Introduction à la théorie des probabilités

Notes de cours et exercices

Lamia Bazzaoui

Cours Semestres d'automne 2023, 2024

Ecole Nationale de Commerce et de Gestion

Contents

Introduction

Ce document constitue un recueil de notes de cours et d'exercices du cours de *Probabilités et Statistique*, tel qu'il a été enseigné au niveau licence au cours des semestres d'automne 2023 et 2024 à l'ENCG de Casablanca.

L'objectif principal de ce support est d'accompagner les étudiantes et étudiants dans l'apprentissage progressif des concepts fondamentaux du calcul des probabilités et des variables aléatoires. Le cours met l'accent sur la compréhension des notions, la maîtrise des outils mathématiques de base, ainsi que sur leur mise en œuvre à travers de nombreux exercices.

Le contenu couvre notamment la théorie des ensembles et des événements, les principes de dénombrement, les axiomes de probabilité, les probabilités conditionnelles et le théorème de Bayes, les variables aléatoires discrètes et continues, les lois usuelles, les vecteurs aléatoires, l'indépendance, ainsi que les notions de covariance et de corrélation.

Ce document est destiné en priorité aux étudiants de premier cycle universitaire. Il a été conçu comme un support pédagogique autonome, permettant à la fois de suivre le cours et de réviser les notions abordées. Le niveau de présentation se veut volontairement progressif et accessible, sans sacrifier la rigueur mathématique.

Ces notes ne prétendent pas à l'exhaustivité et ne remplacent ni les séances de cours ni les travaux dirigés. Elles constituent toutefois une base solide pour l'étude des probabilités, et peuvent servir de référence pour des cours ultérieurs en statistique, économétrie ou analyse des données.

Chapter 1

Rappels sur la théorie des ensembles

La théorie des probabilités est le fondement sur lequel repose la statistique, fournissant des outils pour modéliser ce qui peut être considéré comme un phénomène aléatoire. Grâce à ces modèles, les statisticiens sont en mesure de tirer des conclusions concernant une population partant des données d'un échantillon. Tout comme la statistique s'appuie sur la théorie des probabilités, la théorie des probabilités repose à son tour sur la théorie des ensembles, dont nous revoyons quelques principes dans ce premier chapitre.

1.1 La notion d'ensemble

Un ensemble est une collection d'objets appelés "éléments". L'énoncé $x \in E$ signifie "x appartient à E" et $x \notin E$ veut dire que E ne contient pas x. On peut définir l'ensemble de deux manières:

- en extension (Set-roster notation): en énumérant entre accolades ses éléments (exemple: $E = \{a, b, c\}$, ou bien $A = \{1, 2, 3, \ldots, 100\}$, ou $B = \{1, 2, 3, \ldots\}$). A noter que seule l'appartenance à l'ensemble importe: il n' y a aucun ordre et les répétitions sont possibles (exemple: $\{a, b\} = \{b, a\} = \{a, b, a\}$). La définition en extension est préférée lorsque l'ensemble est petit ou lorsque les éléments peuvent être facilement énumérés.

- en compréhension (Set-builder notation): E est défini comme le sous-ensemble formé des éléments x d'un ensemble connu X vérifiant une certaine propriété $p(x)$. On note $E = \{x \in X : p(x)\}$ ou $E = \{x \in X|p(x)\}$.[1] Exemple: $E = \{x \in \mathbb{N}|\exists n \in \mathbb{N}, x = n^2 + 1\}$.Cette approche est pratique pour définir des ensembles infinis ou lorsqu'il est plus simple de décrire une propriété commune que de lister les éléments.

[1] E est l'ensemble des x dans X tels que $p(x)$ (ou vérifiant $p(x)$)

Quelques ensembles connus:

- $\emptyset$: l'ensemble vide. Il s'agit de l'ensemble qui ne contient aucun élément $\emptyset = \{\}$
- $\mathbb{N}$: ensemble des nombres naturels. $\mathbb{N} = \{0, 1, 2, 3, \dots\}$
- $\mathbb{Z}$: ensemble des nombres entiers relatifs. $\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
- $\mathbb{Q}$: ensemble des nombres rationnels. $\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z}, b \neq 0\}$. Les nombres rationnels sont ceux qui peuvent être exprimés comme le quotient de deux entiers, où le dénominateur n'est pas nul.
- $\mathbb{R}$: ensemble des nombres réels. Cet ensemble représente tous les points sur la ligne numérique continue. Il inclut tous les nombres rationnels et irrationnels.[2]
- $\mathbb{C}$: ensemble des nombres complexes. $\mathbb{C} = \{a + bi | a, b \in \mathbb{R}, i = \sqrt{-1}\}$. Les nombres complexes sont ceux qui comprennent une partie réelle a et une partie imaginaire bi. L'unité imaginaire i est définie comme $\sqrt{-1}$.

Ces ensembles sont caractérisés par la relation suivante[3]

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$$

1.2 Inclusion, égalité et diagramme de Venn

Définition 1

On considère les ensembles A et B. A est un sous-ensemble de B (on note $A \subseteq B$), si et seulement si, chaque élément de A est aussi un élément de B. Formellement:

$$A \subseteq B \Leftrightarrow (\forall x, x \in A \Rightarrow x \in B)$$

Le symbole $\nsubseteq$ veut dire "pas un sous-ensemble de". Par exemple, si $A \nsubseteq B$, cela signifie que: $\exists\ x$ tel que $x \in A$ et $x \notin B$. Un autre concept est celui de sous-ensemble **strict**: A est un sous-ensemble propre ou strict de B (on note $A \subset$ B), si $A \subseteq B$ mais $A \neq B$ ($\exists\ x$ tel que $x \in B$ et $x \notin A$). Le symbole $\not\subset$ veut dire "pas un sous-ensemble propre de".

[2] Les nombres irrationnels sont ceux qui ne peuvent pas être exprimés comme une fraction de deux entiers. Par exemple, $\sqrt{2}, \pi, e$ sont des nombres irrationnels.

[3] L'ensemble vide est un sous-ensemble de tout ensemble.

Définition 2

Soient les ensembles A et B. $A = B$ si et seulement si $A \subseteq B$ et $B \subseteq A$. On note

$$A = B \Leftrightarrow (A \subseteq B \text{ et } B \subseteq A)$$

Quelques propriétés

$A \subseteq A$: puisque chaque élément de A est un élément de A (voir définition 1)

$\emptyset \subseteq A$: $\emptyset$ ne contient aucun élément donc il s'agit d'une *vérité creuse*.

Soient les 3 ensembles A, B et C:

$$(A \subseteq B \text{ et } B \subseteq C) \Rightarrow A \subseteq C$$

Diagramme de Venn

Pour illustrer les relations logiques et quelques opérations entre ensembles, un outil pratique est le diagramme de Venn. Les diagrammes de Venn ont été conçus autour de 1880 par John Venn. Considérons les sous-ensembles A, B et C, ci-dessous quelques exemples de représentations graphiques:

Figure 1: $A \subset B$, A et C n'ont pas d'éléments en commun, B et C ont quelques éléments en commun mais $B \not\subset C$ et $C \not\subset B$.

Figure 2: $A \subseteq B$ et $B \subseteq C$ (voir propriétés)

FIGURE 1.1

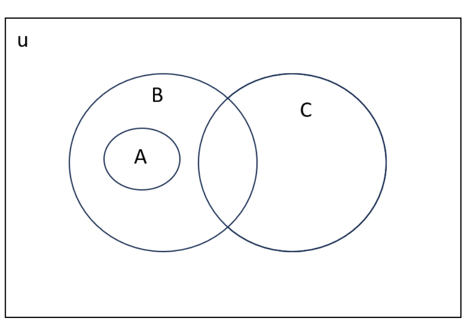
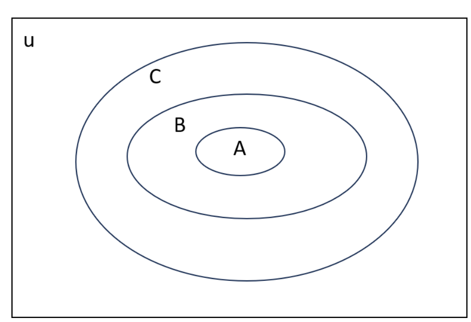


FIGURE 1.2

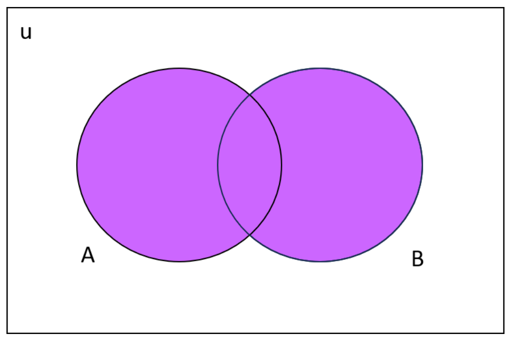


1.3 Union, intersection, différence et complément

1.3.1 Union

Définition 3

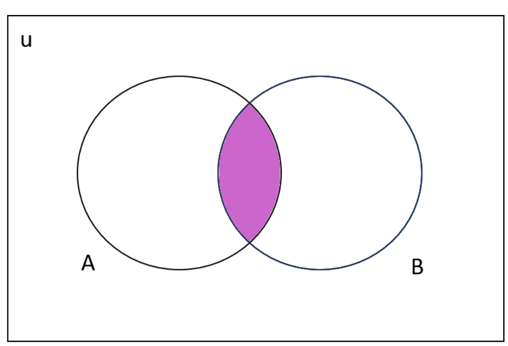
Soient A et B deux ensembles. Il existe un unique ensemble dont les éléments sont ceux de A et B. On l'appelle union de A et B et on le note $A \cup B$. Si A et B correspondent à deux sous-ensembles de Ω (qui est l'ensemble universel u), on définit leur union comme suit: $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ ou } \omega \in B\}$



1.3.2 Intersection

Définition 4

Soient A et B deux ensembles. Il existe un unique ensemble formé par les éléments qui sont communs à A et à B. On l'appelle intersection de A et B et on le note $A \cap B$. Si A et B correspondent à deux sous-ensembles de Ω, on définit leur intersection comme suit: $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ et } \omega \in B\}$

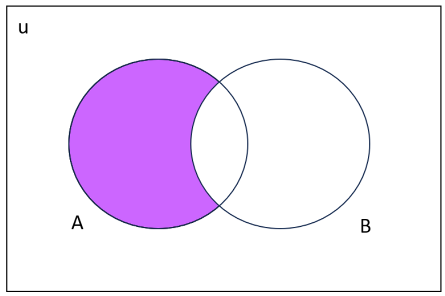


1.3.3 Différence de deux ensembles

Définition 5

Soient A et B deux ensembles. On appelle **différence** de A et B l'ensemble des éléments de A qui n'appartiennent pas à B. On le note $A \backslash B$ (A privé de B). Si A et B sont deux sous-ensembles de Ω, on aura $A \backslash B = A \cap B^c = \{\omega \in \Omega | \omega \in A \text{ et } \omega \notin B\}$

$A \backslash B$



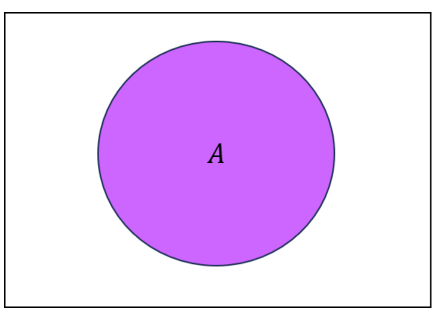
1.3.4 Complémentaire d'un ensemble

Définition 6

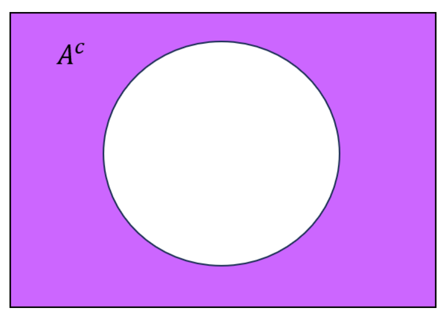
Soit $A \subset E$, le **complémentaire** de A dans E correspond à l'ensemble $E \backslash A$. On le note A^c (on utilise parfois aussi la notation $\bar{A}$).

Si $A \subset \Omega$, le complémentaire de A est défini comme suit: $A^c = \{\omega \in \Omega \mid \omega \notin A\}$.

Ensemble A



Ensemble A^c



1.4 Quelques propriétés

1.4.1 Propriétés liées aux opérations d'union et intersection

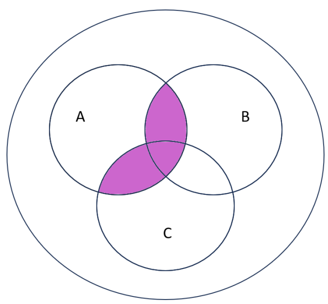
- Les opérations d'union et d'intersection sont **commutatives** et **associatives**:

  $A \cup B = B \cup A$; $A \cap B = B \cap A$

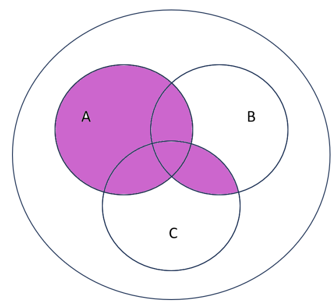
  $(A \cup B) \cup C = A \cup (B \cup C)$; $(A \cap B) \cap C = A \cap (B \cap C)$

- **Distributivité**

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$



$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$



- **Idempotence**

  $A \cup A = A$, $A \cap A = A$

- **Autres propriétés**

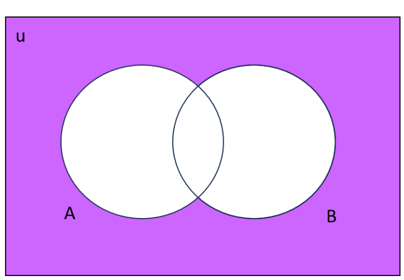
  $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$; $A \cup \Omega = \Omega$, $A \cap \Omega = A$

1.4.2 Propriétés liées à l'ensemble complémentaire

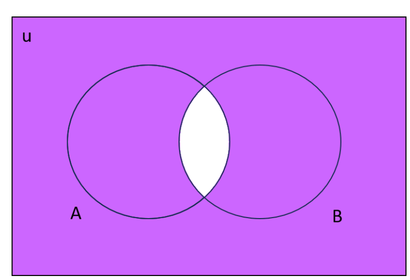
Dans Ω, soit $A \subseteq \Omega$, on a les égalités suivantes: $(A^c)^c = A$, $\emptyset^c = \Omega$, $\Omega^c = \emptyset$, $A \cup A^c = \Omega$, $A \cap A^c = \emptyset$

1.4.3 Loi de De Morgan

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



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



1.5 Extension à une collection d'ensembles

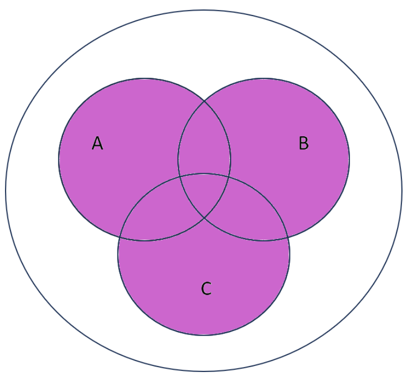
Toute famille non vide d'ensembles $\{A_i, i \in I = \{0, 1, 2, \ldots, n\}\}$ faisant partie de l'ensemble universel Ω possède des unions et des intersections:

- $\bigcup_{i=0}^{n} A_i = \{x \in \Omega | x \in A_i$ pour au moins un $i = 0, 1, 2, \ldots, n\}$
- $\bigcap_{i=0}^{n} A_i = \{x \in \Omega | x \in A_i$ pour chaque $i = 0, 1, 2, \ldots, n\}$

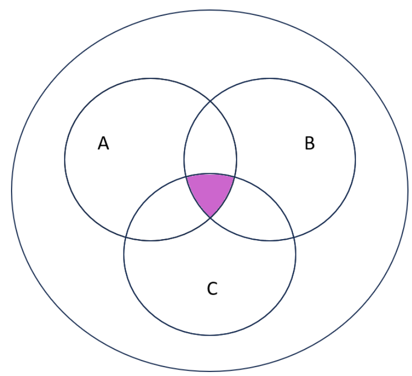
On peut également avoir une extension à l'infini.

Exemple avec I de cardinalité 3

$\bigcup_{i=1}^{3} A_i = A_1 \cup A_2 \cup A_3$



$\bigcap_{i=1}^{3} A_i = A_1 \cap A_2 \cap A_3$



1.6 Partitions

Définition 7

Deux ensembles A, B sont dits **disjoints** lorsqu'ils n'ont aucun élément en commun, c'est-à-dire $A \cap B = \emptyset$

Généralisation à plus de 2 ensembles

Définition 8

Les ensembles $A_1, A_2, A_3, \ldots$ sont dits **mutuellement disjoints** (ou exclusifs) si et seulement si et seulement si pour tout couple d'ensembles distincts A_i et A_j avec $i \neq j$, l'intersection $A_i \cap A_j$ est vide, c'est-à-dire $A_i \cap A_j = \emptyset$. Cela signifie qu'aucun élément ne peut appartenir à plus d'un ensemble parmi $A_1, A_2, A_3, \ldots$.

Définition 9: Partition d'ensemble

Une partition d'un ensemble E est une collection d'ensembles non vides $\{A_1, A_2, \ldots, A_n\}$ tels que:

- Les ensembles A_i sont **mutuellement disjoints**, c'est-à-dire $A_i \cap A_j = \emptyset$ pour $i \neq j$
- L'union de tous les ensembles A_i couvre l'ensemble E, c'est-à-dire $A_1 \cup A_2 \cup \cdots \cup A_n = E$

En d'autres termes, une partition divise l'ensemble E en sous-ensembles qui ne se chevauchent pas et dont l'union reconstitue l'ensemble entier.

Définition 10: l'ensemble puissance (ensemble des parties d'un ensemble)

L'**ensemble puissance** (ou **ensemble des parties**) d'un ensemble E, noté $\mathcal{P}(E)$ est l'ensemble de tous les sous-ensembles possibles de E, y compris l'ensemble vide et l'ensemble E lui-même.

Formellement, si E est un ensemble, l'ensemble puissance $\mathcal{P}(E)$ est défini comme:

$$\mathcal{P}(E) = \{A | A \subseteq E\}$$

Cela signifie que chaque élément de $\mathcal{P}(E)$ est un sous-ensemble de E et le nombre total de sous-enembles est 2^n où n est le nombre d'éléments de E

1.7 Opérations sur les cardinaux

Le cardinal d'un ensemble fini correspond au nombre total de tous ses éléments.

Différentes notations sont utilisées pour désigner le cardinal; par exemple, pour un ensemble E:

$$Card(E) = \#E = |E|$$

Propriétés

Soient A et B deux ensembles inclus dans l'ensemble universel Ω, on a

$$Card(A \cup B) = Card(A) + Card(B) - Card(A \cap B)$$

$$Card(A^c) = Card(\Omega) - Card(A)$$

$$Card(A) = Card(A \backslash B) + Card(A \cap B)$$

Exercises

Exercise 1.1. Redéfinir les ensembles suivants en utilisant la notation en extension :

$$D = \dots \qquad E = \dots \qquad F = \dots \qquad G = \dots \qquad H = \dots \qquad I = \dots$$

Soit l'ensemble $A = \dots$. Est-ce que $\dots \in A$?

Exercise 1.2. Soient les ensembles A, B et C.

Les affirmations suivantes sont-elles vraies ou fausses ? Justifier votre réponse.

Exercise 1.3. Parmi les affirmations suivantes, lesquelles sont vraies ? Expliquer.

Exercise 1.4. Soient les ensembles A et B définis comme suit :

$$A = \dots \qquad B = \dots$$

Démontrer que $A \subset B$ et que $A \cap B = A$.

Exercise 1.5. Soient les ensembles A et B définis comme suit :

$$A = \dots \qquad B = \dots$$

Est-ce que $A = B$?

Exercise 1.6. En utilisant le diagramme de Venn, représenter les cas suivants :

1. $A \cap B = \emptyset$
2. A et C n'ont pas d'éléments en commun, B et C ont quelques éléments en commun
3. $(A \cap B) \cup C$
4. $A \cap (B \cup C)$

Exercise 1.7. Soient les ensembles A, B, C.

Donner $A \cup B$, $A \cap C$, $B \setminus C$.

Si $A \subset B$, quel est $A \cap B$?

En utilisant le diagramme de Venn, représenter $A \cup C$ et $A \cap C$.

Exercise 1.8. On considère que l'ensemble universel est $U = \ldots$. Soit :

$$A = \ldots \qquad B = \ldots \qquad C = \ldots$$

Trouver A^c, B^c, $A \cap B$ et $A \cup C$.

Exercise 1.9. Soient A, B, C et D. Trouver :

$$A \cup B, \quad A \cap C, \quad (A \cup B)^c, \quad A \setminus D$$

Exercise 1.10. Soient A, B, C et D. Trouver :

$$(A \cap B) \cup C, \quad A \cap (B \cup C)$$

Exercise 1.11. Pour chaque entier positif i, soit $A_i = \ldots$.

