
PROBABILITÉS

variables aléatoires discrètes et à densité

Licence de Mathématiques 2ème année

Jean-Christophe BRETON

Université de LA ROCHELLE

Janvier–Mai 2010

Table des matières

1	Langages ensembliste et probabiliste	3
1.1	Opérations entre ensembles	3
1.2	Vocabulaire probabiliste	4
1.3	Dénombrement	6
1.4	Rappel sur les séries	8
2	Mesures de probabilité	13
2.1	Espace de cardinal fini	13
2.2	Espaces infinis dénombrables (par exemple \mathbb{N}, \mathbb{Z})	14
2.3	Espace Ω général	16
3	Indépendance et conditionnement	21
3.1	Conditionnement	21
3.2	Indépendance	25
4	Variables aléatoires discrètes	31
4.1	Variables aléatoires discrètes	31
4.1.1	Définition	31
4.1.2	Loi d'une variable aléatoire discrète	32
4.1.3	Fonction de répartition	33
4.2	Lois discrètes classiques	34
4.2.1	Lois de v.a. finies déjà connues	34
4.2.2	Lois Géométriques	36
4.2.3	Loi de Poisson	37
5	Moment des variables aléatoires discrètes	41
5.1	Espérance d'une v.a.	41
5.1.1	Définitions	41
5.1.2	Espérances classiques	42
5.1.3	Propriétés de l'espérance	42
5.2	Variance d'une va	46

6	Variables aléatoires à valeurs réelles réelles	53
6.1	Généralités	53
6.2	Intégrales impropres	56
6.3	Variables aléatoires réelles à densité	56
6.4	Lois à densité classiques	59
6.4.1	Lois uniformes	60
6.4.2	Lois exponentielles	61
6.4.3	Lois de Cauchy	62
6.4.4	Lois normales ou gaussiennes	62
6.4.5	Lois log-normales	63
6.5	Espérance et variance des lois à densité	64
6.6	Tableau comparatif	68
7	Vecteurs aléatoires	71
7.1	Généralités	71
7.2	Vecteurs aléatoires discrets	72
7.3	Intégrales multiples	73
7.4	Vecteurs aléatoires réels à densité	75
7.5	Variables aléatoires indépendantes	77
7.6	Lois conditionnelles	82
7.6.1	Cas discret	82
7.6.2	Cas continu : densité conditionnelle	83
8	Somme de v.a. indépendantes	85
8.1	Somme de deux v.a. indépendantes	85
8.2	Convergences probabilistes	91
8.3	Loi des grands nombres	93
8.3.1	Loi faible des grands nombres	93
8.3.2	Lemme de Borel-Cantelli	96
8.3.3	Loi forte des grands nombres	97
8.4	Théorème central limite	100

Introduction

Dans la vie courante, il existe de nombreuses expériences dont le résultat n'est pas connu avec certitude. C'est l'objet de la théorie des probabilités que de fournir des modèles mathématiques permettant l'étude d'expériences dont le résultat n'est pas connu ou ne peut pas être prévu avec une totale certitude. Par exemple :

Expérience	Résultat observable
Lancer d'un dé	Un entier $k \in \{1, \dots, 6\}$
Prélèvement de n objets en sortie d'une chaîne de production	Nombre d'objets défectueux dans l'échantillon $\in \mathbb{N}^*$
Questionnaire à 100 questions binaires	Suite ω de 100 réponses $\omega \in \{\text{oui}, \text{non}\}^{100}$
Lancer d'une pièce jusqu'à l'obtention d'un pile	Un entier $k \in \mathbb{N}$: le temps d'attente du premier succès
Mise en service d'une ampoule	Durée de vie $T \in \mathbb{R}_+$
Lancer d'une flèche sur une cible	Point d'impact $M \in \mathbb{R}^2$
Mouvement (Brownien) d'un grain de pollen dans un liquide	Une fonction continue : la trajectoire $x \in \mathcal{C}(\mathbb{R})$
Mélange de deux gaz	Répartition spatiale de deux types de molécules

Le résultat précis de ces expériences n'est en général pas prévisible. Toutefois, l'observation et/ou l'intuition amènent souvent à penser que certaines règles semblent vérifier.

Par exemple si on jette 6000 fois un dé à 6 faces, on s'attend à ce que le nombre d'apparitions de faces « 4 » soit voisin de 1000. De même, si on met en service 100 téléviseurs du même modèle, leurs durées de vie observées seront concentrées autour d'une valeur moyenne.

Lorsqu'un phénomène se répète à l'infini avec des réalisations indépendantes et identiques, ses effets cummulés ont une distribution qui s'approche toujours de la même loi : une loi normale.

La théorie des probabilités permet de donner un sens précis à ces règles.

De façon générale, la théorie des probabilités modélise des situations concrètes et permet de calculer les probabilités d'évènement.

En aval des probabilités, il y a les *statistiques*. Ils se chargent de confronter les modèles probabilistes à la réalité observée pour les valider ou les invalider.

Les statistiques s'occupent par exemple de questions du genre :

- si à un examen sous forme de 100 questions avec réponses binaires, un étudiant a 60 bonnes réponses, est-il légitime de considérer qu'il a fait mieux que répondre au hasard ?

- Si sur un échantillon de 1000 personnes, un sondage donne un candidat à une élection crédité de 54% des voix, peut-on en déduire raisonnablement son élection ?

Dans ce cours, nous verrons les outils probabilistes de base pour calculer des probabilités d'évènements.

Nous définirons les lois classiques et nous étudierons leurs utilisations. Les cas des variables aléatoires discrètes et des variables aléatoires à densité sont traités.

Dans une deuxième partie, nous considérerons les vecteurs aléatoires et les sommes de variables aléatoires (indépendantes).

Nous terminerons avec la loi des grands nombres et le théorème central limite qui sont les premiers résultats fondamentaux des Probabilités.

Chapitre 1

Langages ensembliste et probabiliste

La théorie moderne des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous commençons donc par quelques rappels sur les opérations usuelles entre les ensembles.

1.1 Opérations entre ensembles

Soit Ω un ensemble de base. Considérons A et B deux sous ensembles de Ω .

Un élément ω appartient à A s'écrit $\omega \in A$.

L'ensemble A est inclus dans l'ensemble B s'écrit $A \subset B$.

L'ensemble des points de B qui ne sont pas dans A se note $B \setminus A$.

L'ensemble de tous les points qui ne sont pas dans A est le complémentaire de A , il est noté $A^c = \Omega \setminus A$.

L'ensemble vide \emptyset est l'ensemble qui ne contient aucun point, il s'agit du complémentaire de tout l'espace $\emptyset = \Omega^c = \Omega \setminus \Omega$.

La réunion $A \cup B$ de A et de B est l'ensemble des points qui sont dans A ou dans B .

L'intersection $A \cap B$ de A et de B est l'ensemble des points qui sont dans A et dans B .

Deux ensembles A et B sont dits disjoints si leur intersection est vide $A \cap B = \emptyset$.

Proposition 1.1.1 • *Le complémentaire d'une réunion ou d'une intersection est donné par*

$$(A \cap B)^c = A^c \cup B^c \quad \text{et} \quad (A \cup B)^c = A^c \cap B^c.$$

• *Le complémentaire du complémentaire d'un ensemble est cet ensemble : $(A^c)^c = A$.*

Démonstration : Exercice

Rappelons enfin que de façon générale, pour montrer l'égalité $A = B$ de deux ensembles A et B , il faut (et il suffit de) voir la double inclusion

$$A \subset B \quad \text{et} \quad B \subset A.$$

C'est à dire, montrer que pour tout $\omega \in A$, on a $\omega \in B$ et de la même façon pour tout $\omega' \in B$, on a $\omega' \in A$. On peut éventuellement le faire en une seule étape si on raisonne par équivalence : il faut alors montrer que $\omega \in A$ est équivalent à $\omega \in B$.

1.2 Vocabulaire probabiliste

Dans la suite, l'ensemble de base Ω va nous permettre de décrire une expérience aléatoire. Cet ensemble va représenter l'ensemble des résultats possibles de l'expérience (aléatoire) que l'on étudie. Nous l'appellerons *l'univers des possibles* ou *espace probabilisé*. Les parties de Ω seront appelés des évènements (ou évènements composés), les élément $\omega \in \Omega$ seront les évènements élémentaires, c'est à dire les évènements les plus simples qui ne peuvent pas être exprimés par des évènements encore plus simples.

Exemple : On lance un dé à six face. Le résultat *a priori* est aléatoire et les résultats possibles sont 1, 2, 3, 4, 5, 6. L'espace $\Omega = \{1, 2, 3, 4, 5, 6\}$ décrit bien l'ensemble des résultats. La partie $A = \{1, 4\}$ est un évènement composé : il s'agit de « le résultat est un 1 ou un 4 ». Par contre $\{3\}$ est un évènement élémentaire, « observer un 3 » ne peut pas être décrit par des évènements plus simples.

Avec ce mode de représentation, les opérations logiques sur les évènements que sont « ou », « et », « négation » se traduisent par des opérations ensemblistes : réunion \cup , intersection \cap , complémentaire c . Voici le tableau des correspondances entre ces deux langages :

Notations	Vocabulaire ensembliste	Vocabulaire probabiliste
\emptyset	ensemble vide	évènement impossible
Ω	ensemble plein	évènement certain
ω	élément de Ω	évènement élémentaire
A	sous-ensemble de Ω	évènement
$\omega \in A$	ω appartient à A	le résultat ω est une des réalisations possibles de A
$A \subset B$	A inclus dans B	A implique B
$A \cup B$	réunion de A et B	A ou B
$A \cap B$	intersection de A et B	A et B
A^c	complémentaire de A dans Ω	évènement contraire de A
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

Remarque 1.2.1 Il faut retenir que

- une réunion \cup s'interprète comme un « ou »,
- une intersection \cap s'interprète comme un « et »,
- un complémentaire c s'interprète comme « le contraire de ».

Notez enfin que en mathématiques le « ou » est un ou inclusif alors que dans le langage usuel il s'agit d'un ou exclusif (dessert ou fromage ? c'est l'un ou l'autre mais pas les deux alors qu'avec le « ou » mathématiques, ça pourrait être les deux).

Les opérations sur les ensembles (ou sur les évènements) peuvent faire intervenir plus de deux évènements. Ainsi si A_1, \dots, A_n sont des évènements,

$$\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n$$

est l'ensemble des ω qui sont dans au moins un des A_i . De même

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$$

est l'ensemble des ω qui sont dans tous les A_i . On étend encore ces définitions aux réunions et intersections dénombrables (*i.e.* en nombre infini mais qu'on peut énumérer) :

$$\bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i=1}^{+\infty} A_i = \{\text{réalisation d'au moins un } A_i\},$$

$$\bigcap_{i \in \mathbb{N}} A_i = \bigcap_{i=1}^{+\infty} A_i = \{\text{réalisation de tous les } A_i\}.$$

Rappel (dénombrabilité) : une partie infinie est dénombrable si elle peut être mise en bijection avec \mathbb{N} , c'est à dire si on peut énumérer tous ses éléments. L'ensemble \mathbb{N} , bien sûr, est dénombrable mais \mathbb{Z} , \mathbb{Q} le sont aussi. Par contre $[0, 1]$ ou \mathbb{R} ne le sont pas.

Comme on peut énumérer aussi les éléments d'une partie finie, il est usage d'inclure le cas fini dans le cas dénombrable, même si d'ordinaire, le terme dénombrable est utilisé pour les parties infinies dénombrables.

Ces opérations logiques sur des suites d'évènements sont très utiles pour analyser les évènements complexes : il s'agit de les réexprimer comme réunion, intersection, complémentaire d'évènements plus simples. Il importe donc de bien traduire en langage ensembliste un énoncé et ses enchaînements logiques.

Voilà maintenant un exemple, utile dans de nombreuses situations, de « traduction » en langage ensembliste d'une assertion en français.

Proposition 1.2.1 *Soit $A_i, i \geq 0$, une collection infinie d'ensembles. Alors*

– *À partir d'un certain rang, ω est dans tous les A_i s'écrit*

$$\omega \in \bigcup_{i \geq 0} \bigcap_{j > i} A_j \quad (= \varliminf_i A_i).$$

– ω est dans une infinité de A_i s'écrit

$$\omega \in \bigcap_{i \geq 0} \bigcup_{j > i} A_j \quad (= \overline{\lim}_i A_i).$$

Démonstration :

• Pour le premier point : Soit ω qui, à partir d'un certain rang, est dans tous les A_i . On traduit cela de la façon suivante : il existe un rang i tel que pour tout rang $j > i$, ω est dans A_j . D'après la signification des symboles $\forall, \exists, \cap, \cup$, cela revient à écrire

$$\omega \in \underbrace{\bigcup_{i \geq 0}}_{\text{il existe } i \geq 0} \underbrace{\bigcap_{j > i}}_{\text{pour tout } j > i} \underbrace{A_j}_{\omega \text{ est dans } A_j}.$$

• Pour le second point, dire que ω est dans une infinité de A_i est équivalent à dire que

« pour tout p , il existe $q > p$ avec ω dans A_q . »

En effet, si tel est le cas, ω est bien dans une infinité de A_i car, d'après cette propriété,

- avec $p = 0$, il existe $p_1 > p$ tel que ω est dans A_{p_1}
- avec $p = p_1$, il existe $p_2 > p_1$ tel que ω est dans A_{p_2}
- avec $p = p_2$, il existe $p_3 > p_2$ tel que ω est dans A_{p_3}
- ...
- avec $p = p_n$, il existe $p_{n+1} > p_n$ tel que ω est dans $A_{p_{n+1}}$
- ...

et finalement, ω est dans chaque A_{p_n} , $n \in \mathbb{N}^*$, c'est à dire dans une infinité de A_i . Réciproquement, s'il est dans une infinité de A_i , alors pour tout p , on trouve $q > p$ tel que $\omega \in A_q$; sinon, ce serait qu'il existe p tel que pour $q > p$, ω n'est pas dans A_q . Ou encore : ω ne peut appartenir qu'aux A_i d'indice $i \leq p$, c'est à dire seulement à un nombre fini d'entre eux, ce qui est faux.

Donc, pour ce deuxième point, pour tout p , on trouve $q > p$, tel que $\omega \in A_q$, en langage \forall, \exists , cela s'écrit

$$\omega \in \underbrace{\bigcap_{p \geq 0}}_{\text{pour tout } p \geq 0} \underbrace{\bigcup_{q > p}}_{\text{il existe } q > p} \underbrace{A_q}_{\omega \text{ est dans } A_q}.$$

■

1.3 Dénombrement

Considérons un ensemble $\Omega = \{\omega_1, \dots, \omega_n\}$ de cardinal n .

• **Permutation**

Le nombre de permutations d'un ensemble est le nombre de manières d'ordonner ses éléments. Le nombre de permutations de Ω est $n! = 1 \times 2 \times 3 \times \dots \times n$.

En effet, il s'agit de trouver tous les reordonnements de $\{\omega_1, \dots, \omega_n\}$. On a d'abord n choix pour le premier terme, puis $n - 1$ pour le deuxième puis $n - 2$ puis \dots puis 2 choix pour l'avant dernier et enfin plus qu'un seul pour le dernier. Il y a donc $n \times (n - 1) \times (n - 2) \times \dots \times 2 \times 1 = n!$.

Exercice. Faire la preuve pour $n = 3$ et trouver les $3! = 6$ permutations de $\{A, B, C\}$.

Exemple. Un professeur doit faire passer dans la journée 5 étudiants pour un oral de rattrapage. Il a $5! = 120$ manières de choisir l'ordre de passage.

• **Tirage de p objets (avec remise) dans un ensemble de cardinal n .**

Pour chaque tirage, il y a n objets possibles à tirer, il y a donc en tout $n \times \dots \times n = n^p$ tirages possibles (avec remise) dans un ensemble de cardinal n .

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 manières de choisir de façon indépendante 30 éléments de l'ensemble $\{0, 1, \dots, 20\}$ de cardinal 21. Il y a donc 21^{30} résultats possibles pour l'ensemble de la classe.

• **Arrangement (nombre de tirages ordonnés sans remise).**

On appelle tirage sans remise de p éléments dans un ensemble Ω de cardinal n , tout tirage successif de p éléments de Ω , chaque élément ne pouvant être tiré plus d'une fois.

Bien évidemment, pour qu'un tel tirage puisse exister, il faut avoir $p \leq n$.

Le nombre de tirages sans remise est

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

Remarque 1.3.1 Le nombre $n!/(n - p)!$ s'appelle le nombre d'arrangements, on le note A_n^p . Lorsque $n = p$, on retrouve le nombre de permutations, puisqu'on tire tous les éléments de Ω et qu'en fait, on les a reordonnés.

Exemple. 3500 personnes se présentent au concours de l'agrégation de Mathématiques. 300 places sont mises au concours. Combien y-a-t-il de palmarès possibles (en supposant qu'il n'y ait pas d'*ex-aequo*) ?

Réponse : $3500 \times 3499 \times \dots \times 3202 \times 3201 = \frac{3500!}{3200!}$.

• **Combinaison (nombre de tirages désordonnés sans remise)**

Il s'agit du nombre de parties d'un ensemble Ω possédant p éléments.

C'est exactement le nombre de manières de choisir p objets dans un ensemble de n objets, l'ordre n'ayant pas d'importance.

On sait qu'il y a $n!/(n - p)!$ tirages de p objets lorsque l'on tient compte de l'ordre. Or un tirage (désordonné) donné (où l'ordre n'est pas pris en compte) représente $p!$ tirages où l'ordre est pris en compte (car il y a $p!$ permutations de l'ensemble des p objets du

tirage). Il y a donc $p!$ fois plus de tirages de p objets lorsque l'on tient compte de l'ordre. Finalement, le nombre de tirages (sans tenir compte de l'ordre) est

$$\frac{n!}{p!(n-p)!}.$$

Exemple. Dénombrer le nombre de tirages sans remise de 2 éléments parmi 4 avec ordre puis sans ordre.

Exemple. 3500 personnes se présentent au concours de l'agrégation de Mathématiques. 300 places sont mises au concours. Combien y-a-t-il de promotions possibles ?

Réponse : C_{3500}^{300} . Ici, Ω est l'ensemble des candidats et il s'agit de choisir 300 d'entre eux. On s'intéresse aux différentes promotions possibles, prises dans leur ensemble, sans tenir compte du classement de la promotion.

• Rappelons d'abord la définition des coefficients binomiaux et la formule du binôme de Newton :

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} \quad 0 \leq k \leq n, \quad (a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}.$$

C_n^k s'interprète comme le nombre d'échantillons de taille k dans une population de taille n . Par exemple, si dans une urne de n boules distinctes, on en tire k , il y a C_n^k tirages différents possibles.

Rappelons les propriétés immédiates suivantes pour tout $n \in \mathbb{N}^*$ et $k \leq n$:

$$C_n^k = C_n^{n-k}, \quad C_n^n = C_n^0 = 1, \quad C_n^{n-1} = C_n^1 = n \\ C_n^{k-1} + C_n^k = C_{n+1}^k \quad (\text{triangle de Pascal}).$$

Exercice. Prouver deux fois la formule du binôme de Newton :

- d'abord en utilisant l'interprétation en dénombrement des coefficients binomiaux,
- puis par récurrence sur n avec la relation du triangle de Pascal.

1.4 Rappel sur les séries

Dans le cadre qu'on se fixe dans ce cours (espaces discrets ou réels), les calculs de probabilités font intervenir plus d'outils que ceux de dénombrement (réservé au cadre fini, cf. Terminale ou première année de Licence). On a alors recours à des outils d'analyse qu'il faut connaître : séries numériques, intégrales impropres, quelques notions de développement en séries entières. Ces outils seront développés ultérieurement dans d'autres cours de Mathématiques. On se contente ici de les définir succinctement. Cette approche suffira pour l'utilisation dont on a besoin. Pour plus de précision, on renvoie aux cours concernés.

Séries numériques

Les séries numériques sont des sommes infinies de réels (ou de complexes). Généralement, il s'agit de la somme de tous les termes d'une suite réelle $(a_i)_{i \in \mathbb{N}}$. On la définit (lorsqu'elle existe) comme la limite quand n tend vers $+\infty$ de la somme des n premiers termes :

$$S = \sum_{i=0}^{+\infty} a_i = \lim_{n \rightarrow +\infty} \sum_{i=0}^n a_i. \quad (1.1)$$

Bien sûr, la somme infinie n'a pas toujours un sens (exemple, $\sum_{i=0}^{+\infty} (-1)^i$ n'existe pas car les sommes partielles valent 1 ou 0 selon la parité de n) ou si elle en a un, sa valeur peut être infinie.

Si la limite est bien définie et est finie, on parle alors de série convergente. Sinon on parle de série divergente.

Lorsque $a_i \geq 0$, la limite qui définit la série dans (1.1) existe toujours, seulement, elle peut être égale à $+\infty$.

Pour une suite $(a_i)_i$ de signe quelconque –ou complexe–, la série S est dite absolument convergente si la série $\sum_{i=1}^{+\infty} |a_i|$ des $|a_i|$ est finie. Lorsque tel est le cas, la série S est *a fortiori* convergente : la convergence absolue entraîne la convergence simple (la réciproque étant fausse).

On renvoie au cours d'Analyse 2 pour les différents critères de convergence des séries. On se contente de :

Séries de Riemann : La série de terme général $a_n = 1/n^\alpha$ converge ssi $\alpha > 1$.

Critère de Riemann : Soit $(a_n)_n$ une suite réelle (ou même complexe).

- S'il existe $\alpha > 1$ tel que $\lim_{n \rightarrow +\infty} n^\alpha |a_n| = 0$ alors la série $\sum_{n=0}^{+\infty} a_n$ converge (absolument).
- S'il existe $\alpha < 1$ tel que $\lim_{n \rightarrow +\infty} n^\alpha |a_n| = +\infty$ alors la série $\sum_{n=0}^{+\infty} a_n$ diverge (absolument).

Lorsque l'on considère une suite géométrique (du type $a_{i+1} = \rho a_i$), la somme S_n de ses $n + 1$ premiers termes est connue :

$$S_n = \sum_{i=0}^n a_i = a_0 \frac{1 - \rho^{n+1}}{1 - \rho}$$

si $\rho \neq 1$, sinon, $S_n = n + 1$ (valeur qu'on retrouve en faisant un d.l. à l'ordre 1 en 1 de la formule précédente). La série converge si $|\rho| < 1$ et alors la série vaut

$$S = \lim_{n \rightarrow +\infty} S_n = a_0 \frac{1}{1 - \rho}.$$

Ce type de série est le cas particulier d'autres séries qui définissent des fonctions : les séries entières.

Séries entières

Définition 1.4.1 *Étant donnée une suite $(a_n)_{n \in \mathbb{N}}$, la série entière associée est la série de terme général $u_n = a_n x^n$, c'est donc une fonction de x :*

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n.$$

Cette fonction n'est pas définie pour tout les x . Cependant, on a :

Proposition-définition *Il existe $R \in [0, +\infty]$ appelé rayon de convergence (nul dans le plus mauvais cas, infini dans le meilleur) de la série entière tel que*

- *la série qui définit $f(x)$ converge si $|x| < R$,*
- *la série diverge si $|x| > R$,*
- *pour $|x| = R$, le comportement (convergence ou divergence) dépend de la série étudiée.*

Exemples.

• Avec $a_n = 1$ pour tout $n \geq 0$, on obtient la série géométrique (de raison x) de rayon de convergence $R_1 = 1$:

$$f_1(x) = \sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}.$$

• Avec $a_n = (-1)^{n+1}/n$ pour tout $n \geq 1$, on obtient le développement en série entière de la fonction logarithme (en 1) de rayon de convergence $R_2 = +1$:

$$f_2(x) = \sum_{n=0}^{+\infty} \frac{(-1)^{n+1}}{n} x^n = \ln(1+x).$$

• Avec $a_n = 1/n!$ pour tout $n \geq 0$, on obtient le développement en série entière de la fonction exponentielle de rayon de convergence $R_3 = +\infty$:

$$f_3(x) = \sum_{n=0}^{+\infty} \frac{x^n}{n!} = e^x.$$

• Un cas pathologique : avec $a_n = n!$ pour tout $n \geq 0$, on obtient

$$f_4(x) = \sum_{n=0}^{+\infty} n! x^n$$

qui est de rayon nul. Donc en pratique, cette série n'existe pas (ou si peu : seulement pour $x = 0$).

Un résultat important concernant les séries entières est qu'on les dérive termes à termes sur leur domaine de convergence (*i.e.* pour $|x| < R$) :

$$f'(x) = \left(\sum_{n=0}^{+\infty} a_n x^n \right)' = \sum_{n=0}^{+\infty} a_n (x^n)' = \sum_{n=1}^{+\infty} n a_n x^{n-1} = \sum_{n=0}^{+\infty} (n+1) a_{n+1} x^n \quad \text{pour } |x| < R.$$

On intègre aussi les séries entières termes à termes sur le disque (ouvert) de convergence :

$$\int_0^t f(x) dx = \int_0^t \left(\sum_{n=0}^{+\infty} a_n x^n \right) dx = \sum_{n=0}^{+\infty} a_n \int_0^t x^n dx = \sum_{n=0}^{+\infty} \frac{a_n}{n+1} x^{n+1} = \sum_{n=1}^{+\infty} \frac{a_{n-1}}{n} x^n \quad \text{pour } |x| < R.$$

Ces résultats sont très utiles dans le calcul de séries entières.

Exemples. Retrouver les développements en séries entières en 0 de

$$\frac{1}{1+x}, \quad \ln(1+x), \quad \frac{1}{1+x^2}, \quad \arctan x.$$

Chapitre 2

Mesures de probabilité

2.1 Espace de cardinal fini

Dans le cas où l'espace Ω est fini $\{\omega_1, \dots, \omega_n\}$, la probabilité $\mathbb{P}(A)$ peut se définir pour tout sous ensemble $A \subset \Omega$ et la probabilité \mathbb{P} est donnée par une suite finie $(p_i)_{1 \leq i \leq n}$ qui est la suite des probabilités des évènements élémentaires $\omega_i : p_i = \mathbb{P}(\{\omega_i\})$.

Définition 2.1.1 Soit $\Omega = \{\omega_1, \dots, \omega_n\}$ un ensemble fini à n éléments. On définit une probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ par la donnée d'une suite finie de réels positifs p_i de somme $\sum_{i=1}^n p_i = 1$ données par $\mathbb{P}\{\omega_i\} = p_i$. Pour tout $A \subset \Omega$, $\mathbb{P}(A)$ est alors donnée par

$$\mathbb{P}(A) = \sum_{i: \omega_i \in A} p_i.$$

On rappelle que $\mathcal{P}(\Omega)$ désigne l'ensemble des parties de Ω , et que $\text{card } \mathcal{P}(\Omega) = 2^{\text{card } \Omega}$. Ainsi, si $\Omega = \{a, b, c\}$ alors

$$\mathcal{P}(\Omega) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}.$$

Une mesure de probabilités \mathbb{P} est une fonction définie sur l'ensemble des parties de Ω . Formellement, on a

$$\mathbb{P} : \begin{cases} \mathcal{P}(\Omega) & \rightarrow [0, 1] \\ A & \mapsto \mathbb{P}(A) \end{cases}$$

Remarque 2.1.1 On constate facilement que \mathbb{P} satisfait les propriétés suivantes :

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- si $A \cap B = \emptyset$, alors

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

et plus généralement si A_1, A_2, \dots, A_p sont 2 à 2 disjoints, alors

$$\mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_p) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_p) \quad (\text{propriété de fini-additivité}).$$

Exemple : On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir face au premier lancer et pile au dernier ?

On modélise la situation en prenant $\Omega = \{p, f\}^3$ où p désigne pile et f face. \mathbb{P} est définie sur l'ensemble de toutes les parties de Ω , $\mathcal{P}(\Omega)$. Il y a $8 = 2^3$ triplets de résultats possibles :

$$(p, p, p), (p, p, f), (p, f, p), (f, p, p), (f, f, p), (f, p, f), (p, f, f), (f, f, f).$$

Si on suppose la pièce bien équilibrée, *a priori* chacun de ces triplets est équiprobable (i.e. a la même probabilité de survenir). Comme la somme de leur probabilité doit faire 1, chacun a pour probabilité $1/8$. L'évènement A cherché se décompose en :

$$(f, f, p), (f, p, p).$$

D'où $\mathbb{P}(A) = \mathbb{P}\{(f, f, p), (f, p, p)\} = \mathbb{P}\{(f, f, p)\} + \mathbb{P}\{(f, p, p)\} = 1/8 + 1/8 = 1/4$.

Lorsque l'espace est fini, la probabilité la plus simple est l'*équiprobabilité* qui donne la même probabilité à chaque évènement élémentaire ω_i . C'est la probabilité qui est utilisée en général sur les espaces Ω finis lorsque rien n'est précisé.

Exemple. • Lorsqu'on lance un dé équilibré à 6 faces, il est naturel de choisir l'équiprobabilité qui associe la probabilité $1/6$ à chaque face 1, 2, 3, 4, 5, 6.

- Au loto, la probabilité de chaque boule est $1/49$.

Exercice. On jette deux fois deux dés. Quelle est la probabilité d'avoir au moins une fois au moins un six ?

2.2 Espaces infinis dénombrables (par exemple \mathbb{N}, \mathbb{Z})

Lorsque l'espace Ω est infini dénombrable, la probabilité se définit encore sur tout $\mathcal{P}(\Omega)$: à chaque sous-ensemble A de Ω , une probabilité est associée. Elle est donnée à nouveau par la suite infinie des probabilités des évènements élémentaires ω_i :

Définition 2.2.1 Soit $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ un ensemble infini dénombrable. Une probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ est définie par la donnée d'une suite infinie de réels positifs p_i de somme $\sum_{i=1}^{+\infty} p_i = 1$, donnés par $\mathbb{P}\{\omega_i\} = p_i$. Pour tout $A \subset \Omega$, $\mathbb{P}(A)$ est alors donnée par

$$\mathbb{P}(A) = \sum_{i: \omega_i \in A} p_i.$$

Remarque 2.2.1 Il revient au même dans le cas discret fini ou infini d'indexer à partir de 0 ou de 1. Dans le cas fini, se méfier cependant si l'indexation va de 0 à n que le cardinal de l'ensemble Ω est $n + 1$.

