



INTRODUCTION AUX PROBABILITÉS

Stéphane BALAC - Olivier MAZET

CENTRE DE MATHÉMATIQUES
INSTITUT NATIONAL DES SCIENCES APPLIQUÉES DE LYON

Avant Propos

Ce polycopié correspond au cours de probabilités enseigné par les auteurs en troisième année dans les départements INFORMATIQUE et TÉLÉCOMMUNICATION SERVICES ET USAGES de l'INSA de Lyon.

Ce cours est un cours de probabilités élémentaires dont les pré-requis sont les notions d'analyse de premier cycle. Aucune connaissance de la théorie de la mesure et de l'intégrale de Lebesgue n'est requise. Ces notions sont néanmoins indispensables pour une formalisation mathématique complète d'une probabilité sur un univers continu. Ceci nous a amené dans ce cours à admettre sans démonstration un certain nombre de résultats concernant les probabilités sur un univers continu et à les considérer comme une généralisation de résultats établis dans le cas d'un univers dénombrable.

Tout commentaire concernant ce polycopié sera recueilli avec intérêt à l'une des adresses suivantes :

Stéphane Balac

Centre de Mathématiques
INSA de LYON
21 avenue Capelle
69621 Villeurbanne
stephane.balac@insa-lyon.fr

Olivier Mazet

Centre de Mathématiques
INSA de LYON
21 avenue Capelle
69621 Villeurbanne
olivier.mazet@insa-lyon.fr

Introduction

« Assurances, diagnostic médical, définition du risque nucléaire, produits financiers virtuels aussi bien que sondages, prévisions météorologiques ou économiques, fiabilité des installations industrielles sont des champs d'application de la science du hasard. Sans théorie des probabilités il ne serait pas possible de formaliser les mécanismes de la génétique, pas plus que la mécanique quantique ou la thermodynamique. Dans le domaine des sciences sociales, le poids de ces mathématiques n'a cessé de grandir ; des choix politiques importants concernant les technologies ou l'économie sont soumis à des analyses fondées sur le calcul des probabilités. C'est pourquoi les notions clés de la théorie des probabilités doivent faire partie de notre culture.»¹

Le propos de ce cours est donc de présenter les bases de la théorie des probabilités, et de définir les principaux outils de modélisation qui en découlent.

Les préoccupations de modélisation des jeux de hasard sont fort anciennes. (Hasard : mot d'origine arabe signifiant «jeu de dés»). Il s'agissait de quantifier et de prédire, dans le but évident de gagner ! Pascal fut le premier à élaborer une *théorie des jeux*, qui a évolué ensuite en *théorie des probabilités* avec Laplace, Gauss, Poisson au XVIII^e siècle, puis Poincaré, Borel, Lévy et Kolmogorov au XX^e siècle.

De nos jours, des outils très divers de la théorie des probabilités s'appliquent dans différents domaines : Analyse (théorie du potentiel), Physique (Mécanique Statistique), Génétique (Chaînes de Markov cachées), Informatique (Chaînes de Markov, reconnaissance de forme), Économie (mathématiques financières), etc.

La théorie des probabilités constitue un outil puissant de modélisation mathématique. Un modèle s'applique sur le réel perçu, d'une part pour quantifier, d'autre part pour prédire. Dans le cas des probabilités, la théorie vise à modéliser le hasard, ou autrement dit les phénomènes aléatoires, l'imprévisible.

Il existe deux manières de construire un modèle :

- Manière inductive, fondée sur l'expérience passée, sans en comprendre forcément les causes.
- Manière déductive, en analysant les causes «physiques».

Dans le premier cas, les probabilités utilisent les statistiques, pour déterminer les lois sous-jacentes. Exemple : quelle est la probabilité que la première naissance de l'an 2020 sur la commune de Lyon soit un garçon ? On sait «par expérience» que celle-ci est très proche de $\frac{1}{2}$, mais on connaît très mal les «causes» de ce résultat, qui sont comprises dans un processus biologique complexe.

Dans le deuxième cas, le raisonnement déterministe dit que connaissant les conditions initiales, on est capable de connaître les conditions finales. Mais la somme et la complexité

¹Didier DACUNHA-CASTELLE, *Chemins de l'aléatoire*, Champs-Flammarion, 1996.

d'enchaînement de phénomènes déterministes amènent la pertinence de la modélisation aléatoire. Exemple : lacher d'une pièce de monnaie de 1 mm de hauteur, puis de 1 m. Dans le premier cas, il est fort aisé de prédire le résultat, contrairement au second, alors que les lois physiques et mécaniques mises en cause sont les mêmes !

Table des matières

1 Rappels d'analyse combinatoire	1
1 Notions de dénombrement	1
1.1 Permutations	1
1.2 Arrangements sans répétition	2
1.3 Combinaisons sans répétition	2
1.4 Arrangements avec répétition	3
1.5 Combinaisons avec répétition	4
1.6 Permutation avec répétition	4
1.7 Pratique du dénombrement	5
1.8 Tirages	6
2 Rappels de théorie des ensembles	7
2.1 Sous-ensembles	7
2.2 Réunion et Intersection	9
2.3 Ensemble complémentaire	10
2 Fondements de la théorie des Probabilités	13
1 Espace d'événements	13
1.1 Univers	13
1.2 Événement	13
2 Espace probabilisé	14
2.1 Définition	15
2.2 Propriétés	17
2.3 Exemples	18
3 Probabilités conditionnelles	19
3.1 Problématique	19
3.2 Définition	21
3.3 Formule de Bayes	22
4 Indépendance	24
3 Variables aléatoires réelles	27
1 Définitions	27
1.1 Variable aléatoire	27
1.2 Fonction de répartition	28
1.3 Variables aléatoires discrètes et variables aléatoires continues	30
1.4 Loi d'une variable aléatoire	31
2 Exemples de lois	33

2.1	Variables aléatoires discrètes	33
2.2	Variables aléatoires continues	35
3	Moments d'une variable aléatoire	37
3.1	Espérance	37
3.2	Variance	41
3.3	Autres moments	43
3.4	Espérance et variance pour les v.a.r. de lois usuelles	43
3.5	Inégalités remarquables	45
4	Caractérisation de la loi d'une variable aléatoire	47
4.1	Caractérisation par la fonction de répartition	48
4.2	Autre caractérisation	49
4.3	Fonction caractéristique, fonction génératrice	50
4.4	Fonctions caractéristique et génératrice de certaines lois	53
4	Vecteurs aléatoires	55
1	Couple de variables aléatoires réelles	55
1.1	Lois d'un couple de variables aléatoires réelles	55
1.2	Moments d'un couple de variables aléatoires réelles	57
1.3	Fonction caractéristique et fonction génératrice	59
1.4	Variables aléatoires conditionnelles	59
2	Indépendance de 2 variables aléatoires réelles	61
2.1	Définition	61
2.2	Indépendance et moments	61
2.3	Coefficient de corrélation	63
2.4	Indépendance et fonction caractéristique	64
3	Somme de deux variables aléatoires indépendantes	64
3.1	Loi de la somme de deux variables aléatoires	64
3.2	Fonction caractéristique de la somme de deux variables aléatoires	66
4	Vecteurs aléatoires	67
4.1	Définition	67
4.2	Loi d'un vecteur aléatoire	68
4.3	Exemples de lois pour un vecteur aléatoire	69
4.4	Moments d'un vecteur aléatoire	70
4.5	Fonction caractéristique et fonction génératrice	72
4.6	Indépendance de n variables aléatoires	72
5	Théorèmes limites	75
1	Convergences stochastiques	75
1.1	Convergence en loi	75
1.2	Convergence en probabilité	76
1.3	Convergence presque sûre	77
2	Loi faible des grands nombres	77
3	La loi forte des grands nombres	78
4	Théorème de la limite centrale	79
5	Application : estimation de la moyenne d'une variable aléatoire	80
5.1	Estimations ponctuelles de la moyenne et de la variance	80
5.2	Estimations par intervalle de confiance de la moyenne	81

6 Chaînes de Markov discrètes	83
1 Chaîne de Markov homogène	83
2 Caractérisation d'une chaîne de Markov	87
3 Classification des états d'une chaîne de Markov	89
4 Comportement asymptotique d'une chaîne de Markov	91

Chapitre 1

Rappels d'analyse combinatoire

1 Notions de dénombrement

Pour une grande partie des calculs de probabilités discrètes, on cherche à calculer le nombre d'événements réalisables, le nombre d'événements favorables, etc. D'où la nécessité d'utiliser des outils de dénombrement et de combinatoire. Dans tout ce paragraphe E désigne un ensemble à n éléments que l'on suppose distinguables.

1.1 Permutations

DÉFINITION 1 *On appelle permutation des n éléments de l'ensemble E toute disposition ordonnée de ces n éléments.*

REMARQUE Deux permutations ne diffèrent donc que par l'ordre des n éléments distincts qui la composent.

Le nombre de permutations de n éléments est le nombre de manières possibles d'ordonner ces n éléments.

EXEMPLE Les permutations de l'ensemble $\{1, 2, 3\}$ sont $(1, 2, 3)$, $(1, 3, 2)$, $(2, 1, 3)$, $(2, 3, 1)$, $(3, 1, 2)$, $(3, 2, 1)$.

PROPOSITION 1 *Le nombre de permutations d'un ensemble à n éléments est $n!$.*

DÉMONSTRATION On démontre ce résultat par récurrence : il y a $1!$ manière de permutter 1 élément. Supposons qu'il y en a $(k - 1)!$ pour la permutation de $k - 1$ éléments. Alors étant donnés k éléments, on en choisit 1 parmi k , ce qui donne k possibilités, et il reste $k - 1$ éléments à ordonner, soit $(k - 1)!$ possibilités. Le total fait donc $k(k - 1)! = k!$.

Par convention, on pose $0! = 1$. ○

EXEMPLE DE RÉFÉRENCE : dans une urne contenant n boules distinguables (numérotées), on tire les n boules l'une après l'autre (on s'intéresse à l'ordre), sans les remettre dans l'urne (on n'autorise pas de répétition). Le nombre de tirages possibles est le nombre de permutations de l'ensemble $\{1, 2, \dots, n\}$, c'est-à-dire $n!$.

1.2 Arrangements sans répétition

DÉFINITION 2 *On appelle arrangement sans répétition de p éléments pris parmi les n éléments d'un ensemble E , toute disposition ordonnée de p éléments de E .*

REMARQUE Un arrangement de n éléments pris parmi les n éléments d'un ensemble E est une permutation. Dans un arrangement, on se contente de p éléments pris parmi les n .

EXEMPLE Les arrangements à 2 éléments de l'ensemble $\{1, 2, 3\}$ sont $(1, 2)$, $(1, 3)$, $(2, 1)$, $(2, 3)$, $(3, 1)$, $(3, 2)$.

PROPOSITION 2 *Le nombre d'arrangements sans répétition de p éléments pris dans un ensemble à n éléments est*

$$A_n^p = \frac{n!}{(n-p)!} = n \times (n-1) \times \cdots \times (n-p+1).$$

DÉMONSTRATION Il y a n façons de choisir le 1^{er} élément de l'arrangement parmi les n éléments de l'ensemble. Pour le 2^e élément de l'arrangement, il y a $n-1$ façons de le choisir (puisque il ne doit pas y avoir répétition d'un élément). En itérant, on vérifie qu'il y a $n-p+1$ façons de le choisir p^e élément de l'arrangement. Au total le nombre d'arrangement est donc $n(n-1)\dots(n-p+1)$. \circ

EXEMPLE DE RÉFÉRENCE : dans une urne contenant n boules distinguables (numérotées), On tire les p boules l'une après l'autre (on s'intéresse à l'ordre), sans les remettre dans l'urne (on n'autorise pas de répétition). Le nombre de tirages possibles est le nombre d'arrangements sans répétition à p éléments de l'ensemble $\{1, 2, \dots, n\}$, c'est-à-dire A_n^p .

1.3 Combinaisons sans répétition

DÉFINITION 3 *On appelle combinaison sans répétition de p éléments pris parmi les n éléments d'un ensemble E toute disposition non ordonnée de p éléments de E .*

REMARQUE Deux combinaisons ne diffèrent que par la nature des éléments qui la composent (l'ordre de ces éléments est indifférent).

EXEMPLE Les combinaisons à 2 éléments de l'ensemble $\{1, 2, 3\}$ sont $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$.

PROPOSITION 3 *Le nombre de combinaisons sans répétition de p éléments pris dans un ensemble à n éléments est*

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

DÉMONSTRATION Considérons dans un premier temps le nombre d'arrangements sans répétition de p éléments pris parmi n éléments : $A_n^p = \frac{n!}{(n-p)!}$. Pour une combinaison de p éléments donnée, il y a $p!$ arrangements différents de ces p éléments (correspondant au nombre de permutations des p éléments de la combinaison).

Le nombre de combinaisons sans répétition de p éléments pris parmi n éléments est donc

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

○

PROPOSITION 4 *On a les propriétés suivantes :*

– $C_n^p = C_n^{n-p} = \frac{n!}{p!(n-p)!} \quad \forall (n, p) \in \mathbb{N}^2$.

– $C_n^p = C_{n-1}^{p-1} + C_{n-1}^p \quad \forall (n, p) \in \mathbb{N}^2$.

– *Formule du binôme* : $\forall n \in \mathbb{N}, \forall (a, b) \in \mathbb{R}^2$,

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

EXEMPLE DE RÉFÉRENCE : dans une urne contenant n boules distinguables (numérotées), on tire p boules simultanément (par conséquent sans ordre et sans répétition). Le nombre de tirages possibles est le nombre de combinaisons sans répétition à p éléments de l'ensemble $\{1, 2, \dots, n\}$, c'est-à-dire C_n^p .

1.4 Arrangements avec répétition

DÉFINITION 4 *On appelle arrangement avec répétition de p éléments pris parmi les n éléments d'un ensemble E toute disposition ordonnée de p éléments, non nécessairement distincts, de E .*

EXEMPLE Les arrangements avec répétition à 2 éléments de l'ensemble $\{1, 2, 3\}$ sont $(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3)$.

PROPOSITION 5 *Le nombre d'arrangements avec répétition de p éléments pris dans un ensemble à n éléments est n^p .*

DÉMONSTRATION Il y a p façons de choisir chacun des éléments de l'arrangement parmi les n éléments de l'ensemble. Au total le nombre d'arrangement avec répétition est donc $\underbrace{n \times n \times \dots \times n}_{p \text{ fois}} = n^p$.

○

EXEMPLE DE RÉFÉRENCE : dans une urne contenant n boules distinguables (numérotées), on tire p boules l'une après l'autre (on s'intéresse à l'ordre), en les remettant dans l'urne après chaque tirage (on accepte les répétitions). Le nombre de tirages possibles est le nombre d'arrangements avec répétition à p éléments de l'ensemble $\{1, 2, \dots, n\}$, c'est-à-dire n^p .

1.5 Combinaisons avec répétition

DÉFINITION 5 *On appelle combinaison avec répétition de p éléments pris parmi les n éléments d'un ensemble E toute disposition non ordonnée de p éléments, non nécessairement distincts, de E .*

EXEMPLE Les combinaisons avec répétition à 2 éléments de l'ensemble $\{1, 2, 3\}$ sont $\{1, 1\}, \{1, 2\}, \{1, 3\}, \{2, 2\}, \{2, 3\}, \{3, 3\}$.

PROPOSITION 6 *Le nombre d'arrangements avec répétition de p éléments pris dans un ensemble à n éléments est \mathcal{C}_{n+p-1}^p .*

EXEMPLE DE RÉFÉRENCE : dans une urne contenant n boules distinguables (numérotées), on tire p boules en les remettant dans l'urne après chaque tirage (on accepte les répétitions). On ne s'intéresse pas à l'ordre d'apparition des coules. Le nombre de tirages possibles est le nombre de combinaisons avec répétition à p éléments de l'ensemble $\{1, 2, \dots, n\}$, c'est-à-dire \mathcal{C}_{n+p-1}^p .

1.6 Permutation avec répétition

DÉFINITION 6 *Supposons que les n éléments de l'ensemble E se répartissent en ℓ catégories : il y a n_1 éléments du type 1, n_2 éléments du type 2, ..., n_ℓ éléments du type ℓ (avec $n_1 + n_2 + \dots + n_\ell = n$).*

On appelle permutation avec répétition de n éléments de l'ensemble E toute disposition ordonnée de n éléments où figure n_1 fois un élément du type 1, n_2 fois un élément du type 2, ..., et n_ℓ fois un élément du type ℓ .

REMARQUE La permutation avec répétition n'est pas un cas particulier d'arrangement avec répétition, contrairement au cas «sans répétition». La «répétition» n'agit pas dans le même contexte pour permutation et arrangement.

- Dans le cas des permutations, «répétition» signifie qu'un type d'élément donné de l'ensemble E peut être présent plusieurs fois.
- Dans le cas d'arrangements, «répétition» signifie qu'un même élément de l'ensemble E peut être réutilisé plusieurs fois.

PROPOSITION 7 *Le nombre de permutations avec répétition des n éléments de l'ensemble E (se répartissant en ℓ catégories de la manière suivante : n_1 éléments du type 1, n_2 éléments du type 2, ..., et n_ℓ éléments du type ℓ , avec $n_1 + n_2 + \dots + n_\ell = n$) est*

$$\frac{n!}{n_1! \times \dots \times n_\ell!}.$$

DÉMONSTRATION Il y a en effet $n!$ permutations différentes des n éléments de l'ensemble. Pour obtenir le nombre de permutations avec répétition, il faut considérer le fait que les éléments d'un même type vont fournir une même permutation. Or le nombre de permutations identiques fournies par les éléments de type k est $n_k!$ (nombre de permutation de ces n_k éléments). Au total le nombre de permutations avec répétition est donc $\frac{n!}{n_1! \times \dots \times n_\ell!}$.
○

EXEMPLE DE RÉFÉRENCE : dans une urne contenant n boules de ℓ couleurs différentes se répartissant de la manière suivante : n_1 boules de la couleur c_1 , n_2 boules de la couleur c_2 , ..., n_ℓ boules de la couleur c_ℓ (avec $n_1 + n_2 + \dots + n_\ell = n$), on tire les n boules l'une après l'autre (on s'intéresse à l'ordre), sans les remettre dans l'urne (la répétition joue au niveau des couleurs). Le nombre de tirages possibles distincts est le nombre de permutations des ces n boules avec répétition des ℓ couleurs, c'est-à-dire $\frac{n!}{n_1! \times \dots \times n_\ell!}$.

EXERCICE Quel est le nombre de mots de 6 lettres que l'on peut former à partir des lettres A G E R E A ? (on ne s'intéresse pas au sens du mot !)

1.7 Pratique du dénombrement

Soit une expérience aléatoire \mathcal{E} composée de r expériences successives, la 1^{re} pouvant produire un résultat quelconque parmi n_1 résultats possibles, et la 2^e pouvant produire un résultat quelconque parmi n_2 résultats possibles, et ainsi de suite, la r^e pouvant produire un résultat quelconque parmi n_r résultats possibles. Le nombre total de résultats possibles pour l'expérience aléatoire \mathcal{E} est égal au produit $n_1 \times n_2 \times \dots \times n_r$.

Autrement dit, le nombre de «choix» possibles pour une expérience aléatoire consistant à faire «un choix» **et** «un choix» est obtenu en effectuant le **produit** du nombre de ces choix.

EXEMPLE Calculons le nombre de plaques minéralogiques distinctes disponibles par département dans la numérotation «4 chiffres, 2 lettres».

On choisit donc 4 chiffres parmi 10 avec ordre et avec remise **et** on choisit 2 lettres parmi 26 avec ordre et avec remise. Le nombre total de possibilités est égal au produit $10^4 \times 26^2$, soit 6 760 000.

Si une expérience aléatoire \mathcal{E} peut être réalisée de r manières différentes, la 1^{re} fournissant n_1 résultats distincts possibles, la 2^e fournissant n_2 résultats distincts possibles, etc, la r ^e fournissant n_r résultats distincts possibles, alors le nombre total de résultats possibles pour l'expérience aléatoire \mathcal{E} est égal à la somme $n_1 + n_2 + \dots + n_r$.

Autrement dit, le nombre de «choix» possibles pour une expérience aléatoire consistant à faire «un choix» **ou** «un choix» est obtenu en effectuant la **somme** du nombre de ces choix.

EXEMPLE Dans une urne contient 49 boules numérotées de 1 à 49, on tire simultanément 6 boules. Calculons le nombre de tirages possibles ayant (au moins) 5 numéros entre 1 et 6.

Au moins 5 bons numéros c'est exactement 5 «bons» numéros **ou** exactement 6 «bons» numéros. Il y a 1 seul tirage donnant 6 numéros entre 1 et 6. Le nombre de tirages possibles avec exactement 5 «bons» numéros est $C_6^5 C_{43}^1$ (on choisit sans ordre et sans remise 5 numéros parmi les 6 «bons» numéros **et** on choisit 1 numéro parmi les 43 «mauvais» numéros). Au total, il y a $1 + C_6^5 C_{43}^1$ tirages possibles.

EXERCICE Un problème de dénombrement qu'a résolu PASCAL dans sa correspondance avec le CHEVALIER DE MÉRÉ, est relatif au jeu de «passe-dix» qui se joue avec 3 dés. L'un des joueurs parie que le total des points dépassera 10 («passe-dix»), l'autre qu'il sera inférieur ou égal à 10. Les chances des deux joueurs sont égales. Le CHEVALIER DE MÉRÉ avait observé que le joueur qui parie pour «passe-dix» gagne plus souvent en obtenant 11 points qu'en obtenant 12 points. Or objectait DE MÉRÉ, on peut obtenir 11 points de six manières différentes

$$(6, 4, 1), (6, 3, 2), (5, 5, 1), (5, 4, 2), (5, 3, 3), (4, 3, 3),$$

et on peut également obtenir 12 points de six manières différentes

$$(6, 5, 1), (6, 4, 2), (6, 3, 3), (5, 5, 2), (5, 4, 3), (4, 4, 4).$$

On devrait donc obtenir aussi souvent 11 points que 12 points en observant un grand nombre de parties.

La réponse de PASCAL était fort simple ... et mettait en évidence une erreur de modélisation. Pourriez-vous à votre tour répondre au CHEVALIER DE MÉRÉ ?

1.8 Tirages

La physique (cinétique des gazs, physique nucléaire, ...) fait largement appel aux modèles probabilistes. Ces modèles font souvent référence aux tirages de boules dans des urnes. On considère une urne de n boules distinguables où on effectue p tirages successifs. Combien y a-t-il de tirages possibles ?

	Sans remise	Avec remise
Avec ordre	A_n^p	n^p
Sans ordre	C_n^p	C_{n+p-1}^p

La signification de la notion de **remise** est claire. La notion **d'ordre** revient à différencier toutes les combinaisons des p mêmes boules tirées, en tenant compte de l'ordre dans lequel elles ont été tirées.

2 Rappels de théorie des ensembles

Nous renvoyons aux ouvrages de cours de mathématiques de premier cycle¹ pour un exposé détaillé des résultats de la théorie des ensembles. Nous nous contenterons dans ce paragraphe de rappeler les principales propriétés qui seront utilisées dans la suite de ce cours.

2.1 Sous-ensembles

DÉFINITION 7 Soient A et B deux ensembles. On dit que A est inclus dans B et on note $A \subset B$ si tout élément de A appartient à l'ensemble B . L'ensemble A est alors qualifié de partie ou de sous-ensemble de l'ensemble B .

Pour signifier que A n'est pas inclus dans B , on note $A \not\subset B$.

REMARQUE Pour que l'ensemble A ne soit pas inclus dans B , il faut et il suffit qu'il existe un élément de A qui n'appartienne pas à B . (Par exemple $\mathbb{N} \not\subset \mathbb{R}^*$.)

L'ensemble vide est inclus dans tout ensemble.

Un ensemble est inclus dans lui-même.

PROPOSITION 8 La relation d'inclusion est une relation d'ordre, c'est-à-dire si A, B et C sont trois ensembles, alors

- $A \subset B$ et $B \subset C \Rightarrow A \subset C$ (propriété de transitivité).
- $A \subset B$ et $B \subset A \Rightarrow A = B$ (propriété d'antisymétrie).
- $A \subset A$ (propriété de réflexivité).

DÉFINITION 8 Soient A et B deux ensembles. On dit que les ensembles A et B sont **égaux** et on note $A = B$ si tout élément de l'un des ensembles appartient à l'autre ensemble. Autrement dit, $A = B$ signifie que $A \subset B$ et $B \subset A$.

DÉFINITION 9 Un ensemble Ω est dit fini lorsque le nombre d'éléments qui le composent est un entier naturel. Dans ce cas, le nombre d'éléments est appelé **cardinal de l'ensemble** et est noté $\text{Card}(\Omega)$, ou $\#\Omega$, ou $|\Omega|$.

¹Par exemple «Algèbre et analyse, cours de mathématiques première année avec exercices corrigés», S. Balac et F. Sturm, Presses Polytechniques et Universitaires Romandes, Coll. INSA de Lyon, 2003.

Un ensemble qui n'est pas fini est dit infini. On appelle **singleton** un ensemble composé d'un seul élément. L'ensemble vide est l'ensemble qui ne contient aucun élément. On le note \emptyset . Par convention, $\text{Card}(\emptyset) = 0$.

PROPOSITION 9 *Si $A \subset B$ alors $\text{Card}(A) \leq \text{Card}(B)$.*

DÉFINITION 10 *Un ensemble est dit **dénombrable** s'il existe une bijection entre cet ensemble et l'ensemble des entiers naturels \mathbb{N} .*

DÉFINITION 11 *Soit Ω un ensemble. Les sous-ensembles de Ω définissent un ensemble appelé **ensemble des parties** de Ω et noté $\mathcal{P}(\Omega)$. Il y équivalence entre les assertions $A \in \mathcal{P}(\Omega)$ et $A \subset \Omega$.*

REMARQUE $\mathcal{P}(\Omega)$ contient toujours \emptyset et Ω . Attention, les éléments de $\mathcal{P}(\Omega)$ sont des sous-ensembles de Ω et non pas des éléments de Ω .

EXEMPLE Si $\Omega = \{a, b, c\}$ alors $\mathcal{P}(\Omega) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$. Est-on certain d'avoir énuméré tout les éléments de $\mathcal{P}(\Omega)$? Oui, car ...

PROPOSITION 10 *Soit Ω un ensemble fini de cardinal n . L'ensemble $\mathcal{P}(\Omega)$ des parties de Ω est alors de cardinal 2^n .*

DÉMONSTRATION Effectuons un peu de dénombrement. Il y a 1 ensemble ayant 0 élément, c'est l'ensemble vide. Puisque $\text{Card}(\Omega) = n$, il y a n singleton (ensemble ayant 1 élément). De façon générale, le nombre d'ensembles ayant k éléments correspond au nombre de possibilités de choisir k éléments parmi les n éléments de l'ensemble Ω , ceci sans ordre et sans remise. Il y a donc $\mathcal{C}_n^k = \frac{n!}{k!(n-k)!}$ sous-ensembles de k éléments. Ainsi,

$$\text{Card}(\mathcal{P}(\Omega)) = \sum_{k=0}^n \mathcal{C}_n^k.$$

Or d'après la formule du binôme de Newton, pour tout réel x et y on a

$$(x + y)^n = \sum_{k=0}^n \mathcal{C}_n^k x^k y^{n-k}.$$

Si l'on prend $x = 1$ et $y = 1$ on obtient,

$$(x + y)^n = 2^n = \sum_{k=0}^n \mathcal{C}_n^k.$$

Le résultat est ainsi démontré. ○

Si A est un ensemble, on note \mathbb{I}_A la fonction indicatrice de l'ensemble A définie par

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

2.2 Réunion et Intersection

DÉFINITION 12 Soient Ω un ensemble et A, B deux sous-ensembles de Ω . La réunion des 2 ensembles A et B notée $A \cup B$ est l'ensemble constitué par les éléments de Ω appartenant à A ou à B . Autrement dit

$$A \cup B = \{\omega \in \Omega, \omega \in A \text{ ou } \omega \in B\}.$$

L'intersection des 2 ensembles A et B notée $A \cap B$ est l'ensemble constitué par les éléments de Ω appartenant à A et à B . Autrement dit

$$A \cap B = \{\omega \in \Omega, \omega \in A \text{ et } \omega \in B\}.$$

Si $A \cap B = \emptyset$, on dit que les ensembles A et B sont **disjoints**.