Trouver $\bigcup_i A_i$ et $\bigcap_i A_i$.

Trouver $\bigcup_{i=1}^n A_i$ et $\bigcap_{i=1}^n A_i$.

Exercise 1.12. Soient A et B. Est-ce que A et B sont disjoints ?

Soient A, B et C. Est-ce qu'ils sont mutuellement disjoints ?

Soient A, B, C et D. Est-ce que $\{B, C, D\}$ est une partition de A ?

Trouver l'ensemble puissance de A, noté $\mathcal{P}(A)$.

Exercise 1.13. Une classe comporte 30 étudiants inscrits soit en mathématiques, soit en économie, soit dans les deux.

- 14 sont inscrits en mathématiques
- 10 sont inscrits en mathématiques mais pas en économie

1. Combien sont inscrits à la fois en mathématiques et en économie ?
2. Combien sont inscrits en économie mais pas en mathématiques ?

Exercise 1.14. Dans un examen, 100 étudiants ont obtenu plus de 80% en économie ou en comptabilité.

- 70 ont eu plus de 80% en économie
- 20 ont eu plus de 80% dans les deux matières

Combien d'étudiants ont eu plus de 80% seulement en comptabilité ?

Chapter 2

Dénombrement

2.1 Expériences en plusieurs étapes

2.1.1 Définition

Considérons l'exemple de lancer de deux pièces de monnaie. Les résultats de l'expérience correspondent à Pile ou Face pour chacune des pièces.

$$S = \{(F, F), (P, F), (F, P), (P, P)\}$$

Quatre résultats sont possibles qu'on peut énumérer. La règle de comptage des expériences à plusieurs étapes consiste à dénombrer les résultats possibles sans les énumérer. Cette règle correspond au principe fondamental du dénombrement et est comme suit:

Principe fondamental du dénombrement

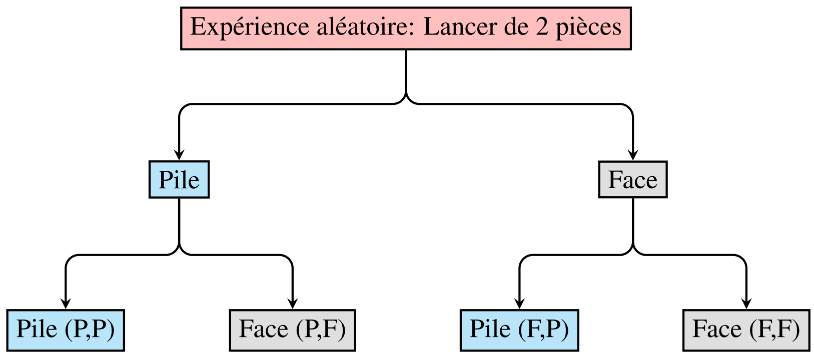
Si une expérience peut être décrite par une séquence de k étapes, avec n_1 résultats possibles à la première étape, n_2 résultats possibles à la seconde étape et ainsi de suite, alors le nombre total de résultats possibles de l'expérience est égal à $(n_1)(n_2)\dots(n_k)$.

Dans l'exemple précédent, on peut considérer les deux lancers comme une expérience à deux étapes: un premier lancer avec deux résultats possibles ($n_1 = 2$), un deuxième lancer ($n_2 = 2$). D'après la règle de comptage, l'expérience a quatre résultats possibles différents ($n_1 \times n_2 = 2 \times 2 = 4$). De la même manière, le nombre de résultats possibles pour 6 lancers de pièces serait de $2^6 = 64$.[1]

[1] Autre exemple: dans le problème des partis vu en introduction du cours, on avait calculé le nombre de résultats possibles pour les manches restantes comme étant 2^{i+j-1}

2.1.2 Diagramme arborescent

On peut visualiser graphiquement une expérience à plusieurs étapes à l'aide d'un diagramme arborescent.



2.1.3 Exemples

Exemple 1

Si l'on doit choisir un mot de passe contenant 5 lettres à partir d'un alphabet de 26 lettres. On aura $26 \times 26 \times 26 \times 26 \times 26 = 26^5 = 11\ 881\ 376$ si on peut répéter la même lettre plusieurs fois. Si les lettres doivent être différentes, alors il y a $26 \times 25 \times 24 \times 23 \times 22 = 7\ 893\ 600$ possibilités.

Exemple 2

Le nombre de codes à 4 chiffres possibles pour une carte bancaire est de $10 \times 10 \times 10 \times 10 = 10\ 000$. Mais si on ne peut pas avoir de chiffres qui se répétent, ce nombre sera de $10 \times 9 \times 8 \times 7 = 5\ 040$.

2.2 Arrangements

2.2.1 Arrangements sans répétition

Soit k et n deux entiers naturels. On appelle arrangement de k éléments parmi n toute application injective[2] de $\mathbb{N}_k$ dans $\mathbb{N}_n$. Puisque la donnée d'une application injective de $\mathbb{N}_k$ dans l'ensemble E revient à choisir k éléments de E avec ordre, un arrangement de k éléments parmi n peut être défini comme un $k-$uplet[3] d'éléments de E, mais sans répétition. On a la proposition suivante:

[2] Voir rappel en annexe 1

[3] En mathématiques, un uplet est une collection ordonnée finie d'objets.

Arrangements sans répétition

Le nombre d'arrangements possibles de k éléments parmi n vaut

$$A_n^k = 0 \text{ si } k > n, \qquad A_n^k = \frac{n!}{(n-k)!} = n(n-1) \times (n-k+1) \text{ si } k \leq n$$

Où $n! = n(n-1)(n-2)\dots(2)(1)$, $k! = k(k-1)(k-2)\dots(2)(1)$ et $0! = 1$,

Notez que $A_n^n = n!$. On a également $A_n^0 = 1$

Exemple 1: Les arrangements de 2 éléments pris dans l'ensemble $\{1,2,3,4\}$ sont: $\{1,2\}$, $\{1,3\}$, $\{1,4\}$,$\{2,1\}$, $\{2,3\}$,$\{2,4\}$, $\{3,1\}$, $\{3,2\}$, $\{3,4\}$,$\{4,1\}$, $\{4,2\}$, $\{4,3\}$. On note qu'il y a au total 12 possibilités. En utilisant la formule, on a:

$$A_4^2 = \frac{4!}{(4-2)!} = \frac{4 \times 3 \times 2!}{2!} = 12$$

Exemple 2: 5 individus se présentent à un concours. L'individu classé premier remportera un prix de 500 dollars et le deuxième un prix de 250 dollars. Combien y a-t-il de remises de prix possibles?

$$A_5^2 = 2! \begin{pmatrix} 5 \\ 2 \end{pmatrix} = \frac{5!}{(5-2)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)} = 20$$

On peut aussi reprendre les exemples précédents:

Pour choisir un mot de passe contenant 5 lettres **distinctes** de l'alphabet de 26 lettres, on aura $26 \times 25 \times 24 \times 23 \times 22 = 7\ 893\ 600$ ou bien $A_{26}^5 = \frac{26!}{(26-5)!} = \frac{26 \times 25 \times 24 \times 23 \times 22 \times 21!}{21!} = 7\ 893\ 600$ possibilités.

Le nombre de codes à 4 chiffres **distincts** possibles pour une carte bancaire est de $10 \times 9 \times 8 \times 7 = 5\ 040$. En utilisant la formule ci-dessus, on aura $A_{10}^4 = \frac{10!}{(10-4)!} = \frac{10 \times 9 \times 8 \times 7 \times 6!}{6!} = 5\ 040$

2.2.2 Arrangements avec répétition

Un arrangement avec répétition de k éléments choisis parmi n est une liste ordonnée avec répétition éventuelle des éléments.

Arrangements avec répétition

Le nombre de ces arrangements est comme suit

$$n \times n \times n \times \dots = n^k$$

Exemple: Un numéro de téléphone comporte 8 chiffres. Combien y a-t-il de numéros de téléphone possibles? Puisque le nombre de chiffres est de 10 (0,1,2...,9), alors on aura $10^8 = 100\,000\,000$.

On peut aussi reprendre les exemples précédents

Si l'on doit choisir un mot de passe contenant 5 lettres à partir d'un alphabet de 26 lettres. On aura $26 \times 26 \times 26 \times 26 \times 26 = 26^5 = 11\ 881\ 376$ si on peut répéter la même lettre plusieurs fois. Ou alors, en utilisant la formule: 26^5

Le nombre de codes à 4 chiffres possibles pour une carte bancaire est de $10 \times 10 \times 10 \times 10 = 10\ 000$. Ou alors, en utilisant la formule 10^4.

Dans ce cas, k n'est pas forcément plus petit que n. De manière générale, le nombre de n-uplets (n^k) correspond au nombre de séries de longueur k composées de n éléments.

2.3 Permutations

La règle de comptage par permutations permet de calculer le nombre de possibilités d'ordonner n éléments parmi n. La méthode de calculer est comme suit

Permutations

On définit une bijection de l'ensemble fini E vers l'ensemble $\{1, 2, 3 \ldots, n\}$. Le nombre de permutations pour E est

$$P = n \times (n-1) \times (n-2) \cdots \times 1 = n!$$

Notez que la permutation est un cas particulier d'arrangement sans répétition A_n^n

Exemple 1: On a un jeu de 4 cartes $E = \{C, D, H, S\}$. Quel est le nombre de permutations possibles?

Suivant la formule, on a $P = 4! = 4 \times 3 \times 2 \times 1 = 24$. Si on veut les lister, on aura

CDHS	CDSH	CHDS	CHSD	CSHD	CSDH
DCHS	DCSH	DHCS	DHSC	DSHC	DSCH
HDCS	HDSC	HCDS	HCSD	HSCD	HSDC
SDHC	SDCH	SHDC	SHCD	SCHD	SCDH

Exemple 2:

On a invité 8 familles amies à une fête. Chaque famille va être installée dans une table à part. Combien y a-t-il de manières de les organiser?

On aura $P = 8! = 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 40\ 320$ dispositions possibles.

2.4 Combinaisons

Soient n et k deux entiers naturels. On appelle combinaison de k objets parmi n toute partie à k éléments d'un ensemble de n objets.

Combinaisons

Le nombre de combinaisons de k objets parmi n vaut

$$C_n^k = \binom{n}{k} = 0 \text{ si } k > n, \qquad C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n}{1}\frac{n-1}{2}\frac{n-2}{3}\cdots\frac{n-k+1}{k} \text{ si } k \leq n$$

On a : $C_n^0 = C_n^n = 1$, $C_n^1 = C_n^{n-1} = n$. On a également[4] $C_n^{n-k} = C_n^k$

Notez qu'on peut aussi utiliser la formule suivante

$$C_n^k = \frac{A_n^k}{k!}$$

Exemple 1: Dans le cadre d'un contrôle qualité, un inspecteur sélectionne aléatoirement deux pièces sur cinq pour tester leur qualité. Combien y a-t-il de combinaisons possibles?

Sur la base de la règle ci-dessus, nous avons

$$C_5^2 = \binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{2 \times 1 \times (3 \times 2 \times 1)} = 10$$

Une expérience aura toujours plus d'arrangements sans répétition que de combinaisons pour un même nombre d'objets sélectionnés, car pour chaque tirage de k objets, il y a $k!$ façons différentes de les ordonner (comme expliqué dans la partie permutations), et donc $C_n^k = \frac{A_n^k}{k!}$.

Exemple 2

Un joueur au loto doit sélectionner les 6 bons numéros parmi 47. Quel est le nombre de combinaisons possibles de 6 chiffres parmi 47?

[4]Pour rappel: on a vu que les valeurs sur le triangle de Pascal sont symétriques

Ce nombre sera de

$$C_{47}^{6} = \frac{47!}{6!(47-6)!} = \frac{47 \times 46 \times 45 \times 44 \times 43 \times 42 \times 41}{6! \times 41} = 10\ 737\ 573$$

2.5 Le binôme de Newton

Le terme C_n^k correspond aux coefficients binomiaux dans la formule du binôme de Newton. La formule du binôme de Newton permet de trouver le développement d'une puissance entière quelconque d'un binôme.

Formule du binôme

$\forall (x, y) \in \mathbb{R}^2, \forall n \in \mathbb{N}$:

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^{n-k} y^k$$

Les coefficients binomiaux peuvent aussi être obtenus à l'aide du triangle de Pascal, en utilisant la formule.

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

Corollaire: Comment calculer la somme des coefficients binomiaux $\sum_{k=0}^{n} C_n^k, \forall n \in \mathbb{N}$?

En prenant $x = y = 1$:

$$\sum_{k=0}^{n} C_n^k \times 1^{n-k} \times 1^k = (1+1)^n = 2^n$$

Cela correspond au nombre de résultats possibles pour une expérience répétée n fois, avec 2 issues possibles à chaque répétition (exemple: lancer de pièce). Ce résultat est aussi visible au niveau du triangle de Pascal (somme de chaque ligne = 2^n)

2.6 Résumé

Le tableau ci-dessous résume les différentes formules de dénombrement

	Avec répétition	**Sans répétition**
Avec ordre	n^p	$A_n^p = p!\binom{n}{p} = \frac{n!}{(n-p)!}$
Sans ordre	$\bar{C}_n^k = \binom{n+k-1}{k}$	$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

Quelques remarques:

- Le cas ”sans ordre et avec répétition” n’a pas été expliqué dans les sections précédentes, mais il correspond au cas des combinaisons avec répétition (un exemple est donné en annexe).

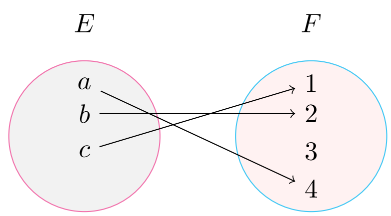
- La permutation est un cas particulier de la case ”avec ordre et sans répétition” ($A_n^n = n!$)

Annexe 1: applications injectives, surjectives, bijectives

Soit l’application $f : E \to F$

Définition d’une application injective

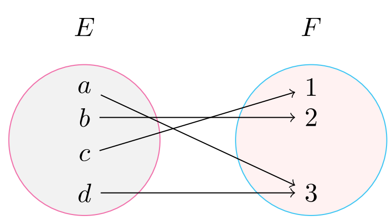
f est **injective** si $\forall x, x' \in E$ on a $(f(x) = f(x') \Rightarrow x = x')$



Cette application est **injective**: notez que chaque élément de F a **au plus** un antécédent

Définition d’une application surjective

f est **surjective** si $\forall\, y \in F, \exists\, x \in E$ tel que $(y = f(x))$

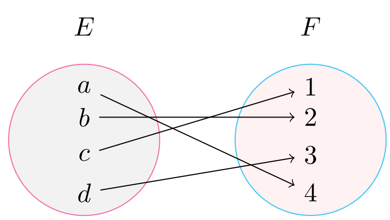


Cette application est **surjective**: notez que chaque élément de F a **au moins** un antécédent.

Définition d’une application bijective

f est **bijective** si $\forall\, y \in F, \exists\, !x \in E$ tel que $(y = f(x))$

Cette application est **bijective** parce que tout élément de F a **un unique** antécédent par f. Elle est à la fois injective et surjective.


E
F
a
b
c
d
1
2
3
4

Annexe 2: un exemple de combinaisons avec répétitions

On veut choisir 2 fruits parmi une variété de 4. Quel est le nombre de combinaisons possibles?

Pomme, Pomme	Pomme, Ananas	Pomme, raisins	Pomme, Poire
Ananas, Ananas	Ananas, raisins	Ananas, Poire	
Raisins, raisins	Raisins, Poire		
Poire, poire			

La solution est $C^2_{4+2-1} = \begin{pmatrix} 4+2-1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix} = 10.$

Si les répétitions n'avaient pas été possibles, on aurait eu $C^2_4 = 6$ (on exclut la première colonne du tableau ci-dessus).

Exercises

Principes additif, multiplicatif et de soustraction

Exercise 2.1. La plaque d'immatriculation d'un véhicule personnel contient typiquement trois blocs :

- un bloc de 5 chiffres,
- un bloc avec une seule lettre arabe (parmi un ensemble de 6 lettres),
- un bloc représentant l'identifiant de la préfecture d'émission (un code allant de 1 à 89).

Quel est le nombre de plaques différentes possibles ?

Exercise 2.2.

1. Un cadenas à code se compose de 5 roues dont chacune peut prendre une valeur de 0 à 9. Combien y a-t-il de codes possibles ?
2. La façade d'un immeuble à 6 étages compte à chaque étage une fenêtre, dont la lumière peut être soit éteinte soit allumée. De combien de manières peut se présenter cette façade ?
3. Soit un compteur binaire à 3 positions. Chaque position peut prendre la valeur 0 ou 1. Combien de configurations différentes sont possibles ?

Exercise 2.3. Soit l'ensemble $E = \{\dots\}$.

1. Combien de mots de 3 lettres peut-on créer à partir des lettres de E si aucune lettre ne peut être utilisée plus d'une fois ?

2. Combien de sous-ensembles peut-on constituer à partir de E ?
3. Combien de mots de 5 lettres peut-on créer à partir des lettres de E si aucune lettre ne peut être répétée et si les lettres *A* et *B* doivent toujours être placées côte à côte ?

Exercise 2.4. Pour créer un compte sur un site web, on vous demande de choisir un mot de passe fort. Il doit :

- contenir 8 ou 9 caractères,
- chaque caractère doit être soit une lettre de l'alphabet latin (majuscule ou minuscule), soit un chiffre,
- contenir au moins un chiffre.

Quel est le nombre de mots de passe possibles ?

Exercise 2.5.

1. Combien y a-t-il de nombres impairs à 3 chiffres, si l'on exclut 0 et que l'on n'admet pas la répétition ?
2. Combien de mots de 7 lettres (d'un alphabet de 26 lettres) contiennent au moins une lettre répétée ?

Formules d'analyse combinatoire

Exercise 2.6. Utiliser la formule appropriée pour répondre aux questions suivantes :

1. On dispose de 4 cartes distinctes. Combien y a-t-il de façons de les ordonner ?
2. Un numéro de téléphone comporte 8 chiffres. Combien de numéros différents sont possibles ?
3. Un joueur au loto doit sélectionner 6 numéros parmi 47. Combien de sélections possibles existe-t-il ?
4. Cinq individus se présentent à un concours. Le premier reçoit un prix de 500 dollars et le deuxième un prix de 250 dollars. Combien de remises de prix possibles existe-t-il ?

Exercise 2.7.

1. Un manager gère une équipe de 4 employés. Chacun doit être affecté à une mission parmi 4. Combien d'affectations différentes sont possibles ?
2. Combien de groupes de 6 personnes peut-on former à partir de 4 garçons et 6 filles :
    - s'ils doivent contenir obligatoirement deux garçons donnés ?

- s'ils doivent contenir exactement deux garçons ?
- s'ils doivent contenir au moins deux garçons ?

3. De combien de façons peut-on disposer 16 livres sur une étagère (8 romans et 8 bandes dessinées) de sorte que deux romans ne soient jamais côte à côte ?
4. Même question avec 15 livres (8 romans et 7 bandes dessinées).

Exercise 2.8. Quinze athlètes participent à une compétition de course de fond aux Jeux Olympiques.

1. Par ≪ résultat ≫, on entend l'attribution des médailles d'or, d'argent et de bronze. Combien de résultats possibles existe-t-il ?
2. Supposons que 3 des 15 athlètes soient Marocains. Combien de résultats impliquent qu'au moins un Marocain remporte une médaille ?

Exercise 2.9. Le Sénat des États-Unis compte 2 sénateurs pour chacun des 50 États.

1. On sélectionne un comité de 8 sénateurs. Combien de comités possibles existe-t-il ? Combien contiennent au moins un des deux sénateurs de l'État de New York ?
2. On forme un groupe de 50 sénateurs. Combien de groupes possibles contiennent exactement un sénateur par État ?

Exercise 2.10. On dispose de n boules. On veut former k groupes contenant respectivement $n_1, n_2, \ldots, n_k$ boules, en utilisant toutes les boules.

De combien de façons peut-on réaliser ce regroupement ?

Review Exercises (Chapters 1 and 2)

Exercise 2.11. Soit Ω l'ensemble représentant les résultats d'un lancer de dé. On définit :

$$A = \{\text{le résultat est pair}\}, \qquad B = \{\text{le résultat est au moins égal à } 4\}.$$

Lister les éléments des ensembles suivants :

$$A, \quad B, \quad A \cap B, \quad A \cup B, \quad A^c, \quad B^c.$$

Exercise 2.12. On considère trois ensembles A, B et C représentés sur un diagramme de Venn.

Compléter le tableau suivant :

Situation	Ensemble E	Régions du diagramme
Au moins deux des événements A, B, C se sont produits		
Au plus deux des événements A, B, C se sont produits		
Aucun des événements A, B, C ne s'est produit		
Tous les événements A, B, C se sont produits		
Exactement un des événements A, B, C s'est produit		
Les événements A et B se sont produits mais pas C		
Soit B s'est produit, soit ni B ni C ne se sont produits		

Exercise 2.13. Soient les ensembles A et B formant une partition de l'ensemble E. Peut-on conclure que A et B sont disjoints ?