On constate encore que \mathbb{P} satisfait les propriétés suivantes :

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
- si $A \cap B = \emptyset$, alors

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

et plus généralement si $A_1, A_2, \dots, A_n, \dots$ est une suite **infinie** d'évènements 2 à 2 disjoints, alors

$$\mathbb{P}\left(\bigcup_{i=1}^{+\infty} A_i\right) = \mathbb{P}(A_1 \cup \dots \cup A_n \cup \dots) = \sum_{i=1}^{+\infty} \mathbb{P}(A_i) \quad (\text{propriété de } \sigma\text{-additivité}).$$

Noter que l'équiprobabilité n'existe plus lorsque l'espace est dénombrable non fini : si on accorde la même probabilité p à chaque ω_i , par σ -additivité, on doit avoir

$$1 = \mathbb{P}(\Omega) = \mathbb{P}\left(\bigcup_{i=1}^{+\infty} \{\omega_i\}\right) = \sum_{i=1}^{+\infty} \mathbb{P}\{\omega_i\} = \sum_{i=1}^{+\infty} p;$$

la somme est alors infinie si $p > 0$ et elle est nulle si $p = 0$, ce qui dans les deux cas est absurde.

Exemple de probabilité définie sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. Soit $a \in \mathbb{R}_+^*$, posons

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

$(p_k)_{k \geq 0}$ est une suite de terme positifs de somme

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

La probabilité d'une partie A de \mathbb{N} est alors donnée par

$$\mathbb{P}(A) = \sum_{k \in A} \frac{e^{-a} a^k}{k!}.$$

Par exemple la probabilité de l'ensemble des nombres pairs $2\mathbb{N} = \{2n, n \in \mathbb{N}\}$ est

$$\mathbb{P}(2\mathbb{N}) = \sum_{k=0}^{+\infty} \frac{e^{-a} a^{2k}}{2k!} = e^{-a} \cosh(a) = e^{-a} \frac{e^a + e^{-a}}{2} = \frac{1 + e^{-2a}}{2}$$

et celle des nombres impairs est $\mathbb{P}(2\mathbb{N} + 1) = 1 - \mathbb{P}(2\mathbb{N}) = \frac{1 - e^{-2a}}{2}$.

On verra au chapitre prochain qu'il s'agit de la loi de Poisson de paramètre a .

2.3 Espace Ω général

On l'a vu dans les cas précédents : la probabilité \mathbb{P} est une fonction qui à un ensemble A associe un nombre compris entre 0 et 1, sa probabilité $\mathbb{P}(A)$. C'est donc une fonction d'ensembles – c'est à dire sur les ensembles. Cette fonction \mathbb{P} doit vérifier un certain nombre de propriétés (poids total égale à 1, σ -additivité). Pour à la fois les satisfaire et être définie avec cohérence, on ne peut pas définir, en général, la probabilité de tous les sous-ensembles $A \subset \Omega$. \mathbb{P} n'est donc pas en général définie sur tout $\mathcal{P}(\Omega)$, l'ensemble de tous les sous-ensembles de Ω . On doit se restreindre à une famille d'évènements $\mathcal{F} \subset \mathcal{P}(\Omega)$ qu'on appellera *famille des évènements observables* (en L3, on parlera de *tribu* ou de σ -algèbre).

On définit alors la probabilité \mathbb{P} sur l'ensemble des évènements observables \mathcal{F} :

Définition 2.3.1 Soient Ω un ensemble, \mathcal{F} une famille d'observables sur Ω . On appelle *probabilité sur (Ω, \mathcal{F})* toute application \mathbb{P} de \mathcal{F} dans $[0, 1]$ qui vérifie :

(i) $\mathbb{P}(\Omega) = 1$,

(ii) (*Propriété de σ -additivité*) Pour toute suite $(A_i)_{i \in \mathbb{N}^*}$ d'observables, deux à deux disjoints, on a

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

On appelle $(\Omega, \mathcal{F}, \mathbb{P})$ un *espace probabilisé* ou *espace de probabilité*.

Remarque 2.3.1 – En pratique, pour vérifier que \mathbb{P} est une probabilité, on se contentera de vérifier que $\mathbb{P}(A) \geq 0$ pour tout observable A , que $\mathbb{P}(\Omega) = 1$ et que \mathbb{P} est additive : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ quand A et B sont disjoints.

- Lorsque l'espace Ω est discret (c'est à dire fini ou dénombrable, par exemple \mathbb{N} ou une partie de \mathbb{N}), tous les ensembles sont observables et on peut choisir $\mathcal{F} = \mathcal{P}(\Omega)$. C'est la raison pour laquelle cette restriction aux familles d'observables n'est jamais apparue lors de cours de Probabilités en espaces finis ou discrets.
- Lorsque l'espace est \mathbb{R} , pour le choix de la famille d'observables, on peut se restreindre aux ensembles qui sont des intervalles.

Exemples. Soit $f(x) = \frac{1}{2}e^{-|x|}$, montrer que

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

définit une mesure de probabilité sur les intervalles de \mathbb{R} .

Finalement, par *observables*, on pourra se contenter de comprendre dans ce cours : n'importe quel ensemble si l'espace Ω est discret et les intervalles si l'espace est \mathbb{R} .

Propriétés des probabilités

Une probabilité satisfait un certain nombre de propriétés de base qu'il faut connaître.

Toute probabilité \mathbb{P} sur (Ω, \mathcal{F}) vérifie les propriétés suivantes :

- $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

En effet $\Omega = A \cup A^c$ avec une réunion disjointe. Par additivité, on a donc

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c).$$

- $\mathbb{P}(\emptyset) = 0$.

En effet $\emptyset = \Omega^c$ donc $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0$.

- Additivité (cas particulier du point (ii) de la définition d'une probabilité) :
 - Si $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$,
 - Si les A_i ($1 \leq i \leq n$) sont deux à deux disjoints,

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

- $\forall A, B \in \mathcal{F}, A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.

En effet $B = (B \setminus A) \cup A$ où la réunion est disjointe. On a donc

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

- $\forall A, B \in \mathcal{F}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

En effet $A \cup B = (B \setminus A) \cup (A \cap B) \cup (A \setminus B)$ où les ensembles sont deux à deux disjoints.

On a donc

$$\mathbb{P}(A \cup B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B). \quad (2.1)$$

Or $A = (A \setminus B) \cup (A \cap B)$ avec une réunion d'ensembles disjoints donc

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

Et de même $B = (B \setminus A) \cup (A \cap B)$ avec une réunion d'ensembles disjoints donc

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

On a donc $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$ et $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$, ce qui conclut en reportant dans (2.1).

- $\forall A \in \mathcal{F}, \forall B \in \mathcal{F}, \mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$,
- $\forall A_1, A_2, \dots, A_n \in \mathcal{F}, \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n)$,
- $\forall A_1, A_2, \dots, A_n, \dots \in \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) \leq \sum_{i=1}^{+\infty} \mathbb{P}(A_i).$$

En effet cela suit pour une réunion de deux ensembles $A \cup B$ du point précédent. Le cas d'une réunion dénombrable est une simple généralisation.

- Propriété de continuité monotone séquentielle

(i) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite croissante d'évènements (i.e. pour tout n $A_n \subset A_{n+1}$) alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}(A) \text{ où } A = \bigcup_{n \in \mathbb{N}^*} A_n. \quad (2.2)$$

(ii) Si $(B_n)_{n \in \mathbb{N}^*}$ est une suite décroissante d'évènements (i.e. pour tout n $B_{n+1} \subset B_n$) alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}(B) \text{ où } B = \bigcap_{n \in \mathbb{N}^*} B_n. \quad (2.3)$$

En effet dans le cas croissant, on a $\bigcup_{k=1}^n A_k = A_n$ et donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(\bigcup_{k=1}^n A_k) = \mathbb{P}(\bigcup_{k=1}^{+\infty} A_k).$$

De même dans le cas décroissant, on a $\bigcap_{k=1}^n A_k = A_n$ et donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \mathbb{P}(\bigcap_{k=1}^n A_k) = \mathbb{P}(\bigcap_{k=1}^{+\infty} A_k).$$

Remarque 2.3.2 En général, on ne peut pas calculer $\mathbb{P}(A \cup B)$ à partir de $\mathbb{P}(A)$ et de $\mathbb{P}(B)$ comme le montre la formule $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$: il faut connaître $A \cap B$, on verra que ceci est lié à l'indépendance ou non des évènements A et B .

Attention, cette formule ne se généralise pas immédiatement pour plus de deux évènements, par exemple pour A, B, C , on a :

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

Plus généralement, on a le résultat suivant (admis) :

Proposition 2.3.1 (Formule de Poincaré) *Pour tout entier $n \geq 2$, et tous évènements A_1, A_2, \dots, A_n , on a :*

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

Définition 2.3.2 *On appelle système complet d'évènements toute suite $A_1, A_2, \dots, A_n, \dots$ d'évènements deux à deux disjoints et dont la somme des probabilités vaut 1 :*

$$\sum_i \mathbb{P}(A_i) = 1.$$

Proposition 2.3.2 (Formule des probabilités totales – version 1) *Étant donné un système complet $\{A_1, A_2, \dots, A_n, \dots\}$, pour tout évènement B , sa probabilité peut se décomposer de la façon suivante :*

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B \cap A_i) = \mathbb{P}(B \cap A_1) + \mathbb{P}(B \cap A_2) + \dots + \mathbb{P}(B \cap A_i) + \dots \quad (2.4)$$

Démonstration : Notons $\Omega_0 = \bigcup_i A_i$, il s'agit d'un évènement de probabilité 1 (par définition de $\{A_i\}_i$ système complet). Observons que les ensembles $B \cap A_i$, $i = 1 \dots, n \dots$, sont deux à deux disjoints : $(B \cap A_i) \cap (B \cap A_j) \subset A_i \cap A_j = \emptyset$ si $i \neq j$. Par σ -additivité, on a maintenant

où la dernière égalité vient de :

$$\begin{aligned} \mathbb{P}(B) - \mathbb{P}(B \cap \Omega_0) &= \mathbb{P}(B \setminus (B \cap \Omega_0)) \\ &= \mathbb{P}(B \cap (\Omega \setminus \Omega_0)) \\ &= \mathbb{P}(B \cap \Omega_0^c) \\ &\leq \mathbb{P}(\Omega_0^c) \\ &= 1 - \mathbb{P}(\Omega_0) \\ &= 1 - 1 = 0. \end{aligned}$$

■

Concrètement le calcul de $\mathbb{P}(B)$ par (2.4) consiste à calculer la probabilité de B en distinguant selon les différents cas possibles dans l'espace Ω (A_1, \dots, A_n, \dots) et à faire la somme obtenue des probabilités dans ces différents cas pour réobtenir la probabilité totale $\mathbb{P}(B)$.

Chapitre 3

Indépendance et conditionnement

3.1 Conditionnement

Le conditionnement a pour objet de répondre à la question suivante : comment se modifie la probabilité d'un évènement lorsque l'on connaît déjà une information supplémentaire ?

Exemple. On choisit au hasard deux chiffres entre 1 et 9. Sachant que la somme obtenue est paire, calculer la probabilité p pour que les deux chiffres soient impairs.

Dans la suite, on fixe un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 3.1.1 Soit B un évènement de probabilité non nulle $\mathbb{P}(B) \neq 0$. Pour tout évènement observable A , on définit la probabilité conditionnelle de A sachant B :

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

L'intérêt de cette notion vient du fait que souvent, compte tenu des informations disponibles dans un problème, il est plus facile d'attribuer une valeur à la probabilité conditionnelle $\mathbb{P}(A|B)$ que de calculer $\mathbb{P}(A \cap B)$ ou $\mathbb{P}(A)$.

Exemple. Une urne U_1 contient 9 boules dont 5 rouges, une urne U_2 contient 5 boules dont 3 rouges. On choisit une urne au hasard et on tire une boule de cette urne. Si cette boule est rouge, calculer la probabilité pour que la boule tirée vienne de l'urne U_1 .

Remarque 3.1.1 – Il importe de bien comprendre la différence entre $\mathbb{P}(A|B)$, la probabilité que A se réalise sachant que B est réalisé et $\mathbb{P}(A \cap B)$, la probabilité que A et B se réalisent simultanément.

- L'écriture $A|B$ ne désigne pas un nouvel ensemble et ne veut rien dire isolément. Seul $\mathbb{P}(A|B)$ a une signification. En fait, on ne modifie pas l'ensemble A mais on change la probabilité en prenant $\mathbb{P}(\cdot|B)$ comme nouvelle probabilité, comme le montre la proposition suivante :

Proposition 3.1.1 Soit dans un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, $B \in \mathcal{F}$ un évènement observable fixé tel que $\mathbb{P}(B) > 0$. Alors la fonction d'ensemble $\mathbb{P}(\cdot|B)$:

$$A \in \mathcal{F} \longrightarrow \mathbb{P}(A|B)$$

est une nouvelle probabilité sur (Ω, \mathcal{F}) .

De ce fait, on dispose pour les probabilités conditionnelles de toutes les propriétés d'une probabilité vues à la proposition ??.

Notons que pour *chaque* observable B de probabilité non nulle, on définit une probabilité conditionnelle $\mathbb{P}(\cdot|B)$. Ce ne sont pas les mêmes probabilités conditionnelles pour des évènements B et B' différents.

On dispose par ailleurs des propriétés suivantes propres aux conditionnements :

Proposition 3.1.2 (Règle des conditionnements successifs) Si A_1, \dots, A_n sont n évènements observables tels que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$, alors

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

Application : Quand $B \subset C$, on a on a pour tout évènement A :

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

Démonstration : Notons d'abord que pour $1 \leq i \leq n-1$, $\bigcap_{j=1}^{n-1} A_j \subset \bigcap_{j=1}^i A_j$ donc

$$0 < \mathbb{P}\left(\bigcap_{j=1}^{n-1} A_j\right) \leq \mathbb{P}\left(\bigcap_{j=1}^i A_j\right)$$

si bien qu'on peut conditionner par $\bigcap_{j=1}^i A_j$ pour tout $1 \leq i \leq n-1$. On a alors par simplifications en cascade.

$$\begin{aligned} & \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \times \dots \times \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}) \\ &= \mathbb{P}(A_1) \times \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \times \frac{\mathbb{P}(A_1 \cap A_2 \cap A_3)}{\mathbb{P}(A_1 \cap A_2)} \times \dots \times \frac{\mathbb{P}(A_1 \cap \dots \cap A_n)}{\mathbb{P}(A_1 \cap \dots \cap A_{n-1})} \\ &= \mathbb{P}(A_1 \cap \dots \cap A_n). \end{aligned}$$

■

Définition 3.1.2 Une famille finie ou non d'ensembles $(A_i)_{i \in I}$ est une partition de Ω si
 – les ensembles sont deux à deux disjoints : $\forall i \neq j, A_i \cap A_j = \emptyset$,

– leur réunion est $\Omega : \Omega = \bigcup_{i \in I} A_i$.

La partition est dite finie, infinie, dénombrable si I est un ensemble d'indices respectivement fini, infini, dénombrable.

On peut aussi calculer la probabilité d'un évènement en conditionnant par tous les cas possibles, c'est l'objet de la proposition suivante :

Proposition 3.1.3 (Formule des probabilités totales – version 2)

– Si $B \in \mathcal{F}$ est tel que $\mathbb{P}(B) \neq 0$ et 1 alors

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c).$$

– Si $B_1, \dots, B_n \in \mathcal{F}$ forment une partition finie de Ω en évènements de probabilités non nulles alors

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) = \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \dots + \mathbb{P}(A|B_n)\mathbb{P}(B_n).$$

– Si $B_1, \dots, B_n, \dots \in \mathcal{F}$ forment une partition dénombrable de Ω en évènements avec $\mathbb{P}(B_i) > 0$ pour tout $i \in \mathbb{N}^*$, alors

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) = \sum_{i=1}^{+\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Démonstration : Les premier et deuxième points sont des cas particuliers du dernier qu'on prouve : comme $(B_i)_{i \in \mathbb{N}^*}$ est une partition de Ω , on a

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in \mathbb{N}^*} B_i \right) = \bigcup_{i \in \mathbb{N}^*} A \cap B_i,$$

cette réunion est disjointe car les B_i le sont :

$$(A \cap B_i) \cap (A \cap B_j) \subset B_i \cap B_j = \emptyset \text{ si } i \neq j.$$

Par σ -additivité de \mathbb{P} , il suit :

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i \in \mathbb{N}^*} A \cap B_i \right) = \sum_{i=1}^{+\infty} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{+\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

■

Exercice. Revenons à l'exemple précédent avec deux urnes où il y a des boules rouges. Calculer la probabilité de tirer une boule rouge.

Au passage, notons le résultat utile suivant qui souligne que pour calculer une probabilité, il suffit parfois de diviser les cas :

Proposition 3.1.4 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, $(B_i)_{i \in \mathbb{N}^*}$ une partition de Ω . Alors, on a

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) = \sum_{i=1}^{+\infty} \mathbb{P}(A \cap B_i).$$

Lorsque l'on sait calculer les probabilités conditionnelles $\mathbb{P}(A|B_i)$ pour tout un système de partition $(B_i)_{i \in I}$, on peut chercher les probabilités conditionnelles avec les conditionnements inverses $\mathbb{P}(B_i|A)$. Elles sont données par :

Proposition 3.1.5 (Formule de Bayes) Soient A un évènement observable de probabilité non nulle et $(B_i)_{i \in I}$ une partition de Ω en évènements de probabilités non nulles. On a

$$\forall j \in I, \quad \mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Le résultat est vrai que I soit un ensemble d'indice fini ou infini dénombrable.

Démonstration : Par définition des probabilités conditionnelles :

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

Il ne reste plus qu'à développer $\mathbb{P}(A)$ par la formule des probabilités totale en

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

■

Exemple. Dans une population, chaque individu a une probabilité 0,2 d'être droitier. On pratique un test de latéralisation sur les individus de cette population. Un individu qui n'est pas droitier a une probabilité 0,6 d'échouer au test. Un individu qui est droitier a une probabilité 0,8 de réussir le test. On fait le test sur une personne. Quelle est la probabilité pour qu'elle ne soit pas droitière sachant que le test est positif ?

Notons D l'évènement « être droitier » et T l'évènement « réussir le test ».

L'énoncé indique que $\mathbb{P}(D) = 0,2$, $\mathbb{P}(T^c|D^c) = 0,6$ et $\mathbb{P}(T|D) = 0,8$. On cherche $\mathbb{P}(D^c|T)$.

Pour cela, on utilise la formule de Bayes (ou celle des probabilités totales)

$$\begin{aligned} \mathbb{P}(D^c|T) &= \frac{\mathbb{P}(D^c \cap T)}{\mathbb{P}(T)} = \frac{\mathbb{P}(T|D^c)\mathbb{P}(D^c)}{\mathbb{P}(T|D^c)\mathbb{P}(D^c) + \mathbb{P}(T|D)\mathbb{P}(D)} = \frac{(1 - 0,6) \times (1 - 0,2)}{(1 - 0,6) \times (1 - 0,2) + 0,8 \times 0,2} \\ &= \frac{0,4 \times 0,8}{0,4 \times 0,8 + 0,8 \times 0,2} = \frac{0,4}{0,4 + 0,2} = \frac{4}{4 + 2} = \frac{4}{6} = \frac{2}{3} \simeq 0,666. \end{aligned}$$

3.2 Indépendance

Il arrive que la connaissance d'un évènement influe sur celle d'un autre par exemple, quand on considère la population française, savoir {habiter près de la mer} influe sur la réalisation de {habiter à La Rochelle}. Parfois, ce n'est pas le cas : {parler une langue étrangère} n'influencera pas, *a priori*, sur {habiter à La Rochelle}. Dans ce dernier cas, on parle d'évènements indépendants.

C'est aussi le cas dans les situations suivantes :

- lors de tirages aléatoires avec remises dans une urne contenant des boules bleues et rouges ou
- lors de deux lancers successifs d'un dé.

Dans le premier cas, la couleur de la boule au premier tirage ne donne aucune information sur le tirage de la seconde.

Dans le cas d'un dé, l'obtention de l'as au premier lancer ne modifie pas la probabilité d'obtention d'un quatre, par exemple, au second lancer.

Voyons un autre exemple.

Exemple : Considérons une population de 100 étudiants composée de 60 étudiants en mathématiques et de 40 en informatique. On étudie deux caractéristiques de ces individus : être attiré par le métier d'enseignant et pratiquer régulièrement un sport. La répartition observée est la suivante :

- 31 matheux veulent devenir enseignant, 29 non,
- 10 informaticiens veulent devenir enseignant, 30 non.
- 24 matheux font du sport, 36 non,
- 16 informaticiens font du sport, 24 non,

Notons $A = \{\text{pratiquer un sport}\}$, $B = \{\text{être matheux}\}$ et $C = \{\text{vouloir devenir enseignant}\}$. On a

$$\mathbb{P}(A) = \frac{24 + 16}{100} = 0,40, \quad \mathbb{P}(B) = 0,60, \quad \mathbb{P}(C) = \frac{31 + 10}{100} = 0,41.$$

Par ailleurs, $A \cap B = \{\text{être matheux et pratiquer un sport}\}$ d'où $\mathbb{P}(A \cap B) = 0,24$.

Et $B \cap C = \{\text{être matheux et vouloir devenir enseignant}\}$ d'où $\mathbb{P}(B \cap C) = 0,31$.

On remarque alors que

$$\mathbb{P}(A \cap B) = 0,24 = 0,40 \times 0,60 = \mathbb{P}(A) \times \mathbb{P}(B). \quad (3.1)$$

et

$$\mathbb{P}(B \cap C) = 0,31 \neq \mathbb{P}(B) \times \mathbb{P}(C) = 0,60 \times 0,41 = 0,246 \quad (3.2)$$

Intuitivement, on comprend bien que la pratique d'un sport est sans rapport avec le fait d'être étudiant en mathématiques ou en informatique. Par contre, la spécialité influe sur l'attrait du métier d'enseignant : comme le nombre de postes au concours d'enseignants est plus important en mathématiques qu'en informatique, il est légitime que davantage d'étudiant en mathématiques envisage sérieusement cette carrière.

On dit alors que A et B sont indépendants tandis que B et C ne le sont pas.

Compte tenu de (3.1) et (3.2), la notion d'indépendance de deux évènements se définit de la façon suivante :

Définition 3.2.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Deux évènements observables A et B sont dits indépendants si

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

Exemple. Reprenons l'exemple du lancer de deux dés. L'espace à considérer est $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ muni de ces parties $\mathcal{P}(\Omega)$ et de d'équiprobabilité \mathbb{P} . Chaque tirage possible a la probabilité $1/36$ car $\#\Omega = 6^2 = 36$.

Soient A : « obtention de l'as au premier lancer » et B : « obtention du 4 au second lancer ».

Le tirage de l'as au premier lancer se décompose en les évènements élémentaires suivants : $\{1, 1\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}$ et est de probabilité

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(\{1, 1\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}) \\ &= \mathbb{P}\{1, 1\} + \mathbb{P}\{1, 2\} + \mathbb{P}\{1, 3\} + \mathbb{P}\{1, 4\} + \mathbb{P}\{1, 5\} + \mathbb{P}\{1, 6\} \\ &= 1/36 + 1/36 + 1/36 + 1/36 + 1/36 + 1/36 \\ &= 1/6. \end{aligned}$$

De même, l'obtention du 4 au second lancer se décompose en $\{1, 4\}, \{2, 4\}, \{3, 4\}, \{4, 4\}, \{5, 4\}, \{6, 4\}$ et est de probabilité

$$\mathbb{P}(B) = \mathbb{P}(\{1, 4\}, \{2, 4\}, \{3, 4\}, \{4, 4\}, \{5, 4\}, \{6, 4\}) = 6/36 = 1/6.$$

L'évènement $A \cap B$ désigne le tirage de l'as au premier et du 4 au second lancer, on a donc $A \cap B = \{1, 4\}$ et $\mathbb{P}(A \cap B) = \mathbb{P}\{1, 4\} = 1/36 = 1/6 \times 1/6 = \mathbb{P}(A) \times \mathbb{P}(B)$. Les évènements A « tirage de l'as au premier lancer » et B « tirage du 4 au second » sont donc indépendants dans $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$.

Intuitivement, les deux lancers sont indépendants donc leurs résultats le sont aussi.

Considérons maintenant C : « la somme des deux résultats est 6 ». L'évènement C se décompose en

$$C = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$$

et est donc de probabilité $\mathbb{P}(C) = 5/36 = 1/6$.

On a $A \cap C = \{(1, 5)\}$ de probabilité $\mathbb{P}(A \cap C) = 1/36 \neq \mathbb{P}(A)\mathbb{P}(C)$.

Les évènements A et C ne sont pas indépendants.

Intuitivement, si on a un 1 au premier lancer pour que finalement la somme des résultats fasse 6, il faut que le second lancer soit important, ce qui explique la non-indépendance de A et de C .

Exemple. On lance deux fois une pièce et on considère les évènements A : « obtenir deux fois le même résultat », B : « avoir une face au premier lancer » et C : « avoir au moins une face ».

En notant p pour pile et f pour face, on a facilement

$$A = \{ff, pp\}, \quad \mathbb{P}(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$\begin{aligned}
B &= \{ff, fp\}, & \mathbb{P}(B) &= \frac{1}{2} \\
C &= \{ff, fp, pf\}, & \mathbb{P}(C) &= \frac{3}{4} \\
A \cap B &= \{ff\}, & \mathbb{P}(A \cap B) &= \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbb{P}(A) \times \mathbb{P}(B) \\
A \cap C &= \{ff\}, & \mathbb{P}(A \cap C) &= \frac{1}{4} \neq \mathbb{P}(A) \times \mathbb{P}(B).
\end{aligned}$$

Les évènements A et B sont donc indépendants alors que A et C ne le sont pas.

Les évènements B et C ne sont pas indépendants car $B \subset C$ et donc $B \cap C = B$ ce qui empêche d'avoir $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$ car $\mathbb{P}(C) \neq 1$.

Remarques.

- Il faut faire attention à ne pas confondre « être indépendants » et « être disjoints ». En particulier deux évènements A et B disjoints ne peuvent pas être indépendants quand ils sont de probabilités non nulles.

C'est clair intuitivement : avoir une information sur A , c'est en avoir une sur B (si A se réalise alors par disjonction B ne peut pas se réaliser).

C'est clair aussi par le calcul car

$$\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0 \neq \mathbb{P}(A) \times \mathbb{P}(B).$$

- Il faut faire attention encore : l'indépendance de deux évènements A et B n'est pas intrinsèque mais dépend de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ utilisé (c'est à dire du choix du modèle) :

Exemple : Une urne contient 12 boules numérotées de 1 à 12. On en tire une au hasard et on considère : $A = \{\text{tirage d'un nombre pair}\}$, $B = \{\text{tirage d'un multiple de 3}\}$. Que dire de A et B ?

L'espace à considérer est $\Omega = \{1, 2, \dots, 12\}$ munie de l'équiprobabilité \mathbb{P} , chaque boule étant équiprobable. On a

$$A = \{2, 4, 6, 8, 10, 12\}, \quad B = \{3, 6, 9, 12\}, \quad A \cap B = \{6, 12\}.$$

On a $\mathbb{P}(A) = 6/12 = 1/2$, $\mathbb{P}(B) = 4/12 = 1/3$ et

$$\mathbb{P}(A \cap B) = \frac{2}{12} = \frac{1}{6} = \frac{1}{2} \times \frac{1}{3} = \mathbb{P}(A) \times \mathbb{P}(B).$$

Les évènements A et B sont indépendants dans l'espace $(\Omega, \mathcal{F}, \mathbb{P})$.

On rajoute maintenant une treizième boule. Que dire de A et B ?

Le modèle –c'est à dire l'espace de probabilité– change. Il faut désormais considérer $\Omega' = \{1, 2, \dots, 12, 13\}$ munie de la nouvelle equiprobabilité \mathbb{P}' . A , B gardent les mêmes descriptions mais leur probabilité sont $\mathbb{P}'(A) = 6/13$, $\mathbb{P}'(B) = 4/13$ et

$$\mathbb{P}'(A \cap B) = \frac{2}{13} \neq \frac{6}{13} \times \frac{4}{13} = \mathbb{P}'(A) \times \mathbb{P}'(B).$$

Les évènements A et B ne sont pas indépendants pour \mathbb{P}' .

Intuitivement : dans le premier cas, la proportion des multiples de 3 est la même chez les pairs et les impairs. Le fait de savoir que la boule tirée est paire ne modifie donc pas notre information sur B . Par contre dans le deuxième cas, la proportion des multiples de 3 est plus élevée chez les pairs que chez les impairs. Le fait de savoir que la boule tirée est paire augmente un peu la probabilité (conditionnelle) que nous pouvons attribuer à B .

Remarque 3.2.1 (Probabilités conditionnelles) La notion d'indépendance est évidemment liée à celle de conditionnement : rappelons que si B est de probabilité non nulle la probabilité conditionnelle de A sachant B est définie par

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

On a alors les équivalences pour A et B de probabilités non nulles :

- $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$ (i.e. A et B indépendants),
- $\mathbb{P}(A|B) = \mathbb{P}(A)$ (le conditionnement par B est sans effet sur A),
- $\mathbb{P}(B|A) = \mathbb{P}(B)$ (le conditionnement par A est sans effet sur B).

Ces équivalences découlent directement des définitions de l'indépendance et des probabilités conditionnelles.

Proposition 3.2.1 *Si A et B sont deux évènements observables indépendants, alors A^c et B , A et B^c , A^c et B^c le sont encore deux à deux.*

Démonstration : Montrons le pour A^c et B : comme

$$A^c \cap B = (\Omega \setminus A) \cap B = (\Omega \cap B) \setminus (A \cap B) = B \setminus (A \cap B),$$

on a :

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B). \end{aligned}$$

■

Exercice. Faire la preuve dans les autres cas.

La notion d'indépendance se généralise à plus de deux évènements avec quelques précautions. Pour le voir, analysons d'abord la situation suivante :

Exemple : On considère une urne qui contient quatre boules : une bleue, une blanche, une rouge et une tricolore. Considérons les évènements suivants

$$\begin{aligned} A &= \{\text{la boule tirée contient du bleu}\}, \\ B &= \{\text{la boule tirée contient du blanc}\}, \end{aligned}$$

$$C = \{\text{la boule tirée contient du rouge}\}.$$

Comme il y a quatre boules et qu'à chaque fois la boule unie de couleur concernée et la boule tricolore conviennent, on a facilement $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 2/4 = 1/2$. Par ailleurs $A \cap B = \{\text{tirage de la boule tricolore}\}$ et

$$\mathbb{P}(A \cap B) = \mathbb{P}\{\text{tirage de la boule tricolore}\} = 1/4 = 1/2 \times 1/2 = \mathbb{P}(A)\mathbb{P}(B).$$

Donc A, B sont indépendants, de même pour A et C , B et C . Les événements A, B, C sont donc *deux à deux* indépendants.