On utilise le symbole \cup_d pour indiquer la réunion de 2 ensembles disjoints. On note

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

et

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

PROPOSITION 11 Soient A et B deux ensembles finis. On a

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B).$$

Si A et B sont disjoints, on a

$$\text{Card}(A \cup_d B) = \text{Card}(A) + \text{Card}(B).$$

PROPOSITION 12 *On a les propriétés suivantes pour l'intersection et la réunion :*

- *Ce sont des lois de composition interne :*
 $si A \in \mathcal{P}(\Omega) \text{ et } B \in \mathcal{P}(\Omega), alors A \cap B \in \mathcal{P}(\Omega) \text{ et } A \cup B \in \mathcal{P}(\Omega).$
- *Ces lois sont commutatives :*
 $\forall A, B \in \mathcal{P}(\Omega) \quad A \cap B = B \cap A, \quad A \cup B = B \cup A.$
- *Ces lois sont associatives :*
 1. $\forall A, B, C \in \mathcal{P}(\Omega) \quad (A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C,$
 2. $\forall A, B, C \in \mathcal{P}(\Omega) \quad (A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C.$
- \emptyset est élément neutre pour l'union et Ω est élément neutre pour l'intersection :
 $\forall A \in \mathcal{P}(\Omega) \quad A \cup \emptyset = A \text{ et } A \cap \Omega = A.$
- *L'intersection est distributive sur la réunion*

$$A \cap \left(\bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i),$$

et la réunion est distributive sur l'intersection

$$A \cup \left(\bigcap_{i=1}^n B_i \right) = \bigcap_{i=1}^n (A \cup B_i).$$

REMARQUE Il existe des correspondances entre les opérateurs ensemblistes et les opérateurs logiques suivants : **et**, **ou**, \exists et \forall

- $(\omega \in A \text{ et } \omega \in B) \iff \omega \in A \cap B$
- $(\omega \in A \text{ ou } \omega \in B) \iff \omega \in A \cup B$
- $\omega \in \bigcap_{i=1}^n A_i \iff \forall i \in \{1, \dots, n\} \omega \in A_i$
- $\omega \in \bigcup_{i=1}^n A_i \iff \exists i \in \{1, \dots, n\} \omega \in A_i$

2.3 Ensemble complémentaire

DÉFINITION 13 *Soient Ω un ensemble et A un sous-ensemble de Ω . On appelle **complémentaire de A dans Ω** , et on note $\mathcal{C}_\Omega A$ ou \bar{A} ou A^c , l'ensemble constitué des éléments de Ω qui n'appartiennent pas à A . Autrement dit,*

$$\mathcal{C}_\Omega A = \{\omega \in \Omega, \omega \notin A\}.$$

Lorsqu'il n'y aura pas d'ambigüité sur l'ensemble Ω de référence, nous utiliserons de préférence la notation A^c .

PROPOSITION 13 Soient A et B deux sous-ensembles de Ω . On a les propriétés suivantes :

- $(A^c)^c = A$.
- $\emptyset^c = \Omega$, $\Omega^c = \emptyset$.
- $A \cap A^c = \emptyset$, $A \cup A^c = \Omega$.
- $(A \cap B)^c = A^c \cup B^c$ et $(A \cup B)^c = A^c \cap B^c$.
- $A \subset B \Rightarrow B^c \subset A^c$.
- $\text{Card}(A^c) = \text{Card}(\Omega) - \text{Card}(A)$.

DÉFINITION 14 Soient A, B deux sous-ensembles d'un ensemble Ω . On appelle **différence des ensembles** A et B , et on note $A \setminus B$, l'ensemble constitué des éléments de A qui n'appartiennent pas à B . Autrement dit,

$$A \setminus B = \{\omega \in \Omega, \omega \in A \text{ et } \omega \notin B\} = A \cap B^c.$$

Chapitre 2

Fondements de la théorie des Probabilités

Dans ce chapitre, nous allons progressivement définir les concepts et outils de construction d'un **modèle probabiliste**.

1 Espace d'événements

1.1 Univers

On appelle *ensemble des possibles*, ou *ensemble des éventualités*, ou encore *univers*, l'ensemble Ω des résultats possibles d'une expérience aléatoire. Les éléments ω de Ω sont appelés *issues*, ou *événements élémentaires*, ou encore *épreuves* de l'expérience aléatoire. L'ensemble Ω peut être fini, infini dénombrable, ou infini non dénombrable.

Exemples d'univers associés à des expériences aléatoires :

Ex₁— Lancer d'une pièce : $\Omega = \{F, P\}$, ensemble fini de cardinal $\text{Card}(\Omega) = 2$.

Ex₂— Lancer d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$, ensemble fini de cardinal $\text{Card}(\Omega) = 6$.

Ex₃— Lancer d'un dé jusqu'à l'obtention d'un 6 : $\Omega = \{1, 2, 3, \dots\} = \mathbb{N}$, ω est le numéro du lancer où l'on obtient 6. Ensemble infini dénombrable.

Ex₄— Choisir au hasard un nombre entre 0 et 1, $\Omega = [0, 1]$: ensemble infini non dénombrable.

1.2 Événement

On appelle *événement* l'ensemble des issues de l'expérience qui vérifient une propriété donnée. C'est donc une partie A de Ω ($A \in \mathcal{P}(\Omega)$).

À titre d'exemple, reprenons les expériences aléatoires énumérées au paragraphe précédent et donnons quelques exemples d'événements associés à ces expériences :

Ex₁— On obtient «face» : $A = \{F\}$.

Ex₂— On obtient un nombre pair : $A = \{2, 4, 6\}$.

Ex₃— On obtient le premier 6 entre le troisième coup et le septième coup (inclus) : $A = \{3, 4, 5, 6, 7\}$.

Ex₄— On choisit un nombre strictement plus grand que $\frac{1}{2}$: $A = \left] \frac{1}{2}, 1 \right]$.

- Deux événements A et B sont dit **incompatibles** si $A \cap B = \emptyset$.
- L'ensemble Ω est un événement appelé **événement certain**.

On associe à l'univers Ω un ensemble d'événements \mathcal{A} appelé **tribu**. Cet ensemble doit vérifier les propriétés suivantes :

- (i) $\Omega \in \mathcal{A}$,
- (ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$,
- (iii) $\forall i \in \mathbb{N}, \quad A_i \in \mathcal{A} \implies \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.

Si les axiomes (i), (ii) et (iii) sont vérifiés, alors le couple (Ω, \mathcal{A}) est appelé **espace d'événements**.

Il n'y a pas unicité de l'espace d'événements (Ω, \mathcal{A}) pour une expérience aléatoire donnée. On peut par exemple associer à un univers Ω les tribus suivantes : $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathcal{A} = \{\emptyset, \Omega\}$, ou encore $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ où A désigne un événement. (On vérifiera à titre d'exercice qu'il s'agit bien de tribus.)

Suivant le choix de la tribu on aura un modèle probabiliste plus ou moins raffiné qui permettra de décrire de manière plus ou moins judicieuse l'expérience aléatoire considérée. Reprenons à titre d'illustration l'exemple Ex₁ du lancer d'une pièce. On a $\Omega = \{P, F\}$. On peut prendre $\mathcal{A} = \{\emptyset, \Omega\}$ mais les seuls renseignements que ce modèle nous apportera concerteront les chances d'obtenir «pile ou face» et les chances de n'obtenir «ni pile ni face» (on ne tient pas compte de la possibilité que la pièce reste sur la tranche!). Ces informations ne sont pas très intéressantes. Il vaut donc mieux «raffiner» le modèle en «grossissant» \mathcal{A} . En prenant $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \Omega, \{P\}, \{F\}\}$, on aura un modèle qui nous donnera plus d'informations sur le résultat de l'expérience : chances de n'obtenir «ni pile ni face», chances d'obtenir «pile ou face», chances d'obtenir «face», chances d'obtenir «pile».

Lorsque Ω est fini ou infini dénombrable, l'ensemble des événements «standard» est $\mathcal{A} = \mathcal{P}(\Omega)$. Lorsque $\Omega = \mathbb{R}$, $\mathcal{P}(\mathbb{R})$ est bien une tribu, mais cette tribu n'est malheureusement pas utilisable dans la théorie classique de la mesure. C'est pourquoi dans le cas où $\Omega = \mathbb{R}$ on choisit pour ensemble des événements «standard» l'ensemble des boréliens.

DÉFINITION 15 *On appelle ensemble des Boréliens, et on note $\mathcal{B}(\mathbb{R})$, l'ensemble des parties de \mathbb{R} engendrées (par intersections et réunions dénombrables) par les intervalles de \mathbb{R} .*

2 Espace probabilisé

L'objectif de ce paragraphe est de définir la façon de «calculer», pour un événement donné, la probabilité qu'il se réalise.

Commençons par présenter une construction intuitive sur l'exemple du lancer de dé. On lance un dé à six faces et on observe le numéro obtenu. On prend pour univers associé à cette expérience aléatoire

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Chaque issue de l'expérience (ou encore chaque événement élémentaire) a «une chance sur six» d'être réalisée (si le dé n'est pas truqué!), donc chaque issue A de l'expérience (par exemple $A = \{3\}$) a une probabilité de $1/6$ de se réaliser. On note $\mathbb{P}(A) = 1/6$.

L'événement «on obtient un nombre pair», noté $B = \{2, 4, 6\}$ a donc «3 chances sur 6» d'être réalisé :

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

L'événement «on obtient un nombre entre 1 et 6» est l'événement *certain*, et a «6 chances sur 6» d'être réalisé : $\mathbb{P}(\Omega) = 1$.

On constate dans cette expérience que chaque événement A a une probabilité d'être réalisé égale à

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)},$$

où $\text{Card}(A)$ désigne le cardinal de l'ensemble A et $\text{Card}(\Omega)$ désigne le cardinal de l'ensemble Ω .

D'après les propriétés de la fonction cardinal, on obtient :

- $\mathbb{P}(\Omega) = \frac{\text{Card}(\Omega)}{\text{Card}(\Omega)} = 1$,
- $\mathbb{P}(A^c) = \frac{\text{Card}(A^c)}{\text{Card}(\Omega)} = \frac{\text{Card}(\Omega) - \text{Card}(A)}{\text{Card}(\Omega)} = 1 - \mathbb{P}(A)$,
- $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$,
- si $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \frac{\text{Card}(A \cup B)}{\text{Card}(\Omega)} = \frac{\text{Card}(A) + \text{Card}(B)}{\text{Card}(\Omega)} = \mathbb{P}(A) + \mathbb{P}(B)$.

On remarquera que cette manière de définir la probabilité d'un événement est étroitement liée d'une part au fait que Ω soit un ensemble fini, d'autre part au fait que chaque événement élémentaire a la même probabilité (on dit qu'il y a *équiprobabilité*). C'est pourquoi on a besoin de définir le concept de «probabilité» applicable dans un cadre beaucoup plus général.

2.1 Définition

DÉFINITION 16 Soit (Ω, \mathcal{A}) un espace d'événements. On appelle probabilité sur (Ω, \mathcal{A}) une application \mathbb{P} définie sur \mathcal{A} vérifiant

1. $\forall A \in \mathcal{A} \quad \mathbb{P}(A) \in [0, 1]$.
2. $\mathbb{P}(\Omega) = 1$.
3. Quelle que soit la suite $(A_i)_{i \in \mathbb{N}}$ d'éléments de \mathcal{A} deux à deux disjoints (i.e. $\forall i, j \in \mathbb{N}$, $A_i \cap A_j = \emptyset$), on a

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

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace probabilisé.

REMARQUE On peut remplacer la condition 3 de la définition 16 par :
quels que soient $A_1, A_2 \in \mathcal{A}$ disjoints ($A_1 \cap A_2 = \emptyset$) on a

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

EXEMPLE On peut vérifier que dans le cas de l'expérience du lancé de dé à six faces où l'univers est

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

l'application $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ définie par $\mathbb{P}(A) = \text{Card}(A)/6$ pour tout $A \in \mathcal{P}(\Omega)$ est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

Cas particuliers importants

1 - Si Ω est un univers fini ou dénombrable, $\Omega = \{\omega_i\}_{i \in I}$, $I \subset \mathbb{N}$, la donnée des valeurs prises par \mathbb{P} sur les singletons (i.e. : la donnée de $\mathbb{P}(\{\omega_i\})$ pour tout $i \in I$) suffit à caractériser l'application \mathbb{P} sur \mathcal{A} . En effet, tout événement $A \in \mathcal{A}$ s'écrit $A = \bigcup_{j \in J} \{\omega_j\}$ où

J est un sous-ensemble de I . D'après la propriété 3 on en déduit que

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{j \in J} \{\omega_j\}\right) = \sum_{j \in J} \mathbb{P}(\{\omega_j\}).$$

2 - Dans le cas où Ω est un univers fini, $\Omega = \{\omega_i\}_{i=1, \dots, n}$, si toutes les issues de l'expérience sont équiprobales, c'est-à-dire si

$$\mathbb{P}(\{\omega_1\}) = \mathbb{P}(\{\omega_2\}) = \dots = \mathbb{P}(\{\omega_n\}) = \frac{1}{n},$$

alors la probabilité \mathbb{P} est appelée **probabilité uniforme** sur Ω . Dans ce cas

$$\forall A \in \mathcal{A} \quad \mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

On dit alors souvent que la probabilité d'un événement est «le rapport du nombre de cas favorables au nombre de cas possibles».

Attention, ce n'est pas parce que l'univers Ω est fini que la probabilité sur l'espace d'événements associé est la probabilité uniforme. Il ne faut surtout pas confondre univers fini et probabilité uniforme ! L'exemple suivant illustre cette mise en garde.

EXEMPLE On sonne à votre porte. Quelle est la probabilité que le visiteur soit un éléphant rose ? Il y a 2 cas possibles : ou bien c'est un éléphant rose ou bien ce n'est pas un éléphant rose. «Le rapport du nombre de cas favorables au nombre de cas possibles» est donc de $1/2$. On comprend mieux la raison pour laquelle les parents recommandent à leurs enfants de ne pas ouvrir la porte à n'importe qui ...

REMARQUE Dans le cas d'un univers infini, par exemple $\Omega = \mathbb{N}$, il ne peut pas y avoir équiprobabilité (autrement dit il est impossible d'associer à l'espace d'événements la probabilité uniforme) car on aurait alors

$$\mathbb{P}(\{i\}) > 0 \quad \forall i \in \mathbb{N}$$

ce qui impliquerait

$$\mathbb{P}(\Omega) = \sum_{i=1}^{+\infty} \mathbb{P}(\{i\}) = +\infty.$$

2.2 Propriétés

PROPOSITION 14 Soient A et B deux événements d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. La probabilité \mathbb{P} est une application qui vérifie les propriétés suivantes :

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$,
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

DÉMONSTRATION 1. Les événements A et A^c sont disjoints et $A \cup A^c = \Omega$. On déduit de la définition 16 que

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

ce qui implique que $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

2. Si A est un sous-ensemble de B , alors $B = A \cup_d (B \setminus A)$ et l'union est disjointe. On déduit de la définition 16 que $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$, avec $\mathbb{P}(B \setminus A) \geq 0$. Ainsi, $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$.

3. Pour tous sous-ensembles A et B de Ω , on a $A = (A \setminus B) \cup_d (A \cap B)$ où l'union est disjointe. On en déduit que $\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)$. Par ailleurs, on a aussi $B = (B \setminus A) \cup_d (B \cap A)$ (où l'union est disjointe) d'où

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

Par conséquent,

$$A \cup B = (A \setminus B) \cup_d (A \cap B) \cup_d (B \setminus A)$$

et

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

○

REMARQUE L'analogie entre les propriétés de la probabilité et celles du cardinal (voir les propositions 11 et 13 du chapitre précédent), déjà mise en exergue au début de ce paragraphe, se poursuit ici.

EXERCICE On considère le cas où Ω est un univers fini et on munit l'espace d'événements $(\Omega, \mathcal{P}(\Omega))$ de la probabilité uniforme, de sorte que

$$\forall A \in \mathcal{P}(\Omega) \quad \mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

En utilisant les propriétés du cardinal d'un ensemble, redémontrer la proposition 14 dans ce cas particulier.

On peut généraliser la propriété 3 de la proposition 14 par la formule de Poincaré.

PROPOSITION 15 (Formule de Poincaré) *Soient A_1, \dots, A_n des événements d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On a*

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) \\ &\quad + \dots + (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) \\ &\quad + \dots + (-1)^{n-1} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right). \end{aligned}$$

EXERCICE Soient A_1, A_2, A_3 trois événements d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Expliciter $\mathbb{P}(A_1 \cup A_2 \cup A_3)$ en fonction de $\mathbb{P}(A_1), \mathbb{P}(A_2), \mathbb{P}(A_3), \mathbb{P}(A_1 \cap A_2), \mathbb{P}(A_1 \cap A_3), \mathbb{P}(A_3 \cap A_2)$ et $\mathbb{P}(A_1 \cap A_2 \cap A_3)$.

Nous admettons la proposition suivante, qui n'aura dans ce cours qu'un intérêt technique, utile pour des démonstrations futures.

PROPOSITION 16 *Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé où Ω est un univers infini et $(A_n)_{n \in \mathbb{N}}$ est une suite d'événements telle que*

$$\forall n, m \in \mathbb{N}, \quad m > n, \quad A_m \subset A_n.$$

On a alors

$$\lim_{n \rightarrow +\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right).$$

2.3 Exemples

Reprendons les exemples du paragraphe 1.2 et définissons les espaces probabilisés associés à chaque expérience.

Ex₁— On choisit $\Omega = \{P, F\}$. On peut prendre $\mathcal{A} = \{\emptyset, \Omega\}$ avec $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(\Omega) = 1$, mais les seuls renseignements que ce modèle nous apporte sont : « il y a toutes les chances

d'obtenir pile ou face» ($\mathbb{P}(\Omega) = \mathbb{P}(\{P, F\}) = 1$), et «il n'y a aucune chance de n'obtenir ni pile ni face» ($\mathbb{P}(\emptyset) = 0$).

Ces informations ne sont pas très intéressantes. Il vaut donc mieux «raffiner» le modèle en «grossissant» $\mathcal{A} : \mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, \Omega, \{F\}, \{P\}\}$, avec $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\{F\}) = 1/2$, $\mathbb{P}(\{P\}) = 1/2$, modèle qui donne plus d'informations sur le résultat de l'expérience.

Ex₂— On prend $\Omega = \{1, 2, 3, 4, 5, 6\}$. On choisit le modèle le plus «complet» : $\mathcal{A} = \mathcal{P}(\Omega)$.

($\text{Card}(\mathcal{A}) = 2^6 = 64$). Si le dé n'est pas truqué, alors $\forall i = 1 \dots 6$, $\mathbb{P}(\{i\}) = 1/6$. La donnée de \mathbb{P} sur les singletons suffit à caractériser \mathbb{P} en entier. En effet

$$\mathbb{P}(\{1, 2\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) = \frac{2}{6} = \frac{1}{3},$$

$$\mathbb{P}(\{1, 2, 4, 6\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{4}{6} = \frac{2}{3},$$

etc.

Ex₃— (Lancer d'un dé jusqu'à l'obtention d'un 6.) On prend $\Omega = \mathbb{N}$, ensemble infini dénombrable et $\mathcal{A} = \mathcal{P}(\Omega)$. On définit \mathbb{P} sur les singletons :

$$\mathbb{P}(\{i\}) = \left(\frac{5}{6}\right)^{i-1} \frac{1}{6},$$

où $(5/6)^{i-1}$ est la probabilité de faire 1, 2, 3, 4 ou 5 aux $(i-1)$ premiers coups, et $1/6$ est la probabilité de faire 6 au i ème coup.

Ici, on vérifie bien que

$$\begin{aligned} \mathbb{P}(\Omega) &= \sum_{i=1}^{+\infty} \left(\frac{5}{6}\right)^{i-1} \frac{1}{6} \\ &= \frac{1}{6} \sum_{i=0}^{+\infty} \left(\frac{5}{6}\right)^i \\ &= \frac{1}{6} \frac{1}{1 - \frac{5}{6}} = 1. \end{aligned}$$

Ex₄— (Choisir «uniformément» un nombre entre 0 et 1). On prend $\Omega = [0, 1]$ ensemble infini non dénombrable et \mathcal{A} est l'ensemble des boréliens de $[0, 1]$. On pose $\mathbb{P}([a, b]) = b - a$, où $[a, b] = [a, b],]a, b], [a, b] \text{ ou }]a, b]$.

On voit sur ce modèle que $\forall c \in \mathbb{R}$, $\mathbb{P}(\{c\}) = 0$.

3 Probabilités conditionnelles

Nous nous plaçons dans ce paragraphe sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

3.1 Problématique

Considérons l'expérience consistant à lancer deux fois un dé. On définit

$$\Omega = \{(i, j) \in \mathbb{N}^2, 1 \leq i \leq 6, 1 \leq j \leq 6\}$$

et $\mathcal{A} = \mathcal{P}(\Omega)$. On s'intéresse à la somme des deux chiffres obtenus.

Soit A l'événement «le total fait 9». On a

$$A = \{(3, 6), (4, 5), (5, 4), (6, 3)\},$$

donc

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{4}{36} = \frac{1}{9}.$$

Soit B l'événement : «on obtient 3 au premier lancer». Alors

$$B = \{(3, j) \in \mathbb{N}^2, 1 \leq j \leq 6\},$$

donc

$$\mathbb{P}(B) = \frac{\text{Card}(B)}{\text{Card}(\Omega)} = \frac{6}{36} = \frac{1}{6}.$$

On se pose maintenant le problème suivant : sachant que B est réalisé, c'est-à-dire sachant que l'on obtient 3 au premier lancer, quelle est la probabilité que le total fasse 9 ? Pour répondre à cette question, il semble nécessaire de modifier notre espace probabilisé.

On définit un nouvel univers (ensemble des possibles) :

$$\Omega' = \{(3, j) \in \mathbb{N}^2, 1 \leq j \leq 6\} = \Omega \cap B$$

et on prend $\mathcal{A}' = \mathcal{P}(\Omega')$. On considère alors l'espace probabilisé $(\Omega', \mathcal{A}', \mathbb{P}')$ où \mathbb{P}' est la probabilité uniforme sur Ω' . L'événement A' «le total fait 9» dans ce nouveau modèle s'écrit

$$A' = \{(3, 6)\} = A \cap B.$$

On a

$$\mathbb{P}'(A') = \frac{\text{Card}(A')}{\text{Card}(\Omega')} = \frac{1}{6} \neq \frac{1}{9}.$$

Remarquons que l'on aurait pu, en calculant $\mathbb{P}(A \cap B)$ dans l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, s'affranchir de cette nouvelle modélisation puisque

$$\mathbb{P}'(A') = \frac{\text{Card}(A')}{\text{Card}(\Omega')} = \frac{\text{Card}(A \cap B)}{\text{Card}(\Omega \cap B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

La notion de probabilité conditionnelle est l'un des concepts fondamentaux de la théorie des probabilités. Elle intervient dans les situations suivantes.

- On cherche à calculer la probabilité de réalisation d'un événement alors qu'une partie de l'information concernant le résultat est connue et on ne souhaite pas modifier la modélisation associée à l'expérience aléatoire considérée. C'est le cas de l'exemple précédent.
- Il est parfois beaucoup plus simple de modéliser une expérience aléatoire en utilisant des probabilités conditionnelles plutôt que d'utiliser toute l'information disponible pour définir l'espace probabilisé (qui peut être difficile à préciser). L'exemple suivant est typique de cette situation. Une urne contient 10 boules blanches, 5 boules jaunes et 10 boules noires. Une boule est tirée dans l'urne au hasard et on constate qu'elle n'est pas noire. Quelle est la probabilité qu'elle soit jaune ?

3.2 Définition

DÉFINITION 17 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $A \in \mathcal{A}$, $B \in \mathcal{A}$ tel que $\mathbb{P}(B) \neq 0$. La «probabilité de A sachant B » est définie par

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

REMARQUE On peut étendre cette définition aux événements de probabilité nulle en posant $\mathbb{P}(A|B) = 0$ si $\mathbb{P}(B) = 0$.

PROPOSITION 17 L'application

$$\mathbb{P}_B : A \in \mathcal{A} \longmapsto \mathbb{P}(A|B)$$

définit une probabilité sur l'espace des événements (Ω, \mathcal{A}) . Cette probabilité est appelée **probabilité conditionnelle sachant B** .

DÉMONSTRATION On vérifie les trois points de la définition 16 :

1. Soit $A \in \mathcal{A}$, alors d'une part

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0,$$

car $\mathbb{P}(A \cap B) \geq 0$ et $\mathbb{P}(B) \geq 0$, d'autre part

$$\mathbb{P}_B(A) \leq 1,$$

car $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ (voir proposition 14).

2. $\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.

3. Si $A_1 \cap A_2 = \emptyset$, alors $(A_1 \cap B) \cap (A_2 \cap B) = \emptyset$, et on peut écrire

$$\begin{aligned} \mathbb{P}_B(A_1 \cup A_2) &= \frac{\mathbb{P}((A_1 \cup A_2) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_1 \cap B) + \mathbb{P}(A_2 \cap B)}{\mathbb{P}(B)} \\ &= \mathbb{P}_B(A_1) + \mathbb{P}_B(A_2). \end{aligned}$$

Ceci est généralisable, par récurrence, à $\bigcup_{i \in \mathbb{N}} A_i$.

○

PROPOSITION 18 Soient $A, B \in \mathcal{A}$. On a les relations suivantes :

1. $\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B)$;
2. $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) \times \mathbb{P}(B) = \mathbb{P}(B|A) \times \mathbb{P}(A)$.

DÉMONSTRATION La première relation résulte du fait que \mathbb{P}_B définit une probabilité sur l'espace des événements (Ω, \mathcal{A}) , voir la proposition 17. On peut également la démontrer directement : puisque $B = (B \cap A) \cup_d (B \cap A^c)$ avec une union disjointe, on a

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

On en déduit que

$$\frac{\mathbb{P}(B \cap A^c)}{\mathbb{P}(B)} = 1 - \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B)}$$

c'est-à-dire que

$$\mathbb{P}(A^c|B) = 1 - \mathbb{P}(A|B).$$

La seconde relation est immédiate en utilisant la définition de la probabilité conditionnelle.

○

3.3 Formule de Bayes

BAYES, Thomas (1702, Londres - 1761, Tunbridge Wells).



Théologien protestant, il s'adonna aux mathématiques sous la houlette de De Moivre. Il sera le premier, avant Laplace, à exposer le problème de la probabilité des causes : calcul de la probabilité d'un événement complexe dont on sait qu'un de ses composants (causes) s'est produit. Il fut membre de la Royal Society.

