Ce document contient les principales notions qui seront vues en cours, mais ne correspond toutefois pas exactement au cours (certains exemples, corrections d'exercices ou démonstrations sont laissés en blanc et ne seront traités qu'en cours) : la présence en cours est donc indispensable pour une bonne acquisition des notions que nous allons aborder.

Introduction

La théorie des Probabilités fournit des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec certitude. Comme premiers exemples, on peut penser au résultat d'un lancer d'un dé, d'une rencontre sportive, à l'évolution du cours d'une action en bourse. Plus généralement, la théorie des Probabilités est un outil mathématique qui intervient dans de nombreux domaines, comme la biologie, la physique, l'économie, la théorie de l'information (ex. : fonctionnement des algorithmes de recherche de Google), et bien sûr les statistiques.

Les Probabilités ne servent cependant pas qu'aux applications. C'est une branche des mathématiques modernes à part entière, qui s'est beaucoup développée depuis quelques dizaines d'années et qui a vu ses chercheurs récompensés à plusieurs reprises par la Médaille Fields. La dernière en date étant cet été 2022, où le probabiliste français Hugo Duminil-Copin a reçu cette prestigieuse récompense.

L'objectif de ce cours est d'aborder la formalisation mathématique qui s'est développée depuis un peu moins d'un siècle pour cette science du hasard. On restera la majeure partie de ce semestre dans un cadre relativement simple, celui des *Probabilités discrètes*. La théorie des Probabilités dans son cadre le plus complet repose sur la théorie de la Mesure et de l'Intégration, qui sera abordée en L3.

Nous aurons besoin d'outils mathématiques tels que la théorie des ensembles, le dénombrement, les séries et les intégrales (ces deux dernières notions seront développées en détails dans le cours d'Analyse 3).

****

C'est un cours qui demande un **travail régulier et profond** : lecture, relecture, apprentissage du cours et résolution d'exercices. Bon travail !

Quelques références pour approfondir et compléter le cours :

- Kortchemski-Mansuy : Probabilités, Prépas scientifiques 1ère et 2ème années. Ce livre contient un résumé du cours et des exercices corrigés.
- Grimmett-Stirzaker : Probability and Random processes (niveau L1 à M1 – on n'abordera ici que les 3 premiers chapitres). Il y a 2 tomes : un tome de cours et un tome d'exercices corrigés.

Table des matières

1 Théorie des ensembles et dénombrement 4

1.1 Catalogue : définitions et notations . . . 4

1.2 Dénombrabilité . . . 5

1.3 Dénombrement . . . 6

1.3.1 Cardinal d'un ensemble fini . . . 6

1.3.2 Permutations, arrangements et combinaisons . . . 7

1.3.3 Formule du binôme de Newton . . . 10

2 Espaces probabilisés 12

2.1 Introduction . . . 12

2.2 Espaces probabilisés : définition et premiers exemples . . . 13

2.3 Probabilité uniforme . . . 17

2.4 Propriétés fondamentales d'une probabilité . . . 18

2.4.1 Propriétés élémentaires d'une probabilité . . . 18

2.4.2 Continuité par limite monotone d'événements, à droite et à gauche 20

2.5 Conditionnement, indépendance . . . 23

2.5.1 Conditionnement . . . 23

2.5.2 Formule des probabilités totales et formule de Bayes . . . 24

2.5.3 Indépendance . . . 27

3 Variables aléatoires discrètes 32

3.1 Définition . . . 32

3.2 Loi d'une variable aléatoire . . . 33

3.3 Variables aléatoires à support fini . . . 35

3.3.1 Lois classiques sur des ensembles finis . . . 35

3.3.2 Espérance d'une v.a. réelle à support fini . . . 38

3.3.3 Théorème de transfert . . . 39

3.3.4 Propriétés de l'espérance. 41
3.3.5 Moments d'une variable aléatoire à support fini 43
3.4 Variables aléatoires à support infini dénombrable 45
3.4.1 Parenthèse analytique 45
3.4.2 Lois classiques sur des ensembles infinis dénombrables 47
3.4.3 Extension de la notion d'espérance à des v.a. réelles à support infini dénombrable 51
3.4.4 Théorème de transfert 53
3.4.5 Propriétés de l'espérance 54
3.4.6 Moments d'une variable aléatoire à valeurs réelles 54

4 Couples et vecteurs aléatoires 56
4.1 Loi du couple, lois marginales 56
4.2 Lois conditionnelles 59
4.3 Espérance d'une fonction du couple 61
4.3.1 Cadre général 61
4.3.2 Covariance 62
4.4 Variables aléatoires indépendantes 64
4.5 Somme de variables aléatoires 68
4.5.1 Loi de la somme de deux variables aléatoires 68
4.5.2 Deux exemples fondamentaux 68
4.5.3 Variance de la somme de deux variables aléatoires 69
4.6 Vecteurs et suites aléatoires 70

5 Inégalités classiques et applications 72
5.1 Inégalités de Markov et Bienaymé-Tchébychev 72
5.2 Application : loi faible des grands nombres 73
5.3 Inégalité de Cauchy-Schwarz 74

6 Introduction aux variables aléatoires continues 75
6.1 Espaces probabilisés : cadre général 75
6.2 Variables aléatoires à valeurs réelles : cadre général 78
6.3 Variables aléatoires réelles continues (ou "à densité") 80
6.3.1 Définition 80
6.3.2 Trois familles de lois continues "classiques" 81
6.3.3 Quelques propriétés des lois continues 84
6.3.4 Notion d'espérance d'une variable aléatoire continue 85

1 Théorie des ensembles et dénombrement

1.1 Catalogue : définitions et notations

On rappelle ici un certain nombre de définitions et notations vues en L1.
Soit E un ensemble.

- Soit A un ensemble. On note $A \subset E$ si tout élément de A est aussi un élément de E. L'écriture mathématique avec des quantificateurs est donc

$$\forall x \in A, x \in E.$$

  On dit alors que A est une *partie* de E, ou que A est un *sous-ensemble* de E.

- On note $\mathcal{P}(E)$ l'ensemble des parties de E, autrement dit

$$\mathcal{P}(E) := \{A : A \subset E\}.$$

  Remarque : l'ensemble vide $\emptyset$ est une partie de E.

  <u>Ex.</u> : si $E := \{1,2,3\}$, $\mathcal{P}(E) := \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$
  Dans cet exemple $\mathcal{P}(E)$ a $8 = 2^3$ éléments.

- Soient $A_1, \ldots, A_n$ des parties de E. On définit l'*intersection* $\cap_{i=1}^n A_i = A_1 \cap \ldots \cap A_n$ des $A_i, 1 \leq i \leq n$ par

$$\cap_{i=1}^n A_i := \{x \in E : x \in A_i, \forall i \in \{1, \ldots, n\}\}.$$

  C'est l'ensemble des points de E qui sont dans tous les $A_i, 1 \leq i \leq n$.

- De même, on définit la *réunion* $\cup_{i=1}^n A_i = A_1 \cup \ldots \cup A_n$ des $A_i, 1 \leq i \leq n$ par

$$\cup_{i=1}^n A_i := \{x \in E : \exists i \in \{1, \ldots, n\}, x \in A_i\}.$$

  C'est l'ensemble des points de E qui appartiennent à au moins un des $A_i, 1 \leq i \leq n$.

- L'intersection et la réunion sont *distributives* l'une par rapport à l'autre : pour toute partie B de E,

$$B \cup (\cap_{i=1}^n A_i) = \cap_{i=1}^n (B \cup A_i) \quad \text{et} \quad B \cap (\cup_{i=1}^n A_i) = \cup_{i=1}^n (B \cap A_i).$$

- Lorsque E est infini, on peut naturellement étendre ces définitions à une famille (infinie) de parties de E $(A_i)_{i \in I}$ – où I désigne ici un ensemble d'indices infini – par :

$$\cap_{i \in I} A_i := \{x \in E : x \in A_i, \forall i \in I\}$$

et

$$\cup_{i\in I} A_i := \{x \in E : \exists i \in I, x \in A_i\}.$$

Les propriétés de distributivité restent valides.

Lorsque $I = \mathbb{N}$, on notera plutôt $\cap_{i=0}^{\infty} A_i$ à la place de $\cap_{i\in\mathbb{N}} A_i$. De même pour l'union.

- Soit $A \subset E$. Le *complémentaire* de A dans E, noté A^c, est

$$A^c := \{x \in E : x \notin A\}.$$

  C'est l'ensemble des points de E qui ne sont pas dans A.

  Ex. : $E = \{1, 2, 3\}$, alors $\{1\}^c = \{2, 3\}$.

  Quelque soit l'ensemble E, et les parties $A, A_1, \ldots, A_n$ de E, on a toujours : $E^c = \emptyset$, $\emptyset^c = E$, $(A^c)^c = A$ et

$$(\cap_{i=1}^n A_i)^c = \cup_{i=1}^n A_i^c \quad \text{et} \quad (\cup_{i=1}^n A_i)^c = \cap_{i=1}^n A_i^c.$$

- Si $A \subset E$, *l'indicatrice de A*, notée $\mathbb{1}_A$, est la fonction :

$$\mathbb{1}_A : \begin{array}{rl} E & \mapsto \{1, 0\} \\ x & \mapsto \begin{cases} 1 \text{ si } x \in A \\ 0 \text{ si } x \notin A. \end{cases} \end{array}$$

- Enfin, si $E_1, E_2, \ldots, E_n$ sont des ensembles, on définit le *produit (cartésien)* de $E_1, E_2, \ldots, E_n$ par

$$E_1 \times E_2 \times \ldots \times E_n := \{(x_1, \ldots, x_n) : x_1 \in E_1, \ldots, x_n \in E_n\}.$$

  Ex. : $E = \{1, 2, 3\}$ et $F = \{a, b\}$, alors $E \times F = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$.

1.2 Dénombrabilité

Définition 1. *Un ensemble E est dit* fini *s'il n'a qu'un nombre fini d'éléments.*

Définition 2. *Un ensemble E est dit* dénombrable *s'il existe une bijection de E dans* $\mathbb{N}$.

Définition 3. *Un ensemble E est dit* au plus dénombrable *s'il est fini ou dénombrable.*

Ex. : $\mathbb{N}, \mathbb{N} \times \mathbb{N}, \mathbb{Q}$ sont dénombrables, mais $[0, 1], \mathbb{R}$ ne sont pas dénombrables.

Nous renvoyons au cours d'Analyse 3 pour plus de détails sur la notion de dénombrabilité.

1.3 Dénombrement

1.3.1 Cardinal d'un ensemble fini

Dans toute cette section, E désigne un ensemble *fini*.

Définition 4. *On appelle* cardinal *de E le nombre d'éléments de E.*
Notation : $\mathrm{Card}(E)$ *ou* $|E|$.

Remarque : $\mathrm{Card}(\emptyset) = 0$.

Proposition 5. *Si E et F sont des ensembles finis alors*

$$\mathrm{Card}(E \times F) = \mathrm{Card}(E) \times \mathrm{Card}(F).$$

Par récurrence, on en déduit que pour tout entier $n \in \mathbb{N}^*$ et tous les ensembles finis $E_1, \ldots, E_n$,

$$\mathrm{Card}(E_1 \times \ldots \times E_n) = \prod_{i=1}^{n} \mathrm{Card}(E_i).$$

Ex. : si $E = \{a, b, c\}$ et $F = \{1, 2\}$, alors $\mathrm{Card}(E \times F) = 6$.

Idée de la démonstration.

$\square$

Théorème 6. *On a :*

$$\mathrm{Card}(\mathcal{P}(E)) = 2^{\mathrm{Card}(E)}.$$

Ex. : si $E = \{1, 2, 3\}$, $\mathrm{Card}(E) = 3$ et $\mathcal{P}(E) := \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.
On a donc bien $\mathrm{Card}(\mathcal{P}(E)) = 8 = 2^3 = 2^{\mathrm{Card}(E)}$.

Démonstration du Théorème 6.

$\square$

Notation factorielle : pour tout entier $n \in \mathbb{N}^*$, on note

$$n! := 1 \times 2 \times \ldots \times (n-1) \times n$$

le produit des n premiers entiers positifs non nuls.

Par convention, on décide que $0\,! := 1$.

1.3.2 Permutations, arrangements et combinaisons

Dans cette section E désigne un ensemble fini de cardinal $n \in \mathbb{N}^*$.

Définition 7. *On appelle* permutation *de E tout n-uplet $(x_1, \ldots, x_n)$ d'éléments distincts de E. Autrement dit, une permutation de E est un* ***classement ordonné*** *des éléments de E.*

Rappel : la notation avec des accolades représente un ensemble, et il n'y a pas de notion d'ordre (ex : $\{a, b\} = \{b, a\}$), tandis que la notation aves des parenthèses représente une suite, donc l'ordre est important (ex : $(a, b) \neq (b, a)$).

Ex. : si $E = \{1, 2, 3\}$, les permutations de E sont :

Proposition 8. *Pour un ensemble E à n éléments, il y a $n!$ permutations de E.*

Démonstration.

$\square$

Exercice. (Anagrammes.)

1. Combien de mots différents peut-on former en changeant de place les lettres du mot PARIS ? (On ne demande pas que ces mots aient un sens.)
2. Combien de mots différents peut-on former en changeant de place les lettres du mot SERIE ? Et du mot ANAGRAMMES ?