D'autre part $\mathbb{P}(A \cap B \cap C) = 1/4$ car $A \cap B \cap C = \{\text{tirage de la boule tricolore}\}$ et donc

$$\mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/2 \times 1/2 \times 1/2 = 1/8. \quad (3.3)$$

Intuitivement, on n'a pas indépendance de A, B, C ensemble et donc l'indépendance deux à deux ne suffit à généraliser à plus de deux événements l'indépendance globale. Ce qu'il manque pour avoir l'indépendance de A, B, C ensemble c'est la vérification de (3.3).

D'où la définition suivante : trois événements A, B, C sont dits mutuellement indépendants (càd indépendants quand on les considère ensemble et non pas seulement deux à deux) si la probabilité de toute intersection d'ensembles pris parmi A, B, C est le produit des probabilités de ces mêmes ensembles :

$$\begin{aligned} \mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B), \\ \mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C), \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C), \\ \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C). \end{aligned}$$

Plus généralement pour n événements :

Définition 3.2.2 *Les n événements observables A_1, \dots, A_n sont mutuellement indépendants si pour toute sous-famille A_{i_1}, \dots, A_{i_p} avec $1 \leq i_1 < \dots < i_p \leq n$, on a*

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_p}) = \mathbb{P}(A_{i_1}) \times \dots \times \mathbb{P}(A_{i_p}).$$

L'indépendance mutuelle implique l'indépendance deux à deux (prendre $p = 2$), la réciproque est fautive comme le justifie l'exemple précédent.

Le même type de remarque qu'à la Proposition 3.2.1 est valable : si un ou plusieurs événements est remplacé par son complémentaire, cela ne change rien à l'indépendance de la famille.

Définition 3.2.3 *Une suite infinie d'événements est dite indépendante si toute sous-famille finie est formée d'événements mutuellement indépendants.*

Souvent dans la suite nous parlerons d'une suite indépendante d'épreuves. Intuitivement, il est clair de quoi il s'agit ; formellement on a :

Définition 3.2.4 *On dit qu'une suite infinie d'épreuves est indépendante si toute suite $(A_i)_{i \geq 1}$ formée d'événements A_i dont la réalisation ne dépend que de la i ème épreuve forme une suite indépendante d'événements.*

Chapitre 4

Variables aléatoires discrètes

Jusqu'à maintenant, on a parlé d'évènements (aléatoires) et de leur probabilité. Or dans la plupart des expériences aléatoires, il s'agit plutôt d'observer un résultat (aléatoire aussi), en général numérique, que des évènements. Par exemple :

- le nombre de bonnes réponses dans un QCM,
- le temps d'obtention du premier six aux dés,
- la durée de vie d'une ampoule électrique.

Le résultat aléatoire est une fonction (au sens mathématique) du hasard $\omega \in \Omega$. L'information apportée par cette expérience est portée par cette fonction appelée variable aléatoire.

4.1 Variables aléatoires discrètes

4.1.1 Définition

Définition 4.1.1 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On appelle variable aléatoire discrète sur $(\Omega, \mathcal{F}, \mathbb{P})$ toute application X :

$$X : \Omega \longrightarrow \mathbb{R}, \quad \omega \longmapsto X(\omega)$$

telle que l'ensemble de ses images $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ est une partie au plus dénombrable de \mathbb{R} . On peut donc numéroter ses éléments par des indices entiers : $X(\Omega) = \{x_1, x_2, \dots, x_k, \dots\}$.

$X(\Omega)$ s'appelle le domaine ou le support de la v.a. X .

Les points x_k du support $X(\Omega)$ de la v.a. X s'appellent les atomes de la loi ou de la v.a. X .

L'évènement $A_k = \{\omega \in \Omega, X(\omega) = x_k\}$ est aussi noté $X^{-1}(\{x_k\})$ (où X^{-1} désigne l'inverse ensembliste) ou encore $\{X = x_k\}$ (dans ce cas, il est implicite que c'est l'ensemble des $\omega \in \Omega$ tel que $X(\omega) = x_k$, nous utiliserons la plupart du temps dans la suite, cette notation implicite).

Dans la suite, nous utiliserons l'abréviation v.a. pour variable(s) aléatoire(s).

Remarquons que l'ensemble des A_k forme une partition de l'espace de probabilité Ω (par partition, on entend une famille d'ensembles deux à deux disjoints et de réunion Ω). On a alors

$$\sum_{k, x_k \in X(\Omega)} \mathbb{P}(A_k) = \sum_{k, x_k \in X(\Omega)} \mathbb{P}(X = x_k) = \mathbb{P}(X \in \{x_1, \dots, x_k, \dots\}) = 1.$$

Dans cette écriture, les sommes sont des séries convergentes si $X(\Omega)$ est infini, des sommes finies si l'ensemble $X(\Omega)$ est fini.

4.1.2 Loi d'une variable aléatoire discrète

L'application X permet de transporter la probabilité \mathbb{P} de Ω en une probabilité \mathbb{P}_X sur \mathbb{R} : on considère pour cela les $\mathbb{P}(X = x_k)$ comme des masses ponctuelles p_k situées en les points x_k de la droite réelle, on définit ainsi une probabilité sur \mathbb{R} (le point x_k a la probabilité p_k). La probabilité, pour cette loi, d'une partie quelconque de \mathbb{R} est alors la somme des masses ponctuelles qu'elle contient.

Définition 4.1.2 Soit X une v.a. discrète sur $(\Omega, \mathcal{F}, \mathbb{P})$. On lui associe sa loi \mathbb{P}_X qui est la probabilité définie sur les parties de \mathbb{R} par

$$p_k = \mathbb{P}_X(\{x_k\}) = \mathbb{P}(X = x_k) = \mathbb{P}(A_k).$$

Pour tout $B \subset \mathbb{R}$:

$$\mathbb{P}_X(B) = \sum_{x_k \in B} \mathbb{P}(X = x_k) = \sum_{x_k \in B} p_k.$$

Les probabilités $p_k = \mathbb{P}(X = x_k)$ sont appelées probabilités ponctuelles de la v.a. X .

Dans la suite, le symbole \sim signifiera « a pour loi ». Par exemple, on notera $X \sim \mathcal{B}(n, p)$ pour signifier que la v.a. X suit la loi binomiale $\mathcal{B}(n, p)$.

Dans ce cas, n'importe quelle partie de \mathbb{R} est observable car la loi est discrète (l'ensemble des observables est $\mathcal{P}(\mathbb{R})$ tout entier).

Notons en particulier que comme $\sum_{k, x_k \in \Omega} p_k = 1$, $\mathbb{P}_X(B)$ est une sous-série d'une série à termes positifs convergente donc convergente : $\mathbb{P}_X(B)$ est donc toujours bien définie pour toute partie $B \subset \mathbb{R}$. Ce ne sera pas aussi simple dans le cas des variables aléatoires réelles (pour lesquelles les observables seront réduits aux intervalles de \mathbb{R}).

Remarque 4.1.1

- Attention, deux v.a. peuvent avoir la même loi sans pour autant être égales. Par exemple si on dispose d'un dé rouge et d'un dé bleu et que X, Y désignent la somme des points obtenus après un lancer respectivement du dé rouge et du dé bleu, X et Y ont la même loi. Pourtant bien sûr, on n'a pas $X = Y$, ce qui équivaudrait à dire que les tirages des deux dés sont nécessairement identiques.
- Désormais, on utilise la notation suivante où le ω est implicite :

$$\mathbb{P}(X \in B) = \mathbb{P}(\omega \in \Omega, X(\omega) \in B).$$

4.1.3 Fonction de répartition

Définition 4.1.3 On appelle fonction de répartition de la v.a. X la fonction F_X définie par :

$$\forall x \in \mathbb{R}, \quad F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \in]-\infty, x]) = \mathbb{P}(X \leq x).$$

On a aussi pour une v.a. X discrète

$$F_X(x) = \sum_{\substack{k, x_k \in X(\Omega), \\ x_k \leq x}} \mathbb{P}(X = x_k) = \sum_{\substack{k, x_k \in X(\Omega), \\ x_k \leq x}} p_k.$$

Proposition 4.1.1 Soit X une v.a. discrète d'ensemble de valeurs $X(\Omega) = \{x_1, x_2, \dots, x_k, \dots\}$ supposé ordonné $x_1 < x_2 < \dots < x_k < \dots$. La fonction de répartition F_X de X est croissante de 0 en $-\infty$ à 1 en $+\infty$, constante sur chaque intervalle $[x_k, x_{k+1}[$ avec un saut p_k en chaque atome x_k .

Notons que F_X détermine complètement la loi de X : les points du support sont les points de sauts de F_X et la probabilité associée est donnée par

$$p_k = F_X(x_k) - F_X(x_{k-1}).$$

Autrement dit $\mathbb{P}_X([a, b]) = F_X(b) - \lim_{t \rightarrow a^-} F_X(t)$. On retrouve donc la loi à partir de F_X .

Démonstration : D'abord F_X est à valeurs positives car une probabilité est toujours positive. Si $s < t$,

$$\begin{aligned} F(t) - F(s) &= \mathbb{P}(X \leq t) - \mathbb{P}(X \leq s) \\ &= \mathbb{P}(X \leq s) + \mathbb{P}(s < X \leq t) - \mathbb{P}(X \leq s) \\ &= \mathbb{P}(s < X \leq t) = \sum_{i \mid s < x_i \leq t} p_i \geq 0 \end{aligned}$$

donc F_X est croissante. Puis si $s < t$ sont dans $[x_k, x_{k+1}[$ alors

$$F(t) - F(s) = \sum_{i \mid s < x_i \leq t} p_i = 0$$

car la somme est vide : il n'y a pas d'atome x_i entre s et t . S'il y en avait un, il serait *a fortiori* entre x_k et x_{k+1} , ce qui est exclu, car par l'indexation, les atomes x_k et x_{k+1} sont consécutifs.

Puis avec $s = x_k$ et $t = x_{k+1}$, on a

$$F(x_{k+1}) - F(x_k) = \sum_{i \mid x_k < x_i \leq x_{k+1}} p_i = \sum_{i \mid x_i \in]x_k, x_{k+1}] } p_i = p_{k+1}$$

car x_{k+1} est le seul atome dans $]x_k, x_{k+1}]$. Il y a donc un saut p_{k+1} en x_{k+1} . Enfin,

$$\lim_{t \rightarrow -\infty} F_X(t) = \lim_{t \rightarrow -\infty} \sum_{i \mid x_i \leq t} p_i = 0$$

car pour $t \leq \inf_k(x_k)$, la somme est vide donc –par convention– nulle. Et

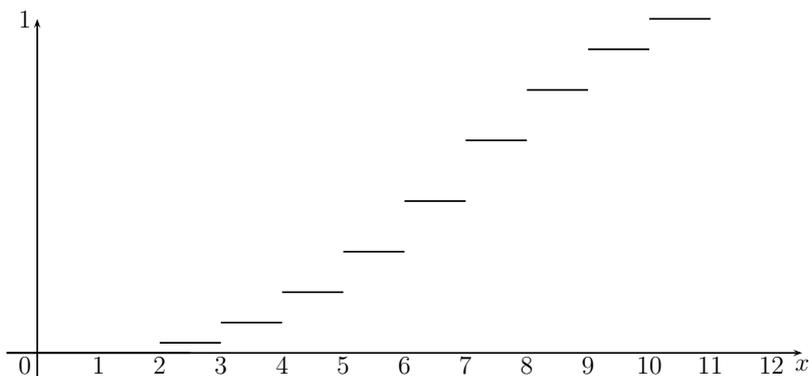
$$\lim_{t \rightarrow +\infty} F_X(t) = \lim_{t \rightarrow +\infty} \sum_{i \mid x_i \leq t} p_i = \sum_i p_i = 1$$

car pour $t \geq \sup_k(x_k)$, la somme devient $\sum_{i \mid x_i \in \mathbb{R}} p_i = 1$. ■

Exemple. Soit S la variable aléatoire qui donne la somme des faces obtenues en lançant deux fois un dé à six faces bien équilibré. La loi de S est donnée par l'ensemble des valeurs possibles $S(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ et les probabilités associées

$$\begin{aligned} \mathbb{P}(S = 2) &= 1/36, & \mathbb{P}(S = 6) &= 5/36, & \mathbb{P}(S = 10) &= 3/36, \\ \mathbb{P}(S = 3) &= 2/36, & \mathbb{P}(S = 7) &= 6/36, & \mathbb{P}(S = 11) &= 2/36, \\ \mathbb{P}(S = 4) &= 3/36, & \mathbb{P}(S = 8) &= 5/36, & \mathbb{P}(S = 12) &= 1/36. \\ \mathbb{P}(S = 5) &= 4/36, & \mathbb{P}(S = 9) &= 4/36, & & \end{aligned}$$

La fonction de répartition est alors donnée par :



4.2 Lois discrètes classiques

4.2.1 Lois de v.a. finies déjà connues

Ont déjà été vues en L1, les v.a. prenant un nombre fini de valeurs. Rappelons les principales telles lois.

Loi de Bernoulli de paramètre p notée $b(p)$. Une v.a. X suit une loi de Bernoulli de paramètre $p \in [0, 1]$ si elle ne prend que deux valeurs, la plupart du temps 0 et 1 avec :

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p := q.$$

Exemple. Pile ou face avec $p = 1/2$ si la pièce est équilibrée, $p \neq 1/2$ si elle est truquée.

Loi equirépartie sur un ensemble fini $\{x_1, \dots, x_n\}$ notée $\mathcal{U}\{x_1, \dots, x_n\}$. Une v.a. X prenant un nombre fini de valeurs x_1, \dots, x_n suit une loi equirépartie quand

$$\mathbb{P}_X(\{x_i\}) = \frac{1}{n}, \quad 1 \leq i \leq n.$$

Exemple. Jet d'un dé (équilibré).

Loi binomiale de paramètres n, p notée $\mathcal{B}(n, p)$. Une v.a. suit une loi binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$ si l'ensemble de ses valeurs possibles est :

$$X(\Omega) = \{0, 1, 2, \dots, n\}$$

et pour tout $k = 0, 1, \dots, n$, on a

$$\mathbb{P}(X = k) = C_n^k p^k (1 - p)^{n-k} \quad (4.1)$$

où $C_n^k = \frac{n!}{k!(n-k)!}$ est le coefficient binomial. Il s'agit bien d'une loi de probabilité car la formule du binôme de Newton (d'où le nom de la loi) donne :

$$\sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

Remarque 4.2.1 Il est souvent pratique de voir cette loi comme celle du nombre de succès obtenus dans une suite de n épreuves répétées indépendantes avec pour chaque épreuve une probabilité p de succès (par exemple des tirages avec remise de n boules rouges dans une urne contenant des boules rouges, en proportion $p \in [0, 1]$, et des boules noires). Ainsi, $\mathbb{P}(X = k)$ est la probabilité d'avoir exactement k succès en n épreuves (dans l'exemple, k boules rouges en n tirages). On en déduit l'explication suivante des différents facteurs de (4.1) :

- p^k est la probabilité des k succès (par indépendance des tirages),
- $(1-p)^{n-k}$ est la probabilité des $n-k$ échecs (pour avoir **exactement** k succès, il faut bien que les $n-k$ autres épreuves soient des échecs),
- et C_n^k pour tenir compte de tous les choix possibles des k épreuves réussies sur les n réalisées (il y a C_n^k tirages différents de n boules comprenant k boules rouges).

Une autre façon de dire la même chose est qu'une v.a. Y de loi binomiale $\mathcal{B}(n, p)$ peut se voir comme la somme de n v.a. X_i indépendantes de loi de Bernoulli $b(p)$:

$$Y = X_1 + \dots + X_n.$$

D'après cette interprétation, la loi binomiale intervient dans les tirages avec remise : elle modélise la v.a. qui compte le nombre de bons tirages en un nombre fixé, n , de tirages. Ici chaque X_i indique si à l'épreuve i on a eu un succès ($X_i = 1$) ou un échec ($X_i = 0$).

Intéressons nous maintenant aux lois des v.a. discrètes prenant un nombre infini de valeurs.

4.2.2 Lois Géométriques

Exemple : Considérons une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0, 1[$. Soit X le numéro aléatoire de l'épreuve où l'on obtient le premier succès. Convenons que si l'on n'obtient jamais de succès, on note $X = +\infty$. Calculer $\mathbb{P}(X = k)$ pour tout $k \in \mathbb{N}^*$. En déduire $\mathbb{P}(X \in \mathbb{N}^*)$ et $\mathbb{P}(X = +\infty)$.

A priori, X prend ses valeurs dans $\{1, 2, \dots, k, \dots\} \cup \{\infty\} = \mathbb{N}^* \cup \{\infty\}$.

Notons $R_i = \{\text{succès à la } i\text{-ème épreuve}\}$, on a :

$$\begin{aligned} \{X = k\} &= \{\text{échecs aux } k-1 \text{ premières épreuves puis succès à la } k\text{-ème}\} \\ &= \left(\bigcap_{i=1}^{k-1} R_i^c \right) \cap R_k. \end{aligned}$$

Par indépendance des épreuves, il suit

$$\mathbb{P}(X = k) = \mathbb{P}\left(\left(\bigcap_{i=1}^{k-1} R_i^c\right) \cap R_k\right) = \prod_{i=1}^{k-1} \mathbb{P}(R_i^c) \times \mathbb{P}(R_k) = (1-p)^{k-1}p.$$

Posons $q := 1-p \in]0, 1[$. Décomposons l'évènement $\{X \in \mathbb{N}^*\}$ en réunion disjointe des $\{X = k\}$ pour $k \in \mathbb{N}^*$, on a alors par σ -additivité de \mathbb{P} :

$$\begin{aligned} \mathbb{P}(X \in \mathbb{N}^*) &= \sum_{k \in \mathbb{N}^*} \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} q^{k-1}p = p \sum_{l=0}^{\infty} q^l \quad (\text{en posant } l = k-1) \\ &= p \frac{1}{1-q} = p \frac{1}{1-(1-p)} = 1 \quad (\text{somme d'une suite géométrique}). \end{aligned}$$

Ainsi, avec probabilité 1, le premier succès intervient au bout d'un nombre fini d'épreuves (car dire $X \in \mathbb{N}^*$, c'est dire que X est un entier donc prend une valeur finie). On en déduit immédiatement

$$\mathbb{P}(X = +\infty) = \mathbb{P}(X \notin \mathbb{N}^*) = 1 - \mathbb{P}(X \in \mathbb{N}^*) = 0.$$

Définition 4.2.1 Une v.a. X suit la loi géométrique de paramètre $p \in]0, 1[$ notée $\mathcal{G}(p)$ si $X(\Omega) = \mathbb{N}^*$ et

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

Exercice. Notons que si X suit une loi $\mathcal{G}(p)$, les probabilités $\mathbb{P}(X > n)$ ont une expression simple qu'on pourra montrer à titre d'exercice :

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

4.2.3 Loi de Poisson

Cette loi intervient dans les processus aléatoires dont les éventualités sont faiblement probables et survenant indépendamment les unes des autres : cas de phénomènes accidentels, d'anomalies diverses, de problèmes d'encombrement (files d'attente), de rupture de stocks, etc.

Définition 4.2.2 On dit qu'une v.a. discrète X suit une loi de Poisson de paramètre $\lambda > 0$ si l'ensemble de ses valeurs possibles est $X(\Omega) = \mathbb{N}$ et

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

La loi de Poisson de paramètre $\lambda > 0$ est notée $\mathcal{P}(\lambda)$.

Rappelons que la fonction exponentielle a un développement en série entière avec un rayon de convergence infini :

$$\forall \lambda > 0, \quad e^\lambda = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}.$$

On vérifie alors facilement qu'on a bien défini une loi de probabilité car

$$\sum_{k=0}^{+\infty} \mathbb{P}(X = k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = e^{-\lambda+\lambda} = 1.$$

Approximation de la loi binomiale par la loi de Poisson

En liaison avec les lois binomiales, on dispose du résultat suivant justifiant que la loi de Poisson approxime la loi binomiale $\mathcal{B}(n, p)$ quand n est « grand » et np est « petit ».

Théorème 4.2.1 (Approximation de la loi de Poisson par la loi binomiale)

Si $(p_n)_{n \in \mathbb{N}^*}$ est une suite de réels de $[0, 1]$ telle que

$$np_n \rightarrow \lambda \in]0, +\infty[, \quad n \rightarrow +\infty$$

alors

$$C_n^k p_n^k (1 - p_n)^{n-k} \longrightarrow e^{-\lambda} \frac{\lambda^k}{k!}, \quad \text{quand } n \rightarrow +\infty.$$

Autrement dit, si X_n est une suite de v.a. de loi binomiale $\mathcal{B}(n, p_n)$ et X une v.a. de loi $\mathcal{P}(\lambda)$ alors pour tout $k \in \mathbb{N}$:

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

« Les lois binomiales convergent vers la loi de Poisson. »

Démonstration : On remplace p_n par son équivalent λ/n . Pour k fixé,

$$\begin{aligned} P(X_n = k) &= C_n^k p_n^k (1 - p_n)^{n-k} = \frac{n!}{k!(n-k)!} p_n^k e^{(n-k) \ln(1-p_n)} \\ &\simeq \frac{n(n-1) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k e^{-(n-k)\lambda/n} \\ &\simeq \frac{1}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k} \lambda^k e^{-\lambda} e^{k\lambda/n} \\ &\simeq e^{-\lambda} \frac{\lambda^k}{k!} = P(X = k). \end{aligned}$$

■

Application pratique. Le théorème 4.2.1 sert de justification théorique à la règle pratique suivante : lorsque n est « grand » et np est « petit », on peut remplacer la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(\lambda)$ où $\lambda = np$.

En général, on considère que lorsque n est de l'ordre de quelques centaines et np est de l'ordre de quelques unités, l'approximation de $\mathcal{B}(n, p)$ par $\mathcal{P}(np)$ est assez bonne.

Intérêt : si n est grand, le calcul des coefficients binomiaux C_n^k est fastidieux, voire impossible. En approchant par la loi de Poisson, le calcul devient assez simple.

Exemple : Le président d'un bureau de vote est né un 1er avril. Il décide de noter le nombre de personnes ayant leur anniversaire le même jour que lui parmi les 500 premiers votants.

La situation peut être assimilée à une suite de 500 épreuves indépendantes répétées avec une probabilité $p = 1/365$ de succès (on néglige les effets des années bissextiles, sinon il faudrait plutôt prendre $p = 4/(3 \times 365 + 366)$). Notons X la variable aléatoire qui compte le nombre de succès. X suit une loi $\mathcal{B}(500, p)$, ainsi :

$$\mathbb{P}(X = k) = C_{500}^k p^k (1 - p)^{500-k}.$$

Comme 500 est « grand » et $np = 500/365 \simeq 1,37$, la règle ci-dessus permet l'approximation par la loi $\mathcal{P}(\lambda)$ avec $\lambda = 500/365$. Voici une comparaison numérique pour les petites valeurs de k :

k	0	1	2	3	4	5
$\mathbb{P}(X = k)$	0,2537	0,3484	0,2388	0,1089	0,0372	0,0101
$\frac{e^{-\lambda} \lambda^k}{k!}$	0,2541	0,3481	0,2385	0,1089	0,0373	0,0102

On constate que les valeurs approchées sont très proches des valeurs réelles.

Application aux files d'attente

Un poste de péage d'une autoroute possède plusieurs guichets. En période de pointe et dans la tranche 7h–9h, on compte 6300 véhicules par heure (c'est à peu près les chiffres sur le périphérique parisien).

Des compteurs à la sortie du péage ont montré qu'un automobiliste met en moyenne 18 secondes pour s'acquitter du montant du péage. On estime qu'il y a risque de saturation (création d'un bouchon) si on compte plus de 10 véhicules en attente à chaque guichet.

On se place désormais dans la tranche 7h–9h.

1) Soit X la v.a. qui désigne le nombre de véhicules présents au péage à un instant donné. Quelle est sa loi ? Quelle est son espérance (nombre moyen de véhicules présents au péage à un instant donné) ?

L'arrivée au péage est un événement de probabilité $p = 18/(2 \times 360) = 0,0025$. Il y a $N = 12600$ véhicules présents dans la tranche horaire. La v.a. X suit la loi binomiale $\mathcal{B}(N, p)$.

2) Dans le cas où il y a 5 guichets, en admettant une égale répartition des véhicules sur chaque guichet et en notant Y le nombre de véhicules se présentant à un guichet donné, justifier que Y suit sensiblement une loi de Poisson et calculer la probabilité de saturation, $\mathbb{P}(Y > 10)$.

Pour un guichet donné Y suit la loi binomiale $\mathcal{B}(N, p')$ avec $p' = p/5$, ce qui s'approxime par la loi de Poisson $\mathcal{P}(Np') = \mathcal{P}(6, 3)$. L'approximation est légitime car Np' est entre 1 et 10.

La probabilité de saturation est $\mathbb{P}(Y > 10) = \mathbb{P}(Y \geq 11) = 1 - \mathbb{P}(Y \leq 10) \simeq 0,056$.

3) On suppose le nombre k de guichets non précisé. Quelle est la valeur minimale à attribuer à k pour que la probabilité de saturation ne dépasse pas 0,01 ?

En procédant par dichotomie (tester $k = 5$ guichets avec $p' = p/5$ puis $k = 10$ guichets avec $p' = p/10$ puis etc), on trouve $k = 7$ guichets.

Notez que la probabilité trouvée en 2) avec 5 guichets est de 6% alors qu'avec 2 guichets de plus elle est divisée par 6. Il n'y a donc pas proportionnalité dans ces calculs de probabilités !

Chapitre 5

Moment des variables aléatoires discrètes

On en vient dans la section suivante aux espérances et variances de variables aléatoires. Ce sont certaines quantités numériques associées à la loi d'une v.a. qui apportent des informations sur cette loi. On traite d'abord le cas des v.a. discrètes, l'analogie s'énoncera pour les v.a. à densité.

5.1 Espérance d'une v.a.

5.1.1 Définitions

Définition 5.1.1 (Espérance d'une v.a. discrète) Soit X une v.a. discrète prenant une infinité de valeurs $x_1, x_2, \dots, x_n, \dots$ si

$$\sum_{k=1}^{\infty} |x_k| \mathbb{P}(X = x_k) < +\infty, \quad (5.1)$$

on définit l'espérance de X par

$$E[X] = \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k) = \sum_{k=1}^{\infty} x_k p_k. \quad (5.2)$$

Remarque 5.1.1

- Si $X(\Omega)$ est fini, la somme dans (5.1) est forcément finie et l'espérance dans (5.2) est forcément définie.
- L'espérance généralise la notion intuitive de moyenne et peut être vue comme le barycentre des points x_k avec le poids donné par leur probabilité ponctuelle $p_k = \mathbb{P}(X = x_k)$. Elle donne donc la valeur moyenne de la v.a. au sens probabiliste.
- L'espérance ne dépend que de la loi de X : si X et Y sont deux v.a. de même loi, alors $E[X] = E[Y]$ car ça ne dépend que des atomes x_k et de leur probabilité p_k . On devrait ainsi plutôt parler de l'espérance de la loi de X .

5.1.2 Espérances classiques

Exemples :

– X v.a. constante ($\exists c \in \mathbb{R}, \forall \omega \in \Omega, X(\omega) = c$), alors son espérance est $E[X] = c \times 1 = c$.