DÉFINITION 18 *On dit que les sous-ensembles A_1, \dots, A_n de l'ensemble Ω forment une partition de Ω si*

1. $\forall (i, j) \in \{1, \dots, n\}^2, \quad i \neq j, \quad A_i \cap A_j = \emptyset,$
2. $\bigcup_{i=1}^n A_i = \Omega.$

THÉORÈME 1 Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soient A_1, \dots, A_n une partition de Ω . On a les relations suivantes :

$$\forall B \in \mathcal{A} \quad \mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B|A_i) \times \mathbb{P}(A_i) \quad (\text{formule des probabilités totales}) ;$$

$$\forall B \in \mathcal{A} \quad \mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k) \times \mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i) \times \mathbb{P}(A_i)} \quad (\text{formule de Bayes}) .$$

DÉMONSTRATION Soient $B \in \mathcal{A}$ et A_1, \dots, A_n une partition de Ω . On a

$$B = B \cap \Omega = B \cap \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n (B \cap A_i).$$

Remarquons que l'union est disjointe puisque pour tout $i, j \in \{1, \dots, n\}$ avec $i \neq j$ on a par associativité de l'intersection

$$(B \cap A_i) \cap (B \cap A_j) = B \cap (\underbrace{A_i \cap A_j}_{= \emptyset}) = \emptyset.$$

On en déduit que

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

D'autre part, pour tout entier k avec $k \in \{1, \dots, n\}$, on a

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

○

REMARQUE La formule de Bayes est souvent utilisée dans le cas $n = 2$ avec $A_2 = A_1^c$, comme l'illustre l'exemple suivant.

EXEMPLE Un étudiant répond à une question à choix multiple, et il doit choisir entre 3 réponses. On suppose que l'étudiant a une chance sur deux de connaître la réponse. Dans le cas où il ne la connaît pas, il coche une réponse au hasard. Sachant que l'étudiant répond juste, quelle est la probabilité qu'il connaisse la réponse ?

On considère les événements $C = \text{«l'étudiant connaît la réponse»}$ et $J = \text{«l'étudiant répond juste»}$. On a alors $\mathbb{P}(C) = 1/2$, $\mathbb{P}(J|C) = 1$, $\mathbb{P}(J|C^c) = 1/3$ et on veut calculer $\mathbb{P}(C|J)$. Si l'on applique la formule de Bayes, on obtient

$$\mathbb{P}(C|J) = \frac{\mathbb{P}(J|C) \times \mathbb{P}(C)}{\mathbb{P}(J|C) \times \mathbb{P}(C) + \mathbb{P}(J|C^c) \times \mathbb{P}(C^c)} = \frac{1 \times \frac{1}{2}}{1 \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{2}} = \frac{3}{4}.$$

Il y a donc 3 chances sur 4 que l'étudiant n'ait pas répondu juste par hasard...

4 Indépendance

DÉFINITION 19 Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et A, B deux éléments de \mathcal{A} . Les événements A et B sont dits indépendants si

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

PROPOSITION 19 Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et A, B deux éléments de \mathcal{A} tel que $\mathbb{P}(B) \neq 0$. Les événements A et B sont dits indépendants si et seulement si

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

DÉMONSTRATION - On a $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Si l'on suppose A et B indépendants on

$$\text{obtient } \mathbb{P}(A|B) = \frac{\mathbb{P}(A) \times \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A).$$

- Réciproquement, si $\mathbb{P}(A|B) = \mathbb{P}(A)$ alors $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$ et par conséquent

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

autrement dit les événements A et B sont indépendants. ○

La notion d'événements indépendants est l'une des premières difficulté du calcul des probabilités. À la question : «qu'est-ce que 2 événements indépendants?», on répond communément «ce sont deux événements qui n'ont pas d'influence l'un sur l'autre». Or cette réponse qui dans le langage courant, paraît claire et satisfaisante, est en réalité imprécise. L'indépendance des événements A et B signifie qu'avoir observé la réalisation de B ne modifie pas la probabilité d'une éventuelle réalisation de A . On ne peut donc parler d'indépendance sans avoir au préalable fixé une probabilité sur l'espace d'événements considéré. Ainsi, considérons l'expérience aléatoire consistant à lancer deux fois un dé et prenons pour univers $\Omega = \Gamma \times \Gamma$ où $\Gamma = \{1, 2, 3, 4, 5, 6\}$ et pour tribu sur Ω l'ensemble $\mathcal{A} = \mathcal{P}(\Omega)$. Intéressons-nous aux événements A : «on obtient 1 ou 2 à chacun des 2 lancers» et B : «la somme des deux résultats est paire». Les événements A et B sont-ils indépendants? Cette question n'a pas de sens! On ne peut pas trancher concernant l'indépendance des événements A et B sans avoir auparavant donné une probabilité de réalisation aux événements A et B et également à l'événement $A \cap B$. Si le dé n'est pas truqué, on munit naturellement l'espace d'événements $(\Omega, \mathcal{P}(\Omega))$ de la probabilité uniforme. On peut vérifier que $\mathbb{P}(A) = 1/9$, $\mathbb{P}(B) = 1/2$ et que $\mathbb{P}(A \cap B) = 1/18$. Dans ce cas, les événements A et B sont effectivement indépendants.

Supposons maintenant que le dé est pipé de sorte que le 1 ait une chance sur deux de sortir, et que chacune des cinq autres faces ait une chance sur dix. On peut vérifier qu'on a alors $\mathbb{P}(A) = 9/25$, $\mathbb{P}(B) = 29/50$ et que $\mathbb{P}(A \cap B) = 13/50$. Dans ce cas, les événements A et B ne sont pas indépendants.

L'indépendance des événements est donc liée à la probabilité considérée, elle même définie par la qualité du dé.

REMARQUES 1. De façon générale, il faut bien retenir que la définition probabiliste de l'indépendance est plus «large» que la notion intuitive.

2. Ne pas confondre événements indépendants et événements incompatibles (ou disjoints). Deux événements A et B incompatibles (ou disjoints) de probabilités non nulles ne sont pas indépendants, car $\mathbb{P}(A \cap B) = \mathbb{P}(\emptyset) = 0$, et $\mathbb{P}(A) \neq 0$, $\mathbb{P}(B) \neq 0$, d'où $\mathbb{P}(A) \times \mathbb{P}(B) \neq 0$. En effet, on a $A \cap B = \emptyset$ si et seulement si $A \subset B^c$, donc « A implique non- B », il y a donc une forte relation de dépendance !

PROPOSITION 20 *Si A et B sont indépendants, alors il en est de même pour A et B^c , pour A^c et B , pour A^c et B^c .*

DÉMONSTRATION Il suffit de montrer que A et B^c sont indépendants, on en déduira les autres par passages successifs au complémentaire. D'une part, on a

$$(A \cap B^c) \cup (A \cap B) = A \cap (B \cup B^c) = A,$$

et $(A \cap B^c) \cap (A \cap B) = \emptyset$. Donc

$$\begin{aligned} \mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A) \times \mathbb{P}(B) \quad (\text{par indépendance de } A \text{ et } B) \\ &= \mathbb{P}(A) \times (1 - \mathbb{P}(B)) = \mathbb{P}(A) \times \mathbb{P}(B^c). \end{aligned}$$

○

DÉFINITION 20 *Les événements $A_1, \dots, A_n \in \mathcal{A}$ sont dits **mutuellement indépendants** si $\forall k \in \{1, \dots, n\}$ et $\forall (i_1, \dots, i_k) \in \mathbb{N}^k$ tel que $1 \leq i_1 < i_2 < \dots < i_k \leq n$, on a*

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

EXEMPLE Pour trois événements A , B et C , cela se traduit par

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$, $\mathbb{P}(A \cap C) = \mathbb{P}(A) \times \mathbb{P}(C)$, $\mathbb{P}(B \cap C) = \mathbb{P}(B) \times \mathbb{P}(C)$,
2. $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$.

REMARQUE Des événements peuvent être 2 à 2 indépendants sans être mutuellement indépendants. Considérons par exemple l'expérience consistant à lancer un dé deux fois et les événements A : «le premier jet a donné un numéro pair», B : «le second jet a donné un numéro impair» et C : «la somme des deux numéros est un chiffre pair». On vérifie que les trois événements sont 2 à 2 indépendants (en exercice) mais pas mutuellement indépendants (on vérifie que les trois événements sont incompatibles $\mathbb{P}(A \cap B \cap C) = 0$ mais $\mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C) \neq 0$).

L'indépendance est parfois difficile à admettre. Chacun sait que, malgré tout ce que l'on peut leur expliquer, les joueurs qui ont vu au casino le rouge sortir 10 fois de suite, penseront que la probabilité de voir le noir sortir à la 11^e partie est accrue. Or il n'en n'est rien. Les parties étant indépendantes, la probabilité de voir le noir sortir à la 11^e partie est toujours de $1/2$ et est égale à la probabilité de voir le rouge sortir. Le résultat d'un million de parties n'apporterait aucune information supplémentaire sur la partie suivante !

Chapitre 3

Variables aléatoires réelles

1 Définitions

1.1 Variable aléatoire

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé modélisant une expérience aléatoire. Il arrive souvent que l'on s'intéresse plus particulièrement à des «conséquences» des résultats de l'expérience, c'est-à-dire à une certaine fonction X des résultats de l'expérience. Par exemple, on lance 2 dés et on s'intéresse à la somme des valeurs des 2 dés. X est alors une application qui au résultat du lancer associe la somme des valeurs des 2 dés. Autrement dit, on s'intéresse à l'ensemble $\Omega' = X(\Omega)$ des conséquences possibles des résultats de l'expérience et on veut calculer la probabilité de certains événements de la tribu \mathcal{A}' associée à Ω' . Pour obtenir cette nouvelle probabilité sur (Ω', \mathcal{A}') on se sert des propriétés de l'application X .

DÉFINITION 21 Soient (Ω, \mathcal{A}) et (Ω', \mathcal{A}') deux espaces d'événements. On appelle variable aléatoire de l'espace d'événements (Ω, \mathcal{A}) vers (Ω', \mathcal{A}') une application

$$X : \Omega \longrightarrow \Omega'$$

telle que

$$\forall A' \in \mathcal{A}', \quad X^{-1}(A') \in \mathcal{A},$$

où $X^{-1}(A') = \{\omega \in \Omega, \ X(\omega) \in A'\}$.

REMARQUE Le terme «variable aléatoire» constitue un abus terminologique puisque l'on définit une application.

EXEMPLE On considère un jeu de lancer de dé où le joueur qui a misé une certaine somme double sa mise si le résultat est 6, perd tout sinon. L'univers associé à cette expérience est $\Omega = \{1, 2, 3, 4, 5, 6\}$. On considère la fonction $X : \Omega \mapsto \{0, 1\}$ définie par : $X(\omega) = 1$ si $\omega = 6$, $X(\omega) = 0$ sinon. Cette fonction met en exergue la seule issue qui intéresse le joueur. On a donc $\Omega' = \{0, 1\}$. On s'intéresse à l'événement «le joueur gagne», c'est-à-dire à $A' = \{1\}$. On doit donc poser $\mathcal{A}' = \mathcal{P}(\Omega')$ de façon à ce que $A' \in \mathcal{A}' = \{\Omega', \emptyset, \{1\}, \{0\}\}$.

Pour que X soit une variable aléatoire, il faut que

$$\begin{aligned} X^{-1}(\Omega') &= \Omega \in \mathcal{A}, \\ X^{-1}(\emptyset) &= \emptyset \in \mathcal{A}, \\ X^{-1}(\{1\}) &= \{6\} \in \mathcal{A}, \\ X^{-1}(\{0\}) &= \{1, 2, 3, 4, 5\} \in \mathcal{A}. \end{aligned}$$

On voit donc par exemple que si $\mathcal{A} = \{\emptyset, \Omega\}$, X n'est pas une variable aléatoire, et de fait, l'information «contenue» dans \mathcal{A} est «le dé est jeté» (Ω) ou «le dé n'est pas jeté» (\emptyset), et ne suffit pas pour déterminer si le joueur a gagné ou non. \mathcal{A} doit au minimum contenir $\{6\}$ et $\{6\}^c$.

On s'intéresse dans ce chapitre à définir de nouveaux outils qui vont permettre de calculer la probabilité que le joueur gagne, c'est-à-dire à calculer la probabilité que $X(\omega) = 1$.

- REMARQUES**
1. En pratique, on s'intéresse souvent au cas où $\Omega' \subset \mathbb{R}$, par exemple $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. La variable aléatoire de l'espace d'événements (Ω, \mathcal{A}) vers (Ω', \mathcal{A}') est alors qualifiée de variable aléatoire réelle (notée v.a.r.).
 2. Un espace d'événements (Ω, \mathcal{A}) est amené à être muni d'une certaine probabilité \mathbb{P} . On désignera par variable aléatoire réelle sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ une variable aléatoire réelle sur l'espace d'événements (Ω, \mathcal{A}) muni de la probabilité \mathbb{P} .

Nous admettons qu'une variable aléatoire réelle est caractérisée de la manière suivante :

PROPOSITION 21 *Une application $X : \Omega \longrightarrow \mathbb{R}$ est une variable aléatoire réelle sur l'espace d'événements (Ω, \mathcal{A}) si*

$$\forall x \in \mathbb{R}, \quad X^{-1}(]-\infty, x]) \in \mathcal{A}.$$

REMARQUE La composée à gauche $g(X)$ d'une v.a.r. X par une fonction «raisonnable¹» g est elle aussi une v.a.r. Par exemple, si X est une v.a.r. sur $(\Omega, \mathcal{A}, \mathbb{P})$, et si g est une fonction continue par morceaux quelconque de \mathbb{R} dans \mathbb{R} , alors $g(X)$ est encore une v.a.r. sur $(\Omega, \mathcal{A}, \mathbb{P})$.

1.2 Fonction de répartition

DÉFINITION 22 *Soit X une v.a.r. sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. L'application*

$$\begin{aligned} F_X : \quad \mathbb{R} &\longrightarrow [0, 1] \\ x &\longmapsto \mathbb{P}(X^{-1}(]-\infty, x])) \end{aligned}$$

est appelée fonction de répartition de X .

¹Rigoureusement, on exigerait seulement que cette fonction soit «mesurable», notion que nous ne définissons pas ici, voir par exemple [Gapaillard].

Rappelons que $\mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(\{\omega \in \Omega, X(\omega) \leq x\})$ ce que, par abus de notation, nous écrirons le plus souvent $\mathbb{P}(X \leq x)$.

PROPOSITION 22 Soit X une v.a.r. sur $(\Omega, \mathcal{A}, \mathbb{P})$. La fonction de répartition associée à X vérifie les propriétés suivantes :

1. $\lim_{x \rightarrow -\infty} F_X(x) = 0$;
2. $\lim_{x \rightarrow +\infty} F_X(x) = 1$;
3. F_X est croissante : $\forall (x, y) \in \mathbb{R}^2 \quad (x \leq y \implies F_X(x) \leq F_X(y))$;
4. F_X est continue à droite.

DÉMONSTRATION 1. On a

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{x \rightarrow -\infty} \mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(X^{-1}(\emptyset)) = \mathbb{P}(\emptyset) = 0.$$

2. On a $\lim_{x \rightarrow +\infty} F_X(x) = \lim_{x \rightarrow +\infty} \mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1$.
3. Si $x \leq y$ alors $]-\infty, x] \subset]-\infty, y]$ et donc $X^{-1}(]-\infty, x]) \subset X^{-1}(]-\infty, y])$. On en déduit que

$$F_X(x) = \mathbb{P}(X^{-1}(]-\infty, x])) \leq \mathbb{P}(X^{-1}(]-\infty, y])) = F_X(y).$$

4. Soit $(u_n)_{n \in \mathbb{N}}$ une suite décroissante de \mathbb{R}^+ tendant vers 0. Montrons² que

$$\lim_{n \rightarrow +\infty} F_X(x + u_n) = F_X(x).$$

On a

$$\begin{aligned} \lim_{n \rightarrow +\infty} F_X(x + u_n) &= \lim_{n \rightarrow +\infty} \mathbb{P}(X \leq x + u_n) \\ &= \mathbb{P}(\bigcap_{n \in \mathbb{N}} [X \leq x + u_n]) \\ &= \mathbb{P}(\{\omega \in \Omega, \forall n \in \mathbb{N}, X(\omega) \leq x + u_n\}) \\ &= \mathbb{P}(\{\omega \in \Omega, X(\omega) \leq x\}) \\ &= F_X(x). \end{aligned}$$

Le passage de la première à la deuxième ligne provient du fait que les événements $A_n = [X \leq x + u_n]$ forment une suite d'événements décroissants puisque la suite $(u_n)_{n \in \mathbb{N}}$ est décroissante, voir la proposition 16 page 18. \circ

²Nous utilisons ici d'une part l'équivalence sur \mathbb{R} entre la continuité et la continuité séquentielle, et d'autre part que de toute suite positive convergeant vers 0 on peut extraire une sous-suite décroissante, voir par exemple [Balac-Sturm].

1.3 Variables aléatoires discrètes et variables aléatoires continues

DÉFINITION 23 Une v.a.r. X est dite discrète si elle ne prend qu'un nombre fini ou dénombrable de valeurs dans \mathbb{R} , c'est-à-dire si $X(\Omega) = \{x_j \in \mathbb{R}, j \in J\}$ où $J \subset \mathbb{N}$.
Dans ce cas la fonction

$$\begin{aligned} p : J &\longrightarrow [0, 1] \\ j &\longmapsto p_j \end{aligned}$$

où $p_j = \mathbb{P}(X = x_j)$ est appelée fonction de masse de la v.a.r. X .

REMARQUES 1. Si X ne prend qu'un nombre fini de valeurs, on considère par convention que $p_j = 0$ pour tout $j \in \mathbb{N} \setminus J$.

2. D'après les points 1 et 2 de la définition 16 du chapitre 2, on a

(a) pour tout $j \in J$ $p_j \geq 0$;

(b) $\sum_{j \in J} p_j = 1$.

PROPOSITION 23 Si X est une v.a.r. discrète prenant les valeurs $\{x_i \in \mathbb{R}, i \in I\}$ avec $I \subset \mathbb{N}$, alors sa fonction de répartition F_X est constante par morceaux et a pour points de discontinuité $\{x_i \in \mathbb{R}, i \in I\}$.

DÉMONSTRATION On suppose que les x_i sont rangés dans l'ordre croissant de leur indice : $x_0 < x_1 < x_2 \dots$ On vérifie alors que :

$$\begin{aligned} \text{si } x < x_0 &\quad \text{alors } F_X(x) = \mathbb{P}(X \leq x) = 0 \\ \text{si } x \in [x_0, x_1[&\quad \text{alors } F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X = x_0) = p_0 \\ \text{si } x \in [x_1, x_2[&\quad \text{alors } F_X(x) = \mathbb{P}(X = x_0 \text{ ou } x_1) \\ &\quad = \mathbb{P}([X = x_0] \cup_d [X = x_1]) = p_0 + p_1. \end{aligned}$$

De manière générale, on vérifie par récurrence que pour tout $k \in \mathbb{N}$,

$$x \in [x_k, x_{k+1}[\implies F_X(x) = \sum_{i=0}^k p_i.$$

La fonction de répartition F_X est donc constante par morceaux et admet pour points de discontinuité $\{x_i, i \in I\}$. ○

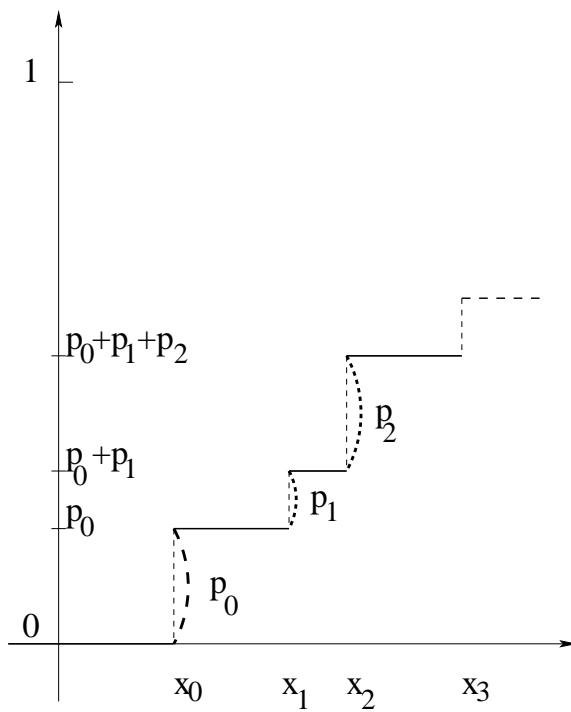


FIG. 3.1 – Allure de la représentation graphique de la fonction de répartition d'une variable aléatoire discrète.

DÉFINITION 24 Une v.a.r. X est dite absolument continue³ si il existe une fonction de \mathbb{R} dans \mathbb{R} , notée f_X , telle que la fonction de répartition F_X de la v.a.r. X admette la représentation intégrale suivante :

$$F_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Ceci est équivalent à dire que F_X est dérivable sur \mathbb{R} , de dérivée f_X . La fonction f_X est appelée **densité** de X . On parle aussi de v.a.r. à densité pour désigner une v.a.r. absolument continue.

REMARQUE Une v.a.r. peut être ni discrète ni absolument continue (F_X est cependant «toujours» dérivable par morceaux), mais nous ne considérerons généralement que des exemples de v.a.r. discrètes ou absolument continues. Nous appellerons souvent ces dernières (abusivement) variables aléatoires continues.

1.4 Loi d'une variable aléatoire

³Par rapport à la mesure de Lebesgue.

PROPOSITION 24 Soient (Ω, \mathcal{A}) et (Ω', \mathcal{A}') deux espaces d'événements. Soient X une variable aléatoire de (Ω, \mathcal{A}) vers (Ω', \mathcal{A}') et \mathbb{P} une probabilité sur (Ω, \mathcal{A}) .

L'application

$$\begin{aligned}\mathbb{P}_X : \mathcal{A}' &\longrightarrow [0, 1] \\ A' &\longmapsto \mathbb{P}(X^{-1}(A'))\end{aligned}$$

définit une probabilité sur l'espace d'événements (Ω', \mathcal{A}') . Cette probabilité est appelée loi de la variable aléatoire X .

DÉMONSTRATION Montrons que \mathbb{P}_X vérifie bien les conditions pour être une probabilité (voir la définition 16 page 15).

- \mathbb{P}_X est bien définie car le fait que X soit une v.a. implique que $X^{-1}(A') \in \mathcal{A}$, et l'image de \mathbb{P}_X est incluse dans celle de \mathbb{P} , c'est-à-dire dans $[0, 1]$.
- $\mathbb{P}_X(\Omega') = \mathbb{P}(X^{-1}(\Omega')) = \mathbb{P}(\Omega) = 1$.
- Soit $(A'_i)_{i \in \mathbb{N}}$ une suite d'événements disjoints de (Ω', \mathcal{A}') , alors

$$\begin{aligned}\mathbb{P}_X\left(\bigcup_{i \in \mathbb{N}} A'_i\right) &= \mathbb{P}\left(X^{-1}\left(\bigcup_{i \in \mathbb{N}} A'_i\right)\right) \\ &= \mathbb{P}\left(\{\omega \in \Omega, X(\omega) \in \bigcup_{i \in \mathbb{N}} A'_i\}\right) \\ &= \mathbb{P}\left(\{\omega \in \Omega, \exists i \in \mathbb{N}, X(\omega) \in A'_i\}\right) \\ &= \mathbb{P}\left(\bigcup_{i \in \mathbb{N}} \{\omega \in \Omega, X(\omega) \in A'_i\}\right) \\ &= \mathbb{P}\left(\bigcup_{i \in \mathbb{N}} X^{-1}(A'_i)\right) \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}(X^{-1}(A'_i)) \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}_X(A'_i).\end{aligned}$$

Le passage de la quatrième à la cinquième ligne résulte du fait que les $X^{-1}(A'_i)$ sont disjoints (car les A'_i le sont et X est une application).

○

Caractérisation de \mathbb{P}_X

Soit X une variable aléatoire réelle sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et soit A un borélien de \mathbb{R} .

- Si X est une v.a.r. discrète, alors

$$\mathbb{P}_X(A) = \sum_{j \in \mathbb{N}} \mathbb{P}(X = x_j) \times \mathbb{I}_A(x_j) = \sum_{j \in \mathbb{N}} p_j \times \mathbb{I}_A(x_j),$$

où l'on rappelle que $\mathbb{I}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$ dénote la fonction indicatrice de l'ensemble A .

- Si X est une v.a.r. continue, alors

$$\mathbb{P}_X(A) = \int_A f_X(t) dt = \int_{\mathbb{R}} f_X(t) \times \mathbb{I}_A(t) dt.$$

La loi d'une variable aléatoire est donc caractérisée par la donnée de

1. la fonction de masse dans le cas d'une v.a.r. discrète ;
2. la fonction de densité dans le cas d'une v.a.r. absolument continue.

REMARQUES 1. À toute v.a.r. discrète est associée une fonction de masse $p = (p_i)_{i \in \mathbb{N}}$ sur \mathbb{N} telle que

$$(i) \forall i \in \mathbb{N}, \quad p_i \in [0, 1];$$

$$(ii) \sum_{i \in \mathbb{N}} p_i = 1.$$

2. À toute v.a.r. continue est associée une fonction de densité f_X telle que

$$(iii) \forall t \in \mathbb{R}, \quad f_X(t) \in \mathbb{R}^+;$$

$$(iv) \int_{\mathbb{R}} f(t) dt = 1.$$

3. Réciproquement, toute fonction $p = (p_i)_{i \in \mathbb{N}}$ vérifiant (i) et (ii) est la fonction de masse d'une v.a.r. discrète, et toute fonction positive dont l'intégrale sur \mathbb{R} est égale à 1 (c'est-à-dire vérifiant (iii) et (iv)) est la densité d'une v.a.r. continue.

2 Exemples de lois

2.1 Variables aléatoires discrètes

Loi de Bernoulli $\mathcal{B}(1, p)$, $p \in]0, 1[$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi de Bernoulli de paramètre p (on note $X \sim \mathcal{B}(1, p)$) si elle est à valeurs dans $\{0, 1\}$ avec

$$\mathbb{P}(X = 1) = p \quad \text{et} \quad \mathbb{P}(X = 0) = 1 - p.$$

Cette loi modélise l'issue d'une expérience où l'on ne s'intéresse qu'au «succès» ou à l'«échec» de l'expérience.

Loi binomiale $\mathcal{B}(n, p)$, $n \in \mathbb{N}^*$, $p \in]0, 1[$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi de binomiale de paramètre (n, p) (on note $X \sim \mathcal{B}(n, p)$) si elle est à valeurs dans $\{0, 1, \dots, n\}$ avec

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

Cette loi modélise une succession de «succès» et d'«échecs», p étant la probabilité du succès. Notons que la loi de Bernoulli est une loi binomiale particulière ($n = 1$).

Loi uniforme sur $\{1, \dots, N\}$ $\mathcal{U}_N, N \in \mathbb{N}^*$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi uniforme sur $\{1, \dots, N\}$ (on note $X \sim \mathcal{U}_N$) si elle est à valeurs dans $\{1, \dots, N\}$ avec

$$\mathbb{P}(X = k) = 1/N \quad \forall k \in \{1, \dots, N\}.$$

Cette loi modélise l'issue d'une expérience où les résultats sont équiprobables.

Loi géométrique $\mathcal{G}(p), \quad p \in]0, 1[$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi géométrique de paramètre p (on note $X \sim \mathcal{G}(p)$) si elle est à valeurs dans \mathbb{N}^* avec

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

Cette loi modélise une série d'«échecs» suivie du premier «succès».

Loi de Poisson $\mathcal{P}(\lambda), \quad \lambda > 0$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi de Poisson de paramètre λ (on note $X \sim \mathcal{P}(\lambda)$) si elle est à valeurs dans \mathbb{N} avec

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

C'est une loi discrète fréquemment utilisée en modélisation, en particulier pour les files d'attente.

Poisson, Siméon Denis (1781, Pithiviers - 1840, Sceaux).



Brillant polytechnicien, élève de Fourier et de Laplace, astronome et physicien, il occupa de nombreux et importants postes d'enseignement et fut élevé à la dignité de pair de France par Louis-Philippe en 1837. On le connaît bien sûr pour sa célèbre loi de probabilités (Théorie du calcul des probabilités, 1838), mais ses travaux portent principalement en électricité, magnétisme, mécanique et mouvements vibratoires (théorie de la chaleur, théorie des ondes) où, introduisant de nombreux concepts mathématiques liés aux équations de Laplace, il apparaît comme le bâtisseur de la physique mathématique moderne (étude, au moyen de la seule analyse mathématique, du comportement d'un phénomène).

