Année universitaire 2000-2001

UNIVERSITÉ D'ORLÉANS

Olivier GARET

MA3.03 : Probabilités et Graphes

Tome I

Probabilités

Table des matières

Chapitre 1

Ensembles

1.1 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, 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, $\underset{i\in I}{\cap} A_i$ 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 $x \in \bigcap_{i \in I} A_i$.”

1.2 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). latex prograp 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.

1.3 Principes de base du dénombrement

1.3.1 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.

1.3.2 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.

1.3.3 Principe de partition

On dit que les ensembles $(A_i)_{i\in I}$ forment une partition de A si l'on a $A = \bigcup\limits_{i\in I} A_i$ 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 1. *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 $\Omega = \bigcup\limits_{i\in I} \Omega_i$, on a $A = A\cap\Omega = \bigcup\limits_{i\in I} A\cap\Omega_i = \bigcup\limits_{i\in I} A_i$. D'autre part, pour $i \neq j$, on a $A_i \cap A_j \subset \Omega_i \cap \Omega_j = \varnothing$, d'où $A_i \cap A_j = \varnothing$. $\square$

1.3.4 Lemme des bergers

Le lemme suivant peut également être utile

Lemme 1 (Lemme des bergers). *Soit ϕ une application* surjective *de A dans D. On suppose qu'il existe un entier $a \geq 1$ tel que*

$$\forall y \in A \quad |\{x \in D; \phi(x) = y\}| = a$$

(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 $y \in A$: $D_y = \{x \in D; \phi(x) = y\}$, les D_y forment clairement une partition de D, d'où

$$|D| = \sum_{y \in A} |D_y| = \sum_{y \in A} a = |A|a.$$

$\square$

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.

1.4 Quelques résultats incontournables

1.4.1 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 $A = \{0, \ldots, 20\}$ des notes possibles. Comme $|A| = 21$ et $|D| = 30$, il y a donc 21^{30} résultats possibles.

1.4.2 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.

1.4.3 Nombre d'injections de D dans A

On pose $|A| = n$ et $|D| = p$. En vertu de la remarque faite en 1.2, 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 : $3500 \times 3499 \times \cdots \times 3201 \times 3200$. Ici D est l'ensemble des rangs, on a donc $D = \{1, \dots, 300\}$ 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.

1.4.4 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

$$\binom{n}{p} = \frac{n(n-1)\dots(n-p+1)}{p(p-1)\dots 1} = \frac{n!}{p!(n-p)!}.$$

Démonstration. Il suffit d'appliquer le lemme des bergers à

- D : ensemble des injections de $\{1,\dots,p\}$ dans Ω
- $A = \mathcal{B}_p(\Omega)$
- ϕ définie par $\phi(f) = \{f(k); k \in \{1,\dots,p\}\}$

On a vu précédemment que $|A| = n(n-1)\dots(n-p+1)$. Il n'est pas difficile de voir que ϕ est surjective. Une partie $\{e_1,\dots,e_p\}$ de Ω étant donnée, combien existe-t-il d'injections (en fait de bijections) de $\{1,\dots,p\}$ dans Ω telles que $\{f(1),\dots,f(p)\} = \{e_1,\dots,e_p\}$? C'est évidemment le nombre d'injections de $\{1,\dots,n\}$ dans $\{e_1,\dots,e_p\}$, 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.

1.4.5 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.

1.5 Équations et inéquations en entiers

Lemme 2. *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 2. *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 :*

$$x_1 + x_2 + \cdots + x_p \leq n \tag{1.1}$$

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 $\{1, \dots, n\}$. □

Théorème 3. *Pour n, p entiers vérifiant $n \geq p$, il existe exactement $\binom{n-1}{p-1}$ p-uplets $(x_1, \dots, x_p) \in (\mathbb{N} \backslash \{0\})^p$ solutions à l'équation :*

$$x_1 + x_2 + \cdots + x_p = n \tag{1.2}$$

Démonstration. Il suffit de remarquer que les solutions $(x_1, \dots, x_p) \in (\mathbb{N} \backslash \{0\})^p$ de l'équation (1.2) sont exactement les solutions de l'inéquation (1.2) qui ne sont pas solutions de l'inéquation

$$x_1 + x_2 + \cdots + x_p \leq n - 1 \tag{1.3}$$

Il y en a donc $\binom{n}{n} - \binom{n-1}{p} = \binom{n-1}{p-1}$.

□

Théorème 4. *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 :*

$$x_1 + x_2 + \cdots + x_p = n \tag{1.4}$$

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 (1.4) et les solutions $(x_1, \ldots, x_p) \in (\mathbb{N} \backslash \{0\})^p$ de l'équation

$$x_1 + x_2 + \cdots + x_p = n + p \tag{1.5}$$

□

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 $\binom{9+4-1}{4-1} = \binom{12}{3} = \frac{9 \times 8 \times 7}{1 \times 2 \times 3} = 84$ répartitions possibles.

Théorème 5. *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 :*

$$x_1 + x_2 + \cdots + x_p \leq n \tag{1.6}$$

Démonstration. La preuve, analogue à celle du théorème précédent, est laissée en exercice. □

Chapitre 2

Formalisme probabiliste, notion d'indépendance

2.1 Introduction

Il y a dans le monde de nombreuses expériences à venir dont on ne peut prévoir l'issue avec certitude.

1. La pièce que je lance tombera-t-elle sur pile ou sur face ?
2. Jacques Chirac sera-t-il réélu Président de la République ?
3. Combien de temps me faudra-t-il pour obtenir ma licence ?

Le présent cours ne prétend pas répondre à ces questions. Si même ce cours vous permettait de tenter une prédiction sur l'un de ces événements et que celle-ci se révélait inexacte, ce ne saurait cependant remettre en question la pertinence de ce cours. Si vous êtes un téléspectateur assidu des soirées électorales, je gage que vous avez déjà entendu des arguments similaires dans la bouche des spécialistes des instituts de sondage.

Pourquoi un résultat isolé ne saurait remettre en question la théorie ? Parce que la théorie est basée sur le postulat suivant : si une même expérience est répétée un grand nombre de fois, la fréquence d'obtention d'un même résultat tend à se stabiliser vers un nombre appelé probabilité théorique du résultat. Une seule expérience ne peut donc prétendre approcher l'infini. Il faut souligner que ce postulat n'est pas un postulat mathématique – comme les postulats d'Euclide ou l'axiome du choix – mais un postulat physique, c'est à dire sur la nature de l'univers. Évidemment, ce postulat a comme origine la pratique : on s'est aperçu que si on lançait 1000 fois une pièce et que l'on divisait le nombre de faces obtenues par 1000, le résultat obtenu était proche de 1/2 – quoique rarement égal à 1/2.

L'objectif d'une théorie des probabilités est de bâtir des objets mathématiques dont au pourra calculer la probabilité – mot dont le sens mathématique reste encore à définir. Pour que la construction alors faite ait un sens, il est nécessaire dans ce cadre théorique, l'on puisse démontrer que si une même expérience est répétée une infinité fois de manière indépendantes , la fréquence d'obtention d'un même résultat converge (en un sens à préciser) vers la probabilité théorique alors définie. Cette vérification est nécessaire si l'on veut pouvoir relier les nombreuses pages de calcul que nous allons faire à la réalité. Cet objectif majeur est l'une des grandes nouveautés de ce cours par rapport à ce que vous avez pu voir dans les années antérieures.

2.2 Des événements aux espaces probabilisés

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 :

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

On dit qu'une partie $\mathcal{F} \subset \mathcal{P}(\Omega)$ est une famille d'observables – on dit aussi 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 $\mathcal{F}$, $\bigcup\limits_{i=1}^{+\infty} A_i \in \mathcal{F}$.

Une fois que Ω et $\mathcal{F}$ sont fixés, on appelle événement tout élément de $\mathcal{F}$. Voyons maintenant la définition d'une probabilité sur $(\Omega, \mathcal{F})$.

On appelle probabilité 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, $P(\bigcup\limits_{i=1}^{+\infty} A_i) = \sum\limits_{i=1}^{+\infty} P(A_i)$.