– X de loi de Bernoulli $b(p)$:

$$E[X] = 0 \times (1 - p) + 1 \times p = p.$$

– X de loi équirépartie sur $\{x_1, \dots, x_n\}$:

$$E[X] = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

– X de loi binomiale $\mathcal{B}(n, p)$:

$$\begin{aligned} E[X] &= \sum_{k=0}^n k C_n^k p^k (1-p)^{n-k} = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{l=0}^{n-1} \frac{(n-1)!}{l!(n-1-l)!} p^l (1-p)^{n-1-l} = np(p + (1-p))^{n-1} \\ &= np. \end{aligned}$$

– X de loi géométrique $\mathcal{G}(p)$: avec la propriété de dérivation des séries entières rappelée en section 1.4 (ou à voir en cours d'analyse) :

$$\begin{aligned} E[X] &= \sum_{k=1}^{+\infty} k q^{k-1} p = p \sum_{k=1}^{+\infty} \left[\frac{d}{dx} (x^k) \right]_{x=q} = p \left[\frac{d}{dx} \left(\sum_{k=0}^{+\infty} x^k \right) \right]_{x=q} \\ &= p \frac{1}{(1-q)^2} = \frac{p}{(1-(1-p))^2} = \frac{1}{p}. \end{aligned}$$

– X de loi de Poisson $\mathcal{P}(\lambda)$:

$$E[X] = \sum_{k=0}^{+\infty} k \mathbb{P}(X = k) = \sum_{k=1}^{+\infty} \frac{e^{-\lambda} \lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{l=0}^{+\infty} \frac{\lambda^l}{l!} = \lambda \quad (\text{avec } l = k - 1).$$

5.1.3 Propriétés de l'espérance

Les principales propriétés des espérances sont données par

Proposition 5.1.1 (Linéarité de l'espérance) *Soient X et Y deux v.a. discrètes admettant des espérances. Alors*

- (1) $E[X + Y] = E[X] + E[Y]$,
- (2) Pour tout réel a , $E[aX] = aE[X]$,

Démonstration : Le deuxième point est clair : en effet si $X(\Omega) = \{x_1, \dots, x_k, \dots\}$ alors $(aX)(\Omega) = \{ax_1, \dots, ax_k, \dots\}$ et comme $\mathbb{P}(aX = ax_k) = \mathbb{P}(X = x_k)$, on a d'abord

$$\sum_{k=1}^{+\infty} |ax_k| \mathbb{P}(aX = ax_k) = |a| \sum_{k=0}^{\infty} |x_k| \mathbb{P}(X = x_k) < +\infty$$

par hypothèse de l'existence de $E[X]$. Puis la même chose sans les valeurs absolues donne :

$$E[aX] = \sum_{k=1}^{+\infty} ax_k \mathbb{P}(aX = ax_k) = a \sum_{k=0}^{\infty} x_k \mathbb{P}(X = x_k) = aE[X].$$

Pour le premier point, posons $Z = X + Y$, supposons d'abord que X et Y prennent un nombre fini de valeurs $\{x_1, \dots, x_n\}$ et $\{y_1, \dots, y_p\}$ alors Z prend aussi un nombre fini de valeurs $\{z_k = x_i + y_j, i = 1, \dots, n, j = 1, \dots, p\}$ et la loi de Z est donnée par

$$\mathbb{P}(Z = z_k) = \sum_{(i,j) | x_i + y_j = z_k} \mathbb{P}(X = x_i, Y = y_j).$$

La condition (5.1) est remplie dans ce cas car Z prend un nombre finie de valeur (et donc la somme dans (5.1) est finie). On a alors

$$E[Z] = \sum_k z_k \mathbb{P}(Z = z_k) \quad (5.3)$$

$$= \sum_k z_k \sum_{x_i + y_j = z_k} \mathbb{P}(X = x_i, Y = y_j)$$

$$= \sum_k \sum_{x_i + y_j = z_k} (x_i + y_j) \mathbb{P}(X = x_i, Y = y_j)$$

$$= \sum_{i,j} (x_i + y_j) \mathbb{P}(X = x_i, Y = y_j) \quad (5.4)$$

$$= \sum_{i,j} x_i \mathbb{P}(X = x_i, Y = y_j) + \sum_{i,j} y_j \mathbb{P}(X = x_i, Y = y_j) \quad (5.5)$$

car $\{(i, j), x_i + y_j = z_k\}_k$ forme une partition de $\{1, \dots, n\} \times \{1, \dots, p\}$. Comme les sommes sont finies, on les calcule dans le sens qu'on veut $\sum_{i,j} = \sum_i \sum_j = \sum_j \sum_i$.

$$E[Z] = \sum_i \sum_j x_i \mathbb{P}(X = x_i, Y = y_j) + \sum_j \sum_i y_j \mathbb{P}(X = x_i, Y = y_j) \quad (5.6)$$

$$= \sum_i x_i \left[\sum_j \mathbb{P}(X = x_i, Y = y_j) \right] + \sum_j y_j \left[\sum_i \mathbb{P}(X = x_i, Y = y_j) \right]$$

$$= \sum_i x_i \mathbb{P}(X = x_i) + \sum_j y_j \mathbb{P}(Y = y_j)$$

$$= E[X] + E[Y]. \quad (5.7)$$

Dans le cas général, X et/ou Y prennent un nombre dénombrable infini de valeurs (i.e. $X(\Omega), Y(\Omega)$ sont infinis). On commence alors à remplacer z_k par $|z_k|$ et $x_i + y_j$ par $|x_i + y_j|$ dans les égalités (5.3)–(5.7). On passe alors de (5.4) à (5.5) par l'inégalité triangulaire $|x_i + y_j| \leq |x_i| + |y_j|$, ce qui donne :

$$\sum_k |z_k| \mathbb{P}(Z = z_k) \leq \sum_i |x_i| \mathbb{P}(X = x_i) + \sum_j |y_j| \mathbb{P}(Y = y_j) < +\infty$$

les deux dernières séries sont finies d'après les hypothèses sur les existences des espérances $E[X]$ et $E[Y]$.

On obtient alors la condition qui garantit l'existence de $E[Z]$ et les égalités (5.3)–(5.7) restent vraies, dans le cas infini dénombrable, par les propriétés des séries doubles absolument convergentes (théorème de Fubini). ■

Exemple : Calcul de l'espérance d'une loi binomiale $\mathcal{B}(n, p)$

Si X suit la loi $\mathcal{B}(n, p)$, alors $E[X] = np$.

En effet, on sait que X a même loi qu'une somme de v.a. X_i de Bernoulli indépendantes de paramètre p :

$$S = \sum_{i=1}^n X_i.$$

Par linéarité de l'espérance, comme $E[X_i] = p$ pour tout $i = 1, \dots, n$, on a

$$E[X] = E[S] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = np.$$

Proposition 5.1.2 (Théorème de transfert, Espérance d'une fonction de va) Soient X une v.a. discrète de domaine $X(\Omega) = \{x_1, \dots, x_k, \dots\}$ et F une fonction numérique sur \mathbb{R} (ou dont l'ensemble de définition contient au moins l'ensemble des valeurs $X(\Omega)$ de X). Alors si $E[F(X)]$ existe, on a

$$E[F(X)] = \sum_{k=1}^{+\infty} F(x_k) \mathbb{P}(X = x_k).$$

Démonstration : Notons $Y = F(X)$, l'ensemble des valeurs prises par la v.a. Y est $\{F(x_1), \dots, F(x_k), \dots\}$ avec éventuellement des répétitions car F n'est pas nécessairement injective. En notant $\{y_1, \dots, y_k, \dots\}$ l'ensemble des valeurs de Y sans répétition (i.e. les y_i sont deux à deux distincts), on a :

$$E[Y] = E[F(X)] = \sum_{i=1}^{+\infty} y_i \mathbb{P}(Y = y_i). \quad (5.8)$$

Pour chaque $i = 1, \dots, k, \dots$, notons $B_i = \{x_k \mid F(x_k) = y_i\}$ l'ensemble des antécédents de y_i par F . Ce sous-ensemble est non vide et au plus dénombrable (si F est injective, cet ensemble est de cardinal 1).

$$\{Y = y_i\} = \bigcup_{k \mid x_k \in B_i} \{X = x_k\} \quad (5.9)$$

en effet $\{Y = y_i\} \subset \bigcup_{k \mid x_k \in B_i} \{X = x_k\}$ car si $\omega \in \{Y = y_i\}$ alors $F(X(\omega)) = Y(\omega) = y_i$. Or il existe k tel que $X(\omega) = x_k$. Comme alors $F(x_k) = y_i$, on a $x_k \in B_i$. Autrement dit, il existe k vérifiant $x_k \in B_i$ tel que $X(\omega) = x_k$, c'est à dire

$$\omega \in \bigcup_{k \mid x_k \in B_i} \{X = x_k\}.$$

Puis si $\omega \in \bigcup_{k, x_k \in B_i} \{X = x_k\}$ alors $Y(\omega) = F(X(\omega)) = F(x_k) = y_i$ car $x_k \in B_i$, donc $\omega \in \{Y = y_i\}$, ce qui justifie l'inclusion réciproque et donc l'égalité (5.9).

Le terme général de la série (5.8) se transforme alors en

$$y_i \mathbb{P}(Y = y_i) = y_i \mathbb{P}\left(\bigcup_{x_k \in B_i} \{X = x_k\}\right) = \sum_{x_k \in B_i} y_i \mathbb{P}(X = x_k) = \sum_{x_k \in B_i} F(x_k) \mathbb{P}(X = x_k).$$

La série précédente est absolument convergente car F est constante sur B_i . Comme les B_i forment une partition de $X(\Omega)$, les propriétés des séries à termes positifs donnent

$$\sum_{k=1}^{+\infty} |F(x_k)| \mathbb{P}(X = x_k) = \sum_{i=1}^{+\infty} \sum_{x_k \in B_i} |F(x_k)| \mathbb{P}(X = x_k) = \sum_{i=1}^{+\infty} |y_i| \mathbb{P}(Y = y_i) < +\infty$$

par hypothèse (existence de $E[Y]$). Ceci légitime le même calcul sans les valeurs absolues et prouve la proposition. ■

Remarque 5.1.2 La proposition appliquée avec $F(x) = |x|$ donne :

$$E[|X|] = \sum_{k=1}^{+\infty} |x_k| \mathbb{P}(X = x_k).$$

La condition de définition des espérances (*cf.* Déf. 5.1.1) pour l'existence de $E[X]$ n'est donc rien d'autre que $E[|X|] < +\infty$.

On a de plus facilement

Proposition 5.1.3 (Espérance et valeurs absolues) *Si $E[X]$ existe, $|E[X]| \leq E[|X|]$.*

Démonstration : car par l'inégalité triangulaire, la valeur absolue d'une somme est majorée par la somme des valeurs absolues.

Proposition 5.1.4 (Positivité de l'espérance)

- Si X a une espérance et $X \geq 0$, alors $E[X] \geq 0$.
- Si X et Y ont des espérances et $X \leq Y$ alors $E[X] \leq E[Y]$.

Démonstration : $X \geq 0$ signifie que pour tout $\omega \in \Omega$, on a $X(\omega) \geq 0$. De même, $X \leq Y$ signifie que pour tout $\omega \in \Omega$, on a $X(\omega) \leq Y(\omega)$.

Il suffit de voir le premier point, le deuxième se voit en appliquant le premier à $Z = Y - X$ et en appliquant la linéarité de l'espérance.

Soit donc $X \geq 0$, l'ensemble des valeurs x_k prises par X est dans \mathbb{R}^+ . $E[X]$ apparaît alors comme une série avec que des termes positifs, elle est *a fortiori* positive. ■

Théorème 5.1.1 (Inégalité de Markov) Si X est une v.a. positive ayant une espérance alors

$$\forall t > 0, \quad \mathbb{P}(X \geq t) \leq \frac{E[X]}{t}.$$

Démonstration : Dans la série $E[X]$, on regroupe les termes en deux paquets selon la position de x_k par rapport à t :

$$\begin{aligned} E[X] &= \sum_{k=1}^{+\infty} x_k \mathbb{P}(X = x_k) = \sum_{k \mid x_k < t} x_k \mathbb{P}(X = x_k) + \sum_{k \mid x_k \geq t} x_k \mathbb{P}(X = x_k) \\ &\geq 0 + t \sum_{k \mid x_k \geq t} \mathbb{P}(X = x_k) = t \mathbb{P}(X \geq t). \end{aligned}$$

car $\{X \geq t\} = \bigcup_{k \mid x_k \geq t} \{X = x_k\}$. ■

5.2 Variance d'une va

Définition 5.2.1 Soit $r \in \mathbb{N}^*$, on appelle moment d'ordre r de la v.a. discrète X le nombre

$$E[|X|^r] = \sum_{k=1}^{+\infty} |x_k|^r \mathbb{P}(X = x_k)$$

On peut alors définir aussi

$$E[X^r] = \sum_{k=1}^{+\infty} x_k^r \mathbb{P}(X = x_k).$$

où $X(\Omega) = \{x_1, \dots, x_k, \dots\}$ est l'ensemble des valeurs de X .

Remarque 5.2.1 Attention, il ne faut pas confondre $E[X^r]$ et $E[X]^r$. La position de l'exposant r par rapport à l'espérance E est cruciale.

Lorsque une v.a. est bornée, il est facile de voir qu'elle a des moments de tous les ordres. De façon générale, l'existence de moment d'ordre r implique celles des moments d'ordre inférieurs.

Proposition 5.2.1 *Si X possède un moment d'ordre r , pour tout $n \leq r$, X en possède un d'ordre n .*

Démonstration : Il s'agit de voir la finitude de

$$E[|X|^n] = \sum_{k=1}^{+\infty} |x_k|^n \mathbb{P}(X = x_k).$$

Regroupons les termes de la série précédente en deux selon le module des x_k par rapport à 1 :

$$E[|X|^n] = \sum_{\substack{k=1, \\ |x_k| \leq 1}}^{+\infty} |x_k|^n \mathbb{P}(X = x_k) + \sum_{\substack{k=1, \\ |x_k| > 1}}^{+\infty} |x_k|^n \mathbb{P}(X = x_k).$$

Pour la première série, on a la majoration

$$\sum_{\substack{k=1, \\ |x_k| \leq 1}}^{+\infty} |x_k|^n \mathbb{P}(X = x_k) \leq \sum_{\substack{k=1, \\ |x_k| \leq 1}}^{+\infty} \mathbb{P}(X = x_k) \leq \sum_{k=1}^{+\infty} \mathbb{P}(X = x_k) = 1$$

car $|x_k|^n \leq 1$. Pour la deuxième, comme $|x_k| > 1$, on a $|x_k|^n \leq |x_k|^r$ et on la majore par

$$\sum_{\substack{k=1, \\ |x_k| \leq 1}}^{+\infty} |x_k|^r \mathbb{P}(X = x_k) \leq \sum_{k=1}^{+\infty} |x_k|^r \mathbb{P}(X = x_k) = E[|X|^r] < +\infty.$$

On a même montré pour $n \leq r$:

$$E[|X|^n] \leq 1 + E[|X|^r]$$

■

Définition 5.2.2 (Variance d'une va) *Soit X une v.a. de domaine $X(\Omega) = \{x_1, \dots, x_k, \dots\}$ et avec un moment d'ordre 2. On appelle respectivement variance de X et écart-type de X les quantités*

$$\begin{aligned} \text{Var}(X) &= E[(X - E[X])^2] = \sum_{k=1}^{+\infty} (x_k - E[X])^2 \mathbb{P}(X = x_k), \\ \sigma_X &= \sqrt{\text{Var}(X)}. \end{aligned}$$

Remarque 5.2.2 – L'espérance permet de trouver la valeur moyenne d'une v.a. à partir de cette valeur, la variance ou l'écart-type permettent de voir la dispersion de la v.a. autour de cette moyenne.

- On introduit l'écart type pour avoir une quantité homogène avec X : si X est une grandeur physique d'une certaine unité, alors σ_X a la même unité, ce n'est pas le cas de $\text{Var}(X)$.
- En quelque sorte, l'écart-type σ_X de X donne l'écart moyen de la v.a. X par rapport à sa valeur moyenne $E[X]$.

Proposition 5.2.2 (Translation et changement d'échelle) *Si X a un moment d'ordre 2 alors*

$$\text{Var}(aX) = a^2 \text{Var}(X), \quad \text{Var}(X + b) = \text{Var}(X).$$

Démonstration :

$$\text{Var}(aX) = E[(aX - E[aX])^2] = E[a^2(X - E[X])^2] = a^2 E[(X - E[X])^2] = a^2 \text{Var}(X)$$

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

■

En particulier avec $a = 0$, on obtient la nullité de la variance d'une va constante. On a le résultat réciproque :

Proposition 5.2.3

$$\text{Var}(X) = 0 \iff X \text{ est presque sûrement constant} \iff X = E[X] \text{ p.s.}$$

Notons $\mu = E[X]$, l'égalité presque sûre (p.s.) $X = \mu$ signifie que $\mathbb{P}(X = \mu) = 1$. De même, X constant presque sûrement signifie qu'il existe une constante telle que presque sûrement X est égale à cette constante.

Démonstration : On a

$$\text{Var}(X) = E[(X - \mu)^2] = \sum_{k=1}^{+\infty} (x_k - \mu)^2 \mathbb{P}(X = x_k).$$

Cette série à termes positifs est nulle si et seulement si chacun de ses termes est nuls c'est à dire si

$$x_k = \mu \quad \text{ou} \quad \mathbb{P}(X = x_k) = 0.$$

Par définition d'un atome, $x_k \in X(\Omega)$, sa probabilité ponctuelle $\mathbb{P}(X = x_k)$ est non nulle. On a donc forcément $x_k = \mu$. En fait, il y a un seul atome et c'est $\mu = E[X]$ et sa probabilité ponctuelle est alors forcément 1. La valeur $\mu = E[X]$ est donc prise par la v.a. X avec probabilité 1 (et il est bien logique que cette valeur soit l'espérance). ■

Proposition 5.2.4 (Formule de Koenig)

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

En pratique, on calcule la variance par cette formule. On commence par calculer $E[X]$ puis $E[X^2]$ pour en déduire la variance par Koenig.

Démonstration : Il suffit de développer $\text{Var}(X)$, notons encore $\mu = E[X]$:

$$\begin{aligned} \text{Var}(X) = E[(X - \mu)^2] &= E[X^2 - 2X\mu + \mu^2] \\ &= E[X^2] - 2E[X\mu] + \mu^2 \\ &= E[X^2] - 2E[X]\mu + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - E[X]^2. \end{aligned}$$

■

Exemples : Quelques variances classiques

- X v.a. constante ($\exists c \in \mathbb{R}, \forall \omega \in \Omega, X(\omega) = c$), alors sa variance est $\text{Var}(X) = 0$.

On retrouve le cas spécial où il n'y a aucune dispersion autour de la moyenne puisque la v.a. est constante, égale à cette moyenne.

- X de loi de Bernoulli $\mathcal{B}(p)$: on a $E[X^2] = 0 \times (1-p) + 1^2 \times p = p$ d'où $\text{Var}(X) = p - p^2$:

$$\text{Var}(X) = p(1 - p).$$

- X de loi equirépartie sur $\{1, \dots, n\}$: on a

$$\begin{aligned} E[X] &= \sum_{k=1}^n k \mathbb{P}(X = k) = \sum_{k=1}^n \frac{k}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2} \\ E[X^2] &= \sum_{k=1}^n k^2 \mathbb{P}(X = k) = \sum_{k=1}^n \frac{k^2}{n} = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6} \\ \text{Var}(X) &= E[X^2] - E[X]^2 = \frac{(n+1)(n-1)}{12}. \end{aligned}$$

- X de loi binomiale $\mathcal{B}(n, p)$:

$$\text{Var}(X) = np(1 - p).$$

En effet $E[X^2] = \sum_{k=0}^n C_n^k k^2 p^k (1-p)^{n-k} = S_q(p)$ où $q = 1-p$ et $S_q(x) = \sum_{k=1}^n C_n^k k^2 x^k q^{n-k}$.
Or

$$\begin{aligned} S_q(x) &= \sum_{k=1}^n C_n^k k^2 x^k q^{n-k} = x \sum_{k=1}^n C_n^k k^2 x^{k-1} q^{n-k} = x \sum_{k=1}^n C_n^k k (x^k)' q^{n-k} \\ &= x \left(\sum_{k=1}^n C_n^k k x^k q^{n-k} \right)' = x \left(x \sum_{k=1}^n C_n^k k x^{k-1} q^{n-k} \right)' \end{aligned}$$

$$\begin{aligned}
&= x \left(x \sum_{k=1}^n C_n^k (x^k)' q^{n-k} \right)' = x \left(x \left(\sum_{k=1}^n C_n^k x^k q^{n-k} \right)' \right)' \\
&= x \left(x \left(\sum_{k=0}^n C_n^k x^k q^{n-k} \right)' \right)' = x (x[(x+q)^n])' \\
&= x (x \times n(x+q)^{n-1})' = xn(x+q)^{n-1} + x^2 \times n(n-1)(x+q)^{n-2}.
\end{aligned}$$

D'où $E[X^2] = S_{1-p}(p) = pn + p^2n(n-1)$ et

$$\text{Var}(X) = pn + p^2n(n-1) - (np)^2 = n(p - p^2) = np(1-p).$$

- X de loi géométrique $\mathcal{G}(p)$:

$$\text{Var}(X) = \frac{1-p}{p^2}.$$

En effet $E[X^2] = \sum_{k=0}^{+\infty} k^2(1-p)^{k-1}p = pS(1-p)$ avec $S(x) = \sum_{k=0}^{+\infty} k^2x^{k-1}$. Puis

$$\begin{aligned}
S(x) &= \sum_{k=0}^{+\infty} k^2x^{k-1} = \sum_{k=1}^{+\infty} k(x^k)' = \left(\sum_{k=1}^{+\infty} kx^k \right)' \\
&= \left(x \sum_{k=1}^{+\infty} kx^{k-1} \right)' = \left(x \sum_{k=1}^{+\infty} (x^k)' \right)' \\
&= \left(x \left(\sum_{k=1}^{+\infty} x^k \right)' \right)' = \left(x \left(\sum_{k=0}^{+\infty} x^k \right)' \right)' \\
&= \left(x \left(\frac{1}{1-x} \right)' \right)' = \left(x \frac{1}{(1-x)^2} \right)' \\
&= \frac{1}{(1-x)^2} + \frac{2x}{(1-x)^3}
\end{aligned}$$

D'où $E[X^2] = pS(1-p) = p \frac{1}{(1-(1-p))^2} + p \frac{2-2p}{(1-(1-p))^3} = \frac{1}{p} + \frac{2-2p}{p^2}$ et

$$\text{Var}(X) = \frac{1}{p} + \frac{2-2p}{p^2} - \frac{1}{p^2} = \frac{p(1-p)}{p^2}.$$

- X de loi de Poisson $\mathcal{P}(\lambda)$:

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

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

Comme $E[X] = \lambda$, on en déduit par la formule de Koenig :

$$\text{Var}(X) = \lambda.$$

Théorème 5.2.1 (Inégalité de Tchebychev) Si $\text{Var}(X)$ existe, on a pour tout $t > 0$

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

Démonstration : Par l'inégalité de Markov, on a

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

■

Application. On jette 3600 fois un dé. Minorer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre 480 et 720.

Notons S le nombre d'apparitions du 1. On peut voir S comme la somme de 3600 v.a. de Bernoulli indépendantes de paramètre $p = 1/6$ (probabilité d'apparition du 1 au cours d'un lancer). Par un raisonnement classique, S suit une loi $\mathcal{B}(3600, p)$. On cherche ici

$$\mathbb{P}(480 < S < 720) = \sum_{k=481}^{719} C_n^k p^k (1-p)^{n-k}.$$

Ce résultat exact ne peut être calculé en pratique, même un ordinateur très puissant ne pouvant calculer tous ces coefficients binomiaux pour des chiffres aussi grands.

On peut penser à approximer la loi $\mathcal{B}(3600, 1/6)$ par $\mathcal{P}(600)$ mais il resterait à calculer

$$\sum_{k=481}^{719} e^{-600} \frac{600^k}{k!},$$

ce qui n'est pas évident.

On a alors recours à l'inégalité de Tchebychev : notons que $E[S] = np = 3600/6 = 600$ et $\text{Var}(X) = npq = 3600 \times 5/6 \times 1/6 = 500$. Remarquons de plus que

$$480 < S < 720 \iff -120 < S - 600 < 120.$$

D'où

$$\begin{aligned}
 \mathbb{P}(480 < S < 720) &= \mathbb{P}(-120 < S - 600 < 120) = \mathbb{P}(|S - 600| < 120) \\
 &= 1 - \mathbb{P}(|S - 600| \geq 120) \\
 &\geq 1 - \frac{500}{120^2} \\
 &\geq 0,95833\dots
 \end{aligned}$$

Remarque 5.2.3 Les valeurs 480 et 720 sont symétriques par rapport à la moyenne 600 de la v.a. considérée, ce sont 600 ± 120 . Ce n'est pas nécessaire : on peut aussi appliquer l'inégalité de Tchebychev sur un intervalle non centré autour de l'espérance. Il suffit pour cela d'utiliser le plus grand intervalle centré sur l'espérance qu'il contient. Ainsi pour minorer $\mathbb{P}(550 < S < 700)$, il suffit de remarquer que

$$550 < S < 700 \quad \Longleftarrow \quad \underbrace{550 < S < 650}_{\text{intervalle centré autour de 600}} \quad \Longleftrightarrow \quad -50 < S - 600 < 50.$$

et

$$\begin{aligned}
 \mathbb{P}(550 < S < 700) &\geq \mathbb{P}(550 < S < 650) = \mathbb{P}(-50 < S - 600 < 50) \\
 &= \mathbb{P}(|S - 600| < 50) \\
 &= 1 - \mathbb{P}(|S - 600| \geq 50) \\
 &\geq 1 - \frac{500}{50^2} = 0,8.
 \end{aligned}$$

Chapitre 6

Variables aléatoires à valeurs réelles réelles

6.1 Généralités

Dans l'observation de grandeurs physiques (longueur, aire, volume, temps, poids), les données à modéliser prennent souvent une infinité de valeurs non nécessairement discrètes. D'où la nécessité de définir des variables aléatoires qui prennent un ensemble diffus de valeurs réelles.

Jusqu'à maintenant, pour les variables aléatoires discrètes, on s'est ramené aux probabilités des points où la v.a. est répartie (les masses ou probabilités ponctuelles $\mathbb{P}(X = x_k)$ en les atomes x_k). Pour les variables continues, les probabilités des points seront la plupart du temps nulles (il y a en quelque sorte trop de points pour qu'ils aient chacun une probabilité ponctuelle non nulle). Par contre les probabilités des intervalles ne s'annulent pas. Il apparaît alors pertinent de baser la théorie des v.a. réelles non plus sur les quantités du type $\mathbb{P}(X = x_k)$ mais sur $\mathbb{P}(X \in [a, b])$ où $[a, b]$ désigne un intervalle réel. La définition qui suit est donc motivée par la nécessité d'attribuer une définition cohérente aux probabilités des ensembles $\{\omega \in \Omega; X(\omega) \in [a, b]\} = \{X \in [a, b]\}$.

Définition 6.1.1 (Variable aléatoire réelle) *On appellera, dans ce cours, variable aléatoire réelle (notée v.a.r.) sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ une fonction $X : \Omega \rightarrow \mathbb{R}$, $\omega \mapsto X(\omega)$ dont le domaine $X(\Omega)$ est un intervalle (qui peut être borné ou une demi-droite ou encore \mathbb{R} tout entier).*

À nouveau, à chaque v.a.r., on associe sa loi. Elle définit une probabilité sur \mathbb{R} :

Définition 6.1.2 *Soit X une v.a.r. sur $(\Omega, \mathcal{F}, \mathbb{P})$. On lui associe la fonction d'ensembles \mathbb{P}_X qu'on considérera sur l'ensemble des intervalles de \mathbb{R} en posant*

$$\forall I \text{ intervalle de } \mathbb{R}, \quad \mathbb{P}_X(I) = \mathbb{P}(\omega \in \Omega; X(\omega) \in I) = \mathbb{P}(X \in I) = \mathbb{P}(X^{-1}(I)).$$

La fonction d'ensemble \mathbb{P}_X ainsi définie est une probabilité sur \mathbb{R} muni de la famille des observables obtenue à partir des intervalles. On l'appelle la loi de la v.a.r. X .

Il est facile de vérifier qu'il s'agit bien d'une probabilité sur \mathbb{R} : en effet, d'abord $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(\Omega) = 1$. Puis si $(I_n)_{n \in \mathbb{N}}$ est une famille d'intervalles de \mathbb{R} deux à deux disjoints, on a :

$$X^{-1}\left(\bigcup_n I_n\right) = \{\omega \in \Omega; X(\omega) \in \bigcup_n I_n\} = \bigcup_n \{\omega \in \Omega; X(\omega) \in I_n\} = \bigcup_n X^{-1}(I_n).$$

Comme les évènements $\{X \in I_n\}$, $n \in \mathbb{N}^*$, sont deux à deux disjoints, il suit par σ -additivité de \mathbb{P} :

$$\mathbb{P}_X\left(\bigcup_n I_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_n I_n\right)\right) = \mathbb{P}\left(\bigcup_n X^{-1}(I_n)\right) = \sum_n \mathbb{P}\left(X^{-1}(I_n)\right) = \sum_n \mathbb{P}_X(I_n).$$

Proposition 6.1.1 *Soit X une v.a.r. Sa loi \mathbb{P}_X est caractérisée par la fonction de répartition $F_X : \mathbb{R} \rightarrow [0, 1]$ définie par*

$$F_X(x) = \mathbb{P}_X(]-\infty, x]) = \mathbb{P}(X \in]-\infty, x]) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Ainsi en pratique pour montrer que deux v.a.r. X, Y ont même loi, il suffit de montrer que pour tous réels a, b on a $\mathbb{P}(X \in [a, b]) = \mathbb{P}(Y \in [a, b])$, ou que X, Y ont même fonction de répartition. La caractérisation de la loi par la fonction de répartition F_X se déduit facilement de

$$\mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a).$$

La fonction de répartition F_X d'une v.a.r. X jouit des mêmes propriétés que celles des v.a. discrètes qu'on énonce de la même façon :

Proposition 6.1.2 *La fonction de répartition d'une v.a. X satisfait :*

- $F_X(t) \geq 0$,
- F_X est croissante,
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$, $\lim_{t \rightarrow +\infty} F_X(t) = 1$,
- En tout point, F_X est continue à droite et a une limite à gauche, c'est à dire

$$\lim_{t \downarrow t_0} F_X(t) = F_X(t_0), \quad \text{et} \quad \lim_{t \searrow t_0} F_X(t) \text{ existe.}$$

Démonstration : Le premier point est clair car une probabilité est toujours positive. Le deuxième vient de ce que $\{X \leq s\} \subset \{X \leq t\}$ si $s \leq t$, car si $\omega \in \Omega$ vérifie $X(\omega) \leq s$ alors il vérifie a fortiori $X(\omega) \leq t$, il suit par croissance de \mathbb{P} :

$$F_X(s) = \mathbb{P}(X \leq s) \leq \mathbb{P}(X \leq t) = F_X(t).$$

Pour le dernier point : en tant que fonction croissante F_X a des limites à droite et à gauche en tout point (résultat élémentaire d'analyse) : en effet, si $h > 0$,

$$F_X(t_0 + h) - F_X(t_0) = \mathbb{P}(t_0 < X \leq t_0 + h) \xrightarrow{h \rightarrow 0} \mathbb{P}(t_0 < X \leq t_0) = \mathbb{P}(\emptyset) = 0$$

(rappelons qu'en passant à la limite, une inégalité devient forcément large).

D'où $\lim_{t \rightarrow t_0^+} F_X(t) = F_X(t_0)$. Alors que

$$F_X(t_0) - F_X(t_0 - h) = \mathbb{P}(t_0 - h < X \leq t_0) \xrightarrow{h \rightarrow 0} \mathbb{P}(t_0 \leq X \leq t_0) = \mathbb{P}(X = t_0) \geq 0$$

éventuellement non nul si X a un atome en t_0 . On a donc $\lim_{t \rightarrow t_0^-} F_X(t) = \mathbb{P}(X < t_0)$ existe.