Exercise 2.14. Soient A et B deux sous-ensembles de S.

1. Est-ce que $A \cap B$ et $A \cap B^c$ sont disjoints ?
2. Est-ce que $A \cup B = S$?
3. Est-ce que A et B^c sont disjoints ?
4. Est-ce que $(A \cap B)^c = A^c \cup B^c$?

Exercise 2.15.

1. Combien y a-t-il de nombres impairs à 3 chiffres si l'on n'admet pas la répétition ?
2. Un numéro de téléphone est composé de 9 chiffres dont le premier ne peut être 0. Combien y a-t-il de numéros possibles ?
3. Combien de ces numéros contiennent au moins un chiffre 5 ?
4. Combien de ces numéros contiennent exactement une fois la séquence ≪ 789 ≫ ?

Exercise 2.16.

1. Dix étudiants doivent être répartis dans quatre chambres : une simple, une double, une triple et une quadruple. Combien de répartitions possibles existe-t-il ?

2. Une entreprise compte six représentants commerciaux. Le meilleur reçoit 50 000 dirhams, le deuxième gagne un voyage en Thaïlande, les autres suivent une formation. Combien de résultats possibles existe-t-il ?
3. Une pizzeria propose 7 garnitures de viande et 6 de légumes. Combien de pizzas peut-on composer avec 3 viandes différentes et 2 légumes différents ?
4. Même question si l'on souhaite 5 garnitures différentes, avec au plus 3 garnitures de viande.

Exercise 2.17. Dans le pays de Marinalia, l'Assemblée compte 15 députés : 8 Libéraux et 7 Conservateurs.

1. Combien de façons existe-t-il de former un comité de 5 personnes composé de 3 Libéraux et 2 Conservateurs ?
2. Combien de façons existe-t-il si l'on désigne en plus un président Libéral et un membre principal Conservateur ?
3. Combien de comités de 5 personnes peut-on former si l'on interdit les comités entièrement Libéraux ou entièrement Conservateurs ?

Exercise 2.18. Une grille 7×5 représente le domicile d'un étudiant (en bas à gauche) et son université (en haut à droite).

1. Combien de chemins possibles permettent de se rendre du domicile à l'université ?
2. Combien de chemins sont possibles si l'étudiant doit passer par un café donné ?
3. Combien de chemins sont possibles s'il souhaite éviter ce café ?

Exercise 2.19.

1. Combien d'anagrammes du mot $\ll$ PROGRAMME $\gg$ peut-on former ?
2. Parmi les mots suivants : COURIR, CRIARD, RONRON, SAVANT, RIRENT, lequel admet le moins d'anagrammes ? Lequel en admet le plus ?

Exercise 2.20. Un jeu de cartes contient 52 cartes (4 couleurs de 13 cartes).

1. Combien de mains de 5 cartes peut-on former ?
2. Combien de mains contiennent exactement 2 Cœurs et 3 Piques ?
3. Combien de mains correspondent à :

- une quinte flush royale ?
- une quinte flush ?
- un carré ?
- un brelan ?

4. Quelles sont les probabilités associées à chacune de ces mains ?

Chapter 3

Axiomes de probabilité

3.1 Introduction

Examinez les exemples suivants:

- Le management d'une entreprise industrielle se demande s'il serait possible de quantifier la probabilité qu'une pièce tirée au hasard de la production ait un défaut de fabrication. La direction de la qualité de l'entreprise effectue une enquête sur les pièces produites qui révèle que sur 10 000 pièces examinées, 200 sont défectueuses. Le responsable conclut alors que la probabilité qu'une pièce tirée au hasard soit défectueuse est de $\frac{200}{10000} = 2\%$
- On considère une expérience de lancer de dé. Les issues possibles sont 1, 2, 3, 4, 5 et 6, donc n = 6. Ces 6 numéros ont tous les mêmes chances d'apparaître. Donc on aura $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$
- Un analyste de performance sportive estime que le Maroc a 10% de chances de décrocher la prochaine coupe du monde de football. Pour définir cette probabilité, il se base sur une impression personnelle.

Ces exemples montrent bien qu'il n'y a pas de méthode unique pour déterminer la probabilité d'un événement. On examinera les différentes définitions de probabilité dans cette section.

3.2 Définitions fréquentiste, classique et subjective

3.2.1 Définition fréquentiste de la probabilité

On considère l'expérience aléatoire de lancer d'une pièce de monnaie. Y aurait-il moyen de savoir si la pièce est truquée ou non? Si on lance la pièce une ou deux fois, quel que soit le résultat, on ne pourra rien conclure. Mais si après plusieurs centaines de lancers, on trouve qu'on tombe deux fois plus souvent sur Pile que sur Face, alors il y a de fortes chances que la pièce soit truquée. C'est donc la répétition de l'expérience qui nous aura fourni de l'information. Et on pourrait en outre se baser sur cette information pour déterminer la probabilité des événements. Par exemple, si au bout de 4000 lancers, on trouve 3000 fois Pile, on pourrait dire que la probabilité d'avoir Pile est de $\frac{3000}{4000} = \frac{3}{4}$. Dans ce cas, la probabilité serait calculée sur la base de la fréquence empirique. Cette approche montre l'intuition derrière la définition fréquentiste de la probabilité

Définition fréquentiste de la probabilité

La probabilité d'un événement est la limite de la fréquence relative des succès lorsque $n \to \infty$, où n est le nombre de répétitions de l'expérience, dans des conditions identiques.[a]

Formellement, si on considère que n_A est le nombre de fois où A est réalisé, la fréquence empirique de A est la fréquence de réalisation de A sur n coups; c'est-à-dire $f_n(A) = \frac{n_A}{n}$ et la probabilité de A est

$$P(A) = \lim_{n \to +\infty} f_n(A)$$

[a] Cette définition est basée sur la même logique que la loi des grands nombres, qui sera discutée à la fin du semestre.

Toutefois, cette définition présente certaines limitations:

- Elle a une applicabilité limitée puisqu'elle est uniquement valable pour une expérience aléatoire qu'il est possible de répéter un grand nombre de fois. Ce n'est pas toujours le cas, exemple: probabilité d'avoir de la pluie aujourd'hui (on ne peut pas répéter aujourd'hui plusieurs fois).

- On ne sait pas comment obtenir cette limite de manière précise ni pourquoi elle devrait exister.

3.2.2 Définition classique

Un exemple tiré de l'histoire

Dans un lancer d'un ensemble de dés non truqués, on n'est pas en mesure de connaitre le résultat à l'avance.[1]. Néanmoins, au XVIIème siècle, des joueurs ont tenté d'établir les chances d'apparition de certains résultats en se basant sur une approche fréquentiste. Ils ont remarqué que certains résultats étaient observés plus fréquemment que d'autres. En particulier, la question suivante a été discutée dans certains ouvrages dont un ouvrage de Galilée (1620):[2]

[1] Voir lien `https://rolladie.net/`

[2] Galilée a rédigé vers 1620 sous le titre Sopra le scoperte dei dadi, un mémoire où il calcule la probabilité que la somme des faces de trois dés soit égale à une certaine valeur, cette question lui ayant été posée par le grand-duc de Toscane.

Dans un lancer de 3 dés non truqués, si on calcule le total des points obtenus pour les 3 dés, on remarque que certains totaux sont observés plus fréquemment que d'autres. Par exemple, sur le long-terme, le total de 10 est observé plus fréquemment que celui du 9.

La réponse fournie par Galilée est comme suit:

Le nombre de résultats possibles pour les 3 lancers est de $6^3 = 216$. Le nombre de scores totaux possibles est en revanche de **16** seulement: 3, 4, 5,..., 18. De ce fait, chaque score total correspond à plusieurs combinaisons possibles. Donc **certains scores totaux correspondent à plus de résultats possibles que d'autres**. Par exemple, le total de 9 est obtenu pour 25 résultats possibles et le total de 10 est obtenu pour 27 résultats possibles.[a]

[a] Voir tableau en annexe

Notez que dans cet exemple, l'événement recherché est celui d'obtenir un score total donné et non pas un résultat particulier donné. Pour calculer la probabilité d'un événement, on va compter le nombre de cas où il est réalisé, ce qui va correspondre à plusieurs résultats de Ω, et non pas un seul. Une illustration est fournie avec un exemple de deux dés (sur l'application html).
Cet exemple montre aussi les limites de l'approche fréquentiste, dans la mesure où la probabilité varie en fonction du nombre de répétitions de l'expérience. De ce fait, une autre approche de calcul des probabilités est considérée. Il s'agit de l'approche classique qui est souvent appliquée dans le cas des jeux de hasard.

Définition classique de la probabilité

Au lieu de se baser sur la fréquence empirique, la probabilité d'un événement selon l'approche classique correspond à

$$\frac{\text{Nombre de cas favorables}}{\text{Nombre de tous les cas possibles}}$$

Toutefois, cette approche n'est valable que lorsque tous les résultats possibles de l'expérience sont **équiprobables**.[3]
Formellement,

[3] Pour n résultats possibles, probabilité de $1/n$ chacun.

Définition classique

Lorsque chaque singleton de l'espace fini Ω a la même chance de réalisation, alors la probabilité P sur cet espace est dite uniforme et on a

$$p_\omega = P(\{\omega\}) = \frac{1}{Card(\Omega)}$$

Plus généralement, si P est une probabilité uniforme, alors:

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

Lorsque cette définition est utilisée, on a recours au **calcul combinatoire** (méthodes de dénombrement vues au Chapitre 2). Comme l'approche précédente, cette définition présente quelques limites

- Une définition circulaire: tous les cas sont également probables veut dire qu'ils ont tous la même probabilité, un concept qu'on essaie justement de définir.

- Une applicabilité limitée: avoir une même probabilité pour tous les résultats possibles n'arrive pas très souvent.[4]

3.3 Une définition subjective

Compte tenu des limites des définitions fréquentiste et classique, une autre approche est parfois utilisée, c'est l'approche subjective:

Définition subjective

La probabilité d'un événement traduit le niveau de confiance qu'on a dans la réalisation de l'événement

Cette définition ne fournit aucune méthode de calcul concrète. Pour déterminer les probabilités, on se base sur toutes les informations disponibles, l'expérience, l'intuition, etc. Elle est utilisée lorsqu'il est irréaliste de supposer que les résultats de l'expérience sont équiprobables (on ne peut pas utiliser la définition classique) et lorsqu'on ne dispose pas de suffisamment de données (on ne peut pas utiliser la méthode fréquentiste).

Exemple: Dans le cadre d'élections présidentielles, le directeur de campagne du candidat Y estime que sa probabilité de remporter les élections est de $P(E_1) = 0.6$. Néanmoins, le candidat Y est moins optimiste et pense que sa probabilité de gagner ne dépasse pas $P(E_1) = 0.42$. Dans les deux cas, l'approche suivie est subjective.

[4]En outre, cette définition ne peut pas être utilisée si Ω a une infinité d'éléments (exemple, lancer une pièce jusqu'à obtenir un Face).

3.4 Définition axiomatique de la probabilité

On va définir un ensemble de propriétés mathématiques que la probabilité devrait satisfaire. **Quelle que soit la méthode utilisée (fréquentiste, classique ou subjective), l'essentiel est que les axiomes de probabilités soient respectés**.

3.4.1 Quelques rappels

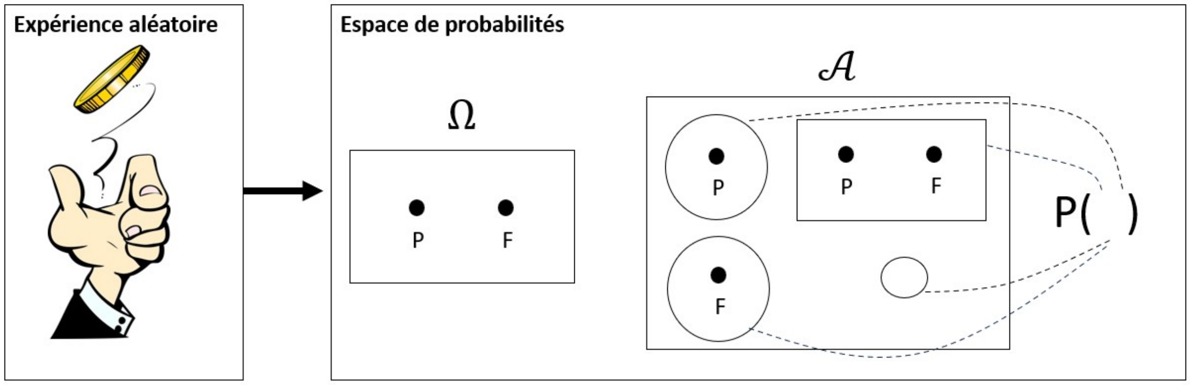
Dans le chapitre introductif, on avait vu les définitions suivantes
L'expérience aléatoire: On appelle **expérience aléatoire** une expérience ε qui, reproduite dans des conditions identiques, peut avoir des résultats différents, résultats qui ne peuvent être prévus par avance.
L'espace-échantillon: L'espace de tous les résultats possibles de l'expérience est appelé **espace-échantillon** (sample space). Il est noté Ω.
L'événement aléatoire: On appelle **événement aléatoire** (associé à l'expérience ε) un sous-ensemble de Ω dont on peut dire au vu de l'expérience s'il est réalisé ou non

3.4.2 Un nouveau concept: σ-algèbre des événements

On veut déterminer des probabilités pour des événements, mais aussi pour leurs compléments, les unions et les intersections d'événements. Donc, on va définir un ensemble qui inclut non seulement les événements individuels, mais aussi différentes combinaisons de ces événements. Cet ensemble est l'ensemble σ-algèbre des événements $\sigma(\Omega) = \mathcal{A}$: σ-algèbre des événements. Il est également appelé "Tribu". Une tribu doit satisfaire les propriétés suivantes:
$\emptyset \in \mathcal{A}$ et $\Omega \in \mathcal{A}$
$A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$
$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
$A_i \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
Les événements sont des sous-ensembles A de Ω tels que $A \in \mathcal{A}$

3.4.3 Un nouveau concept: L'espace de probabilités

L'espace de probabilités est formé de 3 éléments: Ω, σ-algèbre des événements et la fonction de probabilité $(\Omega, \mathcal{A}, P)$. La probabilité est une fonction qui assigne un nombre donné à un événement. Le domaine de définition de cette fonction est $\mathcal{A}$.



3.4.4 Une définition axiomatique de la probabilité

La fonction de probabilités permettra de savoir comment la probabilité de réalisation est distribuée entre les différents événements ou combinaisons d'événements. Elle doit respecter les axiomes suivants (axiomes de Kolmogorov):

Définition axiomatique de la probabilité

On considère une expérience aléatoire avec un espace-échantillon Ω, et un σ-algèbre associé $\mathcal{A}$. La fonction de probabilité est une application $P : \mathcal{A} \to [0, 1]$ qui satisfait les axiomes suivants:

1. Non-négativité $\forall A \in \mathcal{A}, P(A) \geq 0$
2. $P(\Omega)$=1
3. $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
4. $A_i \cap A_j = \emptyset \Rightarrow P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ [a]

[a] Cette propriété est l'additivité dénombrable pour une série infinie d'événements deux-à-deux disjoints (pairwise disjoint)

Pour un événement donné, la probabilité est égale à la somme des probabilités des points d'échantillon qui constituent cet événement. Cependant, dans de nombreux cas, l'identification des éléments de l'échantillon et de leurs probabilités est difficile. Une autre approche est alors nécessaire. Dans ces cas-là, les théorèmes suivants[5] peuvent parfois être utilisés.

3.5 Quelques théorèmes

3.5.1 Théorème 1

Théorème 1

En utilisant les axiomes de probabilités, on peut obtenir les résultats suivants

- $P(\emptyset) = 0$
- $P(A) \leq 1$
- $P(A^c) = 1 - P(A)$

3.5.2 Théorème 2

Théorème 2

- $P(B \cap A^c) = P(B) - P(A \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $A \subset B \Rightarrow P(A) \leq P(B)$

3.6 Evénement presque sûr et événement presque impossible

Evénement presque sûr

En théorie des probabilités, un évènement est dit **presque sûr** si et seulement si il a une probabilité de un $P(A) = 1$. En d'autres mots, l'ensemble des cas où l'évènement ne se réalise pas est de probabilité nulle.

Rappel: l'événement **certain** est l'événement Ω

Evénement presque impossible

Un événement est **presque impossible** si et seulement si il a une probabilité nulle $P(A) = 0$

Rappel: l'événement **impossible** c'est l'ensemble vide $\emptyset$.

[5] Voir démonstration en classe

3.7 Annexe

Ci-après les résultats possibles qui donnent un score de 9 ou de 10 dans une expérience de lancer de 3 dés.

Score de 9		Score de 10	
Résultats possibles	Nombre de permutations	Résultats possibles	Nombre de permutations
6-2-1	6	6-3-1	6
5-3-1	6	6-2-2	3
5-2-2	3	5-4-1	6
4-4-1	3	5-3-2	6
4-3-2	6	4-4-2	3
3-3-3	1	4-3-3	3
	25		27

Exercises

Part A: Axioms and Properties

Exercise 3.1. Soient A, B et C des sous-ensembles de l'ensemble universel U, pas nécessairement disjoints. À l'aide d'un diagramme de Venn, illustrer les événements suivants :

$$A \cup B, \quad A \cap B, \quad A \cap B^c, \quad (A \cup B)^c, \quad A \cap (B \cup C).$$

Exercise 3.2. Des statistiques montrent que dans la ville ensoleillée de Santa Bay :

- 35% des jeunes pratiquent le surf,
- 30% jouent au football,
- 5% pratiquent les deux sports.

1. Quelle est la probabilité qu'un jeune pratique le surf ou le football ?
2. Quelle est la probabilité qu'un jeune ne pratique aucun de ces deux sports ?

Exercise 3.3. On étudie des changements de régimes politiques dans différents pays. On définit les événements :

$$A = \{\text{crise économique}\}, \quad B = \{\text{instabilité politique}\}, \quad C = \{\text{coup militaire}\}.$$

Les probabilités suivantes sont données :

$$P(A) = \dots, \quad P(B) = \dots, \quad P(C) = \dots,$$

$$P(A \cap B) = \dots, \quad P(A \cap C) = \dots, \quad P(B \cap C) = \dots.$$

Calculer :

1. $P\big(A \cap (B \cup C)\big)$,
2. $P\big((A \cup B) \cap C\big)$,
3. $P(A \cup B \cup C)$.

Exercise 3.4. Vrai ou Faux ? Justifier.

1. Soient A et B deux événements avec $P(A) = 0.5$ et $P(B) = 0.9$. Les événements A et B sont disjoints.
2. Soient A, B et C des sous-ensembles disjoints de U. Alors $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

Exercise 3.5. On lance deux dés équilibrés. On définit X comme le résultat du premier dé et Y comme celui du second. Calculer :

$$P(X \text{ et } Y \text{ sont pairs}), \quad P(X + Y \text{ est pair}), \quad P(X < Y).$$

Part B: Counting and Probability Calculations

Exercise 3.6. Parmi cinq hommes et cinq femmes, on forme un comité de quatre personnes. Toutes les combinaisons sont supposées équiprobables.

Calculer les probabilités suivantes :

1. le comité est composé de deux hommes et deux femmes ;
2. le comité compte plus de femmes que d'hommes ;
3. le comité comporte au moins un homme ;
4. M. Ali et Mme Alya sont tous les deux membres du comité.

Exercise 3.7. On considère les cinq prochaines commandes d'un restaurant. Chaque commande peut être soit végétarienne soit non végétarienne, avec la même probabilité. Calculer :

1. la probabilité que les cinq commandes soient du même type ;

2. la probabilité d'avoir exactement trois commandes végétariennes ;

3. la probabilité d'avoir au moins une commande non végétarienne ;

4. la probabilité d'avoir plus de commandes végétariennes que non végétariennes.

Chapter 4

Probabilités conditionnelles et théorème de Bayes

4.1 Conditionnalité en probabilité

Exemple a: Un étudiant doit passer deux examens, un examen en la matière ”Introduction à la microéconomie” et un autre en la matière ”Microéconomie avancée”. On définit les deux événements suivants: A= ”réussir l’examen de microéconomie avancée” et B= ”réussir l’examen d’introduction à la microéconomie”. La connaissance de la réalisation de l’événement B devrait influencer notre estimation de la probabilité $P(A)$. Autrement dit, on va s’intéresser à la probabilité que A soit réalisé, sachant que B a été réalisé.

Exemple b: Soit une expérience aléatoire de lancer d’un dé équilibré. Nous définissons les événements suivants: A= ”obtenir un chiffre pair”; B= ”obtenir un chiffre supérieur à 3”. On va s’intéresser à la probabilité que A soit réalisé sachant que B a été réalisé.