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

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

- $P(\varnothing) = 0$
- $\forall A \in \mathcal{F} \quad P(A^c) = 1 - P(A)$
- $\forall A, B \in \mathcal{F} \quad A \cap B = \emptyset \Longrightarrow P(A \cup B) = P(A) + P(B)$
- $\forall A, B \in \mathcal{F} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $\forall A, B \in \mathcal{F} \quad P(A \cup B) \leq P(A) + P(B)$
- $\forall A, B \in \mathcal{F} \quad A \subset B \Longrightarrow P(A) \leq P(B)$
- $\forall A, B \in \mathcal{F} \quad P(A \cap B) \leq \min(P(A), P(B))$
- $\forall A, B \in \mathcal{F} \quad P(A \cup B) \geq \max(P(A), P(B))$
- Pour toute suite $(A_i)_{1 \leq i \leq n}$ d'éléments de $\mathcal{F}$ deux à deux disjoints, $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.
- Pour toute suite $(A_i)_{1 \leq i \leq n}$ d'éléments de $\mathcal{F}$, $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$.
- 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 $A = \bigcup_{i=1}^{+\infty} A_i$, alors la suite $(P(A_n))_{n \in \mathbb{N}}$ est monotone, croissante, et converge vers $P(A)$.
- 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 $A = \bigcap_{i=1}^{+\infty} A_i$, alors la suite $(P(A_n))_{n \in \mathbb{N}}$ est monotone, décroissante, et converge vers $P(A)$.

2.3 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 6. *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. D'après le théorème 1, 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 $P(A)=\sum_{i\in I}P(A\cap\Omega_i)$. □

2.4 Événements indépendants

2.4.1 Indépendance de deux événements

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

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

Théorème 7. *Soient A et B deux événements observables.*
A est indépendant de B si et seulement si A est indépendant de B^c.

Démonstration. Calculons de deux manières différentes $P(A)$: on a d'une part

$$P(A) \quad = \quad P(A\cap B)+P(A\cap B^c)$$

D'autre part $1=P(B)+P(B^c)$ entraîne $P(A)=P(A)P(B)+P(A)P(B^c)$. On a donc

$$P(A\cap B)+P(A\cap B^c)=P(A)P(B)+P(A)P(B^c).$$

Cette égalité entraîne immédiatement que $P(A\cap B)=P(A)P(B)$ est équivalent à $P(A\cap B^c)=P(A)P(B^c)$. □

Remarque : Les événement de probabilité 0 ou de probabilité 1 sont indépendants de n'importe quels événements – même d'eux-mêmes.

Démonstration. Soit B un événement de probabilité nulle : soit A un événement observable quelconque : on a $P(A\cap B)\le P(B)=0$ donc $P(A\cap B)=0=P(A)P(B)$. B est donc indépendant de n'importe quel événement observable.

Si B est de probabilité 1, B^c est de probabilité 0 et est donc indépendant de n'importe quel événement, d'après la première partie de notre raisonnement. . Le théorème 7 permet alors de conclure. □

2.4.2 Indépendance d'une famille d'événements

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 $I\subset G$:

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

Remarque : si un ou plusieurs des événements est remplacé par son complémentaire, cela ne change rien à l'indépendance de la famille. La preuve est laissée au lecteur ; il faut utiliser le théorème 7 et effectuer une récurrence sur $|I|$.

Chapitre 3

Variables aléatoires

3.1 Généralités

3.1.1 Définition

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. On dit qu'une application X définie sur Ω est à valeurs dans $\mathbb{R}$ est une variable aléatoire si elle vérifie la propriété suivante :

$$\forall t \in \mathbb{R} \quad \{\omega \in \Omega; X(\omega) \leq t\} \in \mathcal{F}.$$

Dès lors, le nombre

$$P(\{\omega \in \Omega; X(\omega) \leq t\})$$

est bien défini. Pour $x \in \mathbb{R}$, on pose

$$F_X(t) = P(\{\omega \in \Omega; X(\omega) \leq t\}).$$

La fonction $t \mapsto F_X(t)$ est appelée fonction de répartition. Nous allons y revenir bientôt.

Comme les écritures du type $P(\{\omega \in \Omega; X(\omega) \leq t\})$ sont assez lourdes, on choisit de les alléger en $P(X \leq t)$. Ainsi, dans la pratique courante et à notre niveau, l'existence d'Ω peut parfois être oubliée.

3.1.2 Loi d'une variable aléatoire

Il est naturel de se demander quels sont les événements relatifs à la variable aléatoire X dont on peut calculer la probabilité. Précisément, pour quels ensembles $A \subset \mathbb{R}$ a-t-on

$$\{\omega \in \Omega; X(\omega) \in A\} \in \mathcal{F} \quad ?$$

Pour répondre à cette question, on admettra le théorème suivant :

Théorème 8. *Il existe une unique famille d'observables (ou tribu) sur $\mathbb{R}$, quel l'on note $\mathcal{B}(\mathbb{R})$, vérifiant les propriétés suivantes*

- *$\mathcal{B}(\mathbb{R})$ contient tous les ouverts de $\mathbb{R}$.*
- *Toute tribu sur $\mathbb{R}$ contenant tous les ouverts de $\mathbb{R}$ contient $\mathcal{B}(\mathbb{R})$.*

$\mathcal{B}(\mathbb{R})$ est appelée tribu borélienne de $\mathbb{R}$. Les éléments de la tribu borélienne sont appelés les boréliens de $\mathbb{R}$. Si X est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$, on a pour tout borélien A de $\mathbb{R}$:

$$\{\omega \in \Omega; X(\omega) \in A\} \in \mathcal{F}$$

Dès lors, le nombre $P(\{\omega \in \Omega; X(\omega) \in A\})$(que l'on note aussi $P(X \in A)$) est bien défini.

Les singletons, les intervalles de $\mathbb{R}$, les ouverts et les fermés de $\mathbb{R}$ sont des boréliens. De même, toute union ou intersection finie ou dénombrable de tels ensembles est un boréliens

Théorème 9. *Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ L'application*

$$\begin{array}{rcl} P_X : \mathcal{B}(\mathbb{R}) & \to & [0,1] \\ A & \mapsto & P(X \in A) \end{array}$$

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

On dit que la probabilité P_X est la loi de la variable aléatoire X sous la probabilité P, ou, plus briévement, que P_X est la loi de X.

3.1.3 Fonction de répartition

On remarque que la fonction de répartition de X est liée à la loi de X par la formule

$$F_X(t) = P_X(]-\infty, t]).$$

Ainsi, lorsque l'on connaît la loi d'une variable aléatoire, on peut retrouver la fonction de répartition. Réciproquement, on admettra que la fonction de répartition détermine de manière unique la loi de la variable aléatoire.

Nous allons maintenant énoncer les propriétés caractéristiques d'une fonction de répartition.

Théorème 10. *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 $s \le t$, on a $]-\infty, s] \subset]-\infty, s]$, d'où

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

- Posons pour $n \ge 1$, $A_n =]-\infty, -n]$, on a $A_{n+1} \subset A_n$ et $\bigcap_{n\ge 1} A_n = \varnothing$, 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

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

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

- Posons pour $n \ge 1$, $A_n =]-\infty, n]$, on a $A_n \subset A_{n+1}$ et $\bigcup_{n\ge 1} A_n = \mathbb{R}$, 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 $P(A_n) \ge< 1-\varepsilon$. Comme F_X est croissante et majorée par 1, on a

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

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

- Soit $t \in \mathbb{R}$, $(t_n)_{n\ge 1}$ une suite décroissante convergeant vers t. Posons pour $n \ge 1$, $A_n =]-\infty, t_n]$, on a $A_{n+1} \subset A_n$ et $\bigcap_{n\ge 1} A_n =]-\infty, t]$, 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.