Corrigé.

Définition 9. *Soit $1 \leq p \leq n$. On appelle* arrangement *de p éléments de E tout p-uplet $(x_1, \dots, x_p)$ d'éléments distincts de E.*

Remarque : là aussi, l'ordre est important.

Ex. : si $E = \{1, 2, 3, 4\}$ et $p = 2$, on a 12 arrangements de 2 éléments :

Proposition 10. *Le nombre d'arrangements de p éléments de E est*

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

Idée de la démonstration. Similaire à celle pour dénombrer le nombre de permutations – à faire en exercice. $\square$

Ceci nous amène à introduire la notation, pour tout entier $n \in \mathbb{N}$ et tout entier p, $0 \leq p \leq n$:

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

Remarque : lorsque $p = n$, un arrangement de p éléments de E correspond à une permutation de E. On remarque que $A_n^n = \frac{n!}{0!} = n!$, ce qui correspond bien au nombre de permutations d'un ensemble à n éléments.

Définition 11. *Soit $1 \leq p \leq n$. On appelle* combinaison *de p éléments de E toute partie de cardinal p de E.*

Remarque : l'ordre ici n'a donc pas d'importance.

Ex. : si $E = \{1, 2, 3, 4\}$ et $p = 2$, on a 6 combinaisons de 2 éléments :

Proposition 12. *Le nombre de combinaisons de p éléments de E est*

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

On notera cette quantité $\binom{n}{p}$:

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

(On pourra aussi utiliser la notation C_n^p, qui est égale à $\binom{n}{p}$ et qu'on trouve dans un certain nombre d'ouvrages.)

Démonstration.

□

Exemple 1. a) Dans une urne contenant 10 boules numérotées de 1 à 10 on tire 3 boules *successivement, sans remise.* On s'intéresse aux numéros des boules obtenus, dans l'ordre. Combien y-a-t-il de tirages successifs ?

Rép. :

b) Dans la même urne, on tire maintenant 3 boules *simultanément.* Combien y-a-t-il de tirages possibles ?

Rép. :

c) Et si on fait un tirage de 3 boules avec remise ?
Rép. :

Exemple 2. a) Combien y-a-t-il de codes de cartes bleues ?
Rép. :

b) Combien y-a-t-il de codes avec des chiffres tous distincts ?
Rép. :

1.3.3 Formule du binôme de Newton

Théorème 13 (Formule du binôme de Newton.)**.** *Soient a, b deux réels et $n \in \mathbb{N}^*$. Alors,*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{1}$$

Remarque : lorsque $n = 2$, on retrouve ainsi l'identité remarquable $(a+b)^2 = a^2 + 2ab + b^2$.

Démonstration.

.

$\square$

2 Espaces probabilisés

2.1 Introduction

Objectif : introduire et étudier des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec certitude. Dans ce cours on se restreindra souvent au cadre fini ou dénombrable (dit aussi cadre "discret"), qui a l'avantage de se présenter plus simplement et donne déjà une bonne idée de nombreux concepts probabilistes. On sortira cependant de temps à autres de ce cadre. La théorie des Probabilités dans son cadre le plus complet repose sur la théorie de la Mesure et de l'Intégration, qui sera abordée en L3.

Pour pouvoir modéliser une expérience aléatoire, on va avoir besoin de deux objets mathématiques :

- **Un ensemble**, en général noté Ω, qui représente l'ensemble de tous les résultats possibles d'une expérience. C'est cet ensemble qui sera souvent, chez nous, fini ou dénombrable.

  Exemples :

- **Une probabilité** (la définition mathématique sera donnée ci-dessous), en général notée $\mathbb{P}$, qui associe à chaque partie de Ω un réel entre 0 et 1 qui représente la probabilité que cette partie se réalise.

  Ex. : Jet d'une pièce. On prend $\Omega = \{P, F\}$. Il y a 4 parties de Ω :

$$\emptyset,\ \Omega,\ \{P\},\ \{F\}.$$

  Si on veut modéliser une pièce équilibrée, on prendra :

- $\mathbb{P}(\emptyset) = 0$ (probabilité de n'obtenir ni un Pile, ni un Face)
- $\mathbb{P}(\{F\}) = 1/2$ (probabilité d'obtenir un Face)
- $\mathbb{P}(\{P\}) = 1/2$ (probabilité d'obtenir un Pile)
- $\mathbb{P}(\{P, F\}) = 1$ (probabilité d'obtenir un Pile ou un Face).

Si on veut modéliser une pièce biaisée, il faudra changer la probabilité. Par exemple, si on sait que la probabilité d'obtenir un Pile est p, où p est un réel fixé de $[0, 1]$, alors la probabilité correspondante sur l'ensemble des parties de Ω est définie par

$$\mathbb{P}(\emptyset) = 0, \mathbb{P}(\{F\}) = 1 - p, \mathbb{P}(\{P\}) = p, \mathbb{P}(\{P, F\}) = 1.$$

Notre objectif est d'introduire un cadre formel pour étudier rigoureusement ces expériences aléatoires.

2.2 Espaces probabilisés : définition et premiers exemples

Extension de la notation somme : soit $(a_n)_{n\geq 1}$ une suite de réels ***positifs***. On note

$$\sum_{k=1}^{\infty} a_k := \lim_{n\to\infty} \sum_{k=1}^{n} a_k.$$

Cette limite a toujours un sens car la suite $(\sum_{k=1}^{n} a_k)_{n\geq 1}$ est croissante (puis les a_k sont positifs), mais cette limite peut-être finie ou infinie.

Ex. : 1)

2)

Définition 14. *Si $A_1, \ldots, A_n$ sont des parties d'un même ensemble Ω, on dit qu'elles sont* 2 à 2 disjointes *si $A_i \cap A_j = \emptyset$ pour tout $i \neq j, 1 \leq i, j \leq n$. De même si on considère une suite infinie $(A_n)_{n\in\mathbb{N}}$ de parties de Ω, on dit qu'elles sont* 2 à 2 disjointes *si $A_i \cap A_j = \emptyset$ pour tout $i \neq j, i, j \geq 0$.*

Définition 15. *Soit* Ω *un ensemble. On appellera* probabilité *sur* $\mathcal{P}(\Omega)$ *toute application*

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$$

telle que :

- $\mathbb{P}(\Omega) = 1$ *et* $\mathbb{P}(\emptyset) = 0$.
- *Pour toute suite* $(A_n)_{n\geq 1}$ *de parties* deux à deux disjointes *de* Ω,

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \tag{2}$$

Le couple $(\Omega, \mathbb{P})$ *est alors appelé* espace probabilisé.

Remarques : 1. Pour toute partie $A \in \mathcal{P}(\Omega)$, $\mathbb{P}(A) \in [0,1]$.
2. La définition ci-dessus ne correspond pas tout-à-fait à la définition d'une probabilité lorsqu'on se place dans le cadre général qui sera abordé en L3. On se limitera cependant à cette définition simplifiée dans ce cours.

On déduit immédiatement de cette définition la propriété suivante :

Proposition 16. *Si* $A_1, \dots, A_n$ *sont des parties de* Ω deux à deux disjointes, *et si* $\mathbb{P}$ *est une probabilité sur* $\mathcal{P}(\Omega)$, *on a :*

$$\mathbb{P}(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mathbb{P}(A_k). \tag{3}$$

Démonstration.

$\square$

Définition 17. *On appelle la propriété (2)* propriété de sigma-additivité. *De même pour la version (3) impliquant un nombre fini de parties.*

Dans ce cours, on supposera souvent que Ω est un ensemble *fini ou dénombrable*.
Dans ce cas on pourra utiliser le résultat suivant :

Proposition 18. *Si Ω est un ensemble fini ou dénombrable, alors*

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}), \quad \forall A \in \mathcal{P}(\Omega).$$

Autrement dit, pour calculer la probabilité de A il suffit de savoir calculer les probabilités de chacun des éléments de A et de les sommer.

Démonstration.

$\square$

Définition 19. *En Probabilités, on appelle les éléments de $\mathcal{P}(\Omega)$ des* événements.

Intuitivement, la probabilité $\mathbb{P}(A)$ représente alors la probabilité que l'événement A se réalise.

Ex. : Jet d'un dé équilibré. On a $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathbb{P}(\{\omega\}) = 1/6$ pour tout $\omega \in \Omega$. D'après la Proposition 18, on a donc

$$\mathbb{P}(A) = \mathrm{Card}(A) \times \frac{1}{6} = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}, \quad \forall A \in \mathcal{P}(\Omega).$$

Par exemple : $\{1, 3, 5\}$ est l'événement "on obtient un résultat impair". La probabilité de cet événement est $\mathbb{P}(\{1, 3, 5\}) = 1/2$; $\{5, 6\}$ est l'événement "on obtient un résultat plus grand que 5". La probabilité de cet événement est $\mathbb{P}(\{5, 6\}) = 1/3$.

Exercice. On dispose d'un dé à 6 faces pipé tel que la probabilité d'obtenir une face soit proportionnelle au chiffre porté par cette face. On lance le dé pipé.

Q1. Construire un espace probabilisé modélisant l'expérience aléatoire.

Q2. Quelle est la probabilité d'obtenir un chiffre pair ?

Corrigé.

Interprétation probabiliste de la théorie des ensembles.

Définition 20. *Soit* $(\Omega, \mathbb{P})$ *un espace probabilisé.*

- *Un événement* $A \in \mathcal{P}(\Omega)$ *est dit* négligeable *si* $\mathbb{P}(A) = 0$.
- *Un événement* $A \in \mathcal{P}(\Omega)$ *est dit* presque sûr *si* $\mathbb{P}(A) = 1$.

Remarque : Un événement négligeable est donc un événement qui a une probabilité nulle de se réaliser. Par exemple l'ensemble vide est négligeable. La réciproque est fausse en général : il y a des espaces probabilisés ayant des ensembles négligeables autres que l'ensemble vide. Par exemple, si $\Omega = \{1, 2\}$ et $\mathbb{P}$ est la probabilité définie sur Ω par $\mathbb{P}(\{1\}) = 0$, $\mathbb{P}(\{2\}) = 1$, $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(\Omega) = 1$ (vous verifierez qu'il s'agit bien d'une probabilité), le singleton $\{1\}$ est négligeable, mais ce n'est pas l'ensemble vide. De même, un événement presque sûr n'est pas nécessairement égal à Ω.

2.3 Probabilité uniforme

On verra dans ce cours plusieurs exemples classiques de probabilités. On commence par présenter dans ce paragraphe un exemple très courant, appelé *probabilité uniforme.*

Pour introduire cette définition, commençons par considérer Ω un ensemble à n éléments, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, et $\mathbb{P}$ une probabilité sur $\mathcal{P}(\Omega)$ telle que chacun de ces n éléments a la *même probabilité de se réaliser* : $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\{\omega_1\})$, $\forall 1 \le i \le n$. Alors, par propriété de sigma-additivité,

$$\sum_{i=1}^{n} \mathbb{P}(\{\omega_i\}) \underset{\text{les ens. } \{\omega_i\} \text{ sont disjoints}}{=} \mathbb{P}(\cup_{i=1}^{n} \{\omega_i\}) = \mathbb{P}(\Omega) = 1.$$

On a donc nécessairement $\mathbb{P}(\{\omega_i\}) = 1/n$, $\forall 1 \le i \le n$ et par conséquent

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{n}, \quad \forall A \subset \Omega.$$

De façon générale, on définit donc :

Définition 21. *Si* Ω *est un ensemble* fini*, la probabilité uniforme sur* $\mathcal{P}(\Omega)$ *est la probabilité définie par*

$$\mathbb{P}(A) = \frac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)}, \quad \forall A \subset \Omega.$$

Ex. : jet d'un dé équilibré, jet d'une pièce équilibrée, jet consécutif de deux pièces équilibrées (chaque configuration $(PP), (PF), (FP), (FF)$ a la même probabilité de se réaliser), etc.

2.4 Propriétés fondamentales d'une probabilité

2.4.1 Propriétés élémentaires d'une probabilité

Proposition 22. *Soit* $(\Omega, \mathbb{P})$ *un espace probabilisé. Alors,*

(i) *Pour tout* $A \in \mathcal{P}(\Omega)$*, on a :*

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

(ii) *Pour tous* $A \in \mathcal{P}(\Omega)$ *et* $B \in \mathcal{P}(\Omega)$ *tels que* $A \subset B$*, on a :*

$$\mathbb{P}(A) \leq \mathbb{P}(B).$$

(iii) *Pour tous* $A \in \mathcal{P}(\Omega)$ *et* $B \in \mathcal{P}(\Omega)$*, on a :*

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

Ceci implique en particulier que $\mathbb{P}(A \cup B) \leq P(A) + \mathbb{P}(B)$*.*

(iv) *Plus généralement, pour tous* $A_1, \ldots, A_n \in \mathcal{P}(\Omega)$*,*

$$\mathbb{P}(\cup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} \mathbb{P}(A_k).$$

<u>Démonstration.</u>

.