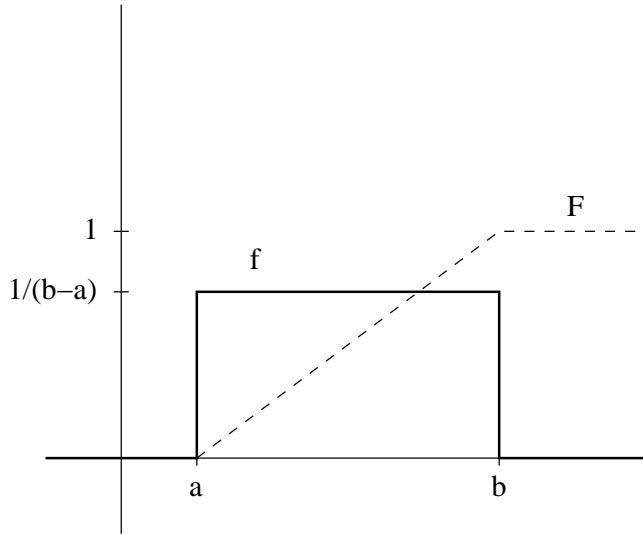


FIG. 3.2 – Représentations graphiques de la densité f et de la fonction de répartition F d'une loi uniforme sur $[a, b]$.

2.2 Variables aléatoires continues

Loi uniforme $\mathcal{U}([a, b])$, $a < b$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi uniforme sur $[a, b]$ (on note $X \sim \mathcal{U}([a, b])$) si elle est à valeurs dans $[a, b]$ et a pour densité la fonction

$$f_X(t) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(t).$$

On vérifie que la fonction de répartition d'une v.a.r. X de loi uniforme est

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b] \\ 1 & \text{si } x \geq b \end{cases}$$

Loi normale $\mathcal{N}(m, \sigma^2)$, $m \in \mathbb{R}$, $\sigma > 0$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi normale de paramètres (m, σ^2) (on note $X \sim \mathcal{N}(m, \sigma^2)$) si elle est à valeurs dans \mathbb{R} et a pour densité la fonction

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right).$$

REMARQUE La fonction de répartition de la loi normale

$$F_X : x \in \mathbb{R} \longmapsto \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) dt$$

n'est pas exprimable à l'aide des fonctions usuelles (puissances, trigonométriques, hyperboliques et leurs fonctions réciproques). Les valeurs numériques prises par la fonction de répartition de la loi normale sont tabulées, voir le tableau page 96.

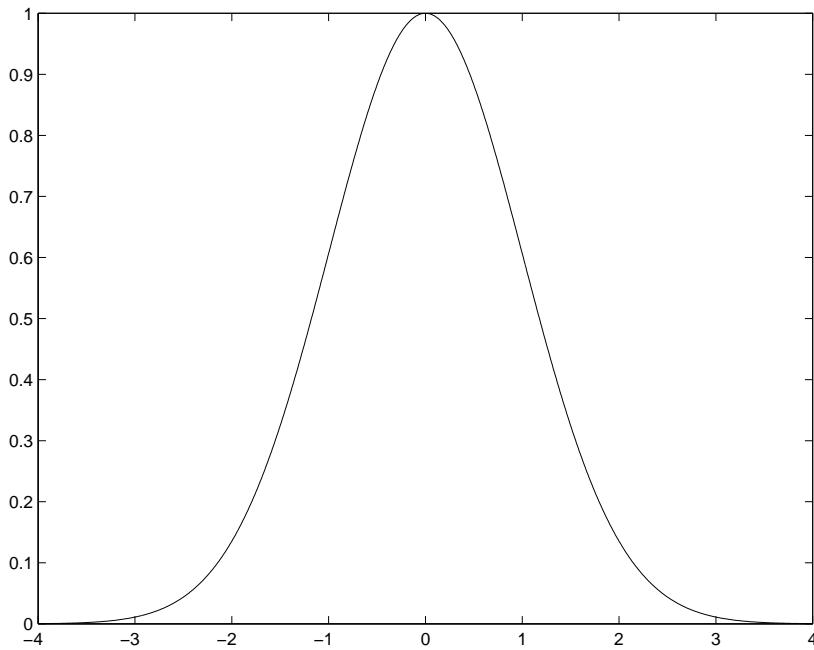


FIG. 3.3 – Représentation graphique de la densité de la loi normale pour $m = 0$ et $\sigma = 1$. Cette courbe est appelée courbe de Gauss.

Loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi exponentielle de paramètre λ (on note $X \sim \mathcal{E}(\lambda)$) si elle est à valeurs dans \mathbb{R}^+ avec pour densité la fonction

$$f_X(t) = \lambda e^{-\lambda t} \mathbb{I}_{\mathbb{R}^+}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t \geq 0 \\ 0 & \text{sinon} \end{cases}.$$

Loi de Cauchy

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi de Cauchy si elle est à valeurs dans \mathbb{R} avec pour densité la fonction

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

CAUCHY, Augustin-Louis (1789, Paris - 1857, Sceaux).



Augustin-Louis Cauchy commence sa carrière comme ingénieur militaire. En 1816, il obtient un poste de professeur à la Faculté des Sciences de Paris et à l'École Polytechnique et entre à l'Académie des Sciences. L'œuvre de Cauchy est considérable, surtout en analyse où il a su donner le cadre rigoureux nécessaire à son développement. Il introduit une notion précise de continuité et élabore une définition rigoureuse de l'intégrale. Son travail concerne tous les domaines des mathématiques, en particulier les équations différentielles, la théorie des groupes et l'algèbre linéaire.

Loi Gamma $\Gamma(a, b)$, $a > 0$, $b > 0$

La v.a.r. X sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ suit une loi Gamma de paramètres (a, b) (on note $X \sim \Gamma(a, b)$) si elle est à valeurs dans \mathbb{R}^+ avec pour densité la fonction

$$f_X(t) = \frac{1}{\Gamma(a)} b^a t^{a-1} e^{-bt} \mathbb{I}_{\mathbb{R}^+}(t)$$

où

$$\Gamma(a) = \int_0^{+\infty} u^{a-1} e^{-u} du, \quad \forall a > 0.$$

REMARQUE La fonction «Gamma» prolonge la fonction «factorielle» sur l'ensemble des réels au sens où $\forall n \in \mathbb{N}$, $\Gamma(n+1) = n!$ et $\forall a > 0$, $\Gamma(a+1) = a\Gamma(a)$. De plus, on a $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

3 Moments d'une variable aléatoire

3.1 Espérance

DÉFINITION 25 Soit X une v.a.r. sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.

1. Si X est une v.a.r. discrète et $X(\Omega) = \{x_j \in \mathbb{R}, j \in J\}$ où $J \subset \mathbb{N}$, on appelle espérance de X et on note $\mathbb{E}(X)$, la moyenne des valeurs prises par X pondérées par leurs probabilités de réalisation, autrement dit, lorsque cette quantité existe,

$$\mathbb{E}(X) = \sum_{j \in J} x_j \times \mathbb{P}(X = x_j).$$

2. Si X est une v.a.r. continue de densité f_X , on appelle espérance de X et on note $\mathbb{E}(X)$, lorsqu'elle existe, la quantité

$$\mathbb{E}(X) = \int_{\mathbb{R}} x \times f_X(x) \, dx.$$

REMARQUE Une v.a.r. n'admet pas nécessairement une espérance comme le montrent les exemples suivants.

1. Considérons une v.a.r. absolument continue X suivant une loi de Cauchy de densité

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

On a bien une densité de probabilité puisque

$$\int_{\mathbb{R}} \frac{1}{\pi(1+t^2)} dt = \frac{1}{\pi} \left[\arctan t \right]_{-\infty}^{+\infty} = \frac{1}{\pi} \left(\frac{\pi}{2} + \frac{\pi}{2} \right) = 1$$

mais la fonction

$$t \in \mathbb{R} \longmapsto t f_X(t) = \frac{t}{\pi(1+t^2)}$$

n'a pas une intégrale généralisée convergente sur \mathbb{R} (voir [Balac-Sturm]) donc la variable aléatoire X n'a pas d'espérance.

2. La série numérique $\sum_{k \in \mathbb{N}^*} \frac{1}{k^2}$ converge. Notons S sa somme (on peut démontrer que $S = \pi^2/6$). Considérons une v.a.r. discrète Y telle que $Y(\Omega) = \mathbb{N}^*$ dont la fonction de masse est

$$p : i \in \mathbb{N}^* \longmapsto \frac{1}{i^2 S}.$$

(On vérifiera à titre d'exercice que la fonction p définit bien une fonction de masse.) Cette v.a.r. n'a pas d'espérance puisque $i \times \mathbb{P}(Y = i) = \frac{1}{i S}$ et que la série de terme général $1/i$ diverge.

Plus généralement, on a le résultat suivant pour l'espérance d'une v.a.r. $g(X)$.

PROPOSITION 25 1. Soit X une v.a.r. discrète sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ avec $X(\Omega) = \{x_j \in \mathbb{R}, j \in J\}$ où $J \subset \mathbb{N}$, et $g : X(\Omega) \rightarrow \mathbb{R}$ une application. Lorsque cette quantité existe, l'espérance de la v.a.r. $g(X)$ est définie par

$$\mathbb{E}(g(X)) = \sum_{j \in J} g(x_j) \times \mathbb{P}(X = x_j).$$

2. Soit X une v.a.r. absolument continue sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ de densité f_X et $g : \mathbb{R} \rightarrow \mathbb{R}$ une application continue par morceaux. Lorsque cette quantité existe, l'espérance de la v.a.r. $g(X)$ est définie par

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(t) \times f_X(t) dt.$$

REMARQUE Comme nous l'avons remarqué à la page 28, l'hypothèse de continuité par morceaux de g est plus restrictive que nécessaire. Il suffit que la fonction soit *mesurable*. Cependant, les notions de la théorie de la mesure n'entrent pas dans le cadre de ce cours. Nous nous contenterons donc en général de l'hypothèse de continuité par morceaux, qui

se trouve être suffisante dans la grande majorité des cas qui nous intéressent. L'existence de la quantité $\mathbb{E}(g(X))$, dans la théorie de la mesure, est garantie par l'*intégrabilité* de $g(X)$, c'est-à-dire par la condition $\int_{\Omega} |g(X(\omega))| \, d\omega < +\infty$.

EXEMPLE Espérance de gain.

Un joueur mise et lance un dé. Il double sa mise si le dé fait 5 ou 6, la récupère s'il fait 4, et la perd s'il fait 1, 2 ou 3. Quelle est l'espérance de gain du joueur ?

On pose $g : \{1, 2, 3, 4, 5, 6\} \rightarrow \mathbb{R}$, avec $g(1) = g(2) = g(3) = -1$, $g(4) = 0$ et $g(5) = g(6) = 1$. D'autre part, on considère l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, où $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{A} = \mathcal{P}(\Omega)$ et \mathbb{P} est la probabilité standard associée au problème, c'est-à-dire l'équiprobabilité. Sur cet espace probabilisé, considérons la v.a.r. X représentant le résultat du lancer. X est donc l'application de Ω dans \mathbb{R} qui à i associe i . L'espérance de gain est

$$\mathbb{E}(g(X)) = \sum_{i=1}^6 g(i) \times \mathbb{P}(X = i) = \frac{1}{6}(-1 - 1 - 1 + 0 + 1 + 1) = -\frac{1}{6}.$$

L'espérance du gain est donc de $-1/6$ fois la mise. Un joueur prudent évitera donc ce jeu !

REMARQUE L'espérance mathématique représente un gain moyen. Ce n'est pas nécessairement le plus probable et il se peut même que le gain ne soit jamais égal à l'espérance. Ainsi, si l'on considère le jeu consistant à lancer un dé 10 fois et où l'on gagne 6 bonbons chaque fois que l'on obtient 1, on a une espérance de gain qui vaut 10 mais on ne pourra jamais gagner 10 bonbons. On pourra en gagner 0, 6, 12, 18, ..., 60 mais pas 10.

En anticipant sur la suite du cours, on peut dire que si l'on jouait un très grand nombre de parties successives de 10 lancers, la moyenne statistique des gains serait de 10 bonbons. C'est en ce sens que 10 bonbons constitue la valeur moyenne de gain par partie, tout en étant un gain impossible à obtenir au cours d'une unique partie.

PROPOSITION 26 *L'espérance possède les propriétés suivantes :*

1. (Linéarité) *Soient X une v.a.r. continue, g_1 et g_2 deux fonctions de \mathbb{R} dans \mathbb{R} continues par morceaux et λ un réel (ou : soient X une v.a.r. discrète, g_1 et g_2 deux fonctions de \mathbb{N} dans \mathbb{N}). On a*

$$\mathbb{E}(g_1(X) + g_2(X)) = \mathbb{E}(g_1(X)) + \mathbb{E}(g_2(X)),$$

$$\mathbb{E}(\lambda g_1(X)) = \lambda \mathbb{E}(g_1(X)).$$

2. (Positivité) *Soient X une v.a.r. continue et g une fonction de \mathbb{R} dans \mathbb{R} positive et continue par morceaux (ou : soient X une v.a.r. discrète et g une fonction de \mathbb{N} dans \mathbb{N} positive). On a*

$$\mathbb{E}(g(X)) \geq 0.$$

3. (Croissance) *Soient X une v.a.r. continue, g_1 et g_2 deux fonctions de \mathbb{R} dans \mathbb{R} continues par morceaux vérifiant⁴ $g_1 \leq g_2$ (ou : soient X une v.a.r. discrète, g_1 et g_2 deux fonctions de \mathbb{N} dans \mathbb{N} vérifiant⁵ $g_1 \leq g_2$). On a*

$$\mathbb{E}(g_1(X)) \leq \mathbb{E}(g_2(X));$$

4. *Si X est une v.a.r. constante sur Ω (on parle de fonction déterministe) et g une fonction quelconque alors,*

$$\mathbb{E}(g(X)) = g(X).$$

DÉMONSTRATION 1. Si X est une v.a.r. continue, alors

$$\begin{aligned} \mathbb{E}(g_1(X) + g_2(X)) &= \int_{\mathbb{R}} (g_1(t) + g_2(t)) f_X(t) dt \\ &= \int_{\mathbb{R}} g_1(t) f_X(t) dt + \int_{\mathbb{R}} g_2(t) f_X(t) dt \\ &= \mathbb{E}(g_1(X)) + \mathbb{E}(g_2(X)) \end{aligned}$$

par linéarité de l'intégrale. De même,

$$\mathbb{E}(\lambda g_1(X)) = \int_{\mathbb{R}} \lambda g_1(t) f_X(t) dt = \lambda \int_{\mathbb{R}} g_1(t) f_X(t) dt = \lambda \mathbb{E}(g_1(X)).$$

Si X est une v.a.r. discrète telle que $X(\Omega) = \{x_j \in \mathbb{R}, j \in J\}$ où $J \subset \mathbb{N}$, alors

$$\begin{aligned} \mathbb{E}(g_1(X) + g_2(X)) &= \sum_{j \in J} (g_1(x_j) + g_2(x_j)) \times \mathbb{P}(X = x_j) \\ &= \sum_{j \in J} g_1(x_j) \times \mathbb{P}(X = x_j) + \sum_{j \in J} g_2(x_j) \times \mathbb{P}(X = x_j) \\ &= \mathbb{E}(g_1(X)) + \mathbb{E}(g_2(X)) \end{aligned}$$

par linéarité de la série. De même,

$$\mathbb{E}(\lambda g_1(X)) = \sum_{j \in J} \lambda g(x_j) \times \mathbb{P}(X = x_j) = \lambda \sum_{j \in J} g(x_j) \times \mathbb{P}(X = x_j) = \lambda \mathbb{E}(g_1(X)).$$

2. Si X est une v.a.r. continue, alors

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(t) f_X(t) dt.$$

Puisque f_X est une fonction positive (c'est une densité de probabilité) et que par hypothèse g est positive, le produit $g \times f_X$ l'est aussi. Le résultat découle alors des propriétés de positivité de l'intégrale. Le raisonnement est le même pour le cas discret, car pour tout entier i , $\mathbb{P}(X = i) \geq 0$.

3. Il suffit de remarquer que : $g_1 \leq g_2 \implies g_2 - g_1 \geq 0$ et appliquer le résultat précédent à la fonction $g = g_2 - g_1$.

4. Si $X = a$ (avec $a \in \mathbb{R}$) alors X est forcément une v.a.r. discrète, et $\mathbb{P}(X = a) = 1$. On a donc $\mathbb{E}(g(X)) = g(a)\mathbb{P}(X = a) = g(a) = g(X)$. \circ

COROLLAIRE 1 Soient X une v.a.r. sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et a un réel. On a

1. $\mathbb{E}(a) = a$ (on considère ici la v.a.r. constante égale à a) ;
2. $\mathbb{E}(X) \geq 0$ si X est à valeurs positives.

3.2 Variance

DÉFINITION 26 La variance de la v.a.r. X est définie, lorsque cette quantité existe, par

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

On peut simplifier l'écriture de $\text{Var}(X)$ en utilisant la propriété de linéarité de l'espérance :

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

Posons pour simplifier, en supposant que cette quantité existe, $m = \mathbb{E}(X)$. On a alors pour tout $a \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}((X - a)^2) &= \mathbb{E}((X - m + m - a)^2) \\ &= \mathbb{E}((X - m)^2 - 2(m - a)(X - m) + (m - a)^2) \\ &= \mathbb{E}((X - m)^2) + 2(m - a)\mathbb{E}(X - m) + \mathbb{E}((m - a)^2) \\ &= \mathbb{E}((X - m)^2) + \mathbb{E}((m - a)^2), \end{aligned}$$

car $\mathbb{E}(X - m) = \mathbb{E}(X) - m = 0$. On remarque que $a = m$ est la valeur qui minimise $\mathbb{E}((X - a)^2)$, et ce minimum est appelé la variance de X .

La variance d'une v.a.r. X est la quantité qui mesure la «dispersion» de X autour de sa moyenne. Plus les valeurs de X sont dispersées, plus sa variance augmente. On peut démontrer de plus que

$$X \text{ est une v.a.r. constante} \iff \text{Var}(X) = 0.$$

PROPOSITION 27 *La variance vérifie les propriétés suivantes. Pour toute v.a.r. X et pour tout réel λ on a*

1. $\text{Var}(X + \lambda) = \text{Var}(X)$;
2. $\text{Var}(\lambda X) = \lambda^2 \text{Var}(X)$;
3. $\text{Var}(X) \geq 0$.

DÉMONSTRATION 1. On a

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

2. On a

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

3. On a $(X - \mathbb{E}(X))^2 \geq 0$, et le résultat découle de la positivité de l'espérance.

○

COROLLAIRE 2 *Si X est une v.a.r. pour laquelle $\mathbb{E}(X)$ et $\text{Var}(X)$ existent, alors la v.a.r. Y définie par*

$$Y = \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}}$$

appelée variable centrée réduite associée à la v.a.r. X , vérifie :

$$\mathbb{E}(Y) = 0,$$

$$\text{Var}(Y) = 1.$$

COROLLAIRE 3 *Si X est une v.a.r. pour laquelle $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$ existent, alors*

$$\mathbb{E}(X)^2 \leq \mathbb{E}(X^2).$$

DÉMONSTRATION On a $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. La positivité de la variance implique la relation $\mathbb{E}(X)^2 \leq \mathbb{E}(X^2)$. ○

L'inégalité donnée au corollaire 3 est un cas particulier de l'inégalité dite *de Jensen* : pour toute fonction ϕ convexe, et pour toute v.a.r. X telle que $\mathbb{E}(X)$ et $\mathbb{E}(\phi(X))$ existent on a

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X)).$$

DÉFINITION 27 Soit X une v.a.r. telle que $\text{Var}(X)$ existe. L'écart-type de X est défini par

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

3.3 Autres moments

On peut généraliser les définitions de l'espérance et de la variance d'une v.a.r.. Pour $p \in \mathbb{N}$, on appelle *moment d'ordre p* de la v.a.r. X , le réel $\mathbb{E}(|X|^p)$, quand il existe. On appelle *moment centré d'ordre p* le réel $\mathbb{E}(|X - \mathbb{E}(X)|^p)$.

Ainsi, $\mathbb{E}(X^2)$ est le moment d'ordre 2, et $\text{Var}(X)$ le moment centré d'ordre 2.

3.4 Espérance et variance pour les v.a.r. de lois usuelles

Loi uniforme $X \sim \mathcal{U}_N$, $N \in \mathbb{N}^*$

On a

$$\mathbb{E}(X) = \sum_{k=1}^N k \frac{1}{N} = \frac{1}{N} \sum_{k=1}^N k = \frac{N+1}{2}.$$

et

$$\text{Var}(X) = \sum_{k=1}^N k^2 \frac{1}{N} = \frac{1}{N} \sum_{k=1}^N k^2 = \frac{N^2 - 1}{12}.$$

On rappelle, voir par exemple [Balac-Sturm], que la somme des N premiers entiers vaut $N(N+1)/2$ et que la somme des carrés des N premiers entiers vaut $N(N^2 - 1)/12$.

Loi de Bernoulli $X \sim \mathcal{B}(1, p)$, $p \in]0, 1[$

On a

$$\mathbb{E}(X) = 0 \times \mathbb{P}(X = 0) + 1 \times \mathbb{P}(X = 1) = p.$$

et

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 0^2 \times \mathbb{P}(X = 0) + 1^2 \times \mathbb{P}(X = 1) - p^2 = p(1 - p).$$

Loi géométrique $X \sim \mathcal{G}(p)$, $p \in]0, 1[$

On rappelle que pour tout $r \in]-1, 1[$ la série géométrique $\sum_{k=1}^{+\infty} r^k$ converge uniformément vers $\frac{1}{1-r}$. On a alors

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

On montre sur le même principe que

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

Loi uniforme $X \sim \mathcal{U}(a, b)$, $a < b$

On a

$$\mathbb{E}(X) = \int_a^b \frac{t}{b-a} dt = \left[\frac{t^2}{2(b-a)} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

L'espérance de X est la moyenne entre a et b !

D'autre part, $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ avec

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

et

$$\mathbb{E}(X)^2 = \left(\frac{a+b}{2} \right)^2 = \frac{a^2 + 2ab + b^2}{4},$$

d'où

$$\text{Var}(X) = \frac{4b^2 + 4ab + 4a^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{a^2 + b^2 - 2ab}{12} = \frac{(b-a)^2}{12}.$$

Loi normale $X \sim \mathcal{N}(m, \sigma^2)$, $m, \sigma \in \mathbb{R}$

On a

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

Le changement de variable $y = \frac{x - m}{\sigma}$ permet d'obtenir la relation

$$\begin{aligned}
 \mathbb{E}(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (\sigma y + m) \exp(-y^2/2) \sigma dy \\
 &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} y \exp(-y^2/2) dy + m \times \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-y^2/2) dy}_{=1} \\
 &= m + \frac{\sigma}{\sqrt{2\pi}} \underbrace{\left[-\exp(-y^2/2) \right]_{-\infty}^{+\infty}}_{=0} \\
 &= m.
 \end{aligned}$$

On vérifie par un calcul analogue que $\mathbb{E}(X^2) = \sigma^2 + m^2$, ce qui permet d'établir que $\text{Var}(X) = \sigma^2$.

3.5 Inégalités remarquables

MARKOV, Andreï Andreïevitch (1856, Ryazan (Russie) - 1922, Saint-Petersbourg).



On doit à cet élève de Tchebychev de très importants travaux en calcul des probabilités et en théorie du potentiel. Il crée l'analyse «markovienne» (macrolinguistique) qui a permis de grands progrès dans le cryptage (à vocation militaire) mais aussi dans l'analyse de documents anciens partiellement effacés. Il a introduit de façon précise les processus aléatoires (chaines de Markov) ce qui lui a permis de donner une démonstration précise du théorème central limite.

PROPOSITION 28 (Inégalité de Markov) *Soit X une v.a.r.. Pour tout réel a strictement positif on a*

$$\mathbb{P}(|X| \geq a) \leq \frac{1}{a} \mathbb{E}(|X|).$$

DÉMONSTRATION Supposons que X est une v.a.r. admettant une densité f_X (le cas général est admis). On a

$$\mathbb{P}(|X| \geq a) = \mathbb{P}(X \in]-\infty, -a] \cup [a, +\infty[) = \int_{-\infty}^{-a} f_X(t) dt + \int_a^{+\infty} f_X(t) dt.$$

Or pour $t \in [a, +\infty[$ on a $\frac{t}{a} \geq 1$, d'où

$$\int_a^{+\infty} f_X(t) dt \leq \int_a^{+\infty} \frac{t}{a} f_X(t) dt.$$

De même, pour $t \in]-\infty, -a]$ on a $|\frac{t}{a}| \geq 1$, d'où

$$\int_{-\infty}^{-a} f_X(t) dt \leq \int_{-\infty}^{-a} \left| \frac{t}{a} \right| f_X(t) dt.$$

Finalement,

$$\begin{aligned} \mathbb{P}(|X| \geq a) &\leq \int_a^{+\infty} \left| \frac{t}{a} \right| f_X(t) dt + \int_{-\infty}^{-a} \left| \frac{t}{a} \right| f_X(t) dt \\ &\leq \int_{-\infty}^{+\infty} \left| \frac{t}{a} \right| f_X(t) dt = \mathbb{E}\left(\left| \frac{X}{a} \right|\right). \end{aligned}$$

○

BIENAYMÉ, Irénée Jules (1796, Paris - 1878, Paris).



Statisticien, inspecteur général des Finances, il appliqua la théorie des probabilités aux calculs financiers et entra à l'Académie des sciences en 1852. On lui doit les premiers travaux sur les variables aléatoires.

PROPOSITION 29 (Inégalité de Bienaym -Tchebychev) *Soit X une v.a.r. et α un réel strictement positif. On a*

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

DÉMONSTRATION Il suffit d'appliquer l'inégalité de Markov à la v.a.r. $Y = (X - \mathbb{E}(X))^2$ en prenant $a = \alpha^2$. On a

$$\mathbb{P}((X - \mathbb{E}(X))^2 \geq \alpha^2) \leq \frac{1}{\alpha^2} \mathbb{E}((X - \mathbb{E}(X))^2) = \frac{\text{Var}(X)}{\alpha^2}.$$

○

TCHEBYCHEV, Pafnouty Lvovitch (1821, Saint-Petersbourg - 1894, Saint-Petersbourg).



Professeur de mathématiques à Saint-Petersbourg où il crée sa propre école de mathématiques où enseignera, par exemple, Markov son élève. Ses travaux portent essentiellement en théorie des nombres et des probabilités. Il fut membre de l'Académie des sciences de Saint-Petersbourg et des plus grandes académies d'Europe (de France, de Berlin, de Londres).

REMARQUE Les inégalités de Markov et de Tchebychev permettent d'avoir une estimation de certaines probabilités alors que la densité de la v.a.r. mise en jeu n'est pas connue. Toutefois, il faut prendre garde au fait que l'on a des inégalités et que la probabilité exacte peut être éloignée de la borne proposée.

EXEMPLE Le nombre de pièces fabriquées dans une usine en une semaine est une v.a.r. d'espérance 50 et de variance 25.

On peut en utilisant l'inégalité de Markov, estimer la probabilité que la production de la semaine à venir dépasse 75 pièces :

$$\mathbb{P}(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}.$$

En utilisant l'inégalité de Bienaymé-Tchebychev, on peut estimer la probabilité que la production de la semaine à venir soit strictement comprise entre 40 et 60 pièces. On a

$$\mathbb{P}(|X - 50| \geq 10) \leq \frac{25}{10^2} = \frac{1}{4}$$

donc $\mathbb{P}(|X - 50| < 10) \geq \frac{3}{4}$.

