $P(\Omega') = 1$ et que

$$\forall \omega \in \Omega' \subset \Omega \quad X_n(\omega) \to X(\omega).$$

On écrit alors $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$.

La convergence presque sûre n'est autre que la convergence presque partout relativement à une mesure de probabilité. On a alors les résultats classiques suivants : si $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$ et $Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} Y$, (avec X et Y dans $\mathbb{R}^d$, $d \geq 1$) alors

- $\forall a \in \mathbb{R} \quad aX_n \xrightarrow[n\to+\infty]{\text{p.s.}} aX$.
- $X_n + Y_n \xrightarrow[n\to+\infty]{\text{p.s.}} X + Y$.
- $\langle X_n, Y_n\rangle \xrightarrow[n\to+\infty]{\text{p.s.}} \langle X, Y\rangle$.

Plus généralement, si $X_1, \dots, X_n, \dots, X$ sont à valeurs dans un ouvert O et que $X_n \xrightarrow[n\to+\infty]{\text{p.s.}} X$, alors pour toute fonction f continue définie sur O, on a $f(X_n) \xrightarrow[n\to+\infty]{\text{p.s.}} f(X)$.

Il peut être intéressant de remarquer que la convergence presque sûre d'une suite de vecteurs aléatoires est équivalente à la convergence presque sûre de chacune des composantes.

5.2.1 Rappels d'analyse

En probabilités, le retour aux ε est très fréquent. Si l'on ne veut pas que cela devienne trop compliqué, il importe de bien connaître les outils d'analyse permettant de simplifier les choses.

Pour toute suite $(x_n)_{n\geq 0}$ de nombre réels, on peut définir

$$\overline{\lim_{n\to+\infty}}\ x_n = \lim_{n\to+\infty} \sup_{k\geq n} x_k$$

et

$$\underline{\lim_{n\to+\infty}}\ x_n = \lim_{n\to+\infty} \inf_{k\geq n} x_k.$$

Ces deux limites existent toujours dans $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty; +\infty\}$. La suite $(x_n)_n$ converge dans $\overline{\mathbb{R}}$ si et seulement si ces deux limites sont égales. Rappelons quelques propriétés des limites supérieures. Pour plus de détails, on pourra se reporter à un cours d'analyse.

– Pour $x \in \mathbb{R}$

$$\lim_{n\to+\infty} x_n = x \iff \overline{\lim_{n\to+\infty}} |x_n - x| = 0$$

–

$$\overline{\lim_{n\to+\infty}} x_n \leq M \iff \forall \varepsilon > 0 \quad \{n : x_n \geq M + \varepsilon\} \text{ est fini} \tag{5.1}$$

$$\overline{\lim_{n\to+\infty}} x_n \geq M \iff \forall \varepsilon > 0 \quad \{n : x_n \geq M - \varepsilon\} \text{ est infini} \tag{5.2}$$

– En prenant la contraposée de (5.1) et(5.2) , on a

$$\overline{\lim_{n\to+\infty}} x_n > M \iff \exists \varepsilon > 0 \quad \{n : x_n \geq M + \varepsilon\} \text{ est infini}$$

$$\overline{\lim_{n\to+\infty}} x_n < M \iff \exists \varepsilon > 0 \quad \{n : x_n \geq M - \varepsilon\} \text{ est fini}$$

–

$$\overline{\lim_{n\to+\infty}} (x_n + y_n) \leq \overline{\lim_{n\to+\infty}} x_n + \overline{\lim_{n\to+\infty}} y_n.$$

Nous verrons en exercice un exemple où l'inégalité est stricte

5.2.2 Limites supérieures, inférieures d'ensembles

Si $(A_n)_{n\geq 0}$ est une suite d'ensembles, on note

$$\overline{\lim_{n\to+\infty}} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$$

et

$$\underline{\lim_{n\to+\infty}} A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k.$$

Ainsi la limite supérieure d'une suites d'ensembles est l'ensemble des points qui appartiennent à une infinité de ces ensembles, tandis que la limite inférieure d'une suite d'ensembles est l'ensemble des points qui appartient à tous ces ensembles à partir d'un certain rang.

Ainsi, dire que $\{n : X_n(\omega) \geq M - \varepsilon\}$ est infini , c'est dire que ω appartient à $\overline{\lim_{n\to+\infty}} \{\omega : X_n(\omega) \geq M - \varepsilon\}$.

On en déduit

$$\{ \overline{\lim_{n\to+\infty}} X_n \geq M\} = \bigcap_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \geq M - \varepsilon\} \tag{5.3}$$

Par ailleurs, dire que

$$\{n : x_n \geq M + \varepsilon\} \text{ est fini },$$

c'est dire qu'à partir d'un certain rang , on a $x_n < M + \varepsilon$. Donc si ω est tel que $\{n : X_n(\omega) \geq M + \varepsilon\}$ est fini , c'est que $\omega \in \underline{\lim}_{n\to+\infty} \{X_n < M + \varepsilon\}$. On en déduit que

$$\{ \overline{\lim_{n\to+\infty}} X_n \leq M\} = \bigcap_{\varepsilon>0} \underline{\lim}_{n\to+\infty} \{X_n < M + \varepsilon\} \tag{5.4}$$

Si on remplace X_n par $-X_n$ et M par $-M$ dans (5.4), on obtient :

$$\{ \overline{\lim_{n\to+\infty}} - X_n \leq -M\} = \bigcap_{\varepsilon>0} \underline{\lim}_{n\to+\infty} \{-X_n < -M + \varepsilon\}$$

Soit

$$\{ \underline{\lim}_{n\to+\infty} X_n \geq M\} = \bigcap_{\varepsilon>0} \underline{\lim}_{n\to+\infty} \{X_n > M - \varepsilon\} \tag{5.5}$$

Et en passant aux complémentaires dans (5.4) et (5.5), on a

$$\{ \overline{\lim_{n\to+\infty}} X_n > M\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \geq M + \varepsilon\} \tag{5.6}$$

et

$$\{ \underline{\lim}_{n\to+\infty} X_n < M\} = \bigcup_{\varepsilon>0} \overline{\lim_{n\to+\infty}} \{X_n \leq M - \varepsilon\} \tag{5.7}$$

Si on fait subir à la formule (5.3) les mêmes transformations qu'à (5.4), on peut obtenir 3 autres formules.

Dans la pratique, comment fait-on si l'on veut montrer que $\overline{\lim_{n\to+\infty}} X_n = M$ presque sûrement ? Comme vous l'avez deviné, on montre $\overline{\lim_{n\to+\infty}} X_n \geq M$ presque sûrement, puis $\overline{\lim_{n\to+\infty}} X_n \leq M$ presque sûrement Comme la suite $\overline{\lim_{n\to+\infty}} \{X_n \geq M - \varepsilon\}$ est monotone en ε, on a

$$\{ \overline{\lim_{n\to+\infty}} X_n \geq M\} = \cap_{\varepsilon>\in\mathbb{Q}_+^*} \overline{\lim_{n\to+\infty}} \{X_n \geq M - \varepsilon\} \tag{5.8}$$

L'avantage est que l'intersection est maintenant dénombrable. Or, on a le résultat classique très utile suivant :

Théorème 51. *L'intersection d'une famille dénombrable d'événements est de probabilité 1 si et seulement si chacun des événements est de probabilité 1.*

Démonstration. Soit D un ensemble d'indexs dénombrable. $(A_n)_{n\in D}$ une famille d'événements indexée par D. On pose $A = \underset{n\in D}{\cap} A_n$. Pour tout n, $A \subset A_n$, donc $P(A) \leq P(A_n)$. Ainsi si $P(A) = 1$, on a pour tout $n \in D$ $P(A_n) = 1$. Réciproquement, on a

$$\begin{aligned} P(A^c) &= P(\underset{n\in D}{\cup} A_n^c) \\ &\leq \sum_{n\in D} P(A_n^c) \\ &\leq \sum_{n\in D} 0 \end{aligned}$$

Donc $P(A) = 1 - P(A^c) = 1 - 0 = 1$. □

Pour prouver que $\overline{\lim_{n\to+\infty}} X_n \geq M$ presque sûrement, il suffit donc de prouver que $\forall a < M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 1$

De la même manière, on voit que pour avoir $\overline{\lim_{n\to+\infty}} X_n \leq M$ presque sûrement, il suffit donc de prouver que $\forall a > M \quad P(\underline{\lim_{n\to+\infty}} \{X_n < a\}) = 1$, on de manière équivalente que $\forall a > M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 0$.

On peut donc énoncer le théorème suivant

Théorème 52. *Soit X_n une suite de variables aléatoires et M un réel. On suppose que*

1. $\forall \quad a < M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 1$
2. $\forall \quad a > M \quad P(\overline{\lim_{n\to+\infty}} \{X_n \geq a\}) = 0$

Alors

$$\overline{\lim_{n\to+\infty}} X_n = M \textit{ presque sûrement.}$$

Le théorème suivant très important en est une application directe

Théorème 53 (Critère fondamental de convergence presque-sûre). *La suite de variables aléatoires X_n converge presque sûrement vers la variable aléatoire X si et seulement si*

$$\forall\varepsilon > 0 \quad P(\overline{\lim_{n\to+\infty}} \{|X_n - X| \ge \varepsilon\}) = 0.$$

Démonstration. Il suffit d'appliquer le lemme précédent à la suite de variables aléatoires $(|X_n - X|)_{n\ge 0}$, avec $M = 0$ et a joue le rôle de ε. □

5.3 Convergence en probabilité

Définition : On dit que (X_n) converge en probabilité vers X si

$$\forall\varepsilon > 0 \qquad \lim_{n\to+\infty} P(|X_n - X| \ge \varepsilon) = 0.$$

5.3.1 Comparaison avec les autres modes de convergence

Convergence dans L^p et convergence en probabilité

Théorème 54. *La convergence dans L^p ($p \ge 1$) implique la convergence en probabilité*

Démonstration.

$$P(\|X_n - X\| \ge \varepsilon) = P(\|X_n - X\|^p \ge \varepsilon^p) \le \frac{\mathbb{E}\|X_n - X\|}{\varepsilon^p}.$$

□

Convergence presque sûre et convergence en probabilité

Théorème 55. *La convergence presque sûre implique la convergence en probabilité.*

Démonstration. Soit $\varepsilon > 0$. D'après le théorème 53, on a

$$P(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = 0.$$

Or, d'après le théorème de continuité séquencielle décroissante, on a

$$P(\overline{\lim_{n\to+\infty}} \{|X_n - X| \geq \varepsilon\}) = \lim_{n\to+\infty} P(\bigcup_{k\geq n} \{|X_k - X| \geq \varepsilon\})$$

Comme

$$0 \leq P(|X_n - X| \geq \varepsilon) \leq P(\bigcup_{k\geq n} \{|X_k - X| \geq \varepsilon\}),$$

on en déduit que

$$\lim_{n\to+\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Comme ε est quelconque, on peut dire que X_n converge en probabilité vers X. □

Théorème 56. *Soit $(X_n)_{n\geq 0}$ une suite convergeant en probabilité vers X. Alors, il existe une sous-suite X_{n_k} telle que $X_{n_k} \xrightarrow[k\to\infty]{p.s.} X$..*

Démonstration. On pose $n_0 = 0$, puis, pour $k \geq 1$:

$$n_k = \inf\{n > n_{k-1}; P(|X_n - X| \geq \frac{1}{k}) \leq \frac{1}{2^k}\}$$

À k fixé, $P(|X_n - X| \geq \frac{1}{k})$ tend vers 0 quand n tend vers l'infini, donc on a bien pour tout k : $n_k < +\infty$.

Maintenant, on a pour tout $k \geq 0$:

$$P(|X_{n_k} - X| \geq \frac{1}{k}) \leq \frac{1}{2^k}$$

Comme la série de terme général converge, le premier lemme de Borel-Cantelli nous permet d'affirmer que

$$P(\overline{\lim_{k\to\infty}} \{|X_{n_k} - X| \geq \frac{1}{k})\}) = 0,$$

ce qui est équivalent

$$P(\underline{\lim_{k\to\infty}} \{|X_{n_k} - X| < \frac{1}{k})\}) = 1,$$

ce qui veut dire que pour presque tout ω, il existe un $k_0(\omega)$ tel que

$$k \geq k_0(\omega) \Longrightarrow |X_{n_k}(\omega) - X(\omega)| < \frac{1}{k}),$$

ce qui implique bien sûr que $X_{n_k}(\omega)$ tend vers $X(\omega)$ pour P-presque tout ω. $\square$

5.3.2 Loi faible des grands nombres

Théorème 57. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires de même loi, admettant un moment d'ordre 2 et deux à deux non corrélées.*
On pose

$$S_n = \sum_{k=1}^{n} X_k \text{ et } M_n = \frac{1}{n} S_n.$$

Alors

1. $M_n \xrightarrow[n\to+\infty]{\mathrm{L}^2} \mathbb{E}X_0|^2 \to 0$. *On dit que M_n converge en moyenne quadratique vers $\mathbb{E}X_0$.*
2. *Et donc $M_n \xrightarrow[n\to+\infty]{\mathbb{P}} \mathbb{E}X_0$.*

Démonstration. $\mathbb{E}M_n = \frac{1}{n}\mathbb{E}S_n = \frac{1}{n}\sum_{k=1}^{n} EX_k = \frac{1}{n}n\mathbb{E}X_0 = \mathbb{E}X_0$. Par conséquent $|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{1}{n^2}\operatorname{Var} S_n$. Comme les X_k sont 2 à 2 non corrélées, on a

$$\operatorname{Var} S_n = \sum_{k=1}^{n} \operatorname{Var} X_k = n \operatorname{Var} X_1.$$

On a donc

$$\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \operatorname{Var} M_n = \frac{\operatorname{Var} X_1}{n}, \tag{5.9}$$

qui tend bien vers zéro. $\square$

5.4 Lemmes de Borel-Cantelli

5.4.1 Premier lemme de Borel-Canteli

Théorème 58. *Soit $(A_n)_{n\geq 1}$ une suite d'événements observables. Si la série de terme général $P(A_n)$ est convergente, alors $P(\overline{\lim_{n\to+\infty}} A_n) = 0$.*

Démonstration. On pose $B_n = \bigcup_{k\geq n} A_k$. la suite (B_n) est décroissante, et l'intersection des (B_n) est, par définition, $\overline{\lim}_{n\to+\infty} A_n$. D'après le théorème de continuité séquentielle décroissante, on a donc

$$0 \leq P(\overline{\lim_{n\to+\infty}} A_n) = \lim_{n\to+\infty} P(B_n).$$

Or

$$P(B_n) = P(\bigcup_{k\geq n} A_k) \leq \sum_{k\geq n} P(A_k) = r_n$$

Comme r_n est le reste d'ordre n d'une série convergente, r_n est de limite nulle, et donc, par comparaison $P(\overline{\lim}_{n\to+\infty} A_n) = 0$. □

5.4.2 Deuxième lemme de Borel-Cantelli

Théorème 59. *Soient $(A_n)_{n\geq 1}$ une suite d'évévements.*
On pose

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k} \textit{ et } N = \sum_{k=1}^{+\infty} \mathbb{1}_{A_k}$$

$$m_n = \sum_{k=1}^{n} P(A_k) = \mathbb{E}N_n$$

$$\textit{On a } \overline{\lim_{n\to+\infty}} A_n = \{N = +\infty\}.$$

Si

$$\lim_{n\to+\infty} m_n = +\infty \textit{ et } \lim_{n\to+\infty} \frac{\operatorname{Var} N_n}{m_n^2} = 0,$$

alors

$$P(\overline{\lim_{n\to\infty}} A_n) = 1.$$

Démonstration. Pour $m_n > a$, on a

$$\begin{aligned} P(N \leq a) &\leq P(N_n \leq a) \\ &\leq P(|N_n - m_n| \geq m_n - a) \leq \frac{\operatorname{Var} N_n}{(m_n - a)^2} \end{aligned}$$

En faisant tendre n vers $+\infty$, on en déduit que

$$\forall a \in \mathbb{N} \quad P(N \le a) = +\infty.$$

Ainsi

$$P(N < +\infty) = P(\lim_{a\to+\infty} \uparrow \{N \le a\}) = \lim_{a\to+\infty} P(N \le a) = \lim_{a\to+\infty} 0 = 0.$$

□

Théorème 60 ($2^{\text{ème}}$ lemme de Borel-Cantelli). *Soit $(A_n)_{n\ge1}$ une suite d'événements indépendants. Si la série de terme général $P(A_n)$ est divergente, alors $P(\overline{\lim}_{n\to+\infty} A_n) = 1$.*

Démonstration. On va appliquer le théorème précédent : comme les $(A_k)_{k\ge1}$ sont indépendants, leurs indicatrices sont des variables aléatoires indépendantes, et donc

$$\mathrm{Var}\, N_n = \sum_{k=1}^n \mathrm{Var}\, \mathbb{1}_{A_k} = \sum_{k=1}^n \mathrm{Var}\, P(A_k)(1-P(A_k) \le \sum_{k=1}^n P(A_k) = m_n$$

Ainsi $\frac{\mathrm{Var}\, N_n}{m_n^2} = \frac{1}{m_n}$. Comme $\lim_{n\to+\infty} m_n = \sum_{k=1}^{+\infty} P(A_k) = +\infty$, le résultat s'ensuit. □

Exercice : La conclusion du $2^{\text{ème}}$ lemme de Borel-Cantelli reste-t-elle vraie si l'on suppose seulement que les $(A_k)_{k\ge1}$ sont deux à deux indépendants ?