$\square$

<u>Exercice.</u> Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Soient A et B deux événements tels que $\mathbb{P}(A) = 2/3$ et $\mathbb{P}(B) = 1/2$. Montrer que

$$\frac{1}{6} \leq \mathbb{P}(A \cap B) \leq \frac{1}{2}$$

et trouver des exemples d'espaces probabilisés et d'événements montrant que les valeurs extrémales peuvent être atteintes.

<u>Corrigé :</u>

.

2.4.2 Continuité par limite monotone d'événements, à droite et à gauche

Proposition 23. *Soit* $(\Omega, \mathbb{P})$ *un espace probabilisé.*

(i) *Soit* $(A_n)_{n\geq 1}$ *une suite* croissante *d'événements, c'est-à-dire que*

$$A_n \subset A_{n+1} \text{ pour tout } n \in \mathbb{N}^*.$$

Alors

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(ii) *Soit* $(B_n)_{n\geq 1}$ *une suite* décroissante *d'événements, c'est-à-dire que*

$$B_{n+1} \subset B_n \quad \text{pour tout} \quad n \in \mathbb{N}^*.$$

Alors

$$\mathbb{P}(\cap_{n=1}^{\infty} B_n) = \lim_{n\to\infty} \mathbb{P}(B_n).$$

Remarque : ces résultats sont faux sans les hypothèses de monotonie. Par exemple, si $\Omega = \{P, F\}$ et $\mathbb{P}$ est la probabilité uniforme sur $\mathcal{P}(\Omega)$, et si on considère les événements $A_1 = \{P\}$, $A_n = \{F\}$ pour $n \geq 2$, alors on a :

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = 1 \neq \frac{1}{2} = \lim_{n\to\infty} \mathbb{P}(A_n),$$

ce qui ne contredit pas (i) puisque la suite (A_n) n'est pas croissante ; de plus,

$$\mathbb{P}(\cap_{n=1}^{\infty} A_n) = 0 \neq \frac{1}{2} = \lim_{n\to\infty} \mathbb{P}(A_n),$$

Ce qui ne contredit pas (ii) puisque la suite (A_n) n'est pas décroissante.

Exemple d'application. On jette une pièce équilibrée une infinité de fois. Pour modéliser cette expérience, on prend comme ensemble Ω l'ensemble des suites à valeurs dans $\{P, F\}$. Cet ensemble est souvent noté $\{P, F\}^\infty$. Il n'est pas dénombrable (on admet ce point). On note A l'événement "on obtient uniquement des Pile".

Q. : Que vaut $\mathbb{P}(A)$?

Rép. :

$\mathbb{P}(A) = 0$. En effet, on remarque que $A = \cap_{n=1}^{\infty} A_n$ où A_n est l'événement "on a obtenu n Pile au cours des n premiers lancers". Clairement $A_{n+1} \subset A_n$ et $\mathbb{P}(A_n) = (1/2)^n$ pour tout n. Donc d'après la Proposition 23 (ii),

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0.$$

Remarque : on a donc ici un autre exemple d'ensemble négligeable qui n'est pas l'ensemble vide : $A = \{(P, P, P, P, P, P, P, \ldots)\}$.

Démonstration de la Proposition 23.

.

$\square$

Exercice. VRAI ou FAUX ? (Justifiez votre réponse.)

1. Si A et B sont deux événements tels que $\mathbb{P}(A) + \mathbb{P}(B) = 1$ alors $B = A^c$.
2. Si A et B sont deux événements tels que $\mathbb{P}(A) = \frac{1}{5}$ et $\mathbb{P}(B) = \frac{1}{4}$ alors nécessairement $\mathbb{P}(A \cup B) \leq \frac{9}{20}$.
3. Si A et B sont deux événements tels que $\mathbb{P}(A) = \frac{1}{5}$ et $\mathbb{P}(B) = \frac{1}{4}$ alors nécessairement $\mathbb{P}(A \cap B) \geq \frac{1}{20}$.
4. Pour toute suite d'événements $(A_n)_{n\geq 1}$, $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Corrigé :

2.5 Conditionnement, indépendance

2.5.1 Conditionnement

La notion de conditionnement est très importante en théorie des Probabilités. Elle intervient lorsque dans le déroulement d'une expérience aléatoire une information partielle est fournie, modifiant ainsi la probabilité d'un événement. Commençons par regarder ce qu'il se passe sur un exemple.

Ex. On jette un dé équilibré : $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathbb{P}$ est la probabilité uniforme sur $\mathcal{P}(\Omega)$. On s'intéresse aux événements $A := \{2\}$, $B := \{2, 4, 6\}$ et $C := \{1, 2\}$. On a : $\mathbb{P}(A) = 1/6$, $\mathbb{P}(B) = 3/6 = 1/2$, $\mathbb{P}(C) = 2/6 = 1/3$.

Supposons qu'on ait une information partielle sur l'expérience : on sait qu'on a obtenu un résultat pair (B est réalisé). Dans ce cas la probabilité que A se réalise sachant que B est réalisé est $1/3$ (probabilité d'obtenir 2 sachant qu'on a obtenu 2, 4 ou 6). On notera cette probabilité $\mathbb{P}(A|B)$.

Autre cas de figure : on sait que C est réalisé. Dans ce cas la probabilité que A se réalise sachant que C est réalisé est $1/2$ (probabilité d'obtenir 2 sachant qu'on a obtenu 1 ou 2). On notera cette probabilité $\mathbb{P}(A|C)$.

Remarquons que dans ces deux cas,

$$\mathbb{P}(A|B) = \frac{1}{3} = \frac{1/6}{3/6} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

et

$$\mathbb{P}(A|C) = \frac{1}{2} = \frac{1/6}{2/6} = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)}.$$

Cet exemple est illustrateur d'un processus général, qui nous amène à la définition suivante.

Dans tout ce qui suit on travaille sur $(\Omega, \mathbb{P})$ un espace probabilisé.

Définition 24. *Soit B un événement tel que $\mathbb{P}(B) > 0$. Alors, pour tout événement A, la* probabilité conditionnelle de A sachant B*, notée $\mathbb{P}(A|B)$, est définie par*

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

Remarque : $A|B$ se lit "A sachant B". Ce n'est pas un événement.

Proposition 25. *Soit B un événement tel que $\mathbb{P}(B) > 0$. Alors, l'application*

$$\begin{array}{llll} \mathbb{P}(\cdot|B) : & \mathcal{P}(\Omega) & \to & [0, 1] \\ & A & \mapsto & \mathbb{P}(A|B) \end{array}$$

est une probabilité sur $\mathcal{P}(\Omega)$.

Démonstration.

$\square$

Remarque importante : On peut appliquer toutes les propriétés d'une probabilité vues précédemment à cette probabilité conditionnelle $\mathbb{P}(\cdot|B)$.

2.5.2 Formule des probabilités totales et formule de Bayes

On va maintenant voir deux formules extrêmement utiles. On commence par une définition.

Définition 26. (i) *On appelle* partition finie de Ω *une suite finie d'événements* $(B_n)_{1\leq n\leq N}$ *($N \in \mathbb{N}$) deux à deux disjoints telle que* $\Omega = \cup_{n=1}^{N} B_n$.
(ii) *De même on appelle* partition infinie de Ω *une suite infinie d'événements* $(B_n)_{n\geq 1}$ *deux à deux disjoints telle que* $\Omega = \cup_{n=1}^{\infty} B_n$.

Théorème 27 (Formule des probabilités totales)**.**

(i) *Soit* $(B_n)_{1\leq n\leq N}$ *une partition finie de* Ω *telle que* $\mathbb{P}(B_n) > 0$ *pour tout* $1 \leq n \leq N$. *Alors,*

$$\mathbb{P}(A) = \sum_{n=1}^{N} \mathbb{P}(A|B_n)\mathbb{P}(B_n), \quad \forall A \in \mathcal{P}(\Omega).$$

(ii) *De même, si* $(B_n)_{n\geq 1}$ *est une partition infinie de* Ω *telle que* $\mathbb{P}(B_n) > 0$ *pour tout* $n \geq 1$,

$$\mathbb{P}(A) = \sum_{n=1}^{\infty} \mathbb{P}(A|B_n)\mathbb{P}(B_n), \quad \forall A \in \mathcal{P}(\Omega).$$

(iii) *Cette formule reste valide sous la forme suivante lorsque certains des événements* B_n *sont négligeables : si* $(B_n)_{n\geq 1}$ *est une partition infinie de* Ω, *on a*

$$\mathbb{P}(A) = \sum_{n \text{ tel que } \mathbb{P}(B_n)>0, n\geq 1} \mathbb{P}(A|B_n)\mathbb{P}(B_n), \quad \forall A \in \mathcal{P}(\Omega).$$

Démonstration.

□

Théorème 28 (Formule de Bayes)**.** *Soient* A, B *deux événements tels que* $\mathbb{P}(A) > 0$ *et* $\mathbb{P}(B) > 0$. *Alors,*

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

Démonstration. Il suffit d'utiliser la définition d'une probabilité conditionnelle :

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

$\square$

Exemple d'utilisation de ces deux formules. A l'oral d'un concours les candidats doivent choisir au hasard une question dans un sabot de 50 questions réparties en 20 questions de maths, 20 questions d'économie et 10 questions d'anglais. (Il est sous entendu que le choix est uniforme ici.) Un candidat se présente à l'oral. La probabilité qu'il réponde correctement s'il choisit une question de maths est $1/3$, si c'est une question d'économie $5/6$, si c'est une question d'anglais 0 (il a fait une impasse).

Q.1. Quelle est la probabilité que le candidat réponde correctement à la question choisie au hasard dans le sabot ?

Q.2. Quelle est la probabilité que la question choisie soit une question d'économie sachant qu'il n'a pas répondu correctement à cette question ?

Il est très important de savoir rédiger une réponse mathématique claire pour ce type de questions, et ne pas se contenter de donner une réponse basée sur une intuition. Il faut donc commencer par formaliser les événements qui nous intéressent, puis identifier en termes de ces événements les probabilités et probabilités conditionnelles connues et celles qui sont à calculer.

.

2.5.3 Indépendance

Soit $(\Omega, \mathbb{P})$ un espace probabilisé et A, B deux événements.

Supposons que $\mathbb{P}(B) > 0$. On a vu qu'en général, lorsqu'on sait que B est réalisé cela change la probabilité que A se réalise. Ce n'est cependant pas toujours le cas, il est possible que

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

(ou de façon équivalente $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, ou encore $\mathbb{P}(B|A) = \mathbb{P}(B)$ si $\mathbb{P}(A) > 0$). Dans ce cas B n'a pas d'impact sur sur la probabilité que A se réalise et vice-versa. On dit alors que "A et B sont indépendants".

Définition 29. (i) *(Indépendance de 2 événements.) Deux événements A et B sont* indépendants *si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

(ii) *(Indépendance d'une suite finie d'événements.) Plus généralement, des événements $A_n, 1 \leq n \leq N$ $(N \in \mathbb{N}^*)$ sont dit indépendants si pour toute partie $I \subset \{1, \ldots, N\}$*

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

(iii) *(Indépendance d'une suite infinie d'événements.) Des événements $A_n, n \geq 1$ sont dit indépendants si pour toute partie* finie *$I \subset \mathbb{N}^*$*

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

Notation : on utilisera la notation $A \perp\!\!\!\perp B$ pour signifier que A et B sont indépendants.

Ex. : si on reprend l'exemple d'une question choisie au hasard lors d'un concours, les événements E et R ne sont pas indépendants car $\mathbb{P}(R \cap E) \neq \mathbb{P}(E)\mathbb{P}(R)$. En effet, $\mathbb{P}(R \cap E) = \mathbb{P}(R|E)\mathbb{P}(E) = 1/3$ tandis que $\mathbb{P}(E)\mathbb{P}(R) = 14/75$.

Exercice. Montrer qu'un événement A est indépendant de lui-même si et seulement si $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

Remarque. Attention aux points (ii) et (iii) de la définition, les égalités doivent être vérifiées pour *toute partie finie* I dans $\{1, \ldots, N\}$ ou $\mathbb{N}^*$ respectivement.

Définition 30 (Indépendance 2 à 2). *On dit que des événements* $A_n, 1 \leq n \leq N$ *sont* indépendants 2 à 2 *si*

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \forall i \neq j, 1 \leq i, j \leq N$$

(définition similaire pour une suite infinie d'événements).

Remarque importante : L'indépendance implique clairement l'indépendance 2 à 2 mais la réciproque est fausse en général : **l'indépendance 2 à 2 n'implique pas l'indépendance.** Prenons, pour illustrer ce point, le cas de 3 événements A, B, C. D'après la définition ils sont indépendants si et seulement si les 4 points suivants sont satisfaits :

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

Dans l'exemple suivant, les événements A, B, C ne sont pas indépendants bien qu'ils soient indépendants 2 à 2.

Ex. : On lance deux dés de façon indépendante, et on s'intéresse eux événements $A =$"le premier dé donne un résultat pair", $B =$"le deuxième dé donne un résultat pair", $C =$"la somme des résultats est paire". Les événements A, B, C ne sont pas indépendants, mais sont indépendants 2 à 2. En effet,

Exercice : Cherchez l'erreur.