4 Caractérisation de la loi d'une variable aléatoire

Nous avons vu que la loi d'une v.a.r. est caractérisée par sa fonction de masse dans le cas d'une v.a.r. discrète ou par sa densité dans le cas d'une v.a.r. continue. Il existe d'autres manières de caractériser la loi d'une v.a.r..

Ces différents outils de caractérisation vont en particulier s'avérer utiles à la résolution du problème-type suivant : soit X v.a.r. de loi connue, soit $h : \mathbb{R} \rightarrow \mathbb{R}$ fonction continue par morceaux, quelle est la loi de la v.a.r. $Y = h(X)$?

4.1 Caractérisation par la fonction de répartition

PROPOSITION 30 *La fonction de répartition caractérise la loi de la v.a.r. associée.*

DÉMONSTRATION Dans le cas où X admet une densité f_X , on a vu que

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

autrement dit que F_X est l'intégrale indéfinie⁶ associée à f_X . Ceci implique que pour tout réel x , $F'_X(x) = f_X(x)$. Puisque la densité caractérise la loi, il en est donc de même pour la fonction de répartition.

Dans le cas où X est une v.a.r. discrète, on a vu qu'il y avait une relation étroite entre la fonction de répartition F_X et la fonction de masse ($p_j, j \in \mathbb{N}$) (voir la proposition 23). On peut en particulier, calculer la fonction de masse grâce à la fonction de répartition : Si $X(\Omega) = \{x_j \in \mathbb{R}, j \in J\}$ avec $J \subset \mathbb{N}$ on a la relation

$$\forall j \in \mathbb{N}, \quad p_j = F_X(x_j) - F_X(x_{j-1}).$$

Là encore, puisque la fonction de masse caractérise la loi, il en est donc de même pour la fonction de répartition. \circ

EXEMPLE Soit X une v.a.r. de loi uniforme $\mathcal{U}([0, 1])$ et Y la v.a.r. définie par $Y = X^2$. Déterminons la loi de Y . La fonction de répartition de X est définie par

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}.$$

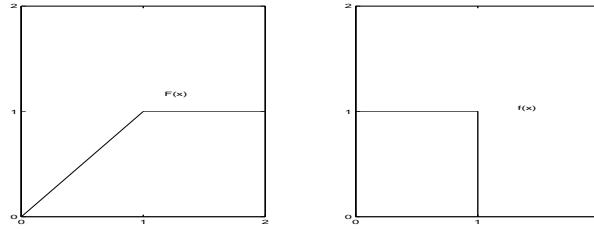


FIG. 3.4 – Fonction de répartition et densité de X

On a

$$F_Y(x) = \begin{cases} \mathbb{P}(Y \leq x) = \mathbb{P}(X^2 \leq x) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

⁶Voir par exemple [Balac-Sturm].

et

$$\mathbb{P}(X \leq \sqrt{x}) = \begin{cases} \sqrt{x} & \text{si } 0 \leq \sqrt{x} \leq 1 \text{ c'est-à-dire si } 0 \leq x \leq 1 \\ 1 & \text{si } \sqrt{x} \geq 1 \text{ c'est-à-dire si } x \geq 1 \end{cases}$$

donc

$$F_Y(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \sqrt{x} & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x \geq 1 \end{cases}.$$

On en déduit la densité de Y :

$$f_Y(x) = F'_Y(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{2\sqrt{x}} & \text{si } 0 < x < 1 \\ 0 & \text{si } x > 1 \end{cases},$$

et on peut poser par exemple $f_Y(0) = f_Y(1) = 0$.

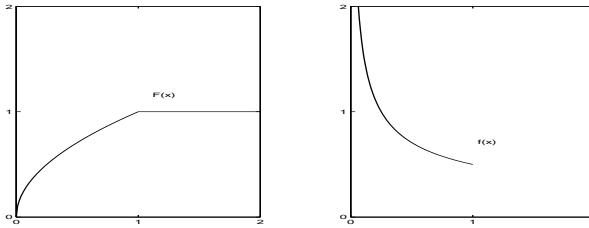


FIG. 3.5 – Fonction de répartition et densité de Y

4.2 Autre caractérisation

Nous nous plaçons ici uniquement dans le cas d'une v.a.r. continue.

PROPOSITION 31 *Soit X une v.a.r. admettant une densité f_X . S'il existe, pour toute fonction ϕ bornée et continue par morceaux sur \mathbb{R} , une fonction f ne dépendant pas de ϕ telle que l'espérance de $\phi(X)$ admette la formule de représentation intégrale*

$$\mathbb{E}(\phi(X)) = \int_{\mathbb{R}} \phi(x)f(x) dx,$$

alors $f = f_X$, autrement dit f est la densité de X .

DÉMONSTRATION Ce résultat est admis. ○

EXEMPLE Soit $X \sim \mathcal{U}([0, 1])$ et $Y = X^2$. La densité de la v.a.r. X est $f_X(x) = \mathbb{I}_{[0,1]}(x)$. Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction bornée et continue par morceaux. Alors

$$\mathbb{E}(\phi(Y)) = \mathbb{E}(\phi(X^2)) = \int_0^1 \phi(x^2) dx = \int_0^1 \phi(y) \frac{1}{2\sqrt{y}} dy,$$

la dernière égalité résultant du changement de variables $y = x^2$ sur $[0, 1]$. Puisque l'on a égalité pour toute fonction ϕ , on en conclut d'après la proposition 31 que

$$f_Y(y) = \frac{1}{2\sqrt{y}} \mathbb{I}_{[0,1]}(y).$$

4.3 Fonction caractéristique, fonction génératrice

Nous allons définir les fonctions les plus utilisées pour la caractérisation de lois : la fonction caractéristique pour une v.a.r. quelconque et la fonction génératrice dans le cas d'une v.a.r. discrète. Nous notons i l'unité imaginaire telle que $i^2 = -1$.

DÉFINITION 28 On appelle fonction caractéristique de la v.a.r. X la fonction de \mathbb{R} dans \mathbb{C} définie par

$$\phi_X(t) = \mathbb{E}(e^{itX}) \quad \forall t \in \mathbb{R}.$$

Nous utiliserons souvent le fait que la fonction caractéristique, comme son nom l'indique, caractérise la loi de la v.a.r. associée (i.e. la donnée de la fonction ϕ_X suffit à déterminer complètement la loi de X).

THÉORÈME 2 Deux v.a.r. définies sur un même espace probabilisé ayant même fonction caractéristique ont même loi.

DÉMONSTRATION Ce résultat est admis. ○

PROPOSITION 32 La fonction caractéristique ϕ_X d'une v.a.r. X possède les propriétés suivantes :

1. ϕ_X est une application continue à valeurs dans $\{z \in \mathbb{C}, |z| \leq 1\}$ et on a $\phi_X(0) = 1$.
2. Si ϕ_X est deux fois dérivable en 0, alors $\mathbb{E}(X)$ et $\mathbb{E}(X^2)$ existent⁷ et

$$\begin{aligned} \mathbb{E}(X) &= -i\phi'_X(0), \\ \mathbb{E}(X^2) &= -\phi''_X(0). \end{aligned}$$

3. Soient a et b deux réels et Y la v.a.r. définie par $Y = aX + b$. La fonction caractéristique ϕ_Y de la v.a.r. Y vérifie

$$\phi_Y(t) = e^{itb} \phi_X(at) \quad \forall t \in \mathbb{R}.$$

DÉMONSTRATION On se place dans le cas où X admet une densité. On a pour tout $t \in \mathbb{R}$,

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

1. La continuité de ϕ_X est admise. On a

$$\phi_X(0) = \int_{\mathbb{R}} f_X(x) dx = 1$$

car f_X est une densité de probabilité. Par ailleurs,

$$|\phi_X(t)| = \left| \int_{\mathbb{R}} e^{itx} f_X(x) dx \right| \leq \int_{\mathbb{R}} |f_X(x)| dx = 1$$

ce qui implique que ϕ_X est une application à valeurs dans $\{z \in \mathbb{C}, |z| \leq 1\}$.

2. Si ϕ_X est deux fois dérivable, nous avons

$$\phi'_X(t) = \int_{\mathbb{R}} (e^{itx})' f_X(x) dx = \int_{\mathbb{R}} ixe^{itx} f_X(x) dx,$$

la première égalité étant admise, d'où

$$\phi'_X(0) = i \int_{\mathbb{R}} xf_X(x) dx = i\mathbb{E}(X),$$

ce qui implique que $\mathbb{E}(X) = -i\phi'_X(0)$. On procède de même pour $\phi''_X(0)$.

3. On a pour tout réel t ,

$$\phi_Y(t) = \int_{\mathbb{R}} e^{it(ax+b)} f_X(x) dx = e^{itb} \int_{\mathbb{R}} e^{i(ta)x} f_X(x) dx = e^{itb} \phi_X(at).$$

○

Dans le cas où X est une v.a.r. discrète, il est souvent plus commode de considérer la fonction génératrice de X plutôt que la fonction caractéristique.

DÉFINITION 29 Soit une X v.a.r. discrète à valeurs entières. On appelle fonction génératrice de X la fonction définie pour tout $s \in [-1, 1]$ par

$$G_X(s) = \mathbb{E}(s^X) = \sum_{j \in \mathbb{N}} s^j \times \mathbb{P}(X = j).$$

PROPOSITION 33 La fonction génératrice d'une v.a.r. entière X a les propriétés suivantes :

1. la série $\sum_{j \in \mathbb{N}} s^j \times \mathbb{P}(X = j)$ est une série qui converge pour tout $s \in [-1, 1]$;
2. G_X est continue sur $[-1, 1]$ et infiniment dérivable sur $[-1, 1]$;

⁷On dit que X est respectivement intégrable et de carré intégrable.

DÉMONSTRATION Ce résultat est admis. ○

THÉORÈME 3 Deux v.a.r. entières définies sur un même espace probabilisé ayant même fonction génératrice ont même loi.

DÉMONSTRATION Ce résultat est admis. ○

PROPOSITION 34 La fonction génératrice d'une v.a.r. entière X a les propriétés suivantes.

$$1. \quad G_X(1) = 1, \quad G_X(0) = \mathbb{P}(X = 0).$$

$$2. \quad G'_X(1) = \mathbb{E}(X), \quad G''_X(1) = \mathbb{E}(X(X - 1)) \text{ et plus généralement } \forall k \in \mathbb{N}^*,$$

$$G_X^{(k)}(1) = \mathbb{E}(X \times (X - 1) \times \cdots \times (X - k + 1)).$$

DÉMONSTRATION 1. On a

$$G_X(1) = \sum_{j \in \mathbb{N}} \mathbb{P}(X = j) = \mathbb{P}\left(\bigcup_{j \in \mathbb{N}} [X = j]\right) = \mathbb{P}(\Omega) = 1.$$

Par ailleurs pour tout $s \in [-1, 1]$,

$$G_X(s) = \sum_{j \in \mathbb{N}} s^j \mathbb{P}(X = j) = \mathbb{P}(X = 0) + \sum_{j \in \mathbb{N}^*} s^j \mathbb{P}(X = j),$$

donc $G_X(0) = \mathbb{P}(X = 0)$.

2. On a pour tout $s \in [-1, 1]$,

$$G'_X(s) = \sum_{j \in \mathbb{N}} j s^{j-1} \mathbb{P}(X = j)$$

donc $G'_X(1) = \sum_{j \in \mathbb{N}} j \mathbb{P}(X = j) = \mathbb{E}(X)$ par définition de l'espérance. De même, pour tout $s \in [-1, 1]$,

$$G''_X(s) = \sum_{j \in \mathbb{N}} j(j-1) s^{j-2} \mathbb{P}(X = j)$$

d'où $G''_X(1) = \sum_{j \in \mathbb{N}} j(j-1) \mathbb{P}(X = j) = \mathbb{E}(X(X - 1))$. On utilise ensuite un raisonnement par récurrence pour établir que

$$G_X^{(k)}(1) = \mathbb{E}(X \times (X - 1) \times \cdots \times (X - k + 1)).$$

○

4.4 Fonctions caractéristique et génératrice de certaines lois

Loi binomiale $X \sim \mathcal{B}(n, p)$

- Fonction caractéristique : pour $t \in \mathbb{R}$

$$\begin{aligned}\phi_X(t) = \mathbb{E}(e^{itX}) &= \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} e^{itk} \\ &= \sum_{k=0}^n C_n^k (pe^{it})^k (1-p)^{n-k} \\ &= (pe^{it} + (1-p))^n,\end{aligned}$$

d'après la formule du binôme.

- Fonction génératrice : pour $s \in [-1, 1]$

$$G_X(s) = (ps + (1-p))^n.$$

Loi de Poisson $X \sim \mathcal{P}(\lambda)$

- Fonction caractéristique : pour $t \in \mathbb{R}$

$$\phi_X(t) = \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{\lambda^k}{k!} e^{itk} = \sum_{k \in \mathbb{N}} e^{-\lambda} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)},$$

car $e^u = \sum_{k=0}^{+\infty} \frac{u^k}{k!}$ pour tout réel u .

- Fonction génératrice : pour $s \in [-1, 1]$

$$G_X(s) = e^{\lambda(s-1)}.$$

Loi normale $X \sim \mathcal{N}(m, \sigma^2)$

Fonction caractéristique : pour $t \in \mathbb{R}$

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

On vérifie alors aisément que

$$\mathbb{E}(X) = -i\phi'_X(0) = m,$$

et que

$$\mathbb{E}(X^2) = -\phi''_X(0) = m^2 + \sigma^2,$$

d'où on déduit que

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

Loi de Cauchy

Cette loi est définie par la densité

$$f_X(t) = \frac{1}{\pi(1+t^2)}, \quad t \in \mathbb{R}.$$

On vérifie que

$$\int_{\mathbb{R}} \frac{1}{\pi(1+t^2)} dt = \frac{1}{\pi} \left[\arctan t \right]_{-\infty}^{+\infty} = 1,$$

mais la fonction

$$t \longmapsto \frac{t}{\pi(1+t^2)}$$

n'est pas intégrable sur \mathbb{R} , donc la variable aléatoire X de densité f_X n'a pas d'espérance. On peut vérifier que $\phi_X(t) = e^{-|t|}$ n'est pas dérivable en 0.

Chapitre 4

Vecteurs aléatoires

Pour épargner au lecteur des lourdeurs de notation pouvant affecter la compréhension, les définitions et résultats concernant les vecteurs aléatoires sont d'abord énoncés dans le cas simple où les vecteurs ont seulement deux composantes, c'est-à-dire le cas des couples aléatoires. Une généralisation aux vecteurs quelconques est donnée au paragraphe 4.

1 Couple de variables aléatoires réelles

On appelle tribu borélienne de \mathbb{R}^2 la tribu engendrée par la famille des ouverts de \mathbb{R}^2 , c'est-à-dire la plus petite tribu contenant tous les ouverts de \mathbb{R}^2 . On appelle borélien de \mathbb{R}^2 un élément de la tribu borélienne de \mathbb{R}^2 .

DÉFINITION 30 Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X l'application

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}^2 \\ \omega &\longmapsto (X_1(\omega), X_2(\omega)). \end{aligned}$$

On dit que X est un **couple aléatoire** si pour tout borélien B de \mathbb{R}^2 , $X^{-1}(B) \in \mathcal{A}$.

On démontre facilement qu'un couple aléatoire (c'est-à-dire un vecteur aléatoire de \mathbb{R}^2) $X = (X_1, X_2)$ est un couple de deux v.a.r. X_1 et X_2 .

1.1 Lois d'un couple de variables aléatoires réelles

DÉFINITION 31 On appelle fonction de répartition du couple de v.a.r. $X = (X_1, X_2)$ l'application $F_X : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par

$$F_X(x_1, x_2) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) \quad \forall (x_1, x_2) \in \mathbb{R}^2$$

où on a noté

$$\begin{aligned} \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2) &= \mathbb{P}([X_1 \leq x_1] \cap [X_2 \leq x_2]) \\ &= \mathbb{P}(\{\omega \in \Omega, X_1(\omega) \leq x_1, X_2(\omega) \leq x_2\}). \end{aligned}$$

PROPOSITION 35 Si $X = (X_1, X_2)$ est un couple de v.a.r. de fonction de répartition F_X alors on a

$$\lim_{x \rightarrow -\infty} F_X(x, x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F_X(x, x) = 1.$$

DÉMONSTRATION La démonstration de ce résultat est analogue à celle de la proposition 22, page 29. \circ

DÉFINITION 32 Le couple de v.a.r. $X = (X_1, X_2)$ défini sur l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est dit discret s'il est à valeurs dans un sous-ensemble $D = X(\Omega)$ fini ou dénombrable de \mathbb{R}^2 . La fonction

$$\begin{aligned} D &\longrightarrow [0, 1] \\ (k_1, k_2) &\longmapsto \mathbb{P}(X_1 = k_1, X_2 = k_2) \end{aligned}$$

est alors appelée **fonction de masse** du couple de v.a.r. $X = (X_1, X_2)$.

DÉFINITION 33 Le couple de v.a.r. $X = (X_1, X_2)$ est dit **absolument continu** de densité f_X , si sa fonction de répartition F_X admet la représentation intégrale suivante :

$$F_X(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_X(t_1, t_2) \, dt_2 \, dt_1.$$

Les fonctions de répartition F_{X_1} et F_{X_2} des v.a.r. X_1 et X_2 sont données en fonction de la fonction de répartition du couple de v.a.r. $X = (X_1, X_2)$ par

$$\begin{aligned} F_{X_1}(x) &= \mathbb{P}(X_1 \leq x, X_2 \in \mathbb{R}) = \lim_{y \rightarrow +\infty} F_X(x, y) \quad \forall x \in \mathbb{R}, \\ F_{X_2}(y) &= \mathbb{P}(X_1 \in \mathbb{R}, X_2 \leq y) = \lim_{x \rightarrow +\infty} F_X(x, y) \quad \forall y \in \mathbb{R}. \end{aligned}$$

Dans le cas où le couple de v.a.r. $X = (X_1, X_2)$ admet pour densité f_X , on a

$$\begin{aligned} F_{X_1}(x) &= \int_{-\infty}^x \int_{-\infty}^{+\infty} f_X(t_1, t_2) \, dt_2 \, dt_1 \quad \forall x \in \mathbb{R}, \\ F_{X_2}(y) &= \int_{-\infty}^y \int_{-\infty}^{+\infty} f_X(t_1, t_2) \, dt_1 \, dt_2 \quad \forall y \in \mathbb{R}. \end{aligned}$$

DÉFINITION 34 Pour $k \in \{1, 2\}$, on appelle $k^{\text{ième}}$ loi marginale du couple de v.a.r. $X = (X_1, X_2)$ la loi de la variable aléatoire réelle X_k .

Si $X = (X_1, X_2)$ est un couple de v.a.r. discret les lois marginales des v.a.r. X_1 et X_2 ont respectivement pour fonctions de masse les applications :

$$p_1 : k_1 \in \mathbb{N} \longmapsto \sum_{k_2 \in \mathbb{N}} \mathbb{P}(X_1 = k_1, X_2 = k_2),$$

$$p_2 : k_2 \in \mathbb{N} \mapsto \sum_{k_1 \in \mathbb{N}} \mathbb{P}(X_1 = k_1, X_2 = k_2).$$

Si $X = (X_1, X_2)$ est un couple de v.a.r. absolument continu de densité f_X , les lois marginales des v.a.r. X_1 et X_2 ont respectivement pour fonctions densités les applications :

$$\begin{aligned} f_{X_1} : x_1 \in \mathbb{R} &\mapsto \int_{\mathbb{R}} f_X(x_1, x_2) dx_2, \\ f_{X_2} : x_2 \in \mathbb{R} &\mapsto \int_{\mathbb{R}} f_X(x_1, x_2) dx_1. \end{aligned}$$

1.2 Moments d'un couple de variables aléatoires réelles

Espérance d'un couple de v.a.r.

DÉFINITION 35 On appelle *espérance* du couple de v.a.r. (X_1, X_2) l'élément de \mathbb{R}^2 défini par

$$\mathbb{E}(X_1, X_2) = (\mathbb{E}(X_1), \mathbb{E}(X_2)).$$

PROPOSITION 36 Soit $X = (X_1, X_2)$ un couple de v.a.r. discret défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $D = X(\Omega)$ sous-ensemble fini ou dénombrable de \mathbb{N}^2 . Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction bornée et continue par morceaux. La v.a.r. discrète $Z = h(X_1, X_2)$ a pour espérance

$$\mathbb{E}(Z) = \sum_{(k_1, k_2) \in D} h(k_1, k_2) \times \mathbb{P}(X_1 = k_1, X_2 = k_2).$$

DÉMONSTRATION Considérons les ensembles finis ou dénombrables suivants,

$$\begin{aligned} D &= X(\Omega) = \left\{ k = (k_1, k_2) \in \mathbb{N}^2, \exists \omega \in \Omega \quad k = X(\omega) \right\}, \\ \Omega' &= h(D) = \left\{ z \in \mathbb{R}, \exists k = (k_1, k_2) \in D \quad z = h(k_1, k_2) \right\}. \end{aligned}$$

On a alors,

$$\begin{aligned} E(Z) &= \sum_{z \in \Omega'} z \mathbb{P}(Z = z) \\ &= \sum_{z \in \Omega'} z \mathbb{P}(h(X_1, X_2) = z) \\ &= \sum_{k \in D} h(k_1, k_2) \mathbb{P}(h(X_1, X_2) = h(k_1, k_2)) \\ &= \sum_{k \in D} h(k_1, k_2) \mathbb{P}((X_1, X_2) = (k_1, k_2)). \end{aligned}$$

○

PROPOSITION 37 Soient $X = (X_1, X_2)$ un couple de v.a.r. absolument continu admettant pour densité l'application f_X de \mathbb{R}^2 dans \mathbb{R} et $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction bornée et continue par morceaux. La v.a.r. $Z = h(X_1, X_2)$ a pour espérance

$$\mathbb{E}(Z) = \int_{\mathbb{R}^2} h(t_1, t_2) \times f_X(t_1, t_2) \, dt_1 \, dt_2$$

lorsque cette intégrale existe.

COROLLAIRE 4 Soient $X = (X_1, X_2)$ un couple de v.a.r. et (a, b) deux réels. On a

$$\mathbb{E}(aX_1 + bX_2) = a\mathbb{E}(X_1) + b\mathbb{E}(X_2).$$

Covariance

DÉFINITION 36 Soit (X_1, X_2) un couple de v.a.r.. Si $\mathbb{E}(X_1)$ et $\mathbb{E}(X_2)$ existent, on appelle **covariance** du couple de v.a.r. (X_1, X_2) le réel

$$\text{Cov}(X_1, X_2) = \mathbb{E}\left((X_1 - \mathbb{E}(X_1)) \times (X_2 - \mathbb{E}(X_2))\right).$$

PROPOSITION 38 La covariance du couple de v.a.r. (X_1, X_2) vérifie les propriétés suivantes :

1. $\text{Cov}(X_1, X_1) = \text{Var}(X_1)$;
2. $\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1) \times \mathbb{E}(X_2)$;
3. l'application $(X_1, X_2) \mapsto \text{Cov}(X_1, X_2) \in \mathbb{R}$ est une forme bilinéaire symétrique : pour toutes v.a.r. X_1, Y_1, X_2, Y_2 et pour tous réels a, b on a
 - (a) $\text{Cov}(aX_1 + bY_1, X_2) = a\text{Cov}(X_1, X_2) + b\text{Cov}(Y_1, X_2)$,
 - (b) $\text{Cov}(X_1, aX_2 + bY_2) = a\text{Cov}(X_1, X_2) + b\text{Cov}(X_1, Y_2)$,
 - (c) $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$.

DÉMONSTRATION

1. Cette relation est évidente d'après la définition de la variance d'une v.a.r..

2. On a

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \mathbb{E}(X_1 X_2 - \mathbb{E}(X_1) X_2 - X_1 \mathbb{E}(X_2) + \mathbb{E}(X_1) \times \mathbb{E}(X_2)) \\ &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1) \times \mathbb{E}(X_2) \end{aligned}$$

d'après les propriétés de linéarité de l'espérance.

3. La bilinéarité de l'application covariance résulte des propriétés de linéarité de l'espérance. Elle est clairement symétrique d'après sa définition.

○

1.3 Fonction caractéristique et fonction génératrice

DÉFINITION 37 On appelle **fonction caractéristique** du couple de v.a.r. $X = (X_1, X_2)$ la fonction ϕ_X définie de \mathbb{R}^2 dans \mathbb{C} par

$$\phi_X(t_1, t_2) = \mathbb{E}(e^{i(t_1 X_1 + t_2 X_2)}).$$

DÉFINITION 38 On appelle **fonction génératrice** du couple de v.a.r. $X = (X_1, X_2)$ la fonction G_X définie de \mathbb{R}^2 dans \mathbb{R} par

$$G_X(s_1, s_2) = \mathbb{E}(s_1^{X_1} s_2^{X_2}).$$

1.4 Variables aléatoires conditionnelles

Soient (X, Y) un couple de v.a.r. discrètes à valeurs dans \mathbb{N}^2 et j un entier tel que $\mathbb{P}(Y = j) \neq 0$. Considérons la fonction

$$p : i \in \mathbb{N} \longmapsto \mathbb{P}(X = i \mid Y = j) \in [0, 1].$$

On a d'une part,

$$\mathbb{P}(X = i \mid Y = j) = \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(Y = j)}$$

et d'autre part

$$P(Y = j) = \sum_{i \in \mathbb{N}} \mathbb{P}(X = i, Y = j) \quad \text{et} \quad \mathbb{P}(X = i) = \sum_{j \in \mathbb{N}} \mathbb{P}(X = i, Y = j).$$

On en déduit que

$$\begin{aligned} \sum_{i \in \mathbb{N}} p(i) &= \sum_{i \in \mathbb{N}} \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(Y = j)} \\ &= \frac{1}{\mathbb{P}(Y = j)} \sum_{i \in \mathbb{N}} \mathbb{P}(X = i, Y = j) = \frac{1}{\mathbb{P}(Y = j)} \mathbb{P}(Y = j) \\ &= 1, \end{aligned}$$

autrement dit que la fonction p définit une fonction de masse.

DÉFINITION 39 (Cas discret) Soient (X, Y) un couple de v.a.r. discrètes à valeurs dans \mathbb{N}^2 et j un entier tel que $\mathbb{P}(Y = j) \neq 0$.

On appelle **loi conditionnelle de X sachant $[Y = j]$** la loi définie par la fonction de masse

$$p : i \in \mathbb{N} \longmapsto \mathbb{P}(X = i | Y = j).$$

On appelle **fonction de répartition conditionnelle de X sachant $[Y = j]$** l'application $F_X^{[Y=j]}$ de \mathbb{R} dans $[0, 1]$ définie pour tout $x \in \mathbb{R}$ par

$$F_X^{[Y=j]}(x) = \mathbb{P}(X \leq x | Y = j) = \frac{\mathbb{P}(X \leq x, Y = j)}{\mathbb{P}(Y = j)}.$$

REMARQUE On peut définir de manière analogue la loi conditionnelle de X sachant $[Y \leq j]$, et plus généralement la loi conditionnelle de X sachant n'importe quel événement dépendant de Y .

Soient (X, Y) un couple de v.a.r. de densité $f_{(X,Y)}$. La seconde loi marginale du couple admet pour densité l'application

$$f_Y : t \in \mathbb{R} \longmapsto \int_{\mathbb{R}} f_{(X,Y)}(s, t) \, ds.$$

Considérons la fonction $f_X^{[Y=y]}$ définie par

$$f_X^{[Y=y]} : s \in \mathbb{R} \longmapsto \frac{f_{(X,Y)}(s, y)}{f_Y(y)}.$$

Il s'agit d'une fonction positive qui vérifie

$$\begin{aligned} \int_{\mathbb{R}} f_X^{[Y=y]}(s) \, ds &= \int_{\mathbb{R}} \frac{f_{(X,Y)}(s, y)}{f_Y(y)} \, ds \\ &= \frac{1}{f_Y(y)} \int_{\mathbb{R}} f_{(X,Y)}(s, y) \, ds = \frac{1}{f_Y(y)} f_Y(y) \\ &= 1. \end{aligned}$$

On en déduit que la fonction $f_X^{[Y=y]}$ définit une densité de probabilité.