5.5 Loi forte des grands nombres

5.5.1 La loi forte des grands nombres

Théorème 61. *Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires deux à deux indépendantes, de même loi μ. On suppose que μ admet un moment d'ordre 1. Alors*

$$\frac{X_1+\dots+X_n}{n} \xrightarrow[n\to+\infty]{p.s.} \mathbb{E}X_1.$$

5.5.2 Probabilités et fréquences asymptotiques

Théorème 62. *Soit $(A_n)_{n\geq 0}$ une suite d'événements observables indépendants de même probabilité p. Pour ω dans l'univers Ω On note $N_n(\omega)$ le nombre d'événements qui sont réalisés parmi $A_1, \dots, A_n$. Ainsi, on a*

$$N_n = \sum_{k=1}^{n} \mathbb{1}_{A_k} \text{ et } f_n = \frac{1}{n} N_n.$$

Alors il existe un événement observable $\tilde{\Omega} \subset \Omega$ avec $P(\tilde{\Omega} \subset \Omega) = 1$ et

$$\forall \omega \in \tilde{\Omega} \subset \Omega \quad f_n(\omega) \to p.$$

Démonstration. Il suffit de poser $X_k = \mathbb{1}_{A_k}$ et d'appliquer le théorème 61. X_k admet bien un moment d'ordre 1 car $0 \leq X_k \leq 1$ et l'on a $\mathbb{E}X_1 = P(A_1) = p$. □

5.6 Exercices sur la convergence presque sûre

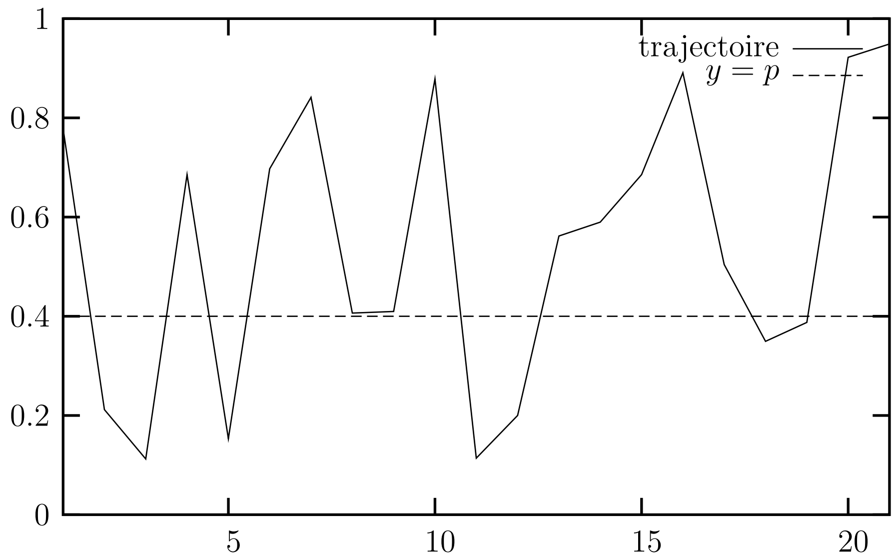
1. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(m, \sigma^2)$. Montrer que la suite $\frac{X_1^2+\cdots+X_n^2}{n}$ convergence presque sûrement et déterminer sa limite.
2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires identiquement distribuées telle qu'il existe $\alpha > 0$ avec $\mathbb{E}\exp(\alpha|X_1|) < +\infty$.
   Montrer que
   $$\lim_{n\to+\infty} \frac{X_n}{(\ln n)^{\frac{3}{2}}} = 0.$$
3. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant une loi exponentielle de paramètre 1.

   Calculer $\overline{\lim}_{n\to+\infty} \frac{X_n}{\ln n}$.
4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0, 1)$.

   (a) Calculer $\overline{\lim}_{n\to+\infty} \frac{X_n}{\sqrt{2\ln n}}$.

   (b) On se donne maintenant une deuxième suite $(Y_n)_{n\geq 1}$ de variables aléatoires indépendantes suivant la loi $\mathcal{N}(0, 1)$, cette deuxième suite étant indépendante de la première. Comparer
   $$\overline{\lim}_{n\to+\infty} \frac{X_n}{\sqrt{2\ln n}} + \overline{\lim}_{n\to+\infty} \frac{Y_n}{\sqrt{2\ln n}} \text{ et } \overline{\lim}_{n\to+\infty} \frac{(X_n+Y_n)}{\sqrt{2\ln n}}.$$

5. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi binomiale $\mathcal{B}(n, \frac{1}{n^{1,01789}})$. Déterminer l'ensemble des valeurs d'adhérences de la suite (X_n)
6. Soit $p \in [0,1]$ et $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. On note T_n le nombre de fois où le graphe associé à (U_n) coupe la droite d'équation $y = p$ avant le temps n. Dans notre exemple, $p = 0.4$ et $T_{20} = 8$.



Montrer que $\frac{T_n}{n}$ converge presque sûrement et déterminer la limite.
7. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires de Bernoulli indépendantes de paramètre $1/2$. On pose
$$M_n = \begin{pmatrix} 2 + X_n & 1 \\ 1 & 2 + X_n \end{pmatrix}$$
et
$$A_n = M_n \times M_{n-1} \times \dots M_2 \times M_1.$$
   (a) Montrer que la suite $(\det A_n)^{1/n}$ converge presque sûrement et déterminer la limite.
   (b) Soit $(x, y) \in \mathbb{R}^2 \backslash \{(0,0)\}$. On pose
   $$X_n = A_n \begin{pmatrix} x \\ y \end{pmatrix}$$
   Montrer que
   $$\|X_n\|^{1/n} \xrightarrow[n\to+\infty]{\text{p.s.}} \begin{cases} \sqrt{2} & \text{si } x + y = 0 \\ \sqrt{12} & \text{si } x + y \neq 0 \end{cases}$$

8. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, où $(\lambda_n)_{n\geq 1}$ est une suite tendant vers 0 en l'infini. Montrer que la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = X_1X_2\dots X_n$ est nulle à partir d'un certain rang.
9. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, avec

$$\sum_{n=1}^{+\infty} \lambda_n^2 < +\infty$$

   Montrer que la suite $(X_n)_{n\geq 1}$ est presque sûrement bornée.
10. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes telles que pour tout n, X_n suive une loi de Poisson de paramètre λ_n, avec

$$\lambda_n = o(\ln n)$$

   Montrer que la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = X_1X_2\dots X_n$ est nulle à partir d'un certain rang.
11. Soit $(X_n)_{n\geq 2}$ une suite de variables aléatoires indépendantes telle que pour tout n, X_n suive une loi de Poisson de paramètre $2\ln n$. Montrer que

$$X_2X_3\dots X_n \xrightarrow[n\to+\infty]{\text{p.s.}} \begin{cases} 0 & \text{avec probabilité } p \\ +\infty & \text{avec probabilité } 1-p \end{cases},$$

   où $p \in]0,1[$.

Chapitre 6

Vecteurs gaussiens

Définition: On dit qu'un vecteur aléatoire $X \in \mathbb{R}^d$ est gaussien si pour tout $a \in \mathbb{R}^d$ la variable aléatoire $\langle X, a\rangle$ est gaussienne .

6.1 Image affine d'un vecteur gaussien

Théorème 63. *L'image d'un vecteur gaussien X d'espérance m_X et de matrice de covariance C_X par une application affine $x \mapsto Ax + b$ est un vecteur gaussien d'espérance $m_Y = Am_X + b$ et de matrice de covariance $C_Y = AC_X A^*$.*

Démonstration. On pose $Y = AX + b$. Soit $a \in \mathbb{R}^d$. $\langle Y, a\rangle = \langle AX, a\rangle + \langle b, a\rangle = \langle X, A^*a\rangle + \langle b, a\rangle$. Comme X est un vecteur gaussien $\langle X, A^*a$ est une variable aléatoire gaussienne. Quand on ajoute une constante à variable aléatoire gaussienne, on obtient une variable aléatoire gaussienne. Ainsi, pour tout A, $\langle Y, a\rangle$ est une variable aléatoire gaussienne, donc Y est un vecteur gaussien. L'expression de l'espérance et de la covariance est une conséquence du théorème 30. □

Corollaire 12. *Si $X = (X_1, \dots, X_d)$ est gaussien, alors pour tout $I \in \{1, \dots, d\}$, le vecteur $(X_i)_{i\in I}$ est gaussien*

Démonstration. $(X_i)_{i\in I}$ est l'image de $X = (X_1, \dots, X_d)$ par l'application linéaire

$$\begin{array}{rcl} \mathbb{R}^{\{1,\dots,d\}} & \to & \mathbb{R}^I \\ (x_i)_{i\in\{1,\dots,d\}} & \mapsto & (x_i)_{i\in D} \end{array}$$

□

6.2 Exemple fondamental

Théorème 64. *Soit $X_1, \dots, X_d$ d variables aléatoires gaussiennes indépendantes. Alors $X = (X_1, \dots, X_d)$ est un vecteur gaussien.*

Démonstration. Pour $k \in \{1, \dots d\}$, on pose $S_k = \sum_{i=1}^k a_X X_k$. On montre par récurrence sur k que S_k est une variable aléatoire gaussienne. Pour $k = 1$, $S_1 = a_1 X_1$: quand on multiplie une variable gaussienne par une constante, on a une variable aléatoire gaussienne. Supposons acquis que S_k est gaussienne : on a $S_{k+1} = S_k + a_{k+1} X_{k+1}$. $a_{k+1} X_{k+1}$ est une variable gaussienne indépendante de S_k, car S_k est $\sigma(X_1, \dots X_k)$-mesurable. Or d'après le théorème 33, la somme de deux variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc S_{k+1} est une variable aléatoire gaussienne.

Comme $\langle X, a\rangle = S_d$ quel que soit a , on en déduit que X est un vecteur gaussien. □

6.3 Lois gaussiennes

Théorème 65. *Soit C une matrice symétrique positive $d \times d$ et $m \in \mathbb{R}^d$. Alors, on peut construire un vecteur gaussien admettant m comme espérance et C comme matrice de covariance.*

Démonstration. Comme C est symétrique,on peut la diagonaliser avec une matrice de passage orthogonale. Comme elle est symétrique, les valeurs propres sont positives. Ainsi, on peut trouver une matrice O orthogonale et des réels positifs tels que

$$C = O \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \lambda_{n-1} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_n \end{pmatrix} O^*$$

Posons

$$A = O \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \sqrt{\lambda_{n-1}} & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \sqrt{\lambda_n} \end{pmatrix} O^*$$

et soient $(X_1, \dots, X_d)$ d variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$: l'espérance de X est $m_X = 0$ sa matrice de covariance $C_X = \mathrm{Id}_{\mathbb{R}^d}$. D'après le théorème 64, X est gaussien donc, d'après le théorème 30, $Y = AX + m$ est un vecteur gaussien d'espérance $A.0 + m = m$ et de covariance $AC_XA^* = AA^* = C$. □

Théorème 66. *Soit X et Y deux vecteurs gaussiens ayant même espérance et même matrice de covariance. Alors X et Y ont même loi.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $V = \langle X, A\rangle$ et $W = \langle Y, A\rangle$. L'espérance de V est $\langle m_X, a\rangle$ est la covariance de V est $\mathrm{Covar}(\langle X, a\rangle, \langle X, a\rangle) = \langle C_X a, a\rangle$. Comme X est gaussien, V est gaussienne, donc $V \sim \mathcal{N}(\langle m_X, a\rangle, \langle C_X a, a\rangle)$. De même $V \sim \mathcal{N}(\langle m_Y, a\rangle, \langle C_Y a, a\rangle)$. Comme $m_X = m_Y$ et $C_X = C_Y$, on en déduit que $V = \langle X, A\rangle$ et $W = \langle Y, A\rangle$ ont même loi. Comme cela est vrai quel que soit a, on déduit du théorème 40que X et Y ont même loi. □

Définition: Soit $m \in \mathbb{R}^d$ et C une matrice $d \times d$ symétrique positive. On note $\mathcal{N}(m, C)$ la loi commune à tous les vecteurs gaussiens admettant m comme espérance et C comme matrice de covariance.

La pertinence de cette définition est assurée par les théorèmes 65 et 66

6.4 Lois gaussiennes et indépendance

Théorème 67. *Soit $d_1, \dots, d_n$ n entiers positifs de somme d. Soit $C_1, \dots C_n$ n matrices symétriques positives, et $m_1, \dots m_n$ n vecteurs C_i étant de taille n_i. Alors*

$$\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \cdots \otimes \mathcal{N}(m_n, C_n) = \mathcal{N}(m, C),$$

avec

$$m = \begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix} \text{ et } C = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}$$

Démonstration. Soit $Y_1, \dots, Y_n$ n vecteurs gaussiens indépendants, avec $Y_1 \sim \Updownarrow_\rangle, \mathcal{C}_\rangle$. On va d'abord montrer que $X = (Y_1, \dots, Y_n)$ est gaussien. Soit $a \in \mathbb{R}^d$. On peut écrire $a = (a_1, \dots, a_d)$, avec a_i de taille d_i. On a

$$\langle X, a \rangle = \sum_{k=1}^{d} \langle Y_k, a_k \langle.$$

Comme Y_k est gaussien, chaque variable aléatoire $\langle Y_k, a_k \langle$ est une variable aléatoire gaussienne. Comme les Y_k sont indépendants, les variables aléatoires $\langle Y_k, a_k \langle$ sont indépendantes. Or on sait qu'une somme de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne, donc $\langle X, a \rangle$ est une variable aléatoire gaussienne. Il s'ensuit que X est un vecteur gaussien. L'expression de l'espérance et de la matrice de covariance ne pose pas de problème, puisque des variables aléatoires indépendantes ne sont pas corrélées. □

Théorème 68. *Soit $d_1, \dots, d_n$ n entiers positifs de somme d. On suppose que $X = (X_1, \dots, X_n)$ est un vecteur gaussien dont la matrice de covariance est diagonale par blocs :*

$$C_X = \begin{pmatrix} C_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & C_n \end{pmatrix}$$

Alors, si l'on pose $Y_1 = (X_1, \dots, X_{d_1})$, $Y_2 = (X_{d_1+1}, \dots, X_{d_1+d_2})$, $Y_n = (X_{d_1+d_2+\dots+d_{n-1}+1}, \dots, X_{d_1+d_2+\dots+d_n})$, les vecteurs $Y_1, \dots Y_n$ sont des vecteurs gaussiens indépendants.