□

3.1.4 Fonction d'une variable aléatoire

Il est très simple de fabriquer, à partir d'une variable aléatoire, de nouvelles variables aléatoires. On admettra le théorème suivant :

Théorème 11. *Soit X une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ et f une fonction continue par morceaux définie sur $X(\Omega)$. Alors, la fonction Y définie par*

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

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

3.1.5 Variables aléatoires indépendantes

On dit que des variables aléatoires $X_1, \dots, X_n$ sont indépendantes si pour tout n-uplet $(A_1, \dots, A_n) \in \mathcal{B}(\mathbb{R})^n$, les événements

$$\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$$

sont indépendants.

Théorème 12. *Des variables aléatoires $X_1, \dots, X_n$ sont indépendantes si et seulement si pour tout n-uplet $(A_1, \dots, A_n) \in \mathcal{B}(\mathbb{R})^n$, on a*

$$P(\bigcap_{i=1}^{n} {}^{n} A_i) = \prod_{i=1}^{n} P(A_i).$$

Démonstration. La preuve est laissée en exercice au lecteur. Une des implications est évidente, l'autre mérite un peu de réflexion. □

3.1.6 Fonctions de variables aléatoires indépendantes

Théorème 13. *Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et $f_1, \dots, f_n$ des fonctions continues par morceaux respectivement définies sur $X_1(\Omega), \dots, X_n(\Omega)$. Alors les variables aléatoires $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes*

La preuve est laissée en exercice.

3.2 Variables aléatoires discrètes

On dit que X est une variable aléatoire discrète s'il existe un ensemble D fini ou dénombrable inclus dans $\mathbb{R}$ tel que $X(\Omega) = D$.

Alors, si 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 est 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 $X(\Omega) = D$. Alors, pour tout $A \in \mathcal{B}(\mathbb{R})$, on a*

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

où l'on a posé $p_i = P(X = i)$.

Démonstration. Si l'on pose pour $i \in D$

$$\Omega_i = \{\omega \in \Omega; X(\omega) = i\},$$

il est facile de voir que les Ω_i forment une partition de Ω. On a donc

$$\{\omega \in \Omega; X(\omega) \in A\} = \bigcup_{i \in D} \Omega_i \cap \{\omega \in \Omega; X(\omega) \in A\})$$

Mais si $i \notin A$, on a $\Omega_i \cap \{\omega \in \Omega; X(\omega) \in A\}) = \varnothing$. Inversement, si $i \in A$, on a $\Omega_i \cap \{\omega \in \Omega; X(\omega) \in A\}) = \Omega_i$ d'où

$$\{\omega \in \Omega; X(\omega) \in A\} = \bigcup_{i \in D; i \in A} \Omega_i,$$

puis

$$P_X(A) = P(\{\omega \in \Omega; X(\omega) \in A\}) = \sum_{i \in D; i \in A} P(\Omega_i) = \sum_{i \in D \cap A} p_i.$$

$\square$

En particulier, on a donc

$$F_X(t) = \sum_{i \in D; i \leq t} p_i.$$

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. On peut prendre $\Omega = D$, $\mathcal{F} = \mathcal{P}(\Omega)$ et $X(\omega) = \omega$ avec

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

(La fin de la preuve est laissée au lecteur). □

3.2.1 Fonction d'une variable aléatoire discrète

Théorème 16. *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 écrira $Y = f(X)$.

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

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

3.2.2 Variables et lois discrètes classiques

3.2.3 Indicatrice d'un évévenement

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

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

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

3.2.4 Loi uniforme sur un ensemble

Soit E un ensemble fini. On dit qu'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 $\{1, 2, 3, 4, 5, 6\}$.

3.2.5 Loi binomiale

On dit qu'une variable aléatoire X suit la loi binomiale de paramètres n et p si l'on a

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

Théorème 17. *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 .

Démonstration. L'événement $\{X = k\}$ se partitionne ainsi :

$$\{X = k\} \quad = \quad \underset{B \in \mathcal{B}_k(\{1,\dots,n\})}{\cup} \; (\underset{i \in B}{\cap} A_i) \cap (\underset{i \in \{1,\dots,n\} \setminus B}{\cap} A_i^c)$$

Or, pour tout $B \in \mathcal{B}_k(\{1, \dots, n\})$, on a

$$\begin{aligned} P((\underset{i \in B}{\cap} A_i) \cap (\underset{i \in \{1,\dots,n\} \setminus B}{\cap} A_i^c)) &= (\prod_{i \in B} P(A_i)).(\prod_{i \in \{1,\dots,n\} \setminus B} P(A_i^c)) \\ &= p^k (1-p)^{n-k} \end{aligned}$$

(On utilise ici l'indépendance des (A_i).) On en déduit

$$P(X=k) = \sum_{B \in \mathcal{B}_k(\{1,\dots,n\})} p^k(1-p)^{n-k} = |\mathcal{B}_k(\{1,\dots,n\})| p^k(1-p)^{n-k} = \binom{n}{k} p^k(1-p)^{n-k}.$$

□

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 $\frac{1}{2}$.

3.2.6 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 18. *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.

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

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

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

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

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

Comme

$$\{X = +\infty\} = \underset{k \ge 1}{\cap} \{X > k\},$$

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

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

□

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

3.2.7 Loi de Poisson

On dit qu'une variable aléatoire X suit la loi géométrique 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

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

3.3 Convolution de variables aléatoires discrètes

Le but de cette section est de déterminer la loi de la somme de deux variables aléatoires discrètes indépendantes.

Théorème 19. *Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, P)$ à valeurs dans $\mathbb{Z}$ On pose $Z = X + Y$. Alors, Z est une variable aléatoire discrète vérifiant*

$$\forall n \in \mathbb{Z} \quad P(Z=n) = \sum_{i \in \mathbb{Z}} P(X=i)P(Y=n-i).$$

Si X et Y sont en fait à valeurs dans $\mathbb{N}$, la formule ci-dessous se simplifie en

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

Démonstration. Soit $n \in \mathbb{Z}$.

$$\begin{aligned}
P(Z=n) &= \sum_{i\in\mathbb{Z}} P(\{Z=n\}\cap\{X=i\}) &&= \sum_{i\in\mathbb{Z}} P(\{X+Y=n\}\cap\{X=i\}) \\
&= \sum_{i\in\mathbb{Z}} P(\{Y+i=n\}\cap\{X=i\}) &&= \sum_{i\in\mathbb{Z}} P(\{Y=n-i\}\cap\{X=i\}) \\
&= \sum_{i\in\mathbb{Z}} P(X=i)P(Y=n-i)
\end{aligned}$$

Si X est à valeurs dans $\mathbb{N}$, $P(X=i)=0$ pour $i<0$. De même, si Y est à valeurs dans $\mathbb{N}$, $P(Y=n-i)=0$ pour $i>n$. On en déduit les simplifications annoncées. □

Chapitre 4

Dépendance, lois conjointes

Dans la vie scientifique courante, il est souvent nécessaire de savoir si des événements sont liés. Quelques exemples :

- Être contrôlé à plus de 110 km/h sur la route Pérenchies-Vervicq et avoir réellement commis un excès de vitesse.
- Avoir la moyenne au partiel et réussir l'examen.
- Se déclarer, lors d'un sondage, hostile à l'immigration et voter pour un parti fasciste.

Il faut bien remarquer que les liens entre ces événements ne sont pas symétriques. Par exemple, il est fort probable que quelqu'un contrôlé à 110km/h ait effectivement oublié son pied sur l'accélérateur, alors que – à l'heure où ses lignes sont écrites ! – il est peu probable que quelqu'un dépassant la limitation de vitesse sur la route Pérenchies-Vervicq se fasse prendre, vu le faible nombre de contrôles.

4.1 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

$$\begin{array}{rcl} P(.|B) : \mathcal{F} & \to & \mathbb{R} \\ A & \mapsto & P(A|B) = \dfrac{P(A \cap B)}{P(B)} \end{array}$$

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

On a évidemment

$$P(A \cap B) = P(B)P(A|B). \tag{4.1}$$

Remarques :

- L'application "probabilité conditionnelle" est une probabilité. Elle vérifie donc toutes les propriétés énoncées au chapitre 2.
- Si $P(B) \neq 0$, alors A est indépendant de B si et seulement si $P(A|B) = P(A)$.