Le reste de la preuve vient maintenant des propriétés de monotonie séquentielle des probabilités vues au chapitre 1 : Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite croissante d'évènements (i.e. pour tout n , $A_n \subset A_{n+1}$) alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}(A) \text{ où } A = \bigcup_{n \in \mathbb{N}^*} A_n. \quad (6.1)$$

Si $(B_n)_{n \in \mathbb{N}^*}$ est une suite décroissante d'évènements (i.e. pour tout n , $B_{n+1} \subset B_n$) alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}(B) \text{ où } B = \bigcap_{n \in \mathbb{N}^*} B_n. \quad (6.2)$$

La limite 0 en $-\infty$ vient alors de (6.2) appliquée à \mathbb{P}_X et $B_n =]-\infty, -n]$ pour lequel $B = \bigcap_{n \in \mathbb{N}^*} B_n = \bigcap_{n \in \mathbb{N}^*}]-\infty, -n] = \emptyset$ et qui donne

$$\lim_{n \rightarrow \infty} F_X(-n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(B_n) = \mathbb{P}_X(\emptyset) = 0.$$

La limite 1 en $+\infty$ vient alors de (6.1) appliquée à \mathbb{P}_X et $A_n =]-\infty, n]$ pour lequel $A = \bigcup_{n \in \mathbb{N}^*} A_n = \bigcup_{n \in \mathbb{N}^*}]-\infty, n] = \mathbb{R}$ et qui donne

$$\lim_{n \rightarrow \infty} F_X(n) = \lim_{n \rightarrow +\infty} \mathbb{P}_X(A_n) = \mathbb{P}_X(\mathbb{R}) = 1.$$

La continuité à droite en x s'obtient aussi en appliquant (2.3) à \mathbb{P}_X et $B_n =]-\infty, x + 1/n]$ avec $B =]-\infty, x]$. On a

$$\lim_{n \rightarrow +\infty} F_X(x + 1/n) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n) = \mathbb{P}_X(B) = F_X(x).$$

■

Signalons que ce résultat admet la réciproque –culturelle– suivante :

Théorème 6.1.1 *Soit F une fonction définie et croissante sur \mathbb{R} . On suppose de plus que F est continue à droite, qu'elle admet une limite à gauche en tout point et qu'elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Alors, il existe un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et une v.a.r. X définie sur cet espace et ayant F pour fonction de répartition.*

6.2 Intégrales impropres

Lorsque l'on intègre une fonction sur un domaine non borné ou sur un domaine où elle possède une singularité (i.e. où elle n'est pas définie), on parle d'intégrale impropre.

Fonctions positives

Dans le cas de l'intégration d'une fonction f sur \mathbb{R}^+ , on définit son intégrale impropre, si elle existe, par

$$\int_0^{+\infty} f(x)dx = \lim_{M \rightarrow +\infty} \int_0^M f(x)dx.$$

Si f est définie sur \mathbb{R} , son intégrale impropre est donnée, si elle existe, par

$$\int_{-\infty}^{+\infty} f(x)dx = \lim_{M \rightarrow +\infty} \int_{-M}^M f(x)dx.$$

Là encore, on parle d'intégrale convergente ou divergente selon que la limite existe et est finie ou non.

Critère d'intégrabilité en $x_0 \in \mathbb{R}$: (souvent $x_0 = 0$) Soit $f : [a, b] \rightarrow \mathbb{R}^+$ et $x_0 \in]a, b[$, si $\exists \alpha < 1$ tel que

$$\lim_{x \rightarrow x_0} (x - x_0)^\alpha f(x) = 0$$

alors f est intégrable au voisinage de x_0 .

Par exemple : $x \mapsto 1/\sqrt{|x-1|}$ est intégrable en 1, $x \mapsto 1/|x-1|$ ne l'est pas.

Critère d'intégrabilité en $+\infty$: Soit $f : [A, +\infty) \rightarrow \mathbb{R}^+$, si $\exists \alpha > 1$ tel que

$$\lim_{x \rightarrow +\infty} x^\alpha f(x) = 0$$

alors f est intégrable en $+\infty$.

Par exemple, $x \mapsto 1/x^a$ est intégrable en $+\infty$ ssi $a > 1$, $x \mapsto e^{-x^2}$ est intégrable en $+\infty$, $x \mapsto 1/\log x$ ne l'est pas.

Fonctions de signe quelconque

Une fonction f de signe quelconque est dite absolument intégrable si la fonction positive $|f|$ est intégrable.

L'absolue intégrabilité entraîne l'intégrabilité (simple).

6.3 Variables aléatoires réelles à densité

C'est essentiellement ce type de v.a.r. que l'on considérera dans la suite de ce cours.

La loi d'une v.a.r. est à densité f si pour tout intervalle de \mathbb{R} , la probabilité d'appartenance de X à cet intervalle s'exprime comme l'intégrale de f sur cet intervalle.

Définition 6.3.1 Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est appelée densité de probabilité si

– f est positive : en tout point t où elle est définie $f(t) \geq 0$,

– f est intégrable sur \mathbb{R} d'intégrale 1 :

$$\int_{-\infty}^{+\infty} f(t) dt = 1.$$

Définition 6.3.2 La v.a.r. X suit la loi de densité f si

$$\forall [a, b] \text{ intervalle de } \mathbb{R}, \quad \mathbb{P}(X \in [a, b]) = \int_a^b f(t) dt.$$

Notons que pour une v.a. X de densité f , la probabilité que X vaille un point est 0, car c'est une intégrale sur un intervalle réduit à un point :

$$\mathbb{P}(X = x_0) = \mathbb{P}(X \in [x_0, x_0]) = \int_{x_0}^{x_0} f(x) dx = 0.$$

Par conséquent, le sens des bornes des intervalles (fermées ou ouvertes) n'est pas important :

$$\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in]a, b]) \quad \text{et} \quad \mathbb{P}(X \leq t) = \mathbb{P}(X < t),$$

en effet la différence est la probabilité que X vaille un point, ce qui est 0.

Il est clair que si Y est une v.a.r. de même loi que X alors elle a aussi la densité f . Il serait plus correct de parler de la densité de la loi.

La densité d'une v.a. réelle joue le rôle pour une v.a. discrète de ses probabilités ponctuelles $\mathbb{P}(X = x_k)$.

Exemples : Quelques exemples de densité de var

$$\begin{aligned} f_1(t) &= \frac{1}{b-a} \mathbf{1}_{[a,b]}(t), & f_2(t) &= \frac{1}{2\sqrt{t}} \mathbf{1}_{[0,1]}(t), \\ f_3(t) &= e^{-t} \mathbf{1}_{[0,+\infty[}(t), & f_4(t) &= \frac{1}{\pi(1+t^2)} \end{aligned}$$

avec la notation suivante pour la fonction indicatrice d'un ensemble A

$$\mathbf{1}_A(t) = \begin{cases} 1 & \text{si } t \in A \\ 0 & \text{si } t \notin A. \end{cases}$$

Généralement, les densités que nous considérerons seront de l'un de deux types suivants

- f est définie et continue sur \mathbb{R} et son intégrale de Riemann généralisée $\int_{-\infty}^{+\infty} f(t) dt$ converge et vaut 1.
- f est définie sur \mathbb{R} privé d'un point ou d'un ensemble fini de point, $a_1 < \dots < a_n$. Sur chaque intervalle ouvert $] -\infty, a_1[$, \dots , $]a_i, a_{i+1}[$, \dots , $]a_n, +\infty[$, f est continue et a une intégrale de Riemann (généralisée ou non) convergente et la somme de ces intégrales vaut 1.

Dans l'exemple ci-dessus f_1, f_2, f_3 sont du deuxième type, f_4 du premier.

Proposition 6.3.1 (Domaine d'une v.a. à densité) *Le domaine d'une v.a.r. à densité coïncide avec le support de sa densité.*

Démonstration : Soit X une v.a.r. de densité f . On note D le support de $f : \forall x \notin D, f(x) = 0$. Pour simplifier, on suppose f continue (ou continue par morceaux), dans ce cas $D = f^{-1}\{\mathbb{R}^*\}$. Pour tout intervalle $I \subset D^c$, on a

$$\mathbb{P}(X \in I) = \int_I f(t) dt = 0$$

car f est nulle sur $I \subset D^c$. Les seuls intervalles que X visite avec une probabilité non nulle sont ceux inclus dans le support D de sa densité.

Réciproquement, si $\mathbb{P}(X \in I) = 0$ alors $\int_I f(t) dt = 0$ et comme f est positive et continue, on a $f(t) = 0$ pour tout $t \in I$. Il suit $I \cap D = \emptyset$. On en déduit $X(\Omega) = D$. ■

Lorsqu'elle existe la densité f est naturellement reliée à la fonction de répartition F_X :

Proposition 6.3.2 *Si X est une v.a.r. de densité f , sa fonction de répartition F_X vérifie :*

- (i) $\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t) dt$.
- (ii) F_X est continue sur \mathbb{R} .
- (iii) Si f est continue au point x_0 , alors F_X est dérivable en x_0 de dérivée $F'_X(x_0) = f(x_0)$.

D'après (ii), la fonction de répartition est continue. De là, vient aussi qu'on parle de variable aléatoire continue pour v.a. à densité.

Démonstration : Puisque X a pour densité f , et comme

$$F_X(b) = \mathbb{P}(X \in]-\infty, b]) = \mathbb{P}(X \in]-\infty, a] \cup]a, b]) = F_X(a) + \mathbb{P}(X \in]a, b]),$$

on a pour tous réels $a < b$:

$$\mathbb{P}(\omega, X(\omega) \in]a, b]) = \mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a) = \int_a^b f(t) dt. \quad (6.3)$$

(i) : Il suffit d'appliquer (6.3) avec $b = x$ fixé et $a = -n$ pour chaque $n \in \mathbb{N}$ tel que $x > -n$. La suite d'évènements

$$A_n = \{\omega, X(\omega) \in]-n, x]\}, \quad n > -x,$$

est croissante pour l'inclusion et de réunion $A = \{\omega, X(\omega) \in]-\infty, x]\} = \{X \leq x\}$. Par la propriété de monotonie séquentielle, on a $\mathbb{P}(A_n) \nearrow \mathbb{P}(A)$, d'où

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(A) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \lim_{n \rightarrow +\infty} \int_{-n}^x f(t) dt = \int_{-\infty}^x f(t) dt$$

en notant que l'intégrale généralisée de la densité f converge en $-\infty$.

(ii) : On fixe $x_0 \in \mathbb{R}$ quelconque. D'abord F_X est continue à droite en tout point car c'est une fonction de répartition et que c'est vrai de toute fonction de répartition (cf. Prop. 6.1.2).

Il reste à voir la continuité à gauche. On se contente de le faire avec l'hypothèse supplémentaire suivante : « il existe $a < x_0$ tel que f soit définie et Riemann intégrable sur tout intervalle $[a, a'] \supset [a, x_0]$ ». On a alors :

$$\lim_{x \searrow x_0} \int_a^x f(t) dt = \int_a^{x_0} f(t) dt,$$

où la deuxième intégrale est soit une intégrale de Riemann ordinaire soit une intégrale de Riemann impropre convergente. On peut réécrire

$$\lim_{x \searrow x_0} (F_X(x) - F_X(a)) = F_X(x_0) - F_X(a).$$

On conclut en rajoutant des deux côtés $F_X(a)$.

(iii) : Comme par hypothèse f est continue en x_0 , elle est définie sur tout un voisinage de x_0 et donc sur un intervalle $[a, b]$ qui contient x_0 . La continuité de f en x_0 s'écrit : $\forall \varepsilon > 0, \exists \delta > 0$ tel que $]x_0 - \delta, x_0 + \delta[\subset]a, b[$ et

$$\forall t \in]x_0 - \delta, x_0 + \delta[, \quad |f(t) - f(x_0)| < \varepsilon.$$

Pour tout h tel que $0 < |h| < \delta$, on a alors $F_X(x_0 + h) - F_X(x_0) = \int_{x_0}^{x_0+h} f(t) dt$. D'où

$$\begin{aligned} |F_X(x_0 + h) - F_X(x_0) - hf(x_0)| &= \left| \int_{x_0}^{x_0+h} f(t) dt - \int_{x_0}^{x_0+h} f(x_0) dt \right| \\ &= \left| \int_{x_0}^{x_0+h} (f(t) - f(x_0)) dt \right| \leq \int_{x_0}^{x_0+h} |f(t) - f(x_0)| dt \\ &\leq h\varepsilon. \end{aligned}$$

En divisant par h puis en faisant $h \rightarrow 0$, on constate que F_X est dérivable en x_0 , de dérivée $f'(x_0)$. ■

6.4 Lois à densité classiques

Définition 6.4.1 La fonction indicatrice d'un ensemble A est

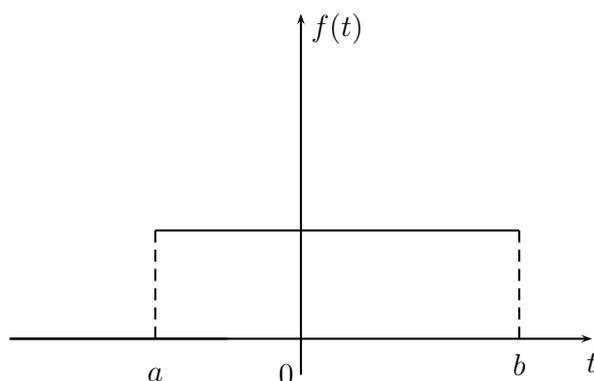
$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{si } x \notin A. \end{cases}$$

6.4.1 Lois uniformes

Définition 6.4.2 La v.a.r. X suit une loi uniforme sur l'intervalle $[a, b]$ ($-\infty < a < b < +\infty$) si elle a une densité f constante sur cet intervalle et nulle en dehors. Sa densité est alors

$$f(t) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(t) = \begin{cases} 1/(b-a) & \text{si } t \in [a, b], \\ 0 & \text{si } t \notin [a, b]. \end{cases}$$

Cette loi est l'équivalent continu de la loi discrète équirépartie. L'allure de la densité d'une v.a. de loi uniforme est :

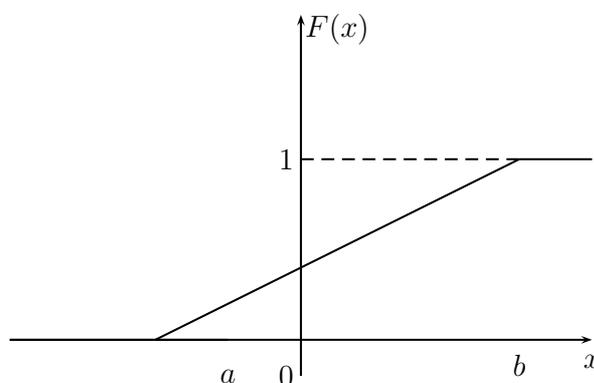


Remarque 6.4.1 Le facteur $1/(b-a)$ permet de normaliser l'intégrale de f sur \mathbb{R} pour que $\int_{-\infty}^{+\infty} f(t) dt = 1$. On comprend bien dès lors pourquoi on ne parle de lois uniformes que sur les intervalles **finis** : si a ou b est infini le facteur de normalisation est nul et la densité f vaut 0 partout. Son intégrale ne peut plus dès lors valoir 1.

Sa fonction de répartition est affine par morceaux :

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{si } -\infty < x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b \leq x < +\infty. \end{cases}$$

L'allure de la fonction de répartition d'une v.a. de loi uniforme est :



Le résultat suivant permet d'éviter des calculs fastidieux pour la probabilité uniforme d'un intervalle.

Proposition 6.4.1 *Si X est une v.a.r. de loi uniforme sur $[a, b]$ alors pour tout intervalle I de \mathbb{R} :*

$$\mathbb{P}(X \in I) = \frac{l([a, b] \cap I)}{l([a, b])}$$

où $l(J)$ désigne la longueur de l'intervalle J ($l([a, b]) = b - a$).

6.4.2 Lois exponentielles

Définition 6.4.3 *Soit a un réel strictement positif. La v.a.r. X suit une loi exponentielle de paramètre a si elle admet pour densité :*

$$f(t) = ae^{-at}\mathbf{1}_{[0, +\infty[}(t).$$

Elle est notée $\mathcal{E}(a)$.

En pratique, à la place de la fonction de répartition, on utilise souvent la fonction de survie G d'une v.a.r. de loi exponentielle

$$G_X(x) = \mathbb{P}(X > x) = 1 - F_X(x) = \begin{cases} 1 & \text{si } x \leq 0, \\ e^{-ax} & \text{si } x \geq 0. \end{cases}$$

Les lois exponentielles sont souvent utilisées pour modéliser des temps d'attente ou des durées de vie. Par exemple, les temps d'attente à partir de maintenant du prochain tremblement de terre, de la prochaine panne d'un appareil, de la prochaine désintégration dans un réacteur nucléaire suivent des lois exponentielles. On verra bientôt que le paramètre a désigne alors l'inverse du temps d'attente moyen.

Une propriété intéressante de ce type de loi est l'absence de mémoire. Cette propriété caractérise les lois exponentielles.

Théorème 6.4.1 (i) *Si la v.a.r. X suit une loi exponentielle alors elle vérifie la propriété d'absence de mémoire :*

$$\forall s \in \mathbb{R}^+, \quad \forall t \in \mathbb{R}^+, \quad \mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s). \quad (6.4)$$

(ii) *Réciproquement, si une v.a.r. X vérifie (6.4) alors elle suit une loi exponentielle.*

Autrement dit si X survit jusqu'en t , sa survie pendant encore s unités de temps est la même qu'une survie de durée s depuis le départ : tout se passe comme si, ce qui se passe de 0 à t est oublié pour survivre encore s unités de temps. C'est à comparer, par exemple, avec la vie humaine qui a une mémoire : pour un homme de 60, la probabilité de vivre encore 30 ans n'est pas la même que pour celle d'un nouveau né (de 0 an).

Démonstration : Pour (i) :

$$\begin{aligned}\mathbb{P}(X > t + s \mid X > t) &= \frac{\mathbb{P}(X > t + s \text{ et } X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-a(t+s)}}{e^{-at}} = e^{-as} = \mathbb{P}(X > s)\end{aligned}$$

d'après l'expression de la fonction de survie. Le point (ii), un peu plus délicat, est admis. ■

6.4.3 Lois de Cauchy

Définition 6.4.4 Une variable aléatoire réelle suit une loi de Cauchy de paramètre $a \in \mathbb{R}_+^*$ si elle admet pour densité :

$$f(t) = \frac{a}{\pi} \frac{1}{a^2 + t^2}.$$

Exercice. Montrer qu'il s'agit bien d'une densité.

6.4.4 Lois normales ou gaussiennes

Elles jouent un rôle capital dans l'étude des lois limites de sommes de variables aléatoires indépendantes (*cf.* le théorème central limite, résultat central comme son nom l'indique en théorie des probabilités). On parle encore de loi gaussiennes.

Définition 6.4.5 On dit que la v.a.r. X suit une loi gaussienne ou normale $\mathcal{N}(m, \sigma^2)$ si elle a pour densité la fonction :

$$f_{m,\sigma} : \mathbb{R} \longrightarrow \mathbb{R}, \quad t \longmapsto \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

La loi normale standard $\mathcal{N}(0, 1)$ est celle de densité $f_{0,1}(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.

Remarque 6.4.2 Cette loi est fondamentale en théorie des probabilités et en statistique : c'est la loi limite de la moyenne dans une suite infinie d'épreuves répétées indépendantes. En pratique elle sert à modéliser les effets additifs de petits phénomènes aléatoires indépendants répétées souvent.

On parle de densité gaussienne pour $f_{m,\sigma}$. Le paramètre m est un paramètre de localisation (c'est la valeur où $f_{m,\sigma}$ atteint son maximum), le paramètre σ est un paramètre d'échelle. Nous verrons que ce sont en fait la moyenne et l'écart-type de la loi.

La fonction $f_{0,1}$ s'appelle la densité normale standard (ou gaussienne standard). Sa courbe représentative est bien connue, il s'agit de la « courbe en cloche » (ou courbe de Gauss) à laquelle il est souvent fait référence. Les courbes des $f_{m,\sigma}$ sont aussi des « courbes en cloche » obtenues par translation et dilatation de celle de $f_{0,1}$ (ce phénomène est lié à la Proposition 6.4.2).

Notez que le facteur $1/\sqrt{2\pi}$ dans les densités gaussiennes a été choisi car on montre que

$$\int_{-\infty}^{+\infty} e^{-t^2/2} dt = \sqrt{2\pi}.$$

Notez encore qu'on peut facilement passer d'une loi normale à la loi standard :

Proposition 6.4.2 *Si la v.a.r. X suit une loi $\mathcal{N}(m, \sigma^2)$, alors $Y := \frac{X - m}{\sigma}$ suit la loi $\mathcal{N}(0, 1)$.*

La v.a. Y s'appelle la v.a. centrée réduite associée à X .

Démonstration : Calculons pour $a < b$ quelconques $\mathbb{P}(a \leq Y \leq b)$:

$$\begin{aligned} \mathbb{P}\left(a \leq \frac{X - m}{\sigma} \leq b\right) &= \mathbb{P}(\sigma a + m \leq X \leq \sigma b + m) \\ &= \int_{\sigma a + m}^{\sigma b + m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t - m)^2}{2\sigma^2}\right) dt. \end{aligned}$$

Il suffit alors de faire le changement de variable $s = (t - m)/\sigma$ pour obtenir

$$\forall a \in \mathbb{R}, \forall b > a, \quad \mathbb{P}(a \leq Y \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s^2}{2}\right) ds,$$

c'est à dire Y suit la loi $\mathcal{N}(0, 1)$. ■

6.4.5 Lois log-normales

Définition 6.4.6 *Une variable aléatoire réelle X suit une loi log-normale si elle admet la densité*

$$f(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{1}{\sigma t\sqrt{2\pi}} \exp\left(-\frac{(\ln t - m)^2}{2\sigma^2}\right) & \text{si } t \geq 0 \end{cases}$$

où $m \in \mathbb{R}, \sigma \in \mathbb{R}^*$.

Cette loi est l'analogie multiplicatif de la loi normale : elle modélise les effets multiplicatifs de phénomènes aléatoires nombreux et indépendants.

La terminologie vient de ce que :

Proposition 6.4.3 *Si X est de loi log-normale alors $\ln(X)$ suit une loi normale et réciproquement.*

Démonstration : En effet X est log-normale si et seulement si pour $x \geq 0$

$$F_X(x) = \int_0^x \frac{1}{\sigma t\sqrt{2\pi}} \exp\left(-\frac{(\ln t - m)^2}{2\sigma^2}\right) dt$$

$$\begin{aligned}
&= \int_{-\infty}^{\ln(x)} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(s-m)^2}{2\sigma^2}\right) ds \quad \text{en posant } s = \ln(t), \\
&= F_Y(\ln(x))
\end{aligned}$$

où Y est une variable aléatoire de loi normale $\mathcal{N}(m, \sigma^2)$.

On a alors

$$F_{\ln X}(x) = \mathbb{P}(\ln X \leq x) = \mathbb{P}(X \leq e^x) = F_X(e^x) = F_Y(\ln e^x) = F_Y(x).$$

La variable $\ln X$ a la même fonction de répartition que Y , variable normale, elle est donc normale de paramètres m, σ^2 . ■

6.5 Espérance et variance des lois à densité

Définition 6.5.1 (Espérance d'une v.a.r. à densité) Si X est une v.a.r. de densité f telle que l'intégrale $\int_{-\infty}^{+\infty} |x|f(x) dx$ converge, on appelle espérance de X le réel (fini)

$$E[X] = \int_{-\infty}^{+\infty} xf(x) dx. \quad (6.5)$$

Remarque 6.5.1 Noter la ressemblance formelle du cas continu avec le cas discret : on a juste remplacé \sum par \int , et il s'agit encore de faire la moyenne des x

- pondérés par la densité $f(x)$ dans le cas continu avec densité,
- pondérés par les probabilités ponctuelles $\mathbb{P}(X = x_k)$ dans le cas discret.

Il faut retenir que la densité du cas continu est l'équivalent des probabilités ponctuelles du cas discret.

Exemples : espérance de v.a. réelles à densité

- Si X suit une loi uniforme sur $[a, b]$ son espérance est

$$E[X] = \frac{1}{b-a} \int_{-\infty}^{+\infty} t \mathbf{1}_{[a,b]}(t) dt = \frac{1}{b-a} \int_a^b t dt = \frac{1}{b-a} \left[\frac{t^2}{2} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

- Si X suit une loi exponentielle de paramètre $a > 0$, son espérance est (en intégrant par parties)

$$\begin{aligned}
E[X] &= \int_{-\infty}^{+\infty} ate^{-at} \mathbf{1}_{\mathbb{R}^+}(t) dt = \int_0^{+\infty} at e^{-at} dt = \left[-te^{-at} \right]_0^{+\infty} + \int_0^{+\infty} e^{-at} dt \\
&= 0 + \left[\frac{e^{-at}}{a} \right]_0^{+\infty} = 1/a.
\end{aligned}$$

• Si X suit une loi de Cauchy de paramètre a alors, l'espérance n'est pas définie. En effet,

$$\int_{-\infty}^{+\infty} \frac{a|t| dt}{\pi(a^2 + t^2)} = +\infty,$$

car $t/(a^2 + t^2) \simeq_{+\infty} 1/t$ qui n'est pas intégrable en $+\infty$ (on fait de même en $-\infty$). Finalement, la condition d'existence de l'espérance n'est pas remplie.

• Si X suit une loi normale $\mathcal{N}(m, \sigma^2)$, son espérance vaut $E[X] = m$: comme $Y := \frac{X - m}{\sigma}$ suit une loi $\mathcal{N}(0, 1)$ (cf. la proposition 6.4.2), il suffit de voir d'après la linéarité (justifiée dans la proposition 6.5.1 à venir) $E[Y] = 0$ c'est à dire

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-t^2/2} dt = 0$$

ce qui est clair car comme l'intégrand est impair :

$$\int_{-\infty}^0 te^{-t^2/2} dt = - \int_0^{+\infty} se^{-s^2/2} ds \quad \text{avec le changement de variables } s = -t$$

et donc

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} te^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} te^{-t^2/2} dt - \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} se^{-s^2/2} ds = 0.$$

(Au passage, noter que la convergence de l'intégrale ne pose pas de problème grâce au facteur $e^{-t^2/2}$ qui assure que les critères de convergence sont vérifiés car par exemple $t^2 e^{-t^2/2} \rightarrow 0, t \rightarrow \pm\infty$, cf. page 56.)

Les principales propriétés des espérances ont déjà été vues précédemment pour les v.a. discrètes. Elles ont leurs analogues pour des v.a. à densité. On se contente de les citer, les preuves étant essentiellement de simples adaptations de celles déjà vues.

Proposition 6.5.1 (Linéarité de l'espérance) Soient X et Y deux v.a.r. admettant des espérances. Alors

- (1) $E[X + Y] = E[X] + E[Y]$,
- (2) Pour tout réel a , $E[aX] = aE[X]$.

Démonstration : On prouve seulement le 2) pour $a > 0$:

$$F_{aX}(x) = \mathbb{P}(aX \leq x) = \mathbb{P}(X \leq x/a) = F_X(x/a) = \int_{-\infty}^{x/a} f_X(t) dt$$

En dérivant, on déduit la densité de aX : $f_{aX}(t) = F'_{aX}(t) = \frac{1}{a} f_X(t/a)$ et

$$E[aX] = \int_{\mathbb{R}} u f_{aX}(u) du = \int_{\mathbb{R}} \frac{u}{a} f_X\left(\frac{u}{a}\right) du = a \int_{\mathbb{R}} x f_X(x) dx$$

avec le changement de variable $x = u/a$. ■

Proposition 6.5.2 (Théorème de transfert : Espérance d'une fonction de v.a.) Soient X une v.a.r. de densité f et F une fonction numérique sur \mathbb{R} (ou dont l'ensemble de définition contient au moins l'ensemble des valeurs de X) continue par morceaux. Alors si $E[|F(X)|]$ existe, on a

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

On a de plus facilement avec la fonction numérique $F(x) = |x|$:

Proposition 6.5.3 (Espérance et valeurs absolues) $E[|X|] = \int_{-\infty}^{+\infty} |x|f(x)dx$. Et si $E[X]$ existe, $|E[X]| \leq E[|X|]$.

Notons que la condition de définition des espérances (cf. Déf. 6.5.1) pour l'existence de $E[X]$ n'est donc à nouveau rien d'autre que $E[|X|] < +\infty$.

Proposition 6.5.4 (Positivité de l'espérance)

- Si X a une espérance et $X \geq 0$, alors $E[X] \geq 0$.
- Si X et Y ont des espérances et $X \leq Y$ alors $E[X] \leq E[Y]$.

Démonstration : Si $X \geq 0$ alors $\mathbb{P}(X \leq 0) = \int_{-\infty}^0 f(t) dt = 0$. Comme f est positive et d'intégrale nulle sur \mathbb{R}^- alors $f(t) = 0$ si $t < 0$. On a donc

$$E[X] = \int_{-\infty}^{+\infty} tf(t) dt = \int_0^{+\infty} tf(t) dt.$$

Comme sur \mathbb{R}^+ , $t \geq 0$ et f est positive, l'intégrale précédente est positive.

Pour la deuxième partie, appliquer la première à la v.a. positive $Y - X$ puis utiliser la linéarité de l'espérance. ■

Soulignons qu'on dispose toujours de l'inégalité de Markov :

Théorème 6.5.1 (Inégalité de Markov) Si X est une v.a.r. positive à densité ayant une espérance alors

$$\forall t > 0, \quad \mathbb{P}(X \geq t) \leq \frac{E[X]}{t}.$$

Démonstration : Si on note f la densité de la v.a. X , on a $f(t) = 0$ si $t < 0$ car X est à valeurs positives. Puis

$$\begin{aligned} E[X] &= \int_0^{+\infty} xf(x)dx = \int_0^t xf(x)dx + \int_t^{+\infty} xf(x)dx \\ &\geq 0 + \int_t^{+\infty} xf(x)dx \geq t \int_t^{+\infty} f(x)dx \\ &\geq t\mathbb{P}(X \in [t, +\infty)) = t\mathbb{P}(X \geq t). \end{aligned}$$

On redéfinit aussi les moments :

Définition 6.5.2 Soit $r \in \mathbb{N}^*$, on appelle moment d'ordre r de la v.a.r. X de densité f le nombre

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

Lorsqu'il est fini, on considère aussi

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

Lorsque une v.a. est bornée, il est facile de voir qu'elle a des moments de tous les ordres. De façon générale, l'existence de moment d'ordre r implique celles des moments d'ordre inférieurs. En effet, on montre comme dans le cas discret (cf. Prop. 5.2.1) que pour $n \leq r$:

$$E[|X|^n] \leq 1 + E[|X|^r].$$

On définit en particulier la variance et l'écart-type de la même façon que pour les v.a. discrètes.

$$\text{Var}(X) = E[(X - E[X])^2], \quad \sigma_X = \sqrt{\text{Var}(X)}.$$

On dispose des mêmes propriétés que dans le cas des v.a. discrètes :

–

$$\text{Var}(aX) = a^2 \text{Var}(X), \quad \text{Var}(X + b) = \text{Var}(X), \quad \forall a, b \in \mathbb{R}.$$

– **Formule de Koenig :**

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

– **Inégalité de Tchebychev :** si $\text{Var}(X)$ existe, on a

$$\mathbb{P}(|X - E[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}, \quad \forall t > 0.$$

Exemples : variances des lois à densités usuelles

• Si X suit une loi uniforme sur $[a, b]$ sa variance est

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

En effet

$$\begin{aligned} E[X^2] &= \frac{1}{b - a} \int_a^b t^2 dt = \frac{b^3 - a^3}{3(b - a)} = \frac{a^2 + ab + b^2}{3} \\ \text{Var}(X) &= \frac{a^2 + ab + b^2}{3} - \frac{(a + b)^2}{4} = \frac{(b - a)^2}{12}. \end{aligned}$$

• Si X suit une loi exponentielle de paramètre a , sa variance est $1/a^2$: en intégrant par parties (deux fois),

$$\begin{aligned} E[X^2] &= \int_0^{+\infty} at^2 e^{-at} dt = [-t^2 e^{-at}]_0^{+\infty} + \int_0^{+\infty} 2te^{-at} dt \\ &= 0 - 2 \left[\frac{te^{-at}}{a} \right]_0^{+\infty} + 2 \int_0^{+\infty} \frac{e^{-at}}{a} dt = 0 - 2 \left[\frac{e^{-at}}{a^2} \right]_0^{+\infty} = 2/a^2 \\ \text{Var}(X) &= E[X^2] - E[X]^2 = 2/a^2 - 1/a^2 = 1/a^2. \end{aligned}$$

• Si X suit une loi de Cauchy alors elle n'a pas de moment d'ordre 2 car n'en a déjà pas d'ordre 1 (pas de moyenne). Elle n'admet donc pas de variance.