Q. : *On lance une pièce équilibrée 3 fois, de façon indépendante. Quelle est la probabilité d'obtenir 3 fois le même résultat, c'est-à-dire 3 Pile ou 3 Face ?*

$\rightsquigarrow$ Alex dit : "De toute façon, on aura au moins deux résultats identiques, 2 Pile ou 2 Face. Il faut ensuite que le 3ème soit égal aux 2 autres, ce qui arrive avec probabilité 1/2. Donc la réponse est 1/2."

$\rightsquigarrow$ Tandis que Barbara argumente ainsi : "On énumère tous les résultats ordonnés possibles : PPP ; PPF ; PFP ; PFF ; FPP ; FPF ; FFP ; FFF. Chaque cas a la même probabilité de se réaliser, 1/8, donc finalement la probabilité cherchée est $\mathbb{P}(\{PPP\}) + \mathbb{P}(\{FFF\}) = 1/4$.

Qui a raison ?

Lemme 31 (L'indépendance passe au complémentaire.). *Soient* (A_n) *une suite (finie ou infinie) d'événements indépendants. A partir de cette suite, on définit une autre suite en posant, pour tout* n*, ou bien*

$$B_n := A_n$$

ou bien

$$B_n := A_n^c.$$

Alors les événements de la suite (B_n) *sont indépendants.*

Une démonstration sera donnée en TD pour le cas d'une suite de 2 éléments. Le cas général est laissé en exercice.

<u>Exercice.</u> On suppose qu'à chaque naissance la probabilité d'avoir une fille (respectivement un garçon) est $\frac{1}{2}$ (respectivement $\frac{1}{2}$). On suppose également que dans une même famille, les sexes des enfants sont déterminés de façon indépendante. Une famille a deux enfants. Quelle est la probabilité que les 2 soient des garçons sachant qu'au moins un des 2 est un garçon ?

Exercice. Démontrer, à l'aide d'exemples, que :

(1) Deux événements peuvent être disjoints et pas indépendants.
(2) Deux événements peuvent être indépendants et pas disjoints.

Exercice. Soit Ω un ensemble fini dont le cardinal est un *nombre premier* et soit $\mathbb{P}$ la probabilité uniforme sur $\mathcal{P}(\Omega)$. Soit A et B deux événements distincts chacun de $\emptyset$ et Ω. Démontrer que A et B ne sont pas indépendants.

3 Variables aléatoires discrètes

3.1 Définition

On ne sera pas toujours intéressés par le résultat complet d'une expérience aléatoire, mais souvent par une conséquence de celle-ci, c'est-à-dire par une fonction du résultat de cette expérience. Ceci nous amène au concept de variable aléatoire (abréviation v.a.), central en théorie des probabilités.

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

Définition 32. *Soit E un ensemble. Une* variable aléatoire discrète *à valeurs dans E est une fonction $X : \Omega \to E$ telle que son image $X(\Omega)(= \{X(\omega), \omega \in \Omega\})$ est finie ou dénombrable.*
On appelle $X(\Omega)$ le support *de X.*
On dit que la variable aléatoire est réelle *si $E \subset \mathbb{R}$.*

Le mot *discrète* fait référence au fait que $X(\Omega)$ est fini ou dénombrable. Ceci implique qu'on pourra toujours supposer dans ce qui suit que E est dénombrable (quitte à prendre E plus restreint), ce qu'on fera souvent. Dans ce chapitre toutes les variables aléatoires sont discrètes, ce qui sera sous-entendu dans la suite : on n'utilisera que la dénomination variable aléatoire, pour simplifier.

Remarque : la définition d'une variable aléatoire est plus complexe si on se place dans le cadre général de la théorie des Probabilités.

Ex.1.

Ex.2.

Notation : pour une v.a. X à valeurs dans E et $A \in \mathcal{P}(E)$, on note $\{X \in A\}$ l'événement image réciproque de A par X, c'est-à-dire

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

Pour simplifier, on notera $\mathbb{P}(X \in A)$ sa probabilité, plutôt que $\mathbb{P}(\{X \in A\})$.

3.2 Loi d'une variable aléatoire

Théorème 33. *Soit $(\Omega, \mathbb{P})$ un espace probabilisé et $X : \Omega \to E$ une variable aléatoire à valeurs dans E, E dénombrable. L'application*

$$\begin{array}{llll} \mathbb{P}_X : & \mathcal{P}(E) & \to [0,1] \\ & A & \mapsto \mathbb{P}_X(A) := \mathbb{P}(X \in A) \end{array}$$

est une probabilité sur $\mathcal{P}(E)$. On l'appelle la loi *de la variable aléatoire X.*

Démonstration.

$\square$

Soit $X : \Omega \to E$ une variable aléatoire discrète. Puisque $X(\Omega)$ est fini ou dénombrable on peut numéroter ses éléments, $x_1, x_2, x_3, \ldots$. Pour tout $k \in \mathbb{N}^*$, on note

$$\{X = x_k\} \text{ l'événement } \{\omega \in \Omega : X(\omega) = x_k\} = X^{-1}(\{x_k\})$$

et $\mathbb{P}(X = x_k)$ la probabilité de cet événement. On remarque que $\cup_{x_k \in X(\Omega)}\{X = x_k\} = \Omega$ et que ces événements sont deux à deux disjoints, donc

$$\sum_{x_k \in X(\Omega)} \mathbb{P}(X = x_k) = \mathbb{P}(X \in X(\Omega)) = 1.$$

Si $X(\Omega)$ est fini, la somme ci-dessus est une somme classique. Sinon elle est à interpréter comme la somme $\sum_{k=1}^{\infty} \mathbb{P}(X = x_k) := \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(X = x_k)$.

Proposition 34. *La loi de X est entièrement caractérisée par son support $X(\Omega)$ et les probabilités $\mathbb{P}(X = x_k)$, $k \in \mathbb{N}^*$.*

Démonstration.

$\square$

Donc, en pratique, donner la loi de X, v.a. discrète, c'est donner son support et les probabilités $\mathbb{P}(X = x)$ pour tout $x \in X(\Omega)$.

Remarque (admise) : à partir de n'importe quel ensemble fini $E = \{x_1, \ldots, x_k\}$ (respectivement dénombrable $E = \{x_i, i \geq 1\}$) et de n'importe quelle suite de réels positifs $(p_i)_{1\leq i\leq k}$ (respectivement $(p_i, i \geq 1)$) telle que

$$\sum_{i=1}^{k} p_i = 1 \quad \left(\text{respectivement } \sum_{i=1}^{\infty} p_i = 1\right)$$

on peut construire un espace probabilisé $(\Omega, \mathbb{P})$ et une v.a. $X : \Omega \to E$ telle que $\mathbb{P}(X = x_i) = p_i$ pour tout i, $1 \leq i \leq k$ (respectivement pour tout $i \geq 1$).

Ex.

Remarque importante : à chaque variable aléatoire on associe une probabilité, sa “loi”. Mais il ne faut pas confondre une variable aléatoire et sa loi : **deux variables aléatoires différentes peuvent avoir la même loi.** Par exemple, jetons une pièce équilibrée et introduisons X la v.a. qui vaut 1 si on obtient Pile et 0 sinon, et Y la v.a. qui vaut 0 si on obtient un Pile et 1 sinon. Clairement $X \neq Y$ (et même $X = 1-Y$). Pourtant X et Y ont la même loi : $X(\Omega) = Y(\Omega) = \{0,1\}$ et $\mathbb{P}_X(0) = \mathbb{P}_X(1) = 1/2$ et $\mathbb{P}_Y(0) = \mathbb{P}_Y(1) = 1/2$. (Cette loi s'appelle la loi de Bernoulli de paramètre 1/2, voir le paragraphe suivant.)

3.3 Variables aléatoires à support fini

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

Dans toute cette section, on ne considèrera que des variables aléatoires à support fini.

3.3.1 Lois classiques sur des ensembles finis

1. Loi de Bernoulli : Soit $p \in [0,1]$. On dit qu'une v.a. X suit une loi de Bernoulli de paramètre p si :

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

Notation : $X \sim \mathrm{Ber}(p)$.

Ex. : 1. Lancer d'une pièce : avec probabilité $p \in [0,1]$ on obtient Pile et $1-p$ Face. La variable aléatoire correspondant au nombre de Pile obtenus suit une loi de Bernoulli de paramètre p.

2. Si $A \in \mathcal{P}(\Omega)$, $X := \mathbf{1}_A$ (fonction indicatrice) suit une loi de Bernoulli de paramètre $\mathbb{P}(A)$; en effet $X(\Omega) = \{0,1\}$ et $\{X = 1\} = A$.

2. Loi binomiale : Soit $n \in \mathbb{N}^*$ et $p \in [0,1]$. On dit qu'une v.a. X suit une loi binomiale de paramètres (n,p) si :

- $X(\Omega) = \{0,1,2,\ldots,n-1,n\}$
- $\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, $\forall k \in X(\Omega)$.

Remarque : ceci définit bien une loi de probabilité puisque

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

Notation : $X \sim \mathrm{Bin}(n, p)$.

Interprétation.

3. Loi uniforme : On dit qu'une v.a. X suit une loi uniforme sur un ensemble fini $\{x_1, \ldots, x_n\}$ si $\mathbb{P}(X = x_i) = 1/n$ pour tout $1 \leq i \leq n$. Autrement dit la loi de X, $\mathbb{P}_X$, est la probabilité uniforme sur $\mathcal{P}(\{x_1, \ldots, x_n\})$ – par abus de langage, on dira aussi probabilité uniforme sur $\{x_1, \ldots, x_n\}$.

Vocabulaire : on dit aussi que $\mathbb{P}_X$ est *équiprobable* sur $\{x_1, \ldots, x_n\}$.

<u>Exercice</u> : VRAI ou FAUX ?

1. Si X est une v.a. à valeurs dans $\{0, 1\}$ alors X suit la loi de Bernoulli de paramètre $\mathbb{P}(X = 1)$.
2. Si $X \sim \text{Bin}(n, p)$ alors $n - X \sim \text{Bin}(n, p)$.
3. Si X est une v.a. de loi uniforme sur $\{0, 1, \ldots, n\}$ alors $n - X$ suit la loi uniforme sur $\{0, 1, \ldots, n\}$.

3.3.2 Espérance d'une v.a. réelle à support fini

Dans toute cette partie, on ne considèrera que des v.a. à valeurs *réelles*.
Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire. On suppose que l'ensemble $X(\Omega)$ est fini et on note $x_1, x_2, \dots, x_k$ ses éléments.

Définition 35. *On appelle* espérance de X *le nombre* $\mathbb{E}[X]$ *défini par*

$$\mathbb{E}[X] := \sum_{i=1}^{k} x_i \mathbb{P}(X = x_i).$$

C'est donc la moyenne des valeurs prises par X, pondérée par les probabilités correspondantes.

Notation. On remarque que l'ordre dans lequel on numérote les éléments de $X(\Omega)$ n'est pas important car l'espérance $\mathbb{E}[X]$ ne dépend pas de cet ordre. Afin d'éviter d'avoir à numéroter les éléments de $X(\Omega)$, on utilisera (souvent) l'écriture

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$$

(on somme sur tous les x parcourant $X(\Omega)$ les réels $x\mathbb{P}(X = x)$.)

Ex. 1. Si $X \sim \mathrm{Ber}(p)$. Alors,

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

Ex. 2. Si X est uniforme sur $\{x_1, \dots, x_k\}$. Alors,

$$\mathbb{E}[X] = \sum_{i=1}^{k} x_i \cdot \frac{1}{k} = \frac{\sum_{i=1}^{k} x_i}{k}$$

(on retrouve donc la moyenne au sens usuel).

Ex. 3. Si $X \sim \mathrm{Bin}(n, p)$. Alors, $\mathbb{E}[X] = np$.
Démonstration.

$\square$

3.3.3 Théorème de transfert

Soit X une v.a. à support fini et $f : X(\Omega) \to \mathbb{R}$ une fonction. Alors $f \circ X : \Omega \to \mathbb{R}$ est une v.a. réelle à support fini. On veut calculer son espérance. Une première approche est de calculer sa loi (ce qui peut être fastidieux), puis l'espérance associée. On peut simplifier les choses en utilisant le résultat suivant, qui permet un calcul plus direct.

Théorème 36 (Théorème de transfert.)**.** *On a :*

$$\mathbb{E}[f(X)] = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x)$$

(où, à nouveau, la somme signifie qu'on parcourt l'ensemble des éléments de $X(\Omega)$; puisque cet ensemble est fini la somme est bien définie).

(Formule extrêmement utile en pratique.)

<u>Démonstration.</u>

$\square$

Exemple d'application. Soit X une v.a. réelle à support fini, disons $\{x_1, \dots, x_n\}$. Alors, pour tout réel positif a,

$$\mathbb{E}[X^a] = \sum_{i=1}^{n} x_i^a \mathbb{P}(X = x_i).$$

Par exemple,

- Si $X \sim \mathrm{Ber}(p)$ ($p \in [0, 1]$), alors $\mathbb{E}[X^2] = 1^2 \cdot \mathbb{P}(X = 1) + 0^2 \cdot \mathbb{P}(X = 0) = p + 0 = p$.
- Si $X \sim \mathrm{Bin}(n, p)$ ($n \in \mathbb{N}^*$, $p \in [0, 1]$), alors $\mathbb{E}[X^2] = n(n-1)p^2 + np$.