4.1.1 Définition de la probabilité conditionnelle

Probabilité conditionnelle

Soit $A, B \in \mathcal{A}$ et $P(B) > 0$

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

Corollaires

$$P(A \cap B) = P(A|B) \times P(B)$$

$$P(A \cap B) = P(B|A) \times P(A)$$

Les probabilités conditionnelles font l'objet de plusieurs paradoxes dont certains sont discutés dans les exemples qui suivent.

4.1.2 Exemple 1: Le paradoxe des 3 prisonniers

3 prisoniers A, B et C sont condamnés à mort. On décide de gracier l'un d'entre eux mais l'identité de la personne graciée est gardée secrète pour quelques jours. Le gardien n'a pas le droit de la révéler.

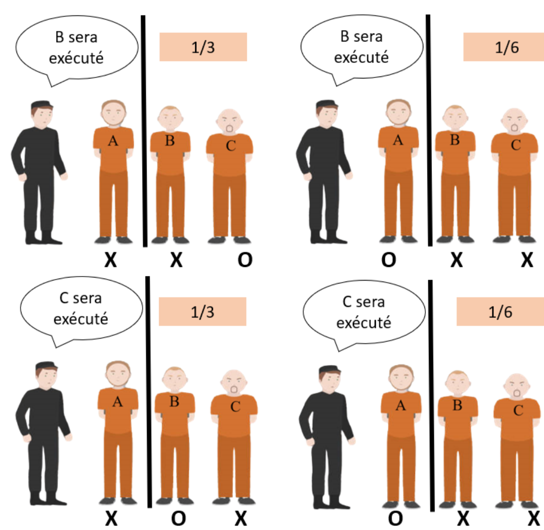
Le prisonier A pose au gardien la question suivante: "Qui de B ou C sera exécuté?"

Le gardien répond: "B sera exécuté"

A pense que désormais, sa chance d'être sauvé est de 1/2. Mais le gardien pense qu'il n'a fourni aucune information utile. Qui des deux a raison?

L'espace-échantillon Ω est comme suit

i	ω_i =(gracié, désigné)	$P(\omega_i)$
1	(A,B)	1/6
2	(A,C)	1/6
3	(B,C)	1/3
4	(C,B)	1/3



On considère l'événement

W ={"B est désigné"}={(A,B), (C,B)}.

Quelle sera alors $P(A|W)$?

$$P(A|W) = \frac{P(A \cap W)}{P(W)} = \frac{P(\{(A,B),(A,C)\} \cap \{(A,B),(C,B)\})}{P(W)}$$

$$= \frac{P(\{(A,B)\})}{P(W)} = \frac{1/6}{1/6 + 1/3} = 1/3$$

Par conséquent, on peut voir que $P(A|W) = P(A) = 1/3$ et donc l'information fournie par le gardien n'est pas utile pour le prisonnier A.

4.1.3 Exemple 2: Le problème de Monty Hall

Le problème de Monty Hall est une énigme mathématique inspirée du jeu télévisé américain "Let's make a deal". Le jeu oppose un présentateur à un candidat. Le candidat est placé face à 3 portes fermées: A, B et C. Derrière l'une d'entre elles, se trouve une voiture et derrière chacune des deux autres se trouve une chèvre. Si le candidat choisit la porte derrière laquelle se trouve la voiture, il pourra l'emporter.

Le candidat choisit une porte au hasard, par exemple A, mais elle est maintenue fermée. Ensuite, le présentateur ouvre l'une des deux portes restantes, qui ne cache pas la voiture. Il offre au candidat l'option de changer d'avis et de choisir l'autre porte.

Le candidat devrait-il modifier son choix?[1]

Stratégie 1: Garder le choix initial (porte A)

$$\Omega = \{\text{"voiture derrière porte A"}, \text{"voiture derrière porte B"}, \text{"voiture derrière porte C"}\}$$

$$P(gagner) = 1/3$$

[1] Voir simulation sur le lien `https://www.rossmanchance.com/applets/2021/montyhall/Monty.html`et exemple de calcul des fréquences sur fichier html

Stratégie 2: Changer de porte

$$\Omega = \{\omega_1 = \text{"voiture derrière porte A, présentateur ouvre B, le candidat choisit C"},$$
$$\omega_2 = \text{"voiture derrière porte A, présentateur ouvre C, le candidat choisit B"},$$
$$\omega_3 = \text{"voiture derrière porte B, présentateur ouvre C, le candidat choisit B"},$$
$$\omega_4 = \text{"voiture derrière porte C, présentateur ouvre B, le candidat choisit C"}\}$$

$$P(gagner) = P(\{\omega_3, \omega_4\}) = 2/3$$

On peut trouver le même résultat en utilisant des probabilités conditionnelles. On définit deux événements: A= "Voiture derrière porte A" et B= "Le présentateur ouvre la porte B". On a

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

On a $P(A \cap B) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$
Et $P(B) = P(B \cap A) + P(B \cap A^c) = \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{2}$
Donc

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

De ce fait

$$P(A^c|B) = 1 - P(A|B) = \frac{2}{3}$$

4.1.4 Exemple 3: Détection de phytovirus dans les plantes

Pour lutter contre certains virus de plantes, une détection précoce est un enjeu majeur qui permet une meilleure efficacité du traitement des plantes contaminées. Une technique récente a été mise en place en vue de détecter les phytopathogènes à travers un test conduit en laboratoire. On définit les événements suivants:
$E =$ Le test est positif.[2]
2 scénarios possibles:

- A_1: La plante est infectée
- A_2: La plante n'est pas infectée

On a $A_1 \cup A_2 = \Omega$, et $A_1 \cap A_2 = \emptyset$: A_1 et A_2 forment une **partition** de l'espace-échantillon. On donne les probabilités:

$$P(E|A_1) = 99.9\%$$

[2] Le test est positif= le test indique que la plante est contaminée

$$P(E|A_2) = 0.01\%$$

On conduit le test pour une plante et il s'avère positif. Est-ce que la plante est contaminée? On va examiner la probabilité $P(A_1|E)$. Par définition, on a

$$P(A_1|E) = \frac{P(A_1 \cap E)}{P(E)}$$

On ne connait pas $P(A_1 \cap E)$. Toutefois, la définition de probabilité conditionnelle nous donne

$$P(A_1 \cap E) = P(E \cap A_1) = P(E|A_1)P(A_1)$$

$P(A_1)$ est la **probabilité à priori** de l'infection. Dans des conditions environnementales saines, on considère que $P(A_1) = 0.01\%$. Pour $P(E)$, on peut l'exprimer comme suit[3]

$$P(E) = P((E \cap A_1) \cup (E \cap A_2)) = P(E \cap A_1) + P(E \cap A_2)$$

Donc

$$P(A_1|E) = \frac{P(E|A_1)P(A_1)}{\sum_{i=1}^{2} P(E|A_i)P(A_i)}$$

En calculant

$$P(A_1|E) = \frac{99.9 \times 0.01}{99.9 \times 0.01 + 0.01 \times 99.99} \approx 50\%$$

Notez bien que $P(E|A_1) = 99.9\%$ alors que $P(A_1|E) \approx 50\%$. On peut illustrer ce résultat par ce qui suit:

Sur 10000 plantes, une seule est infectée (taux de base de 0.01%). Cette plante infectée a 99.9% de chances d'avoir un test positif. Et parmi les 9999 plantes non infectées, une seul aura un test positif (0.01% de taux de faux positifs).

Si une plante a un résultat positif au test, quelle est la probabilité pour qu'elle soit infectée? Sur 10000 plantes, 2 auront un test positif. L'une de ces deux est réellement infectée. Donc, pour un test positif, il y a 50% de chances que la plante soit infectée.

Cet exemple est une application du théorème de Bayes, expliqué dans ce qui suit.

4.2 Théorème de Bayes

En pratique, il arrive souvent qu'on commence l'analyse avec des probabilités initiales (à priori) concernant des événements donnés. Ensuite, on obtient de nouvelles informations qui nous permettent de calculer des probabilités révisées, dites probabilités à postériori. Le théorème de Bayes permet d'effectuer ces calculs.

[3] Cette formule est connue sous le nom de la **la formule des probabilités totales**

4.2.1 Théorème de Bayes

Ce théorème s'applique lorsqu'il y a une **partition de l'espace-échantillon**[4] (comme dans l'exemple précédent) et un événement aléatoire donné.

Théorème

Soit $A_1, A_2, \ldots$ une partition de l'espace-échantillon, et soit B un événement aléatoire. Alors:

$$\forall i, P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

4.2.2 Exemple 4: Tests de produits laitiers

Les produits laitiers commerciaux, tels que les produits laitiers cultivés doivent obligatoirement être testés pour la contamination microbienne. La multinationale "Teraplak" qui opère dans l'industrie laitière a développé sa propre technologie brevetée pour tester ses produits avant leur mise en vente. Les produits déclarés non conformes dans le cadre la procédure de test sont mis au rebut. Récemment, le responsable R&D de l'entreprise a avisé le management de l'entreprise qu'il était urgent de revoir la technologie de test employée. Pour quelle raison? L'explication est fournie dans ce qui suit:

Pour un produit testé, on va examiner la probabilité: $P(NC|TP)$ (NC: non conforme; TP: test positif). On dispose des données suivantes:

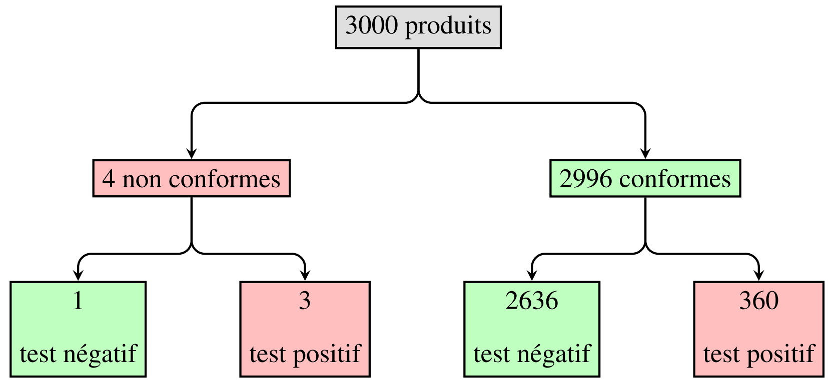
Pour un produit donné, la probabilité de ne pas être conforme est de $P(NC) = 1/700$ (probabilité à priori). Le test ne détecte pas forcèment l'ensemble des cas de produits contaminés, on a $P(TP|NC) = 0.73$. Il y a également un taux de faux positifs de $P(TP|C) = 0.12$ (C: conforme). On a donc

$$P(TP) = P(TP|NC) \times P(NC) + P(TP|C) \times P(C) \approx 12.1$$

$$P(NC|TP) = P(TP|NC) \times P(NC)/P(TP) = 0.73 \times 1/700/0.121 \approx 0.0086 < 1\%$$

Cela veut dire que si le test est positif pour un produit donné, la probabilité qu'il soit effectivement non conforme est en réalité inférieure à 1%! Pour illustrer ces résultats, si l'on prend un groupe de 3000 produits, on obtient la répartition suivante:

[4] Evénements mutuellement exclusifs et leur union correspond à l'espace-échantillon en entier.



Ceci montre que parmi les 363 produits pour lesquels le test est positif, uniquement 3 sont réellement contaminés. De ce fait, l'entreprise supporte un coût important en mettant au rebut une grande quantité de produits conformes.

4.2.3 Exemple 5: Transmission d'un code morse

Dans une transmission de code morse classique "." et "-" apparaissent avec une proportion de 3 contre 4

$$P(.\text{ envoyé}) = 3/7 \qquad P(-\text{ envoyé}) = 4/7$$

Supposons qu'il y ait une interférence et qu'on ait

$$P(-\text{ reçu}|.\text{ envoyé}) = 1/8 \qquad P(.\text{ reçu}|-\text{ envoyé}) = 1/8$$

Si on reçoit un ".", quelle est la probabilité que "." ait été réellement envoyé? On peut utiliser la formule de Bayes:

$$P(.\text{ envoyé}|.\text{ reçu}) = P(.\text{ reçu}|.\text{ envoyé})\frac{P(.\text{ envoyé})}{P(.\text{ reçu})}$$

On a $P(.\text{ envoyé}) = 3/7$, $P(.\text{ reçu}|.\text{ envoyé}) = 7/8$.[5] Et

$$P(.\text{ reçu}) = P(.\text{ reçu} \cap .\text{ envoyé}) + P(.\text{ reçu} \cap -\text{ envoyé})$$

$$= P(.\text{ reçu}|.\text{ envoyé})P(.\text{ envoyé}) + P(.\text{ reçu}|-\text{ envoyé})P(-\text{ envoyé})$$

$$= 7/8 \times 3/7 + 1/8 \times 4/7 = 25/56$$

[5] Remarque: On a utilisé une formule générale qui est de $P(A^c|B) = 1 - P(A|B)$

Remplaçant dans la formule de Bayes

$$P(.\text{ envoyé}|.\text{ reçu}) = \frac{(7/8)\times(3/7)}{25/56} = \frac{21}{25} = 0.84$$

4.3 Indépendance en probabilités

4.3.1 Indépendance

Définition

A et B sont indépendants si et seulement si on a[a]

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

[a] Remarque: dans le cas de l'indépendance, on a $P(A|B) = P(A)$ et $P(B|A) = P(B)$

Théorème

Si A et B sont indépendants, alors

- A et B^c
- A^c et B
- A^c et B^c

sont indépendants

4.3.2 Indépendance et incompatibilité pour 2 événements

Il ne faut pas confondre indépendance et incompatibilité

- A et B sont incompatibles si $A \cap B = \emptyset$
- A et B sont indépendants si $P(A \cap B) = P(A)P(B)$

Exemple

On considère une expérience aléatoire de 2 lancers de pièce, l'espace-échantillon Ω est

$$\Omega = \{(P,P),(P,F),(F,P),(F,F)\}$$

Soient les événements suivants

- A_1: le premier résultat est P

- A_2: le deuxième résultat est F
- A_3: le premier résultat est F

Quelle est la relation entre A_1, A_2 et A_3?

On a

- $A_1 = \{PP, PF\}$
- $A_2 = \{PF, FF\}$
- $A_3 = \{FP, FF\}$
- $A_1 \cap A_2 = \{PF\}$
- $A_1 \cap A_3 = \emptyset$
- $A_2 \cap A_3 = \{FF\}$

- $P(A_1) = 1/2$
- $P(A_2) = 1/2$
- $P(A_3) = 1/2$
- $P(A_1 \cap A_2) = 1/4$
- $P(A_2 \cap A_3) = 1/4$

Par conséquent, on conclut que

- A_1 et A_2 sont indépendants
- A_1 et A_3 sont incompatibles
- A_2 et A_3 sont indépendants

4.3.3 Indépendance de plus de 2 événements

Définition

$A_1, \ldots, A_n$ sont mutuellement indépendants si et seulement si, pour un groupe quelconque $A_{i_1}, \ldots, A_{i_k}$, on a

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

Exemple

On considère l'expérience aléatoire de lancer de deux dés équilibrés. On a ce qui suit

$$\Omega = \{(1,1), (1,2), \ldots, (6,6)\}$$

$$A = \{\text{on obtient des résultats doubles}\} = \{(1,1), (2,2) \ldots, (6,6)\}$$

$$B = \{\text{La somme est entre 7 et 10}\} = \{(1,6), (2,5), \ldots, (5,5)\}$$

$$C = \{\text{La somme est soit 2, 7 ou 8}\} = \{(1,1), (3,4), (4,4), \ldots\}$$

Est-ce que les événements A, B et C sont indépendants?

On a:

$$P(A) = 1/6 \quad P(B) = 1/2 \quad P(C) = 1/3$$

$$P(A \cap B \cap C) = P(\{(4,4)\}) = 1/36 = P(A) \times P(B) \times P(C)$$

$$P(B \cap C) = P(\{\text{la somme est de 7 ou 8}\}) = 11/36 \neq P(B) \times P(C)$$

On a donc l'indépendance des 3 événements mais pas l'indépendance mutuelle.

Exercises

Part A: Conditional Probability and Bayes' Theorem

Exercise 4.1. On lance deux dés équilibrés à 4 faces. Les 16 résultats possibles sont équiprobables.

1. Trouver la probabilité que les deux dés montrent le même nombre (un double).
2. Sachant que la somme obtenue est inférieure ou égale à 4, trouver la probabilité conditionnelle d'obtenir un double.
3. Trouver la probabilité qu'au moins un des deux dés montre un 4.
4. Sachant que les deux dés montrent des nombres différents, trouver la probabilité conditionnelle qu'au moins un des deux dés montre un 4.

Exercise 4.2. Un funambule avance sur une corde. À chaque pas, il avance d'un pas avec probabilité p et recule d'un pas avec probabilité $1 - p$. Les pas sont indépendants.

1. Quelle est la probabilité qu'après deux pas, le funambule soit revenu à son point de départ ?
2. Quelle est la probabilité qu'après trois pas, il se trouve un pas en avant par rapport au point de départ ?
3. Sachant qu'après trois pas il a avancé d'un pas, quelle est la probabilité que son premier pas ait été en avant ?

Exercise 4.3. Dans une transmission de code Morse, les symboles $\ll$. $\gg$ et $\ll$ – $\gg$ apparaissent respectivement avec des proportions de $3/7$ et $4/7$.

Supposons qu'il existe une interférence telle que :

$$P(\text{reçu . } | \text{ envoyé .}) = \ldots, \qquad P(\text{reçu . } | \text{ envoyé } -) = \ldots.$$

Si un symbole ≪ . ≫ est reçu, quelle est la probabilité qu'un ≪ . ≫ ait réellement été envoyé ?

Exercise 4.4. Les produits laitiers doivent être testés pour vérifier l'absence de contamination microbienne. L'entreprise *TPK* utilise un test de détection avant la mise en vente. On dispose des informations suivantes :

- la probabilité qu'un produit soit non conforme est $P(N) = \dots$;
- la probabilité que le test détecte correctement un produit non conforme est $P(T \mid N) = \dots$;
- le taux de faux positifs est $P(T \mid N^c) = \dots$.

1. Déterminer la probabilité $P(N \mid T)$.
2. Interpréter cette probabilité et expliquer pourquoi elle pourrait justifier une révision de la technologie de test.
3. Illustrer votre raisonnement pour un échantillon de 3000 produits.

Part B: Independence and Mutual Independence

Exercise 4.5. On considère une pièce de monnaie particulière. Lorsqu'on la lance deux fois :

- le premier lancer donne Face avec une certaine probabilité ;
- le second lancer donne toujours le même résultat que le premier.

Les seuls résultats possibles sont donc FF et PP, équiprobables.

Les deux événements considérés sont-ils indépendants ? Justifier.

Exercise 4.6. On veut créer un mot de 3 lettres à partir des lettres a, b, c. L'espace-échantillon contient toutes les permutations ainsi que les triples. On définit les événements A, B et C correspondants à certaines positions de lettres.

Peut-on dire que A, B et C sont mutuellement indépendants ?

Exercise 4.7. On considère l'expérience de trois lancers de pièce. Peut-on dire que les événements

$$A = \{\text{le premier lancer donne Face}\}, \quad B = \{\text{le deuxième lancer donne Face}\}, \quad C = \{\text{le troisième lancer donne Face}\}$$

sont mutuellement indépendants ?

Chapter 5

Variables aléatoires discrètes

5.1 Définition et types de variables aléatoires

Dans les chapitres précédents, on avait vu les définitions suivantes

L'expérience aléatoire

L'espace-échantillon Ω

L'événement aléatoire

Sigma-algèbre des événements $\mathcal{A}$ et axiomes de la fonction de probabilité P

L'espace de probabilités $(\Omega, \mathcal{A}, P)$

5.1.1 Définition d'une variable aléatoire

Définition

Soit un espace de probabilités $(\Omega, \mathcal{A}, P)$, une variable aléatoire sur cet espace est une application X de Ω dans $\mathbb{R}$ qui attribue des valeurs numériques à des événements abstraits.

$$X : \Omega \mapsto \mathbb{R}$$

$$\mathcal{A} \mapsto \mathcal{B}$$

Exemples

- Dans une expérience de lancer d'une pièce, on assigne la valeur 1 au résultat {P} et 0 à {F}.

- Dans une expérience de lancer de deux dés, on crée une variable aléatoire qui prend comme valeur la somme des résultats obtenus sur les deux dés.

- Dans une expérience de lancer de fléchette, on attribue un score de 10 si la cible au centre du plateau est atteinte et un score de 1 si une cible extérieure est atteinte.
En d'autres termes, une variable aléatoire peut être considérée comme une expérience dont les résultats sont des nombres (une description numérique des résultats d'une expérience).

5.1.2 Fonctions de distribution

La fonction de distribution de probabilité d'une variable aléatoire décrit comment sont distribuées les probabilités en fonction des valeurs de la variable aléatoire. Différentes notations peuvent être utilisées:

- La fonction de distribution est définie comme $P(X = x)$ pour une valeur spécifique x. Elle peut également être dénotée par $P_X(x)$, $P(X(\omega) = x)$, $P(\{\omega : X(\omega) = x\})$ ou $P(X^{-1}(x))$.
- Pour un ensemble B, on a la notation $P(X(\omega) \in B)$ ou bien $P(X(\omega) \in B)$, $P(\omega : X(\omega) \in B)$ ou $P(X^{-1}(B))$.