Démonstration. On voit que X a même espérance et même matrice de covariance que le vecteur aléatoire considéré au théorème précédent. Comme X et le vecteur aléatoire considéré au théorème précédent sont tous deux gaussiens, ils ont tous deux la même loi, donc la loi de X est $\mathcal{N}(m_1, C_1) \otimes \mathcal{N}(m_2, C_2) \cdots \otimes \mathcal{N}(m_n, C_n)$, ce qui signifie que les Y_i sont indépendants et que pour tout i, on a $Y_i \sim \mathcal{N}(m_i, C_i)$. □

En particulier, on a le corollaire suivant :

Corollaire 13. *Si le vecteur gaussien $X = (X_1, \dots, X_d)$ a une matrice de covariance dont tous les termes non-diagonaux sont nuls, alors $X_1, \dots, X_d$ sont des variables aléatoires indépendantes.*

6.5 Lois gaussiennes à densité

Théorème 69. *Soit C une matrice symétrique définie positive et $m \in \mathbb{R}^d$. La loi sur $\mathbb{R}^d$ $\mathcal{N}(m,C)$ admet comme densité par rapport à la mesure de Lebesgue la fonction*

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp(-\frac{1}{2}\langle C^{-1}(y-m), y-m\rangle).$$

Démonstration. On reprend les notations de la preuve du théorème 65. On a $X \sim \mathcal{N}(m,C)$. Par rapport au cas général, on gagne le fait que C est définie positive, ce qui implique que les λ_i sont strictement positifs, et donc que A est inversible. Comme X est composé de n vecteurs aléatoires indépendants à densité, la densité de X est le produit des densités, soit

$$f_X(x) = \prod_{k=1}^{d} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x_k^2}{2}) = \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle x, x\rangle).$$

D'après le théorème 32, $Y = AX + m$ admet comme densité

$$\frac{1}{\det A} f_X(A^{-1}(y-m)) = \frac{1}{\det A} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\langle A^{-1}(y-m), A^{-1}(y-m)\rangle).$$

Or

$$\begin{aligned}
\langle A^{-1}(y-m), A^{-1}(y-m)\rangle &= \langle (A^{-1})^* A^{-1}(y-m), y-m\rangle \\
&= \langle (A^{-1})^* A^{-1}(y-m), y-m\rangle \\
&= \langle (AA^*)^{-1}(y-m), y-m\rangle \\
&= \langle C^{-1}(y-m), y-m\rangle
\end{aligned}$$

D'autre part, $\det C = \det AA^* = (\det A)^2$, donc $\det A = \sqrt{\det C}$. On en déduit que la densité de Y (c'est à dire de la loi de Y) est

$$f_{m,C}(y) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det C}} \exp(-\frac{1}{2}\langle C^{-1}(y-m), y-m\rangle).$$

□

6.6 Fonction caractéristique des vecteurs gaussiens

Théorème 70. *En dimension quelconque, la fonction caractéristique de la loi normale $\mathcal{N}(m,C)$ est*

$$x \mapsto \exp(i\langle x, m\rangle \exp(-\frac{1}{2}\langle Cx, x\rangle)$$

Démonstration.

$$\mathbb{E}\exp(i\langle X,t\rangle) = \mathbb{E}\exp(iY) = \varphi_Y(1),$$

où $Y = \langle X,t\rangle$. comme X est gaussien de covariance C et d'espérance m, Y est gaussien de covariance $\langle Cx,x\rangle$ et d'espérance $\langle x,m\rangle$. On déduit donc le résultat de la formule précédente. □

6.7 Exercices sur les vecteurs gaussiens

1. Soient U et V deux variables aléatoires gaussiennes centrées et f une fonction croissante. Montrer que

$$\mathbb{E}\|U\|^2 < \mathbb{E}\|V\|^2 \Longrightarrow \mathbb{E}f(\|U\|) \leq \mathbb{E}f(\|V\|)$$

2. Soit X,Y deux variables aléatoires indépendantes suivant $\mathcal{N}(0,1)$. On pose

$$U = \begin{pmatrix} 2X \\ 2X \end{pmatrix} \text{ et } V = \begin{pmatrix} 2X \\ \sqrt{5}Y \end{pmatrix}.$$

   Montrer que U et V sont gaussiens centrés, puis que l'on a

$$\mathbb{E}\|U\|^2 < \mathbb{E}\|V\|^2,$$

   tandis que

$$\mathbb{E}\|U\|^4 > \mathbb{E}\|V\|^4.$$

   Comparer avec le résultat de l'exercice précédent.

3. Soit X un vecteur gaussien de matrice de covariance C. Montrer que

$$\mathbb{E}\|X\|_2^2 = \mathrm{Tr}C.$$

4. (a) Montrer qu'il existe un vecteur gaussien dont la matrice de covariance est

$$C = \begin{pmatrix} 5 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 2 \end{pmatrix}$$

   (b) Déterminer

$$\sup_{a\in\mathbb{R}^3} \frac{\mathbb{E}|\langle X,a\rangle|}{\|a\|_2}.$$

5. Montrer qu'on peut trouver des variables aléatoires X_1, X_2, X_3 telles que pour tout $i \in \{1,2,3\}$ $X_i \sim \mathcal{N}(0,3)$ et que $\text{Covar}(X_1,X_2) = \text{Covar}(X_1,X_3) = \text{Covar}(X_3,X_2) = 1$. Calculer la variance de $X_1 + 2X_2 + 3X_3$.
6. (a) Soit X et Y deux variables aléatoires indépendantes suivant la loi $\mathcal{N}(0,1)$. Calculer la matrice de covariance du couple $(X, \alpha X + \beta Y)$
   (b) Soit (Z,T) un vecteur gaussien centré, avec $\text{Var}\, Z = 1$. Montrer qu'il existe α et β tels que la matrice de covariance de (Z,T) soit
   $$\begin{pmatrix} 1 & \alpha \\ \alpha & \alpha^2+\beta^2 \end{pmatrix}.$$
   En déduire que $\mathbb{E}Z^2T^2 = 3\alpha^2 + \beta^2 = \text{Var}\, Z + 2\,\text{Covar}(Z,T)^2$,
   (c) Soit (Z,T) est un vecteur gaussien centré quelconque – on ne suppose plus que $\text{Var}\, Z = 1$ –, montrer
   $$\mathbb{E}Z^2T^2 = \text{Var}\, Z\, \text{Var}\, T + 2\,\text{Covar}(Z,T)^2.$$
   (d) Soit (Z,T) un vecteur gaussien de matrice de covariance
   $$\begin{pmatrix} 2 & 1 \\ 1 & 5 \end{pmatrix}.$$
   Calculer $\mathbb{E}Z^2T^2$.
   (e) Soit X un vecteur gaussien de matrice de covariance C. Montrer que
   $$\mathbb{E}\|X\|^4 = 2\text{Tr}C^*C + (\text{Tr}C)^2.$$
7. Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0,C)$ et α est un réel vérifiant $\alpha < \rho(C)^{-1}$, où $\rho(C)$ est le rayon spectral de C. Montrer que
   $$\mathbb{E}\exp(\frac{\alpha}{2}\|X\|_2^2) = \prod_i (1-\alpha\lambda_i)^{-1/2}.$$
   où les λ_i sont les valeurs propres de C.
8. (a) Soit X un vecteur gaussien dans $\mathbb{R}^n$ de loi $\mathcal{N}(0,I_n)$ et O une matrice orthogonale. Montrer que $Y = OX$ a la même loi que X.
   (b) On appelle loi du chi-deux à n degrés de liberté et on note $\chi^2(n)$ la loi de $\|X\|_2^2$. Montrer que $\chi^2(1)$ admet la densité
   $$x \mapsto \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-\frac{x}{2}}$$
   En déduire que $\chi^2(n)$ est une loi à densité.
9. Soit A la matrice d'un projecteur orthogonal de rang r et X une variable aléatoire suivant la loi $\mathcal{N}(0,A)$. Montrer que $\|X\|_2^2$ suit la loi $\chi^2(r)$.

Chapitre 7

Convergence en loi

7.1 Convergence en loi

7.1.1 Définition

On dit qu'une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilités sur $\mathbb{R}^d$ converge faiblement vers la mesure de probabilité μ lorsque pour toute fonction f continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\ d\mu_n = \int_{\mathbb{R}^d} f\ d\mu.$$

Par extension, on dit qu'une suite de variables aléatoires X_n converge en loi vers la variable aléatoire X (ou vers la loi μ) si la suite de mesures P_{X_n} converge faiblement vers P_X (ou vers la loi μ).

Ainsi, dire que X_n converge en loi vers X signifie que pour toute fonction continue bornée, $\mathbb{E}f(X_n)$ converge vers $\mathbb{E}f(X)$.

Rappel : si μ et ν sont deux mesures telles que pour toute fonction f continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a $\int_{\mathbb{R}^d} f\ d\mu = \int_{\mathbb{R}^d} g\ d\mu$, alors $\mu = \nu$.

On en déduit immédiatement l'unicité de la limite pour la convergence en loi.

Théorème 71. *Soit g une fonction continue définie sur $\mathbb{R}^d$. Si $(X_n)_{n\geq 1}$ converge en loi vers X, alors la suite $(Y_n)_{n\geq 1}$ définie par $Y_n = g(X_n)$ converge en loi vers $g(X)$.*

Démonstration. Soit f une fonction continue bornée. $\mathbb{E}f(Y_n) = \mathbb{E}f(g(X_n)) = \mathbb{E}(f\circ g)(X_n)$. Comme f et g sont continues, $f\circ g$ est continue. Comme f est bornée, $f\circ g$ bornée. $f\circ g$ est continue, bornée et $(X_n)_{n\geq 1}$ converge en loi vers X, donc $\mathbb{E}(f\circ g)(X_n)$ converge vers $\mathbb{E}(f\circ g)(X) = \mathbb{E}f(g(X))$, ce qui achève la preuve. $\square$

Corollaire 14. *Soient* $(X_n)_{n\geq 1}$ *et* $(Y_n)_{n\geq 1}$ *deux suites de vecteurs aléatoires à valeurs dans* $\mathbb{R}^d$*. Si* $((X_n, Y_n))_{n\geq 1}$ *converge en loi vers* (X, Y)*, alors*
- $X_n + Y_n$ *converge en loi vers* $X + Y$*.*
- $X_n Y_n$ *converge en loi vers* XY*.*

Démonstration. Il suffit d'appliquer le théorème précédent à la fonction continue $(x, y) \mapsto x + y$ et à la fonction continue $(x, y) \mapsto xy$. □

7.1.2 Premiers exemples

Un critère de convergence en loi

Lemme 4 (Théorème de Scheffé). *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré;* $f, (f_n)_{n\geq 1} : \Omega \to \mathbb{R}$ *des applications intégrables par rapport à* μ *telles que*
- $f_n \xrightarrow[n\to+\infty]{\mu p.p.} f$.
- $\lim_{n\to\infty} \int_\Omega f_n dx = \int_\Omega f dx$.

Alors $f_n \xrightarrow[n\to+\infty]{L^1(\mu)} f$.

Démonstration. On a

$$\begin{cases} f + f_n = \max(f, f_n) + \min(f, f_n) \\ |f - f_n| = \max(f, f_n) - \min(f, f_n)| \end{cases}$$

Alors

$$f + f_n - |f - f_n| = 2\min(f, f_n),$$

d'où

$$|f - f_n| = f + f_n - 2\min(f, f_n) = -f + f_n + 2(f - \min(f, f_n)).$$

Ainsi

$$\|f_n - f\|_{L^1(\mu)} = -\int_\Omega f(x) d\mu(x) + \int_\Omega f_n(x) d\mu(x) + 2\int_\Omega (f - \min(f, f_n))(x) d\mu(x)$$

D'après la deuxième hypothèse, $-\int_\Omega f(x)d\mu(x) + \int_\Omega f_n(x)d\mu(x)$ tend vers 0 lorsque n tend vers $+\infty$. D'autre part, on a $0 \leq f - \min(f, f_n) \leq f$ et $f - \min(f, f_n) \xrightarrow[n\to+\infty]{\mu \text{ p.s.}} 0$. D'après le théorème de convergence dominée, on a donc $\int_\Omega f - \min(f, f_n)\, d\mu \xrightarrow[n\to+\infty]{} 0$, d'où le résultat.

□

Corollaire 15. *Soit* $(\Omega, \mathcal{A}, \mu)$ *un espace mesuré. Soient* $\nu, (\nu_n)_{n\geq 1}$ *des mesures de probabilités sur* $(\Omega, \mathcal{A}, \mu)$ *admettant les densités* $f, (f_n)_{n\geq 1}$ *par à* μ*. On suppose que* $f_n \xrightarrow[n\to+\infty]{\mu\ p.p.} f$*. Alors* ν_n *converge faiblement vers* ν*.*

Démonstration. Soit g une fonction continue bornée sur Ω. On a

$$\begin{aligned} |\int g(x)\, d\nu_n(x) - \int g(x)\, d\nu(x)| &= |\int g(x)f(x)\, d\mu(x) - \int g(x)f_n(x)\, d\mu(x)| \\ &\leq \int |g(x)(f(x)-f_n(x))|\, d\mu(x) \\ &\leq \|g\|_\infty \int |f(x)-f_n(x)|\, d\mu(x), \end{aligned}$$

qui tend vers 0 d'après le théorème de Scheffé. Comme cette convergence a lieu pour toute fonction continue bornée g, on peut dire que ν_n converge en loi vers ν. □

Corollaire 16. *Soit $X, (X_n)_{n\geq 1}$ des variables aléatoires discrètes à valeurs dans un ensemble dénombrable D. On suppose que*

$$\forall k \in D \qquad \lim_{n\to+\infty} P(X_n = k) = P(X = k)$$

Alors $(X_n)_{n\geq 1}$ converge en loi vers X.

Démonstration. Il suffit de remarquer que $P_X, P_{X_1}, \dots, P_{X_n} \dots$ ont une densité par rapport à la mesure de comptage sur D et appliquer le corollaire précédent. □

Convergence de la loi binomiale vers la loi de Poisson

Théorème 72. *Soit, pour $n \geq 1$ une variable aléatoire X_n suivant une loi binomiale de paramètres n et p_n. On suppose que l'on a*

$$\lim_{n\to\infty} np_n = \lambda > 0$$

Alors, la suite X_n converge en loi vers la loi de Poisson $\mathcal{P}(\lambda)$.

Démonstration. D'après le corollaire 16, il suffit de montrer que pour tout entier $k \geq 0$, $P(X_n = k)$ converge vers $e^{-\lambda}\frac{\lambda^k}{k!}$. On a

$$P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \binom{n}{k} p_n^k (1-p_n)^n (1-p_n)^{-k}$$

On a $\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} \sim \frac{n^k}{k!}$, $p_n^k \sim (\lambda/n)^k = \lambda^k n^{-k}$. D'autre part, $\ln(1-p_n)^n = n\ln(1-p_n) = n(-p_n + o(p_n)) \sim -np_n \sim -\lambda$. Ainsi $\ln(1-p_n)^n$ converge vers $-\lambda$ donc $(1-p_n)^n$ converge vers $e^{-\lambda}$. En mettant bout à bout les équivalents, on obtient le résultat souhaité. □

Application pratique Si n est “grand” et np ”pas trop grand”, on peut remplacer la loi binomiale par une loi de Poisson. D'après une grand-mère statisticienne, n est grand à partir de 30 et np n'est pas trop grand jusqu'à 10. Ce théorème peut etre interprété de la manière suivante : la loi de Poisson est une bonne modélisation pour le nombre de fois où un événement rare survient.

Convergence de la loi hypergéométrique vers la loi binomiale

Théorème 73. *Soit, pour $j \geq 1$ une variable aléatoire X_j suivant une loi hypergéométrique de paramètres $\mathcal{H}(N_j, n_j, k)$. On suppose que N_j tend vers l'infini et que l'on a*

$$\lim_{j\to\infty} \frac{n_j}{N_j} = p.$$

Alors, la suite $(X_j)_{j\geq 1}$ converge en loi vers la loi binomiale $\mathcal{B}(k, p)$.

Démonstration. D'après le corollaire 16, il suffit de montrer que pour tout entier $i \geq 0$, $P(X_j = i)$ converge vers $\binom{k}{i} p^i (1-p)^{k-i}$. On a

$$P(X_j = i) = \frac{\binom{n_j}{i}\binom{N_j - i}{k-i}}{\binom{N_j}{k}} \sim \frac{\frac{n_j^i}{i!}\frac{(N_j - n_j)^{k-i}}{(k-i)!}}{\frac{N_j^k}{k!}}.$$

Or

$$\frac{\binom{n_j}{i}\binom{N_j - i}{k-i}}{\binom{N_j}{k}} \sim \frac{\frac{n_j^i}{i!}\frac{(N_j - n_j)^{k-i}}{(k-i)!}}{\frac{N_j^k}{k!}} = \frac{k!}{i!(k-i)!}(\frac{n_j}{N_j})^i(1 - \frac{n_j}{N_j})^{k-i},$$

qui converge vers $\binom{k}{i} p^i (1-p)^{k-i}$ lorsque j tend vers l'infini, ce qui achève la preuve.