Démonstration.

$\square$

Exercice. Soit $n \in \mathbb{N}^*$ et X une variable aléatoire telle que $X(\Omega) = [\![0; n]\!]$ et

$$\mathbb{P}(X = k) = c\binom{n}{k}, \quad \forall k \in [\![0; n]\!],$$

où c est un réel qui ne dépend pas de k.

1. Déterminer c et l'exprimer comme une fonction simple de n.
2. Calculer $\mathbb{E}[X]$ (exprimer le résultat comme une fonction simple de n).
3. Calculer $\mathbb{E}\left[\frac{1}{X+1}\right]$ (exprimer le résultat comme une fonction simple de n).

.

3.3.4 Propriétés de l'espérance.

Proposition 37 (Propriétés de l'espérance.)**.** *Soient X et Y des variables aléatoires réelles à support fini.*

(i) (Linéarité) *Pour tous réels a, b,*

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

(ii) (Positivité) *Si $X(\omega) \geq 0$ pour tout $\omega \in \Omega$, alors $\mathbb{E}[X] \geq 0$.*

<u>*Remarque.*</u> *Ce résultat reste vrai sous l'hypothèse plus faible*

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) < 0\}) = 0.$$

(iii) (Croissance) *Si $X(\omega) \geq Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbb{E}[X] \geq E[Y]$.*

<u>*Remarque.*</u> *Ce résultat reste vrai sous l'hypothèse plus faible*

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) < Y(\omega)\}) = 0.$$

(iv) *Soit c un réel. Si $X(\omega) = c$ pour tout $\omega \in \Omega$, alors $\mathbb{E}[X] = c$.*

<u>*Remarque.*</u> *Ce résultat reste vrai sous l'hypothèse plus faible*

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) = c\}) = 1.$$

(v) (Inégalité triangulaire) $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Démonstration.

$\square$

3.3.5 Moments d'une variable aléatoire à support fini

Soit X une variable aléatoire réelle à support fini.

Définition 38. *Soit $n \in \mathbb{N}^*$. On appelle* moment d'ordre n *de X la quantité* $\mathbb{E}[X^n]$.

Définition 39. *On appelle* variance *de X la quantité*

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

et écart-type *de X le nombre* $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Remarques.

- La variance et l'écart-type d'une v.a. sont toujours positives.
- D'après la formule de transfert (appliquée à la fonction f définie par $f(x) = (x - \mathbb{E}[X])^2$),

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_{x \in X(\Omega)} (x - \mathbb{E}[X])^2 \mathbb{P}(X = x). \tag{4}$$

- La variance et l'écart-type mesurent l'écart de X à son espérance. D'un point de vue physique, $\sigma(X)$ a la même dimension que la v.a. X.

Sous la forme (4), la variance peut parfois être longue à calculer. On préfèrera en pratique la formule suivante.

Proposition 40. $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Démonstration.

$\square$

Ex. 1. $X \sim \mathrm{Ber}(p)$. On a,

$$\mathrm{Var(X)} = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$$

d'après les valeurs calculées précédement.

Ex. 2. $X \sim \text{Bin}(n,p)$. On a,

$$\text{Var(X)} = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p).$$

Lorsqu'une v.a. est une *fonction affine* d'une autre ($Y = aX + b$, avec a, b deux réels), on peut facilement calculer son espérance (utiliser la linéarité) et aussi sa variance grâce à la formule suivante :

Proposition 41. $\text{Var}(aX + b) = a^2 \text{Var}(X)$.

Démonstration.

$\square$

Exercice. Calculer la variance d'une loi uniforme sur $\{1, \ldots, n\}$, puis d'une loi uniforme sur $\{2, 4, 6, \ldots, 2n\}$ (ens. des nombres pairs de 2 à $2n$), et enfin d'une loi uniforme sur $\{1, 3, 5, \ldots, 2n-1\}$ (ensemble des nombres impairs de 1 à $2n-1$). Commentaires ?

.

3.4 Variables aléatoires à support infini dénombrable

3.4.1 Parenthèse analytique

Remarque.

- Soit $(u_i)_{i\geq 0}$ une suite de réels *positifs*. On a déjà utilisé la notation

$$\sum_{i=0}^{+\infty} u_i = \lim_{n\to\infty} \sum_{i=0}^{n} u_i.$$

  Cette limite peut-être finie ou infinie, mais a toujours un sens.

- Soit $(v_i)_{i\geq 0}$ une suite de réels quelconques. Dans ce cas la limite quand $n \to \infty$ de $\sum_{i=0}^{n} v_i$ peut ne pas exister. Cela dit, on sait (et on admet dans ce cours) que si $\sum_{i=0}^{\infty} |v_i| < \infty$ alors $\lim_{n\to\infty} \sum_{i=0}^{n} v_i$ existe et est finie. On la note alors $\sum_{i=0}^{\infty} v_i$.

3.4.1.1 Séries géométriques.

Proposition 42. *Pour tout $q \in]-1,1[$, la suite $(\sum_{k=0}^{n} q^k)_{n\geq 0}$ converge vers $1/(1-q)$. Pour ces valeurs de q, on utilisera la notation*

$$\sum_{k=0}^{\infty} q^k = \lim_{n\to\infty} \sum_{k=0}^{n} q^k = \frac{1}{1-q}.$$

Démonstration.

$\square$

On admet le fait suivant : on a, dans ce cas particulier, le droit d'intervertir les signes $\sum_{k=0}^{\infty}$ et dérivée sur l'intervalle $]-1,1[$ (attention, dans un cadre général, l'interversion $\sum_{k=0}^{\infty}$ et dérivée peut amener à un résultat faux – voir le cours d'analyse pour plus détails), c'est-à-dire :

Proposition 43 (Admise)**.** *Sur l'intervalle* $]-1,1[$,

$$\left(\sum_{k=0}^{\infty} q^k\right)' = \sum_{k=0}^{\infty} (q^k)'$$

et de même pour les dérivées d'ordre supérieur (dérivée seconde, troisième, etc.).

Par conséquent,

Proposition 44. *Pour tout* $q \in]-1,1[$,

$$\frac{1}{(1-q)^2} = \sum_{k=1}^{\infty} kq^{k-1},$$

et

$$\frac{2}{(1-q)^3} = \sum_{k=2}^{\infty} k(k-1)q^{k-2}.$$

Ces égalités sous-entendent en particulier que les suites $n \mapsto \sum_{k=0}^{n} kq^{k-1}$ *et* $n \mapsto \sum_{k=0}^{n} k(k-1)q^{k-2}$ *ont une limite finie pour tout* $q \in]-1,1[$.

Démonstration.

$\square$

3.4.1.2 Séries exponentielles.

Proposition 45 (Admise)**.** *Pour tout $x \in \mathbb{R}$, la suite $\left(\sum_{k=0}^{n} \frac{x^k}{k!}\right)_{n \geq 0}$ est convergente et converge vers e^x. On utilisera la notation*

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = \lim_{n \to \infty} \sum_{k=0}^{n} \frac{x^k}{k!} = e^x.$$

3.4.2 Lois classiques sur des ensembles infinis dénombrables

Loi géométrique. Soit $p \in]0,1[$. On dit qu'une v.a. X suit une loi géométrique de paramètre p si :

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

Ceci définit bien la loi d'une v.a. sur $\mathbb{N}^*$ puisque $(1-p)^{k-1}p \geq 0$ pour tout $k \geq 1$ et

$$\begin{aligned}
\sum_{k=1}^{\infty}(1-p)^{k-1}p &= \lim_{n\to\infty}\sum_{k=1}^{n}(1-p)^{k-1}p \\
&\underset{i=k-1}{=} p\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-p)^i \\
&= p\sum_{i=0}^{\infty}(1-p)^i = \frac{p}{1-(1-p)} \\
&= 1.
\end{aligned}$$

<u>Notation :</u> $X \sim$ Géo(p).

<u>Interprétation.</u> C'est la loi du *temps de premier succès.*

<u>Ex.</u>

.

Loi de Poisson. Soit $\lambda > 0$ un réel. On dit qu'une v.a. X suit une loi de Poisson de paramètre λ si :

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

Ceci définit bien une loi de probabilité sur $\mathbb{N}$: tous les termes sont positifs et

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

Notation : $X \sim \mathcal{P}(\lambda)$.

On l'appelle parfois la *loi des événements rares*, pour la raison suivante :

Approximation Binomiale-Poisson.

Soit $\lambda > 0$ et $n \in \mathbb{N}^*$, $n > \lambda$. Soit $X_n \sim \mathrm{Bin}\Big(n, \frac{\lambda}{n}\Big)$. On s'intéresse au comportement de la loi de X_n quand $n \to \infty$.

.

Finalement, on a montré que :

Proposition 46 (Approximation Binomiale-Poisson)**.** *Pour tout $k \in \mathbb{N}$,*

$$\mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \mathbb{P}(X = k)$$

où $X \sim \mathcal{P}(\lambda)$.

Ce résultat s'interprète ainsi : si on a un grand nombre (n) d'événements qui ont chacun une très petite probabilité de se réaliser (λ/n), alors le nombre d'evénements effectivement réalisés suit asymptotiquement une loi de Poisson de paramètre λ. Cette loi est beaucoup utilisée en pratique dans des problèmes de modélisation en biologie, télécommunications, finance, ...

Exercice. Vrai ou Faux ?

1. Si $X \sim \text{Géo}(p)$, $p \in]0,1[$, alors $\mathbb{P}(X > n) = (1-p)^n$ pour tout $n \in \mathbb{N}$.
2. Si X est une variable aléatoire de loi géométrique telle que $\mathbb{P}(X = 1) = 3/5$, alors $\mathbb{P}(X = 3) = 12/125$.
3. Si $X \sim \text{Géo}(p)$, $p \in]0,1[$, alors la probabilité que X est paire est égale à $1/2$.
4. Si $X \sim \mathcal{P}(1)$ alors $\mathbb{P}(X \geq 2) = 1 - 2/e$.
5. Si X est une v.a. à valeurs dans $\mathbb{N}$ telle que $\mathbb{P}(X = k) = \frac{1}{2^k k}\mathbb{P}(X = k-1)$ pour tout $k \in \mathbb{N}^*$, alors X suit nécessairement une loi de Poisson de paramètre $1/2$.

3.4.3 Extension de la notion d'espérance à des v.a. réelles à support infini dénombrable

Soit X une variable aléatoire discrète **à valeurs réelles** telle que $X(\Omega)$ **est fini ou dénombrable**. Une difficulté dans le cas dénombrable est que la notion d'espérance peut ne pas être définie. Lorsqu'elle existe elle doit être manipulée avec précaution.

<u>Notations.</u> Soit X une v.a. à support *infini dénombrable*, disons $X(\Omega) = \{x_i, i \in \mathbb{N}\}$. On utilisera la notation

$$\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) := \sum_{i=0}^{\infty} |x_i| \mathbb{P}(X = x_i).$$

Cette quantité peut être finie (et nécessairement positive) ou infinie. Si elle est finie, on a vu ci-dessus que la quantité $\sum_{i=0}^{\infty} x_i \mathbb{P}(X = x_i)$ est bien définie et finie. On utilisera alors la notation

$$\sum_{x \in X(\Omega)} x \mathbb{P}(X = x) := \sum_{i=0}^{\infty} x_i \mathbb{P}(X = x_i).$$

Si Xest une v.a. à support *fini*, $X(\Omega) = \{x_i, 1 \leq i \leq n\}$, on utilisera les notations

$$\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) := \sum_{i=1}^{n} |x_i| \mathbb{P}(X = x_i)$$

et

$$\sum_{x \in X(\Omega)} x \mathbb{P}(X = x) := \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i).$$

Définition 47. *Soit X une v.a. à support fini ou dénombrable telle que $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < \infty$. On appelle alors* espérance de X *la quantité, notée $\mathbb{E}[X]$, définie par*

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \mathbb{P}(X = x).$$

Définition 48. *On dit que $X \in L_1(\Omega)$ si $\sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < \infty$.*

La notion d'espérance est donc bien définie pour toute v.a. dans $L_1(\Omega)$.

Ex.1. Soit $X \sim \text{Geo}(p)$, $p \in]0,1[$. On a $X(\Omega) = \mathbb{N}^*$, $\mathbb{P}(X = i) = (1-p)^{i-1}p$, $\forall i \geq 1$. On va montrer que $X \in L_1(\Omega)$ et $\mathbb{E}[X] = 1/p$.

Ex.2. Soit $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$. On a $X(\Omega) = \mathbb{N}$ et $\mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}$, $\forall i \geq 0$. On va montrer que $X \in L_1(\Omega)$ et $\mathbb{E}[X] = \lambda$.

Ex.3. Exemple de v.a. n'ayant pas d'espérance :

3.4.4 Théorème de transfert

On considère toujours X une variable aléatoire à support fini ou dénombrable, et $f : X(\Omega) \to \mathbb{R}$ une fonction. Sous nos hypothèses, $f(X)$ est une v.a. à support fini ou dénombrable et à valeurs réelles.