• Si X suit une loi normale $\mathcal{N}(m, \sigma^2)$, sa variance vaut $\text{Var}(X) = \sigma^2$, en effet comme $Y := \frac{X - m}{\sigma}$ est de loi $\mathcal{N}(0, 1)$ et que $\text{Var}(X) = \sigma^2 \text{Var}(Y)$ (cf. Prop. 6.4.2), il suffit de voir $\text{Var}(Y) = E[Y^2] - E[Y]^2 = 1$. Or $E[Y] = 0$ et

$$\begin{aligned} E[Y^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-t^2/2} dt \\ &= \frac{1}{\sqrt{2\pi}} [-te^{-t^2/2}]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-t^2/2} dt \\ &= 0 + 1 = 1. \end{aligned}$$

On a bien $E[Y^2] = 1$, $\text{Var}(Y) = 1$, $\text{Var}(X) = \sigma^2$.

6.6 Tableau comparatif des formules pour des v.a. discrètes et continues à densité

Lorsque les intégrales et les séries concernées sont absolument convergentes, on a le tableau comparatif suivant entre le cas discret et le cas continu :

X	Variable discrète	Variable à densité f
$X(\Omega)$	$\{x_1, x_2, \dots, x_k, \dots\}$	\mathbb{R} ou un intervalle
$\mathbb{P}(a \leq X \leq b)$	$\sum_{a \leq x_k \leq b} \mathbb{P}(X = x_k)$	$\int_a^b f(t) dt$
$F(x) = \mathbb{P}(X \leq x)$	$\sum_{x_k \leq x} \mathbb{P}(X = x_k)$	$\int_{-\infty}^x f(t) dt$
$E[X]$	$\sum_{k=1}^{+\infty} x_k \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} t f(t) dt$
$E[g(X)]$	$\sum_{k=1}^{+\infty} g(x_k) \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} g(t) f(t) dt$
$E[X^2]$	$\sum_{k=1}^{+\infty} x_k^2 \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} t^2 f(t) dt$
$\text{Var}(X)$	$\sum_{k=1}^{+\infty} (x_k - E[X])^2 \mathbb{P}(X = x_k)$	$\int_{-\infty}^{+\infty} (t - E[X])^2 f(t) dt$

Chapitre 7

Vecteurs aléatoires

Dans des situations où interviennent plusieurs variables aléatoires, le calcul de la probabilité d'un événement dont la réalisation dépend des valeurs de ces variables doit faire intervenir ces variables considérées dans *leur ensemble* et non chacune isolément. Cela amène ainsi à étudier une nouvelle notion : celle de vecteur aléatoire.

7.1 Généralités

Définition 7.1.1 Soient X, Y des v.a. définies sur le même espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. L'application

$$\Omega \longrightarrow \mathbb{R}^2, \quad \omega \longmapsto (X(\omega), Y(\omega))$$

est appelé couple aléatoire, on le note (X, Y) . Les variables aléatoires X et Y sont alors appelées ses marginales.

Définition 7.1.2 De même, si X_1, X_2, \dots, X_n sont n variables aléatoires, sur le même espace $(\Omega, \mathcal{F}, \mathbb{P})$, on définit le vecteur aléatoire (X_1, \dots, X_n) comme l'application

$$\Omega \longrightarrow \mathbb{R}^n, \quad \omega \longmapsto (X_1(\omega), \dots, X_n(\omega)).$$

La v.a. X_i est appelée la i -ème marginale du vecteur. Pour $n = 2$, on retrouve les couples aléatoires.

Le couple aléatoire (X, Y) permet de transporter la probabilité \mathbb{P} de l'espace Ω sur l'espace \mathbb{R}^2 . Rappelons qu'un produit cartésien $A \times B$ de deux ensembles $A, B \subset \mathbb{R}$ désigne l'ensemble suivant de \mathbb{R}^2 :

$$A \times B = \{(a, b) \text{ tel que } a \in A \text{ et } b \in B\}.$$

Définition 7.1.3 La loi $\mathbb{P}_{X,Y}$ du couple (X, Y) est la probabilité définie sur l'ensemble des produits d'intervalles $I \times J$ de \mathbb{R}^2 par $\forall I, J$ intervalles de \mathbb{R}

$$\begin{aligned} \mathbb{P}_{X,Y}(I \times J) &= \mathbb{P}(\omega \in \Omega; (X(\omega), Y(\omega)) \in I \times J) \\ &= \mathbb{P}(X \in I, Y \in J). \end{aligned} \tag{7.1}$$

Remarque 7.1.1 (pour des v.a. discrètes)

- À nouveau, s'il s'agit de couple de v.a. discrètes, il n'est pas utile de se restreindre à définir $\mathbb{P}_{X,Y}$ sur les produits d'intervalles $I \times J$. On peut définir la loi $\mathbb{P}_{X,Y}$ sur tout $\mathcal{P}(\mathbb{R}^2)$, l'ensemble des parties de \mathbb{R}^2 .
- Dans le cas de v.a. X, Y discrètes, il est facile de voir que la loi du couple (X, Y) est caractérisée par les probabilités ponctuelles $\mathbb{P}_{X,Y}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$ pour tout $x_i \in X(\Omega), y_j \in Y(\Omega)$ dans les domaines de valeurs de X et de Y .

7.2 Vecteurs aléatoires discrets

Rappelons encore que pour X, Y v.a. discrètes, les lois de X et de Y sont définies sur toutes les parties de \mathbb{R} , celle du couple sur toutes les parties de \mathbb{R}^2 . Le résultat suivant montre qu'on retrouve la loi des v.a. marginales à partir de celle d'un couple.

Proposition 7.2.1 *Si (X, Y) est un couple aléatoire de v.a. discrètes de domaine $(X, Y)(\Omega) = \{(x_1, y_1), \dots, (x_i, y_i), \dots\}$, les domaines des marginales X, Y s'obtiennent par projection :*

$$X(\Omega) = p_1((X, Y)(\Omega)) = \{x_1, \dots, x_i, \dots\}, \quad Y(\Omega) = p_2((X, Y)(\Omega)) = \{y_1, \dots, y_i, \dots\}$$

où p_1, p_2 sont les première et seconde projections

$$p_1 : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto x \end{cases}, \quad p_2 : \begin{cases} \mathbb{R}^2 & \rightarrow \mathbb{R} \\ (x, y) & \mapsto y \end{cases}.$$

Les lois marginales $\mathbb{P}_X, \mathbb{P}_Y$ (i.e. les lois de X et de Y , ses marginales) sont données par :

$$\begin{aligned} \forall x_i \in X(\Omega), \quad \mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) &= \sum_{y_j \in Y(\Omega)} \mathbb{P}(X = x_i, Y = y_j), \\ \forall y_j \in Y(\Omega), \quad \mathbb{P}_Y(y_j) = \mathbb{P}(Y = y_j) &= \sum_{x_i \in X(\Omega)} \mathbb{P}(X = x_i, Y = y_j). \end{aligned}$$

Démonstration : Il suffit de faire la preuve pour le domaine et les probabilités ponctuelles de X . Or pour i fixé $\{X = x_i\}$ est la réunion de la famille dénombrable d'événements deux à deux disjoints $\{X = x_i, Y = y_j\}$ pour tous les j tels que $y_j \in Y(\Omega)$ car $\{\omega \in \Omega \mid Y(\omega) = y_j\}_j$ est une partition de Ω . On conclut alors par σ -additivité de \mathbb{P} :

$$\begin{aligned} \mathbb{P}(X = x_i) &= \mathbb{P}\left(\{X = x_i\} \cap \bigcup_j \{Y = y_j\}\right) \\ &= \mathbb{P}\left(\bigcup_j \{X = x_i, Y = y_j\}\right) = \sum_{y_j \in Y(\Omega)} \mathbb{P}(X = x_i, Y = y_j). \end{aligned}$$

Puis $\{x_1, \dots, x_i, \dots\}$ et $\{y_1, \dots, y_j, \dots\}$ sont bien d'une part les projections de $(X, Y)(\Omega)$ sur les premier et second facteurs de $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ et d'autre part les domaines de X et de Y . ■

Remarque 7.2.1 La connaissance du couple (X, Y) permet de connaître les lois marginales de X et de Y . Il est important de comprendre que la réciproque est fautive : les lois marginales ne permettent pas de reconstruire la loi du couple (X, Y) en général. C'est possible dans le cas particulier où X et Y sont indépendantes comme nous le verrons bientôt.

Exemples : On donne le tableau des probabilités ponctuelles $\mathbb{P}(X = x_i, Y = y_j)$ d'un vecteur aléatoire discret (X, Y) :

$X \setminus Y$	$y_1 = -1$	$y_2 = 2$	$y_3 = 3$	$y_4 = 5$	
$x_1 = 0$	0,1	0,05	0,15	0	0,3
$x_2 = 2$	0,05	0,2	0,05	0,1	0,4
$x_3 = 3$	0,1	0	0,1	0,1	0,3
	0,25	0,25	0,3	0,2	1

On en déduit la loi de X : $X(\Omega) = \{0, 2, 3\}$ et

$$\mathbb{P}(X = 0) = 0,3, \quad \mathbb{P}(X = 2) = 0,4, \quad \mathbb{P}(X = 3) = 0,3$$

et celle de Y : $Y(\Omega) = \{-1, 2, 3, 5\}$ et

$$\mathbb{P}(Y = -1) = 0,25, \quad \mathbb{P}(Y = 2) = 0,25, \quad \mathbb{P}(Y = 3) = 0,3, \quad \mathbb{P}(Y = 5) = 0,2.$$

Notons qu'il n'y a pas unicité des couples (X, Y) donnant les mêmes marginales. Ainsi, le couple suivant est différent du précédent mais partage les mêmes marginales.

$X \setminus Y$	$y_1 = -1$	$y_2 = 2$	$y_3 = 3$	$y_4 = 5$	
$x_1 = 0$	0,1	0,1	0	0,1	0,3
$x_2 = 2$	0,1	0,1	0,1	0,1	0,4
$x_3 = 3$	0,05	0,05	0,2	0	0,3
	0,25	0,25	0,3	0,2	1

7.3 Intégrales multiples

Pour exprimer les lois de vecteurs aléatoires réels, on a besoin d'intégrales multiples.

Pour les fonctions de plusieurs variables $F : \mathbb{R}^n \rightarrow \mathbb{R}$, on considère des intégrales multiples $\int_{\mathbb{R}^n} F(x_1, \dots, x_n) dx_1 \dots dx_n$. Pour les calculer, on se ramène à des intégrales simples imbriquées grâce au théorème de Fubini.

Théorème 7.3.1 (Fubini en dimension 2) Soit $F : [a, b] \times [c, d] \rightarrow \mathbb{R}$ telle que l'une ou l'autre des conditions suivantes est vérifiée :

- F est positive : $\forall (x, y) \in [a, b] \times [c, d], F(x, y) \geq 0$, (Fubini-Tonelli)
- $|F|$ est intégrable sur le pavé $[a, b] \times [c, d]$ (Fubini) :

$$\int_{[a,b] \times [c,d]} |F(x, y)| dx dy < +\infty, \quad (7.2)$$

alors

$$\int_{[a,b] \times [c,d]} F(x, y) dx dy = \int_c^d \int_a^b F(x, y) dx dy = \int_a^b \int_c^d F(x, y) dy dx.$$

Si F est positive, on peut intervertir directement les intégrations (par la version Fubini-Tonelli du théorème). Si F ne l'est pas, il faut vérifier (7.2) en calculant l'intégrale double de $|F|$. Pour cela, on peut appliquer par exemple la version Fubini-Tonelli à la fonction positive $|F|$ pour se ramener à des intégrales simples.

Des changements de variables sont souvent utiles pour calculer des intégrales multiples. En particulier le changement de variables en polaire qui consiste à passer de (x, y) représentant des coordonnées cartésiennes dans un repère orthonormée à (r, θ) les coordonnées polaires correspondantes. Ces coordonnées polaires sont données par :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases}, \quad r \in [0, +\infty[, \theta \in [0, 2\pi[.$$

On remplace alors $dx dy$ par $r dr d\theta$ car le jacobien du changement de variables est r . Ainsi :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x, y) dx dy = \int_0^{2\pi} \int_0^{+\infty} F(r \cos \theta, r \sin \theta) r dr d\theta.$$

Exemple : Normalisation de la loi normale $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$.

Notons $I = \int_{-\infty}^{+\infty} e^{-x^2/2} dx$ et montrons que $I^2 = 2\pi$. On a

$$\begin{aligned} I^2 &= \int_{-\infty}^{+\infty} e^{-x^2/2} dx \times \int_{-\infty}^{+\infty} e^{-y^2/2} dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2} e^{-y^2/2} dx dy = \int \int_{\mathbb{R} \times \mathbb{R}} e^{-(x^2+y^2)/2} dx dy \\ &= \int_0^{2\pi} \int_0^{+\infty} e^{-r^2/2} r dr d\theta \\ &= \int_0^{2\pi} d\theta \int_0^{+\infty} r e^{-r^2/2} dr = 2\pi \left[-e^{-r^2/2} \right]_0^{+\infty} = 2\pi \end{aligned}$$

où on a utilisé le théorème de Fubini à la 2ème ligne puis on a fait un changement de variables en polaire à la 3ème ligne.

7.4 Vecteurs aléatoires réels à densité

La notion est la même que celle des densités des v.a.r. adaptée au cas multidimensionnel. La loi d'un vecteur aléatoire de dimension n est de densité f si pour tous intervalles I_1, I_2, \dots, I_n de \mathbb{R} , la probabilité d'appartenance du vecteur aléatoire au produit cartésien $I_1 \times I_2 \times \dots \times I_n$ de ces intervalles s'exprime comme l'intégrale multiple de f sur ce produit d'intervalles.

Définition 7.4.1 Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est appelée densité de probabilité (en dimension n) si

- f est positive : en tout point où elle est définie, $f(t_1, \dots, t_n) \geq 0$,
- f est intégrable sur \mathbb{R}^n d'intégrale 1 :

$$\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(t_1, \dots, t_n) dt_1 \dots dt_n = 1.$$

Définition 7.4.2 Le vecteur aléatoire (X_1, \dots, X_n) suit la loi de densité f si pour tous intervalles $[a_i, b_i]$, $i = 1, \dots, n$

$$\mathbb{P}\left((X_1, \dots, X_n) \in [a_1, b_1] \times \dots \times [a_n, b_n]\right) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_n}^{b_n} f(t_1, t_2, \dots, t_n) dt_1 \dots dt_n.$$

À nouveau, le sens des bornes dans les intervalles (ouvertes ou fermées) n'est pas important.

À nouveau encore, la densité caractérise la loi : si (Y_1, \dots, Y_n) a même loi que (X_1, \dots, X_n) alors ce vecteur a la même densité et réciproquement.

Proposition 7.4.1 Si (X, Y) est un couple aléatoire de loi de densité f , ses lois marginales $\mathbb{P}_X, \mathbb{P}_Y$ sont données par :

$$\begin{aligned} \forall [a, b] \text{ intervalle, } \mathbb{P}_X([a, b]) = \mathbb{P}(X \in [a, b]) &= \int_a^b \int_{-\infty}^{+\infty} f(x, y) dx dy, \\ \forall [a, b] \text{ intervalle, } \mathbb{P}_Y([a, b]) = \mathbb{P}(Y \in [a, b]) &= \int_a^b \int_{-\infty}^{+\infty} f(x, y) dy dx. \end{aligned}$$

Autrement dit, la loi de X est de densité $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$, celle de Y est de densité $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$.

Démonstration : La preuve est une application directe du théorème de Fubini-Tonelli sur les intégrales doubles une fois qu'on a remarqué que

$$\begin{aligned} \mathbb{P}_X([a, b]) &= \mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in [a, b], Y \in \mathbb{R}) = \mathbb{P}_{(X, Y)}([a, b] \times \mathbb{R}) \\ &= \int_{[a, b] \times \mathbb{R}} f(x, y) dx dy = \int_a^b \left(\int_{\mathbb{R}} f(x, y) dy \right) dx = \int_a^b f_X(x) dx \end{aligned}$$

avec la densité annoncée $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$. Il s'applique sans problème car par définition d'une densité, f est positive (et même intégrable sur \mathbb{R}^2). Idem pour Y . ■

Remarque 7.4.1 À nouveau la connaissance de la loi du couple permet d'en déduire celle des lois marginales, la réciproque est en général fautive.

Exemples : • Considérons $f(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y)$. Il s'agit bien d'une densité car f est positive et

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) \, dx dy &= \frac{1}{3} \int \int_{\mathbb{R}^2} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) \, dx dy \\ &= \frac{1}{3} \int \int_{\mathbb{R}^2} \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[-1,2]}(y) \, dx dy \\ &= \frac{1}{3} \underbrace{\int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) \, dx}_{=1} \underbrace{\int_{-\infty}^{+\infty} \mathbf{1}_{[-1,2]}(y) \, dy}_{=2-(-1)=3} \\ &= 1. \end{aligned}$$

Considérons un couple (X, Y) de loi de densité f . La loi de X est alors de densité donnée par :

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) dy = \frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[0,1]}(x) \times \mathbf{1}_{[-1,2]}(y) dy \\ &= \mathbf{1}_{[0,1]}(x) \times \underbrace{\frac{1}{3} \int_{-\infty}^{+\infty} \mathbf{1}_{[-1,2]}(y) dy}_{=1} \\ &= \mathbf{1}_{[0,1]}(x). \end{aligned}$$

De la même façon, $f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y)$.

• Montrer que $f(x, y) = \lambda \mu e^{-\lambda x - \mu y} \mathbf{1}_{\mathbb{R}^+ \times \mathbb{R}^+}(x, y)$ est la densité d'un couple (X, Y) de \mathbb{R}^2 . Montrer que X est de loi $\mathcal{E}(\lambda)$ et Y de loi $\mathcal{E}(\mu)$.

• Montrer que

$$f(x, y) = \frac{e^{-\frac{x^2}{2\sigma^2} - \frac{y^2}{2(\sigma')^2}}}{2\pi}$$

est la densité d'un couple (X, Y) de \mathbb{R}^2 . Montrer que X est de loi $\mathcal{N}(0, \sigma^2)$ et Y de loi $\mathcal{N}(0, (\sigma')^2)$.

• Montrer que

$$f(x, y) = ye^{-xy} \mathbf{1}_{\mathbb{R}^+}(x) \mathbf{1}_{[0,1]}(y)$$

est la densité d'un couple (X, Y) de \mathbb{R}^2 . Montrer que X est de loi donnée par la densité

$$f_X(x) = \frac{1 - e^{-x} - xe^{-x}}{x^2} \mathbf{1}_{\mathbb{R}^+}(x)$$

et Y de loi uniforme sur $[0, 1]$.

- Soit $f(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}$. Il s'agit d'une densité car

$$\begin{aligned} \int \int_{\mathbb{R}^2} f(x, y) dx dy &= \int \int_{\mathbb{R}^2} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dx dy}{3\pi} \\ &= \int \int_{\mathbb{R}^2} e^{-\frac{(x+y)^2+4y^2}{6}} \frac{dx dy}{3\pi} = \int \int_{\mathbb{R}^2} e^{-\frac{(x+y)^2}{2 \times 3}} e^{-\frac{4y^2}{2 \times 3}} \frac{dx dy}{3\pi} \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2 \times 3}} dx \right) e^{-\frac{4y^2}{2 \times 3}} \frac{dy}{3\pi} = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-\frac{z^2}{2 \times 3}} dz \right) e^{-\frac{4y^2}{2 \times 3}} \frac{dy}{3\pi} \\ &= \int_{\mathbb{R}} \sqrt{2\pi \times 3} e^{-\frac{4y^2}{2 \times 3}} \frac{dy}{3\pi} = \int_{\mathbb{R}} e^{-\frac{y^2}{2 \times (3/4)}} \frac{dy}{\sqrt{2\pi \times 3/4}} = 1 \end{aligned}$$

en utilisant la normalisation de la loi normale $\mathcal{N}(0, \sigma^2)$: $\int_{\mathbb{R}} e^{-\frac{t^2}{2\sigma^2}} dt = \sqrt{2\pi\sigma^2}$.

Considérons un couple (X, Y) de densité f , alors X est de densité

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x, y) dy = \int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dy}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(\frac{1}{\sqrt{5}}x + \sqrt{5}y)^2 + 4x^2/5}{6}} \frac{dy}{3\pi} \\ &= \int_{\mathbb{R}} e^{-\frac{(\frac{1}{\sqrt{5}}x + \sqrt{5}y)^2}{6}} e^{-\frac{4x^2}{30}} \frac{dy}{3\pi} = e^{-\frac{4x^2}{30}} \int_{\mathbb{R}} e^{-\frac{z^2}{2 \times 3}} \frac{dz}{3\pi\sqrt{5}} = e^{-\frac{4x^2}{30}} \frac{\sqrt{2\pi \times 3}}{3\pi\sqrt{5}} \\ &= \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}. \end{aligned}$$

La marginale Y est de densité :

$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f(x, y) dx = \int_{\mathbb{R}} e^{-\frac{x^2+2xy+5y^2}{6}} \frac{dx}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(x+y)^2+4y^2}{6}} \frac{dx}{3\pi} = \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2 \times 3}} e^{-\frac{4y^2}{6}} \frac{dx}{3\pi} \\ &= e^{-\frac{4y^2}{6}} \int_{\mathbb{R}} e^{-\frac{(x+y)^2}{2 \times 3}} \frac{dx}{3\pi} = e^{-\frac{4y^2}{6}} \frac{\sqrt{2\pi \times 3}}{3\pi} = \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}. \end{aligned}$$

Les marginales X et Y sont donc de lois $\mathcal{N}(0; 15/4)$ et $\mathcal{N}(0; 3/4)$.

Comme pour la proposition 6.5.2, on a :

Proposition 7.4.2 (Théorème de transfert) *Si (X, Y) est un couple de v.a.r. de densité $f(x, y)$ alors pour F une fonction numérique continue sur \mathbb{R}^2 , on a*

$$E[F(X, Y)] = \int_{\mathbb{R}^2} F(x, y) f(x, y) dx dy. \quad (7.3)$$

7.5 Variables aléatoires indépendantes

L'indépendance pour des v.a. finies a déjà été vue en L1. Il s'agit maintenant de voir le cas de v.a. discrètes pas nécessairement finies et de v.a. à densité.

Dans la suite, on traite simultanément le cas des v.a. discrètes et des v.a. à densité. On énoncera les résultats avec la restriction I intervalle de \mathbb{R} pour les ensembles considérés. On a besoin de cette restriction pour les v.a. à densité. Par contre, pour les v.a. discrètes, ce n'est pas nécessaire et il est possible de prendre des sous-ensembles A quelconques de \mathbb{R} .

Définition 7.5.1 (Indépendance de deux va) Deux v.a. X, Y sont dites indépendantes si pour I, J intervalles de \mathbb{R} , les évènements $\{X \in I\}, \{Y \in J\}$ sont indépendants :

$$\forall I, J \text{ intervalles, } \mathbb{P}(X \in I, Y \in J) = \mathbb{P}(X \in I) \times \mathbb{P}(Y \in J)$$

ce qui s'écrit encore en termes de loi $\mathbb{P}_{X,Y}(I \times J) = \mathbb{P}_X(I) \times \mathbb{P}_Y(J)$: la loi du couple est le « produit » des lois marginales.

Définition 7.5.2 (Indépendance d'une famille finie de va) Les m variables aléatoires X_1, \dots, X_m sont dites (mutuellement) indépendantes si pour tout intervalles I_1, \dots, I_m , les évènements $\{X_1 \in I_1\}, \dots, \{X_m \in I_m\}$ sont mutuellement indépendants : $\forall I_i$ intervalles, $i = 1, \dots, m$,

$$\mathbb{P}(X_1 \in I_1, \dots, X_m \in I_m) = \mathbb{P}(X_1 \in I_1) \dots \mathbb{P}(X_m \in I_m).$$

Remarque 7.5.1 Pour l'indépendance de n évènements, il faut tester toutes les **sous familles** des n évènements et pas seulement la famille entière ou l'indépendance deux à deux. Pour une famille de n va, il suffit de tester **la famille** des n v.a. toutes ensembles. L'apparente différence est due au fait que le test pour n v.a. contient les tests pour toutes les sous familles : il suffit par exemple de prendre $I_k = \mathbb{R}$ pour faire le test sur la famille de $(n-1)$ v.a. où on a exclu la k -ème v.a. Car dire $X_k \in \mathbb{R}$, c'est ne rien dire sur X_k et donc faire comme s'il n'y avait aucune contrainte sur X_k . On comprend bien dès lors que l'on peut tester toutes les sous familles avec des choix adéquats de $I_k = \mathbb{R}$.

Définition 7.5.3 (Indépendance d'une suite de va) Une suite $(X_i)_{i \in \mathbb{N}}$ de v.a. est dite indépendante si toute sous-suite finie de $(X_i)_{i \in \mathbb{N}}$ est indépendante au sens de la définition 7.5.2.

Proposition 7.5.1

– Les v.a. discrètes X et Y sont indépendantes si et seulement si

$$\forall x_i \in X(\Omega), \forall y_j \in Y(\Omega), \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j). \quad (7.4)$$

– Les v.a.r. X, Y de densités respectives f et g sont indépendantes si et seulement si le couple (X, Y) est de densité le produit tensoriel $f \otimes g : \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto f(x)g(y)$.

Démonstration : • Il est clair que l'indépendance implique (7.4) : il suffit de prendre $I = \{x_i\}$ et $J = \{y_j\}$ dans la définition 7.5.1.

Réciproquement, si (7.4) est vérifié, alors pour deux parties A et B quelconques de \mathbb{R} , on a

$$\begin{aligned}
 \mathbb{P}(X \in A, Y \in B) &= \mathbb{P}((X, Y) \in A \times B) \\
 &= \sum_{(x_i, y_j) \in A \times B} \mathbb{P}((X, Y) = (x_i, y_j)) \\
 &= \sum_{(x_i, y_j) \in A \times B} \mathbb{P}(X = x_i, Y = y_j) \\
 &= \sum_{(x_i, y_j) \in A \times B} \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j) \\
 &= \sum_{x_i \in A} \mathbb{P}(X = x_i) \sum_{y_j \in B} \mathbb{P}(Y = y_j) \\
 &= \mathbb{P}(X \in A) \mathbb{P}(Y \in B).
 \end{aligned}$$

• Dans le cas de v.a.r. à densité, si X et Y sont indépendantes, la définition des densités et la définition 7.5.1 donnent :

$$\begin{aligned}
 \mathbb{P}_{(X, Y)}([a, b] \times [c, d]) = \mathbb{P}(X \in [a, b], Y \in [c, d]) &= \mathbb{P}(X \in [a, b]) \mathbb{P}(Y \in [c, d]) \\
 &= \int_a^b f(x) dx \int_c^d g(y) dy \\
 &= \int_a^b \int_c^d f(x)g(y) dx dy
 \end{aligned}$$

par le théorème de Fubini, ce qui montre que $f(x)g(y)$ est densité du couple (X, Y) . Réciproquement, si le couple a pour densité $(f \otimes g)(x, y) = f(x)g(y)$ alors pour tous intervalles $[a, b]$, $[c, d]$:

$$\begin{aligned}
 \mathbb{P}_{(X, Y)}([a, b] \times [c, d]) &= \int_a^b \int_c^d f(x)g(y) dx dy = \int_a^b f(x) dx \int_c^d g(y) dy \\
 &= \mathbb{P}(X \in [a, b]) \mathbb{P}(Y \in [c, d]),
 \end{aligned}$$

ce qui justifie l'indépendance de X et de Y . ■

Remarque 7.5.2 Une conséquence importante : si on connaît les lois de X et de Y , des variables supposées **indépendantes**, on peut reconstruire la loi du couple (X, Y) à partir des marginales par (7.4) dans le cas discret ou par le produit tensoriel $f \otimes g$ des densités dans le cas à densité. **Insistons sur le fait que ce n'est pas vrai en général quand X et Y ne sont pas indépendantes.**

Dans les deux exemples de la page 73, X et Y ne sont pas indépendantes car par exemple pour le premier :

$$\mathbb{P}(X = 2, Y = 2) = 0,2, \quad \text{tandis que} \quad \mathbb{P}(X = 2) \times \mathbb{P}(Y = 2) = 0,4 \times 0,25 = 0,1.$$

Et pour le second :

$$\mathbb{P}(X = 3, Y = 5) = 0, \quad \text{tandis que} \quad \mathbb{P}(X = 3) \times \mathbb{P}(Y = 5) = 0,3 \times 0,2 = 0,06.$$

Exemples :

• On donne le tableau de la loi d'un couple (X, Y) en donnant les probabilités ponctuelles $\mathbb{P}(X = x_i, Y = y_j)$:

$X \setminus Y$	y_1	y_2	y_3	
x_1	0,12	0,08	0,20	0,4
x_2	0,18	0,12	0,30	0,6
	0,3	0,2	0,5	= 1

On vérifie ici que X et Y sont indépendantes car pour tout $i = 1, 2$ et $j = 1, 2, 3$, on a

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j).$$

• Considérons le couple (X, Y) de loi donnée par la densité $f_{(X,Y)}(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y)$. On a vu que X et Y avaient pour densité $f_X(x) = \mathbf{1}_{[0,1]}(x)$ et $f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y)$. On a alors

$$f_{(X,Y)}(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y) = \mathbf{1}_{[0,1]}(x) \times \frac{1}{3} \mathbf{1}_{[-1,2]}(y) = f_X(x) f_Y(y).$$

Les variables X et Y sont donc indépendantes.