□

7.1.3 Théorème du porte-manteau

Théorème 74. *Les propositions suivantes sont équivalentes*

1. *μ_n converge faiblement vers μ.*
2. *Pour toute fonction f uniformément continue bornée de $\mathbb{R}^d$ dans $\mathbb{R}$, on a*

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f \ d\mu_n = \int_{\mathbb{R}^d} f \ d\mu.$$

3. *Pour tout fermé F, $\mu(F) \geq \overline{\lim_{n\to+\infty}} \mu_n(F)$.*

4. *Pour tout ouvert O, $\mu(O) \leq \underline{\lim_{n\to+\infty}} \mu_n(O)$.*

5. *Pour tout borélien A dont la frontière ∂A vérifie $\mu(\partial A) = 0$, on a $\lim_{n\to+\infty} \mu_n(A) = \mu(A)$.*

Démonstration. On va prouver successivement (3) $\iff$ (4), (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (5) $\Longrightarrow$ (1).

– Pour (3) $\iff$ (4), il suffit de remarquer que

$$\begin{aligned}
\sup_{O\textbf{ouvert}} \big(\mu(O) - \underline{\lim_{n\to+\infty}} \mu_n(O)\big) &= \sup_{F\textbf{fermé}} \big(\mu(F^c) - \underline{\lim_{n\to+\infty}} \mu_n(F^c)\big) \\
&= \sup_{F\textbf{fermé}} \big((1-\mu(F)) - \underline{\lim_{n\to+\infty}} (1-\mu_n(F))\big) \\
&= \sup_{F\textbf{fermé}} \big(-\mu(F) - \underline{\lim_{n\to+\infty}} \mu_n(F)\big) \\
&= \sup_{F\textbf{fermé}} \big(-\mu(F) + \overline{\lim_{n\to+\infty}} \mu_n(F)\big)
\end{aligned}$$

Ainsi, si l'un des sup est négatif, l'autre aussi.

– Preuve de (2) $\Longrightarrow$ (3).
Soit F un fermé de $\mathbb{R}^d$. On pose $d_F(x) = d(x,F) = inf(|y-x|; y \in F)$ et, pour $\varepsilon > 0$, G_ε est la fonction continue par morceaux définie par $G_\varepsilon(x) = (1 - \frac{|x|}{\varepsilon})\mathbb{1}_{[-\varepsilon,+\varepsilon]}(x)$. On a $G_\varepsilon \circ d_F \geq \mathbb{1}_F$, donc

$$\overline{\lim_{n\to+\infty}} \int G_\varepsilon \circ d_F \; d\mu_n \geq \overline{\lim_{n\to+\infty}} \int \mathbb{1}_F \; d\mu_n$$

Comme $G_\varepsilon \circ d_F$ est uniformément continue (c'est la composée d'une application 1-lipschitzienne par une application $\frac{1}{\varepsilon}$-lipschitzienne), on a

$$\overline{\lim_{n\to+\infty}} \int G_\varepsilon \circ d_F \; d\mu_n = \int G_\varepsilon \circ d_F \; d\mu,$$

d'où

$$\forall \varepsilon > 0 \quad \int G_\varepsilon \circ d_F \; d\mu \geq \overline{\lim_{n\to+\infty}} \mu_n(F)$$

Or $\int G_\varepsilon \circ d_F \ d\mu = \int G_\varepsilon \ d\mu_{d_F}$ Lorsque ε tend vers 0, G_ε converge vers l'indicatrice de 0, donc par convergence dominée, $\int G_\varepsilon \circ d_F \ d\mu = \int G_\varepsilon \ d\mu_{d_F}$ converge vers

$$\int \mathbb{1}_{\{0\}} \ d\mu_{d_F} = \mu_{d_F}(\{0\}) = \mu(d_F = 0) = \mu(F).$$

– Preuve de $(3,4) \Longrightarrow 5$ On a $\overset{\circ}{A} \subset A \subset \overline{A}$ D'où

$$\varlimsup_{n\to+\infty} \mu_n(A) \leq \varlimsup_{n\to+\infty} \mu_n(\overline{A}) \leq \mu(\overline{A}).$$

D'autre part

$$\varliminf_{n\to+\infty} \mu_n(A) \geq \varliminf_{n\to+\infty} \mu_n(\overset{\circ}{A}) \geq \mu(\overset{\circ}{A}).$$

Ainsi, on a

$$\mu(\overset{\circ}{A}) \leq \varliminf_{n\to+\infty} \mu_n(A) \leq \varlimsup_{n\to+\infty} \mu_n(A) \leq \mu(\overline{A}).$$

Comme $\mu(\overline{A}) - \mu(\overset{\circ}{A}) = \mu(\partial A) = 0$, la suite $(\mu_n(A))_{n\geq 1}$ a une limite supérieure qui coïncide avec sa limite inférieure, donc converge vers $\mu(\overline{A}) = \mu(\overset{\circ}{A})$, c'est à dire vers $\mu(A)$, car $\mu(\overset{\circ}{A}) \leq \mu(A) \leq \mu(\overline{A})$.

– Preuve de (5) $\Longrightarrow$ (1) L'idée est d'approcher la fonction f par une somme d'indicatrice d'ensembles dont la frontière est de mesure nulle. On va commencer par exhiber un grand ensemble de mesure nulle.
Soit T une base de $\mathbb{R}$ vu comme un $\mathbb{Q}$-espace vectoriel. Pour $t \in T$ et $k \in \{1,\dots,d\}$, notons

$$\mathcal{P}_t^k = \{x \in \mathbb{R}^d; x_k - t \in \mathbb{Q}\}.$$

$\mathcal{P}_t^k$ est une réunion d'hyperplans orthogonaux au k-ième vecteur de base. À k fixé, les ensembles $(\mathcal{P}_t^k)_{t\in[0,1[}$ sont disjoints. Comme μ est une probabilité, l'ensemble des $t \in [0,1[$ tels que $\mu(\mathcal{P}_t^k) > 0$ est au plus dénombrable. Comme T n'est pas dénombrable, il existe t_k tel que $\mu(\mathcal{P}_{t_k}^k) = 0$. Pour A entier, on pose $B_p = \{x \in \mathbb{R}^d; \|x - t\|_\infty \leq A$. Soit $\varepsilon > 0$. On peut trouver A tel que $\mu(B_A) \geq 1 - \frac{\varepsilon}{2}$. D'après ce qui précède, $\mu(\partial B_A) = 0$. On peut donc trouver n_0 tel que

$$n \geq n_0 \Longrightarrow \mu_n(B_A) \geq 1 - \varepsilon.$$

B_A est compact. On note ω_f le module de continuité de la restriction de f à B_A :
$\omega_f(\eta) = \sup(|f(x) - f(y); \|x-y\|_\infty \le \eta)$. Pour $p \ge 1$, on pose

$$f_p(x) = 1_{B_A}(x)\|f(t_1 + \frac{\mathbf{Ent}(p(x-t_1))}{p}, \dots, t_d + \frac{\mathbf{Ent}(p(x-t_d))}{p})$$

On a

$$|f_p(x) - f(x)| \le \omega_f(1/p) + 2\|f\|_\infty \mathbb{1}_{B_A^c}(x)$$

Ainsi

$$\forall p \in \mathbb{N} \quad \int_{\mathbb{R}^d} |f_p(x) - f(x)| \le \omega_f(1/p) + 2\|f\|_\infty \frac{\varepsilon}{2}$$

$$\forall p \in \mathbb{N} \quad \overline{\lim_{n\to\infty}} \int_{\mathbb{R}^d} |f_p(x) - f(x)| \le \omega_f(1/p) + 2\|f\|_\infty \varepsilon$$

Soit p_0 un entier tel que $\omega_f(1/p_0) \le \varepsilon$: on a

$$\int_{\mathbb{R}^d} |f_{p_0}(x) - f(x)| \; d\mu \le \varepsilon + 2\|f\|_\infty \frac{\varepsilon}{2}$$

et

$$\overline{\lim_{n\to\infty}} \int_{\mathbb{R}^d} |f_{p_0}(x) - f(x)| \; d\mu_n \le \varepsilon + 2\|f\|_\infty \varepsilon.$$

On a

$$f_{p_0} = \sum_{x \in (t + \frac{1}{p_0}\mathbb{Z}^d) \cap B_A} f(x) \mathbb{1}_{x + [0, \frac{1}{p_0}[^d}$$

Ainsi, f_p s'écrit comme une combinaison linéaire (finie !) d'indicatrices de parties de $\mathbb{R}^d$ dont la frontière est de μ-mesure nulle. On a donc

$$\lim_{n\to+\infty} \int f_p \; d\mu_n = \int f_p \; d\mu.$$

Comme

$$\int f \; d\mu - \int f \; d\mu_n = (\int f \; d\mu - \int f_p \; d\mu) + (\int f_p \; d\mu - \int f_p \; d\mu_n) + (\int f_p \; d\mu_n - \int f \; d\mu_n),$$

on a

$$\begin{aligned} \overline{\lim_{n\to\infty}} \; |\textstyle\int_{\mathbb{R}^d} f(x) \; d\mu(x) - \int_{\mathbb{R}^d} f(x) \; d\mu_n(x)| &\le \textstyle\int_{\mathbb{R}^d} |f(x) - f_{p_0}(x)| \; d\mu(x) \\ &\quad + \overline{\lim_{n\to\infty}} \; |\textstyle\int_{\mathbb{R}^d} f_{p_0}(x) \; d\mu(x) - \int_{\mathbb{R}^d} f_{p_0}(x) \; d\mu_n(x)| \\ &\quad + \overline{\lim_{n\to\infty}} \; \textstyle\int_{\mathbb{R}^d} |f_{p_0}(x) - f(x)| \; d\mu_n(x) \\ &\le (\varepsilon + \|f\|_\infty \varepsilon) + 0 + \varepsilon + (2\|f\|_\infty \varepsilon) \\ &\le \varepsilon(2 + 3\|f\|_\infty). \end{aligned}$$

Comme ε est arbitraire, on en déduit que $\int_{\mathbb{R}^d} f(x)\ d\mu_n(x)$ converge vers $\int_{\mathbb{R}^d} f(x)\ d\mu(x)$, et cela pour toute fonction continue bornée f : (1) est donc vérifié.

□

Corollaire 17. *Soit X_n une suite de variables aléatoires réelles convergeant en loi vers une variable aléatoire X. Alors pour tout point x où F_X est continue, $F_{X_n}(x)$ tend vers $F(x)$ lorsque n tend vers l'infini.*

Cette dernière conséquence de la convergence en loi est très utile, par exemple en statistique.

Démonstration. Posons $A =]-\infty, x]$. A est un borélien dont la frontière $\{a\}$ est telle que $P_X(\{a\}) = 0$, car a est un point de continuité de F_X. Comme $P_{X_n}(F) = F_{X_n}(x)$ et $P_X(F) = F_X(x)$, il suffit donc d'appliquer le théorème du porte-manteau pour conclure. □

Le théorème suivant est également très utile :

Théorème 75. *Si une suite $(\mu_n)_{n\geq 1}$ de mesures de probabilités sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ est telle que pour toute fonction f continue positive à support compact de $\mathbb{R}^d$ dans $\mathbb{R}$, on a*

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} f\ d\mu_n = \int_{\mathbb{R}^d} f\ d\mu,$$

alors μ_n converge faiblement vers μ.

Démonstration. Soit $\varepsilon > 0$. Notons G_A la fonction continue valant 1 sur $]-\infty, A/2]$, 0 sur $[A, +\infty[$, et affine sur $[A/2, A]$ et $H_A = G_A \circ \|.\|$. On a

$$\int_{\mathbb{R}^d} H_A\ d\mu \geq \mu(B(0, A/2)),$$

de sorte que si l'on prend A suffisamment grand, on a

$$\int_{\mathbb{R}^d} G_A(\|x\|)\ d\mu \geq 1 - \varepsilon.$$

Comme H_A est continue positive à support compact, $\int_{\mathbb{R}^d} H_A\ d\mu_n$ converge vers $\int_{\mathbb{R}^d} H_A\ d\mu$. Ainsi, il existe n_0 tel que

$$n \geq n_0 \Longrightarrow \int_{\mathbb{R}^d} G_A(\|x\|)\ d\mu_n \geq 1 - 2\varepsilon,$$

d'où

$$n \geq n_0 \Longrightarrow \mu_n(B(0, A)) \geq 1 - 2\varepsilon.$$

Soit maintenant f une fonction continue positive. On a

$$\int f\,d\mu_n - \int f\,d\mu_n = (\int fH_{2A}\,d\mu_n - \int fH_{2A}\,d\mu_n) + (\int f(1-H_{2A})\,d\mu_n - \int f(1-H_{2A})\,d\mu_n)$$

Ainsi

$$|\int f\ d\mu_n - \int f\ d\mu_n| \le (\int fH_{2A}\ d\mu_n - \int fH_{2A}\ d\mu_n) + \varepsilon\|f\|_\infty$$

Comme fH_{2A} est une fonction continue positive à support compact, $\int fH_{2A}\ d\mu_n$ converge vers $\int fH_{2A}\ d\mu$, d'où

$$\overline{\lim_{n\to+\infty}}\ |\int f\ d\mu_n - \int f\ d\mu_n| \le \varepsilon\|f\|_\infty.$$

Comme ε est arbitraire, on en déduit que $\int f\ d\mu_n$ converge vers $\int f\ d\mu$.

Le passage à une fonction continue bornée de signe quelconque ne pose pas de problème, car $f = \max(f,0) - \max(-f,0)$ et le résultat s'ensuit par linéarité. □

7.1.4 Lien avec les autres modes de convergence

Théorème 76. *Soit $(X_n)_{n\ge1}$ une suite de variables aléatoires, X une variable aléatoire*

1. *Si X_n converge en probabilité vers X, alors X_n converge en loi vers X.*
2. *Si X_n converge en loi vers une constante a (c'est à dire vers une masse de Dirac δ_a, alors X_n converge en probabilité vers a.*

Démonstration. 1. Soit f une fonction continue bornée. On pose $x_n = \mathbb{E}f(X_n)$. La suite $(x_n)_{n\ge1}$ est bornée. Soit a une valeur d'adhérence de $(x_n)_{n\ge1}$, avec $\lim_{k\to+\infty} x_{n_k} = a$. Comme X_{n_k} converge en probabilité vers X, on peut (d'après le théorème 56) en extraire une sous-suite $X_{n_{m_k}}$ telle que $\lim_{k\to+\infty} X_{n_{m_k}} = X$ presque sûrement. Par continuité $\lim_{k\to+\infty} f(X_{n_{m_k}}) = f(X)$ presque sûrement. D'après le théorème de convergence dominée, on a donc $\lim_{k\to+\infty} \mathbb{E}f(X_{n_{m_k}}) = \mathbb{E}f(X)$, c'est à dire $\lim_{k\to+\infty} x_{n_{m_k}} = \mathbb{E}f(X)$. Or $x_{n_{m_k}}$ est une sous-suite de (x_{n_k}) qui converge elle-même vers a, donc $a = \mathbb{E}f(X)$. $(x_n)_{n\ge1}$ est une suite bornée dont $\mathbb{E}f(X)$ est l'unique valeur d'adhérence, donc elle converge vers $\mathbb{E}f(X)$. Comme pour toute fonction continue bornée f, $\mathbb{E}f(X_n)$ converge vers $\mathbb{E}f(X)$, on peut dire que $(X_n)_{n\geq 1}$ converge en loi vers X.

2. Soit $\varepsilon > 0$. On pose $F = \{x \in \mathbb{R}^d ; \|a - x| \geq \varepsilon\}$. Ainsi $P(\|X_n - a\| \geq \varepsilon) = P_{X_n}(F)$. F est fermé et P_{X_n} converge faiblement vers δ_a, donc, d'après le théorème du porte-manteau,

$$\overline{\lim_{n\to\infty}}\ P_{X_n}(F) \leq \delta_a(F) = 0.$$

Ainsi, pour tout $\varepsilon > 0$, $P(\|X_n - a\| \geq \varepsilon)$ tend vers 0 quand n tend vers l'infini : on peut dire que $(X_n)_{n\geq 1}$ converge en probabilité vers a.