4.1.1 Conditionnements en chaîne

Si A, B sont deux événements observables avec $A \subset B$ et $P(B) \neq 0$, la formule (4.1) devient

$$P(A) = P(B)P(A|B). \tag{4.2}$$

On a la généralisation suivante :

Théorème 20. *Soient $n \geq 2$ et $E_1, \ldots, E_n$ des événements observables vérifiant*

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

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

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

Démonstration. La formule se montre par récurrence sur n. Pour $n = 2$, c'est une conséquence immédiate de (4.2). 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 21. *Soient $n \geq 2$ et $A_1, \ldots, A_n$ des événements observables avec $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Alors*

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

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

4.1.2 Conditionnement par tous les cas possibles

Ceci est une reformulation à l'aide des probabilités conditionnelles du théorème 6.

Théorème 22. *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 6, on a

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

□

4.1.3 Formule de Bayes

Théorème 23. *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

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

4.2 Vecteurs aléatoires

4.2.1 Définition

Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé, et n un entier strictement positif. On dit qu'une application X définie sur Ω est à valeurs dans $\mathbb{R}^n$ est un vecteur aléatoire si elle vérifie la propriété suivante :

$$\forall t\in\mathbb{R}^n \quad \{\omega\in\Omega; X(\omega)\in[t_1,+\infty[\times[t_2,+\infty[\times\cdots\times[t_n,+\infty[\times\}\in\mathcal{F}.$$

On admet le théorème suivant :

Théorème 24. *Soit X une application de Ω et à valeurs dans $\mathbb{R}^n$. On note*

$$X(\omega) = (X_1(\omega),\dots,X_n(\omega)).$$

Alors X est un un vecteur aléatoire si et seulement si pour tout $i\in\{1,\dots,n\}$, X_i est une variable aléatoire.

4.2.2 Loi marginale

Soit $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ un vecteur aléatoire. On appelle loi marginale de X suivant la i ème coordonnée la loi de X_i.

4.2.3 Couple aléatoire à valeurs dans un ensemble fini

Soit Y, Z deux variables aléatoires définies sur $(\Omega, \mathcal{F}, P)$. On suppose que $Y(\Omega) = \{y_1, \dots, y_q\}$ et $Z(\Omega) = \{z_1, \dots, z_p\}$. On définit alors une matrice M de taille $q \times p$ par

$$\forall i \in \{1, \dots, q\} \quad \forall j \in \{1, \dots, p\} \quad m_{i,j} = P(\{Y = y_i\} \cap \{Z = z_j\})$$

Alors, la matrice M a les propriétés suivantes :

- $\forall i \in \{1, \dots, q\} \quad \forall j \in \{1, \dots, p\} \quad m_{i,j} \geq 0$.
- $\sum_{i=1}^{q} \sum_{j=1}^{p} m_{i,j} = 1$.

Elle permet de retrouver les lois marginales P_Y et P_Z, puisqu'on a les formules suivantes :

-
$$\forall i \in \{1, \dots, q\} \quad P(Y = y_i) = \sum_{j=1}^{p} P(\{Y = y_i\} \cap \{Z = z_j\}) = \sum_{j=1}^{p} m_{i,j}.$$
-
$$\forall j \in \{1, \dots, p\} \quad P(Z = z_j) = \sum_{i=1}^{q} P(\{Y = y_i\} \cap \{Z = z_j\}) = \sum_{i=1}^{m} m_{i,j}.$$

On voit ainsi que la connaissance des lois marginales ne permet pas de retrouver la loi du couple, puisque les marginales donnent les sommes des lignes et les sommes des colonnes, mais pas les nombres de la matrice.

Exemple : Soit Y, Z deux variables aléatoires indépendantes prenant les valeurs 0 et 1 avec probabilité $1/2$.

La matrice associée à la loi de (Y, Z) est

$$\begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{pmatrix}$$

La matrice associée à la loi de (Y, Y) est

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

On voit que les deux vecteurs aléatoires admettent les mêmes marginales, mais pas les mêmes lois de couple.

Chapitre 5

Variables aléatoires à densité

On dit que X est une variable aléatoire à densité s'il existe une fonction f positive et continue par morceaux sur $\mathbb{R}$ telle que

$$\forall t \in \mathbb{R} \quad F_X(t) = \int_{-\infty}^{t} f(x)\ dx.$$

On dit que f est la densité de (la loi de) la variable aléatoire

5.1 Propriétés

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

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

5.2 Lois usuelles

5.2.1 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

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

5.2.2 Loi gaussienne de paramètres m et σ^2

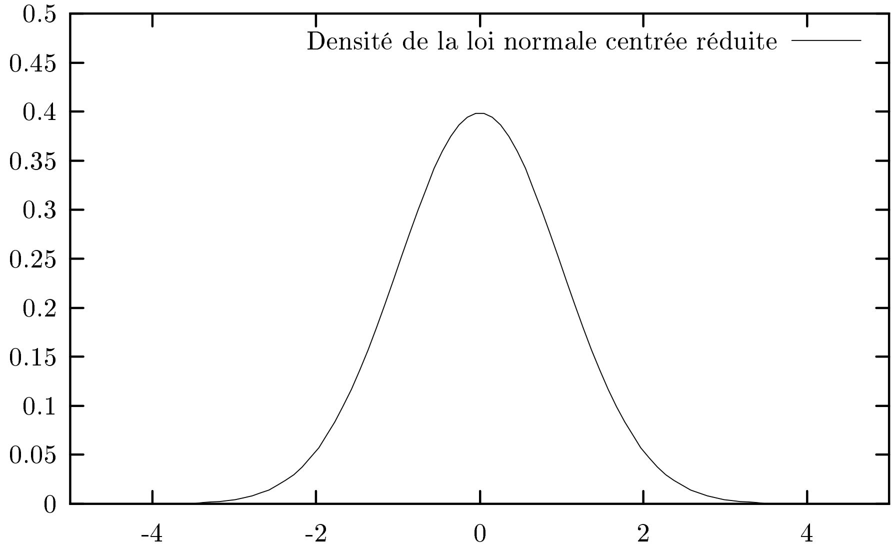
Soit $m \in \mathbb{R}$ et $\sigma^2 > 0$. On dit qu'une variable aléatoire X suit la loi gaussienne $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ^2 si elle admet la densité

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

On emploie également parfois le mot "normale" à la place de "gaussienne" : ces deux mots signifient exactement la même chose.

On dit qu'une variable gaussienne est centrée lorsque $m = 0$.
On dit qu'une variable gaussienne est centrée lorsque $\sigma^2 = 1$.

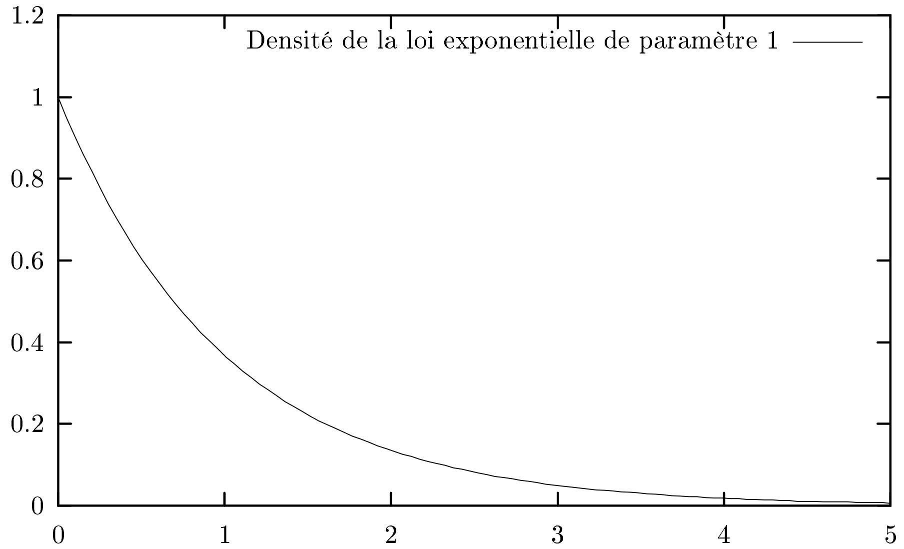
Remarques importantes : Si $X \sim \mathcal{N}(m, \sigma^2)$, alors $aX + b \sim \mathcal{N}(m + b, a^2\sigma^2)$. En particulier, si $X \sim \mathcal{N}(m, \sigma^2)$, alors $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$; et si $X \sim \mathcal{N}(0, 1)$, alors $\sigma X + m \sim \mathcal{N}(m, \sigma^2)$.