• Soit (X, Y) le couple aléatoire de loi donnée par la densité $f_{(X,Y)}(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}$. On a vu que les densités marginales sont

$$f_X(x) = \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}, \quad f_Y(y) = \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}.$$

On a alors

$$f_X(x) f_Y(y) = \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}} \times \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}} \neq \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}} = f_{(X,Y)}(x, y).$$

Dans ce cas, X et Y ne sont pas indépendantes.

Proposition 7.5.2 Soient X, Y des v.a. indépendantes, F, G des fonctions dont les domaines de définition contiennent respectivement $X(\Omega)$ et $Y(\Omega)$. Alors les v.a. $F(X)$ et $G(Y)$ sont indépendantes.

Démonstration : Plaçons nous dans le cas de v.a. discrètes. Rappelons que $F(X)$ désigne l'application $F \circ X$ définie par

$$F \circ X : \Omega \longrightarrow \mathbb{R}, \quad \omega \mapsto F(X(\omega)).$$

Il s'agit bien d'une v.a. discrète car l'ensemble de ses valeurs est $\{F(x_1), \dots, F(x_k), \dots\}$ si celui de X est $\{x_1, \dots, x_k, \dots\}$. De même celui de $G(Y)$ est $\{G(y_1), \dots, G(y_k), \dots\}$. Pour prouver l'indépendance de $F(X)$ et $G(Y)$, il suffit d'après la proposition 7.5.1, de voir pour $t \in F(X)(\Omega)$ et $s \in G(Y)(\Omega)$:

$$\mathbb{P}(F(X) = t, G(Y) = s) = \mathbb{P}(F(X) = t) \mathbb{P}(G(Y) = s).$$

Or

$$\begin{aligned} \mathbb{P}(F(X) = t, G(Y) = s) &= \sum_{\substack{i:F(x_i)=t \\ j:G(y_j)=s}} \mathbb{P}(X = x_i, Y = y_j) \\ &= \sum_{\substack{i:F(x_i)=t \\ j:G(y_j)=s}} \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j) \\ &= \sum_{i:F(x_i)=t} \mathbb{P}(X = x_i) \sum_{j:G(y_j)=s} \mathbb{P}(Y = y_j) \\ &= \mathbb{P}(F(X) = t) \mathbb{P}(G(Y) = s). \end{aligned}$$

■

Proposition 7.5.3 Soient X et Y des v.a. indépendantes et F, G des fonctions numériques $\mathbb{R}^2 \rightarrow \mathbb{R}$. Alors quand toutes les quantités sont bien définies, on a

$$E[F(X)G(Y)] = E[F(X)] E[G(Y)].$$

Démonstration : Par exemple si X et Y sont des v.a.r. de densités f et g , d'après la proposition 7.4.2, (X, Y) est de densité $f(x)g(y)$ et

$$\begin{aligned} E[F(X)G(Y)] &= \int \int_{\mathbb{R}^2} F(x)G(y)f(x)g(y) \, dx dy \\ &= \int \int_{\mathbb{R}^2} F(x)f(x) G(y)g(y) \, dx dy \\ &= \int_{-\infty}^{+\infty} F(x)f(x) \, dx \times \int_{-\infty}^{+\infty} G(y)g(y) \, dy \\ &= E[F(X)] E[G(Y)]. \end{aligned}$$

■

Remarque 7.5.3 En particulier pour X et Y des v.a.r. indépendantes, quand les espérances sont bien définies :

$$E[XY] = E[X] E[Y]. \quad (7.5)$$

7.6 Lois conditionnelles

7.6.1 Cas discret

Définition 7.6.1 Soit (X, Y) un couple de variables aléatoires discrètes sur $(\Omega, \mathcal{F}, \mathbb{P})$ avec X, Y de domaines respectifs $X(\Omega), Y(\Omega)$. Pour $y \in Y(\Omega)$ tel que $\mathbb{P}(Y = y) \neq 0$, on appelle loi conditionnelle de X sachant $Y = y$, l'application définie sur $X(\Omega)$ par

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

De même si $x \in X(\Omega)$ est tel que $\mathbb{P}(X = x) \neq 0$, on appelle loi conditionnelle de Y sachant $X = x$ l'application définie sur $Y(\Omega)$ par

$$\mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

Notons que la loi conditionnelle de X sachant $Y = y$ notée $\mathbb{P}_X(\cdot|Y = y)$ et qui à un évènement A associe

$$\mathbb{P}_X(A|Y = y) = \mathbb{P}(X \in A|Y = y) = \frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)}$$

est en fait une probabilité sur \mathbb{R} . Elle vérifie donc toutes les propriétés d'une probabilité.

Proposition 7.6.1 Soit (X, Y) un couple de variables aléatoires discrètes. Alors pour tous x, y , on a :

$$\mathbb{P}(X = x, Y = y) = \begin{cases} \mathbb{P}(X = x|Y = y) \mathbb{P}(Y = y) & \text{si } \mathbb{P}(Y = y) \neq 0. \\ 0 & \text{si } \mathbb{P}(Y = y) = 0. \end{cases}$$

Démonstration : Si $\mathbb{P}(Y = y) \neq 0$, par définition :

$$\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

et la conclusion s'impose.

Si $\mathbb{P}(Y = y) = 0$ alors $\mathbb{P}(X = x, Y = y) = \mathbb{P}(\{X = x\} \cap \{Y = y\}) \leq \mathbb{P}(Y = y) = 0$. ■

Proposition 7.6.2 Si X, Y sont des v.a. indépendantes alors la loi conditionnelle de X sachant $Y = y$ est la même que celle de X :

$$\mathbb{P}_X(\cdot|Y = y) = \mathbb{P}_X.$$

Le conditionnement par une v.a. indépendante est sans effet.

Démonstration : En effet, pour tout A , par indépendance de X et de Y

$$\begin{aligned} \mathbb{P}_X(A|Y = y) &= \mathbb{P}(X \in A|Y = y) = \frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(X \in A)\mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} \\ &= \mathbb{P}(X \in A) = \mathbb{P}_X(A). \end{aligned}$$

■

Exemple : Une variable aléatoire Y suit la loi de Poisson $\mathcal{P}(\lambda)$. La loi conditionnelle de X sachant $Y = m$ est la loi binomiale de paramètres m et p . Déterminer la loi de X .

On détermine d'abord la loi jointe du couple (X, Y) : pour des entiers n, m , on a

$$\mathbb{P}(X = n, Y = m) = \begin{cases} \mathbb{P}(X = n|Y = m) \mathbb{P}(Y = m) & \text{si } \mathbb{P}(Y = m) \neq 0, \\ 0 & \text{si } \mathbb{P}(Y = m) = 0. \end{cases}$$

Comme $\mathbb{P}(Y = m) = \frac{e^{-\lambda} \lambda^m}{m!} > 0$ mais $\mathbb{P}(X = n|Y = m) = 0$ si $n > m$, on a :

$$\mathbb{P}(X = n, Y = m) = \begin{cases} C_m^n p^n (1-p)^{m-n} \frac{e^{-\lambda} \lambda^m}{m!} = p^n (1-p)^{m-n} \frac{e^{-\lambda} \lambda^m}{n!(m-n)!} & \text{si } n \leq m, \\ 0 & \text{si } n > m. \end{cases}$$

On en déduit la loi de X : pour $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}(X = n) &= \sum_{m=0}^{+\infty} \mathbb{P}(X = n, Y = m) = \sum_{m=n}^{+\infty} \mathbb{P}(X = n, Y = m) \\ &= \sum_{m=n}^{+\infty} p^n (1-p)^{m-n} \frac{e^{-\lambda} \lambda^m}{n!(m-n)!} \\ &= \frac{p^n e^{-\lambda}}{n!} \sum_{m=n}^{+\infty} (1-p)^{m-n} \frac{\lambda^m}{(m-n)!} = \frac{p^n e^{-\lambda}}{n!} \sum_{k=0}^{+\infty} (1-p)^k \frac{\lambda^{k+n}}{k!} \quad \text{avec } k = m - n \\ &= \frac{(p\lambda)^n e^{-\lambda}}{n!} \sum_{k=0}^{+\infty} \frac{((1-p)\lambda)^k}{k!} = \frac{(p\lambda)^n e^{-\lambda}}{n!} e^{\lambda(1-p)} = \frac{(p\lambda)^n e^{-p\lambda}}{n!}. \end{aligned}$$

La variable X suit donc la loi de Poisson $\mathcal{P}(p\lambda)$.

7.6.2 Cas continu : densité conditionnelle

Définition 7.6.2 Soit (X, Y) un couple de variables aléatoires réelles de densité $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. On définit la densité conditionnelle de X sachant $Y = y$ par

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}$$

où $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$ est la densité (marginale) de Y .

La loi conditionnelle de X sachant $Y = y$ est alors définie par cette densité $f_{X|Y=y}$:

$$\forall I \text{ intervalle de } \mathbb{R}, \quad \mathbb{P}(X \in I | Y = y) = \int_I f_{X|Y=y}(x) dy = \int_I \frac{f(x, y)}{f_Y(y)} dy.$$

La densité conditionnelle $f_{X|Y=y}$ est une fonction de la seule variable x . Par contre, y est seulement un paramètre de la fonction.

Exemple : Reprenons le couple (X, Y) de loi donnée par la densité

$$f(x, y) = \frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}.$$

On a vu que X et Y sont de loi $\mathcal{N}(0; 15/4)$ et $\mathcal{N}(0; 3/4)$ avec les densités

$$f_X(x) = \frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}, \quad f_Y(y) = \frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}.$$

La loi de X conditionnellement à $\{Y = y\}$ est alors de densité

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)} = \frac{\frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}}{\frac{1}{\sqrt{3\pi/2}} e^{-\frac{4y^2}{6}}} = \frac{1}{\sqrt{6\pi}} e^{-\frac{x^2+2xy+y^2}{6}}.$$

Celle de Y conditionnellement à $\{X = x\}$ est de densité

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)} = \frac{\frac{1}{3\pi} e^{-\frac{x^2+2xy+5y^2}{6}}}{\frac{1}{\sqrt{15\pi/2}} e^{-\frac{4x^2}{30}}} = \frac{1}{\sqrt{6\pi/5}} e^{-\frac{x^2+10xy+25y^2}{30}}.$$

Comme pour la Proposition 7.6.2 dans le cas discret, on a dans le cas avec des densités :

Proposition 7.6.3 *Si les variables aléatoires X et Y sont indépendantes de densité f_X et f_Y alors les densités conditionnelles sont les densités marginales :*

$$f_{X|Y=y}(x) = f_X(x) \quad \forall y, \quad \text{et} \quad f_{Y|X=x}(y) = f_Y(y) \quad \forall x.$$

À nouveau le conditionnement est sans effet car les variables sont indépendantes.

Démonstration : Comme X et Y sont indépendantes, le couple (X, Y) est de densité $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$. On a alors :

$$\begin{aligned} f_{X|Y=y}(x) &= \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x), \\ f_{Y|X=x}(y) &= \frac{f_{(X,Y)}(x, y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y). \end{aligned}$$

Exemple : Si on considère (X, Y) de loi donnée par la densité

$$f(x, y) = \frac{1}{3} \mathbf{1}_{[0,1] \times [-1,2]}(x, y)$$

alors X et Y sont de densité $f_X(x) = \mathbf{1}_{[0,1]}(x)$ et $f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y)$. On a bien $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$ et donc les lois conditionnelles sont

$$f_{X|Y=y}(x) = f_X(x) = \mathbf{1}_{[0,1]}(x), \quad f_{Y|X=x}(y) = f_Y(y) = \frac{1}{3} \mathbf{1}_{[-1,2]}(y).$$

Chapitre 8

Somme de v.a. indépendantes

Les sommes de variables aléatoires interviennent souvent en probabilité. Elles permettent de modéliser les effets conjugués de plusieurs phénomènes. Lorsqu'en plus les v.a. sont indépendantes et de même loi, on modélise l'effet cumulé d'un phénomène récurrent dont les réalisations sont indépendantes. Plusieurs résultats très importants, des théorèmes limites, leur sont consacrés. En fait, on s'intéresse souvent à la moyenne arithmétique de variables aléatoires indépendantes et de même loi $(X_i)_{i \in \mathbb{N}^*}$:

$$\frac{1}{n} \sum_{i=1}^n X_i. \quad (8.1)$$

Le plus important résultat est la *loi des grands nombres* (LGN) qui énonce la convergence de la moyenne arithmétique (8.1) vers l'espérance de la loi. Puis le théorème central limite (TCL) qui précise (en un certain sens la LGN).

8.1 Somme de deux v.a. indépendantes

Variables aléatoires discrètes

Commençons d'abord par étudier le cas de la somme de deux variables aléatoires discrètes.

Proposition 8.1.1 *Soient X, Y deux v.a. discrètes indépendantes à valeurs entières (i.e. avec $X(\Omega) \subset \mathbb{N}, Y(\Omega) \subset \mathbb{N}$). La loi de $X + Y$ est donnée par :*

$$\begin{aligned} \forall n \in \mathbb{N}, \quad \mathbb{P}(X + Y = n) &= \sum_{i+j=n} \mathbb{P}(X = i) \mathbb{P}(Y = j) \\ &= \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i). \end{aligned}$$

Démonstration : Comme X et Y sont à valeurs entières, il en est de même pour $X + Y$. Sa loi est caractérisée par les probabilités $\mathbb{P}(X + Y = n)$. Pour les calculer, il suffit de décomposer l'évènement $\{X + Y = n\}$ en la réunion de tous les évènements $\{X = i, Y = j\}$ tels que $i + j = n$. Il suit alors

$$\begin{aligned} \mathbb{P}(X + Y = n) &= \mathbb{P}\left(\bigcup_{(i,j);i+j=n} \{X = i, Y = j\}\right) \\ &= \sum_{i+j=n} \mathbb{P}(X = i, Y = j) \end{aligned} \quad (8.2)$$

$$= \sum_{i+j=n} \mathbb{P}(X = i) \mathbb{P}(Y = j), \quad (8.3)$$

ce qui prouve la proposition. ■

Remarque 8.1.1 Si X et Y ne sont pas indépendantes le passage de (8.2) à (8.3) n'est plus valable, on peut seulement écrire (8.2). On voit ainsi que l'on peut toujours calculer la loi de $X + Y$ si on connaît celle du couple (X, Y) par (8.2). Par contre, le calcul de cette loi à partir de celles de X et de Y n'est pas possible en général, il faut en plus l'indépendance pour avoir (8.3).

Exemple : Si X, Y sont des v.a. indépendantes de lois binomiales $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$, alors $X + Y$ suit une loi binomiale $\mathcal{B}(n + m, p)$.

En effet, on sait que X , de loi $\mathcal{B}(n, p)$, peut être vue comme une somme de n v.a. indépendantes ε_i , $1 \leq i \leq n$, de loi de Bernoulli $b(p)$

$$\varepsilon_i = 0 \text{ ou } 1, \quad \text{avec } \mathbb{P}(\varepsilon_i = 1) = p, \mathbb{P}(\varepsilon_i = 0) = 1 - p.$$

De même Y est somme de m v.a. indépendantes $\tilde{\varepsilon}_j$, $1 \leq j \leq m$, de loi $b(p)$. Comme $(\varepsilon_i)_{i=1,\dots,n}$ et $(\tilde{\varepsilon}_j)_{j=1,\dots,m}$ sont indépendantes,

$$X + Y = \varepsilon_1 + \dots + \varepsilon_n + \tilde{\varepsilon}_1 + \dots + \tilde{\varepsilon}_m$$

est une somme de $n + m$ v.a. de Bernoulli $b(p)$ indépendantes. $X + Y$ suit donc la loi $\mathcal{B}(n + m, p)$.

On peut aussi le faire directement : pour $i = 0, \dots, n + m$, on a

$$\begin{aligned} \mathbb{P}(X + Y = i) &= \sum_{j=0}^i \mathbb{P}(X = j, Y = i - j) = \sum_{j=0}^i \mathbb{P}(X = j) \mathbb{P}(Y = i - j) \\ &= \sum_{j=0}^i C_n^j p^j (1 - p)^{n-j} C_m^{i-j} p^{i-j} (1 - p)^{m-i+j} \\ &= p^i (1 - p)^{n+m-i} \sum_{j=0}^i C_n^j C_m^{i-j} \end{aligned}$$

$$= C_{n+m}^i p^i (1-p)^{n+m-i}$$

en utilisant l'identité

$$\sum_{j=0}^i C_n^j C_m^{i-j} = C_{n+m}^i$$

qu'on prouve en développant de deux façons $(1+x)^{n+m} = (1+x)^n(1+x)^m$.

Exemple : Soient X, Y des v.a. indépendantes suivant des lois de Poisson de paramètres respectifs α et β . Quelle est la loi de $S = X + Y$?

Les lois de X et Y sont données par

$$\mathbb{P}(X = i) = \frac{e^{-\alpha} \alpha^i}{i!}, \quad \mathbb{P}(Y = j) = \frac{e^{-\beta} \beta^j}{j!}, \quad i, j \in \mathbb{N}.$$

Comme X et Y sont indépendantes, on a en utilisant la formule du binôme de Newton :

$$\begin{aligned} \mathbb{P}(S = n) &= \sum_{i=0}^n \mathbb{P}(X = i) \mathbb{P}(Y = n - i) = \sum_{i=0}^n \frac{e^{-\alpha} \alpha^i}{i!} \frac{e^{-\beta} \beta^{n-i}}{(n-i)!} \\ &= \frac{e^{-(\alpha+\beta)}}{n!} \sum_{i=0}^n C_n^i \alpha^i \beta^{n-i} \\ &= \frac{e^{-(\alpha+\beta)} (\alpha + \beta)^n}{n!}. \end{aligned}$$

Ainsi $S = X + Y$ suit la loi de Poisson de paramètre $\alpha + \beta$.

Variables aléatoires à densité

Avant de voir la loi de la somme de deux v.a. réelles indépendantes à densité, introduisons la notion de convolution :

Définition 8.1.1 La convolution de deux fonctions f et g réelles est la fonction $f * g$ sur \mathbb{R} donnée par

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(x-y)g(y) dy.$$

On parle encore de la convolée $f * g$ de f et de g .

Remarque 8.1.2 On a $f * g = g * f$ (c'est immédiat en faisant le changement de variable y donne $x - y$ dans l'intégrale de la définition)

Dans la proposition 8.1.1, on a défini un analogue discret de la convolution de la définition 8.1.1, comme en témoigne le résultat suivant qui exprime la loi de $X + Y$ par la convolution des densités.

Proposition 8.1.2 Soient X, Y deux v.a.r. indépendantes et de densités f et g . La loi de $X + Y$ est donnée par :

$$\forall [a, b] \text{ intervalle, } \mathbb{P}(X + Y \in [a, b]) = \int_a^b (f * g)(x) dx$$

Autrement dit, $X + Y$ a pour densité la fonction $f * g$.

Démonstration : Soient $a < b$, comme (X, Y) est de densité $(x, y) \mapsto f(x)g(y)$, on a

$$\mathbb{P}(X + Y \in [a, b]) = \int_{(x,y); x+y \in [a,b]} f(x)g(y) dx dy.$$

On fait le changement de variable $(x, y) \longrightarrow (t, s) = (x, x + y)$. Comme (x, y) varie dans \mathbb{R}^2 de façon que $x + y \in [a, b]$, t décrit tout \mathbb{R} et s décrit $[a, b]$. On a alors :

$$\mathbb{P}(X + Y \in [a, b]) = \int_a^b \int_{-\infty}^{+\infty} f(t)g(s-t) \times \underbrace{1}_{\text{Jac}} \times dt ds = \int_a^b (f * g)(s) ds,$$

car le jacobien du changement de variable est

$$\text{Jac} = \begin{vmatrix} \frac{\partial t}{\partial x} & \frac{\partial s}{\partial x} \\ \frac{\partial t}{\partial y} & \frac{\partial s}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1,$$

ce qui prouve la proposition. ■

Remarque 8.1.3 A nouveau on connaît bien la loi de la somme $X + Y$ si X et Y sont indépendantes, sinon, il faut connaître la loi du couple (X, Y) et sa densité $h(x, y)$ si elle existe pour avoir la loi de $X + Y$ par

$$\mathbb{P}(X + Y \in [a, b]) = \int_{(x,y); x+y \in [a,b]} h(x, y) dx dy = \int_a^b \int_{-\infty}^{+\infty} h(x, y-x) dx dy.$$

Exemples : • Soient X, Y des v.a. indépendantes suivant des lois exponentielles de paramètres respectifs a et b . Quelle est la loi de $S = X + Y$?

Les lois de X et Y sont de densités

$$f(x) = ae^{-ax} \mathbf{1}_{[0, +\infty[}(x), \quad g(y) = be^{-by} \mathbf{1}_{[0, +\infty[}(y).$$

Comme X et Y sont indépendantes, la densité de $X + Y$ est, si $a \neq b$:

$$(f * g)(x) = \int_{-\infty}^{+\infty} g(y)f(x-y) dy$$

$$\begin{aligned}
&= \int_0^{+\infty} b e^{-by} a e^{-a(x-y)} \mathbf{1}_{[0,+\infty[}(x-y) dy \\
&= a b e^{-ax} \mathbf{1}_{\{x \geq 0\}} \int_0^x e^{(a-b)y} dy \\
&= \frac{a b e^{-ax}}{a-b} (e^{(a-b)x} - 1) \mathbf{1}_{\{x \geq 0\}} \\
&= \frac{a b}{a-b} (e^{-bx} - e^{-ax}) \mathbf{1}_{\{x \geq 0\}}
\end{aligned}$$

où à la 3ème ligne on a utilisé $\mathbf{1}_{[0,+\infty[}(x-y) \mathbf{1}_{[0,+\infty[}(y) = \mathbf{1}_{[0,+\infty[}(x) \mathbf{1}_{[0,x]}(y)$. Si $a = b$, la densité est

$$\begin{aligned}
f * g(x) &= \int_{-\infty}^{+\infty} g(y) f(x-y) dy = a^2 \int_{-\infty}^{+\infty} e^{-ay} \mathbf{1}_{\{y \geq 0\}} e^{-a(x-y)} \mathbf{1}_{\{x-y \geq 0\}} dy \\
&= a^2 \mathbf{1}_{\{x \geq 0\}} \int_0^x e^{-ay} e^{-a(x-y)} dy = a^2 \mathbf{1}_{\{x \geq 0\}} e^{-ax} \int_0^x dy = a^2 x \mathbf{1}_{\{x \geq 0\}} e^{-ax}.
\end{aligned}$$

• Soient X_1 de loi $\mathcal{N}(m_1, \sigma_1^2)$ et X_2 de loi $\mathcal{N}(m_2, \sigma_2^2)$ alors $X_1 + X_2$ est de loi normale $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Pour simplifier (?!?) les calculs qui suivent, prenons $m_1 = m_2 = 0$, et notons f_1 et f_2 les densités de X_1 et de X_2 . Celle de $X_1 + X_2$ est donnée par

$$\begin{aligned}
f_1 * f_2(x) &= \int_{-\infty}^{+\infty} f_1(t) f_2(x-t) dt = \int_{-\infty}^{+\infty} e^{-t^2/(2\sigma_1^2)} e^{-(x-t)^2/(2\sigma_2^2)} \frac{dt}{\sqrt{2\pi\sigma_1^2} \sqrt{2\pi\sigma_2^2}} \\
&= \int_{-\infty}^{+\infty} \exp\left(-\frac{(\sigma_1^2 + \sigma_2^2)t^2 - 2\sigma_1^2 x t + \sigma_1^2 x^2}{2\sigma_1^2 \sigma_2^2}\right) \frac{dt}{2\pi\sigma_1\sigma_2} \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)^{1/2} t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} x)^2 - \frac{\sigma_1^4}{(\sigma_1^2 + \sigma_2^2)} x^2 + \sigma_1^2 x^2}{2\sigma_1^2 \sigma_2^2}\right) dt \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)^{1/2} t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} x)^2 + \frac{\sigma_1^2 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)} x^2}{2\sigma_1^2 \sigma_2^2}\right) dt \\
&= \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{((\sigma_1^2 + \sigma_2^2)^{1/2} t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} x)^2}{2\sigma_1^2 \sigma_2^2}\right) dt \\
&= \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2\sigma_1^2 \sigma_2^2}\right) \frac{du}{(\sigma_1^2 + \sigma_2^2)^{1/2}}
\end{aligned}$$

avec le changement de variable $u = (\sigma_1^2 + \sigma_2^2)^{1/2} t - \frac{\sigma_1^2}{(\sigma_1^2 + \sigma_2^2)^{1/2}} x$.

Puis d'après la normalisation de la loi normale $\mathcal{N}(0, \sigma_1^2 \sigma_2^2)$, la dernière intégrale vaut

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{u^2}{2\sigma_1^2 \sigma_2^2}\right) \frac{du}{(\sigma_1^2 + \sigma_2^2)^{1/2}} = \frac{\sqrt{2\pi\sigma_1^2 \sigma_2^2}}{(\sigma_1^2 + \sigma_2^2)^{1/2}}.$$

On a finalement :

$$f_1 * f_2(x) = \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \frac{\sqrt{2\pi\sigma_1^2 \sigma_2^2}}{(\sigma_1^2 + \sigma_2^2)^{1/2}} = \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}}.$$

On a obtenu la densité de la loi $\mathcal{N}(0, \sigma_1^2 + \sigma_2^2)$.

Proposition 8.1.3 (Variance d'une somme de v.a. indépendantes) *Si X, Y sont des v.a. indépendantes alors*

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

Remarque 8.1.4 Notez que la même propriété est vraie pour l'espérance (qui est linéaire) **sans** hypothèse d'indépendance alors qu'en général, c'est faux pour la variance si X et Y ne sont pas indépendantes. Par exemple

$$\text{Var}(X + X) = \text{Var}(2X) = 2^2 \text{Var}(X) = 4 \text{Var}(X) \neq 2 \text{Var}(X) = \text{Var}(X) + \text{Var}(X).$$

Démonstration :

$$\begin{aligned} \text{Var}(X + Y) &= E[(X + Y)^2] - (E[X + Y])^2 \\ &= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\ &= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2 \\ &= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2E[XY] - 2E[X]E[Y] \\ &= \text{Var}(X) + \text{Var}(Y) \end{aligned}$$

car par indépendance de X et Y , on a la propriété (7.5) : $E[XY] = E[X]E[Y]$. ■

Exemple : Soient $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ des variables normales indépendantes. Retrouvons les paramètres de la loi de $Y = X_1 + X_2$.

On a vu que Y suit une loi normale, pour connaître les paramètres, il s'agit de connaître $E[X_1 + X_2] = E[X_1] + E[X_2] = m_1 + m_2$ et $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) = \sigma_1^2 + \sigma_2^2$. D'où

$$Y = X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

Retenons des exemples précédents que :

Proposition 8.1.4 *Soient $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ des v.a. normales indépendantes. Alors*

$$X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

8.2 Convergences probabilistes

Les variables aléatoires X_n sont des applications de Ω vers \mathbb{R} , et pour des applications, le mode de convergence le plus naturel est celui de la convergence pour chaque $\omega \in \Omega$ de la suite de réels $X_n(\omega)$ vers le réel $X(\omega)$.

$$\forall \omega \in \Omega, \quad X_n(\omega) \rightarrow X(\omega), \quad n \rightarrow +\infty.$$

Il s'agit de la convergence simple d'une suite d'applications vue en analyse. Malheureusement, en probabilité, ce type de convergence est trop restrictif : on ne peut raisonnablement demander à *tous* les $X_n(\omega)$ de converger (i.e. pour *tous* les $\omega \in \Omega$).

Par contre, il est plus raisonnable de demander que l'ensemble des ω pour lesquels ça n'arrive pas soit de probabilité nulle (ou au moins petite). Ceci nous amène aux notions de convergences *presque sûre* et en *probabilité*.

Définition 8.2.1 Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires et X une v.a. définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que X_n converge presque sûrement (p.s.) vers X si l'ensemble des ω tels que $X_n(\omega)$ converge vers $X(\omega)$ a pour probabilité 1, c'est à dire :

$$\mathbb{P}(\omega \in \Omega \mid X_n(\omega) \rightarrow X(\omega)) = 1.$$

On la note $X_n \xrightarrow{p.s.} X$.

Rappelons qu'un évènement de probabilité 1 n'est pas nécessairement égale à tout l'espace Ω . Il peut même y avoir une infinité d'éléments dans son complémentaire. Seulement, ce complémentaire est (du point de vue de la probabilité \mathbb{P}) négligeable.

Dans la convergence presque sûre, si on se fixe $\varepsilon > 0$, le rang n_0 à partir duquel $X_n(\omega)$ est à moins de ε de $X(\omega)$ dépend à la fois de ε et de ω : $n_0 = n_0(\varepsilon, \omega)$. Généralement, on ne sait pas comment $n_0(\varepsilon, \omega)$ dépend de ω . De ce fait la convergence presque sûre est essentiellement une convergence théorique.

Par exemple, si on suppose que X_n est une v.a. dont la réalisation dépend de n épreuves répétées, savoir que X_n converge presque sûrement vers X ne permet pas de prédire un nombre (non aléatoire, c'est à dire qui ne dépend pas de ω) n d'épreuves à partir duquel $|X_n(\omega) - X(\omega)| \leq \varepsilon$ si ce n'est pour presque tous les $\omega \in \Omega$, même pour 99% ou 95% d'entre eux. Or cette question a une grande importance pratique pour le statisticien. C'est l'une des raisons de l'introduction de la convergence en probabilité qui permet de répondre à cette question lorsque l'on connaît la vitesse de convergence selon ce mode.

Définition 8.2.2 (Convergence en probabilité) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires et X une v.a. définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que X_n converge en probabilité vers X si :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

On la note $X_n \xrightarrow{\mathbb{P}} X$.