□

7.2 Convergence en loi sur $\mathbb{R}^n$ grâce aux fonctions caractéristiques

7.2.1 Critère de convergence

Théorème 77 (Lévy). *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures et μ une mesure donnée sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Alors la suite $(\mu_n)_{n\geq 1}$ converge faiblement vers μ si et seulement si*

$$\forall t \in \mathbb{R}^d \quad \lim_{n\to+\infty} \varphi_{\mu_n}(t) = \varphi_\mu(t)$$

CE THÉROÈME EST ADMIS.

7.2.2 Théorème de continuité de Lévy

Théorème 78 (théorème de continuité de Lévy). *Soit $(\mu_n)_{n\geq 1}$ une suite de mesures, φ une fonction donnée. Si*

$$\forall t \in \mathbb{R}^d \quad \lim_{n\to+\infty} \varphi_{\mu_n}(t) = \varphi(t)$$

et que φ est continue en 0, alors, il existe une unique mesure μ telle que $\varphi = \varphi_\mu$ et la suite $(\mu_n)_{n\geq 1}$ converge faiblement vers μ.

Démonstration. Pour montrer l'existence d'une mesure μ telle que $\varphi = \varphi_\mu$, nous allons utiliser le théorème de Bochner. Comme φ_{μ_n} est une fonction caractéristique $\varphi_{\mu_n}(0) = 1$, d'où à la limite $\varphi(0) = 1$. Par hypothèse, φ est continue en 0. Il reste donc à montrer que φ est de type positif. Soient $e_1, \dots, e_p$

des éléments quelconques de $\mathbb{R}^d$; on va démontrer que la matrice A de taille $d\times d$ définie par $a_{k,l} = \varphi(e_k - e_l)$ est hermitienne positive. Pour $n \geq 1$, on note A_n la matrice de taille $d \times d$ définie par $a^n_{k,l} = \varphi_{\mu_n}(e_k - e_l)$. Soit $x \in \mathbb{C}^d$. On a $\langle A_n x, x\rangle \geq 0$, car A_n est hermitienne positive. Lorsque n tend vers l'infini A_n converge vers A car φ_{μ_n} converge vers φ. Ainsi $\langle Ax, x\rangle = \lim_{n\to\infty} \langle A_n x, x\rangle \geq 0$, donc A est hermitienne positive. Ainsi, d'après le théorème de Bochner, il existe une unique mesure μ telle que $\varphi = \varphi_\mu$ et le premier théorème de Lévy permet d'affirmer que μ_n converge faiblement vers μ. □

Ce dernier théorème peut être intéressant si la loi limite est une loi nouvelle, inconnue. L'appliquer lorsque la loi est une loi bien connue est assez maladroit.

7.2.3 Une application du théorème de Lévy

Le résultat qui suit peut être démontré sans l'aide du théorème de Lévy, mais ce dernier théorème en rend la preuve particulièrement simple.

Théorème 79. *Si μ_n tend faiblement vers μ et ν_n faiblement vers ν, alors la suite $(\mu_n \otimes \nu_n)_{n\geq 1}$ tend faiblement vers $\mu \otimes \nu$.*

Démonstration. Soit $(s,t) \in \mathbb{R}^d \times \mathbb{R}^d$. On a $\varphi_{\mu_n\otimes\nu_n}(s,t) = \varphi_{\mu_n}(s)\varphi_{\nu_n}(t)$. Comme μ_n converge faiblement vers μ, $\varphi_{\mu_n}(s)$ converge vers $\varphi_\mu(s)$. De même, $\varphi_{\nu_n}(s)$ converge vers $\varphi_\nu(t)$. Ainsi $\varphi_{\mu_n\otimes\nu_n}(s,t)$ converge vers $\varphi_\mu(s)\varphi_\nu(t) = \varphi_{\mu\otimes\nu}(s,t)$, donc d'après le théorème de Lévy, la suite $(\mu_n \otimes \nu_n)_{n\geq 1}$ tend faiblement vers $\mu \otimes \nu$. □

Théorème 80. *Si μ_n tend faiblement vers μ et ν_n faiblement vers ν, alors la suite $(\mu_n * \nu_n)_{n\geq 1}$ tend faiblement vers $\mu * \nu$.*

Démonstration. Soient (X_n, Y_n) de loi $\mu_n \otimes \nu_n$, (X, Y) de loi $\mu \otimes \nu$. D'après le théorème 79, (X_n, Y_n) converge en loi vers (X, Y), donc d'après le corollaire 14, $X_n + Y_n$ converge en loi vers $X + Y$. Mais la loi de $X_n + Y_n$ est $\mu_n * \nu_n$ et la loi de $X + Y$ est $\mu\nu$, donc le résultat est démontré. □

7.3 Théorème de la limite centrale

7.3.1 Théorème de la limite centrale en dimension 1

En dimension 1, le théorème s'énonce comme suit :

Théorème 81. *Soit $X_1, \dots, X_n$ une suite de variables aléatoires réelles i.i.d. admettant un moment d'ordre 2. On note alors m l'espérance et σ^2 la variance communes à ces variables. Alors*

$$\frac{(X_1 + \dots X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, \sigma^2).$$

Démonstration. On pose $S_n = (X_1 + \dots X_n) - nm = \sum_{k=1}^n (X_k - m)$. Notons φ la fonction caractéristique de $X_1 - m$. Comme les variables aléatoires $X_1 - m, \dots, X_n - m$ sont indépendantes, la fonction caractéristique de $S_n/\sqrt{n}$ vaut

$$\varphi_{S_n/\sqrt{n}}(t) = \varphi_{S_n}(\frac{t}{\sqrt{n}}) = \prod_{k=1}^n \varphi_{X_k - m}(\frac{t}{\sqrt{n}}) = \varphi(\frac{t}{\sqrt{n}})^n.$$

D'après le théorème de Lévy, pour montrer que $S_n/\sqrt{n}$ converge en loi vers $\mathcal{N}(0, \sigma^2)$, il suffit de montrer que

$$\forall t \in \mathbb{R} \qquad \lim_{n \to +\infty} \varphi(\frac{t}{\sqrt{n}})^n = \exp(-\frac{\sigma^2}{2} t^2),$$

car $t \mapsto \exp(-\frac{\sigma^2}{2} t^2)$ est la fonction caractéristique de la loi $\mathcal{N}(0, \sigma^2)$.

Pour cela, on va utiliser le développement limité

$$\varphi(x) = 1 - \frac{\sigma^2}{2} x^2 + o(x^2) \tag{7.1}$$

établi au corollaire 11 (voir fonction caractéristique et moments).

L'introduction du logarithme complexe peut être évitée en remarquant que pour des nombres complexes z et u de module inférieur ou égal à 1, on a

$$\forall n \in \mathbb{N} \quad |z^n - u^n| = |(z - u)(\sum_{k=0}^{n-1} z^k u^{n-k})| \le n|z - u|.$$

Il s'ensuit que

$$\begin{aligned} |\varphi(\frac{t}{\sqrt{n}})^n - \exp(-\frac{\sigma^2}{2} t^2)| &= |\varphi(\frac{t}{\sqrt{n}})^n - \exp(-\frac{\sigma^2}{2n} t^2)^n| \\ &\le n|\varphi(\frac{t}{\sqrt{n}}) - \exp(-\frac{\sigma^2}{2n} t^2)|, \end{aligned}$$

Or, on a d'une part $\exp(-\frac{\sigma^2}{2n} t^2) = 1 - \frac{\sigma^2}{2n} t^2 + o(1/n)$, et d'autre part d'après l'équation (7.1) $\varphi(\frac{t}{\sqrt{n}})) = 1 - \frac{\sigma^2}{2n} t^2 + o(1/n)$. Ainsi $n|\varphi(\frac{t}{\sqrt{n}}) - \exp(-\frac{\sigma^2}{2n} t^2)| = o(1)$, ce qui achève la preuve.

□

7.3.2 Théorème de la limite centrale en dimension d

Théorème 82. *Soit $X_1, \dots, X_n$ une suite i.i.d. de vecteurs aléatoires à valeurs dans $\mathbb{R}^p$. On suppose que $\mathbb{E}\|X_1\|^2 < +\infty$. On note alors m l'espérance et C la matrice des covariances. Alors*

$$\frac{(X_1 + \dots X_n) - nm}{\sqrt{n}} \Longrightarrow \mathcal{N}(0, C).$$

En fait, le théorème 82 peut être vu comme une conséquence du théorème 81. On va pour celà énoncer un lemme comparable dans son esprit au theorème 40.

Lemme 5. *Soit $(X_n)_{n\geq 1}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ et X un vecteur aléatoire à valeurs dans $\mathbb{R}^d$. Si pour tout $a \in \mathbb{R}^d$, $\langle X_n, a\rangle$ converge en loi vers $\langle X, a\rangle$, alors X_n converge en loi vers X.*

Démonstration. Soit $a \in \mathbb{R}^d$. On pose $Y_n = \langle X_n, a\rangle$. On a

$$\varphi_{X_n}(a) = \mathbb{E}e^{i\langle X_n, a\rangle} = \mathbb{E}e^{iY_n} = \varphi_{Y_n}(1)$$

Par hypothèse, Y_n converge en loi vers $\langle X, a\rangle$, donc $\varphi_{Y_n}(1)$ converge vers $\varphi_{\langle X,a\rangle}(1) = \mathbb{E}e^{i\langle X,a\rangle} = \varphi_X(a)$. Donc $\varphi_{X_n}(a)$ converge vers $\varphi_X(a)$. Comme c'est vrai pour tout $a \in \mathbb{R}^d$, le théorème de Lévy nous dit que $(X_n)_{n\geq 1}$ converge en loi vers X. □

On peut maintenant prouver le théorème 82.

Démonstration. Soit X un vecteur aléatoire suivant $\mathcal{N}(0, C)$. On pose

$$S_n = (X_1 + \dots X_n) - nm = \sum_{k=1}^{n} (X_k - m).$$

D'après le lemme précédent, il nous suffit de montrer que pour tout $a \in \mathbb{R}^d$ $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge en loi vers $\langle X, a\rangle$.

Fixons $a \in \mathbb{R}^d$ et posons $Y_n = \langle X_n - m, a\rangle = a^*(X_n - m)$. Les Y_n sont des variables aléatoires indépendantes qui suivent toutes la même loi, la loi image de P_{X_1} par $x \mapsto \langle x - m, a\rangle = a^*(x - m)$. D'après le théorème 30, leur espérance est 0 et leur variance $a^*Ca = \langle Ca, a\rangle$. Ainsi, d'après le théorème 81, la suite $\frac{\sum_{k=1}^n Y_k}{\sqrt{n}}$ converge en loi vers $\mathcal{N}(0, \langle Ca, a\rangle)$.

Mais

$$\frac{\sum_{k=1}^n Y_k}{\sqrt{n}} = \langle \frac{S_n}{\sqrt{n}}, a\rangle,$$

et, encore d'après le théorème 30, la loi de $\langle X, a\rangle$ est précisément $\mathcal{N}(0, \langle Ca, a\rangle)$. $\langle \frac{S_n}{\sqrt{n}}, a\rangle$ converge donc en loi vers $\langle X, a\rangle$, ce qui achève la preuve.

□

7.4 Exercices sur la convergence en loi

1. Soit $(X_\lambda)_{\lambda>0}$ une famille de variables aléatoires telles que pour tout $\lambda > 0$, X_λ suive une loi de Poisson $\mathcal{P}(\lambda)$. Montrer que la suite
$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}}$$
converge faiblement vers la loi $\mathcal{N}(0,1)$ lorsque λ tend vers l'infini.

2. (a) Soit $(U_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi uniforme sur $[0,1]$. Pour $x > 0$, on pose
$$X_n^x = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq \frac{x}{n}\}}.$$
Montrer que $(X_n)_{n\geq 1}$ converge en loi et déterminer la loi limite.

   (b) Soit $(u_n)_{n\geq 0}$ une suite croissante, de limite $\ell \in \mathbb{R}$. Pour $x > 0$, on pose
$$f(x) = \sum_{k=0}^{+\infty} e^{-x}\frac{x^k}{k!}u_k.$$
Montrer que f est une fonction croissante et déterminer sa limite en $+\infty$.

3. *Une suite convergente est tendue*
Soit $(\mu_n)_{n\geq 1}$ une suite de lois de probabilités convergeant faiblement vers la loi de probabilité μ. Montrer que la suite $(\mu_n)_{n\geq 1}$ est tendue, c'est à dire que pour tout $\varepsilon > 0$, on peut trouver un compact K_ε tel que
$$\forall n \in \mathbb{N}^* \quad \mu_n(K) \geq 1 - \varepsilon.$$

4. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires convergeant en loi vers une variable aléatoire X, des suite de réels $(a_n)_{n\geq 1}$ et $(b_n)_{n\geq 1}$ convergeant respectivement vers les réels a et b. Montrer que la suite de variables aléatoires $(a_nX_n+b_n)_{n\geq 1}$ converge en loi vers la variable aléatoire $aX + b$.

5. Interpréter la quantité $e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$ comme une probabilité. À l'aide du théorème de la limite centrale, démontrer la relation
$$\lim_{n\to+\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \frac{1}{2}.$$

6. Jean joue au jeu suivant : sur chaque case d'un plateau carré de taille $n \times n$, il dispose une pièce de 1 €, les côtés visibles étant choisies au hasard (c'est à dire avec équiprobabilité), de manière indépendante. Ensuite, il tire au hasard un nombre X entre 1 et n. Deux possibilités s'offrent alors à lui
   – soit retourner tous les pions de la colonne X,
   – soit retourner tous les pions de la ligne X.
   Son but est de maximiser le nombre de faces. On suppose que Jean agit intelligemment ; on note alors F_n (resp P_n) le nombre de faces (resp de piles) dans la configuration alors obtenue. On pose $D_n = F_n - P_n$.
   Montrer que
$$\mathbb{E}D_n \underset{n\to\infty}{\sim} \sqrt{\frac{4n}{\pi}},$$
   mais que
$$\lim_{n\to\infty} P(D_n \geq 0) = \frac{1}{2}.$$
7. *Une preuve probabiliste de la formule de Stirling*
   Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes suivant la loi exponentielle de paramètre 1. On pose
$$S_n = \sum_{k=0}^{n} X_k.$$
   (a) Montrer que $\frac{S_n-n}{\sqrt{n}}$ converge en loi vers la loi $\mathcal{N}(0,1)$.
   (b) Montrer que S_n suit la loi $\Gamma(n+1,1)$. En déduire que la densité de $\frac{S_n-n}{\sqrt{n}}$ s'écrit $g_n(x) = a_n h_n(x)$, avec
$$a_n = \frac{n^{n+1/2}e^{-n}\sqrt{2\pi}}{\Gamma(n+1)}$$
   et
$$h_n(x) = a_n \mathbb{1}_{[-\sqrt{n},+\infty[}\frac{1}{\sqrt{2\pi}}e^{-\sqrt{n}x}(1+\frac{x}{\sqrt{n}})^n.$$
   (c) Montrer que
$$\lim_{n\to+\infty}\int_0^1 g_n(x)\ dx = \int_0^1 \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ dx$$
   (d) Soit Ψ la fonction définie sur $[-1/2, 1/2]$ par
$$\forall x \in [-1/2, +1/2] \quad \ln(1+x) = x - \frac{x^2}{2} + x^3\Psi(x).$$

Montrer que

$$x \geq -\sqrt{n} \Longrightarrow h_n(x) - \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} (e^{\frac{x^3}{n^{1/2}} \psi(\frac{x}{\sqrt{n}})} - 1)$$

En déduire que $h_n(x)$ converge uniformément vers $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ sur tout compact.

(e) Montrer que

$$\lim_{n \to +\infty} \int_0^1 h_n(x)\ dx = \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\ dx.$$

(f) En déduire la formule de Stirling :

$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n}.$$

Annexe A

Rappels de dénombrement

A.1 Rappels de vocabulaire ensembliste

Un ensemble Ω est constitué de points, tous distincts. On dit qu'un ensemble A est inclus dans Ω, et l'on écrit $A \subset \Omega$ lorsque tous les éléments de A appartiennent à Ω.

On rappelle que l'ensemble vide (noté $\varnothing$) ne contient aucun élément et est inclus dans tous les ensembles

Pratiquement, si l'on veut montrer le résultat $A \subset \Omega$, la preuve ressemblera donc à "Soit $x \in A$...(raisonnement)...donc $x \in \Omega$. Comme on a choisi x quelconque dans A, on peut conclure que $A \subset \Omega$."