5.2.3 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

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



Remarque : Si X suit une loi exponentielle de paramètre $a > 0$, alors pour tout $c > 0$, cX suit une loi exponentielle de paramètre $\frac{a}{c}$.

5.3 Vecteurs aléatoires de $\mathbb{R}^2$ à densité

Soit $X = (X_1, X_2)$ un vecteur aléatoire. On dit que X est un vecteur aléatoire à densité s'il existe une fonction $f(x_1, x_2)$ positive et continue par morceaux sur $\mathbb{R}$ telle que

$$\forall (t_1, t_2) \in \mathbb{R}^2 \quad P(\{X_1 \leq t_1\} \cap \{X_2 \leq t_2\}) = \int_{-\infty}^{t_1} \Big(\int_{-\infty}^{t_2} f(x, y)\, dy \Big)\, dx.$$

On dit que f est la densité (de la loi) du vecteur aléatoire

5.4 Propriétés

Soit $X = (X_1, X_2)$ une variable aléatoire de densité f. f a les propriétés suivantes :

- $\forall a, b, c, d \in \mathbb{R} \quad a \leq b; c \leq d \Longrightarrow \quad \int_a^b \big(\int_c^d f(x, y)\, dy \big)\, dx = P(\{a \leq X_1 \leq b\} \cap \{c \leq X_2 \leq d\})$

5.5 Vecteurs à densité et indépendance

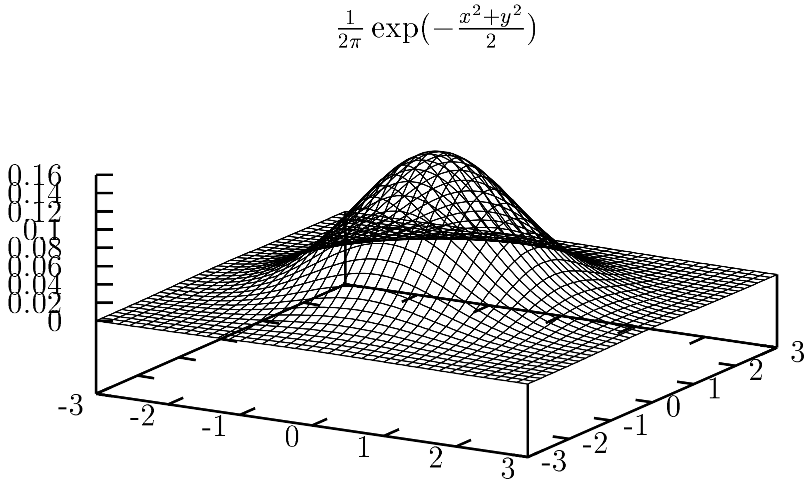
Théorème 25. *Soit X_1, X_2 deux variables aléatoires indépendantes, admettant chacune une densité. On note f_1 la densité de X_1 et f_2 la densité de X_2.*

Alors $X = (X_1, X_2)$ est un vecteur aléatoire à densité de densité :

$$(x_1, x_2) \mapsto f_1(x_1) f_2(x_2)$$

La preuve est laissée au lecteur.

La figure ci-dessous représente la densité d'un couple de variables aléatoires gaussiennes centrées réduites indépendantes.



5.6 Densité des lois marginales

Théorème 26. *Soit $X = (X_1, X_2)$ un vecteur aléatoire à densité de densité : $g(x_1, x_2)$. Alors la variable aléatoire X_1 admet la densité*

$$d_1(x_1) = \int_{-\infty}^{+\infty} g(x_1, x_2)\ dx_2,$$

et la variable aléatoire X_2 admet la densité

$$d_2(x_2) = \int_{-\infty}^{+\infty} g(x_1, x_2)\ dx_1.$$

5.7 Loi conditionnelle

Soit $X = (X_1, X_2)$ un vecteur aléatoire à densité de densité : $g(x_1, x_2)$. On note comme précédemment d_1 la densité de X_1 et d_2 la densité de X_2.

Soit $x_2 \in \mathbb{R}$ tel que $d_2(x_2) > 0$. On appelle loi conditionnelle de X_1 sachant $X_2 = x_2$ la loi sur $\mathbb{R}$ de densité

$$x_1 \mapsto \frac{f(x_1, x_2)}{d_2(x_2)}.$$

Ainsi, on note

$$P(X_1 \in [a,b] | X_2 = x_2) = \frac{\int_a^b g(x_1, x_2)\ dx_2}{\int_{-\infty}^{+\infty} g(x_1, x_2)\ dx_2},$$

$$P(X_1 \geq a | X_2 = x_2) = \frac{\int_a^{+\infty} g(x_1, x_2)\ dx_2}{\int_{-\infty}^{+\infty} g(x_1, x_2)\ dx_2},$$

$$P(X_1 \leq b | X_2 = x_2) = \frac{\int_{-\infty}^{b} g(x_1, x_2)\ dx_2}{\int_{-\infty}^{+\infty} g(x_1, x_2)\ dx_2}.$$

Comme d'habitude avec les lois à densité, on peut remplacer les inégalités larges par des inégalités strictes sans rien changer.

5.8 Convolution

Le but de cette section est de déterminer la loi de la somme de deux variables aléatoires indépendantes dont l'une au moins est à densité.

5.8.1 Convolution d'une loi discrète et d'une loi à densité

On s'intéresse ici au cas où l'une des deux loi est à densité, et l'autre discrète. Afin d'éviter des difficultés techniques, on se limitera au cas où la variable discrète ne prend qu'un nombre fini de valeurs.

Théorème 27. *Soient X et Y deux variables aléatoires indépendantes sur $(\Omega, \mathcal{F}, P)$ vérifiant*

- *Y est une variable de densité f.*

– $X(\Omega) = \{x_1, \ldots, x_n\}$ *et* $\forall i \in \{1, \ldots, n\} \quad P(X = x_i) = p_i$.

On pose $Z = X + Y$*. Alors,* Z *est une variable aléatoire de densité*

$$x \mapsto \sum_{i=1}^{n} p_i f(x - x_i).$$

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

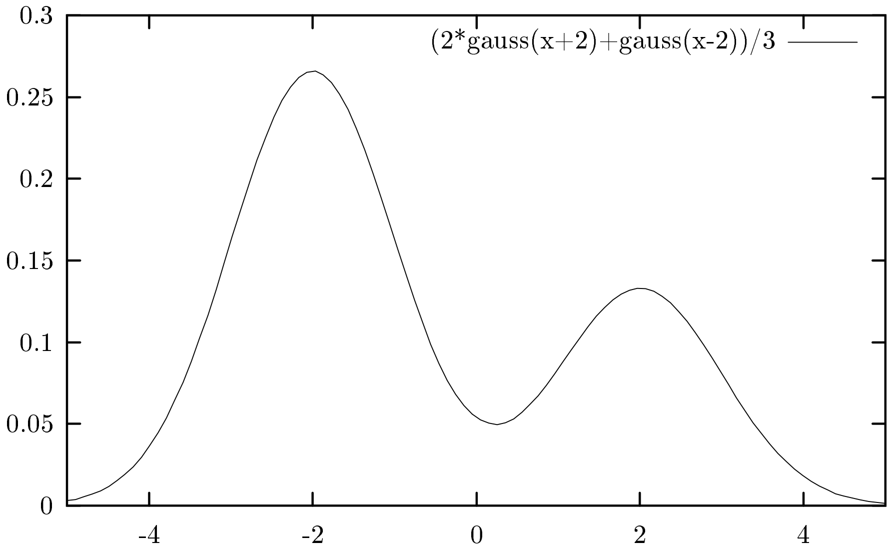
$$\begin{aligned}
P(Z \leq t) &= \sum_{i=1}^{n} P(\{Z \leq t\} \cap \{X = x_i\}) = \sum_{i=1}^{n} P(\{X + Y \leq t\} \cap \{X = x_i\}) \\
&= \sum_{i=1}^{n} P(\{Y + x_i \leq t\} \cap \{X = x_i\}) = \sum_{i=1}^{n} P(\{Y \leq t - x_i\} \cap \{X = x_i\}) \\
&= \sum_{i=1}^{n} P(X = x_i) P(Y \leq t - x_i) \\
&= \sum_{i=1}^{n} p_i \int_{-\infty}^{t - x_i} f(x)\, dx \\
&= \sum_{i=1}^{n} p_i \int_{-\infty}^{t} f(x - x_i)\, dx \\
&= \int_{-\infty}^{t} \sum_{i=1}^{n} p_i f(x - x_i)\, dx
\end{aligned}$$