Remarque
$(X(\Omega), \mathcal{B}, P_X)$ correspond également à un espace de probabilité.

5.1.3 Types de variables aléatoires

Les variables aléatoires peuvent être discrètes ou continues.

Variable aléatoire discrète

Une variable aléatoire qui peut prendre soit un nombre fini de valeurs, soit un ensemble infini de valeurs dénombrables, telles que $0, 1, 2, \ldots$ est dite variable aléatoire discrète.
Exemples
Une variable X qui représente le nombre de produits commandés par un client à une entreprise
Une variable X qui prend la valeur 0 si échec, et 1 si réussite
Remarque: Une variable discrète ne correspond pas nécessairement à un entier naturel. Il peut également s'agir d'une variable qui prend des valeurs négatives ou d'une proportion. Par exemple: la proportion de produits défectueux dans un lot de 100 produits: 1/100, 2/100,...100/100.

Variable aléatoire continue

Une variable aléatoire qui peut prendre ses valeurs numériques dans un intervalle ou une suite d'intervalles est appelée variable aléatoire continue.
Exemples

Une variable X qui représente le temps écoulé en minutes entre deux appels téléphoniques.

Une variable X qui représente le poids d'une marchandise.

Remarque: On peut également avoir des variables qui sont discrètes sur un intervalle de valeurs et continues sur d'autres intervalles.

5.2 Distribution de probabilité discrète

Pour le cas d'une variable discrète X, la distribution de probabilité est définie par une fonction $f_X(x)$, appelée la **fonction de masse de probabilité**[1] telle que

$$f_X(x) = P(X = x), \forall x \in \chi$$

où χ représente le support de la fonction. Cette fonction donne la probabilité que la variable aléatoire prenne une valeur spécifique pour l'ensemble des valeurs possibles. Une propriété importante de la fonction de masse de probabilité est

$$\sum_x f_X(x) = 1$$

$$f_X(x) \geq 0, \quad \forall x$$

Puisque pour chaque résultat de l'expérience ω, X ne peut prendre qu'une seule valeur x. Les événements $\{X = x\}$ sont disjoints et constituent un système complet (exhaustif) d'événements. De ce fait, la somme des probabilités est égale à 1.

La **fonction de distribution cumulative** (ou bien la **fonction de répartition**) F_X d'une variable est

$$F_X(t) = P(X \leq t), t \in \mathbb{R}$$

[1]Contrairement à la variable continue pour laquelle on a une fonction de densité de probabilité

Théorème

$F(x)$ est une fonction de distribution cumulative si et seulement si

- $\lim_{t \to -\infty} F(t) = 0$, $\lim_{t \to \infty} F(t) = 1$
- F est une fonction non-décroissante
- F est continue à droite: $\lim_{t \to x_0^+} F(t) = F(x_0)$

Remarques

- Une variable aléatoire est continue si et seulement si $F(t)$ est continue (c'est-à-dire également continue à gauche). Elle est discrète si $F(t)$ est une fonction en escalier (step function).

- X et Y sont dites identiquement distribuées si et seulement si $F_X = F_Y$.[2].

- La fonction de masse de probabilité est une fonction de distribution pour les variables aléatoires discrètes uniquement (la fonction de densité est utilisée pour les variables continues). En revanche, la fonction de répartition est définie quelque soit le type de la variable aléatoire.

5.3 Espérance et variance

Les caractéristiques les plus importantes d'une distribution sont sa localisation et sa dispersion, mesurées respectivement par l'espérance et la variance.

5.3.1 Espérance

L'espérance d'une variable aléatoire discrète correspond à sa moyennne. La variable aléatoire X peut prendre différentes valeurs avec différentes probabilités. Pour cette raison, la moyenne de X n'est pas juste une moyenne des valeurs comprises dans la plage de X mais une moyenne pondérée (pas les probabilités) de toutes ces valeurs

$$E(X) = \sum_x x \, P(X = x) = \mu$$

Exemples

- Soit une variable aléatoire qui prend les valeurs 0 ou 1 avec des probabilités P(0)=P(1)=1/2. L'espérance est égale à $E(X) = 1/2 \times 0 + 1/2 \times 1 = 1/2$

- Soit une variable aléatoire qui prend les valeurs 0 ou 1 avec des probabilités P(0)=0.75 et P(1)=0.25. L'espérance est égale à $E(X) = 0.75 \times 0 + 0.25 \times 1 = 0.25$

[2]Cette propriété sera discutée dans le chapitre sur les vecteurs aléatoires

5.3.2 Variance

La variance d'une variable aléatoire est définie comme suit

$$Var\left(X\right) = E[(X - E(X))^2]$$

D'après cette définition, la variance ne peut pas être négative. De plus, elle est seulement nulle lorsque $X = E(X)$ pour toutes les valeurs x, c'est-à-dire lorsque X est une constante (et donc a zéro variabilité). Une autre approche utilisée pour le calcul de la variance est comme suit:

$$Var\left(X\right) = E\left(X^2\right) - E(X)^2 = \sigma^2$$

Exemple

On considère deux individus A et B. L'individu A reçoit soit 48 messages soit 52 messages par jour, avec la même probabilité de 1/2. L'individu B reçoit soit 0 soit 100 messages par jour, avec la même probabilité de 1/2. Si X est le nombre de messages reçus par A et Y le nombre de messages reçus par B par jour, on a $E(X) = E(Y) = 50$. Néanmoins, la variable Y est caractérisée par une variabilité plus forte. En calculant la variance, on trouve que $V(X) = 4$ et $V(Y) = 2500$

5.3.3 Ecart-type

L'écart-type est la racine carrée de la variance

$$\sigma = \sqrt{Var\left(X\right)}$$

5.3.4 Propriétés importantes

L'espérance est une application linéaire

$$E[aX + b] = aE(X) + b$$

La variance est quadratique

$$V[aX] = a^2V(X)$$

Elle est aussi invariante par ajout d'une constante: l'ajout d'une constante n'affecte pas cette mesure

$$V[aX + b] = a^2V(X)$$

5.4 Exemple d'une variable aléatoire discrète

5.4.1 Fonction de masse de probabilité et fonction de répartition

Pour l'expérience de lancer de pièce 3 fois, avec valeur de 1 pour le résultat {F} et 0 pour {P}, on a: Le domaine de

i	1	2	3	4	5	6	7	8
ω_i	FFF	FFP	FPF	PFF	PPF	PFP	FPP	PPP
$X(\omega_i)$	3	2	2	2	1	1	1	0
$P(\omega_i)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

X est $\chi = \{0, 1, 2, 3\}$ et la fonction de masse de probabilité est:

$P(\omega : X(\omega) = 3) = 1/8$

$P(\omega : X(\omega) = 2) = 3/8$

$P(\omega : X(\omega) = 1) = 3/8$

$P(\omega : X(\omega) = 0) = 1/8$

La fonction de répartition pour une variable discrète est une **fonction échelon**

$$F_x(t) = \begin{cases} 0 & t < 0 \\ 1/8 & 0 \leq t < 1 \\ 1/2 & 1 \leq t < 2 \\ 7/8 & 2 \leq t < 3 \\ 1 & 3 \leq t \end{cases}$$

Notez que

- F_X est définie sur $\mathbb{R}$ et pas juste $\chi = \{0, 1, 2, 3\}$ (par exemple, on peut avoir la valeur de $F_X(1.5)$)
- F_X saute à chaque x_i d'une valeur égale à $P(X = x_i)$
- $F_X(t) = 0$ pour $t < 0$ puisque X ne peut pas prendre de valeur négative
- $F_X(t) = 1$ pour $t \geq 3$ puisque $X \leq 3$ presque sûrement (pour rappel, l'événement presque sûr a une probabilité égale à 1).

5.4.2 Espérance et variance

En reprenant l'exemple précédent:

$$E[X] = 1 * 3/8 + 2 * 3/8 + 3 * 1/8 = 1.5$$

i	1	2	3	4	5	6	7	8
ω_i	FFF	FFP	FPF	PFF	PPF	PFP	FPP	PPP
$X(\omega_i)$	3	2	2	2	1	1	1	0
$P(\omega_i)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

x	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$E[X^2] = 1 * 3/8 + 2^2 * 3/8 + 3^2 * 1/8 = 3$$

$$V[X] = E\left(X^2\right) - E(X)^2 = 3 - 1.5^2 = 0.75$$

5.5 Lois usuelles discrètes

5.5.1 Loi uniforme

Une distribution de probabilité suit une loi uniforme lorsque toutes les valeurs prises par la variable aléatoire sont **équiprobables**. Si n est le nombre de valeurs différentes prises par la variable aléatoire X qui suit cette loi, la fonction de masse de probabilité est

$$\forall i, \quad P(X = x_i) = \frac{1}{n}$$

L'espérance

$$E(X) = \frac{n+1}{2}$$

La variance

$$V(X) = \frac{n^2-1}{12}$$

Exemple: La distribution des résultats obtenus suite à un lancer d'un dé équilibré suit une loi uniforme avec $n = 6$ et $x_i = 1, 2, 3, 4, 5, 6$. L'espérance correspond à $E(X) = \frac{1}{6}\sum_{i=1}^{6} i = 3.5$ et $V(X) = \frac{1}{6}\sum_{i=1}^{6} i^2 - E(X)^2 = 2.92$

5.5.2 Loi de Bernoulli

La distribution de Bernoulli décrit une variable aléatoire qui résulte d'une expérience à 2 issues possibles appelées "succès" et "échec". Une variable aléatoire discrète suit une distribution de Bernoulli de paramètre p si sa fonction de masse de probabilité est défini comme suit

$$f_X(x) = \begin{cases} p & \text{si } x = 1, \\ 1-p & \text{si } x = 0, \end{cases}$$

$$f_X(x) = p^x(1-p)^{1-x} \text{pour } x = 0, 1$$

avec $0 < p < 1$. X est une variable aléatoire binaire, qui prend la valeur 1 avec probabilité p et la valeur 0 avec probabilité $q = 1 - p$.

Par exemple, on peut définir:

- Une variable aléatoire X qui prend la valeur 1 lorsque le prix d'une action augmente et 0 lorsque le prix de cette action diminue.

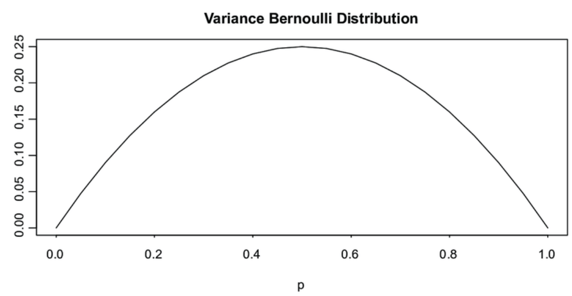
- Une variable X qui prend la valeur 1 si un produit est défectueux et 0 si un produit ne l'est pas.

L'espérance et la variance pour cette distribution peuvent être calculées comme suit:

$$E(X) = 0 \times (1-p) + 1 \times p = p = P(X = 1),$$

$$var(X) = 0^2 \times (1-p) + 1^2 \times p - p^2 = p(1-p),$$

La variance d'une loi de Bernoulli a une valeur maximale pour $p = 1/2$ comme illustré sur la figure ci-après:



5.5.3 Loi binomiale

La loi binomiale est associée à une expérience binomiale.

Expérience binomiale

Une expérience binomiale possède les propriétés suivantes

1. L'expérience est une série de n tirages identiques
2. Deux événements sont possibles à chaque tirage: un succès, un échec
3. La probabilité de succès, notée p, ne se modifie pas d'un tirage à l'autre (même chose pour la probabilité de l'échec notée $1 - p$)
4. Les tirages sont indépendants

Les propriétés 2, 3 et 4 décrivent un processus de Bernoulli. En rajoutant la propriété 1, on a une expérience binomiale. La propriété 3 est appelée hypothèse de stationnarité.
L'intérêt d'une expérience binomiale est de connaître le nombre de succès intervenant au cours de n tirages. Soit X le nombre de succès obtenus en n tirages. X peut prendre les valeurs $0, 1, 2, 3, \ldots, n$. Il s'agit d'une variable aléatoire discrète.

La fonction de probabilité binomiale

Une variable aléatoire suit une distribution binomiale, dénoté par $B(n, p)$, si sa fonction de masse de probabilité est

$$f_X(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} \text{pour } x = 0, 1, \ldots, n$$

Avec $n \geq 1$ et $0 < p < 1$.
Une variable aléatoire binomiale peut prendre $n + 1$ valeurs entières possibles. C'est une distribution sur support fini. Cette loi est appelée loi binomiale parce qu'elle fait intervenir le coefficient binomial C_n^x.
On peut noter qu'une variable qui suit la loi binomiale peut être représentée comme **la somme de variables indépendantes suivant une distribution de Bernoulli** $X = X_1 + \cdots + X_n$. Par exemple, dans un lancer d'une pièce de monnaie n fois de suite et en considérant qu'un Face est un succès. Pour une expérience individuelle, Face a une probabilité de p de se produire. Combien de Face peut-on obtenir à partir de n répétitions? Soit X_i le nombre de Faces obtenu dans le $i^{\text{ème}}$ lancer et X le nombre total de Faces dans les n lancers. Dans ce cas, on a

$$X = \sum_{i=1}^{n} X_i$$

où X_i est une séquence de variables aléatoires iid suivant une loi de Bernoulli(p). On peut alors démontrer que $X = \sum_{i=1}^{n} X_i$ suit une distribution binomiale $B(n, p)$ de la manière suivante. Si on définit l'événement $A_i = \{X_i = 1 \text{ à la ième expérience}\}$, alors X peut être considérée comme le nombre de succès dans n répétitions de l'expérience puisqu'en cas d'échec $X_i = 0$. La probabilité d'avoir un nombre k de succès peut être obtenue pour un exemple distinct de séquence de succès comme suit:

$$P(A_1 \cap A_2 \cap A_3 \cap \ldots A_{n-1}^c \cap A_n^c) = p^k (1-p)^{n-k}$$

Toutefois, il n'est pas nécessaire d'avoir $A_1, \ldots, A_k$ réalisés et $A_{k+1}, \ldots, A_n$ non réalisés. L'essentiel est d'avoir k expériences parmi les n qui donnent un résultat $X_i = 1$. De ce fait

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

Le coefficient binomial C_n^k **donne une indication sur le nombre de manières possibles d'ordonner les k succès parmi les n expériences**. Xi peut prendre les valeurs 0 et 1, mais X peut prendre les valeurs de 0 jusqu'à n. On peut également démontrer que la somme des probabilités données par cette fonction est égale à 1.

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

Parce que la formule du binôme de Newton donne $\sum_{k=0}^{n} C_n^k x^k y^{n-k} = (x+y)^n$. En remplaçant x par p et y par $1-p$, on trouve 1.

Enfin, on peut calculer l'espérance comme suit:

$$E(X) = E(X_1 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = p + p + \cdots + p = np$$

Pour la variance, on obtient (en tenant compte de la propriété d'indépendance)

$$Var(X) = Var(X_1 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n) = np(1-p)$$

Exemple: on considère l'exercice 7 de la série 3 (chapitre 3). L'événement A était défini comme suit: $\ll$ au moins un 6 apparait dans 4 lancers d'un dé à 6 faces $\gg$. Cette expérience peut être représentée comme 4 répétitions d'une expérience de Bernoulli avec $p = 1/6$. Si on définit X comme la variable représentant le nombre de 6 dans 4 lancers, on a $X \sim B(4, 1/6)$. Donc on peut utiliser la formule de la loi binomiale pour calculer $P(A) = P(X > 0) = 1 - P(X = 0) = 1 - C_4^0 (1/6)^0 (1 - 1/6)^{4-0} = 1 - (5/6)^4$. De la même manière, dans cet exercice, on peut calculer la probabilité de l'événement B: $\ll$ au moins un double 6 apparait dans 24 lancers d'un couple de dés à 6 faces $\gg$. On considère la variable Y comme représentant le nombre de double 6s obtenus en 24 lancers. On a $Y \sim B(24, 1/36)$, donc $P(B) = P(Y > 0) = 1 - P(Y = 0) = 1 - C_{24}^0 (1/36)^0 (1 - 1/36)^{24-0} = 1 - (35/36)^4$

Remarque: Il est d'usage d'utiliser une table de la loi binomiale pour obtenir les valeurs de la fonction de répartition pour cette distribution.

5.5.4 Loi de Poisson

La distribution de Poisson décrit un processus qui vérifie les conditions suivantes

- La probabilité de réalisation d'un événement au cours d'une période T est proportionnelle à T
- La réalisation d'un événement au cours d'une période T est indépendante de ce qui s'est passé antérieurement
- La probabilité d'obtenir n réalisations dans une même période diminue très vite à mesure que n augmente

Caractéristiques de la loi de Poisson

La fonction de masse de probabilité de cette loi est comme suit

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Le paramètre λ de cette distribution correspond à la fréquence, nombre moyen de ces événements. L'espérance est comme suit

$$E(X) = \lambda$$

La variance est

$$Var(X) = \lambda$$

La distribution de Poisson est souvent utilisée pour décrire des expériences qui impliquent l'attente d'un événement (attente d'un bus, de l'arrivée de clients, etc). Dans ce cas, il y a un lien entre le temps d'attente et la probabilité d'occurrence de l'événement qui va augmenter plus la période est longue. Cette distribution s'applique donc aussi à des événements qui peuvent rarement se produire de manière simultanée ou dans un laps de temps court (par exemple: offres de travail, appels téléphoniques, pannes de courant électrique, inondations...).

Exemple: On considère le cas d'un opérateur téléphonique qui peut gérer au maximum 5 appels toutes les 3 minutes. Quelle est la probabilité de n'avoir aucun appel la minute suivante? Quelle est la probabilité d'avoir au moins deux appels la minute suivante?

On définit la variable X comme étant le nombre d'appels en une minute. X a une distribution de Poisson avec un paramètre $\lambda = 5/3$. De ce fait, on obtient

$$P(\text{ n'avoir aucun appel la minute suivante}) = P(X = 0) = e^{-5/3}\frac{(5/3)^0}{0!} = 0.189$$

Et

$$P(\text{avoir au moins deux appels la minute suivante}) = P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

$$= 1 - 0.189 - e^{-5/3}\frac{(5/3)^1}{1!} = 0.496$$

Remarque: Il est d'usage d'utiliser une table de la loi de Poisson pour obtenir les valeurs de la fonction de répartition pour cette distribution.

Approximation d'une distribution binomiale par la loi de Poisson

La distribution de Poisson peut être utilisée de manière efficace comme approximation de la loi binomiale pour le cas d'un nombre de répétitions d'expérience n qui est large et une probabilité de succès p qui est petite. Cette approximation est adéquate pour $n \geq 30$ et $p \leq 0.05$. Elle devient de plus en plus juste à mesure que n devient plus large. Dans ce cas, $np = \lambda$ qui est de quelques unités.

Formellement, on peut noter que

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\lambda}} C_n^x p^x (1-p)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

Cette approximation peut être justifiée comme suit: Soit $Y \sim B(n,p)$ et $X \sim Pois(\lambda)$. On peut noter que $P(Y=y)$ peut s'écrire de manière récursive comme suit

$$P(Y=y) = C_n^y p^y (1-p)^{n-y} = \frac{(n-y+1)}{y}\frac{p}{1-p}P(Y=y-1)$$

Avec

$$P(Y=y-1) = C_n^{y-1} p^{y-1} (1-p)^{n-y+1}$$

Donc

$$P(Y=y) = \frac{(n-y+1)}{y}\frac{p}{1-p}P(Y=y-1) = \frac{np-p(y-1)}{y-py}P(Y=y-1)$$

Mais si p est très petite, on pourra écrire $\frac{np-p(y-1)}{y-py} \approx \frac{np}{y}$ en ignorant les termes multipliés par p. En même temps, on peut aussi écrire X sous une forme récursive comme suit:

$$P(X=x) = e^{-\lambda}\frac{\lambda^x}{x!} = \frac{\lambda}{x}P(X=x-1)$$

De ce fait, en posant $\lambda = np$, on obtient une équivalence entre l'expression de probabilité de la loi binomiale et de celle de Poisson.

$$P(Y=y) = \frac{np-p(y-1)}{y-py}P(Y=y-1) \approx \frac{np}{y}P(Y=y-1) = \frac{\lambda}{y}P(Y=y-1)$$

Il reste à confirmer l'équivalence entre $P(Y=y-1)$ et $P(X=x-1)$. Puisqu'on a la relation entre Y et $Y-1$, on peut déduire les probabilités à partir de la valeur 0. Il suffit donc de démontrer que $P(Y=0) \approx P(X=0)$. On a

$$P(Y=0) = (1-p)^n = \left(1-\frac{np}{n}\right)^n = \left(1-\frac{\lambda}{n}\right)^n$$

On a la propriété $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$. De ce fait, si n est suffisamment large

$$P(Y=0) = \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda} = P(X=0)$$

L'approximation par la loi de Poisson est donc valide pour les valeurs larges de n et les valeurs petites de p. En utilisant cette approximation, le calcul devient plus simple comme dans l'exemple qui suit.