Théorème 49. *On a l'équivalence* $\sum_{x \in X(\Omega)} |f(x)|\mathbb{P}(X = x) < \infty \Leftrightarrow f(X) \in L_1(\Omega)$. *Si ces conditions sont réalisées,*

$$\mathbb{E}[f(X)] = \sum_{x \in X(\Omega)} f(x)\mathbb{P}(X = x).$$

<u>Ex.1</u> Soit $X \sim \mathrm{Geo}(p)$, $p \in]0, 1[$. Alors $X^2 \in L_1(\Omega)$ et $\mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}$.

<u>Ex.2.</u> Soit $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$. Alors $X^2 \in L_1(\Omega)$ et $\mathbb{E}[X^2] = \lambda^2 + \lambda$.

3.4.5 Propriétés de l'espérance

Soient X et Y des variables aléatoires à valeurs réelles, à supports finis ou dénombrables.

Théorème 50. *Si X, Y sont des v.a. appartenant à $L_1(\Omega)$ alors $aX + bY \in L_1(\Omega)$, pour tous réels a, b et les propriétés de l'espérance vues dans le cas fini sont encore valables : linéarité, positivité, croissance, inégalité triangulaire.*

3.4.6 Moments d'une variable aléatoire à valeurs réelles

A nouveau X désigne une variable aléatoire à valeurs réelles, à support fini ou dénombrable.

Définition 51. *Soit $n \in \mathbb{N}^*$. On dit que X a un moment d'ordre n si $X^n \in L_1(\Omega)$ (ou, de façon équivalente d'après le théorème de transfert, si $\sum_{x \in X(\Omega)} |x|^n \mathbb{P}(X = x) < \infty$). Dans ce cas on appelle $\mathbb{E}[X^n]$ le* moment d'ordre n *de X.*

Notation : on note $X \in L_n(\Omega)$ si $X^n \in L_1(\Omega)$.

Remarque : si X est à support fini, $X \in L_n(\Omega)$ pour tout entier positif n.

Proposition 52. *Soient $k \leq n$ des entiers strictement positifs. Si $X \in L_n(\Omega)$ alors $X \in L_k(\Omega)$.*

Démonstration.

$\square$

Remarque : $L_1(\Omega) \not\subset L_2(\Omega)$. En effet,

.

Définition 53. *Si $X \in L_2(\Omega)$, alors on définit la* variance *de X par*

$$\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$$

*et l'*écart-type *de X par $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.*

Remarquons que ces deux quantités sont bien finies pour toute variable aléatoire $X \in L_2(\Omega)$.

Attention, si $X \notin L_2(\Omega)$ la variance n'existe pas (une v.a. peut avoir une espérance mais pas de variance, ou ni espérance, ni variance).

Pour toute v.a. $X \in L_2(\Omega)$, les propriétés de la variance vues dans le cas fini restent valables :

- $\mathrm{Var(X)} = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ pour tous réels a, b.

Ex.1. Soit $X \sim \mathrm{Geo}(p)$, $p \in]0, 1[$. On a déjà vu que $X \in L_2(\Omega)$, $\mathbb{E}[X^2] = \frac{2}{p^2} - \frac{1}{p}$ et $\mathbb{E}[X] = 1/p$. Donc $\mathrm{Var}(X) = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

Ex.2. Soit $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$. On a déjà vu que $X \in L_2(\Omega)$, $\mathbb{E}[X^2] = \lambda^2 + \lambda$ et $\mathbb{E}[X] = \lambda$. Donc $\mathrm{Var}(X) = \lambda$.

4 Couples et vecteurs aléatoires

Souvent, on ne s'intéressera pas seulement à une variable aléatoire mais à plusieurs variables aléatoires simultanément, par exemple au poids X et la taille Y d'un individu choisi au hasard dans une population. Dans ce chapitre on va développer en détails la théorie des *couples* aléatoires discrets, puis on verra comment la généraliser aux *vecteurs* et *suites* aléatoires.

Dans ce qui suit, on travaille sur un espace probabilisé $(\Omega, \mathbb{P})$. Toutes les variables aléatoires qu'on considère sont définies sur cet espace et sont discrètes.

4.1 Loi du couple, lois marginales

Dans toute cette partie X désigne une variable aléatoire à valeurs dans un ensemble E et Y une variable aléatoire à valeurs dans un ensemble F. La variable aléatoire (X,Y) définie par

$$(X,Y): \begin{array}{l} \Omega \to E \times F \\ \omega \mapsto (X(\omega), Y(\omega)) \end{array}$$

est donc une variable aléatoire discrète à valeurs dans l'ensemble $E \times F$.

<u>Rappel</u> :
La loi de (X,Y) est la probabilité $\mathbb{P}_{(X,Y)} : \mathcal{P}(E\times F) \to [0,1], A \in \mathcal{P}(E\times F) \mapsto \mathbb{P}((X,Y) \in A)$.
La loi de X est la probabilité $\mathbb{P}_X : \mathcal{P}(E) \to [0,1]$, $B \in \mathcal{P}(E) \mapsto \mathbb{P}(X \in B)$.
La loi de Y est la probabilité $\mathbb{P}_Y : \mathcal{P}(F) \to [0,1]$, $C \in \mathcal{P}(F) \mapsto \mathbb{P}(Y \in C)$.

Définition 54. *La variable aléatoire (X,Y) est appelée* couple aléatoire *de marginales X et Y :*

- $\mathbb{P}_{(X,Y)}$ *est la loi du couple (X,Y), on utilise aussi parfois le vocabulaire* loi jointe *pour désigner cette loi.*
- $\mathbb{P}_X$ *et* $\mathbb{P}_Y$ *sont appelées les lois* marginales *du couple (X,Y).*

Comme ce sont des variables aléatoires discrètes, ces lois sont entièrement caractérisées par les valeurs qu'elles attribuent aux singletons. En particulier, donner la loi de (X,Y) c'est donner à la fois $(X,Y)(\Omega) = X(\Omega) \times Y(\Omega)$ et les probabilités $\mathbb{P}((X,Y) = (x,y))$ pour tous les $x \in X(\Omega)$ et $y \in Y(\Omega)$.

<u>Notation</u> : Pour tout couple $(x,y) \in (X,Y)(\Omega)$, on notera

$$\mathbb{P}(X = x, Y = y) := \mathbb{P}((X,Y) = (x,y)) = \mathbb{P}(\{X = x\} \cap \{Y = y\}).$$

Proposition 55. *Les lois marginales peuvent se calculer à partir de la loi du couple :*

- *Pour tout $x \in X(\Omega)$, $\mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x, Y = y)$.*

- *Pour tout* $y \in Y(\Omega)$, $\mathbb{P}(Y = y) = \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = y)$.

Démonstration.

$\square$

Exercice. Une urne contient 5 boules : 2 bleues, 2 vertes et 1 rouge. On en prélève trois simultanément. On note X le nombre de boules bleues et Y le nombre de boules rouges dans cet échantillon de trois boules. Déterminer la loi de (X, Y), puis la loi de X et la loi de Y.

.

Remarque importante : pour tout couple aléatoire, la loi de (X, Y) caractérise la loi de X et la loi de Y mais la réciproque est fausse en général.

Ex.

Exercice. Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{N}^* \times \{0,1\}$ de loi donnée par

$$\mathbb{P}(X = i, Y = 0) = \frac{1}{3 \cdot 2^i} \quad \text{et} \quad \mathbb{P}(X = i, Y = 1) = \frac{1}{3 \cdot 2^{i-1}} \quad \forall i \in \mathbb{N}^*.$$

Déterminer ses lois marginales.

4.2 Lois conditionnelles

Soit (X, Y) un couple de variables aléatoires.

Définition 56. *Soit $x \in X(\Omega)$ tel que $\mathbb{P}(X = x) > 0$.*

(i) *On appelle* loi conditionnelle de Y sachant $X = x$ *la probabilité sur $\mathcal{P}(\Omega)$, notée $\mathbb{P}_{Y|X=x}$, définie par :*

$$\mathbb{P}_{Y|X=x}(\{y\}) = \mathbb{P}(Y = y|X = x) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}, \quad \forall y \in Y(\Omega).$$

(C'est donc la loi de Y sachant l'information "$X = x$".)

(ii) *Si de plus $Y \in L_1(\Omega)$, on appelle* espérance conditionnelle de Y sachant $X = x$ *le nombre, noté $\mathbb{E}[Y|X = x]$, défini par :*

$$\mathbb{E}[Y|X = x] = \sum_{y \in Y(\Omega)} y\mathbb{P}(Y = y|X = x).$$

Remarque : Si $Y \in L_1(\Omega)$, la somme ci-dessus définissant $\mathbb{E}[Y|X = x]$ a bien un sens puisque

$$\begin{aligned}\sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y|X = x) &= \sum_{y \in Y(\Omega)} |y| \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)} \\ &\leq \frac{1}{\mathbb{P}(X = x)} \sum_{y \in Y(\Omega)} |y| \mathbb{P}(Y = y) < \infty.\end{aligned}$$

Pour les $y \in Y(\Omega)$ tels que $\mathbb{P}(X = y) > 0$, on définit de façon symétrique la loi de X sachant que $Y = y$ par

$$\mathbb{P}_{X|Y=y}(\{x\}) := \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(Y = y)}, \quad \forall x \in X(\Omega),$$

et l'espérance conditionnelle associée lorsque $X \in L_1(\Omega)$.

Ex. On reprend l'exemple de l'urne à 5 boules de la partie précédente. La loi conditionnelle de Y sachant $X = 2$ est donnée par

$$\mathbb{P}_{Y|X=2}(\{0\}) = \frac{\mathbb{P}(X = 2, Y = 0)}{\mathbb{P}(X = 2)} = \frac{2/10}{3/10} = \frac{2}{3}$$

et

$$\mathbb{P}_{Y|X=2}(\{1\}) = \frac{\mathbb{P}(X = 2, Y = 1)}{\mathbb{P}(X = 2)} = 1/3.$$

C'est une loi de Bernoulli de paramètre 1/3. On en déduit que $\mathbb{E}[Y|X = 2] = 1/3$.
On montre de même (exercice) que la loi conditionnelle de Y sachant $X = 1$ est une loi de Bernoulli de paramètre 2/3, et la loi conditionnelle de Y sachant $X = 0$ est une loi de Bernoulli de paramètre 1.

Exercice. Soit X une v.a. de loi de Poisson de paramètre $\lambda > 0$ et Y une v.a. à valeurs dans $\mathbb{N}$ telle que pour tout $n \in \mathbb{N}$, la loi de Y sachant $X = n$ est une loi binomiale de paramètres (n, p), où $p \in [0, 1]$ est fixé. Quelle est la loi de Y ?

.

4.3 Espérance d'une fonction du couple

4.3.1 Cadre général

Soit (X, Y) un couple aléatoire discret et $f : X(\Omega) \times Y(\Omega) \to \mathbb{R}$ une fonction.
On rappelle le théorème de transfert dans ce cadre.

Théorème 57. *La variable aléatoire $f(X, Y)$ est dans $L_1(\Omega)$ si et seulement si*

$$\sum_{(x,y)\in X(\Omega)\times Y(\Omega)} |f(x, y)|\mathbb{P}((X, Y) = (x, y)) < \infty.$$

Dans ce cas,

$$\mathbb{E}[f(X, Y)] = \sum_{(x,y)\in X(\Omega)\times Y(\Omega)} f(x, y)\mathbb{P}((X, Y) = (x, y)).$$

Remarque. Si X et Y sont à support fini, alors $f(X, Y)$ (quelque soit la fonction f) l'est aussi et dans ce cas $f(X, Y)$ est dans $L_1(\Omega)$.

Ex. (X, Y) est un couple aléatoire de loi donnée par $X(\Omega) = \{-1, 0, 1\}, Y(\Omega) = \{2, 5\}$ et

$$\mathbb{P}((X, Y) = (-1, 2)) = 1/2; \quad \mathbb{P}((X, Y) = (0, 2)) = 1/8; \quad \mathbb{P}((X, Y) = (1, 2)) = 0$$

et

$$\mathbb{P}((X, Y) = (-1, 5)) = 0; \quad \mathbb{P}((X, Y) = (0, 5)) = 1/4; \quad \mathbb{P}((X, Y) = (1, 5)) = 1/8.$$

On veut calculer $\mathbb{E}[XY]$.

.

4.3.2 Covariance

Définition 58. *Soient X, Y deux variables aléatoires* réelles *appartenant à $L_2(\Omega)$. On appelle* covariance *de X et Y le nombre*

$$\operatorname{Cov}(X,Y) := \mathbb{E}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right].$$

Si de plus $\operatorname{Var}(X) \neq 0, \operatorname{Var}(Y) \neq 0$*, on appelle* coefficient de corrélation *le nombre*

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

Remarques :
1) ces quantités sont bien définies car si $X, Y \in L_2(\Omega)$ alors $(X-\mathbb{E}[X])(Y-\mathbb{E}[Y]) \in L_1(\Omega)$ (admis)
2) $\operatorname{Cov}(X,X) = \operatorname{Var}(X)$
3) $\operatorname{Cov}(X,Y) = \operatorname{Cov}(Y,X)$
4) $\operatorname{Cov}(\lambda X,Y) = \lambda \operatorname{Cov}(X,Y)$ pour tout réel λ
5) $\operatorname{Cov}(X,Y+Z) = \operatorname{Cov}(X,Y) + \operatorname{Cov}(X,Z)$

Parce qu'elle vérifie les propriétés 4) et 5), on dit que la covariance est *bilinéaire.*

Proposition 59. *Soient X, Y deux v.a. réelles appartenant à $L_2(\Omega)$. Alors*

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Si $X = Y$ on retrouve la formule pour la variance. La démonstration est semblable, on développe le produit :

Démonstration.