Comme l'égalité est valide pour tout $t \in \mathbb{R}$, on en déduit que Z admet $x \mapsto \sum_{i=1}^{n} p_i f(x - x_i)$ comme densité. $\square$

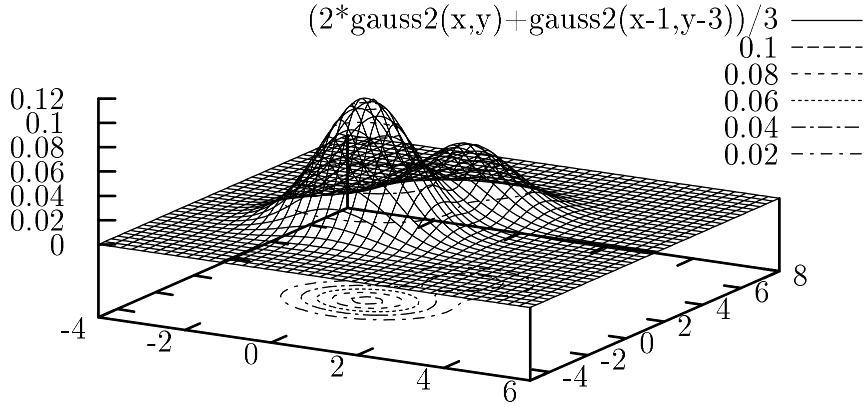
Exemple : Ci-dessous, le graphe de la densité de Z lorsque

– Y suit une loi normale centrée réduite

– X est telle que $P(X = -2) = 2/3$ et $P(X = 2) = 1/3$.



Convolution d'une loi gaussienne et d'une loi discrète



Je laisse le soin au lecteur de deviner deux lois sur $\mathbb{R}^2$ dont le graphe de la convolée est représenté ci-dessus – indication : l'une des deux est un couple de gaussiennes indépendantes.

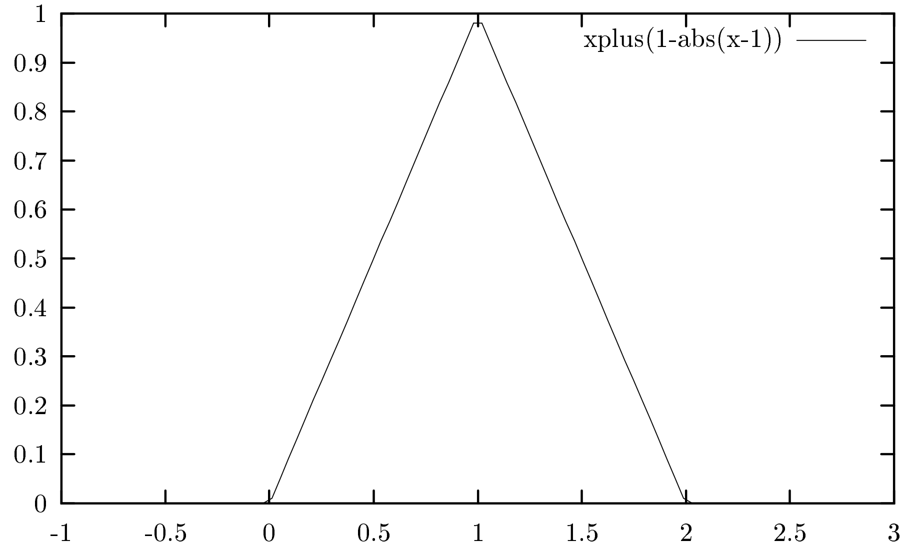
5.8.2 Convolution de deux lois à densité

On admettra le résultat suivant :

Théorème 28. *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*

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

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



Chapitre 6

Moments, lois des grands nombres

6.1 Quelques indications sur la construction de l'espérance

Soit $(\Omega, \mathcal{F}, P)$. On note $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ l'ensemble des variables aléatoires X vérifiant

$$\exists a \geq 0 \quad P(|X| \leq a) = 1.$$

On admet le premier résultat suivant :

Théorème 29. *Il existe une unique application linéaire ϕ de $\mathcal{L}^\infty(\Omega, \mathcal{F}, P)$ dans $\mathbb{R}$ vérifiant*

- $\forall A \in \mathcal{F} \quad \phi(\mathbb{1}_A) = P(A)$.
- $\forall X \in \mathcal{L}^\infty \quad P(X \geq 0) = 1 \Longrightarrow \phi(X)$

Gardons provisoirement ϕ comme notation pour l'application définie ci-dessus. On dit qu'une variable aléatoire X est intégrable si elle vérifie

$$\begin{cases} \lim\limits_{n \to +\infty} \phi(\min(X^+, n)) < +\infty \\ \lim\limits_{n \to +\infty} \phi(\min(X^-, n)) < +\infty \end{cases}$$

où l'on a posé $X^+ = \max(X, 0)$ et $X^- = \max(-X, 0)$, de sorte que $X = X^+ - X^-$.

On note $\mathcal{L}^1(\Omega, \mathcal{F}, P)$ l'ensemble des variables aléatoires intégrables.

Alors, pour toute variable aléatoire intégrable X, on pose

$$\mathbb{E}X = \lim_{n \to +\infty} \phi(\min(X^+, n)) - \lim_{n \to +\infty} \phi(\min(X^-, n)).$$

$\mathbb{E}X$ est appelée espérance de X.

6.2 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.
- $\forall X \in \mathcal{L}^1 \quad P(X \geq 0) = 1 \Longrightarrow \mathbb{E}X \geq 0$
- $\forall X \in \mathcal{L}^1 \quad P(X \leq a) = 1 \Longrightarrow \mathbb{E}X \leq a$.
- $\forall X \in \mathcal{L}^1 \quad P(X \geq b) = 1 \Longrightarrow \mathbb{E}X \geq b$.
- $\forall X \in \mathcal{L}^1 \quad P(X = a) = 1 \Longrightarrow \mathbb{E}X = a$.
- $\forall X \in \mathcal{L}^1 \quad P(|X| \leq a) = 1 \Longrightarrow \mathbb{E}|X| \leq a$.
- Soient X, Y deux variables aléatoires vérifiant $0 \leq X \leq Y$. Si Y est intégrable, alors X est intégrable.
- $\forall X, Y \in \mathcal{L}^1 \quad P(X \leq Y) = 1 \Longrightarrow \mathbb{E}X \leq \mathbb{E}Y$.

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

Théorème 30. *Soit X est une variable aléatoire discrète, et D un ensemble D 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

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

Il admet la généralisation suivante, appelée *théorème de transfert*.

Théorème 31. *Soit X est une variable aléatoire discrète, et D un ensemble D 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

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

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

Théorème 32. *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

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

Voici maintenant le théorème de transfert pour les variables à densité

Théorème 33. *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

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

Remarque importante : Si la densité de X est paire, 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 $f(x) = f(-x)$. □

6.5 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 3. *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 1. *$\mathcal{L}^2(\Omega, \mathcal{F}, P)$ 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. □

6.5.1 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

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

On appelle variance de X le nombre

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

On appelle écart-type de X le nombre

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

On a les propriétés suivantes

- Covar $(X, Y) =$ Covar (Y, X).
- $\forall a, b \in \mathbb{R}$ Covar $(X - a, Y - b) =$ Covar (X, Y).