Exemple: Un typographe commet une erreur en moyenne une fois pour chaque 500 mots saisis. Une page contient environ 300 mots. Quelle est la probabilité que le nombre d'erreurs soit inférieur ou égal à 2 dans 5 pages?

On considère que saisir un mot est une expérience de Bernoulli avec succès="le mot contient une erreur" et échec= "le mot ne contient aucune erreur". La probabilité de succès $p = 1/500$. Sur 5 pages, on a 1500 mots. On peut donc définir une variable X correspondant au nombre d'erreurs sur 5 pages. Cette variable suit une loi binomiale $X \sim B(1500, 1/500)$. Pour répondre à la question, on calcule la probabilité suivante:

$$P(X \leq 2) = \sum_{x=0}^{2} C_{1500}^{x}\left(\frac{1}{500}\right)^x\left(\frac{499}{500}\right)^{1500-x} = 0.4230$$

Le calcul est plus simple si on procède par une approximation par la loi de Poisson[3] avec $\lambda = 1500 \times (1/500) = 3$. On obtient

$$P(X \leq 2) \approx e^{-3}\left(1+3+\frac{3^2}{2}\right) = 0.4230$$

5.5.5 Loi géométrique

On considère une série d'expériences aléatoires de Bernoulli. La variable aléatoire du nombre d'expériences avant d'obtenir un premier succès suit une distribution géométrique. Une variable aléatoire suivant cette loi peut prendre des valeurs entières de 1 jusqu'à l'infini. Le paramètre de cette distribution est p (probabilité d'un succès).

Exemple: un recruteur passe un entretien à un ensemble de candidats à un poste. Le nombre de candidats interviewés avant que le poste ne soit pourvu suit une distribution géométrique.

Caractéristiques de la loi géométrique

La fonction de masse de probabilité de la loi géométrique est comme suit:

$$P(X=x) = P\{\text{le premier succès se produit à la xième expérience}\} = (1-p)^{x-1}p, \quad x = 1, 2, \ldots$$

[3] On a $n \geq 30$ et $p \leq 0.05$

Même si la valeur de X peut aller jusqu'à l'infini, on peut vérifier que la somme des probabilités est égale à 1 puisque:

$$\sum_{x=1}^{\infty}(1-p)^{x-1}p = p\frac{1}{1-(1-p)} = 1$$

On a utilisé dans ce cas la formule de la somme des n premiers termes d'une suite géométrique (d'où l'appelation de loi "géométrique").[4]

L'espérance de cette distribution est[5]

$$E(X) = \frac{1}{p}$$

La variance correspond à

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

Le paradoxe de Saint-Pétersbourg

En probabilités, le paradoxe de Saint-Pétersbourg concerne une variable aléatoire dont la valeur est, très probablement, petite, mais dont l'espérance est infinie. Ce paradoxe a été énoncé pour la première fois par Daniel Bernoulli. Il décrit un jeu dont le déroulement est comme suit: le joueur mise un montant équivalent au montant qu'il désire gagner (par exemple, un montant de 100). Si le joueur gagne une manche, il emporte le gain et arrête de jouer, sinon il double sa mise et continue de jouer. Le nombre de manches à jouer est une variable aléatoire qui suit une distribution géométrique. Elle peut aller jusqu'à l'infini, mais en moyenne, elle sera finie avec $E(X) = 1/p$.

Toutefois, si on considère la variable Y du montant maximal à miser. Ce montant correspond à $Y = 100 \times 2^{X-1}$ (où le montant 100 correspond à la mise initiale). L'espérance de la variable Y est comme suit

$$E(Y) = \sum_x (100 \times 2^{x-1})P_X(x) = 100p\sum_{x=1}^{\infty}(2(1-p))^{x-1} = \begin{cases} \frac{100p}{2(1-p)} & \text{si } p > 1/2 \\ +\infty & \text{si } p \leq 1/2 \end{cases}$$

De ce fait, même si la variable Y est finie, son espérance peut être infinie si $p \leq 1/2$. C'est le paradoxe de Saint-Pétersbourg.

Exercises

Exercise 5.1. Parmi les représentations graphiques suivantes, lesquelles correspondent à des fonctions de répartition valides ?

[4] Pour rappel, cette formule est $u_0\frac{1-q^{n+1}}{1-q}$

[5] Remarque: la somme des termes de la série $s(q) = \sum_0^{\infty} q^x = (1-q)^{-1}$. La dérivée de premier ordre de cette somme est $s'(q) = \sum_0^{\infty} x(q)^{x-1} = ((1-q)^{-1})' = \frac{1}{(1-q)^2}$. Donc $E(X) = p\frac{1}{(1-q)^2} = 1/p$

1. (a)

2. (b)

3. (c)

4. (d)

Exercise 5.2. Un serveur tente de transférer deux fichiers vers un ordinateur.

- le premier fichier est transféré avec probabilité 0.4 ;

- le second fichier est transféré indépendamment avec probabilité 0.3.

Soit X la variable aléatoire représentant le nombre de fichiers transférés avec succès.

1. Donner la fonction de masse de probabilité de X.

2. Donner la fonction de répartition de X et la représenter graphiquement.

Exercise 5.3. On lance deux dés équilibrés.

$$X = \text{la somme des résultats,} \qquad Y = \text{la différence absolue des résultats.}$$

1. Donner les fonctions de masse de probabilité de X et Y.

2. Calculer l'espérance et la variance de X et de Y.

Exercise 5.4. La fonction de masse de probabilité $p(x)$ est définie pour $x \in \{\dots\}$. Calculer l'espérance et la variance de cette distribution.

Exercise 5.5. Dans un tirage au sort, les gains possibles (en dirhams) et leurs probabilités sont donnés par le tableau suivant :

Montant	0	10	500	1000	5000
Probabilité	0.35	0.50	0.11	0.03	0.01

Calculer l'espérance et l'écart-type du gain.

Exercise 5.6. Deux joueurs A et B jouent à un jeu de lancer de pièces. Le joueur A mise 5 Dhs et le joueur B mise 1 Dh. Les règles du jeu sont décrites dans le tableau suivant :

Pièce de A	Pièce de B	Résultat	Gain de A	Gain de B
Face	Face	FF	6	0
Face	Pile	FP	6	0
Pile	Face	PF	1	5
Pile	Pile	PP	0	6

Soient X le gain (en Dhs) du joueur A et Y le gain (en Dhs) du joueur B.

1. Donner les fonctions de masse de probabilité de X et Y.

2. Calculer $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.

Exercise 5.7. Une entreprise de télécommunications étudie les pannes de réseau. La variable aléatoire X représente le nombre de pannes par jour, avec :

x	$P(X = x)$
0	0.7
1	0.2
2	0.1

1. Calculer l'espérance et la variance de X.

2. Soit Y la perte quotidienne en dollars, sachant que chaque panne coûte 500 dollars. Calculer l'espérance et la variance de Y.

Exercise 5.8.

1. Soit X une variable aléatoire prenant les valeurs $-1, 0, 1, 2$ avec la même probabilité. Soit $Y = X^2$. Calculer $\mathbb{E}[Y]$.

2. Une variable aléatoire X est caractérisée par $\mathbb{E}[X] = \mu$ et $\mathrm{Var}(X) = \sigma^2$. Donner $\mathbb{E}[aX + b]$ et $\mathrm{Var}(aX + b)$.

3. Soit $Z = aX + b$. Quelle est $\mathrm{Var}(Z)$?

Exercises

Exercise 5.9. Un centre de copie compte 12 machines d'impression : 7 machines neuves et 5 machines en mauvais état. Un client confie deux travaux d'impression exécutés simultanément.

1. Quelle est la probabilité que les deux machines sélectionnées soient neuves ?
2. Quelle est la probabilité qu'une tâche soit exécutée par une machine neuve et l'autre par une machine en mauvais état ?

On suppose maintenant que le client se rend au centre pendant 6 jours consécutifs et confie chaque jour deux travaux simultanés. On définit un succès comme l'événement où les deux travaux sont exécutés par des machines neuves. Soit X la variable aléatoire représentant le nombre de succès obtenus.

1. Quelles sont les valeurs possibles de X ?
2. Quelle est la loi de distribution de X ? Donner sa fonction de masse de probabilité.
3. Calculer $\mathbb{E}[X]$.
4. Calculer $P(X = x)$ pour chaque valeur possible de X.

Exercise 5.10. Un agent immobilier rencontre 5 clients. La probabilité de conclure une vente à chaque rencontre est $p = 0.40$.

1. Trouver la probabilité qu'au plus une vente soit conclue.
2. Trouver la probabilité que le nombre de ventes conclues soit compris entre 2 et 4 inclus.
3. Représenter graphiquement la fonction de masse de probabilité.

Exercise 5.11. On lance une pièce de monnaie jusqu'à obtenir *Face*. Soit p la probabilité d'obtenir Face lors d'un lancer. Soit X la variable aléatoire représentant le nombre de lancers nécessaires pour obtenir Face.

1. Donner $P(X = 1)$ et $P(X > 1)$.
2. Donner l'expression de $P(X = x)$ en fonction de p et x.
3. Identifier la loi de X et donner son espérance et sa variance.

Exercise 5.12. On considère deux avions :

- un biréacteur B ;
- un quadriréacteur Q.

Chaque réacteur tombe en panne avec probabilité p, indépendamment des autres.

Soient :

$$X = \text{nombre de réacteurs en panne sur } B, \qquad Y = \text{nombre de réacteurs en panne sur } Q.$$

1. Quelles sont les valeurs possibles de X et Y ?
2. Quelle est la loi de distribution de X et Y ?
3. Donner leurs fonctions de masse de probabilité.
4. Un avion peut voler si au moins la moitié de ses réacteurs fonctionne. Donner les probabilités de vol réussi pour B et Q.
5. Quel avion est le plus sûr selon la valeur de p ?

Exercise 5.13. On considère la transmission de messages électroniques. Chaque message est transmis correctement avec probabilité $p = 0.97$. Soit X le nombre de messages correctement transmis sur 200 messages envoyés.

1. Quelle est la loi de X ?
2. Calculer la probabilité qu'au moins 195 messages soient transmis sans erreur.

Exercise 5.14. Une entreprise possède 100 machines, dont 50 nouvelles. On suppose que la probabilité de panne d'une machine nouvelle est $p = 0.04$ et que les pannes sont indépendantes. Soit X le nombre de machines nouvelles tombant en panne durant l'année.

1. Quelle est la loi de X ? Donner sa fonction de masse.
2. Calculer $P(X = 5)$.
3. Calculer l'espérance et la variance de X.
4. Peut-on approximer X par une loi de Poisson ? Si oui, donner cette approximation.
5. Calculer la probabilité d'avoir au plus 3 pannes.

Exercise 5.15. Les clients arrivent à un centre de photocopie à un rythme moyen de 2 clients toutes les 5 minutes. On suppose que les arrivées sont indépendantes. Soit X le nombre d'arrivées sur une période de 5 minutes. Calculer $P(X > 2)$.

Exercise 5.16.

1. Les clients arrivent à un guichet automatique à un rythme moyen de 2 personnes par minute. Calculer la probabilité qu'au plus 3 personnes arrivent en une minute.

2. Le nombre d'incidents mensuels dans une installation suit une loi de Poisson de paramètre $\lambda = 2.6$. Calculer la probabilité d'avoir moins de 2 incidents en un mois et la probabilité d'en avoir plus de 3.

Chapter 6

Vecteurs aléatoires discrets

6.1 Vecteur aléatoire: définition

6.1.1 Définition

Définition

Un vecteur aléatoire est une application de Ω dans $\mathbb{R}^n$ qui attribue un vecteur de nombres à chaque élément de Ω

6.1.2 Exemple 1

On considère l'exemple de lancer de deux dés avec X la variable qui représente la somme des résultats obtenus et Y la valeur absolue de la différence entre les deux résultats:

X= somme; $Y = |\text{Différence}|$

ω	$P(\omega)$	$X(\omega)$	$Y(\omega)$
(1,1)	1/36	2	0
(1,2)	1/36	3	1
...			
(6,6)	1/36	12	0

6.2 Probabilités jointes

6.2.1 Définition

Définition

Pour deux variables aléatoires discrètes X et Y, on définit la fonction de masse de probabilité jointe comme

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

On a

$$P((X,Y) \in A) = \sum_{(x,y)\in A} f_{X,Y}(x,y)$$

Deux résultats immédiats de la définition

$$f_{X,Y} \geq 0 \quad \forall (x,y) \in \mathbb{R}^2$$

$$\sum_{(x,y)\in\mathbb{R}^2} f_{X,Y}(x,y) = 1$$

6.2.2 Exemple 1

Dans le cas de l'exemple de lancer de deux dés, on obtient

$$f(5,3) = P(X = 5, Y = 3) = P(\{(4,1),(1,4)\}) = 2/36 = 1/18$$

$$f(6,0) = P(X = 6, Y = 0) = P(\{(3,3)\}) = 1/36$$

$$P(X = 7, Y \leq 4) = f(7,1) + f(7,3) = 1/18 + 1/18 = 1/9$$

La fonction de probabilités jointes complète est comme suit

Y X	2	3	4	5	6	7	8	9	10	11	12
0	1/36		1/36		1/36		1/36		1/36		1/36
1		1/18		1/18		1/18		1/18		1/18	
2			1/18		1/18		1/18		1/18		
3				1/18		1/18		1/18			
4					1/18		1/18				
5						1/18					

6.2.3 Exemple 2

On considère la distribution jointe suivante:

$$f(0,0) = f(0,1) = 1/6$$

$$f(1,0) = f(1,1) = 1/3$$

$$f(x,y) = 0 \text{ pour tout autre } (x,y)$$

On peut vérifier que f est une fonction de probabilité jointe valide. Puisqu'on a les probabilités jointes, on peut utiliser f pour les calculs sans se référer à Ω. Par exemple, $P(X = Y) = f(0,0) + f(1,1) = 1/2$

6.2.4 Exemple 3

On considère une expérience de lancer de deux pièces de monnaie ($\Omega = \{PP, FF, PF, FP\}$)

On définit X comme une variable aléatoire qui prend la valeur 1 si au moins un "Pile" est obtenu et 0 sinon, et Y comme une variable aléatoire qui prend la valeur 1 si au moins un "Face" est obtenu et 0 sinon.

Alors le résultat {FF} correspondra à (0,1), {FP} à (1,1), {PF} à (1,1) et {PP} à (1,0).

A quoi correspond P(X=0,Y=0)? $\Rightarrow$ cette probabilité est nulle

On peut résumer les probabilités des résultats dans le tableau qui suit

Y \ X	0	1
0	0	1/4
1	1/4	1/2

Ces probabilités représentent la distribution de probabilité jointe $f_{X,Y}$, une fonction définie de $\mathbb{R}^2$ vers $\mathbb{R}$

Nous pouvons également dériver les distributions marginales de X et Y

Y \ X	0	1	
0	0	1/4	**1/4**
1	1/4	1/2	**3/4**
	1/4	**3/4**	**1**

6.3 Probabilités marginales

6.3.1 Définition

Définition

Pour deux variables aléatoires discrètes X et Y, on définit la fonction de probabilité marginale pour X comme

$$f_X(x) = \sum_{y \in \mathbb{R}} f_{X,Y}(x, y)$$

Démonstration

$$f_X(x) = P(X = x) = P(X = x, Y \in \mathbb{R})$$

$$= \sum_{(x,y) \in A_x} f_{X,Y}(x, y)$$

$$= \sum_{y \in \mathbb{R}} f_{X,Y}(x, y)$$

6.3.2 Exemple 1

L'exemple précédent 1 de deux lancers de dés:

Y X	2	3	4	5	6	7	8	9	10	11	12	
0	1/36		1/36		1/36		1/36		1/36		1/36	3/18
1		1/18		1/18		1/18		1/18		1/18		5/18
2			1/18		1/18		1/18		1/18			4/18
3				1/18		1/18		1/18				3/18
4					1/18		1/18					2/18
5						1/18						1/18
	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	

6.3.3 Exemple 2

Pour l'exemple précédent 2, les probabilités marginales $f_X(x)$ et $f_Y(y)$ peuvent être obtenues comme suit:

$$f_X(x) = \begin{cases} 2/6 & x = 0 \\ 2/3 & x = 1 \end{cases}$$

$$f_Y(y) = \begin{cases} 3/6 & y = 0 \\ 3/6 & y = 1 \end{cases}$$

Ces deux fonctions sont des fonctions de masse de probabilité valides.

Remarque: On peut retrouver les probabilités marginales à partir des probabilités jointes. Mais l'inverse n'est pas vrai: on ne peut pas retrouver les probabilités jointes à partir des probabilités marginales.

6.4 Distribution identique

6.4.1 Définition

Définition

On considère les variables aléatoires X, Y. X et Y sont identiquement distribuées (on note $X \sim Y$) si et seulement si $\forall A \in \mathcal{B}, P(X \in A) = P(Y \in A)$

6.4.2 Exemple 3

On considère l'exemple 3 ci-dessus de lancer de deux pièces de monnaie. Est-ce que $X \sim Y$?

Notez que

$$f_X(x) = \begin{cases} 1/4 & x = 0 \\ 3/4 & x = 1 \end{cases} \qquad f_Y(y) = \begin{cases} 1/4 & y = 0 \\ 3/4 & y = 1 \end{cases}$$

Donc $X \sim Y$. On peut dire que $X \sim Ber(3/4)$ (même chose pour Y)

6.4.3 Exemple 4

On considère une expérience de 3 lancers de pièces. On considère la variable X = nombre de Faces et Y =nombre de Piles.

i	1	2	3	4	5	6	7	8
ω_i	FFF	FFP	FPF	PFF	PPF	PFP	FPP	PPP
$X(\omega_i)$	3	2	2	2	1	1	1	0
$Y(\omega_i)$	0	1	1	1	2	2	2	3
$P(\omega_i)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Si X représente les résultats comprenant les résultats pour un nombre donné de "Face"et Y les résultats avec le même nombre de "Pile", alors $X \sim Y$. On a $X \sim Y$ (mais notez que $\forall\, \omega \in \Omega$, $X \neq Y$)

De manière générale, deux variables aléatoires X et Y qui sont identiquement distribuées ont la même fonction de répartition F_X and F_Y

6.5 Indépendance

6.5.1 Définition

Définition

Les variables aléatoires X et Y sont indépendantes si et seulement si

$$f(x,y) = f_X(x)f_Y(y)$$

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

Si ces variables aléatoires sont indépendantes, la connaissance de X ne nous donne pas d'information sur Y. Autrement dit,

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f_X(x)f_Y(y)}{f_X(x)} = f_Y(y)$$

Avec la fonction de probabilité conditionnelle définie comme suit:

$$f(y|x) = f(x,y)/f_X(x), \quad \forall x : f_X(x) > 0$$

Exemple: lancer une pièce de monnaie une fois et lancer la même pièce une deuxième fois

6.5.2 Exemple 1

Pour calculer l'ensemble des probabilités jointes et l'ensemble des produits de probabilités marginales peut demander du temps. Néanmoins, si n trouve un seul exemple de cas où la définition d'indépendance n'est pas vérifiée, on pourra dire que les deux variables X et Y de l'exemple 1 ne sont pas indépendantes.

Par exemple, on peut noter que $P(X = 2, Y = 1) = 0$, pourtant $P(X = 2) = 1/36$ et $P(Y = 1) = 5/18$. Le produit de ces probabilités marginales ne peut donc pas être nul. On conclut que X et Y ne sont pas indépendantes.

Remarque: le domaine (valeurs possibles) de Y varie en fonction des valeurs de X dans cet exemple. Donc, on peut s'attendre à ce que les deux variables soient dépendantes.

6.5.3 Exemple 2

On reprend l'exemple 2 ci-dessus. Est-ce que X et Y sont indépendantes?

Pour répondre à la question, on doit vérifier si $f_{X,Y} = f_X \times f_Y$.

On a les probabilités jointes

$$f(0,0) = f(0,1) = 1/6$$

$$f(1,0) = f(1,1) = 1/3$$

$$f(x,y) = 0 \text{ pour tout autre } (x,y)$$

Et les probabilités marginales

$$f_X(x) = \begin{cases} 2/6 & x = 0 \\ 2/3 & x = 1 \end{cases}$$

$$f_Y(y) = \begin{cases} 3/6 & y = 0 \\ 3/6 & y = 1 \end{cases}$$

Donc

$$f_{X,Y}(0,0) = 1/6 = f_X(0)f_Y(0)$$

$$f_{X,Y}(0,1) = 1/6 = f_X(0)f_Y(1)$$

$$f_{X,Y}(1,0) = 1/3 = f_X(1)f_Y(0)$$

$$f_{X,Y}(1,1) = 1/3 = f_X(1)f_Y(1)$$

On confirme que X et Y sont indépendantes.