DÉFINITION 40 (Cas continu) Soient (X, Y) un couple de v.a.r. de densité $f_{(X,Y)}$ et f_Y la densité de Y . Pour $y \in \mathbb{R}$ fixé tel que $f_Y(y) \neq 0$, la fonction $f_X^{[Y=y]}$ définie par

$$f_X^{[Y=y]}(s) = \frac{f_{(X,Y)}(s, y)}{f_Y(y)}$$

est une densité de probabilité appelée **densité conditionnelle de X sachant $[Y = y]$** .

2 Indépendance de 2 variables aléatoires réelles

2.1 Définition

DÉFINITION 41 Deux v.a.r. X et Y sont dites indépendantes si

$$\forall x \in \mathbb{R}, \forall y \in \mathbb{R}, \quad \mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x) \times \mathbb{P}(Y \leq y).$$

REMARQUE Si F_X , F_Y désignent les fonctions de répartition des v.a.r. X et Y et $F_{(X,Y)}$ la fonction de répartition du couple (X, Y) , on peut reformuler la définition de la manière suivante,

$$\forall x \in \mathbb{R}, \forall y \in \mathbb{R}, \quad F_{(X,Y)}(x, y) = F_X(x) \times F_Y(y).$$

PROPOSITION 39 1. Les deux v.a.r. discrètes à valeurs entières X et Y sont indépendantes si et seulement si

$$\forall i \in \mathbb{N}, \forall j \in \mathbb{N}, \quad \mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i) \times \mathbb{P}(Y = j).$$

2. Soient $f_{(X,Y)}$ la densité du couple de v.a.r. absolument continu (X, Y) et f_X et f_Y les densités marginales des v.a.r. X et Y . Les v.a.r. X et Y sont indépendantes si et seulement si

$$\forall s \in \mathbb{R}, \forall t \in \mathbb{R}, \quad f_{(X,Y)}(s, t) = f_X(s) \times f_Y(t).$$

2.2 Indépendance et moments

PROPOSITION 40 Soient X , Y deux v.a.r. indépendantes. On a,

1. $\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y)$;
2. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

DÉMONSTRATION 1. Supposons que X et Y sont deux v.a.r. continues de lois de densité f_X et f_Y respectivement. Puisque ces 2 v.a.r. sont indépendantes, la loi jointe du couple (X, Y) a pour densité $f_{(X,Y)} : (x, y) \in \mathbb{R}^2 \mapsto f_X(x) \times f_Y(y)$. Ainsi,

$$\begin{aligned} E(XY) &= \int_{\mathbb{R}^2} xy f_{(X,Y)} \, dx \, dy \\ &= \int_{\mathbb{R}^2} xy f_X(x) \times f_Y(y) \, dx \, dy \\ &= \left(\int_R x f_X(x) \, dx \right) \times \left(\int_R y f_Y(y) \, dy \right) = E(X) \times E(Y). \end{aligned}$$

Le cas de v.a.r. discrètes se démontre d'une manière analogue. On a

$$\begin{aligned}
 E(XY) &= \sum_{(k_1, k_2) \in \mathbb{N}^2} k_1 k_2 \mathbb{P}(X_1 = k_1, X_2 = k_2) \\
 &= \sum_{k_1 \in \mathbb{N}} \sum_{k_2 \in \mathbb{N}} k_1 k_2 \mathbb{P}(X_1 = k_1) \times \mathbb{P}(X_2 = k_2) \\
 &= \sum_{k_1 \in \mathbb{N}} k_1 \mathbb{P}(X_1 = k_1) \times \sum_{k_2 \in \mathbb{N}} k_2 \mathbb{P}(X_2 = k_2) \\
 &= E(X) \times E(Y).
 \end{aligned}$$

2. On a

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

Or puisque X et Y sont indépendantes on a $\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y)$, donc

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

○

Cette proposition nous donne une condition suffisante pour que l'espérance d'un produit soit égal au produit des espérances de 2 v.a.r.. La condition n'est pas nécessaire et vérifier que $\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y)$ n'assure pas que les deux v.a.r. X et Y sont indépendantes. Par contre en prenant la contraposée de cette assertion, on obtient un critère pour établir que deux v.a.r. X et Y ne sont pas indépendantes; il suffit que $\mathbb{E}(XY) \neq \mathbb{E}(X) \times \mathbb{E}(Y)$.

Le lien profond existant entre l'indépendance et la «séparation des variables» est plus généralement exprimé dans le théorème suivant.

THÉORÈME 4 Soient X, Y deux v.a.r. et g, h deux fonctions de \mathbb{R} dans \mathbb{R} . Si les v.a.r. X et Y sont indépendantes alors les deux v.a.r. $g(X)$ et $h(Y)$ sont indépendantes et

$$\mathbb{E}(g(X)h(Y)) = \mathbb{E}(g(X)) \times \mathbb{E}(h(Y))$$

dès que ces quantités existent.

DÉMONSTRATION Ce résultat est admis.

○

2.3 Coefficient de corrélation

L'indépendance des 2 v.a.r. X et Y implique que $\mathbb{E}(XY) = \mathbb{E}(X) \times \mathbb{E}(Y)$. Il est possible d'exploiter cette propriété pour mesurer le «taux de dépendance» des deux v.a.r. au moyen d'un réel appelé *coefficient de corrélation*.

On note $\sigma_X = \sqrt{\text{Cov}(X, X)} = \sqrt{\text{Var}(X)}$ (voir la définition 27 page 43).

DÉFINITION 42 *On appelle coefficient de corrélation de X et Y le réel défini par*

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

PROPOSITION 41 *Pour toutes v.a.r. X et Y , on a $|\rho(X, Y)| \leq 1$.*

De plus, le coefficient de corrélation $\rho(X, Y)$ est un indicateur de la façon dont X et Y sont corrélées, au sens où,

- si X et Y sont indépendantes, alors $\rho(X, Y) = 0$;
- s'il existe un couple $(a, b) \in \mathbb{R}^2$ tel que $Y = aX + b$, alors $|\rho(X, Y)| = 1$ (la réciproque est vraie).

DÉMONSTRATION Rappelons que pour tout réel λ on a $\text{Var}(X + \lambda Y) = 0$ si et seulement si $X + \lambda Y$ est constante (voir la remarque de la page 42). Calculons, pour $\lambda \in \mathbb{R}$, la variance de $X + \lambda Y$,

$$\begin{aligned} \text{Var}(X + \lambda Y) &= \mathbb{E}((X + \lambda Y - \mathbb{E}(X + \lambda Y))^2) \\ &= \mathbb{E}((X - \mathbb{E}(X) + \lambda(Y - \mathbb{E}(Y)))^2) \\ &= \mathbb{E}((X - \mathbb{E}(X))^2 + \lambda^2(Y - \mathbb{E}(Y))^2 + 2\lambda(X - \mathbb{E}(X)) \times (Y - \mathbb{E}(Y))) \\ &= \mathbb{E}((X - \mathbb{E}(X))^2) + \lambda^2 \mathbb{E}((Y - \mathbb{E}(Y))^2) \\ &\quad + 2\lambda \mathbb{E}((X - \mathbb{E}(X)) \times (Y - \mathbb{E}(Y))) \\ &= \text{Var}(X) + \lambda^2 \text{Var}(Y) + 2\lambda \text{Cov}(X, Y). \end{aligned}$$

On sait que $\forall \lambda \in \mathbb{R}$, $\text{Var}(X + \lambda Y) \geq 0$, voir la proposition 27 page 42. Le polynôme

$$P(\lambda) = \lambda^2 \text{Var}(Y) + 2\lambda \text{Cov}(X, Y) + \text{Var}(X)$$

ne peut donc pas posséder 2 racines réelles distinctes (sans quoi il serait à valeurs négatives sur un intervalle de longueur non nulle). On en déduit que son discriminant $\Delta = \text{Cov}(X, Y)^2 - \text{Var}(X) \times \text{Var}(Y)$ est nécessairement négatif. On a

$$\begin{aligned} \text{Cov}(X, Y)^2 - \text{Var}(X) \times \text{Var}(Y) \leq 0 &\iff \text{Cov}(X, Y)^2 \leq \text{Var}(X) \times \text{Var}(Y) \\ &\iff |\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y \\ &\iff |\rho(X, Y)| \leq 1, \end{aligned}$$

ce qui achève la démonstration. L'égalité est réalisée lorsque $\text{Var}(X + \lambda Y) = 0$ c'est-à-dire lorsque $X + \lambda Y$ est constante.

Enfin, si X et Y sont indépendantes alors $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \times \mathbb{E}(Y) = 0$. \circ

ATTENTION Ce n'est pas parce que $\rho(X, Y) = 0$ que X et Y sont indépendantes. Par exemple, on définit les deux variables aléatoires $U = X + Y$ et $V = X - Y$ où X et Y sont deux variables aléatoires indépendantes de loi de Bernoulli $\mathcal{B}(1, p)$. On peut vérifier que le coefficient de corrélation du couple (U, V) vaut 0 mais que les variables aléatoires U et V ne sont pas indépendantes.

REMARQUE Soient X et Y deux v.a.r. ; on a

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}((X + Y - \mathbb{E}(X + Y))^2) \\ &= \mathbb{E}((X - \mathbb{E}(X) + Y - \mathbb{E}(Y))^2) \\ &= \mathbb{E}((X - \mathbb{E}(X))^2 + (Y - \mathbb{E}(Y))^2 + 2(X - \mathbb{E}(X)) \times (Y - \mathbb{E}(Y))) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).\end{aligned}$$

Donc si X et Y ne sont pas corrélées (c'est le cas si elles sont indépendantes, voir la proposition 40) alors

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

Ce résultat se généralise pour un nombre quelconque de variables aléatoires.

2.4 Indépendance et fonction caractéristique

L'indépendance de deux variables aléatoires peut s'exprimer à l'aide des fonctions caractéristiques et le cas échéant des fonctions génératrices.

THÉORÈME 5 1. Soit (X, Y) un couple aléatoire de fonction caractéristique $\phi_{(X,Y)}$. Soient ϕ_X et ϕ_Y les fonctions caractéristiques des deux v.a.r. X et Y .
Les v.a.r. X et Y sont indépendantes si et seulement si

$$\forall s, t \in \mathbb{R}, \quad \phi_{(X,Y)}(s, t) = \phi_X(s) \times \phi_Y(t),$$

2. Soit (X, Y) un couple aléatoire discret de fonction génératrice $G_{(X,Y)}$. Soient G_X et G_Y les fonctions génératrices des deux v.a.r. X et Y .

Les v.a.r. X et Y sont indépendantes si et seulement si

$$\forall s, t \in [-1, 1], \quad G_{(X,Y)}(s, t) = G_X(s) \times G_Y(t).$$

3 Somme de deux variables aléatoires indépendantes

3.1 Loi de la somme de deux variables aléatoires

PROPOSITION 42 Soient X et Y deux v.a.r. indépendantes. La loi de la v.a.r. $X + Y$ est obtenue en effectuant le **produit de convolution** des lois de X et de Y , autrement dit,

1. si X et Y sont des v.a.r. discrètes, on a $\forall k \in \mathbb{N}$,

$$\mathbb{P}(X + Y = k) = \sum_{i \in \mathbb{N}} \mathbb{P}(X = k - i) \times \mathbb{P}(Y = i);$$

2. si X et Y sont des v.a.r. continues de densités respectives f_X et f_Y , on a $\forall x \in \mathbb{R}$,

$$f_{X+Y}(x) = \int_{\mathbb{R}} f_X(x - t) \times f_Y(t) dt.$$

DÉMONSTRATION Soient X et Y deux v.a.r. discrètes à valeurs dans \mathbb{N} , indépendantes. On a

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \mathbb{P}\left(\underbrace{\bigcup_{i \in \mathbb{N}} [Y = i] \cap [X + Y = k]}_{= \Omega}\right) \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}([Y = i] \cap [X = k - i]) \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}(X = k - i, Y = i) \\ &= \sum_{i \in \mathbb{N}} \mathbb{P}(X = k - i) \times \mathbb{P}(Y = i), \end{aligned}$$

car X et Y sont indépendantes. Ce qui démontre la proposition dans le cas discret.

La démonstration s'étend **formellement** dans le cas des variables continues. Nous ne nous apesantirons pas sur la justification de l'existence et l'explicitation des dérivées des fonctions intégrales. Soient X et Y deux v.a.r. continues, indépendantes, de densités respectives f_X et f_Y . On note $f_{(X,Y)}$ la densité du couple (X, Y) . La fonction de répartition de la v.a.r. $X + Y$ vérifie

$$\begin{aligned} F_{X+Y}(z) &= \mathbb{P}(X + Y \leq z) \\ &= \iint_{\Delta} f_{(X,Y)}(x, y) dx dy \quad \text{où } \Delta = \{(x, y) \in \mathbb{R}^2, x + y \leq z\} \end{aligned}$$

Or les v.a.r. sont indépendantes donc

$$\begin{aligned} F_{X+Y}(z) &= \iint_{\Delta} f_X(x) f_Y(y) dx dy. \\ &= \int_{\mathbb{R}} f_Y(y) \left(\int_{-\infty}^{z-y} f_X(x) dx \right) dy \\ &\quad \text{car pour } y \text{ fixé } x \text{ varie entre } -\infty \text{ et } z - y \\ &= \int_{\mathbb{R}} f_Y(y) \left(\int_{-\infty}^z f_X(t - y) dt \right) dy \\ &\quad \text{par le changement de variable } t = x + y. \end{aligned}$$

On peut en déduire que la fonction de densité de la v.a.r. $X + Y$ vérifie

$$\begin{aligned} f_{X+Y}(z) &= \frac{d}{dz} F_{X+Y}(z) \\ &= \int_{\mathbb{R}} f_Y(y) \left(\frac{d}{dz} \left\{ \int_{-\infty}^z f_X(t-y) dt \right\} \right) dy \\ &= \int_{\mathbb{R}} f_Y(y) f_X(z-y) dy. \end{aligned}$$

○

COROLLAIRE 5 Soient X, Y deux v.a.r. continues de densités respectives f_X et f_Y et h une fonction continue par morceaux. On a

$$\mathbb{E}(h(X + Y)) = \int_{\mathbb{R}} h(t) \times (f_X * f_Y)(t) dt$$

où $*$ désigne le produit de convolution des fonctions f_X et f_Y .

On a donc

$$\mathbb{E}(h(X + Y)) = \int_{\mathbb{R}} h(t) \times \int_{\mathbb{R}} f_X(x-t) \times f_Y(t) dt.$$

3.2 Fonction caractéristique de la somme de deux variables aléatoires

PROPOSITION 43 Si X et Y sont deux v.a.r. indépendantes de fonctions caractéristiques ϕ_X et ϕ_Y alors la fonction caractéristique ϕ_{X+Y} de la v.a.r. $X + Y$ est donnée par

$$\phi_{X+Y}(t) = \phi_X(t) \times \phi_Y(t) \quad \forall t \in \mathbb{R}.$$

DÉMONSTRATION On a

$$\begin{aligned} \mathbb{E}(e^{it(X+Y)}) &= \mathbb{E}(e^{itX} e^{itY}) \\ &= \mathbb{E}(e^{itX}) \mathbb{E}(e^{itY}) \text{ par indépendance de } X \text{ et } Y \\ &= \phi_X(t) \phi_Y(t). \end{aligned}$$

○

PROPOSITION 44 Si X et Y sont deux v.a.r. discrètes indépendantes de fonctions génératrices G_X et G_Y alors la fonction génératrice G_{X+Y} de $X + Y$ est donnée par

$$G_{X+Y}(s) = G_X(s) \times G_Y(s) \quad \forall s \in [-1, 1].$$

COROLLAIRE 6 Soient X et Y deux v.a.r. indépendantes suivant une loi de Poisson de paramètres respectifs λ et μ . La v.a.r. $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.

DÉMONSTRATION On a $\forall k \in \mathbb{N}$,

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{et} \quad \mathbb{P}(Y = k) = e^{-\mu} \frac{\mu^k}{k!}.$$

On peut calculer les fonctions génératrices de X et de Y . On a pour $s \in [-1, 1]$,

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k \in \mathbb{N}} s^k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k \in \mathbb{N}} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{s\lambda} = e^{\lambda(s-1)},$$

et par un calcul analogue on vérifie que $G_Y(s) = e^{\mu(s-1)}$. On obtient donc pour fonction génératrice de $X + Y$,

$$G_{X+Y}(s) = G_X(s) G_Y(s) = e^{\lambda(s-1)} e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}.$$

On a donc montré que la v.a.r. $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$. ○

Les définitions et résultats énoncés dans ce paragraphe pour la somme de deux v.a.r. s'étendent à la somme de n v.a.r., voir le paragraphe 4.6.

4 Vecteurs aléatoires

4.1 Définition

On appelle tribu borélienne de \mathbb{R}^n , $n \in \mathbb{N}^*$, la tribu engendrée par la famille des ouverts de \mathbb{R}^n , c'est-à-dire la plus petite tribu contenant les ouverts de \mathbb{R}^n . On appelle borélien de \mathbb{R}^n un élément de la tribu borélienne de \mathbb{R}^n .

DÉFINITION 43 Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X l'application

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longmapsto (X_1(\omega), \dots, X_n(\omega)). \end{aligned}$$

On dit que X est un **vecteur aléatoire** de \mathbb{R}^n si pour tout borélien B de \mathbb{R}^n , $X^{-1}(B) \in \mathcal{A}$.

On démontre facilement que chaque composante X_k , $k = 1, \dots, n$, d'un vecteur aléatoire est une variable aléatoire réelle $X_k : \Omega \longrightarrow \mathbb{R}$.

DÉFINITION 44 On appelle fonction de répartition du vecteur aléatoire X l'application $F_X : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$F_X(x_1, \dots, x_n) = \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

où on a noté

$$\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \mathbb{P}([X_1 \leq x_1] \cap \dots \cap [X_n \leq x_n]).$$

PROPOSITION 45 Si X est un vecteur aléatoire de fonction de répartition F_X alors on a

$$\lim_{x \rightarrow -\infty} F_X(x, \dots, x) = 0,$$

et

$$\lim_{x \rightarrow +\infty} F_X(x, \dots, x) = 1.$$

DÉMONSTRATION La démonstration de ce résultat est analogue à celle de la proposition 22, page 29. \circ

4.2 Loi d'un vecteur aléatoire

DÉFINITION 45 Un vecteur aléatoire X défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ est dit discret s'il est à valeurs dans un sous-ensemble $D = X(\Omega)$ fini ou dénombrable de \mathbb{R}^n . La fonction

$$\begin{aligned} D &\longrightarrow [0, 1] \\ (k_1, \dots, k_n) &\longmapsto \mathbb{P}(X_1 = k_1, \dots, X_n = k_n) \end{aligned}$$

est appelée fonction de masse du vecteur aléatoire X .

DÉFINITION 46 Un vecteur aléatoire X est dit absolument continu de densité f_X , si sa fonction de répartition F_X admet la représentation intégrale suivante :

$$F_X(x_1, \dots, x_n) = \int_{\Delta} f_X(t_1, \dots, t_n) dt_1 \dots dt_n,$$

où

$$\Delta =]-\infty, x_1] \times]-\infty, x_2] \times \dots \times]-\infty, x_n] \subset \mathbb{R}^n.$$

REMARQUES 1. On a

$$\int_{\Delta} f_X(t_1, \dots, t_n) dt_1 \dots dt_n = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n) dt_n \dots dt_1.$$

2. Si F_X est de classe \mathcal{C}^n sur \mathbb{R}^n , alors la fonction de répartition est liée à la fonction densité par la relation

$$f_X(t_1, \dots, t_n) = \frac{\partial^n}{\partial t_1 \dots \partial t_n} F_X(t_1, \dots, t_n).$$

Par abus de langage, on utilise le terme vecteur aléatoire continu au lieu de vecteur aléatoire absolument continu.

Comme dans le cas d'une variable aléatoire, on peut vérifier que la fonction de répartition ainsi que la fonction de masse dans le cas discret ou la fonction de densité dans le cas continu, définissent et caractérisent la loi du vecteur aléatoire X . Ces fonctions permettent aussi de déterminer les lois de chacune des variables aléatoires réelles définies comme composantes du vecteur X .

DÉFINITION 47 On appelle $k^{\text{ème}}$ loi marginale $k \in \{1, \dots, n\}$ du vecteur aléatoire $X = (X_1, \dots, X_n)$ de \mathbb{R}^n la loi de la variable aléatoire réelle X_k .

PROPOSITION 46 La fonction de répartition de la v.a.r. X_k est donnée pour $x \in \mathbb{R}$ par

$$\begin{aligned} F_{X_k}(x) &= \mathbb{P}(X_k \leq x) \\ &= \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_{k-1} \in \mathbb{R}, X_k \leq x, X_{k+1} \in \mathbb{R}, \dots, X_n \in \mathbb{R}) \\ &= \lim_{y \rightarrow +\infty} F_X(y, \dots, y, x, y, \dots, y), \end{aligned}$$

où x figure à la $k^{\text{ème}}$ place dans $(y, \dots, y, x, y, \dots, y)$.

De plus, dans le cas où le vecteur aléatoire continu X admet pour densité f_X , on a

$$F_{X_k}(x) = \mathbb{P}(X_k \leq x) = \int_{-\infty}^x \left(\int_{\mathbb{R}^{n-1}} f_X(t_1, \dots, t_n) \, dt_1 \dots dt_{k-1} \, dt_{k+1} \dots dt_n \right) \, dt_k.$$

DÉMONSTRATION Laissée au lecteur. ○

4.3 Exemples de lois pour un vecteur aléatoire

Loi multinomiale $\mathcal{M}(n, p_1, \dots, p_k)$, où $n \in \mathbb{N}^*$ et $p_i \in]0, 1[\ \forall i \in \{1, \dots, k\}$.

La fonction de masse est définie par

$$\mathbb{P}(X_1 = \eta_1, \dots, X_k = \eta_k) = \frac{n!}{\eta_1! \times \dots \times \eta_k!} p_1^{\eta_1} \times \dots \times p_k^{\eta_k}, \quad \forall \eta = (\eta_1, \dots, \eta_k) \in \mathbb{N}^k,$$

avec $\sum_{i=1}^k \eta_i = n$ et $\sum_{i=1}^k p_i = 1$.

La loi multinomiale est la généralisation de la loi binomiale. On a

$$\mathcal{B}(n, p) = \mathcal{M}(n, p, 1 - p).$$

Loi normale n -dimensionnelle $\mathcal{N}(m, \Sigma)$ où $m = (m_1, \dots, m_n) \in \mathbb{R}^n$ et Σ est une matrice carrée d'ordre n , symétrique définie positive.

La fonction densité est définie pour tout $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ par

$$f_X(t_1, \dots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(t - m)^T \Sigma^{-1}(t - m)\right),$$

où $(t - m)^T$ désigne le vecteur colonne de composantes $t_i - m_i$, $i \in \{1, \dots, n\}$ et Σ^{-1} l'inverse de la matrice Σ .

Cette loi est la généralisation de la loi normale $\mathcal{N}(\mu, \sigma)$ où $m = (\mu)$ et $\Sigma = (\sigma)$.

4.4 Moments d'un vecteur aléatoire

On peut généraliser au cas d'un vecteur aléatoire la notion de moment introduite pour les variables aléatoires.

DÉFINITION 48 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On appelle espérance du vecteur aléatoire X , le vecteur

$$\mathbb{E}(X_1, \dots, X_n) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n)) \in \mathbb{R}^n.$$

Autrement dit, l'espérance d'un vecteur aléatoire est le vecteur constitué des espérances de chacune des v.a.r. coordonnées.

PROPOSITION 47 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire discret défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $D = X(\Omega)$ sous-ensemble fini ou dénombrable de \mathbb{R}^n . Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction bornée et continue par morceaux. L'espérance de la v.a.r. $Z = h(X_1, \dots, X_n)$ est donnée par

$$\mathbb{E}(Z) = \sum_{k=(k_1, \dots, k_n) \in D} h(k_1, \dots, k_n) \mathbb{P}(X_1 = k_1, \dots, X_n = k_n).$$

DÉMONSTRATION Ce résultat est admis. ○

PROPOSITION 48 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ admettant pour densité l'application f_X de \mathbb{R}^n dans \mathbb{R} . Soit $h : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction bornée et continue par morceaux. L'espérance de la v.a.r. $Z = h(X_1, \dots, X_n)$ est donnée par

$$\mathbb{E}(Z) = \int_{\mathbb{R}^n} h(t_1, \dots, t_n) f_X(t_1, \dots, t_n) \, dt_1 \dots dt_n$$

lorsque cette intégrale existe.

DÉMONSTRATION Ce résultat est admis. ○

PROPOSITION 49 Soient X, Y deux vecteurs aléatoires de \mathbb{R}^n et $A \in \mathcal{M}_n(\mathbb{R})$ une matrice carrée d'ordre n à coefficients réels. On a

1. $\mathbb{E}(AX) = A \mathbb{E}(X)$;
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

DÉMONSTRATION Laissée au lecteur. ○

La notion de variance d'une v.a.r. se généralise pour un vecteur aléatoire. On est amené à définir une matrice de covariance.

DÉFINITION 49 Soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire de \mathbb{R}^n . On appelle matrice de covariance de X la matrice

$$\Gamma_X = \left(\text{Cov}(X_i, X_j) \right)_{i,j=1,\dots,n}$$

La notion de matrice de covariance d'un vecteur aléatoire est une généralisation naturelle de la variance d'une v.a.r.. On a en effet

$$\Gamma_X = \mathbb{E} \left((X - \mathbb{E}(X))(X - \mathbb{E}(X))^T \right).$$

REMARQUE Pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$ et pour tout vecteur $B \in \mathbb{R}^n$, le vecteur aléatoire $Y = AX + B$ a pour matrice de covariance

$$\begin{aligned} \Gamma_Y &= \mathbb{E} [(AX + B - A\mathbb{E}(X) - B)(AX + B - A\mathbb{E}(X))^T] \\ &= \mathbb{E} \left(A(X - \mathbb{E}(X))(X - \mathbb{E}(X))^T A^T \right) \\ &= A \Gamma_X A^T, \end{aligned}$$

où A^T désigne la matrice transposée de A .

4.5 Fonction caractéristique et fonction génératrice

On peut étendre au cas des vecteurs aléatoires la notion de fonction caractéristique.

DÉFINITION 50 Soit X un vecteur aléatoire de \mathbb{R}^n . On appelle **fonction caractéristique** de X la fonction ϕ_X définie pour tout $\tau = (t_1, \dots, t_n) \in \mathbb{R}^n$ par

$$\phi_X(\tau) = \mathbb{E} \left(\exp \left(i \sum_{k=1}^n t_k X_k \right) \right) = \mathbb{E} (e^{i(\tau \cdot X)}).$$

PROPOSITION 50 Pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$ et pour tout vecteur $B \in \mathbb{R}^n$, le vecteur aléatoire $Y = AX + B$ a pour fonction caractéristique

$$\phi_Y(\tau) = e^{i(\tau \cdot B)} \phi_X(\tau A).$$

DÉMONSTRATION La preuve procède d'un simple calcul,

$$\begin{aligned} \phi_{AX+B}(\tau) &= \mathbb{E} (e^{i(\tau \cdot (AX+B))}) \\ &= e^{i(\tau \cdot B)} \mathbb{E} (e^{i\tau \cdot (AX)}) \\ &= e^{i(\tau \cdot B)} \mathbb{E} (e^{i(\tau A) \cdot X}) \\ &= e^{i(\tau \cdot B)} \phi_X(\tau A). \end{aligned}$$

○

Attention à ne pas confondre $\tau \cdot B$, produit scalaire entre deux vecteurs, avec τA , produit matriciel : $\tau \in \mathbb{R}^n$, $B \in \mathbb{R}^n$, $\tau \cdot B \in \mathbb{R}$ et $\tau A \in \mathbb{R}^n$.