- Var $(X+Y) = \text{Var } X + \text{Var } Y + 2\text{Covar } (X,Y)$.
- Covar $(X,Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$.
- Var $X = \mathbb{E}X^2 - (\mathbb{E}X)^2$.
- $|\mathbb{E}XY|^2 \leq \mathbb{E}X^2\mathbb{E}Y^2$ (inégalité de Cauchy-Schwarz)
- $|\text{Covar } (X,Y)| \leq \sigma(X)\sigma(Y)$.

Démonstration. – Évident.

- Si l'on pose $\overline{X} = X - \mathbb{E}X$, on a Covar $(X,Y) = \mathbb{E}\overline{X}\,\overline{Y}$.
  Or $\overline{X-a} = X-a-\mathbb{E}(X-a) = X-a-\mathbb{E}X+\mathbb{E}a = X-a-\mathbb{E}X+a = \overline{X}$.
  De même pour Y. La formule s'ensuit.
- Var $(X+Y) = \text{Covar } (X+Y, X+Y) = \mathbb{E}\overline{X+Y}^2$. Il est facile de voir que $\overline{X+Y} = \overline{X} + \overline{Y}$.
  On a donc

$$\begin{aligned}
\text{Var } X+Y = \mathbb{E}(\overline{X}+\overline{Y})^2 &= \mathbb{E}\overline{X}^2 + \overline{Y}^2 + 2\overline{XY} \\
&= \text{Covar } (X,X) + \text{Covar } (Y,Y) + 2\text{Covar } (X,Y) \\
&= \text{Var } X + \text{Var } Y + 2\text{Covar } (X,Y).
\end{aligned}$$

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

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

- Il suffit d'appliquer la formule précédente avec $X = Y$.
- Considérons, pour $t \in \mathbb{R}$:

$$f(t) = \mathbb{E}(tX+Y)^2.$$

  Comme $(tX+Y)^2 \geq 0$, on a $f(t) \geq 0$ pour tout t. Or

$$f(t) = t^2\mathbb{E}X^2 + 2t\mathbb{E}XY + \mathbb{E}Y^2$$

  est un polynôme du second degré. Comme il ne change pas de signe, son discriminant réduit est négatif :

$$\Delta' = b'^2 - ac = (\mathbb{E}XY)^2 - \mathbb{E}X^2\mathbb{E}Y^2 \leq 0,$$

  d'où le résultat.

– Il suffit d'appliquer l'inégalité de Cauchy-Schwarz à $\overline{X}$ et $\overline{Y}$, puis de prendre la racine carrée.

□

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

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

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

6.5.2 Espérance et indépendance

On admettra le théorème suivant très important :

Théorème 34. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors, leur produit XY est une variable aléatoire intégrable et l'on a*

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

Corollaire 2. *Soient X, Y deux variables aléatoires intégrables indépendantes. Alors X et Y ne sont pas corrélées.*

Démonstration. On a 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 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 3. *Soient X, Y deux variables aléatoires indépendantes de carré intégrable. Alors on a*

$$\text{\textit{Var}}\ (X + Y) = \text{\textit{Var}}\ X + \text{\textit{Var}}\ Y.$$

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

6.6 Calcul des premiers moments des lois discrètes usuelles

6.6.1 Indicatrice d'un événement

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

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

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 $\mathbb{1}_A^2 = \mathbb{1}_A$. Maintenant, on a

- $\mathbb{E}\mathbb{1}_A = P(A)$.
-

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

6.6.2 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 $A_1, \dots, A_n$ 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

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

6.6.3 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 $\mathrm{Var}\ X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{1}{p} + \frac{2(1-p)}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

6.6.4 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 $\mathrm{Var}\ X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

6.7 Calcul des premiers moments des lois à densité usuelles

6.7.1 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 $\mathrm{Var}\ X = \mathbb{E}X^2$.

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

- $\mathbb{E}Y = \mathbb{E}\frac{a+b}{2} + \frac{b-a}{2}\mathbb{E}X = \frac{a+b}{2}$.

- $\text{Var } Y = (\frac{b-a}{2})^2 \text{Var } X = \frac{(b-a)^2}{12}$

6.7.2 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 $\text{Var } X = \mathbb{E}X^2 = 1$.

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 $\text{Var } Y = \sigma^2 \text{Var } X = \sigma^2$.

6.7.3 Lois exponentielles

Lemme 4. *Soit X une variable aléatoire suivant une loi exponentielle de paramètre 1 et f une fonction croissante de classe C^1 de $\mathbb{R}_+$ dans $\mathbb{R}_+$ telle que $f'(X)$ soit intégrable et que $\lim\limits_{t \to +\infty} e^{-t} f(t) = 0$.*
Alors $f(X)$ est intégrable et on a la formule l'intégration par parties suivante :

$$\mathbb{E}f(X) = f(0) + \mathbb{E}f'(X)$$

Démonstration. Il suffit d'écrire

$$\begin{aligned}\int_0^A e^{-t}f(t)\ dt &= \big[-e^{-t}f(t)\big]_0^A + \int_0^A e^{-t}f'(t)\ dt \\ &= -e^{-A}f(A) + f(0) + \int_0^A e^{-t}f'(t)\ dt\end{aligned}$$

En faisant tendre A vers $+\infty$, on obtient

$$\int_0^{+\infty} e^{-t}f(t)\ dt = f(0) + \int_0^{+\infty} e^{-t}f'(t)\ dt,$$

soit

$$\mathbb{E}f(X) = f(0) + \mathbb{E}f'(X)$$

(Remarquer l'importance du fait que $f'(X)$ et $f(X)$ sont des variables aléatoires positives). □

On applique d'abord le lemme à $f(x) = x$: on obtient $\mathbb{E}X = 0 + \mathbb{E}1 = 1$. Ensuite on l'applique à $f(x) = x^2$: on obtient $\mathbb{E}X^2 = 0 + \mathbb{E}2X = 2\mathbb{E}X = 2$. On a donc $\mathrm{Var}\ X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 2 - 1^2 = 1$.

Passons au cas général. Soit $a > 0$. La variable aléatoire $Y = \frac{1}{a}X$ suit la loi exponentielle de paramètre a.
On a donc $\mathbb{E}Y = \frac{1}{a}\mathbb{E}X = \frac{1}{a}$ et $\mathrm{Var}\ Y = \frac{1}{a^2}\mathrm{Var}\ X = \frac{1}{a^2}$.

6.8 Lois des grands nombres

6.8.1 Inégalité de Markov

Théorème 35. *Soit X une variable aléatoire positive, intégrable. Alors, on a*

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

Démonstration. On a les inégalités

$$\forall \omega \in \Omega \quad a1\!\!1_{\{X \ge a\}}(\omega) \le X(\omega)1\!\!1_{\{X \ge a\}}(\omega) \le X(\omega).$$

Démontrons la première inégalité :

– si $X(\omega) \ge a$, on a $1\!\!1_{\{X \ge a\}}(\omega) = 1$, et donc l'inégalité à démontrer est exactement l'hypothèse $X(\omega) \ge a$.
– si $X(\omega) < a$, on a $1\!\!1_{\{X \ge a\}}(\omega) = 0$, et donc l'inégalité à démontrer se réduit à $0 \le 0$.