$\square$

Interprétation. Intuitivement la covariance est une mesure de la variation simultanée de X et Y.

• $\operatorname{Cov}(X, Y) > 0$ signifie que lorsque X est plus grande que sa moyenne (respectivement plus petite) Y a tendance à être plus grande que sa moyenne (respectivement plus petite)

• tandis que $\operatorname{Cov}(X, Y) < 0$ signifie que lorsque X est plus grande que sa moyenne (respectivement plus petite) Y a tendance à être plus petite que sa moyenne (respectivement plus grande).

Le coefficient de corrélation mesure les mêmes variations, mais a l'avantage d'être adimensionnel et est toujours compris entre -1 et 1.

Proposition 60. *Soient $X, Y \in L_2(\Omega)$. Alors*

$$|\rho(X, Y)| \leq 1.$$

Proposition admise pour l'instant, on verra une démonstration dans le chapitre suivant.

Définition 61. *X et Y sont dites* décorrélées *si* $\operatorname{Cov}(X, Y) = 0$. *Sinon elles sont dites* corrélées.

Ex. On reprend l'exemple du tirage de 3 boules parmi 5 boules colorées des parties précédentes.

Q. Que vaut $\operatorname{Cov}(X, Y)$? Interpréter le résultat.

.

4.4 Variables aléatoires indépendantes

Définition 62. *Deux variables aléatoires discrètes X et Y sont dites* indépendantes *si, pour tout $x \in X(\Omega)$ et tout $y \in Y(\Omega)$, les événements $\{X = x\}$ et $\{Y = y\}$ sont indépendants (autrement dit $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$).*

Notation. $X \perp\!\!\!\perp Y$.

Ex. On reprend l'exemple de l'urne à 5 boules. Les variables aléatoires X et Y ne sont pas indépendantes, car

$$\mathbb{P}(X = 2, Y = 1) = \frac{1}{10} \neq \frac{3}{10} \times \frac{6}{10} = \mathbb{P}(X = 2)\mathbb{P}(Y = 1).$$

Remarque importante : si X et Y sont indépendantes (et seulement dans ce cas), alors on peut déterminer la loi $\mathbb{P}_{(X,Y)}$ du couple (X, Y) à partir de ses marginales $\mathbb{P}_X$ et $\mathbb{P}_Y$ via la formule

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

Lemme 63. *Soient X, Y deux v.a. réelles indépendantes, appartenant à $L_2(\Omega)$. Alors $XY \in L_1(\Omega)$ et :*

- $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$
- *par conséquent,* $\mathrm{Cov}(X, Y) = 0$.

Démonstration.

$\square$

Attention : la réciproque est fausse, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ n'implique pas en général que X et Y sont indépendantes.

Ex.

Proposition 64. *Soient X, Y deux variables aléatoires indépendantes, G, H deux ensembles et $f : X(\Omega) \to G$, $g : Y(\Omega) \to H$ deux fonctions. Alors les variables aléatoires $f(X)$ et $g(Y)$ sont aussi indépendantes.*

Démonstration.

$\square$

Un exemple d'application si X et Y sont des couples aléatoires : Si $X = (X_1, X_2)$ et $Y = (Y_1, Y_2)$ sont des couples *indépendants* alors :

- X_1 et Y_1 sont indépendantes (appliquer la proposition précédente aux projections sur les premières corrdonnées, $f : (x_1, x_2) \mapsto x_1$ et $g : (y_1, y_2) \mapsto y_1$)
- X_2 et Y_2 sont indépendantes (appliquer la proposition précédente aux projections sur les deuxièmes coordonnées)
- X_1 et Y_2 sont indépendantes (appliquer la proposition précédente à la projection sur la première coordonnée et la projection sur la deuxième coordonnée) et symétriquement X_2 et Y_1 sont indépendantes.

Exercice. Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{N}^* \times \mathbb{N}^*$, dont la loi jointe est définie par :

$$\mathbb{P}(X = i, Y = j) = a^i b^j, \quad a, b > 0.$$

1. Quelles conditions doivent remplir a et b ?
2. Calculer les lois marginales de X et Y.
3. Calculer $\mathbb{P}(X = Y)$.
4. X et Y sont-elles indépendantes ?
5. A quelles conditions sur $c > 0$ et $d > 0$ l'espérance de la variable e^{cX+dY} existe-t-elle ? (Exprimez ces conditions en fonction de a et b.) Lorsque cette espérance existe, la calculer.

4.5 Somme de variables aléatoires

4.5.1 Loi de la somme de deux variables aléatoires

Proposition 65. (i) *Soient X, Y deux variables aléatoires réelles. La loi de la somme $S = X + Y$ est donnée par :*

- $S(\Omega) = \{s \in \mathbb{R} \text{ t.q. } s = x + y, \text{pour un } x \in X(\Omega) \text{ et un } y \in Y(\Omega)\}$
- $\mathbb{P}(S = s) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega) \text{ t.q. } x+y=s} \mathbb{P}(X = x, Y = y), \quad \forall s \in S(\Omega).$

(ii) *En particulier, si X et Y sont indépendantes,*

$$\mathbb{P}(S = s) = \sum_{(x,y) \in X(\Omega) \times Y(\Omega) \text{ t.q. } x+y=s} \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \forall s \in S(\Omega).$$

Cas particulier : si X et Y ne prennent que des valeurs entières positives, alors $X + Y$ prend aussi des valeurs entières positives et pour tout $n \in \mathbb{N}$

$$\mathbb{P}(X + Y = n) = \sum_{i=0}^{n} \mathbb{P}(X = i, Y = n - i).$$

Démonstration de la proposition.

□

4.5.2 Deux exemples fondamentaux

Exemple 1. Soit $p \in [0, 1]$ et $n, m \in \mathbb{N}^*$. Si $X \sim \text{Bin}(n, p)$ et $Y \sim \text{Bin}(m, p)$ sont deux variables aléatoires *indépendantes*, alors $X + Y \sim \text{Bin}(n + m, p)$.

.

Exemple 2. Si $X \sim \mathcal{P}(\lambda)$, $\lambda > 0$, et $Y \sim \mathcal{P}(\mu)$, $\mu > 0$ sont deux variables aléatoires *indépendantes*, alors $X + Y \sim \mathcal{P}(\lambda + \mu)$.

4.5.3 Variance de la somme de deux variables aléatoires

Théorème 66. *Soit X et Y deux variables aléatoires réelles appartenant à $L_2(\Omega)$.*

(i) *Alors,* $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$.

(ii) *Par conséquent, si de plus X, Y sont indépendantes,* $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$.

<u>Démonstration.</u>

Ces formules sont extrêmement utiles en pratique !

4.6 Vecteurs et suites aléatoires

Définition 67. *Si $X_1, \ldots, X_n$ sont des variables aléatoires à valeurs respectivement dans $E_1, \ldots, E_n$, on appelle* vecteur aléatoire $(X_1, \ldots, X_n)$ *l'application*

$$\begin{array}{rcl} \Omega & \to & E_1 \times \ldots \times E_n \\ \omega & \mapsto & (X_1(\omega), \ldots, X_n(\omega)) \end{array}$$

La variable aléatoire X_i est la ième marginale du vecteur.

On définit de même une suite aléatoire en considérant une infinité dénombrable de variables aléatoires $X_i, i \in \mathbb{N}^*$.

Définition 68. (i) *Des variables aléatoires $X_1, \ldots, X_n$ sont dites* indépendantes *si pour tout n-uplet $(x_1, \ldots, x_n) \in X_1(\Omega) \times \ldots \times X_n(\Omega)$, les événements $\{X_1 = x_1\}, \ldots, \{X_n = x_n\}$ sont indépendants.*
(ii) *On dit d'une suite de variables aléatoires $X_i, i \geq 1$ qu'elles sont* indépendantes *si toute sous-suite finie est indépendante au sens ci-dessus.*

Notation : Si des variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et ont toutes la même loi, on dit qu'elles sont *indépendantes et identiquement distribuées*, ce qu'on abrège en i.i.d.

Les résultats vus dans la section précédente se généralisent alors de la façon suivante :

Proposition 69. *Si $X_1, \ldots, X_n$ sont des variables aléatoires indépendantes et $f_i : X_i(\Omega) \to F_i$, $1 \leq i \leq n$ des fonctions, alors les v.a. $f_1(X_1), \ldots, f_n(X_n)$ sont indépendantes.*

Proposition 70. (Indépendance par paquets) *Soit $X_1, \ldots, X_n$ des variables aléatoires indépendantes et $(I_j, j \in J)$ une partition de $\{1, \ldots, n\}$. Alors les variables aléatoires $(X_i, i \in I_j), j \in J$ sont indépendantes.*

Proposition 71. (Somme de variables aléatoires binomiales indépendantes) *Soit $p \in [0,1]$ et $n_i, 1 \leq i \leq n$ des entiers positifs. Si $X_1, \ldots, X_n$ sont des v.a. indépendantes t.q. $X_i \sim \mathrm{Bin}(n_i, p)$, $1 \leq i \leq n$, alors*

$$\sum_{i=1}^{n} X_i \sim \mathrm{Bin}\left(\sum_{i=1}^{n} n_i, p\right).$$

En particulier, si on prend $n_i = 1, 1 \leq i \leq n$, on voit que la somme de n variables aléatoires indépendantes de loi de Bernoulli de paramètre p suit une loi $\mathrm{Bin}(n, p)$.

Proposition 72. (Somme de variables aléatoires de Poisson indépendantes) *Si* $X_1, \ldots, X_n$ *sont des variables aléatoires indépendantes telles que* $X_i \sim \mathcal{P}(\lambda_i)$ *(*$\lambda_i > 0$*),* $1 \leq i \leq n$*, alors*

$$\sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} \lambda_i\right).$$

Proposition 73. (Variance d'une somme)
(i) *Si* $X_1, \ldots, X_n$ *sont des variables aléatoires appartenant toutes à* $L_2(\Omega)$ *alors*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{(i,j)\in\{1,\ldots,n\}^2, i\neq j} \mathrm{Cov}(X_i, X_j).$$

(ii) *Si de plus ces variables aléatoires sont indépendantes, on a*

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Application : la variance d'une loi $\mathrm{Bin}(n, p)$ est $np(1-p)$: il suffit pour cela d'écrire une telle variable aléatoire comme la somme de n variables aléatoires de loi $\mathrm{Ber}(p)$ indépendantes et de se rappeler qu'une telle variable aléatoire de Bernoulli a une variance égale à $p(1-p)$.

Exercice. Vrai ou Faux ?

1. Si $X \sim \mathrm{Ber}(p)$ alors $\mathrm{Cov}(X, X^2) = p(1-p)$.
2. Si X et Y sont dans $L_2(\Omega)$ et telles que $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ alors elles sont indépendantes.
3. Si X et Y sont indépendantes et dans $L_2(\Omega)$ alors $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.
4. Si $X \sim \mathrm{Bin}(2; 0,5)$ et $Y \sim \mathrm{Bin}(3; 0,5)$ alors $X + Y \sim \mathrm{Bin}(5; 0,5)$.
5. Si $X \sim \mathrm{Bin}(2; 0,2)$ et $Y \sim \mathrm{Bin}(2; 0,3)$ sont indépendantes alors $X+Y \sim \mathrm{Bin}(2; 0,5)$.

5 Inégalités classiques et applications

Soit $(\Omega, \mathbb{P})$ un espace probabilisé. Dans toute cette partie il est sous-entendu que les variables aléatoires considérées sont définies sur cet espace et discrètes.

5.1 Inégalités de Markov et Bienaymé-Tchébychev

Théorème 74. *Soit X une variable aléatoire réelle.*

(i) ***Inégalité de Markov.*** *On suppose que $X \in L_1(\Omega)$. Alors, pour tout réel $\varepsilon > 0$,*

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X|]}{\varepsilon}.$$

(ii) ***Inégalité de Bienaymé-Tchébychev.*** *On suppose que $X \in L_2(\Omega)$. Alors, pour tout réel $\varepsilon > 0$,*

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

Démonstration.

L'inégalité de Bienaymé-Tchébychev est utile en pratique pour contrôler l'écart de X à son espérance.

5.2 Application : loi faible des grands nombres

Théorème 75 (Loi faible des grands nombres)**.** *Soit $X_i, i \geq 1$ une suite de variables aléatoires réelles* indépendantes *et de même loi. On suppose qu'elles appartiennent à $L_1(\Omega)$. Alors, pour tout réel $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Ce résultat dit que dans un certain sens, la moyenne des X_i converge vers leur espérance commune lorsque $n \to \infty$. Il existe une version plus forte de ce résultat, la *Loi forte des grands nombres* qui dit que sous les hypothèses du théorème précedent,

$$\mathbb{P}\left(\frac{X_1 + \ldots + X_n}{n} \underset{n\to\infty}{\longrightarrow} \mathbb{E}[X_1]\right) = 1.$$

C'est un des deux grands théorèmes en Probabilités, avec le Théorème Central Limite (TCL). Sa démonstration est considérablement plus compliquée que celle de la loi faible.