6.5.4 Exemple 3

Dans l'exemple 3 ci-dessus, est-ce que $X \perp\!\!\!\perp Y$?

Est-ce qu'on a $f(x,y) = f_X(x)f_Y(y)$? Par exemple, on a $P(X = 0, Y = 0) = 0$. Néanmoins, $P(X = 0) * P(Y =$

Y \ X	0	1	
0	0	1/4	**1/4**
1	1/4	1/2	**3/4**
	1/4	**3/4**	**1**

$0) = 1/4.1/4 = 1/16$. Donc X et Y ne sont pas indépendantes.

Un contre-exemple : on change la distribution jointe. Par exemple, si on avait plutôt eu la distribution jointe suivante:

Y \ X	0	1	
0	1/16	3/16	**1/4**
1	3/16	9/16	**3/4**
	1/4	**3/4**	**1**

Dans ce cas, X and Y **sont i.i.d. (indépendantes et identiquement distribuées)**

6.6 Vecteurs aléatoires: l'espérance et la covariance

6.6.1 L'espérance de la somme de deux variables aléatoires

Définition

Soient deux variables aléatoires X et Y

$$E(X+Y) = E(X) + E(Y)$$

6.6.2 L'espérance du produit de deux variables aléatoires

On peut calculer l'espérance d'une fonction de deux variables aléatoires de la manière suivante:

Définition

Soient deux variables aléatoires X et Y et une fonction $g(X,Y)$

$$E(g(X,Y)) = \sum_{(x,y)\in\mathbb{R}^2} g(x,y) f_{X,Y}(x,y)$$

Exemple 1: On considère que $g(X,Y) = X \times Y$

Dans l'exemple 1 de lancer des deux dés, on calcule $E[XY]$

$$E[XY] = \sum_{(x,y)\in\mathbb{R}^2} xy f_{X,Y}(x,y)$$

En reprenant les valeurs du tableau, on trouve:

$$E[XY] = 2\times 0\times 1/36 + 4\times 0\times 1/36 + \cdots + 8\times 4\times 1/18 + 7\times 5\times 1/18 = 13\times 11/18$$

6.6.3 Exemple 3

On considère la fonction $g(X,Y) = X.Y$ pour l'exemple 3, l'espérance $E[g(X,Y)]$ sera

$$\begin{aligned} E[g(X,Y)] = E[X.Y] &= g(0,0)*f(0,0) + g(0,1)*f(0,1) \\ &+ g(1,0)*f(1,0) + g(1,1)*f(1,1) = \frac{1}{2} \end{aligned}$$

6.6.4 La covariance

Définition

Le comportement conjoint de deux variables aléatoires peut être décrit par la covariance

$$Cov(X,Y) = \sigma_{X,Y} = E[(X - E(X))(Y - E(Y))]$$

Ou bien

$$Cov(X,Y) = E(X.Y) - E(X)E(Y)$$

La corrélation

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

On peut aussi démontrer que la variance

$$Var(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{X,Y}$$

Enfin, si X, Y sont **indépendantes**, alors

$$Cov(X,Y) = 0$$

6.6.5 Exemple 3

On a $E(X) = 0*1/4+1*3/4 = 3/4$ et $E(Y) = 0*1/4+1*3/4 = 3/4$. Donc $E(X)*E(Y) = 3/4*3/4 = 9/16$.

Donc, la covariance est non-nulle.

En revanche, dans le contre-exemple, on a $E[X.Y] = 9/16$. Donc la covariance est nulle.

6.7 Echantillon aléatoire

Les définitions précédentes peuvent être étendues au cas de plus de 2 variables aléatoires.

Définition

L'ensemble de variables aléatoires $\{X_1, ..., X_n\}$ est un échantillon aléatoire de taille n de la population suivant la distribution $f(\theta), \theta \in \Theta$ si et seulement si:

- $\{X_i\}$ sont mutuellement **indépendants**
- $X_i \sim f(\theta), \forall i$

Notation: $X_1, ..., X_n$ sont **i.i.d.**, avec $X_i \sim f$

On peut calculer une statistique sur la base d'un échantillon aléatoire

Définition

Une statistique Y_n est une fonction de l'échantillon aléatoire $\{X_i, i = 1, \ldots, n\}$:

$$Y_n = T(X_1, \ldots, X_n)$$

Exemple: on peut calculer la moyenne empirique des variables aléatoires $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Cette moyenne est également une variable aléatoire.

6.8 Processus stochastique

Les processus stochastiques sont souvent utilisés pour la modélisation de l'incertitude dans le temps.

Définition

Un processus stochastique est une séquence de variables aléatoires qui évoluent au fil du temps.

La notation utilisée est comme suit:

$$\{X_t\}, \quad t \in \mathbb{R}$$

Exemple: on lance un dé à chaque seconde sur une certaine durée. La séquence des numéros obtenus représente un processus stochastique. Un autre exemple qui est courant est celui des séries temporelles (comme l'évolution du cours d'une action en bourse).

Remarque: certains processus stochastiques ne sont pas définis par rapport au temps. Exemples: processus stochastiques spatiaux $\{X_s\}, \ s \in \mathbb{R}^2$ (variation de la variable est dans l'espace comme la fluctuation de la température sur différentes régions), processus stochastiques définis par rapport à des événements non ordonnés $\{X_n\}, \ n \in \mathbb{N}$ (par exemple, une séquence de jeux où à chaque tour les joueurs peuvent décider de participer ou non, et ceux qui participent obtiendront des résultats aléatoires), etc.

Exercises

Random Vectors and i.i.d. Variables

Exercise 6.1. Vous disposez des informations suivantes sur deux entreprises :

	Cruise Inc	AI Ltd
Prix de l'action	20$	50$
Espérance du rendement	1$	2.5$
Écart-type	0.5$	1$

On suppose que les rendements des deux actions sont indépendants. Vous disposez d'un budget de 10 000$ et hésitez entre les stratégies suivantes :

1. acheter 500 actions de Cruise Inc ;
2. acheter 200 actions de AI Ltd ;
3. acheter 250 actions de Cruise Inc et 100 actions de AI Ltd.

Calculer l'espérance et la variance du gain pour chacun des portefeuilles. Quelle stratégie choisissez-vous ?

Exercise 6.2. Un magasin analyse les habitudes d'achat de ses clients.

$$X = \text{nombre de produits achetés} \in \{1, 2, 3, 4\}, \qquad Y = \text{temps passé (en minutes)} \in \{10, 20, 30, 40\}.$$

La fonction de masse conjointe est donnée par :

$X \backslash Y$	10	20	30	40
1	1/25	2/25	3/25	1/25
2	2/25	1/25	2/25	1/25
3	3/25	1/25	1/25	2/25
4	1/25	2/25	1/25	1/25

1. Calculer les distributions marginales de X et Y.
2. Les variables sont-elles identiquement distribuées ? Indépendantes ?
3. Soit $Z = XY$. Calculer $\mathbb{E}[Z]$.
4. Calculer $\text{Cov}(X, Y)$.

Exercise 6.3. On classe des étudiants selon :

$$X = \text{moyenne annuelle (GPA)} \in \{1, 2, 3, 4\}, \qquad Y = \text{heures de télévision par jour} \in \{0, 1, 2\}.$$

Les probabilités jointes sont données par :

$Y \backslash X$	1	2	3	4
0	0.05	0.034	0.02	0.00
1	0.14	0.12	0.18	0.20
2	0.04	0.04	0.089	0.091

1. Quelle est la probabilité qu'un étudiant ne regarde pas la télévision ?
2. Calculer l'espérance et la variance de X et de Y.
3. Calculer et interpréter le coefficient de corrélation entre X et Y.

Exercise 6.4. Soient X et Y deux variables aléatoires discrètes indépendantes telles que $X, Y \in \{0, 1\}$. On définit :

$$S = X + Y, \qquad T = X - Y.$$

1. Donner les valeurs de S et T pour chacun des cas : $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$.
2. Regrouper les résultats dans un tableau.
3. Donner les lois conjointes et marginales de S et T.
4. Les variables S et T sont-elles indépendantes ?

Exercise 6.5. Les variables aléatoires X et Y prennent les valeurs :

$$X \in \{0, 1, 2\}, \qquad Y \in \{0, 1\}.$$

La fonction de masse conjointe est définie par une constante c.

1. Déterminer la valeur de c.
2. Calculer les distributions marginales de X et de Y.

Conditioning of Random Variables

Exercise 6.6. Soit Y une variable aléatoire uniforme sur $\{0, 1, 2\}$ et soit $B = \{0, 2\}$.

1. Calculer $\mathrm{Var}(Y)$.
2. Calculer $P(Y = y \mid B)$ pour $y \in \{0, 1, 2\}$.
3. Calculer $\mathbb{E}(Y \mid B)$.
4. Calculer $\mathrm{Var}(Y \mid B)$.

Exercise 6.7. On dispose de deux pièces indépendantes :

- pièce A : $P(\text{Face}) = 1/2$;
- pièce B : $P(\text{Face}) = 1/3$.

On choisit une pièce au hasard : $P(A) = 1/4$, $P(B) = 3/4$, puis on la lance jusqu'à obtenir Face pour la première fois.

1. Calculer l'espérance du nombre de lancers nécessaires.
2. On lance ensuite la pièce A jusqu'à Face, puis la pièce B. Calculer l'espérance du nombre total de lancers.

Exercise 6.8. On définit :

$$X = \text{nombre de produits vendus}, \qquad Y = \text{tranche horaire de vente}.$$

La fonction de masse conjointe est donnée par un tableau.

1. Calculer les distributions marginales.
2. Les variables X et Y sont-elles indépendantes ?
3. Calculer $P(X = x \mid Y = y)$ pour toutes les valeurs possibles.
4. Vérifier que ces probabilités conditionnelles somment à 1.

Exercise 6.9. Soit X une variable aléatoire suivant une loi géométrique de paramètre p.

1. Montrer la propriété de mémoire nulle :

$$P(X > m + n \mid X > m) = P(X > n).$$

2. En utilisant l'espérance totale, montrer que $\mathbb{E}[X] = 1 + (1-p)\mathbb{E}[X]$.

3. Retrouver $\mathbb{E}[X]$.

4. Si $p = 1/2$, donner la valeur de $\mathbb{E}[X]$.

Exercise 6.10. Le couple (X, Y) prend avec probabilité égale l'une des valeurs :

$$(0,1),\ (1,0),\ (-1,0),\ (0,-1).$$

On sait que $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.

1. Calculer $\mathrm{Cov}(X, Y)$.

2. Est-il vrai que $\mathbb{E}[XY] = 0$?

3. La condition $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ implique-t-elle l'indépendance ?

Chapter 7

Variables aléatoires continues

7.1 Caractéristiques d'une variable aléatoire continue

7.1.1 Fonction de densité de probabilité

Définition

La **fonction de densité de probabilité** d'une variable aléatoire **continue** X est

$$f_X : P(a \leq X \leq b) = \int_a^b f(x)dx$$

Notez que pour une variable continue, la probabilité ne peut pas être calculée sur des valeurs spécifiques mais sur des intervalles. C'est dû au fait que $P(X = x) = 0, \forall x \in \mathbb{R}$ (parce que $\int_x^x f_X(t)dt = 0$). Par conséquent, on peut considérer que $P\left(X \in]a,b[\right) = P\left(X \in [a,b]\right) = P\left(X \in]a,b]\right) = P\left(X \in [a,b[\right)$.

Une propriété importante de la fonction de densité de probabilité est

$$f_X : \int_{-\infty}^{+\infty} f_X(x)dx = 1$$

$$f_X(x) \geq 0, \forall x$$

$f_X(x) \geq 0$ parce que F_X est croissante, donc sa dérivée est positive.

7.1.2 Fonction de répartition

Définition

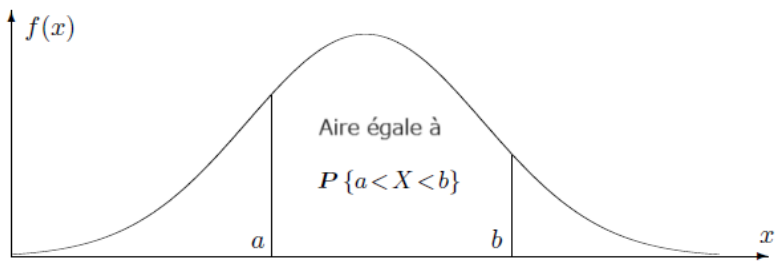
La fonction de répartition est comme suit

$$F_X(x) = P(X \leq x) = P(X < x) = \int_{-\infty}^{x} f_X(t)\,dt$$

Contrairement aux variables discrètes, la loi de répartition est une loi continue. La fonction de densité de probabilité définie plus haut correspond à la dérivée de la fonction de répartition

$$f(x) = F'(x)$$

De ce fait, la densité peut se calculer à partir de la différence de la primitive aux deux bornes de l'intégrale $\int_a^b f(x)dx = F(b) - F(a) = P(a < X < b)$. Cette intégrale correspond à l'aire au-dessous de la courbe de densité entre les deux bornes, comme illustré sur la figure ci-dessous:



7.1.3 Espérance et variance

Définitions

Espérance pour une variable aléatoire continue

$$E(X) = \int x f_X(x)\,dx$$

La définition de la variance ne change pas

$$Var(X) = E[(X - E(X))^2] = E\left(X^2\right) - E(X)^2 = \sigma^2$$

Notez que l'approche est la même que pour une variable continue en remplaçant la fonction de masse de probabilité par la fonction de densité et l'opérateur de somme par une intégrale.

7.1.4 Vecteurs aléatoires continus

Fonctions jointes de distribution cumulative et de densité

Définitions

La fonction de distribution cumulative jointe

$$F_{(X,Y)}(x,y) = P\{X \leq x \cap Y \leq y\}$$

La fonction de densité jointe

$$f_{(X,Y)}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x,y)$$

Indépendance

Définitions

Les variables aléatoires continues X et Y sont indépendantes si et seulement si

$$f(x,y) = f(x)f(y)$$

Où les fonctions marginales sont définies comme:

$$f(x) = \int f(x,y)dy$$

$$f(y) = \int f(x,y)dx$$

Covariance

Définition

La covariance est définie comme suit entre deux variables aléatoires continues X et Y est comme suit

$$Cov(X,Y) = E((X - E(X))(Y - E(Y))) = \int\int (x - E(X))(y - E(Y))f(x,y)dxdy$$

$$= \int\int (xy)f(x,y)dxdy - E(X)E(Y)$$

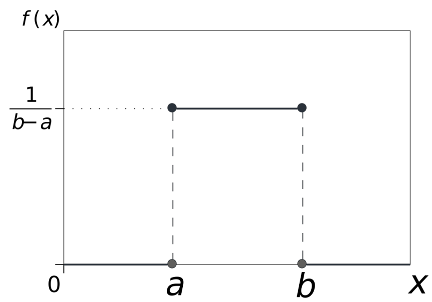
7.2 Lois usuelles continues

7.2.1 La loi uniforme

C'est une distribution de probabilité définie sur un intervalle spécifique où chaque point a une probabilité égale. C'est-à-dire chacune des valeurs possibles de la variable aléatoire a la même chance d'être choisie. On prend l'exemple d'une distribution uniforme sur l'intervalle $[a, b]$

La fonction de densité de probabilité est constante sur l'intervalle considéré

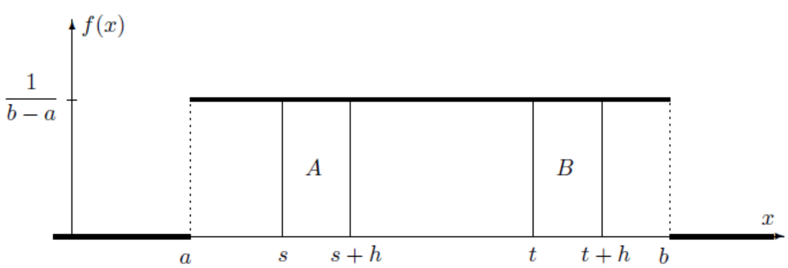
$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & sinon \end{cases}$$



Avec la condition que $|b - a|$ soit un nombre fini. La fonction de répartition est comme suit:

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Propriété uniforme: cette propriété implique que la probabilité ne dépend que de la longueur de l'intervalle et non pas de sa position. Par exemple, sur la figure ci-dessous, les probabilités A et B sont égales.



Exemple: un vol est prévu d'arriver à 17:00. Le temps d'arrivée effectif est distribué de manière uniforme et continue entre 16:50 et 17:10. De ce fait, la probabilité que le vol arrive avant ou après 17:00 est la même. La probabilité que le vol arrive avant 16:55 ou après 17:05 est la même, etc.

L'espérance pour la loi uniforme continue est comme suit:

$$E\left(X\right)=\int_{a}^{b}x\frac{1}{b-a}dx=\frac{1}{2}\left(a+b\right)$$

Et la variance

$$Var\left(X\right)=E\left(X^{2}\right)-\left(E\left(X\right)\right)^{2}=\int_{a}^{b}x^{2}\frac{1}{b-a}dx-\frac{1}{4}(a+b)^{2}=\frac{1}{12}(b-a)^{2}$$

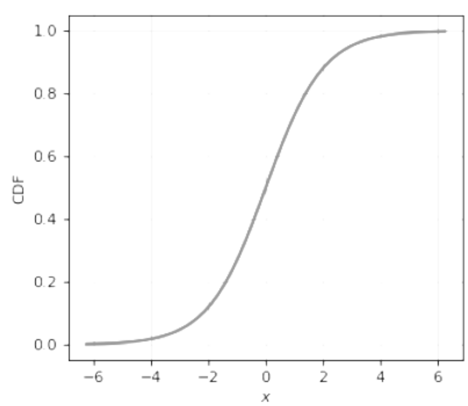
Remarque: la loi uniforme standard est la loi uniforme avec les paramètres $a=0$ et $b=1$

7.2.2 La distribution logistique

La fonction de répartition de la forme standard de la distribution logistique est comme suit:

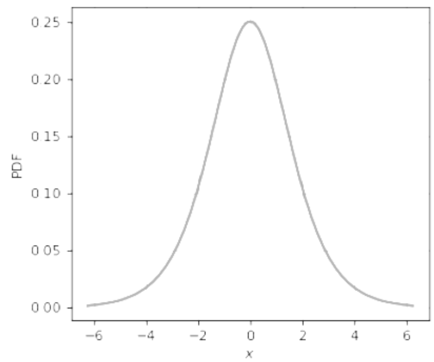
$$F_X\left(x\right)=\frac{1}{1+e^{-x}}$$

Conformément aux propriétés vues dans la section 1, on peut voir que lorsque x tend vers $+\infty$, la fonction tend vers 1 et lorsque x tend vers $-\infty$, la fonction tend vers 0. Cette fonction est aussi continue et croissante en x.



La dérivée de la fonction de répartition est la fonction de densité de probabilité:

$$f_X\left(x\right)=\frac{d}{dx}F_X(x)=\frac{e^{-x}}{\left(1+e^{-x}\right)^{2}}$$



On peut calculer la probabilité à partir des deux fonctions comme suit

$$P(a < X < b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx = \int_{a}^{b} f_X(x)dx$$

Mais si par exemple, on cherche $F_X(a)$ avec $a = 0$, on aura

$$F_X\,(0) = \frac{1}{1 + e^{-0}} = \frac{1}{2}$$

$$P(X \leq 0) = \int_{-\infty}^{0} \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{1}{2}$$

Ce qui correspond à toute la zone à gauche de 0.

Remarque

On peut chercher également l'inverse: quelle est la valeur de x telle que $P(X \leq x) = 1/2$? On a $x : F_X(x) = 1/2 \Rightarrow F_X^{-1}(1/2) = 0$. La fonction inverse F_X^{-1} de la fonction de répartition s'appelle la **fonction quantile**. Dans le présent exemple, on a $F_X\,(x) = \frac{1}{1+e^{-x}}$ et donc $F_X^{-1}(p) = ln(p/(1-p))$. Par exemple, si $P(X \leq x) = 0.975$, alors $F_X^{-1}(0.975) \approx 3.66$

7.2.3 La loi exponentielle

La distribution exponentielle est souvent utilisée pour la modélisation du temps: temps d'attente, temps de défaillance, temps entre des appels téléphoniques, etc. Il existe un lien avec la loi de Poisson dans la mesure où la durée de temps entre les événements rares décrits par la loi de Poisson suit une distribution exponentielle.

La fonction de densité de probabilité

$$f(x) = \lambda e^{-\lambda x}, \quad \text{pour } x > 0$$

Le paramètre λ représente la fréquence dans un intervalle de temps donné (comme dans la loi de Poisson).