Si A est inclus dans Ω, on dit que A est un sous-ensemble, ou encore une partie de Ω.

Si A et B sont des parties de Ω, l'ensemble $A \cup B$ est constitué des éléments de Ω qui sont dans A ou dans B, éventuellement dans les deux. Plus généralement, si I est un ensemble quelconque et $(A_i)_{i \in I}$ une famille de parties de Ω indexé par I, $\underset{i \in I}{\cup} A_i$ est constitué des points de Ω qui sont dans au moins un des A_i.

Pratiquement, si l'on veut montrer le résultat $x \in \underset{i \in I}{\cup} A_i$, , la preuve ressemblera donc à "...(raisonnement)...Il existe donc $i_0 \in I$ tel que $x \in A_{i_0}$. Donc $x \in \underset{i \in I}{\cup} A_i$."

Si A et B sont des parties de Ω, l'ensemble $A \cap B$ est constitué des éléments de Ω qui sont dans A et dans B. Plus généralement, si I est un ensemble quelconque et $(A_i)_{i \in I}$ une famille de parties de Ω indexé par I, $\underset{i \in I}{\cap} A_i$ est constitué des points de Ω qui sont dans tous les A_i.

Pratiquement, si l'on veut montrer le résultat $x \in \bigcap_{i \in I} A_i$, , la preuve ressemblera donc à “Soit $i \in I$...(raisonnement)...Donc $x \in A_i$. Comme i est quelconque, on a donc $x \in \bigcap_{i \in I} A_i$.”

A.2 Applications et cardinaux : définitions et notations

Pour A, D deux ensembles non vides quelconques, on note A^D ou $\mathcal{F}(D, A)$ l'ensemble des fonctions de D (ensemble de départ) vers A (ensemble d'arrivée). Soit f une application de A dans D. On dit que f est
- *injective* si $\forall x, y \in D \quad x \neq y \Longrightarrow f(x) \neq f(y)$
- *surjective* si $\forall z \in A \quad \exists x \in D \quad f(x) = z$.
- *bijective* si elle elle à fois injective et surjective.

Une application injective (resp. surjective, bijective) est une injection (resp. surjection, bijection).

Une bijection d'un ensemble Ω dans lui-même est appelée *permutation* de Ω.

Un ensemble est Ω dit *fini* si
- ou bien c'est l'ensemble vide. ($\varnothing$)
- ou bien il existe un entier n tel qu'il existe une bijection entre Ω et $\{1, \dots, n\}$.

Cet entier n est unique : on l'appelle le *cardinal* de l'ensemble Ω. On le note $|\Omega|$. De manière intuitive, c'est le nombre d'éléments de Ω.

Le cardinal de l'ensemble vide est zéro.

Pour Ω fini de cardinal n, et $p \in \{0, \dots, n\}$, on note $\mathcal{B}_p(\Omega)$ l'ensemble des parties de Ω de cardinal p. Par exemple $\mathcal{B}_2(\{a, b, c\}) = \{\{a, b\}, \{b, c\}, \{a, c\}\}$. De même, on note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω, quelque cardinal qu'elles aient. Par exemple $\mathcal{P}(\{a, b, c\}) = \{\varnothing, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$.

Soit A et D deux ensembles finis. On admettra les résultats suivants :
- Il existe (au moins) une bijection de D dans A si et seulement si $|A| = |D|$.
- Il existe (au moins) une injection de D dans A si et seulement si $|A| \geq |D|$.
- Il existe (au moins) une surjection de D dans A si et seulement si $|A| \leq |D|$.

Le premier des trois résultats énoncés est évidemment le plus utilisé lorsque l'on veut des dénombrements exacts, alors que les deux autres sont plutôt utilisés dans les cas trop complexes où l'on peut juste espérer des encadrements.

A.3 Principes de base du dénombrement

A.3.1 Principe de bijection

Dans la pratique, lorsque l'on veut compter les éléments d'un ensemble, on montre que cet ensemble est en bijection avec un ensemble dont on connaît (par coeur !) le nombre d'éléments. La section suivante énoncera un certain nombre de résultats qu'il faut connaître.

A.3.2 Principe d'indépendance

C'est juste la formule

$$|A \times B| = |A|.|B| \quad .$$

Considérée isolément, elle peut paraître sans intérêt mais elle est souvent utilisée en association avec le principe de bijection.

A.3.3 Principe de partition

On dit que les ensembles $(A_i)_{i\in I}$ forment une partition de A si l'on a $A = \underset{i\in I}{\cup} A_i$ et $i \neq j \Longrightarrow A_i \cap A_j = \varnothing$. On a alors

$$|A| = \sum_{i\in I} |A_i|.$$

Le résultat élémentaire suivant peut souvent être utile.

Théorème 83. *Soit Ω un ensemble quelconque, I un ensemble d'index fini ou dénombrable et $(\Omega_i)_{i\in I}$ une partition de Ω. Alors, si l'on pose $A_i = A\cap\Omega_i$, les ensembles $(A_i)_{i\in I}$ forment une partition de A.*

Démonstration. Comme $\Omega = \underset{i\in I}{\cup} \Omega_i$, on a $A = A\cap\Omega = \underset{i\in I}{\cup} A\cap\Omega_i = \underset{i\in I}{\cup} A_i$. D'autre part, pour $i \neq j$, on a $A_i \cap A_j \subset \Omega_i \cap \Omega_j = \varnothing$, d'où $A_i \cap A_j = \varnothing$. $\square$

A.3.4 Lemme des bergers

Le lemme suivant peut également être utile

Lemme 6 (Lemme des bergers). *Soit φ une application* surjective *de A dans D. On suppose qu'il existe un entier $a \geq 1$ tel que*

$$\forall y \in A \quad |\{x \in D ; \varphi(x) = y\}| = a$$

(autrement dit si tout élément de A a exactement a antécédents), on a

$$|A| = \frac{|D|}{a}.$$

Démonstration. On applique le principe de partition avec $I = A$: si l'on pose, pour $y \in A$: $D_y = \{x \in D ; \varphi(x) = y\}$, les D_y forment clairement une partition de D, d'où

$$|D| = \sum_{y \in A} |D_y| = \sum_{y \in A} a = |A|a.$$

□

Le nom du lemme est du à la procédure prétendûment employée par les bergers chaldéens pour compter le nombre de leurs moutons : il s'agit de compter le nombre de pattes et de diviser par 4. Dans cet exemple A est l'ensemble des moutons, D l'ensemble des pattes de mouton, et φ l'application qui a une patte associe le mouton auquel elle appartient.

A.4 Quelques résultats incontournables

A.4.1 Nombre d'applications de D dans A

Il existe exactement $|A|^{|D|}$ applications de D dans A, ce qui peut s'écrire

$$|A^D| = |A|^{|D|}.$$

On pose $|A| = n$ et $|D| = p$. Un cas particulier important est celui où l'on a $D = \{1, \ldots, p\}$. Or, un p-uplet $(x_1, x_2, \ldots, x_p)$ dont les composantes sont des éléments de A peut être considéré comme la donnée d'une application de $\{1, \ldots, p\}$ dans A. Le nombre de p-uplets $(x_1, x_2, \ldots, x_p)$ dont les composantes sont des éléments de A est donc n^p.

Exemple : Un professeur note chaque étudiant d'une classe de 30 étudiants par une note entière de 0 à 20. Le nombre de résultats possibles est le nombre de fonctions de l'ensemble D des étudiants dans l'ensemble $A = \{0, \dots, 20\}$ des notes possibles. Comme $|A| = 21$ et $|D| = 30$, il y a donc 21^{30} résultats possibles.

Remarque : : au lycée, vous avez vu ce résultat sous la dénomination "choix indépendant (avec remise) de p objets dans un ensemble de cardinal $|A| = n$."

A.4.2 Nombre de permutations de Ω

On pose $|\Omega| = n$. Le nombre de permutations de Ω est

$$n(n-1)\dots 1.$$

On note $n! = n(n-1)\dots 1$.

Remarque : $n!$ se lit "*factorielle n*" .

Exemple : Un professeur doit faire passer dans la journée 5 étudiants à l'oral de contrôle. Il a 5 !=120 manières de choisir l'ordre dans lequel il va les interroger.

A.4.3 Nombre d'injections de D dans A

On pose $|A| = n$ et $|D| = p$. En vertu de la remarque faite en A.2, il existe une injection de D si et seulement si $p \le n$. Alors, le nombre d'injections de A est

$$n(n-1)\dots(n-p+1).$$

Démonstration. Soit n un entier. On pose $A = \{1, \dots, n\}$ et on note $\mathcal{I}_p$ l'ensemble des injections de $\{1, \dots, p\}$ dans A. On va montrer par récurrence sur $p \in \{1, \dots, n\}$ que $|\mathcal{I}_p| = \frac{n!}{(n-p)!}$. Il est évident que $|\mathcal{I}_1| = 1 = \frac{n!}{(n-1)!}$. Considérons l'application

$$R_p : \mathcal{I}_{p+1} \to \mathcal{I}_p$$

qui à chaque injection de $\{1, \dots, p+1\}$ dans A associe sa restriction à $\{1, \dots, p\}$. Un peu de réflexion montre que

$$\forall f \in \mathcal{I}_p \quad |\{g \in \mathcal{I}_{p+1}; R_p(g) = f\}| = n - p$$

D'après le lemme des bergers, on a donc

$$|\mathcal{I}_{p+1}| = (n-p)|\mathcal{I}_p|.$$

Cette identité permet d'achever la preuve par récurrence. □

Remarques :

– Comme on l'a vu dans la preuve, ce nombre peut s'écrire aussi $\frac{n!}{(n-p)!}$.
– Lorsque $n = p$, on trouve $n!$. En fait, une injection entre deux ensembles de même cardinaux est une bijection.

Exemple : 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises aux concours. Combien y a-t-il de palmarès possibles, en supposant qu'il n'y ait pas d'ex-æquos ?
Réponse : $3500 \times 3499 \times \cdots \times 3202 \times 3201$. Ici D est l'ensemble des rangs, on a donc $D = \{1, \ldots, 300\}$ et A l'ensemble des candidats (donc $|A| = 300$). On compte bien le nombre d'applications injectives puisqu'une même personne ne peut avoir deux rangs différents.

A.4.4 Nombre de parties de Ω possédant p éléments

On pose $|\Omega| = n$. Par définition, on note $\binom{n}{p}$ le nombre de parties à p éléments d'un ensemble de n éléments. Il s'agit donc de calculer $|\mathcal{B}_p(\Omega)|$. On va montrer que

$$\binom{n}{p} = \frac{n(n-1)\ldots(n-p+1)}{p(p-1)\ldots 1} = \frac{n!}{p!(n-p)!}.$$

Démonstration. Il suffit d'appliquer le lemme des bergers à
– D : ensemble des injections de $\{1, \ldots, p\}$ dans Ω
– $A = \mathcal{B}_p(\Omega)$
– φ définie par $\varphi(f) = \{f(k); k \in \{1, \ldots, p\}\}$

On a vu précédemment que $|A| = n(n-1)\ldots(n-p+1)$. Il n'est pas difficile de voir que φ est surjective. Une partie $\{e_1, \ldots, e_p\}$ de Ω étant donnée, combien existe-t-il d'injections (en fait de bijections) de $\{1, \ldots, p\}$ dans Ω telles que $\{f(1), \ldots, f(p)\} = \{e_1, \ldots, e_p\}$? C'est évidemment le nombre d'injections de $\{1, \ldots, n\}$ dans $\{e_1, \ldots, e_p\}$, c'est à dire $p!$. Le lemme des bergers s'applique donc avec $a = p!$, d'où le résultat. □

Exemple : 3500 personnes se présentent au concours de l'Agrégation de Mathématiques. 300 places sont mises aux concours. Combien y a-t-il de listes alphabétiques des reçus possibles ? Réponse : $\binom{3500}{300}$. Ici Ω est l'ensemble des candidats et $p = 300$ le nombre de reçus.

A.4.5 Nombre total de parties de Ω

Le nombre total de parties de Ω est $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$.

Démonstration. Il suffit de remarquer que l'application

$$\begin{array}{ccc} \mathcal{P}(\Omega) & \to & \{0;1\}^{\Omega} \\ A & \mapsto & \mathbb{1}_A \end{array}$$

est une bijection. On rappelle que pour $A \subset \Omega$, l'application $\mathbb{1}_A$ (appelée indicatrice de A) est définie sur Ω par

$$\mathbb{1}_A(x) = \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A \end{cases}$$

□

Exemple : 200 étudiants se présentent à un examen. Combien y a-t-il de listes alphabétique des reçus possibles ? Réponse : 2^{200}. Ici Ω est l'ensemble des candidats. La grande différence avec l' exemple précédent est qu'ici, le nombre de reçus n'est pas fixé à l'avance.

A.5 Équations et inéquations en entiers

Lemme 7. *Soit n, p entiers. Si $n \geq p$, il existe exactement $\binom{n}{p}$ applications strictement croissantes de $\{1, \dots, p\}$ dans $\{1, \dots, n\}$. Sinon, il n'en existe aucune.*

Démonstration. Une application strictement croissante étant injective, il est nécessaire que $n \geq p$. Mais se donner une suite strictement croissantes de p éléments pris dans $\{1, \dots, n\}$ revient à choisir une partie de $\{1, \dots, n\}$ possédant p éléments, puis à les ordonner dans l'ordre naturel. Or $\binom{n}{p}$ est, par définition, le nombre de parties de $\{1, \dots, n\}$ possédant p éléments, d'où le résultat. □

Exemple : Un enseignant devrait faire un cours de 70 pages en 7 séances. Combien y a-t-il de progressions possibles, en admettant qu'à chaque séance, l'enseignant progresse d' un nombre entier strictement positif de pages, mais sans être astreint à terminer le programme ?
Réponse : une progression correspond donc à une fonction strictement croissante de $\{1, \dots, 7\}$ dans $\{1, \dots, 70\}$ qui au numéro de chaque cours associe le numéro de la dernière page étudiée à ce cours : il y a donc $\binom{70}{7}$ progressions possibles.

Théorème 84. *Pour n, p entiers vérifiant $n \geq p$, il existe exactement $\binom{n}{p}$ p-uplets $(x_1, \dots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ solutions à l'inéquation :*

$$x_1 + x_2 + \cdots + x_p \leq n \tag{A.1}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \dots, x_p) \mapsto (x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + x_2 + \cdots + x_p)$$

réalise une bijection entre les solutions recherchées de l'inéquation et l'ensemble des suites croissantes de p éléments à valeurs dans $\{1, \ldots, n\}$. $\square$

Théorème 85. *Pour n, p entiers vérifiant $n \geq p$, il existe exactement $\binom{n-1}{p-1}$ p-uplets $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ solutions à l'équation :*

$$x_1 + x_2 + \cdots + x_p = n \tag{A.2}$$

Démonstration. Il suffit de remarquer que les solutions $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation (A.2) sont exactement les solutions de l'inéquation (A.2) qui ne sont pas solutions de l'inéquation

$$x_1 + x_2 + \cdots + x_p \leq n - 1 \tag{A.3}$$

Il y en a donc $\binom{n}{p} - \binom{n-1}{p} = \binom{n-1}{p-1}$.

$\square$

Théorème 86. *Pour n, p entiers positifs, il existe exactement $\binom{n+p-1}{p-1}$ p-uplets $(x_1, \ldots, x_p) \in \mathbb{N}^p$ solutions à l'équation :*

$$x_1 + x_2 + \cdots + x_p = n \tag{A.4}$$

Démonstration. Il suffit de remarquer que l'application

$$(x_1, \ldots, x_p) \quad \mapsto (x_1 + 1, \ldots, x_p + 1)$$

réalise une bijection entre les solutions $(x_1, \ldots, x_p) \in \mathbb{N}^p$ de l'équation (A.4) et les solutions $(x_1, \ldots, x_p) \in (\mathbb{N}\backslash\{0\})^p$ de l'équation

$$x_1 + x_2 + \cdots + x_p = n + p \tag{A.5}$$

$\square$

Exemple : 4 listes se présentent aux élections syndicales étudiantes où 9 sièges sont à pourvoir. Combien y a-t-il de répartitions des sièges possibles. Réponse : il s'agit de compter les solutions en entiers positifs ou nuls à l'équation

$$x_1 + x_2 + x_3 + x_4 = 9,$$

x_k représente le nombre d'élus de la liste k. Il y a donc $\binom{9+4-1}{4-1} = \binom{12}{3} = \frac{12\times11\times10}{1\times2\times3} = 220$ répartitions possibles.