Montrer la deuxième inégalité se résume à remarquer que $0 \leq 1\!\!1_{\{X \geq a\}}(\omega) \leq 1$. On prend alors l'espérance et on obtient

$$\mathbb{E}a1\!\!1_{\{X \geq a\}} \leq \mathbb{E}X.$$

Comme $\mathbb{E}a1\!\!1_{\{X \geq a\}} = a\mathbb{E}1\!\!1_{\{X \geq a\}} = aP(X \geq a)$, on obtient l'inégalité désirée. □

6.8.2 Inégalité de Tchebytchef

Théorème 36. *Soit X une variable aléatoire admettant un moment d'ordre 2. Alors, on a*

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

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 $\mathbb{E}Y = \text{Var } X$, l'inégalité s'ensuit. □

6.8.3 Loi faible des grands nombres

Théorème 37. *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

$$S_n = \sum_{k=1}^{n} X_k \ et\ M_n = \frac{1}{n} S_n.$$

Alors

1. $\mathbb{E}|M_n - \mathbb{E}X_0|^2 \to 0$. *On dit que M_n converge en moyenne quadratique vers $\mathbb{E}X_0$.*
2. $\forall \varepsilon > 0 \quad P(|M_n - \mathbb{E}X_0| \geq \varepsilon) \to 0$. *On dit que M_n converge en probabilité vers $\mathbb{E}X_0$.*

Démonstration. 1. $\mathbb{E}M_n = \frac{1}{n}\mathbb{E}S_n = \frac{1}{n}\sum_{k=1}^{n} EX_k = \frac{1}{n}n\mathbb{E}X_0 = \mathbb{E}X_0$. Par conséquent $|M_n - \mathbb{E}X_0|^2 = \text{Var } M_n = \frac{1}{n^2}\text{Var } S_n$. Comme les X_k sont 2 à 2 non corrélées, on a

$$\text{Var } S_n = \sum_{k=1}^{n} \text{Var } X_k = n\text{Var } X_1.$$

On a donc

$$\mathbb{E}|M_n - \mathbb{E}X_0|^2 = \text{Var } M_n = \frac{\text{Var } X_1}{n}, \tag{6.1}$$

qui tend bien vers zéro.

2. D'après l'inégalité de Tchebitchev, on a

$$P(|M_n - \mathbb{E}X_0| \geq \varepsilon) \leq \frac{\text{Var } M_n}{\varepsilon^2},$$

qui tend vers zéro d'après (6.1).

□

6.8.4 Premier lemme de Borel-Canteli

Définition : Soit $(A_n)_{n\geq 1}$ une suite d'événements observables. On pose

$$\overline{\lim_{n\to+\infty}} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k.$$

Théorème 38. *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 $P(\overline{\lim_{n\to+\infty}} A_n) = 0$. □

6.8.5 Une loi forte des grands nombres

Théorème 39. *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes de même loi, admettant un moment d'ordre 4 (c'est à dire que $\mathbb{E}X_1^4$ est fini). On pose*

$$S_n = \sum_{k=1}^{n} X_k \text{ et } M_n = \frac{1}{n} S_n.$$

Alors il existe un événement observable $\tilde{\Omega} \subset \Omega$ avec $P(\tilde{\Omega} \subset \Omega) = 1$ et

$$\forall \omega \in \tilde{\Omega} \subset \Omega \quad M_n(\omega) \to \mathbb{E}X_1.$$

On dit alors que M_n converge presque sûrement vers $\mathbb{E}X_1$.

Démonstration. On va prouver le théorème dans le cas où $\mathbb{E}X_1 = 0$. Le cas général se déduit par translation.

Pour $\varepsilon > 0$, on pose $D_\varepsilon = \overline{\lim_{n\to+\infty}} \{|M_n| \geq \varepsilon\}$. On veut utiliser le lemme de Borel-Canteli pour montrer que $P(D_\varepsilon) = 0$. Il suffit donc de montrer que la série de terme général $P(|M_n| \geq \varepsilon)$ converge.
Or

$$P(|M_n| \geq \varepsilon) = P(|S_n| \geq n\varepsilon) = P(|S_n^4| \geq n^4\varepsilon^4).$$

D'après l'inégalité de Markov, on obtient

$$P(|M_n| \geq \varepsilon) \leq \frac{\mathbb{E}S_n^4}{n^4\varepsilon^4}.$$

Il s'agit donc maintenant d'estimer $\mathbb{E}S_n^4$.

$$\begin{aligned}
S_n^4 &= (X_1 + \cdots + X_n)^4 \\
&= \sum_{\gamma \in \{1,\ldots,n\}^4} \prod_{k=1}^{n} X_k \\
&= \left\{ \begin{array}{rrlrrl}
 & M(4) & \sum_{1\leq i\leq n} X_i^4 & + & M(1,3) & \sum_{1\leq i<j\leq n} X_i^3 X_j \\
+ & M(2,2) & \sum_{1\leq i<j\leq n} X_i^2 X_j^2 & + & M(2,1,1) & \sum_{1\leq i<j<k\leq n} X_i^3 X_j X_k \\
+ & M(1,1,1,1) & \sum_{1\leq i<j<k<l\leq n} X_i X_j X_k X_l & , & &
\end{array} \right.
\end{aligned}$$

où $M(i_1, \ldots, i_p)$ désigne de suites de 4-uplets $\gamma \in \{1, \ldots, n\}^4$ prenant i_1 fois la valeur v_1, … , i_p fois la valeur v_p.

On a $M(4)=1, M(1,3)=3, M(2,2)=6, M(1,1,1,1)=24$. (Plus généralement $M(i_1,\dots,i_p)=\frac{(i_1\cdots+i_p)!}{i_1!\dots i_p!}$. Ce sont les coefficients multinomiaux). En utilisant la linéarité de l'espérance, puis l'indépendance des X_i, on a

$$\mathbb{E}S_n^4 = \begin{cases} & \sum_{1\le i\le n} \mathbb{E}X_i^4 & + \quad 3 \sum_{1\le i<j\le n} \mathbb{E}X_i^3\mathbb{E}X_j \\ + \quad 6 & \sum_{1\le i<j\le n} \mathbb{E}X_i^2\mathbb{E}X_j^2 & + \quad 12 \sum_{1\le i<j<k\le n} \mathbb{E}X_i^3\mathbb{E}X_j\mathbb{E}X_k \\ + \quad 24 & \sum_{1\le i<j<k<l\le n} \mathbb{E}X_i\mathbb{E}X_j\mathbb{E}X_k\mathbb{E}X_l & , \end{cases}$$

Comme les X_k sont centrées, on a

$$\begin{aligned} \mathbb{E}S_n^4 &= \sum_{1\le i\le n} \mathbb{E}X_i^4 + 6 \sum_{1\le i<j\le n} \mathbb{E}X_i^2\mathbb{E}X_j^2 \\ &= n\mathbb{E}X_1^4 + 6\binom{n}{2}(\mathbb{E}X_1^2)^2 \\ &= n\mathbb{E}X_1^4 + 3n(n-1)(\mathbb{E}X_1^2)^2 \\ &\le Mn + 3Mn(n-1) \\ &\le 3Mn^2, \end{aligned}$$

où l'on a posé $M=\max((\mathbb{E}X_1^2)^2, \mathbb{E}X_1^4)$. [1] On a donc

$$P(|M_n|\ge\varepsilon) \le \frac{\mathbb{E}S_n^4}{n^4\varepsilon^4} \le \frac{3M}{n^2\varepsilon^4}.$$

Comme la série de terme général $\frac{3M}{n^2\varepsilon^4}$ est convergente, le lemme de Borel-Canteli permet d'affirmer que $P(D_\varepsilon)=0$ Posons maintenant

$$D = \overset{+\infty}{\underset{n=1}{\cup}} D_{1/n}.$$

On a $P(D)=0$ (exercice laissé au lecteur !) Maintenant on pose $\tilde{\Omega}=D^c$. On alors $P(\tilde{\Omega})=1-P(D)=1$.

Maintenant, un peu de réflexion devrait vous convaincre que pour tout $\omega\in\tilde{\Omega}$, $M_n(\omega)$ tend vers zéro en l'infini. (Cet effort de réflexion, même s'il peut être long, sera rentable en analyse et en probabilités !) □

[1] Montrer que $M=(\mathbb{E}X_1^2)^2$.

6.8.6 Probabilités et fréquences asymptotiques

Théorème 40. *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 précédent. X_k admet bien un moment d'ordre 4 car $0 \leq X_k \leq 1$ et l'on a $\mathbb{E}X_1 = P(A_1) = p$. □