DÉFINITION 51 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire discret de \mathbb{R}^n . On appelle **fonction génératrice** de X la fonction G_X définie pour tout $(s_1, \dots, s_n) \in [0, 1]^n$ par

$$G_X(s_1, \dots, s_n) = \mathbb{E} (s_1^{X_1} \dots s_n^{X_n}).$$

4.6 Indépendance de n variables aléatoires

DÉFINITION 52 Les n v.a.r. X_1, \dots, X_n sont dites **mutuellement indépendantes** si $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$, les événements $[X_i \leq x_i], i = 1, \dots, n$ sont mutuellement indépendants.

PROPOSITION 51 1. Soient X_1, \dots, X_n n v.a.r. mutuellement indépendantes de fonctions caractéristiques $\phi_{X_i}, i = 1, \dots, n$. La fonction caractéristique de la v.a.r. $S = \sum_{i=1}^n X_i$ est donnée pour tout réel t par

$$\phi_S(t) = \prod_{i=1}^n \phi_{X_i}(t),$$

2. Soient X_1, \dots, X_n n v.a.r. discrètes mutuellement indépendantes de fonctions génératrices $G_{X_i}, i = 1, \dots, n$. La fonction génératrice de la v.a.r. $S = \sum_{i=1}^n X_i$ est donnée pour tout $s \in [0, 1]$ par

$$G_S(s) = \prod_{i=1}^n G_{X_i}(s).$$

DÉMONSTRATION Ce résultat se démontre par récurrence en utilisant les résultats des propositions 43 et 44. \circ

COROLLAIRE 7 La v.a.r. définie comme la somme de n v.a.r. mutuellement indépendantes de même loi de Bernoulli de paramètre p suit une loi binomiale de paramètre (n, p) .

PROPOSITION 52 Soient X_1, \dots, X_n n v.a.r. mutuellement indépendantes.

1. Si pour $k \in \{1, \dots, n\}$, X_k suit une loi normale de paramètres (μ_k, σ_k^2) alors la v.a.r. définie comme la somme de ces n v.a.r. suit une loi normale de paramètres (μ, σ^2) où

$$\mu = \sum_{k=1}^n \mu_k \quad \text{et} \quad \sigma^2 = \sum_{k=1}^n \sigma_k^2.$$

2. Si les n v.a.r. suivent une même loi normale centrée réduite $\mathcal{N}(0, 1)$ alors la v.a.r. $T = \sum_{k=1}^n X_k^2$ suit la loi du Chi-Carré à n degrés de liberté.

PROPOSITION 53 (Identité de Wald¹) Soit $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ un espace probabilisé discret. Soit N une v.a.r. discrète définie sur Ω à valeurs dans \mathbb{N}^* de fonction génératrice G_N . Soient X_1, \dots, X_N , des v.a.r. définies sur Ω , indépendantes, toutes de même loi, de fonction génératrice commune G_X . On considère la v.a.r. S définie par

$$S : \omega \in \Omega \longmapsto S(\omega) = \sum_{i=1}^{N(\omega)} X_i(\omega).$$

La v.a.r. S admet pour fonction génératrice

$$\forall s \in [0, 1], \quad G_S(s) = G_N(G_X(s)).$$

DÉMONSTRATION Ce résultat est admis. ○

COROLLAIRE 8 L'espérance et la variance de la v.a.r. S sont donnés par

$$\mathbb{E}(S) = \mathbb{E}(X_i) \mathbb{E}(N),$$

et

$$\text{Var}(S) = \mathbb{E}(N) \text{Var}(X_i) + \text{Var}(N) \mathbb{E}(X_i)^2.$$

DÉMONSTRATION On a

$$\begin{aligned} \mathbb{E}(S) &= G'_S(1) \\ &= G'_N(G_{X_i}(1)) G'_{X_i}(1) \quad \text{dérivation d'une fonction composée} \\ &= G'_N(1) \mathbb{E}(X_i) \\ &= \mathbb{E}(N) \mathbb{E}(X_i). \end{aligned}$$

Puis,

$$\begin{aligned} \text{Var}(S) &= G''_S(1) + G'_S(1) - (G'_S(1))^2 \\ &= G''_S(1) + \mathbb{E}(S) - \mathbb{E}(S)^2. \end{aligned}$$

Or d'après l'identité de Wald,

$$G''_S(1) = G''_N(G_{X_i}(1)) G'_{X_i}(1)^2 + G'_N(G_{X_i}(1)) G''_{X_i}(1).$$

Il suffit d'appliquer le résultat précédent pour conclure après calculs et simplifications. ○

¹WALD, Abraham (1902, Kolozsvár (Hongrie) - 1950, Travancore (Inde)).

Chapitre 5

Théorèmes limites

Soit E un événement donné relatif à une expérience aléatoire ayant une probabilité $\mathbb{P}(E)$ inconnue. On souhaite obtenir une estimation de la valeur de $\mathbb{P}(E)$. Par exemple, on dispose d'un dé que l'on sait être truqué mais on ignore de quelle manière et on s'intéresse à la probabilité d'obtenir un six avec ce dé.

Supposons que l'on répète l'expérience aléatoire de manière indépendante un nombre N de fois et que la probabilité $\mathbb{P}(E)$ reste constante au cours des différentes répétitions de l'expérience. Pour $i \in \mathbb{N}, 1 \leq i \leq N$, on désigne par X_i la v.a.r. qui vaut 1 si l'événement E survient lors du i^{e} tirage et 0 sinon. Les v.a.r. X_i suivent une même loi de Bernoulli de paramètre $p = \mathbb{P}(E)$ et on a

$$\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = \mathbb{P}(E) = p \quad \forall i \in \{1, \dots, N\}.$$

On introduit la v.a.r. $\overline{X_N} = \frac{1}{N}(X_1 + \dots + X_N)$ appelée *moyenne empirique* de l'événement E . La v.a.r. $\overline{X_N}$ suit une loi binomiale de paramètre $(N, p/N)$ et on a $\mathbb{E}(\overline{X_N}) = p$. Intuitivement, si l'on répète un grand nombre de fois l'expérience, on s'attend à ce que $\overline{X_N}$ donne une approximation de $\mathbb{P}(E)$. Dans le cas de notre dé truqué, si après 1000 lancers on a observé 251 fois le six, l'intuition nous suggère que la probabilité d'obtenir un six avec ce dé est proche de 1/4.

Il est alors légitime de se poser un certain nombre de questions. Quelle est la précision du résultat si l'on approche $\mathbb{P}(E)$ par la valeur de $\overline{X_N}$ obtenue pour la série de répétitions de l'épreuve effectuée ? Est-on certain que la valeur de $\overline{X_N}$ obtenue converge vers $\mathbb{P}(E)$ lorsque N «tend vers l'infini» ? En quel sens ? etc ... Les théorèmes limites précisent les choses du point de vue mathématique et apportent des réponses à ces questions.

1 Convergences stochastiques

1.1 Convergence en loi

DÉFINITION 53 Soient X une v.a.r. de fonction de répartition F_X et $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. de fonctions de répartitions respectives $(F_{X_n})_{n \in \mathbb{N}}$. On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si pour tout réel x tel que F_X est continue en x , on a

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x).$$

On note alors

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X.$$

REMARQUE Contrairement aux autres modes de convergence stochastique (convergence en probabilité, convergence presque sûre, voir paragraphes suivants), il n'est pas nécessaire ici que les variables aléatoires soient définies sur un même espace probabilisé.

PROPOSITION 54 La suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si on a la relation suivante entre les fonctions caractéristiques ϕ_X et ϕ_{X_n} des v.a.r. X et X_n ,

$$\forall t \in \mathbb{R}, \quad \lim_{n \rightarrow +\infty} \phi_{X_n}(t) = \phi_X(t).$$

De manière équivalente la suite $(X_n)_{n \in \mathbb{N}}$ converge en loi vers X si et seulement si pour toute fonction h de \mathbb{R} dans \mathbb{R} , bornée et continue par morceaux,

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

DÉMONSTRATION Ce résultat est admis. ○

1.2 Convergence en probabilité

DÉFINITION 54 Soient X une v.a.r. et $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0.$$

On note

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X.$$

PROPOSITION 55 Si la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X en probabilité, alors la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X en loi.

DÉMONSTRATION Ce résultat est admis. ○

REMARQUE La réciproque est fausse en général. Elle n'est vraie que si X_n converge en loi vers une constante réelle.

1.3 Convergence presque sûre

DÉFINITION 55 Soient X une v.a.r. et $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Soit $A \in \mathcal{A}$ l'ensemble des éventualités $\omega \in \Omega$ telles que la suite numérique $(X_n(\omega))_{n \in \mathbb{N}}$ converge vers $X(\omega)$.

On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge presque sûrement vers X si $\mathbb{P}(A) = 1$. On note

$$X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X.$$

PROPOSITION 56 Si la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X presque sûrement, alors la suite $(X_n)_{n \in \mathbb{N}}$ converge vers X en probabilité.

2 Loi faible des grands nombres

Le théorème suivant appelé «loi faible des grands nombres» confirme, sous certaines hypothèses, qu'intuitivement ou expérimentalement, on observe que lorsque l'on répète un grand nombre de fois une expérience aléatoire, la fréquence d'apparition d'un événement est «proche» de sa probabilité.

THÉORÈME 6 Soient X_1, \dots, X_n n v.a.r. de même loi et non corrélées. On suppose que ces v.a.r. admettent une espérance m et une variance σ^2 finies.

La v.a.r. $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ converge en probabilité vers m , i.e.

$$\overline{X_n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} m.$$

DÉMONSTRATION On pose $S_n = \sum_{i=1}^n X_i$. Montrons que $\forall \epsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{S_n}{n} - m \right| > \epsilon \right) = 0.$$

Appliquons l'inégalité de Bienaymé-Tchebychev à la v.a.r. S_n/n :

$$\forall \epsilon > 0, \quad \mathbb{P} \left(\left| \frac{S_n}{n} - \mathbb{E} \left(\frac{S_n}{n} \right) \right| > \epsilon \right) \leq \frac{\text{Var} (S_n/n)}{\epsilon^2}.$$

Or

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{mn}{n} = m$$

et

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

car les v.a.r. sont non corrélées par hypothèse. On en déduit que

$$0 \leq \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Or

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow +\infty} \frac{\sigma^2}{n\epsilon^2} = 0,$$

d'où

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right| > \epsilon\right) = 0.$$

La v.a.r. $\overline{X_n} = \frac{S_n}{n}$ converge donc en probabilité vers m . ○

3 La loi forte des grands nombres

THÉORÈME 7 Soient X_1, \dots, X_n n v.a.r. de même loi et indépendantes. On suppose que ces v.a.r. admettent une espérance m finie.

La v.a.r. $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ converge presque sûrement vers m , i.e. $\exists \Omega_0 \subset \Omega$ avec $\mathbb{P}(\Omega_0^c) = 0$ tel que $\forall \omega \in \Omega_0$,

$$\lim_{n \rightarrow +\infty} \overline{X_n}(\omega) = m.$$

DÉMONSTRATION Ce résultat est admis. ○

Comparaison entre la loi forte des grands nombres et la loi faible des grands nombres

Ces 2 résultats nous assurent la convergence de la suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ vers X mais il n'y a pas convergence «au même sens». (Voir par exemple les séries de fonctions où on définit 4 types de convergence : simple, absolue, uniforme, normale.)

4 Théorème de la limite centrale

On rappelle que si $Y \sim \mathcal{N}(a, b)$, alors $\frac{Y - a}{\sqrt{b}} \sim \mathcal{N}(0, 1)$. D'autre part, si X_1, \dots, X_n sont n v.a.r. indépendantes de même loi $\mathcal{N}(m, \sigma^2)$, alors $S_n = \sum_{i=1}^n X_i$ suit une loi $\mathcal{N}(nm, n\sigma^2)$. Ce résultat peut se démontrer au moyen de la fonction caractéristique. On en déduit que

$$T_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\overline{X_n} - m}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Le théorème de la limite centrale indique que dans le cas où les X_n ne suivent pas une loi normale, le comportement de la somme «centrée réduite» T_n s'approche toutefois de plus en plus de la loi $\mathcal{N}(0, 1)$ quand n tend vers l'infini.

THÉORÈME 8 Soient X_1, \dots, X_n n v.a.r. indépendantes de même loi, de moyenne m et de variance σ^2 . La v.a.r. $S_n = \sum_{i=1}^n X_i$ vérifie,

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y,$$

où $Y \sim \mathcal{N}(0, 1)$.

DÉMONSTRATION Intéressons-nous à la fonction caractéristique de la v.a.r. S_n ,

$$\begin{aligned} \phi_{S_n}(t) &= \phi_{X_1}(t)^n \\ &= \mathbb{E}(e^{itX_1})^n \\ &= \left(\mathbb{E}\left(1 + itX_1 + \frac{i^2 t^2}{2} X_1^2 + o(t^2)\right) \right)^n \\ &= \left(1 + it\mathbb{E}(X_1) + \frac{i^2 t^2}{2} \mathbb{E}(X_1^2) + o(t^2)\right)^n. \end{aligned}$$

D'où par changement de variables et à t fixé, on a

$$\phi_{T_n}(t) = e^{-it\frac{\sqrt{nm}}{\sigma}} \left(1 + \frac{it}{\sigma\sqrt{n}} \mathbb{E}(X_1) - \frac{t^2}{2n\sigma^2} \mathbb{E}(X_1^2) + o\left(\frac{1}{n}\right)\right)^n,$$

on prend le logarithme de chaque membre,

$$\ln[\phi_{T_n}(t)] = -it\frac{\sqrt{nm}}{\sigma} + n \left(\frac{itm}{\sigma\sqrt{n}} - \frac{t^2}{2n\sigma^2} (\sigma^2 + m^2) + \frac{t^2m^2}{\sigma^2n} + o\left(\frac{1}{n}\right) \right),$$

et on constate que le second membre tend vers $-t^2/2$ quand $n \rightarrow +\infty$, d'où

$$\lim_{n \rightarrow +\infty} \phi_{T_n}(t) = e^{-t^2/2}$$

qui est la fonction caractéristique de la loi $\mathcal{N}(0, 1)$. Comme la fonction caractéristique caractérise la loi d'une v.a.r., le résultat est démontré. \circ

REMARQUE Le théorème de la limite centrale indique que la somme d'un grand nombre de v.a.r. indépendantes de même loi suit approximativement une loi normale. Il fournit un moyen simple pour le calcul approché de probabilités d'événements faisant intervenir une somme de v.a.r.. De plus il explique le fait empirique que bien des phénomènes naturels admettent une «distribution normale». Dans le cas d'une somme de v.a. de Bernoulli, le théorème de la limite centrale a été démontré par de Moivre¹ en 1733.

PROPOSITION 57

1. Soient X_1, \dots, X_n n v.a.r. indépendantes de loi de Bernoulli de paramètre p . La v.a.r. $S_n = \sum_{i=1}^n X_i$ suit une loi binomiale $\mathcal{B}(n, p)$ et la suite $(S_n)_{n \in \mathbb{N}}$ converge en loi vers une loi normale de paramètres $(np, np(1 - p))$.
2. Soient X_1, \dots, X_n n v.a.r. indépendantes de loi de Poisson de paramètre λ . La v.a.r. $S_n = \sum_{i=1}^n X_i$ suit une loi de Poisson de paramètre $n\lambda$ et converge en loi vers une loi normale de paramètres $(n\lambda, n\lambda)$.

REMARQUE En pratique, on approche une loi binomiale $\mathcal{B}(n, p)$ par une loi normale $\mathcal{N}(np, np(1 - p))$ dès que $np \approx 20$ et $np(1 - p) \approx 20$. On approche une loi de Poisson de paramètre λ par une loi normale $\mathcal{N}(n\lambda, n\lambda)$ dès que $n\lambda \approx 20$.

5 Application : estimation de la moyenne d'une variable aléatoire

La distribution exacte d'une variable X dans une population est généralement inconnue à moins que l'on ait une connaissance exhaustive de la population. Dans le cas d'une population de grande taille, il est impossible en pratique d'observer tous les individus de la population pour déterminer exactement la fonction densité de X . Ne connaissant pas la fonction densité de X , on ne peut calculer les caractéristiques principales de la variable X , en particulier son espérance et sa variance.

La seule information disponible sur une variable, en dehors des hypothèses théoriques, est fournie par des observations sur un échantillon pris dans cette population. Si les observations de la variable X sont faites au hasard et de manière indépendantes les unes des autres, l'échantillon devrait être assez représentatif de la population. Il est alors justifié de vouloir estimer les caractéristiques de la variable X dans la population par les caractéristiques de la variable X dans l'échantillon.

5.1 Estimations ponctuelles de la moyenne et de la variance

Considérons une variable X dont la moyenne μ et la variance σ^2 dans une population sont inconnues. Pour estimer ces valeurs, on fait n observations indépendantes x_1, \dots, x_n dans

¹DE MOIVRE, ABRAHAM (Vitry, 1667- Londres, 1754).

les mêmes conditions. La moyenne dans l'échantillon est

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

et la variance dans l'échantillon est

$$s_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2.$$

On utilise alors \bar{x} et s_X^2 comme valeurs estimées de μ et σ^2 dans la population.

Pour n assez grand, une telle estimation est justifiée par la loi des grands nombres. Désignons par X_1, \dots, X_n les variables aléatoires correspondant respectivement à l'observation sur le premier individu de l'échantillon, ..., sur le n^{e} individu de l'échantillon (de sorte que $X_1(\omega) = x_1, \dots, X_n(\omega) = x_n$). Les variables aléatoires X_1, \dots, X_n sont de même loi (celle de X), d'espérance commune μ et de variance commune σ^2 . Elles sont indépendantes si l'échantillon est correctement fabriqué. D'après la loi des grands nombres, la v.a.r. $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ converge presque sûrement et en probabilité vers μ .

En appliquant la loi des grands nombres à la variable \bar{X}^2 , on établit que la v.a.r. $\bar{X}_n^* = \frac{1}{n} \sum_{i=1}^n X_i^2$ converge presque sûrement et en probabilité vers μ^2 et que par conséquent $\bar{X}_n^* - \bar{X}_n^2$ converge presque sûrement et en probabilité vers σ^2 .

5.2 Estimations par intervalle de confiance de la moyenne

En estimant la moyenne μ dans une population par la moyenne \bar{x} dans un échantillon, on fait en général une erreur. La valeur estimée n'est qu'approximative. On a souvent recours à des estimations par intervalle : il s'agit de déterminer un intervalle $[\bar{x} - d, \bar{x} + d]$ autour de la valeur de la moyenne empirique \bar{x} de sorte qu'on puisse affirmer, avec un degré de confiance fixé, que la moyenne μ de la variable X dans la population se trouve dans l'intervalle $[\bar{x} - d, \bar{x} + d]$. Le paramètre d est appelé la *marge d'erreur*. La probabilité γ que μ appartienne à l'intervalle $[\bar{x} - d, \bar{x} + d]$ est appelée *seuil de confiance*.

On peut donc construire un intervalle de confiance pour μ de 2 manières.

1. Choisir une marge d'erreur d puis déterminer le seuil de confiance γ .
2. Choisir un seuil de confiance γ puis déterminer la marge d'erreur d correspondante.

La relation entre la marge d'erreur d et le seuil de confiance γ est

$$\mathbb{P}(|\bar{X} - \mu| \leq d) = \gamma.$$

On a

$$\begin{aligned} \gamma = \mathbb{P}(|\bar{X} - \mu| \leq d) &= \mathbb{P}\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{d}{\sigma/\sqrt{n}}\right) \\ &= \mathbb{P}\left(-\frac{d}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{d}{\sigma/\sqrt{n}}\right). \end{aligned}$$

D'après le théorème de la limite centrale, si n est grand, la variable aléatoire $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ converge en loi vers une variable aléatoire de loi normale centrée réduite. On a donc

$$\gamma \approx 2 F_{\mathcal{N}(0,1)} \left(\frac{d}{\sigma/\sqrt{n}} \right) - 1 \quad \text{et} \quad d \approx \frac{\sigma}{\sqrt{n}} F_{\mathcal{N}(0,1)}^{-1} \left(\frac{\gamma + 1}{2} \right),$$

où $F_{\mathcal{N}(0,1)}$ désigne la fonction de répartition de la loi normale centrée réduite.

Dans le cas où la variance σ^2 est inconnue, on l'estime par la variance empirique dans l'échantillon.

Chapitre 6

Chaînes de Markov discrètes

Très souvent quand on étudie un phénomène qui dépend du hasard, il y a lieu de prendre en compte l'évolution de ce phénomène au cours du temps. Le modèle probabiliste correspondant est appelé **processus stochastique**.

Nous nous intéressons dans ce chapitre aux processus pour lesquels l'espace des temps est un espace dénombrable (on parle alors de **chaîne de Markov**). Enfin, nous nous restreindrons aussi au cas où l'espace d'états de ces processus est dénombrable (et même fini!), c'est-à-dire le cas des chaînes de Markov discrètes.

1 Chaîne de Markov homogène

Dans ce chapitre E désigne un ensemble fini. Sans perte de généralité, nous supposerons que E est l'ensemble $\{1, \dots, N\}$. On note \mathcal{E} l'ensemble des parties de E . On se place sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

DÉFINITION 56 Une **chaîne de Markov** sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) est une suite de v.a.r. $(X_k)_{k \in \mathbb{N}}$ définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans (E, \mathcal{E}) vérifiant $\forall n \in \mathbb{N}$, $\forall (e_0, \dots, e_n, e_{n+1}) \in E^{n+2}$,

$$\mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n, \dots, X_0 = e_0) = \mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n).$$

Cette relation s'appelle «propriété de Markov». Elle traduit le fait que le futur du processus ne dépend du passé qu'à travers le présent ou encore comme l'indique la proposition 58 que conditionnellement au présent, futur et passé sont indépendants.

Pour $k \in \mathbb{N}$, X_k correspond à l'état de la chaîne de Markov à l'étape k (dans une unité de temps donnée). L'ensemble E est ainsi nommé **l'espace d'états**.

PROPOSITION 58 Une suite $(X_k)_{k \in \mathbb{N}}$ de v.a.r. sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) est une chaîne de Markov si et seulement si $\forall n \in \mathbb{N}$, $\forall (e_0, \dots, e_{n+1}) \in E^{n+2}$,

$$\begin{aligned} \mathbb{P}(X_{n+1} = e_{n+1}, X_{n-1} = e_{n-1}, \dots, X_0 = e_0 \mid X_n = e_n) &= \mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n) \\ &\quad \times \mathbb{P}(X_{n-1} = e_{n-1}, \dots, X_0 = e_0 \mid X_n = e_n). \end{aligned}$$

DÉMONSTRATION On a

$$\begin{aligned}
 & \mathbb{P}(X_{n+1} = e_{n+1}, X_n = e_n, \dots, X_0 = e_0) \\
 &= \mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n, \dots, X_0 = e_0) \times \mathbb{P}(X_n = e_n, \dots, X_0 = e_0) \\
 &= \mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n) \times \mathbb{P}(X_n = e_n, \dots, X_0 = e_0) \quad \text{d'après la définition 56} \\
 &= \mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n) \times \mathbb{P}(X_{n-1} = e_{n-1}, \dots, X_0 = e_0 \mid X_n = e_n) \times \mathbb{P}(X_n = e_n).
 \end{aligned}$$

D'où

$$\begin{aligned}
 & \mathbb{P}(X_{n+1} = e_{n+1}, X_{n-1} = e_{n-1}, \dots, X_0 = e_0 \mid X_n = e_n) \\
 &= \frac{\mathbb{P}(X_{n+1} = e_{n+1}, X_n = e_n, \dots, X_0 = e_0)}{\mathbb{P}(X_n = e_n)} \\
 &= \mathbb{P}(X_{n+1} = e_{n+1} \mid X_n = e_n) \times \mathbb{P}(X_{n-1} = e_{n-1}, \dots, X_0 = e_0 \mid X_n = e_n).
 \end{aligned}$$

○

Pour $(m, n) \in \mathbb{N}^2$, $n > m$, les probabilités conditionnelles $P(X_n = e_n \mid X_m = e_m)$ sont appelées **probabilités de transition** ou **probabilités de passage** en $n-m$ étapes. Elles donnent la probabilité que le système passe de l'état e_m à l'état e_n pendant l'intervalle de temps de m à n .

DÉFINITION 57 On dit qu'une chaîne $X = (X_k)_{k \in \mathbb{N}}$ est une **chaîne de Markov homogène** si les probabilités de transition en ℓ étapes sont les mêmes indépendamment de l'instant n de la première transition. Autrement dit si $\forall n \in \mathbb{N}, \forall \ell \in \mathbb{N}, \ell < n, \forall (i, j) \in E^2$

$$\mathbb{P}(X_n = j \mid X_{n-\ell} = i) = P(X_\ell = j \mid X_0 = i).$$

REMARQUE Pour $\ell = 1$, on obtient en particulier que $\forall n \in \mathbb{N}^*, \forall (i, j) \in E^2$

$$\mathbb{P}(X_n = j \mid X_{n-1} = i) = \mathbb{P}(X_1 = j \mid X_0 = i).$$

DÉFINITION 58 On appelle **matrice de transition** la matrice $G = (G_{ij})_{i,j \in E} \in \mathcal{M}_N(\mathbb{R})$ d'éléments

$$G_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i), \quad i, j = 1, \dots, N.$$

PROPOSITION 59 La matrice de transition G est une matrice stochastique, c'est-à-dire une matrice dont les éléments vérifient les 2 propriétés suivantes :

1. $G_{ij} \geq 0 \quad \forall (i, j) \in E^2$;
2. $\forall i \in E, \sum_{j \in E} G_{ij} = 1$ (autrement dit la somme des éléments d'une ligne de la matrice vaut 1).

DÉMONSTRATION Ces propriétés sont évidentes, puisque par définition les éléments G_{ij} sont des probabilités et que

$$\forall i \in E, \quad \sum_{j \in E} G_{ij} = \sum_{j \in E} \mathbb{P}(X_1 = j \mid X_0 = i) = \sum_{j \in E} \mathbb{P}_{[X_0=i]}(X_1 = j)$$

où on a noté $\mathbb{P}_{[X_0=i]}$ la probabilité conditionnelle sachant l'événement $[X_0 = i]$. Puisque les événements $[X_1 = j], j \in E$ sont disjoints, on a

$$\sum_{j \in E} G_{ij} = \mathbb{P}_{[X_0=i]} \left(\bigcup_{j \in E} [X_1 = j] \right) = \mathbb{P}_{[X_0=i]}(\Omega) = 1.$$

○

PROPOSITION 60 Soit $X = (X_k)_{k \in \mathbb{N}}$ une chaîne de Markov homogène de matrice de transition G . Pour $n \in \mathbb{N}$, on désigne par G^n la puissance n^e de la matrice G , c'est-à-dire $G^n = \underbrace{G \times G \times \dots \times G}_{n \text{ fois}}$. La probabilité de transition en n étapes de l'état i à l'état j est donnée par,

$$P(X_n = j \mid X_0 = i) = G_{ij}^n,$$

où G_{ij}^n désigne l'élément d'indices (i, j) de la matrice G^n .

REMARQUE Par commodité, on a noté

$$G_{ij}^n = (G^n)_{ij}$$

l'élément d'indices (i, j) de la matrice G^n produit de la matrice G par elle-même n fois. Il ne faut pas confondre cette quantité avec $(G_{ij})^n$ le produit n fois avec lui-même de l'élément G_{ij} d'indices (i, j) de la matrice G , qui dans le cas général est différent.