Remarque 8.2.1 Il faut bien comprendre que quand X_n converge en probabilité vers X , il est toujours possible que pour certain $\omega \in \Omega$, $X_n(\omega)$ s'écarte de $X(\omega)$ même quand n est grand. Mais, c'est de moins en moins probable, c'est à dire que cela arrive pour peu de $\omega \in \Omega$: la probabilité que X_n soit distant de plus de $\varepsilon > 0$ de X est de plus en plus faible.

Proposition 8.2.1 *La convergence presque sûre entraîne la convergence en probabilité.*

Démonstration : Soit X_n convergeant presque sûrement vers X . L'évènement

$$\Omega' = \left\{ \omega \in \Omega; \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \right\}$$

est de probabilité 1. Fixons $\varepsilon > 0$, et définissons

$$\Omega'_\varepsilon = \left\{ \omega \in \Omega, \exists m_0 = m_0(\omega), \forall n \geq m_0, |X_n(\omega) - X(\omega)| < \varepsilon \right\}.$$

Il est clair que $\Omega' \subset \Omega'_\varepsilon$ et donc $\mathbb{P}(\Omega'_\varepsilon) = 1$. Par traduction des opérateurs logiques \forall et \exists en opérateur ensemblistes \cap , \cup , on exprime facilement :

$$\Omega'_\varepsilon = \bigcup_{m_0 \in \mathbb{N}} \bigcap_{n \geq m_0} \left\{ \omega \in \Omega; |X_n(\omega) - X(\omega)| < \varepsilon \right\}.$$

Posons

$$A_k = \left\{ \omega \in \Omega; \forall n \geq k \quad |X_n(\omega) - X(\omega)| < \varepsilon \right\} = \bigcap_{n \geq k} \left\{ \omega \in \Omega; |X_n(\omega) - X(\omega)| < \varepsilon \right\}.$$

Il est clair que la suite d'ensembles $(A_k)_{k \in \mathbb{N}}$ est croissante ($A_k \subset A_{k+1}$) pour l'inclusion et de réunion Ω'_ε . Par continuité monotone de \mathbb{P} , on a

$$\lim_{k \rightarrow +\infty} \mathbb{P}(A_k) = \mathbb{P}\left(\bigcup_k A_k\right) = \mathbb{P}(\Omega'_\varepsilon) = 1.$$

D'où $\forall \eta > 0$, $\exists k_0$ tel que pour $k \geq k_0$, $\mathbb{P}(A_k) \geq 1 - \eta$. En particulier, la traduction de $\mathbb{P}(A_{k_0}) \geq 1 - \eta$ donne :

$$\forall n \geq k_0, \quad \mathbb{P}(|X_n - X| < \varepsilon) > 1 - \eta,$$

ce qui justifie la convergence en probabilité de X_n vers X . ■

Remarque 8.2.2 La réciproque n'est pas vraie. Cependant, si X_n converge vers X en probabilité, on peut montrer qu'il existe une sous-suite de X_n qui converge presque sûrement vers X .

8.3 Loi des grands nombres

8.3.1 Loi faible des grands nombres

La loi des grands nombres est la formulation rigoureuse des faits intuitifs suivants : si on lance un « grand » nombre de fois une pièce en l'air, il y aura en moyenne 50% de piles. De même, si on lance un « grand » nombre de fois un dé à 6 faces en l'air, il y aura en moyenne 1/6-ème des faces qui seront, par exemple, des 4 (si la pièce et le dé sont équilibrés). Il existe deux versions de la LGN : la faible où on énonce la convergence *en probabilité* et la forte avec la convergence *presque sûre*.

Théorème 8.3.1 (Loi faible des grands nombres) *Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires (deux à deux) indépendantes et de même loi avec un moment d'ordre 2. Alors*

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} E[X_1], \quad n \rightarrow +\infty.$$

La LGN énonce la convergence (en probabilité) de la moyenne arithmétique M_n vers la moyenne probabiliste $E[X_1]$.

Elle est encore vraie en supposant seulement l'existence du moment d'ordre 1 : $E[|X_1|] < +\infty$.

Démonstration : Ici, la v.a. limite est la constante $E[X_1]$ ($= E[X_i]$ pour tout i car les v.a. X_i ont même loi, donc même espérance). Il s'agit de vérifier

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - E[X_1] \right| \geq \varepsilon \right) = 0.$$

Posons $M_n = \frac{1}{n} \sum_{i=1}^n X_i$, par linéarité, on a $E[M_n] = \frac{1}{n} \sum_{i=1}^n E[X_i] = E[X_1]$. D'autre part, par indépendance des X_i , on a grâce à la proposition 8.1.3 :

$$\begin{aligned} \text{Var}(M_n) &= \text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) = \frac{1}{n^2} \text{Var} \left(\sum_{i=1}^n X_i \right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_1) = \frac{1}{n^2} \times (n \text{Var}(X_1)) = \frac{\text{Var}(X_1)}{n}. \end{aligned}$$

L'inégalité de Tchebychev appliquée à M_n donne alors pour tout $\varepsilon > 0$:

$$\begin{aligned} \forall n \in \mathbb{N}^*, \quad \mathbb{P} \left(\left| \frac{1}{n} \left(\sum_{i=1}^n X_i \right) - E[X_1] \right| \geq \varepsilon \right) &= \mathbb{P}(|M_n - E[M_n]| \geq \varepsilon) \\ &\leq \frac{\text{Var}(M_n)}{\varepsilon^2} \end{aligned}$$

$$\leq \frac{\text{Var}(X_1)}{n\varepsilon^2}. \quad (8.4)$$

On conclut en faisant tendre n vers $+\infty$. ■

Remarque 8.3.1 Plus que la convergence, nous avons obtenu la *vitesse* de convergence : d'après (8.4) elle est en $1/n$. Si on connaît $\text{Var}(X_1)$, on peut donc pour une proportion donnée, trouver un rang n_0 tel que pour $n \geq n_0$ et pour cette proportion de $\omega \in \Omega$, on ait la moyenne arithmétique $\frac{1}{n} \sum_{i=1}^n X_i$ à moins de ε de l'espérance $E[X_1]$.

Souvent, on se trouve dans le cas particulier où les v.a. considérées sont de loi de Bernoulli, la LGN se réécrit alors :

Corollaire 8.3.1 Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de Bernoulli de même paramètre p . Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} p, \quad n \rightarrow +\infty.$$

Démonstration : La LGN (théorème 8.3.1) s'applique car $E[X_i^2] = p < \infty$ et elle donne le résultat car $E[X_i] = p$ quand $X_i \sim b(p)$.

C'est ce résultat qui formalise le résultat intuitif sur le lancer des dés ou des pièces : avec

$$X_i = \begin{cases} 1 & \text{si on obtient le 4 au } i\text{-ème lancer} \\ 0 & \text{si on n'obtient pas le 4 au } i\text{-ème lancer} \end{cases} = \mathbf{1}_{\{\text{obtenir le 4 au } i\text{-ème lancer}\}},$$

on a $X_i \sim b(1/6)$ et $p = 1/6$ et $\frac{1}{n} \sum_{i=1}^n X_i$ désigne la fréquence d'apparition du 4 sur les n premiers lancers qui tend vers $1/6$ d'après le corollaire 8.3.1.

Application : estimation d'une proportion inconnue

On se propose d'estimer le paramètre p inconnu d'une loi de Bernoulli en observant un grand nombre de fois un phénomène aléatoire de loi de Bernoulli $b(p)$, c'est à dire en observant les valeurs d'une suite de v.a. $X_i(\omega)$ indépendantes et de loi de Bernoulli $b(p)$.

Considérons une urne comportant des boules rouges en proportion inconnue p et des boules vertes (en proportion $1 - p$).

D'après la LGN, un grand nombre de tirages de boules dans l'urne donnera une estimation de la proportion p en comptant (la fréquence du) nombre de boules rouges ainsi tirées. Seulement, quel est le nombre raisonnable de boules à tirer pour avoir une réponse assez précise ?

On effectue n tirages d'une boule avec remise. Notons

$$X_i = \begin{cases} 1 & \text{si la boule tirée est rouge} \\ 0 & \text{si la boule tirée est verte} \end{cases} = \mathbf{1}\{\text{la boule est rouge au } i\text{-ème tirage}\}.$$

Désignons toujours par M_n la moyenne arithmétique des n premières v.a. X_i . Ici cela correspond à la fréquence d'apparition des boules rouges lors des n premiers tirages. D'après la loi faible des grands nombres (ou plutôt son corollaire 8.3.1 pour les proportions), M_n converge en probabilité vers p :

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} p, \quad n \rightarrow +\infty.$$

On v.a. donc estimer p par M_n pour n assez grand.

En fait, on observe *une* valeur particulière $M_n(\omega)$ calculée à partir des n tirages réellement effectués mais peut-être que cette valeur particulière n'est pas une bonne approximation de p : imaginez qu'on ne tire que des boules rouges, on aurait $M_n(\omega) = 1$ qui n'est sans doute pas une bonne approximation de p ; ce qui nous sauve, c'est qu'un tel tirage est peu probable.

Mais alors, la question pratique qui se pose est de donner un intervalle fourchette I à partir de l'observation de $M_n(\omega)$ pour p et de contrôler le risque (toujours possible) que p ne soit pas du tout dans l'intervalle I proposé.

Pour cela, on dispose de l'inégalité de Tchebychev, qui pour M_n s'écrit :

$$\mathbb{P}(|M_n - p| \geq t) \leq \frac{\text{Var}(X_1)}{nt^2} = \frac{p(1-p)}{nt^2} \leq \frac{1}{4nt^2}.$$

car $p(1-p) \leq 1/4$ en majorant sur $[0, 1]$ la fonction $x \mapsto x(1-x)$ qui atteint son maximum en $1/2$ où il vaut $1/4$. D'où

$$\mathbb{P}(p \in]M_n - t, M_n + t]) = \mathbb{P}(M_n - t < p < M_n + t) = 1 - \mathbb{P}(|M_n - p| \geq t) \geq 1 - \frac{1}{4nt^2}. \quad (8.5)$$

En pratique, on fait n tirages, on observe $M_n(\omega)$ et on dit que $I =]M_n(\omega) - t, M_n(\omega) + t[$ est un intervalle de confiance (ou fourchette). L'équation (8.5) permet de voir que la probabilité de se tromper (*i.e.* en fait $p \notin I$) est majorée par $1/(4nt^2)$.

Si on se fixe un seuil d'erreur $\alpha \in]0, 1[$ (en général proche de 0 pour que l'erreur soit faible), on trouve t_α tel que $1/(4nt_\alpha^2) = \alpha$ et l'intervalle $I_\alpha =]M_n(\omega) - t_\alpha, M_n(\omega) + t_\alpha[$ est l'intervalle de confiance au niveau $1 - \alpha$: on peut annoncer que p est dans l'intervalle $I_\alpha =]M_n(\omega) - t_\alpha, M_n(\omega) + t_\alpha[$ avec un risque d'erreur de α .

Exemple (Sondage) : Avant le second tour d'une élection, opposant les candidats D et G, un institut de sondage interroge au hasard 1000 personnes dans la rue. On note p la proportion d'électeurs décidés à voter pour G dans la population totale et on suppose l'échantillon de personnes interrogées représentatif. Dans l'échantillon sondé, cette proportion est égale à 0,54. Proposer un intervalle de confiance pour p avec un risque d'erreur de 5%.

Le sondage peut être assimilé à une suite de 1000 tirages de boules avec remise (la réponse d'un électeur de l'échantillon correspondant au tirage d'une boule d'une certaine couleur selon son choix de vote), on est alors ramené à la situation de l'exemple précédent. Ici, la fréquence observée du choix du candidat G sur les 1000 électeurs est $M_{1000}(\omega) = 0,54$ et l'intervalle de confiance est

$$I =]0,54 - t; 0,54 + t[$$

avec un niveau de confiance supérieur à

$$1 - 1/(4 \times 1000 \times t^2).$$

Ici, on veut un seuil de confiance d'au moins 0,95, il faut alors

$$1 - \frac{1}{4000 \times t^2} \geq 0,95 \iff \frac{1}{4000 \times t^2} \leq 0,05 \iff t \geq \frac{1}{10\sqrt{2}} \simeq 0,0707.$$

Avec $t = 0,071$, on trouve l'intervalle de confiance $I =]0,469 ; 0,611[$. On constate en particulier qu'une zone de l'intervalle de confiance correspond à une proportion inférieure à $1/2$, pour lequel G ne serait pas élu alors que la proportion observée semblait lui garantir l'élection.

On ne peut donc pas garantir l'élection de G avec une probabilité d'erreur inférieure à 5%.

Combien de personnes faut-il alors interroger pour donner une fourchette à $\pm 1\%$ avec un seuil de 95% ?

Repartons de (8.5), avec une fourchette de $t = 0,01$. On veut un seuil de confiance d'au moins 0,95 donc un risque d'erreur $\alpha \leq 0,05$:

$$\frac{1}{4n \times 0,01^2} \leq 0,05.$$

On trouve $n = 50\,000$, ce qui donne au sondage un coût prohibitif. En gros, pour améliorer la précision d'un facteur 10, il faut interroger 100 fois plus de personnes et donc multiplier les coûts par 100.

8.3.2 Lemme de Borel-Cantelli

Définition 8.3.1 Soit (A_n) une suite d'événements observables. On pose

$$\overline{\lim}_{n \rightarrow +\infty} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k,$$

$$\underline{\lim}_{n \rightarrow +\infty} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k.$$

On parle respectivement de limites supérieure et inférieure de la suite d'ensembles $(A_n)_n$.

L'ensemble $\overline{\lim}_{n \rightarrow +\infty} A_n$ désigne l'ensemble des $\omega \in \Omega$ qui sont dans une infinité d'ensembles A_i .

L'ensemble $\underline{\lim}_{n \rightarrow +\infty} A_n$ désigne l'ensemble des $\omega \in \Omega$ qui sont dans tous les ensembles A_i à partir d'un certain rang.

Notons de plus que $\underline{\lim}_{n \rightarrow +\infty} A_n \subset \overline{\lim}_{n \rightarrow +\infty} A_n$.

Théorème 8.3.2 (Premier lemme de Borel-Cantelli) *Soit $(A_n)_{n \geq 1}$ une suite d'évènements observables. Si la série suivante converge*

$$\sum_{n=1}^{+\infty} \mathbb{P}(A_n) < +\infty,$$

alors

$$\mathbb{P}(\overline{\lim}_{n \rightarrow +\infty} A_n) = 0.$$

Démonstration : Posons $B_n = \bigcup_{k \geq n} A_k$. La suite $(B_n)_n$ est décroissante ($B_{n+1} \subset B_n$) et l'intersection des B_n est $\overline{\lim}_{n \rightarrow +\infty} A_n$. D'après le théorème de continuité monotone séquentielle (cf. Proposition ??), on a

$$\mathbb{P}(\overline{\lim}_{n \rightarrow +\infty} A_n) = \mathbb{P}\left(\bigcap_n B_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(B_n).$$

Or

$$\mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k=n}^{+\infty} \mathbb{P}(A_k) := r_n.$$

Comme r_n est le reste d'une série convergente, r_n est de limite nulle et donc

$$\mathbb{P}(\overline{\lim}_{n \rightarrow +\infty} A_n) = 0. \quad \blacksquare$$

Remarque 8.3.2 Le deuxième lemme de Borel-Cantelli complète le premier : si la série $\sum_n \mathbb{P}(A_n)$ diverge et qu'en plus les A_n sont des évènements indépendants alors la limite supérieure des A_n est de probabilité 1.

8.3.3 Loi forte des grands nombres

Il existe une version de la loi des grands nombres pour la convergence presque sûre, on parle de la loi forte (car la convergence presque sûre est plus forte que celle en probabilité d'après la proposition 8.2.1) :

Théorème 8.3.3 (Loi forte des grands nombres) Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi avec un moment d'ordre quatre (i.e. $E[X_1^4] < +\infty$). Alors

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} E[X_1].$$

Réciproquement, si $\frac{1}{n} \sum_{i=1}^n X_i$ converge presque sûrement vers c quand $n \rightarrow +\infty$ alors les variables ont un moment d'ordre 1, $E[|X_1|] < +\infty$ et leur espérance est $E[X_i] = c$.

Remarque 8.3.3 En fait, il suffit qu'un moment d'ordre 1 existe. Mais on se contente de la preuve dans le cas où le moment d'ordre 4 existe (c'est déjà assez compliqué).

Démonstration : Il suffit de prouver le théorème quand $E[X_1] = 0$, le cas général s'obtenant par translation. Posons

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_n = \sum_{i=1}^n X_i.$$

Soit $\varepsilon > 0$ et $D_\varepsilon = \overline{\lim}_{n \rightarrow +\infty} \{|M_n| \geq \varepsilon\}$.

On va utiliser le lemme de Borel-Cantelli pour montrer que $\mathbb{P}(D_\varepsilon) = 0$.

On conclura alors en montrant que $D = \bigcup_n D_{1/n}$ est de probabilité nulle. En effet on a $D = \{M_n \not\rightarrow 0\}$, donc le résultat est acquis si on montre que D est de probabilité nulle.

Afin d'utiliser le lemme de Borel-Cantelli, on montre la convergence de la série de terme général $\mathbb{P}(|M_n| \geq \varepsilon)$. Or

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

Par l'inégalité de Markov, on a alors

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

Il reste à estimer $E[S_n^4]$.

$$\begin{aligned} S_n^4 &= (X_1 + X_2 + \dots + X_n)^4 \\ &= \sum_{k_1, k_2, k_3, k_4 \in \{1, \dots, n\}^4} X_{k_1} X_{k_2} X_{k_3} X_{k_4} \\ &= \begin{cases} M(4) \sum_{i=1}^n X_i^4 & + M(1, 3) \sum_{1 \leq i < j \leq n} X_i^3 X_j \\ + M(2, 2) \sum_{1 \leq i < j \leq n} X_i^2 X_j^2 & + M(2, 1, 1) \sum_{1 \leq i < j < k \leq n} X_i^2 X_j X_k \\ + M(1, 1, 1, 1) \sum_{1 \leq i < j < k < l \leq n} X_i X_j X_k X_l \end{cases} \end{aligned}$$

où $M(i_1, \dots, i_p)$ désigne le nombre de 4-uplets (u_1, \dots, u_4) de $\{1, \dots, n\}$ en prenant i_1 fois la valeur u_1, \dots, i_p fois la valeur u_p .

La linéarité et l'indépendance des X_i donnent alors $E[S_n^4] =$

$$\left\{ \begin{array}{l} M(4) \sum_{i=1}^n E[X_i^4] \\ + M(2, 2) \sum_{1 \leq i < j \leq n} E[X_i^2] E[X_j^2] \\ + M(1, 1, 1, 1) \sum_{1 \leq i < j < k < l \leq n} E[X_i] E[X_j] E[X_k] E[X_l] \end{array} \right. + \begin{array}{l} M(1, 3) \sum_{1 \leq i < j \leq n} E[X_i^3] E[X_j] \\ + M(2, 1, 1) \sum_{1 \leq i < j < k \leq n} E[X_i^2] E[X_j] E[X_k] \end{array}$$

Comme $E[X_i] = 0$, les deuxième, quatrième et cinquième termes sont nuls. Comme on montre que $M(4) = 1$, $M(2, 2) = 6$, on obtient

$$\begin{aligned} E(S_n^4) &= \sum_{i=1}^n E[X_i^4] + 6 \sum_{1 \leq i < j \leq n} E[X_i^2] E[X_j^2] \\ &= nE[X_1^4] + 6C_n^2 (E[X_1^2])^2 \\ &= nE[X_1^4] + 3n(n-1)(E[X_1^2])^2 \\ &\leq Mn + 3Mn(n-1) \\ &\leq 3Mn^2 < +\infty \end{aligned}$$

où on a posé $M = \max(E[X_1^2]^2, E[X_1^4])$. On a alors

$$\mathbb{P}(|M_n| \geq \varepsilon) \leq \frac{E[S_n^4]}{n^4 \varepsilon^4} \leq \frac{3M}{n^2 \varepsilon^4}.$$

Comme $3M/n^2 \varepsilon^4$ est le terme général d'une série convergente, $\mathbb{P}(|M_n| \geq \varepsilon)$ aussi. Le lemme de Borel-Cantelli s'applique et donne $\mathbb{P}(D_\varepsilon) = 0$. Posons alors

$$D = \bigcup_{p=1}^{+\infty} D_{1/p}$$

On a $\mathbb{P}(D) = 0$ car D est réunion dénombrable d'ensembles $D_{1/p}$ de probabilités nulles. Prenons alors

$$\Omega_0 := D^c = \bigcap_{p \geq 1} D_{1/p}^c = \bigcap_{p \geq 1} \bigcup_{k \geq 1} \bigcap_{n \geq k} \{|M_n| \leq 1/p\}.$$

On a $\mathbb{P}(\Omega_0) = 1$ et pour tout $\omega \in \Omega_0$, par traduction dans le langage logique des symboles ensemblistes, pour tout $p \in \mathbb{N}^*$, il existe un entier k tel que pour tout $n \geq k$ $|M_n| \leq 1/p$.

On a donc M_n qui converge presque sûrement vers 0; ce qui achève la preuve de la LGN forte. \blacksquare

8.4 Théorème central limite

Rappelons deux résultats essentiels pour les v.a. normales.

- Si X est de loi $\mathcal{N}(m, \sigma^2)$ alors pour tout $\alpha \in \mathbb{R}$, αX est de loi $\mathcal{N}(\alpha m, \alpha^2 \sigma^2)$.

Puis on peut toujours se ramener à la loi normale standard car $\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1)$.

Considérer $\frac{X - m}{\sigma}$ s'appelle centrer et réduire la v.a. X .

- Si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ sont indépendantes alors

$$X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2).$$

Théorème 8.4.1 (Théorème central limite) *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, de moyenne m et de variance σ^2 (i.e. avec un moment d'ordre deux fini). Notons M_n les moyennes arithmétiques*

$$M_n = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i,$$

et Z_n les variables centrées réduites associées :

$$Z_n = \frac{M_n - m}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(M_n - m)}{\sigma}.$$

Alors pour tout intervalle $[a, b]$, on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(a \leq Z_n \leq b) = \int_a^b \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt. \quad (8.6)$$

On dit que la loi de la v.a. $Z_n = \frac{\sqrt{n}(M_n - m)}{\sigma}$ converge en loi vers la loi normale standard $\mathcal{N}(0, 1)$.

Remarque 8.4.1 – Ce théorème justifie le rôle fondamental de la loi normale : si $(X_n)_n$ est une suite de v.a. indépendantes de même loi centrée ($E[X_1] = m = 0$) et réduite (de variance $\sigma = 1$) alors (8.6) se réécrit :

$$\forall a < b, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(a \leq \sqrt{n}M_n \leq b) = \int_a^b \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt.$$

Autrement dit les sommes renormalisées se comportent asymptotiquement comme la loi normale. De façon générale, l'écart entre les moyennes arithmétiques et l'espérance (écart qui tend vers 0 par la LGN) se comporte après normalisation comme la loi normale.

- **En pratique** : lorsque l'on considère un grand nombre de v.a. indépendantes et de même loi X_1, \dots, X_n , on approxime leur somme S_n ou leur moyenne M_n par des variables normales suivantes :

$$S_n \text{ "}\sim\text{" } \mathcal{N}(nm, \sigma^2 n), \quad M_n \text{ "}\sim\text{" } \mathcal{N}(m, \sigma^2/n).$$

En particulier si X_n est de loi $\mathcal{B}(n, p)$, on peut voir X_n comme une somme de n v.a. indépendantes de loi de Bernoulli $b(p)$. D'après la remarque précédente, on a

Proposition 8.4.1 (Moivre-Laplace) *La loi binomiale $\mathcal{B}(n, p)$ s'approxime par la loi normale $\mathcal{N}(np, np(1-p))$ lorsque n est grand.*

On a donc deux approximations possibles pour les lois binomiales $\mathcal{B}(n, p)$: celle par une loi de Poisson $\mathcal{P}(np)$ lorsque n est grand, p petit et np de l'ordre de quelques unités et celle par $\mathcal{N}(np, np(1-p))$ lorsque n est grand. Seule la pratique permet de décider laquelle des deux est la meilleure approximation.

- Le TCL est fondamental en statistique pour l'obtention d'intervalles de confiance. Il est à l'origine de beaucoup d'approximation de lois et permet de se ramener à la loi normale pour laquelle on dispose de tables des valeurs.

Exemple : Un joueur lance une pièce équilibrée : lorsqu'il obtient pile, il gagne 100 Euros, lorsqu'il obtient face, il perd 100 Euros. Estimer le nombre maximal de lancers à effectuer pour que ce joueur ait plus de 95 chances sur 100 de perdre au plus 2000 Euros.

Notons n le nombre de lancers effectués, la v.a. X_n égale au nombre de piles obtenus sur les n premiers lancers suit une loi $\mathcal{B}(n, 1/2)$ et le gain (algébrique) vaut :

$$G_n = 100 \times X_n - 100 \times (n - X_n) = 200X_n - 100n.$$

On cherche alors n tel que $\mathbb{P}(G_n \geq -2000) \geq 0,95$. Or $\{G_n \geq -2000\} = \{X_n - n/2 \geq -10\}$.

Comme X_n de loi binomiale, peut être vue comme une somme $X_n = \epsilon_1 + \dots + \epsilon_n$ de v.a. de loi $b(1/2)$, on peut approximer la loi de X_n , d'après le TCL par la loi normale $\mathcal{N}(\frac{n}{2}, \frac{n}{4})$

et donc celle de $X_n^* = \frac{X_n - n/2}{\sqrt{n/4}}$ par la loi $\mathcal{N}(0, 1)$.

Chercher n tel que $\mathbb{P}(G_n \geq -2000) = \mathbb{P}(X_n - n/2 \geq -10) \geq 0,95$ revient à estimer n tel que

$$\mathbb{P}(\mathcal{N}(0, 1) \geq -20/\sqrt{n}) \geq 0,95 \quad \text{ou par symétrie de la loi} \quad \mathbb{P}(\mathcal{N}(0, 1) \leq 20/\sqrt{n}) \geq 0,95.$$

La table de la loi $\mathcal{N}(0, 1)$ donne alors

$$\frac{20}{\sqrt{n}} = 1,65 \quad \text{c'est à dire} \quad n = \left(\frac{20}{1,65}\right)^2 = 146.$$

Exemple : On lance 3600 fois un dé. Évaluer la probabilité que le nombre d'apparitions du 1 soit compris entre 540 et 660.

Soit S le nombre d'apparitions du 1. S suit la loi $\mathcal{B}(3600, 1/6)$ et donc sa loi peut être approchée par celle de $\tilde{S} \sim \mathcal{N}(600, 500)$.

Mais comme $X_0 = \frac{\tilde{S} - 600}{\sqrt{500}} \sim \mathcal{N}(0, 1)$, on a

$$\begin{aligned} \mathbb{P}(540 \leq S \leq 660) &\simeq \mathbb{P}(540 \leq \tilde{S} \leq 660) \\ &= \mathbb{P}\left(\frac{540 - 600}{\sqrt{500}} \leq \frac{\tilde{S} - 600}{\sqrt{500}} \leq \frac{660 - 600}{\sqrt{500}}\right) \\ &= \mathbb{P}(-2, 68 \leq X_0 \leq 2, 68) \\ &= \mathbb{P}(X_0 \leq 2, 68) - \mathbb{P}(X_0 \leq -2, 68) \\ &= 2\mathbb{P}(X_0 \leq 2, 68) - 1 \\ &\simeq 0, 9926. \end{aligned}$$

Exemple : Une entreprise emploie 500 personnes qui déjeunent à la cantine à l'un ou l'autre des deux services avec une probabilité égale de manger au premier ou au second service. Si le gérant veut avoir une probabilité supérieure à 95% de disposer d'assez de couverts, combien devra-t-il en prévoir à chacun des deux services ?

On commence par numéroter les 500 personnes de 1 à 500 et on note pour chacune X_i la variable aléatoire qui vaut 1 si la i ème personne choisit le premier service (avec probabilité $1/2$) et 0 sinon. Les X_i sont donc des v.a. de Bernoulli $b(1/2)$.

Nous cherchons k le nombre minimal de couverts à disposer à chaque service, sinon 500 couverts conviennent sans prendre le moindre risque.

Le nombre de personnes déjeunant au premier service est $S_{500} = \sum_{i=1}^{500} X_i$ de loi $\mathcal{B}(500, 1/2)$. Le nombre de personnes déjeunant au second service est $500 - S_n$ (on suppose que tout le monde mange exactement une fois).

Le problème revient à chercher le plus petit k tel que

$$\mathbb{P}(S_{500} \leq k, 500 - S_{500} \leq k) \geq 0, 95$$

c'est à dire

$$\mathbb{P}(500 - k \leq S_{500} \leq k) \geq 0, 95.$$

D'après le théorème de Moivre-Laplace, on peut approcher la loi de S_{500} par $\mathcal{N}(250, 125)$.

Notons X_0 une v.a. suivant une telle loi, on a $\frac{X_0 - 250}{\sqrt{125}} \sim \mathcal{N}(0, 1)$. Notons F_0 la fonction de répartition de X_0 (pour laquelle on dispose d'une table des valeurs approchées). On a alors

$$\begin{aligned} \mathbb{P}(500 - k \leq S_{500} \leq k) &\simeq \mathbb{P}(500 - k \leq X_0 \leq k) \\ &= \mathbb{P}\left(\frac{250 - k}{\sqrt{125}} \leq X_0 \leq \frac{k - 250}{\sqrt{125}}\right) \\ &= F_0\left(\frac{k - 250}{\sqrt{125}}\right) - F_0\left(\frac{250 - k}{\sqrt{125}}\right) \end{aligned}$$

$$= 2F_0\left(\frac{k-250}{\sqrt{125}}\right) - 1.$$

Pour obtenir une probabilité d'au moins 0,95, il faut que $F_0\left(\frac{k-250}{\sqrt{125}}\right) \geq 0,975$, ce qui d'après la table de la loi normale standard est vrai pour

$$\frac{k-250}{\sqrt{125}} \geq 1,96 \text{ c'est à dire } k \geq 250 + 1,96\sqrt{125} \simeq 271,91.$$

Il faut donc au minimum 272 couverts à chacun des deux services pour qu'avec une probabilité de 95%, chacun puisse manger au service de son choix.

En acceptant les 5% de risque, il y a moyen de réaliser une économie considérable en place et en mobilier.