<u>Démonstration</u>.

<u>Ex.</u> On jette une pièce une infinité de fois. On suppose que la probabilité d'obtenir “Pile” est $p \in [0, 1]$ à chaque lancer, et que les lancers sont indépendants. On pose $X_i = 1$ si on obtient Pile au $i$ème lancer, $X_i = 0$ sinon, $\forall i \geq 1$. La variable aléatoire $\overline{X}_n$ est donc le nombre moyen de Pile obtenus après les n premiers lancers. D'après la loi faible des grands nombres, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\overline{X}_n \notin]p - \varepsilon, p + \varepsilon[\right) \underset{n\to\infty}{\longrightarrow} 0.$$

5.3 Inégalité de Cauchy-Schwarz

Théorème 76 (Inégalité de Cauchy-Schwarz)**.** *Soient X, Y deux variables aléatoires réelles appartenant à $L_2(\Omega)$. Alors $XY \in L_1(\Omega)$ et*

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}.$$

<u>Démonstration</u>

Corollaire 77. *Soient X, Y deux v.a. réelles appartenant à $L_2(\Omega)$. Alors $|\rho(X, Y)| \leq 1$.*

6 Introduction aux variables aléatoires continues

6.1 Espaces probabilisés : cadre général

Exemples de motivation : lancer d'une fléchette sur un disque ; théorème central limite

Pour pouvoir modéliser une expérience aléatoire dans un cadre *général*, nous aurons besoin de *trois* éléments :

- Un **ensemble**, souvent noté Ω, qui représente l'ensemble de tous les résultats possibles d'une expérience. Dans ce cadre général, n'importe quel ensemble peut être considéré, par exemple $\mathbb{R}$, un intervalle $[a, b]$, $\mathbb{R}^n$, un disque de $\mathbb{R}^2$ (cas de la fléchette), et bien sûr aussi tous les ensembles finis ou dénombrables avec lesquels nous avons travaillé jusqu'à présent.
- Une **tribu**, souvent notée $\mathcal{A}$, qui correspond à l'ensemble des parties de Ω auxquelles on pourra attribuer une probabilité.
- Une **probabilité** sur $\mathcal{A}$, souvent notée $\mathbb{P}$, qui associe à chaque élément de $\mathcal{A}$ un réel entre 0 et 1 qui représente la probabilité que cette partie se réalise.

Définissons formellement les notions de tribu sur un ensemble Ω et de probabilité sur une tribu :

Définition 78. *Soit Ω un ensemble. On appelle* tribu *sur Ω toute famille $\mathcal{A}$ de sous-ensembles de Ω vérifiant :*

(i) $\emptyset \in \mathcal{A}$

(ii) *si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$* [stabilité par passage au complémentaire]

(iii) *si $A_n \in \mathcal{A}, \forall n \in \mathbb{N}^*$, alors $\cup_{n\geq 1} A_n \in \mathcal{A}$* [stabilité par union dénombrable].

Définition 79. *Les éléments de $\mathcal{A}$ s'appelle des* événements.

Exemples de tribus sur un ensemble Ω (quelconque) :

Remarque : en général, si Ω est un ensemble fini ou dénombrable, on travaillera avec la tribu $\mathcal{P}(\Omega)$.

Notons les conséquences suivantes de la définition d'une tribu $\mathcal{A}$ sur Ω :

- (i) et (ii) impliquent que Ω appartient à la tribu $\mathcal{A}$.
- (ii) et (iii) impliquent que $\mathcal{A}$ est *stable par intersection dénombrable* : si $A_n \in \mathcal{A}, \forall n \in \mathbb{N}^*$, alors $\cap_{n\geq 1} A_n \in \mathcal{A}$.
  En effet, par (ii), $A_n^c \in \mathcal{A}, \forall n \in \mathbb{N}^*$ et par conséquent, par (iii), $\cup_{n\geq 1} A_n^c \in \mathcal{A}$. Puis en utilisant à nouveau (ii), $\cap_{n\geq 1} A_n = (\cup_{n\geq 1} A_n^c)^c \in \mathcal{A}$.
- (i) et (iii) impliquent que $\mathcal{A}$ est *stable par union finie* :
  il suffit de prendre $A_n = \emptyset, \forall n \geq n_0$, n_0 étant fixé (et de même, $\mathcal{A}$ est *stable par intersection finie*).

Définition 80. *Soit* Ω *un ensemble et* $\mathcal{A}$ *une tribu sur* Ω*. On appellera* probabilité *sur* $\mathcal{A}$ *toute application*

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

telle que :

- $\mathbb{P}(\Omega) = 1$ *et* $\mathbb{P}(\emptyset) = 0$.
- *Pour toute suite* $(A_n)_{n\geq 1}$ *de parties* deux à deux disjointes *d'éléments de* $\mathcal{A}$,

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Définition 81. *Le triplet* $(\Omega, \mathcal{A}, \mathbb{P})$ *est alors appelé* espace probabilisé.

Remarque importante : toutes les propriétés d'une Probabilité que nous avons dans le cadre discret : propriétés fondamentales, continuité à droite et à gauche, restent valables dans ce cadre plus général. Par exemple :

- $\forall A \in \mathcal{A}$, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- $\forall A, B \in \mathcal{A}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- etc.

Notions de tribu borélienne et de mesure de Lebesgue :

6.2 Variables aléatoires à valeurs réelles : cadre général

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

Définition 82. *Une variable aléatoire à valeurs réelles (v.a.r.) est une fonction* $X : \Omega \to \mathbb{R}$ *telle que, pour tout* $x \in \mathbb{R}$*, l'ensemble* $\{X \leq x\} \in \mathcal{A}$.

On pourra alors attribuer une probabilité à tous les événements $\{X \leq x\}, x \in \mathbb{R}$.

Définition 83. *Etant donnée une variable aléatoire à valeurs réelles* X*, on lui associe sa fonction de répartition* F_X *définie par*

$$\begin{aligned} F_X : \quad & \mathbb{R} \to [0,1] \\ & x \mapsto F_X(x) := \mathbb{P}(X \leq x) \end{aligned}$$

Exemples : tracer la fonction de répartition d'une v.a. X suivant une loi Ber(1/3), puis d'une v.a. Y suivant une loi uniforme sur $\{1, 2, 4\}$.

Proposition 84 (Caractérisation d'une fonction de répartition)**.** *Soit* X *une v.a.r. et* F_X *sa fonction de répartition. Alors :*

- F_X *est croissante sur* $\mathbb{R}$
- $F_X(x) \underset{x\to\infty}{\to} 1$ *et* $F_X(x) \underset{x\to-\infty}{\to} 0$
- F_X *est continue à droite :* $F_X(x+\varepsilon) \underset{\varepsilon\downarrow 0}{\to} F_X(x)$*, pour tout* $x \in \mathbb{R}$.

Réciproquement, si une fonction $F : \mathbb{R} \to \mathbb{R}$ *vérifie les trois points ci-dessus alors il existe une v.a.r. dont* F *est la fonction de répartition.*

Démonstration.

Exercice : les fonctions suivantes, définies sur $\mathbb{R}$, sont-elles des fonctions de répartition ?

1. $F(x) = \exp(-x)$
2. $F(x) = 1 - \exp(-x)$
3. $F(x) = \frac{[x]}{1+2[x]}$ si $x \geq 0$, $F(x) = 0$ si $x < 0$.

.

Proposition 85 (admise)**.** *Une fonction de répartition caractérise la loi : si X et Y sont deux v.a.r. telles que $F_X(x) = F_Y(x)$ pour tout $x \in \mathbb{R}$, alors X et Y ont la même loi.*

6.3 Variables aléatoires réelles continues (ou "à densité")

Dans toute cette partie, $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé.

6.3.1 Définition

Définition 86. *Une v.a.r. est dite* continue *si sa fonction de répartition peut s'écrire sous la forme*

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

pour une fonction $f : \mathbb{R} \to \mathbb{R}$ positive et continue par morceaux.
La fonction f est alors appelée la densité *de X.*

<u>Remarques importantes :</u>

.

▶ En pratique, donner la loi d'une variable aléatoire continue c'est donner 1) soit sa densité, 2) soit sa fonction de répartition, 3) soit reconnaître une loi "classique".

6.3.2 Trois familles de lois continues "classiques"

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle continue.

Définition 87. *Soit $\lambda \in]0, +\infty[$. On dit que X suit une* loi exponentielle de paramètre λ *si elle admet pour densité la fonction f définie par*

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

Dans ce cas, sa fonction de répartition F_X est donnée par $F_X(x) = 1 - \exp(-\lambda x)$ si $x \geq 0$, et $F_X(x) = 0$ si $x < 0$.

<u>Notation :</u> $X \sim \mathcal{E}\mathrm{xp}(\lambda)$.

Cette loi est souvent utilisée pour modéliser le temps qui s'écoule entre des événements imprévisibles (tremblement de terre, pannes internet, durée de vie d'une ampoule).

Cette loi a une *propriété d'absence de mémoire* : pour tous réels $t \geq 0, s \geq 0$,

$$\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t).$$

<u>Démonstration.</u>

.

Définition 88. *Soient* $\mu \in \mathbb{R}$ *et* $\sigma^2 \in]0, \infty[$. *On dit que* X *suit une loi* normale (ou gaussienne) de paramètres (μ, σ^2) *si elle admet pour densité la fonction* f *définie par*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad \forall x \in \mathbb{R}.$$

Notation : $X \sim \mathcal{N}(\mu, \sigma^2)$.

Définition 89. *Lorsque* $\mu = 0$ *et* $\sigma^2 = 1$ *on dit que* X *suit une* loi normale standard *ou* loi normale centrée réduite.

C'est une loi particulièrement importante car elle intervient dans la description asymptotique de sommes de variables aléatoires indépendantes et de même loi. C'est le *Théorème Central Limite*, qui sera vu précisément pendant le cours de Statistiques du deuxième semestre.

Remarque :

.

Définition 90. *Soient a, b des réels, tels que $a < b$. On dit que X suit une loi* uniforme sur $[a, b]$ *si elle admet pour densité la fonction f définie par*

$$f(x) = \begin{cases} \frac{1}{b-a} & si\ x \in [a, b] \\ 0 & sinon. \end{cases}$$

<u>Notation</u> : $X \sim \mathcal{U}[a, b]$.

<u>Exercice</u> : calculer et tracer la fonction de répartition d'une v.a. suivant la loi $\mathcal{U}[a, b]$.

<u>Exercice</u> : si $X \sim \mathcal{U}[0, 1]$, quelle est la loi de X^2 (vous commencerez par calculer sa fonction de répartition, puis en déduirez sa densité)

.

6.3.3 Quelques propriétés des lois continues

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle continue, de densité f.

Proposition 91. *Alors,*

- *Sa fonction de répartition* F_X *est continue sur* $\mathbb{R}$.
- $\mathbb{P}(X = x) = 0$ *pour tout* $x \in \mathbb{R}$.
- *pour tous réels* a, b *tels que* $a < b$,

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b) = \int_a^b f(t)\mathrm{d}t.$$

Démonstration.

Remarque : il existe des v.a.r. qui ne sont ni discrètes, ni continues.

6.3.4 Notion d'espérance d'une variable aléatoire continue

Soit X une variable aléatoire continue de densité f.

Définition 92. *Si* $\int_{-\infty}^{\infty} tf(t)\mathrm{d}t$ *est* absolument convergente, *on dit que* X *admet une* espérance. *Cette espérance, notée* $\mathbb{E}[X]$, *est alors définie par*

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} tf(t)\mathrm{d}t.$$

La proposition et le théorème suivants sont admis.

Proposition 93. *Les propriétés de linéarité, positivité, inégalité triangulaire, etc., que l'on a vues pour l'espérance de variables aléatoires discrètes restent valides pour l'espérance de variables aléatoires continues.*

Théorème 94 (Théorème de transfert)**.** *Soit* $g : \mathbb{R} \to \mathbb{R}$ *une fonction continue telle que* $\int_{-\infty}^{\infty} g(t)f(t)\mathrm{d}t$ *est absolument convergente. Alors la variable aléatoire* $g(X)$ *admet une espérance et celle-ci peut se calculer ainsi :*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(t)f(t)\mathrm{d}t.$$

Ex. : si $\int_{-\infty}^{\infty} t^2 f(t)\mathrm{d}t$ est absolument convergente alors X^2 admet une espérance et

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} t^2 f(t)\mathrm{d}t.$$

Définition 95. *Si* $\int_{-\infty}^{\infty} t^2 f(t)\mathrm{d}t$ *est absolument convergente, on définit la variance de* X *par*

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

En développant le carré et en utilisant la linéarité de l'espérance, on voit que

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} t^2 f(t)\mathrm{d}t - \left(\int_{-\infty}^{\infty} t f(t)\mathrm{d}t\right)^2.$$

<u>Exercice</u> : calculer l'espérance et la variance d'une variable aléatoire X :

1. suivant une loi uniforme sur $[a; b]$, avec $a < b$
2. suivant une loi exponentielle de paramètre $\lambda > 0$
3. suivant une loi normale centrée réduite.