DÉMONSTRATION On utilise un raisonnement par récurrence. Pour $n = 1$ c'est la définition. Supposons que $\mathbb{P}(X_n = k \mid X_0 = i) = G_{ik}^n$ pour $i, k \in E$. Calculons,

$$\begin{aligned} \mathbb{P}(X_{n+1} = j \mid X_0 = i) &= \frac{\mathbb{P}(X_{n+1} = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\ &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in E} \mathbb{P}(X_{n+1} = j, X_0 = i, X_n = k) \\ &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in E} \mathbb{P}(X_{n+1} = j \mid X_0 = i, X_n = k) \\ &\quad \times \mathbb{P}(X_0 = i, X_n = k). \end{aligned}$$

Or puisque $(X_k)_{k \in \mathbb{N}}$ est une chaîne de Markov

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i, X_n = k) = \mathbb{P}(X_{n+1} = j \mid X_n = k),$$

donc

$$\begin{aligned}
 \mathbb{P}(X_{n+1} = j \mid X_0 = i) &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in E} \mathbb{P}(X_{n+1} = j \mid X_n = k) \times \mathbb{P}(X_0 = i, X_n = k) \\
 &= \sum_{k \in E} \mathbb{P}(X_{n+1} = j \mid X_n = k) \times \mathbb{P}(X_n = k \mid X_0 = i) \\
 &= \sum_{k \in E} G_{ik}^n G_{kj}.
 \end{aligned}$$

Cette dernière relation correspond à l'élément d'indice (i, j) de la matrice définie comme le produit $G^n \times G$, voir un cours d'algèbre linéaire de premier cycle. On a donc

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i) = G_{ij}^{n+1}.$$

○

EXEMPLE On considère 2 urnes et 4 boules réparties dans les 2 urnes. L'expérience aléatoire consiste à déplacer une boule d'une des urnes, la boule étant choisie au hasard uniformément parmi les 4, dans l'autre urne. Une chaîne est définie par la famille de v.a.r. $X_k : \Omega \rightarrow E$ où $E = \{0, 1, 2, 3, 4\}$ et $X_k(\omega)$ est le nombre de boules dans la première urne à l'étape k .

G_{ij} est la probabilité d'avoir $(j-1)$ boules dans la première urne sachant qu'avant l'échange des boules il y en avait $(i-1)$. On a $G_{11} = 0$ car la probabilité d'avoir 0 boule dans l'urne sachant qu'il n'y en avait aucune auparavant est nulle. Lors de l'échange, la boule a forcément été prise dans la seconde urne (qui seule contenait des boules) pour être mise dans la première (qui contenait alors nécessairement exactement une boule). Par suite $G_{12} = 1, G_{13} = G_{14} = G_{15} = 0$.

De même $G_{21} = \frac{1}{4}$. C'est la probabilité d'avoir 0 boule dans la première urne sachant qu'il y en avait une auparavant, c'est-à-dire la probabilité de choisir la boule de la première urne parmi les 4 boules qu'il y a en tout.

On vérifie que l'on a pour matrice de transition

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

EXEMPLE Pierre et Jacques jouent avec un dé à 6 faces. Pierre dispose de a euros au départ et Jacques de b euros. A chaque lancer Pierre donne un euro à Jacques sauf si le dé tombe sur la face six. Si un six est obtenu, Jacques donne 5 euros à Pierre (ou la totalité de sa fortune lorsque celle-ci est strictement inférieure à 5 euros). Le jeu s'arrête lorsque la mise initiale est épuisée. On s'intéresse à l'évolution de la fortune de Pierre au cours du jeu : on note X_k la somme dont dispose Pierre après le k^{e} lancer.

La suite $(X_k)_{k \in \mathbb{N}}$ définit une chaîne de Markov dont l'espace des états est $E = \{0, \dots, a +$

$b\}$. Soit Y_k la v.a.r. représentant le gain théorique de Pierre au cours du k^{e} lancer (c'est-à-dire que le gain réel peut être moindre si Jacques est presque ruiné!). On a $\mathbb{P}(Y_k = -1) = 5/6$ et $\mathbb{P}(Y_k = 5) = 1/6$. Ainsi,

$$\begin{aligned}\mathbb{P}(X_{n+1} = i-1 \mid X_n = i) &= \mathbb{P}(Y_{n+1} = -1) = 5/6, \\ \mathbb{P}(X_{n+1} = i+5 \mid X_n = i) &= \mathbb{P}(Y_{n+1} = 5) = 1/6 \text{ pour } i \leq a+b-5, \\ \mathbb{P}(X_{n+1} = a+b \mid X_n = i) &= \mathbb{P}(Y_{n+1} = 5) = 1/6 \text{ pour } i > a+b-5, \\ \mathbb{P}(X_{n+1} = 0 \mid X_n = 0) &= \mathbb{P}(X_{n+1} = a+b \mid X_n = a+b) = 1,\end{aligned}$$

les autres probabilités de transition étant nulles.

Si l'on prend $a = b = 5$, on obtient la matrice de transition suivante,

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 5/6 & 0 & 0 & 0 & 0 & 0 & 1/6 & 0 & 0 & 0 \\ 0 & 5/6 & 0 & 0 & 0 & 0 & 0 & 1/6 & 0 & 0 \\ 0 & 0 & 5/6 & 0 & 0 & 0 & 0 & 0 & 1/6 & 0 \\ 0 & 0 & 0 & 5/6 & 0 & 0 & 0 & 0 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 0 & 0 & 0 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 5/6 & 0 & 0 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5/6 & 0 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5/6 & 0 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5/6 & 1 \end{pmatrix}$$

2 Caractérisation d'une chaîne de Markov

Nous allons montrer qu'une chaîne de Markov homogène est caractérisée par la donnée d'une loi initiale (correspondant à la loi de la v.a.r. X_0) et d'une matrice stochastique G .

PROPOSITION 61 (équation de Chapman¹-Kolmogorov)

Soit $X = (X_k)_{k \in \mathbb{N}}$ une chaîne de Markov homogène d'espace d'états E . $\forall i, j \in E, \forall n, m \in \mathbb{N}$, on a,

$$\mathbb{P}(X_{m+n} = j \mid X_0 = i) = \sum_{k \in E} \mathbb{P}(X_m = j \mid X_0 = k) \times \mathbb{P}(X_n = k \mid X_0 = i).$$

DÉMONSTRATION On a

$$\begin{aligned}\mathbb{P}(X_{n+m} = j \mid X_0 = i) &= \frac{\mathbb{P}(X_{n+m} = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\ &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in E} \mathbb{P}(X_{n+m} = j, X_0 = i, X_n = k)\end{aligned}$$

¹CHAPMAN, Sydney (1888, Eccles (GB) - 1970, Boulder (USA)). .

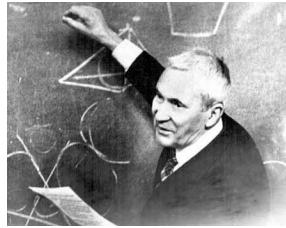
$$\begin{aligned}
 &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in E} \mathbb{P}(X_{n+m} = j \mid X_0 = i, X_n = k) \\
 &\quad \times \mathbb{P}(X_0 = i, X_n = k) \\
 &= \frac{1}{\mathbb{P}(X_0 = i)} \sum_{k \in E} \mathbb{P}(X_{n+m} = j \mid X_n = k) \times \mathbb{P}(X_0 = i, X_n = k) \\
 &= \sum_{k \in E} \mathbb{P}(X_m = j \mid X_0 = k) \times \mathbb{P}(X_n = k \mid X_0 = i).
 \end{aligned}$$

La dernière égalité est obtenue en utilisant le fait que $(X_k)_{k \in \mathbb{N}}$ est une chaîne de Markov homogène et donc que

$$\mathbb{P}(X_{n+m} = j \mid X_0 = i, X_n = k) = \mathbb{P}(X_{n+m} = j \mid X_n = k) = \mathbb{P}(X_m = j \mid X_0 = k).$$

○

KOLMOGOROV, Andreï Nicolaiévitch (1903, Tambov (Russie) - 1987, Moscou).



Professeur à l'université de Moscou, il est principalement connu pour avoir fondé une théorie axiomatique des probabilités (1933) à partir d'axiomes simples définissant une probabilité et faisant usage du concept de tribu (définies par Borel) et des théories récentes de la mesure et du calcul intégral au sens de Lebesgue. Ses travaux portèrent également sur l'analyse de Fourier et les systèmes dynamiques. Pour tout savoir sur Kolmogorov : <http://www.kolmogorov.com/>.

À ce stade, seules des probabilités conditionnelles ont été considérées. Ce sont les probabilités de transition d'un état vers un autre état dans l'espace d'états E en une ou plusieurs étapes. Si l'on désire connaître la loi de la variable X_n pour un entier n donné, il est nécessaire de préciser la loi de la variable X_0 . Cette variable est appelée condition initiale de la chaîne.

On désigne par Π la fonction de masse de la v.a.r. discrète X_0 ,

$$\begin{aligned}
 \Pi : E &\longrightarrow [0, 1] \\
 k &\longmapsto \pi_k = \mathbb{P}(X_0 = k)
 \end{aligned}$$

On définit le vecteur ligne $\pi = (\pi_1, \dots, \pi_N)$.

PROPOSITION 62 Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène de matrice de transition G . Pour tout entier n , la fonction de masse de la v.a.r. discrète X_n est l'application $\Pi^{(n)}$ définie par

$$\begin{aligned}
 \Pi^{(n)} : E &\longrightarrow [0, 1] \\
 k &\longmapsto \pi_k^{(n)}
 \end{aligned}$$

où $\pi_k^{(n)}$ est la k^e composante du vecteur ligne $\pi^{(n)}$ obtenu en effectuant le produit du vecteur ligne π par la matrice G^n . Autrement dit $\pi^{(n)} = \pi G^n$, et $\pi^{(0)} = \pi$.

DÉMONSTRATION Pour $i \in E$, en utilisant la formule de Bayes, on a en effet,

$$\begin{aligned}\mathbb{P}(X_n = i) &= \sum_{k \in E} \mathbb{P}(X_n = i \mid X_0 = k) \mathbb{P}(X_0 = k) \\ &= \sum_{k \in E} G_{ki}^n \pi_k.\end{aligned}$$

Cette dernière égalité représente le produit du vecteur ligne π par la matrice G^n , voir [Balac-Sturm]. \circ

Montrons maintenant que la donnée d'une matrice de transition G et de la loi π de la condition initiale X_0 permet de caractériser une chaîne de Markov homogène $X = (X_n)_{n \in \mathbb{N}}$, c'est-à-dire de déterminer la loi des vecteurs (X_0, \dots, X_n) pour tout entier n .

PROPOSITION 63 *Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène sur un espace d'états E , de matrice de transition G et de condition initiale X_0 ayant pour fonction de masse l'application Π de E dans $[0, 1]$.*

Pour tout entier n et pour tout $(n+1)$ -uplet $(e_0, \dots, e_n) \in E^{n+1}$, on a

$$\mathbb{P}(X_0 = e_0, \dots, X_n = e_n) = \pi_{e_0} G_{e_0, e_1} G_{e_1, e_2} \dots G_{e_{n-1}, e_n}.$$

DÉMONSTRATION Le résultat se démontre par récurrence sur n . Pour $n = 0$, l'égalité correspond à la définition de la fonction de masse Π . Supposons donc la relation vraie pour un entier k donné. On peut alors écrire,

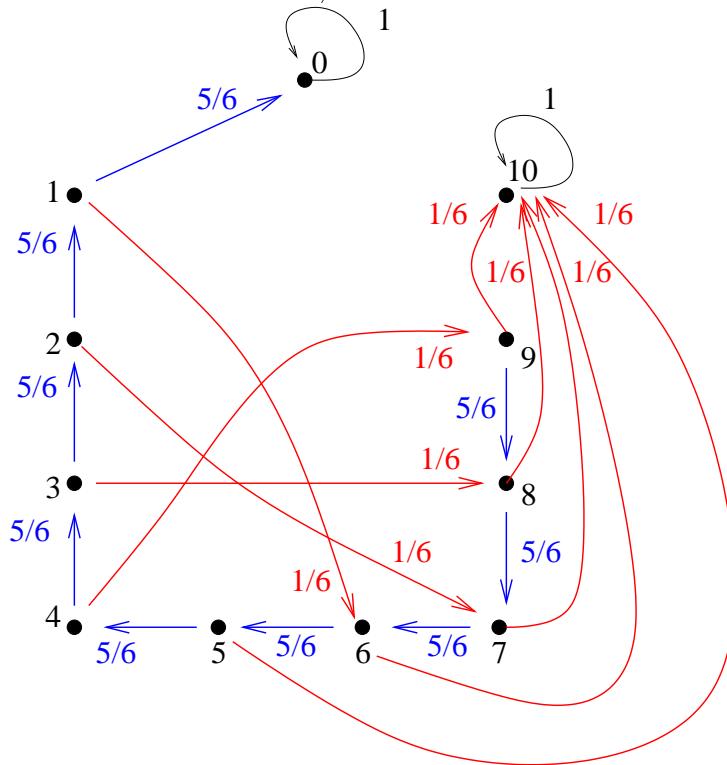
$$\begin{aligned}\mathbb{P}(X_0 = e_0, \dots, X_{k+1} = e_{k+1}) &= \mathbb{P}(X_{k+1} = e_{k+1} \mid X_0 = e_0, \dots, X_k = e_k) \\ &\quad \times \mathbb{P}(X_0 = e_0, \dots, X_k = e_k) \\ &= \mathbb{P}(X_{k+1} = e_{k+1} \mid X_k = e_k) \mathbb{P}(X_0 = e_0, \dots, X_k = e_k) \\ &= G_{e_k, e_{k+1}} \mathbb{P}(X_0 = e_0, \dots, X_k = e_k) \\ &= G_{e_k, e_{k+1}} G_{e_{k-1}, e_k} \dots G_{e_0, e_1} \pi_{e_0}.\end{aligned}$$

\circ

3 Classification des états d'une chaîne de Markov

Dans la définition que nous avons donnée d'une chaîne de Markov, l'évolution du processus au cours du temps à partir d'un état donné est entièrement décrite par la matrice des probabilités de transition. On peut aussi voir une chaîne de Markov comme un ensemble d'états entre lesquels s'effectuent des transitions. Certaines transitions sont possibles (probabilité de transition strictement positive) alors que d'autres sont impossibles (probabilité de transition nulle). Ceci nous amène à vouloir visualiser une chaîne de Markov en représentant chaque état par un sommet et chaque transition par un arc. Ce schéma est appelé

diagramme sagittal. Dans le cas de l'exemple précédent du lancer de dé, le diagramme sagittal associé à la chaîne est le suivant,



La probabilité que le système soit dans l'état i au temps t et dans l'état j au temps $t+n$:

$$G_{ij}^n = \mathbb{P}(X_{t+n} = j \mid X_t = i) = \mathbb{P}(X_n = j \mid X_0 = i).$$

On dit que l'état j est accessible à partir de l'état i et on note $i \rightarrow j$, si la probabilité de passer de i à j est non nulle :

$$\exists n \in \mathbb{N}, G_{ij}^n > 0.$$

Ceci signifie qu'il existe un chemin entre i et j . On dit que les états i et j communiquent et on note $i \leftrightarrow j$, si chacun d'eux est accessible à partir de l'autre : $i \rightarrow j$ et $j \rightarrow i$. La relation de communication entre deux états est réflexive, symétrique (par définition) et transitive, c'est donc une relation d'équivalence. Il est ainsi possible de construire une partition des états d'une chaîne de Markov en classes d'équivalence telle que tous les états d'une classe communiquent entre eux et que deux états appartenant à deux classes différentes ne communiquent pas. Une classe est dite transitoire s'il est possible d'en sortir mais dans ce cas, le processus ne pourra plus jamais y revenir. Une classe est dite récurrente s'il est impossible de la quitter. Si une classe récurrente est composée d'un seul état, cet état est dit absorbant. Un état i absorbant est donc tel qu'une fois dans cet état on ne peut le quitter.

En terme de probabilités de transition, ceci signifie que $\forall k \neq i, G_{ik} = 0$ et donc $G_{ii} = 1$. Les états absorbants sont très particuliers puisqu'ils constituent des états terminaux du système. Il est intéressant d'étudier les probabilités d'absorption, c'est-à-dire les probabilités que le système finisse par atteindre un tel état. Les états d'une classe transitoire sont appelés états transitoires alors que les états d'une classe récurrente sont appelés états

récurrents. Un état absorbant est donc un type particulier d'état récurrent. Une chaîne de Markov pour laquelle il n'existe qu'une seule classe (forcément récurrente) est dite **irréductible**. Ceci signifie que tous les états communiquent,

$$\forall (i, j) \in E^2, \exists n \in \mathbb{N}^*, G_{ij}^n > 0.$$

Une chaîne de Markov est **apériodique** si

$$\forall i \in E, \text{PGDC}\left(\left\{n \in \mathbb{N}, G_{ii}^n > 0\right\}\right) = 1.$$

Lorsque la chaîne est irréductible et apériodique on dit que la chaîne est **ergodique**.

4 Comportement asymptotique d'une chaîne de Markov

Après l'étude du comportement d'une chaîne de Markov homogène en n transitions, nous désirons étudier le comportement asymptotique de celle-ci lorsque l'entier n tend vers $+\infty$. Considérons une chaîne de Markov homogène $X = (X_n)_{n \in \mathbb{N}}$ sur un espace d'états E , de matrice de transition G et de condition initiale X_0 ayant pour fonction de masse Π . Nous avons établi à la proposition 62 que la loi de la v.a.r. discrète X_n s'obtenait à partir du vecteur π et de la matrice G par la relation de récurrence,

$$\pi^{(n)} = \pi^{(n-1)} G, \quad \pi^{(0)} = \pi.$$

DÉFINITION 59 On appelle **loi de probabilité invariante** de la chaîne de Markov homogène $X = (X_n)_{n \in \mathbb{N}}$ une fonction de masse $\Xi : k \in E \mapsto \xi_k \in [0, 1]$ où le vecteur $\xi = (\xi_1, \dots, \xi_N)$ est solution du système linéaire

$$\xi = \xi G.$$

Une telle loi existe toujours. La matrice de transition qui est une matrice stochastique admet la valeur propre 1 et le vecteur dont toutes les composantes sont égales à 1 est le vecteur propre à droite associé à cette valeur propre. Le vecteur ξ associé à une loi de probabilité invariante Ξ est donc vecteur propre à gauche associé à la valeur propre 1. L'équation $\xi = \xi G$ se formule de manière équivalente en disant que ξ^T est vecteur propre de G^T associé à la valeur propre 1. Lorsqu'une loi de probabilité invariante Ξ est choisie comme loi pour la variable X_0 , on obtient alors

$$\xi^{(1)} = \xi G = \xi, \quad \xi^{(2)} = \xi^{(1)} G = \xi G = \xi,$$

puis par récurrence, pour tout entier n ,

$$\xi^{(n)} = \xi^{(n-1)} G = \xi G = \xi.$$

Ceci signifie que la loi de la v.a.r. X_n pour tout entier n est la fonction de masse Ξ .

Nous cherchons désormais des conditions portant sur la matrice de transition G sous lesquelles la loi invariante Ξ est unique. Autrement dit, on cherche à quelles conditions l'équation $\xi = \xi G$ admet une unique solution.

THÉORÈME 9 Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène, irréductible et apé-riodique, sur un espace d'états E , de matrice de transition G .

1. Il existe une unique loi de probabilité Ξ invariante. De plus, $\forall i \in E$, $\xi_i > 0$.
2. Pour tout $(i, j) \in E^2$, on a

$$\lim_{n \rightarrow +\infty} G_{ij}^n = \xi_j.$$

3. Quelle que soit la loi de X_0 , la suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la loi de probabilité invariante Ξ .

REMARQUE D'après le théorème 9, on a

$$\lim_{n \rightarrow +\infty} \mathbb{P}(X_{m+n} = j \mid X_n = i) = \xi_j.$$

Cette limite est donc indépendante de i . Ceci nous amène à faire l'approximation suivante lorsque n est grand

$$\mathbb{P}(X_{m+n} = j \mid X_n = i) \approx \mathbb{P}(X_{m+n} = j).$$

En d'autres termes, nous pouvons considérer que deux états de la chaîne séparés par une longue histoire sont indépendants. On peut démontrer que la limite est atteinte à une vitesse exponentielle et cette vitesse de convergence se caractérise par le coefficient

$$\rho = \max \left\{ |\lambda|, \lambda \in \text{Sp}(G) \text{ et } \lambda \neq 1 \right\}.$$

Annexes

Lois discrètes classiques

Nom de la loi	Ensemble E des valeurs	Densité $\mathbb{P}(X = k), k \in E$.	Espérance	Variance
Loi uniforme \mathcal{U}_N $N \in \mathbb{N}^*$	$\{1, \dots, N\}$	$1/N$	$(N + 1)/2$	$(N^2 - 1)/12$
Loi de Bernoulli $\mathcal{B}(p)$ $p \in]0, 1[$	$\{0, 1\}$	$\begin{aligned}\mathbb{P}(X = 1) &= p \\ \mathbb{P}(X = 0) &= 1 - p\end{aligned}$	p	$p(1 - p)$
Loi binomiale $\mathcal{B}(n, p)$ $n \in \mathbb{N}^*$ $p \in]0, 1[$	$\{0, 1, \dots, n\}$	$C_n^k p^k (1 - p)^{n-k}$	np	$np(1 - p)$
Loi géométrique $\mathcal{G}(p)$ $p \in]0, 1[$	\mathbb{N}^*	$p(1 - p)^{k-1}$	$1/p$	$(1 - p)/p^2$
Loi de Poisson $\mathcal{P}(\lambda)$ $\lambda > 0$	\mathbb{N}	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
Loi binomiale négative $\mathcal{J}(r, p)$ $n \in \mathbb{N}^*$ $p \in]0, 1[$	\mathbb{N}	$C_{k+r-1}^k p^r (1 - p)^k$	$r(1 - p)/p$	$r(1 - p)/p^2$
Loi hypergéométrique $\mathcal{H}(N, n, p)$ $N, n \in \mathbb{N}^*$ $p \in]0, 1[$	$\{0, \dots, n\}$	$\frac{C_{Np}^k C_{N(1-p)}^{n-k}}{C_N^n}$	np	$np(1 - p) \frac{N - n}{N - 1}$
Loi de Pascal $\mathcal{P}(r, p)$	$k \in \mathbb{N}$ avec $k \geq r$	$C_{k-1}^{r-1} p^r (1 - p)^{k-r}$	r/p	$r(1 - p)/p^2$

Lois continues classiques

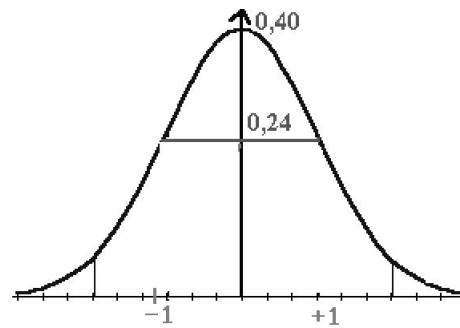
Nom de la loi	Ensemble E des valeurs	Densité $f(x)$ si $x \in E$ 0 si $x \notin E$	Espérance	Variance
Loi Uniforme $\mathcal{U}([a, b])$ $a < b$	$[a, b]$	$\frac{1}{b - a}$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Loi exponentielle $\mathcal{E}(\lambda)$ $\lambda > 0$	\mathbb{R}^+	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Loi normale $\mathcal{N}(m, \sigma^2)$	\mathbb{R}	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right)$	m	σ^2

Table de la fonction de répartition de la loi normale

Si X est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0, 1)$, la table donne la valeur de la fonction de répartition de X en x , $F(x) = \mathbb{P}(X \leq x)$.

La valeur de x s'obtient par addition des nombres inscrits en marge.

Pour $x < 0$, on a $F(x) = 1 - F(|x|)$.



x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7793	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

Bibliographie

1. BALAC, S. & STURM, F. Algèbre et analyse, cours de mathématiques première année avec exercices corrigés. *Presses Polytechniques et Universitaires Romandes*, coll. INSA de Lyon, 2003.
2. BRIANE, M. & PAGÈS, G. Théorie de l'intégration, licence de mathématiques. *Vuibert*, 2000
3. CYGANOWSKI, S., KLOEDEN, P. & OMBACH, J. From elementary probability to stochastic differential equations with Maple. *Springer Verlag Universitext*, 2001.
4. DELMAS, J.P. Probabilités et télécommunications, exercices et problèmes. commentés, *Masson*, coll. *Enseignement de la Physique*, 1987.
5. DELMER, F. Les séries, cours de mathématiques. *Dunod*, 1995.
6. GAPAILLARD, J. Intégration pour la licence. *Masson*, 1997.
7. GRINSTEAD, C. & SNELL, L. Introduction to Probability. *American Mathematical Society*, 1997.
8. MCGEE, V. Principes de statistiques. *Vuibert*, 1975.
9. ROSS, S. Initiation aux probabilités. *Presses Polytechniques et Universitaires Romandes*, 1987.

Index

arrangement avec répétition, 3
arrangement sans répétition, 2

Bayes
 formule de, 22
 notice biographique, 22

Bienaymé
 notice biographique, 46
Bienaymé-Tchebychev, inégalité, 46

Cauchy
 notice biographique, 37

chaîne de Markov, 83
chaîne de Markov
 apériodique, 91
 homogène, 84
 irréductible, 91

Chapman
 notice biographique, 87
Chapman-Komogorov, équation de, 87

combinaison avec répétition, 4
combinaison sans répétition, 2

convergence
 en loi, 76
 en probabilité, 76
 presque sûre, 77

corrélation, 63
 coefficients, 63

couple de v.a.r.
 discret, 56
 espérance, 57
 fonction de masse, 56
 fonction de répartition, 55

covariance, 58
 matrice de, 71

dénombrable, 8
dénombrément, 1
densité, 31
 conditionnelle, 60

écart-type, 43
équiprobabilité, 15
espérance, 38
 de gain, 39
espace d'états, 83
espace probabilisé, 14
événement, 13
 élémentaire, 13
 certain, 14, 15
 espace d', 14

fonction
 caractéristique, 50
 de masse, 30
 de répartition, 28
 conditionnelle, 60
 générateur, 51
fonction caractéristique
 vecteur aléatoire, 59, 72

fonction génératrice
 vecteur aléatoire, 59, 72
fonction indicatrice, 32

hasard, 5

indépendance
 variables aléatoires, 61, 72

indépendants
 événements, 24
indicatrice, 32
issue, 13

Kolmogorov
 notice biographique, 88

limite centrale, théorème, 79
loi, 32
 binomiale, 33, 53, 73
 conditionnelle
 cas continu, 60
 cas discret, 60

de Bernoulli, 33, 43
de Cauchy, 36, 54
de Poisson, 34, 53, 67
exponentielle, 36
géométrique, 34, 44
Gamma, 37
marginale d'un couple de v.a.r, 56
marginale d'un vecteur aléatoire, 69
multinomiale, 69
normale, 35, 44, 53
 n-dimensionnelle, 70
uniforme, 34, 35, 43, 44, 48, 50
loi faible des grands nombres, 77
loi forte des grands nombres, 78

Markov
 inégalité, 45
 notice biographique, 45
 chaîne de, 83
matrice
 de transition, 84
 stochastique, 84
modélisation, 5
modèle probabiliste, 5, 13
moment, 43
 centré, 43
moyenne
 empirique, 75
 vecteur aléatoire, 70

partition, 22
permutation, 1
permutation avec répétition, 4
Poincaré (formule), 18
Poisson
 notice biographique, 34
probabilité
 conditionnelle, 21
processus stochastique, 83
produit de convolution, 65

Tchebychev
 notice biographique, 47
tirage, 6
 avec ordre, 7
 avec remise, 7
univers, 13

variable aléatoire, 27
 centrée réduite, 42
 conditionnelle, 59
 continue, 30, 31
 discrète, 30
 réelle, 28
 somme, 64
variance, 41
vecteur aléatoire, 55, 67
 discret, 68
 fonction de masse, 68
 fonction de répartition, 68

Wald
 identité, 74
 notice biographique, 74