Théorème 87. *Pour n, p entiers positifs, il existe exactement $\binom{n+p}{p}$ p-uplets $(x_1, \ldots, x_p) \in \mathbb{N}^p$ solutions à l'inéquation :*

$$x_1 + x_2 + \cdots + x_p \leq n \tag{A.6}$$

Démonstration. La preuve, analogue à celle du théorème précédent, est laissée en exercice. $\square$

A.6 Formule de Poincaré (aussi appelée formule du crible)

Cette formule est très utile en combinatoire, son application la plus classique étant le calcul du nombre de permutations sans point fixe (nombre de dérangements).

Pour tous ensembles $A_1, A_2, ..., A_n$, on a

$$
\begin{aligned}
|\bigcup_{i=1}^{n} A_i| &= \sum_{B\in\mathcal{P}(\{1,...,n\})\setminus\varnothing} (-1)^{1+\mathrm{Card}(B)} |\cap_{j\in B} A_j| && \text{(A.7)}\\
&= \sum_{i=1}^{n} |A_i| - \sum_{1\le i_1<i_2\le n} |A_{i_1}\cap A_{i_2}| + \cdots && \text{(A.8)}\\
&\quad \cdots + (-1)^{k+1} \sum_{1\le i_1<i_2<...<i_k\le n} |A_{i_1}\cap\cdots\cap A_{i_k}| + \cdots && \text{(A.9)}\\
&\quad \cdots + (-1)^{n+1} |A_1\cap\cdots\cap A_n|. && \text{(A.10)}
\end{aligned}
$$

Exemple : Pour $n = 3$, on a

$$|A_1\cup A_2\cup A_3| = |A_1|+|A_2|+|A_3|-|A_1\cap A_2|-|A_2\cap A_3|-|A_1\cap A_3|+|A_1\cap A_2\cap A_3|.$$

Démonstration. On pourrait prouver la formule par récurrence sur n, mais c'est plutôt lourd. On préférera une preuve probabiliste (voir plus loin). □

A.7 Exercices

1. Combien existe-t-il de mots de n lettres construits avec l'alphabet $\{a; b\}$ et ne comportant pas deux "a" consécutifs ?
   Indication : montrer que si u_n est le nombre de tels mots se terminant par "a" et v_n est le nombre de tels mots se terminant par "b", on a la récurrence

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_n \\ v_n \end{pmatrix}$$

2. On donne un ensemble Ω de 15 entiers compris entre 1000 et 2000. Montrer que l'on peut en extraire deux sous-ensembles disjoints non-vides A et B tels que la somme des éléments de A soit égale à la somme des éléments de B.

Indication : montrer que l'application

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \to \mathbb{N} & \\ A & \mapsto & \sum_{x\in A} x \end{array}$$

n'est pas injective.

Raffiner le raisonnement pour montrer que si l'on remplace 15 par 14, le résultat est encore vrai.

Indication : remplacer $\mathcal{P}(\Omega)$ par $\bigcup_{i=5}^{8} \mathcal{B}_i(\Omega)$.

3. Au tapis vert, il faut choisir 4 cartes (as,roi, dame, valet, dix, neuf, huit ou sept) par couleur (pique, coeur, carreau ou trèfle). On appelle grille l'un des résultats possibles.
   exemple de grille : as de pique, dame de cœur, 7 de carreau, 7 de trèfle.
   Dénombrer les grilles contenant :
   (a) la dame de cœur
   (b) une dame et une seulement
   (c) deux dames et deux seulement
   (d) aucune dame
   (e) au moins une dame
   (f) l'as de pique et une dame seulement
4. 5 prévenus sont amenés à choisir un avocat dans une liste de 10 avocats commis d'office.
   (a) Combien y a-t-il de choix possibles ?
   (b) Combien y a-t-il de choix tels que les 5 prévenus choisissent le même anocats ?
   (c) Combien y a-t-il de choix tels que 2 avocats soient appelés ?
5. De combien de manières peut-on disposer n personnes autour d'une table ronde ? Précision : seule la position relative à la maîtresse de maison importe, une rotation ne change donc rien.
   Si $n = 2k$, avec k hommes et k femmes, combien de dispositions permettent une alternance des sexes ?
6. Dans une assemblée de n personnes, on note la date d'anniversaire de chacun.
   (a) Donner le cardinal de toutes les répartitions possibles.
   (b) Calculer le cardinal des ensembles suivants :

   i. tous les participants ont la meme date d'anniversaire.
   ii. Pierre et Paul ont la meme date d'anniversaire.
   iii. deux personnes ont la meme date d'anniversaire.
   iv. toutes les personnes ont des dates d'anniversaires différentes.

7. On considère l'ensemble Ω des suites de n chiffres (les chiffres sont pris dans $0, 1, \ldots, 9$). Combien vaut $|\Omega|$? Combien y-a-il de chiffres comportant un nombre pair de zéros ?
8. Combien y a-t-il de nombres à trois chiffres (les chiffres sont pris dans $0, 1, \ldots, 9$) dont la somme des chiffres est égale à 22 ?
   Indication : noter
$$\Omega_{22} = \{(x_1, x_2, x_3) \in \mathbb{N}^3; x_1 + x_2 + x_3 = 22\},$$
$$B = \{(x_1, x_2, x_3) \in \{0, \ldots, 9\}^3; x_1 + x_2 + x_3 = 22\},$$
$$M_1 = \{x \in \Omega_{22}; x_1 \geq 10\},$$
$$M_2 = \{x \in \Omega_{22}; x_2 \geq 10\},$$
$$M_3 = \{x \in \Omega_{22}; x_3 \geq 10\}.$$
   Remarquer que $\Omega \backslash B = M_1 \cup M_2 \cup M_3$. Ensuite, si l'on définit Ω_{12} et Ω_2 de manière analogue à Ω_{22}, on pourra exhiber une bijection entre Ω_{12} et M_1 et entre Ω_2 et $M_1 \cap M_2$.

Annexe B

Indications

B.1 Exercices de théorie de la mesure

1. (a) Exprimer C en fonction de A_2 et A_5.
   (b) Exprimer B en fonction des $(A_p)_{p\in\mathcal{P}}$.
2. Remarquer qu'une union dénombrable d'événements est de probabilité nulle si et seulement si chacun est de probabilité nulle.
3. (a) Remarquer qu'aucun entier n'a une infinité de diviseurs premiers.
   (b) On pourra remarquer que

   $$(1-\frac{1}{p_i})^{-1} \geq \sum_{k=0}^{N} p_i^k,$$

   où N est choisi tel que $p_i^{N+1} > p_n$.
   (c) Utiliser la continuité séquentielle décroissante.
   (d) Utiliser le résultat de l'exercice précédent.
4.

B.2 Exercices sur les lois

1. Si X désigne le nombre de fois où l'on a obtenu le nombre choisi, le gain est $X - 1\!\!1_{X=0}$.
2. (a) Pensez à discuter suivant les positions relatives de n et k.
   (b) $\mathbb{Q}$ est dense dans $\mathbb{R}$.
   (c) Utiliser le principe de partition
   (d) Écrire $P(D) = 1$, avec D bien choisi.

   (e) Trivialiser la question !
3. Prendre X et Y deux variables indépendantes suivant chacune une loi de Bernoulli de paramètre 1/2 et poser $Z = |X - Y|$.
4. On prend $\Omega = \{0,1\}\times\{0,1\}$, $\mathcal{F} = \mathcal{P}(\omega)$ $\mathcal{C} = \{(0,0),(0,1)\},\{(0,0),(1,0)\}\}$. $P = \frac{1}{2}(\delta_0 + \delta_1) \otimes \frac{1}{2}(\delta_0 + \delta_1)$, $Q = \frac{1}{2}(\delta_{(0,0)} + \delta_{(1,1)})$
5. Traduire les événements considérés en fonction de Y
6. Poser $x = AM$ et résoudre l'inéquation.
7. Dire que le maximum de n nombres ne dépasse pas x, c'est dire que chacun ne dépasse pas x.
8. Remarquer que $1 - m_n = \max(1 - X_1, \dots, 1 - X_n)$.
9. À k fixé, il faut déterminer les valeurs de X qui sont telles que $Y = k$.
10. Représenter graphiquement les domaines considérés.
11. (a) On pourra montrer que X est $\mathcal{T}$-mesurable, puis que les A_p sont $\sigma(X)$-mesurables.
    (b) On montrera que l'ensemble cherché est $\{\omega \in \mathbb{N}; X(\omega) = 2 \times 3 \times 5 \times 11$.
    (c) Le sens direct est le plus simple. Pour la réciproque, remarquer que si A_n est $\mathcal{T}$-mesurable, il doit contenir $\{\omega \in \mathbb{N}^*; X(\omega) = X(n)\}$.
    (d) On montrera que $A_m \cap A_n = A_{m \wedge n}$.
    (e) Remarquer que pour p assez grand $1 - \frac{1}{p^{2s}} \geq e^{-\frac{2}{p^{2s}}}$.
12. (a) On pourra montrer que $\mathcal{Q}$ est la tribu de queue associée à la famille $(\mathcal{A}_n)_{n\geq 1}$.
    (b) Un ensemble d'entiers est infini si et seulement si il contient au moins un entier plus grand que n'importe quel entier fixé à l'avance.

    Ainsi, on pourra montrer que pour tout n_0, $A = \bigcap_{n\geq n_0} \bigcup_{k\geq n} A_k$.
13. Pour la première formule, faire une intégration par parties ; pour la deuxième, procéder par récurrence.
14. On pourra écrire $e^{-\frac{x^2}{2}} = \frac{1}{x}(xe^{-\frac{x^2}{2}})$ afin de procéder à une intégration par parties.

B.3 Exercice sur les espérances

1. Une probabilité est l'indicatrice d'une espérance.
2. Interpréter le membre de gauche comme la valeur absolue d'une covariance.

3. Effectuer une transformation d'Abel.
4. Appliquer le théorème de transfert, et penser à la forme canonique des polynômes du second degré.
5. Appliquer le théorème de transfert.
6. On pourra commencer par supposer que la loi est centrée (c'est à dire que $a + b = 0$) et faire une majoration simple). On s'y ramènera dans le cas général.
7. Fonction de répartition, ou transformation : vous avez le choix !
8. On peut raisonner en termes de loi.
9. S'inspirer de l'exercice précédent et utiliser une transformation.
10. Appliquer l'exercice précédent.
11. Bien observer que X et Y sont indépendantes.
12. Remarquer que tout se passe comme si (X, Y) suivait la loi uniforme sur $T = \{(x, y) \in \mathbb{R}^2; 0 < y < x < 1\}$.

    Si x et y sont solutions réelles de $x^2 - Sx + P = 0$, alors $|x - y| = \sqrt{S^2 - 4P}$
13. (a) Identifier la loi de S_n et appliquer le théorème de transfert.

    (b) i. Remarquer que $[0, 1]$ est compact.

    ii. Remarquer que

    $$B_n(x) - f(x) = \int_{|\frac{S_n}{n} - x| \leq \varepsilon} f(\frac{S_n}{n}) - f(x)\, dP + \int_{|\frac{S_n}{n} - x| > \varepsilon} f(\frac{S_n}{n}) - f(x)\, dP$$

    En déduire que la suite des polynômes B_n converge vers f uniformément sur $[0, 1]$.
14. Si on note C_x l'ensemble des cases controlées par la dame x, et $C = \bigcup_{x=1}^{7} C_x$, on peut minorer $|C|$ grâce aux inégalités de Bonferroni.
15. (a) On prendra $\Omega = \mathcal{B}_r(\{1, \ldots, n\})$.

    (b) Utiliser les relations entre l'espérance et la queue de distribution, puis utiliser la relation de récurrence du triangle de Pascal.
16. (a) On pourra remarquer qu'une rotation est une application linéaire isométrique.

    (b) Remarquer que $|X - Y| = \sqrt{2}U$.

B.4 Exercices sur les fonctions génératrices et les fonctions caractéristiques

1. (a) Remarquer que $t^Z = \mathbb{1}_A t^X + \mathbb{1}_{A^c} t^Y$.

   (b) Remarquer que le score est une variables aléatoire fabriquée suivant le principe de la première question.

2. (a) Remarquer que

$$\forall s \in [0,1] \quad s^S = \sum_{n=1}^{+\infty} \mathbb{1}_{T=n} s^{S_n}.$$

   (b) Utiliser les liens entre la fonction génératrice et l'espérance.

3. Utiliser la première question de l'exercice précédent.

4. (a) Commencer par déterminer la fonction génératrice de K_1L_1. On pourra s'inspirer de l'exercice 1.

   (b) S'inspirer de l'exercice 2.

   (c) Relire le cours.

5. (a) On trouvera $f^n \circ f$, où $f(z) = (1+z)/2$.

   (b) Remarquer que $f^n \circ f = (f \circ f)^n$

6. Penser à la fonction caractéristique.

7. Regarder la liste des fonctions caractéristiques connues, ou/et chercher l'équation fonctionnelle que doit vérifier φ_X.

8. Utiliser les liens entre fonction caractéristique et moments.

9. Utiliser les liens entre fonction caractéristique et moments.

10. (a) Pour tout borélien A, on a $P_X(A) = P_X(A \cap [0,+\infty[) + P_X(A \cap]-\infty, 0[)$.

    (b) Appliquer le théorème de transfert.

11. (a) On pourra utiliser le théorème de Fubini et se servir de la valeur de la fonction caractéristique d'une loi gaussienne.

    (b) On pourra remarquer que (U,V) a même loi que $(U,-V)$.

12. (a) Passer en coordonnées sphériques.

    (b) Montrer que $\varphi_\mu(t)$ ne dépend que de $\|t\|$, puis écrire $\varphi_\mu(t)$ sous la forme $\varphi_\mu(t) = \frac{1}{\|t\|^{2,5}} \int_0^{\|t\|} f(x)\, dx$.

B.5 Exercices sur la convergence presque sûre

1. Appliquer la loi des grands nombres
2. Pour tout $\varepsilon > 0$, appliquer le lemme de Borel-Cantelli aux événements $\{\frac{X_n}{(\ln n)^{\frac{3}{2}}} > \varepsilon\}$.
3. Discuter en fonction de A la nature de la série de terme général $P(\frac{X_n}{\ln n} > A)$.
4. S'inspirer de l'exercice précédent en utilisant un équivalent pour la queue de la gaussienne.
5. Pour une suite à veleurs entières, les valeurs d'adhérences sont les valeurs qui sont prises une infinité de fois.
6. On a $T_n = \sum_{k=2}^n 1\!\!1_{\{U_{n-1}<p\}\Delta\{U_n<p\}}$. On pourra découper T_n en deux sommes de variables aléatoires indépendantes.
7. (a) Passer au logarithme.
   (b) Écrire M_n sous la forme $M_n = PD_nP^{-1}$ et introduire la norme $\|x\|_* = \|P^{-1}x\|_\infty$.
8. On pourra remarquer que pour tout entier n, l'événement "$(Y_n)_{n\ge 1}$ est nulle à partir d'un certain rang" contient l'événement $\{X_n = 0\}$.
9. On pourra montrer que $X_n \in \{0,1\}$ à partir d'un certain rang.
10. On pourra montrer que $P(Y_n \neq 0) \le \exp(-\sum\limits_{k=1}^{n} \exp(-\lambda_k))$.
11. On pourra montrer que $(X_n)_{n\ge 2}$ ne prend la valeur 1 qu'un nombre fini de fois. Enfin, on montrera que
$$1 - p = \prod_{n=2}^{+\infty} (1 - \frac{1}{n^2}).$$

B.6 Exercices sur les vecteurs gaussiens

1. Remarquer que $\frac{\sigma(U)}{\sigma(V)}V$ a même loi que U.
2. On rappelle que si X suit $\mathcal{N}(0,1)$, $\mathbb{E}X^4 = 3$.
3. Trouver une matrice orthogonale O tel que la matrice de covariance de OX soit diagonale.
4. (a) Déterminer le spectre de C.
   (b) On pourra commencer par trouver une constante K telle que
$$\forall a \in \mathbb{R}^3 \quad \mathbb{E}|\langle X, a\rangle| = (\mathbb{E}|\langle X, a\rangle|_2^2)^{1/2}.$$

5. On pourra montrer qu'il existe un vecteur gaussien dont la matrice de covariance est
$$\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$
6. (a) Relire le cours !
   (b) Se souvenir qu'une matrice de covariance est symétrique positive.
   (c) Si (Z,T) est gaussien $(\frac{Z}{\sigma(Z)}, T)$ aussi.
   (d) Utiliser la question précédente.
   (e)
$$\mathbb{E}\|X\|^4 = \mathbb{E}\Big(\sum_{k=1}^{n} X_k^2\Big)^2.$$
7. Trouver une matrice orthogonale O tel que la matrice de covariance de OX soit diagonale.
8. (a) Revoir l'image d'un vecteur gaussien par une transformation affine.
   (b) On pourra utiliser un changement de variables.
9. On pourra diagonaliser A dans une base orthogonale.

B.7 Exercices sur la convergence en loi

1. Utiliser le premier théorème de Levy.
2. (a) Commencer par déterminer la loi de X_n.
   (b) Pour montrer la croissance, on pourra remarquer que $f(x) = \lim_{n\to+\infty} \mathbb{E}u \circ X_n^x$. Pour déterminer la limite, il est commode de se ramener au cas où $(u_n)_{n\geq 0}$ est à valeurs positives et remarquer qu'alors, pour tout n, il existe un polynôme P_n tel que
$$\forall x > 0 \quad f(x) \geq (1 - \frac{P_n(x)}{e^x})u_n.$$
3. On pourra commencer par déterminer R tel que $\mu(B(0,R)) \geq 1 - \frac{\varepsilon}{2}$, puis utiliser le théorème du porte-manteau.
4. On pourra considérer la loi du triplet (a_n, b_n, X_n).
5. On pourra considérer une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires indépendantes de même loi de Poisson de paramètre 1, poser $S_n = \sum_{k=1}^{n} X_k$ et regarder la quantité $P(\frac{S_n - \mathbb{E}S_n}{\sqrt{n}} \leq 0)$.

6. Notons $m_{x,y} = 1$ si la pièce en (x, y) est face, -1 sinon. Si on note $C'_n = \sum_{k \neq X} M_{k,X}$ et $L'_n = \sum_{k \neq X} M_{X,k}$, on a

$$D_n = \sum_{(k,l) \in \{1,;n\} \backslash \{X\}} m_{k,l} + m_{X,X} + |C'_n - L'_n|.$$

7. (a) On pourra utiliser le théorème de la limite centrale et le résultat de l'exercice 4.

   (b) Se souvenir qu'une loi exponentielle est une loi Gamma.

   (c) Utiliser le théorème de la limite centrale.

   (d) Remarquer que Ψ est bornée.

   (e) Intégrer.

   (f) Cela revient à montrer que (a_n) tend vers 1.

Annexe C

Problèmes

C.1 Densité naturelle des couples premiers entre eux

Notations et résultats admis

- Si E est un ensemble fini, $\mathcal{P}(E)$ désigne l'ensemble des parties de E et $|E|$ le cardinal de E. Si E est non-vide, alors il existe une unique mesure P sur $(E, \mathcal{P}(E))$ vérifiant $\forall x \in E \quad P(\{x\}) = \frac{1}{|E|}$. Cette mesure est une mesure de probabilité, appelée (mesure de) probabilité uniforme sur E. Elle vérifie la propriété suivante :

$$\forall A \in \mathcal{P}(E) \quad P(A) = \frac{|A|}{|E|}.$$

- Pour $x \in \mathbb{R}$, on appelle partie entière de x et on note $\mathrm{Ent}(x)$ l'unique entier n vérifiant $n \leq x < n+1$.
- $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$.
- Lorsque n et k sont des entiers, l'écriture $k|n$ signifie "k divise n".
  Si les séries de terme général (u_n) et (v_n) sont absolument convergentes, alors on a l'identité

$$\Big(\sum_{n\geq 1} u_n\Big)\Big(\sum_{n\geq 1} v_n\Big) = \sum_{n\geq 1}(u * v)_n,$$

  où l'on a posé

$$(u * v)_n = \sum_{d|n} u_d v_{n/d}.$$

Par exemple

$$(u * v)_6 = u_1v_6 + u_2v_3 + u_3v_2 + u_6v_1.$$

– Formule de Poincaré (aussi appelée formule du crible) Pour tous événemements $A_1, A_2, ..., A_n$ sous la probabilité P

$$\begin{aligned}
P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{B\in\mathcal{P}(\{1,\dots,n\})\setminus\varnothing} (-1)^{1+\mathrm{Card}(B)} P\big(\bigcap_{j\in B} A_j\big) \\
&= \sum_{i=1}^{n} P(A_i) - \sum_{1\le i_1<i_2\le n} P(A_{i_1}\cap A_{i_2}) + \cdots \\
&\quad \cdots + (-1)^{k+1} \sum_{1\le i_1<i_2<\dots<i_k\le n} P(A_{i_1}\cap\cdots\cap A_{i_k}) + \cdots \\
&\quad \cdots + (-1)^{n+1} P(A_1\cap\cdots\cap A_n).
\end{aligned}$$

Question préliminaire

Soit donnée une famille de nombres réels $a(k,n)$ pour $k \geq 1, n \geq 1$ entiers. On suppose qu'il existe une suite de nombres réels positifs $(c_k)_{k\geq 1}$ avec les propriétés :

$$\forall (n,k) \in \mathbb{N}^* \times \mathbb{N}^* \quad |a(k,n)| \leq c_k, \qquad \sum_{k=1}^{+\infty} c_k < +\infty.$$

On suppose que pour tout $k \geq 1$, la limite suivante existe :

$$\lim_{n\to+\infty} a(k,n) = a(k,\infty).$$

Montrer que les deux séries $s_n = \sum_{k=1}^{+\infty} a(k,n)$ et $s = \sum_{k=1}^{+\infty} a(k,\infty)$ convergent absolument et que l'on a $\lim_{n\to+\infty} s_n = s$, soit

$$\lim_{n\to+\infty} \sum_{k=1}^{+\infty} a(k,n) = \sum_{k=1}^{+\infty} a(k,\infty)$$

I

Soit $k \in \mathbb{N}^*$. On pose $E_k = \{0,1\}^k$.

Soit $A_k = \{x \in E_k;\ \sum_{i=1}^{k} x_i \text{ est pair}\}$ et $B_k = \{x \in E_k;\ \sum_{i=1}^{k} x_i \text{ est impair}\}$.

1. Montrer que l'application

$$\begin{array}{rcl} \Psi : E_k & \to & E_k \\ (x_1, x_2, \ldots, x_k) & \mapsto & (1 - x_1, x_2, \ldots, x_k) \end{array}$$

   réalise une bijection de A_k dans B_k.
2. Montrer que pour tout entier naturel k non nul

$$|A_k| = |B_k| = 2^{k-1}.$$

3. k étudiants assistent à un cours de mathématiques. Quelle est la probabilité pour que le nombre d'étudiants qui s'ennuyent soit pair ? Préciser les hypothèses simplificatrices que vous choisirez de prendre.
   Indication : Repérer l'état de lassitude des étudiants en associant à chaque étudiant un élément de $\{0, 1\}$ suivant l'intérêt qu'il porte au cours.

II

Soit $n \in \mathbb{N}, n \geq 2$. On décompose n en produits de facteurs premiers :

$$n = \prod_{i=1}^{k} p_i^{\alpha_i},$$

où pour tout i dans $\{1, ..., k\}$, p_i est un nombre premier et α_i un entier strictement positif, les p_i étant deux à deux distincts. On pose

$$\mu(n) = \begin{cases} 0 & \text{si il existe } i \text{ tel que } \alpha_i > 1, \\ (-1)^k & \text{sinon.} \end{cases}$$

On pose également $\mu(1) = 1$.
On a ainsi défini une fonction $\mu : \mathbb{N}^* \to \{-1, 0, 1\}$.
Cette fonction est appelée fonction de Moëbius.
On pose

$$\Omega_n = \{0, \ldots, \alpha_1\} \times \{0, \ldots, \alpha_2\} \times \cdots \times \{0, \ldots, \alpha_k\}$$

et on note Ω'_n l'ensemble des diviseurs positifs de n. On note Q la mesure de probabilité uniforme sur Ω_n et Q' la mesure de probabilité uniforme sur Ω'_n.

1. Montrer que $|\Omega_n| = \prod_{i=1}^{k} (\alpha_i + 1)$.

2. Soit f l'application :
$$\begin{array}{rcl} \Omega_n & \to & \mathbb{N} \\ (\nu_1, \ldots, \nu_k) & \mapsto & \prod_{i=1}^{k} p_i^{\nu_i} \end{array}$$
Montrer que Q' est la mesure image de Q par f.
3. On pose $Z_k = \Omega_n \backslash E_k$. Montrer que pour toute fonction g de Ω_n dans $\mathbb{R}$, on a
$$\int_{\Omega_n} g(\omega)\ dQ(\omega) = \int_{A_k} g(\omega)\ dQ(\omega) + \int_{B_k} g(\omega)\ dQ(\omega) \int_{Z_k} g(\omega)\ dQ(\omega)$$
4. Montrer $\forall \nu = (\nu_1, \ldots, \nu_k) \in \Omega_n$
$$(\mu \circ f)(\nu) = \begin{cases} (-1)^{\left(\sum_{i=1}^{k} \nu_i\right)} & \text{si } \nu \in E_k, \\ 0 & \text{sinon.} \end{cases}$$
5. En utilisant le théorème d'intégration par rapport à une mesure image – appelé aussi théorème de transfert – , montrer que
$$\int_{\Omega'_n} \mu(x)\ dQ'(x) = Q(A_k) - Q(B_k).$$
6. Montrer $\forall n \in \mathbb{N}^*$
$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{si } n = 1, \\ 0 & \text{sinon.} \end{cases}$$
7. On choisit au hasard - avec équiprobabilité - un diviseur de 2000. Quelle est la probabilité pour que ce nombre soit divisible par un carré autre que 1 ?

III

Pour n fixé, soit Ω l'ensemble des couples d'entiers $(u, v) \in \{1, ..., n\}^2$. On suppose que tous les couples sont équiprobables. On note donc P_n la probabilité définie sur $\mathcal{P}(\Omega)$ par
$$P_n(A) = \frac{1}{n^2}|A|.$$
On note p_n la probabilité pour que deux entiers pris "au hasard" entre 1 et n soient premiers entre eux.
Pour $k \in \mathbb{N}^*$, on pose $F_k = \{(u, v) \in \Omega;\ k|u \text{ et } k|v\}$.

1. Montrer $P_n(F_k) = (\mathrm{Ent}(n/k))^2/n^2$.
2. Montrer que si k et l sont premiers entre eux, alors
$$P_n(F_k \cap F_l) = P_n(F_{kl}).$$
Généraliser au cas de m entiers deux à deux premiers entre eux.
3. Soient $\beta_1, ..., \beta_k$ les nombres premiers compris entre 2 et n.
On pose $F = \bigcup_{i=1}^{k} F_{\beta_i}$.
Montrer $p_n = 1 - P_n(F)$.
4. En utilisant la formule de Poincaré, montrer
$$p_n = \sum_{d=1}^{n} \mu(d) \frac{1}{n^2} (\mathrm{Ent}(\frac{n}{d}))^2.$$
Indication : il pourra être utile de remarquer que pour tout entier d compris entre 1 et n, $\mu(d)$ est non-nul si et seulement si d peut s'écrire comme produit d'éléments deux à deux distincts pris dans l'ensemble $\{\beta_1, \dots, \beta_k\}$.
5. (a) On pose
$$a(k,n) = \begin{cases} \mu(k)\frac{1}{n^2}(\mathrm{Ent}(\frac{n}{k}))^2 & \text{si } 1 \leq k \leq n, \\ 0 & \text{sinon.} \end{cases}$$
Montrer
$$\forall (n,k) \in (\mathbb{N}^*)^2 \quad |a(k,n)| \leq \frac{1}{k^2}$$
ainsi que
$$\forall k \in \mathbb{N}^* \quad \lim_{n \to \infty} a(k,n) = \frac{\mu(k)}{k^2}.$$
(b) Établir que la suite $(p_n)_{n \geq 2}$ converge vers
$$p = \sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2}.$$
6. Calculer p.
Indication : on pourra appliquer le résultat sur les séries absolument convergentes admis en introduction avec $u_n = \frac{\mu(n)}{n^2}$ et $v_n = \frac{1}{n^2}$, puis utiliser II.6.
7. Que pensez vous de l'affirmation suivante :
« La probabilité pour que deux entiers pris au hasard soient premiers entre eux est $6/\pi^2$. » ?

C.2 Preuve de la loi forte des grands nombres

Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives de même loi deux à deux indépendantes, admettant un moment d'ordre un. On pose

$$S_n = \sum_{i=1}^{n} X_i \text{ et } Q_n = \frac{1}{n} S_n.$$

On considère également les variables aléatoires tronquées : $Y_i = X_i 1\!\!1_{X_i \leq i}$ et les sommes et quotients associés : $S_n^* = \sum_{i=1}^{n} Y_i$ et $Q_n^* = S_n^*/n$.

1. Montrer
$$\operatorname{Var} S_n^* \leq \sum_{i=1}^{n} \mathbb{E} Y_i^2.$$
En déduire
$$\operatorname{Var} S_n^* \leq n \mathbb{E} X_1^2 1\!\!1_{X_1 \leq n}.$$
2. Soit $\beta > 1$. On note u_n l'entier le plus proche de β^n. Montrer $u_n \sim \beta^n$. En déduire
$$\sum_{n=N}^{+\infty} \frac{1}{u_n} = O(\frac{1}{u_n}).$$
3. Montrer qu'il existe une constante $C > 0$ telle que
$$\forall N \geq 1 \quad \sum_{n=N}^{+\infty} \frac{1}{u_n} \leq C \frac{1}{u_n}.$$
4. Montrer
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n} \leq \mathbb{E} X_1^2 \sum_{n=1}^{+\infty} \frac{1}{u_n} 1\!\!1_{X_1 \leq u_n},$$
puis
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* \leq \mathbb{E} X_1^2 \sum_{n: u_n \geq X_1} \frac{1}{u_n}$$
5. En déduire que
$$\sum_{n=1}^{+\infty} \operatorname{Var} Q_{u_n}^* < +\infty.$$
6. Montrer que
$$Q_{u_n}^* - \mathbb{E} Q_{u_n}^* \to 0 \text{ p.s.}$$

7. Montrer que
$$\lim_{n\to+\infty} \mathbb{E}Y_n = \mathbb{E}X_1.$$
8. Montrer que $\lim \mathbb{E}Q_n^* = \mathbb{E}X_1$. (Indication : on pourra utiliser le théorème de Césaro)
9. En déduire que
$$Q_{u_n}^* \to \mathbb{E}X_1 \text{ p.s.}$$
10. Montrer que la série de terme général $P(X_n \neq X_n^*)$ converge.
11. A l'aide du lemme de Borel-Cantelli, montrer que pour presque tout ω, il existe un $n_0(\omega)$ tel que les suites $(X_n(\omega))$ et $(X_n^*(\omega))$ coincident à partir du rang $n_0(\omega)$.
12. Montrer que
$$Q_{u_n} \to \mathbb{E}X_1 \text{ p.s.}$$
13. Si $u_n \leq k \leq u_{n+1}$, montrer que
$$\frac{u_n}{u_{n+1}} Q_{u_n} \leq Q_k \leq \frac{u_{n+1}}{u_n} Q_{u_{n+1}}.$$
14. En déduire que
$$\frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k \leq \overline{\lim}_{k\to+\infty} Q_k \leq \beta\mathbb{E}X_1 \text{ p.s.}$$
15. On note
$$\Omega_\beta = \{\omega \in \Omega; \frac{1}{\beta}\mathbb{E}X_1 \leq \underline{\lim}_{k\to+\infty} Q_k(\omega) \leq \overline{\lim}_{k\to+\infty} Q_k(\omega) \leq \beta\mathbb{E}X_1\}.$$
Montrer que
$$P(\bigcap_{n=1}^{+\infty} \Omega_{1+1/n}) = 1.$$
16. Montrer que
$$\lim_{k\to+\infty} Q_k = \mathbb{E}X_1 \text{ p.s.}.$$
17. Montrer que le résultat de la dernière question demeure vrai si l'on ne suppose plus que les (X_n) sont des variables positives.

Index