La fonction de répartition

$$F(x) = \int_0^x f(t)dt = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad (x > 0),$$

L'espérance

$$E(X) = \int t f(t)dt = \int_0^\infty t\lambda e^{-\lambda t} dt = \frac{1}{\lambda}$$

La variance

$$Var(X) = \int t^2 f(t)dt - E^2(X)$$

$$= \int_0^\infty t^2 \lambda e^{-\lambda t} dt - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Exemple: On envoie des travaux d'impression à une imprimante à raison de 3 travaux par heure:

1. Quelle est l'espérance du temps entre les travaux d'impression?

2. Quelle est la probabilité que le prochain travail soit envoyé dans un intervalle de 5 minutes.

Réponses: On a $\lambda = 3$ pour une heure

1. $E(T) = 1/\lambda = 1/3$ heures, donc 20 minutes en moyenne.

2. On utilise la définition de la fonction de répartition: $P(T < 1/12\,\text{heures}) = F(1/12) = 1 - e^{-\lambda(1/12)} = 1 - e^{-1/4} = 0.2212$

7.2.4 La loi normale

La loi normale (distribution Gaussienne) joue un rôle fondamental en statistique, pour plusieurs raisons:

- De nombreux phénomènes naturels, tels que la taille des individus, les erreurs de mesure, la température, niveau de pollution, plusieurs variables biologiques suivent une distribution normale

- Théorème Central Limite: Ce théorème stipule que la somme ou la moyenne d'un grand nombre d'échantillons indépendants issus de n'importe quelle distribution de probabilité converge vers une distribution normale.

- Les propriétés mathématiques de la loi normale, telles que la symétrie et la forme en cloche facilitent l'analyse.

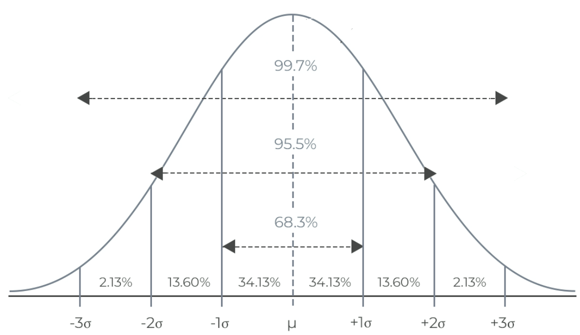
- Inférence statistique: de nombreux tests statistiques classiques reposent sur l'hypothèse de normalité des données. Ces tests sont utilisés dans plusieurs domaines pour faire des études et des prévisions.

Caractéristiques

La fonction de densité de probabilité pour une variable aléatoire X qui suit une loi normale de paramètres μ et σ^2 est comme suit:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Notation $X \sim N(\mu, \sigma^2)$

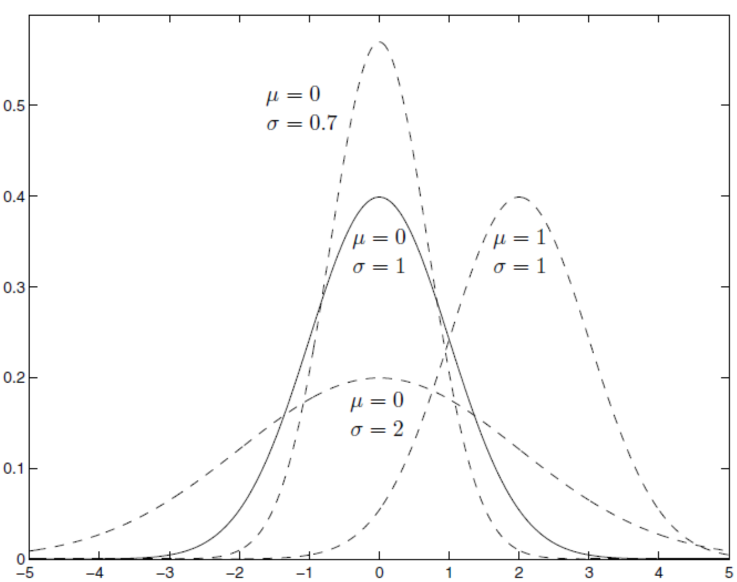


Dans le cas de la distribution normale $P(|X-\mu| \geq 2\sigma) \approx 0.0455$

Les deux paramètres de la loi normale sont son espérance et sa variance. Ils définissent la forme de la courbe

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$



La loi normale standard

La loi normale standard correspond à une loi normale avec paramètres $\mu = 0$ et $\sigma = 1$. On l'appelle aussi la loi normale centrée réduite. La variable utilisée est souvent dénotée par Z, avec

$$Z = \frac{X - \mu}{\sigma}$$

On utilise habituellement une table de la loi normale centrée réduite pour obtenir des probabilités, que ce soit pour Z ou pour X. Par exemple: $P(Z < 1.35) = \Phi(1.35) = 0.9115$, où $\Phi()$ dénote la fonction de répartition de la loi normale centrée réduite. Cette valeur peut être retrouvée au niveau de la table.

En revanche, si on a une variable aléatoire $X \sim N(900, 200^2)$ et qu'on cherche $P(600 < X < 1200)$, on doit d'abord transformer la variable en une variable centrée réduite puis chercher la probabilité au niveau de la table de la loi normale standard.

$$P(600 < X < 1200) = P\left(\frac{600 - 900}{200} < \frac{X - \mu}{\sigma} < \frac{1200 - 900}{200}\right) = P(-1.5 < Z < 1.5)$$

$$= 0.9332 - 0.0668 = 0.8664$$

7.3 L'inégalité de Chebychev

7.3.1 L'inégalité de Chebychev

L'inégalité de Chebychev

Soit X une variable aléatoire et $g(x)$ une fonction non-négative. Alors $\forall r > 0$,

$$P(g(X) \geq r) \leq \frac{E(g(X))}{r}$$

Démonstration du théorème (pour X comme variable continue)

D'après la définition de l'espérance, on a

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

Si on calcule l'intégrale sur un intervalle plus petit: par exemple, l'intervalle où $g(x) \geq r$, on obtient une valeur plus petite

$$\int_{-\infty}^{+\infty} g(x) f_X(x) dx \geq \int_{x: g(x) \geq r} g(x) f_X(x) dx$$

Puisque $g(x) \geq r$, si on multiplie la densité par r au lieu de $g(x)$, on obtient une valeur plus petite

$$\int_{x:g(x)\geq r} g(x) f_X(x) dx \geq r \int_{x:g(x)\geq r} f_X(x) dx = rP(g(X) \geq r)$$

De ce fait, on obtient

$$P(g(X) \geq r) \leq \frac{E(g(X))}{r}$$

7.3.2 Un résultat important de l'inégalité de Chebychev

On considère le cas $g(x) = \frac{(x-\mu)^2}{\sigma^2}$, $\mu = E[x]$, $\sigma^2 = V[X]$, $r = t^2$

En utilisant l'inégalité de Chebychev, on a

$$P\left(\frac{(X-\mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{1}{t^2} E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{1}{t^2}$$

Parce que $E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{E\left[(X-\mu)^2\right]}{\sigma^2} = 1$ puisque le numérateur du second terme est égal à la variance σ^2.

Après simplification, on obtient

Résultat important

$$P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$$

Par exemple, pour $t = 2$, on obtient

$$P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = 0.25$$

Cela implique que

$$P(|X-\mu| \leq 2\sigma) \geq 0.75$$

Ce résultat est valable **quelle que soit la distribution de** X

7.3.3 Exemple de la loi normale

Soit $X \sim \mathcal{N}(\mu, \sigma)$

$$P(|X-\mu| \leq 2\sigma) = P(|X-\mu|/\sigma \geq 2)$$

$$= P(|Z| \geq 2) = 2\frac{1}{\sqrt{2\pi}} \int_2^{\infty} e^{-\frac{x^2}{2}} dx \approx 0.0455$$

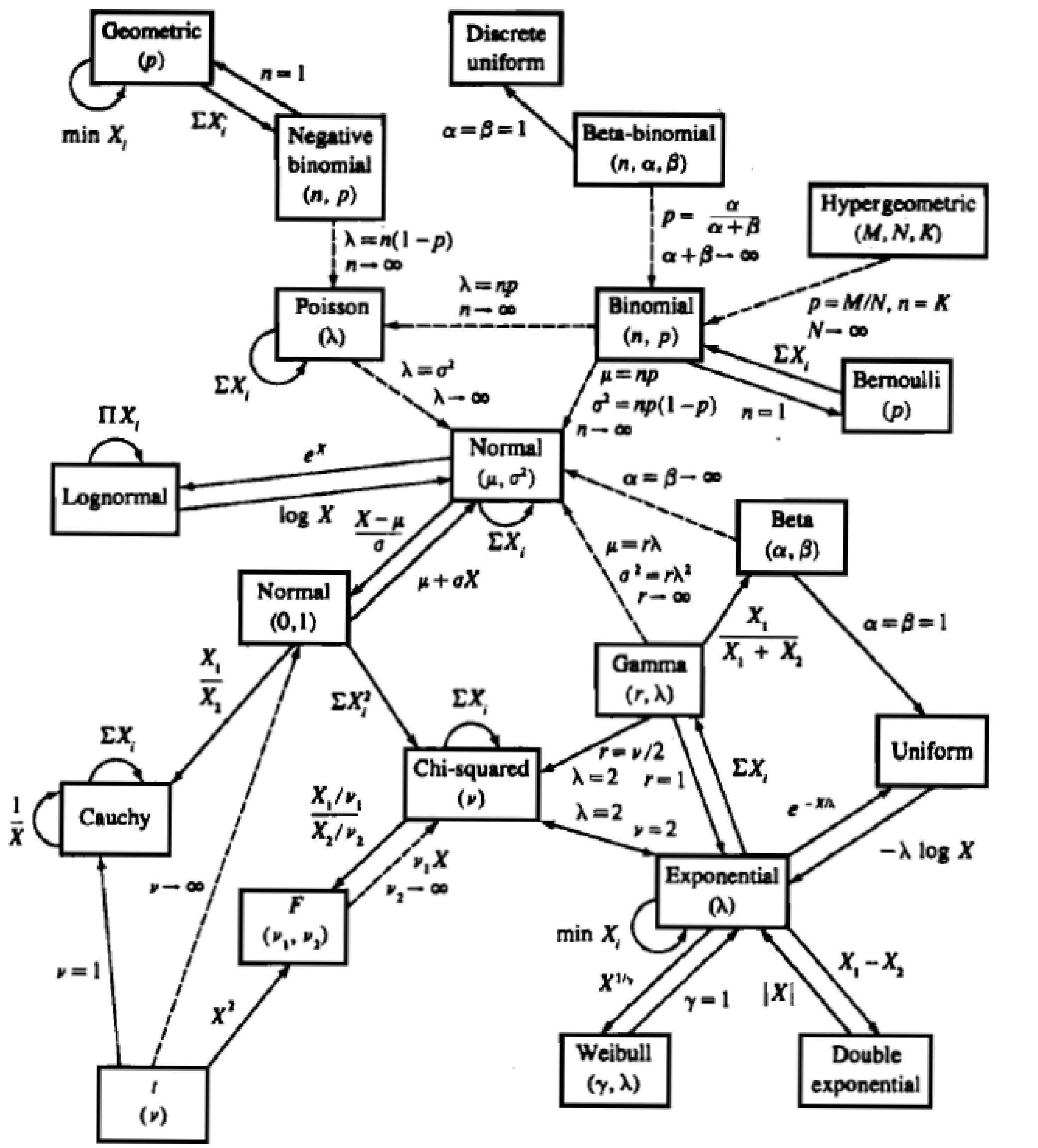
Donc, si $Z \sim \mathcal{N}(0,1)$, alors $P(Z \leq 2) \approx 0.977$.

De même, quel est le quantile z tel que $P(Z \leq z) = 0.975$? C'est $z \approx 1.96$

7.4 Annexe

Le tableau ci-dessous est extrait de l'ouvrage de Casella & Berger et montre les liens qui existent entre les distributions:

TABLE OF COMMON DISTRIBUTIONS 627



Relationships among common distributions. Solid lines represent transformations and special cases, dashed lines represent limits. Adapted from Leemis (1986).

Source: Casella & Berger, Statistical Inference (2002)

Exercises

Exercise 7.1. Soit X une variable aléatoire continue dont la fonction de densité est de la forme

$$f_X(x) = \begin{cases} c\,(\dots) & \text{si } x \in (\dots), \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer la valeur de la constante c.
2. Calculer $P(\dots)$.
3. Calculer $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

Exercise 7.2. On considère la fonction de densité suivante :

$$f_X(x) = \begin{cases} c\,(\dots) & \text{si } x \in (\dots), \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer la valeur de la constante c.
2. Calculer $P(\dots)$.
3. Calculer $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

Exercise 7.3. Soit X une variable aléatoire continue de densité

$$f_X(x) = \begin{cases} cxe^{-x^2} & \text{si } x \in \mathbb{R}, \\ 0 & \text{sinon.} \end{cases}$$

1. Déterminer la valeur de la constante c.
2. Calculer $\mathbb{E}[X]$.
3. Calculer $\mathbb{E}[X^2]$ et en déduire $\mathrm{Var}(X)$.

Exercise 7.4. Le temps d'arrivée d'un train à une gare est supposé uniforme sur l'intervalle $[0, 30]$ minutes après 9h00.

1. Donner la densité $f_X(x)$.

2. Donner la fonction de répartition $F_X(x)$.

3. Calculer la probabilité que le train arrive entre 9h10 et 9h20.

4. Calculer l'espérance et la variance de X.

5. Si un passager arrive à 9h15, quelle est la probabilité qu'il ait manqué le train ?

Exercise 7.5. Soit $X \sim \mathcal{U}[1,3]$ et $Y = X^2$.

1. Calculer $P(Y < 4)$ et en déduire $P(Y > 4)$.

2. Calculer $\mathbb{E}[X]$ et $\mathrm{Var}(X)$.

Exercise 7.6. Soit X une variable aléatoire exponentielle de paramètre λ.

1. Calculer $P(X > 2)$.

2. Donner l'expression de $F_X(x)$ pour $x \geq 0$.

3. Donner l'expression de $F_X(x)$ pour $x < 0$.

4. Calculer $P(1 < X < 3)$.

Exercise 7.7. Soit $Z \sim \mathcal{N}(0,1)$. Calculer les probabilités suivantes à l'aide de la table de la loi normale :

$$P(Z < \dots), \quad P(Z > \dots), \quad P(a < Z < b).$$

Exercise 7.8. Soit $Z \sim \mathcal{N}(0,1)$. Trouver z tel que :

1. $P(Z < z) = 0.8944$,

2. $P(Z > z) = 0.8944$,

3. $P(Z > z) = 0.1685$,

4. $P(Z < z) = 0.0401$.

Exercise 7.9. La durée de vie d'une ampoule suit une loi normale de moyenne $\mu = 1200$ heures et d'écart-type $\sigma = 250$ heures.

1. Calculer la probabilité que la durée de vie soit comprise entre 900 et 1300 heures.

Exercise 7.10. Soit X une variable aléatoire de moyenne μ et de variance σ^2.

1. À l'aide de l'inégalité de Tchebychev, donner une borne supérieure pour $P(|X-\mu| \geq 2\sigma)$ et $P(|X-\mu| \geq 3\sigma)$.

Soit Y une variable aléatoire de moyenne 50 et d'écart-type 5.

1. Montrer que $P(40 \leq Y \leq 60)$ est minorée par une certaine valeur.
2. Comparer ce résultat avec le cas où $Y \sim \mathcal{N}(50, 25)$.

Chapter 8

Convergence

8.1 La convergence

La convergence est un concept fondamental en statistique, qui sous-tend de nombreuses méthodes et théorèmes, tels que la Loi des Grands Nombres et le Théorème Central Limite. Ce concept permet de comprendre comment certaines caractéristiques des échantillons aléatoires convergent vers certaines valeurs à mesure que la taille de l'échantillon augmente. Il existe plusieurs types de convergence.

8.1.1 Convergence en distribution

La distribution d'une séquence d'échantillons se rapproche d'une distribution limite.

8.1.2 Définition

Définition

La forme la plus faible de convergence est la convergence en distribution. Elle est définie comme suit: Soit le processus stochastique X_n

$$X_n \xrightarrow{d} X$$

Si et seulement si[a]

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x) \ \ \forall x$$

Avec F_X la fonction de répartition de X est continue

[a] On dit aussi convergence en loi et on note $X_n \xrightarrow{\mathcal{L}} X$

8.1.3 Exemple

Soient des variables aléatoires X_i qui sont iid et suivent une distribution uniforme sur $[0,1]$, avec $i = 1, \ldots, n$. On définit une variable $X_{(n)} = max_{1 \le i \le n} X_i$. Il s'agit d'un processus stochastique qui dépend de n. On veut vérifier si cette variable converge en distribution. La première étape est de trouver sa fonction de répartition:

$$P(X_{(n)} \le t) = \begin{cases} 0 & t < 0 \\ t^n & 0 \le t < 1 \\ 1 & t \ge 1 \end{cases}$$

La valeur t^n est obtenue comme suit

$$P(max(X_1, X_2, \ldots, X_n) \le t) = P(X_1 \le t, \ldots, X_n \le t)$$

Puisque les X_i sont indépendantes, on peut calculer ces probabilités comme un produit des $P(X_i \le t)$. Partant de la définition de la fonction de répartition de la loi uniforme, on a $P(X_1 \le t) = \cdots = P(X_n \le t) = t$, lorsque $0 \le t < 1$.

Ensuite, on cherche la limite de cette fonction pour $n \to \infty$ et on vérifie si cette limite existe et si elle correspond à une fonction de répartition:

$$\lim_{n\to\infty} P(X_{(n)} \le t) = \begin{cases} 0 & t < 0 \\ 0 & 0 \le t < 1 \\ 1 & t \ge 1 \end{cases}$$

Cette fonction vérifie les conditions définies d'une fonction de répartition (probabilités entre 0 et 1, non-décroissante). Il s'agit d'une fonction de distribution cumulative de la constante 1. Donc:

$$X_{(n)} \xrightarrow{d} 1$$

8.1.4 Convergence en probabilité

La probabilité d'une différence significative entre un échantillon et une valeur théorique diminue à mesure que la taille de l'échantillon augmente.

8.1.5 Définition

Définition

La convergence en probabilité est plus forte que la convergence en distribution. Soit le processus stochastique X_n, on a

$$X_n \xrightarrow{p} X$$

Si et seulement si $\forall \epsilon > 0$

$$\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1$$

Remarque: si $X_n \xrightarrow{p} X$, alors $X_n \xrightarrow{d} X$

8.1.6 Exemple

Soient des variables aléatoires X_i qui sont iid et suivent une distribution uniforme sur $[0, 1]$, avec $i = 1, \ldots, n$. On définit une variable $X_{(n)} = max_{1 \leq i \leq n} X_i$. Puisqu'on a déjà vérifié la convergence en distribution vers 1, on peut juste vérifier si on a la convergence en probabilité vers 1.

En utilisant la définition, on cherche la probabilité suivante, $\forall \epsilon > 0$

$$P(|X_{(n)} - 1| < \epsilon) = 1 - P(|X_{(n)} - 1| \geq \epsilon)$$

On a

$$P(|X_{(n)} - 1| \geq \epsilon) = P(X_{(n)} \geq 1 + \epsilon) + P(X_{(n)} \leq 1 - \epsilon)$$

Mais les X_i sont définies sur $[0, 1]$, donc la première probabilité est 0

$$P(|X_{(n)} - 1| \geq \epsilon) = 0 + P(X_{(n)} \leq 1 - \epsilon)$$

La deuxième probabilité peut être obtenue à partir de la formule de la fonction de répartition d'une distribution uniforme. Et puisque les variables X_i sont iid, la probabilité jointe va être égale au produit des probabilités respectives:

$$P(X_{(n)} \leq 1 - \epsilon) = P(X_i \leq 1 - \epsilon, i = 1, \ldots, n) = (1 - \epsilon)^n$$

Pour $n \to \infty$, cette probabilité va tendre vers zéro car $1 - \epsilon$ est inférieur à 1. De ce fait, on obtient:

$$X_{(n)} \xrightarrow{p} 1$$

8.1.7 Convergence presque sûre

8.1.8 Définition

Définition

Une forme plus forte de convergence est la convergence presque sure qui est définie comme suit: Soit le processus stochastique X_n, on a

$$X_n \xrightarrow{p.s.} X$$

Si et seulement si $\forall \epsilon > 0$

$$P(\lim_{n\to\infty} |X_n - X| < \epsilon) = 1$$

Remarque: si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{p} X$ et donc $X_n \xrightarrow{d} X$

8.1.9 Exemple

Soit l'espace-échantillon $\Omega = [0, 1]$, avec une distribution de probabilité uniforme. On définit

$$X_n(\omega) = \omega + \omega^n$$

Et

$$X(\omega) = \omega$$

On veut vérifier la convergence presque sûre de X_n vers X. On sait que lorsque $n \to \infty$:

$$\forall\omega \in [0, 1[\,,\ \ \omega^n \to 0$$

Donc

$$X_n(\omega) \to X(\omega)$$

Toutefois, pour $\omega = 1$, $X_n(\omega) = 2$, $\forall n$. Donc, $X_n(1)$ ne converge pas vers $X(1)$.

Mais sinon, $P([0, 1[) = 1$, d'où $X_n \to X$ sur un ensemble de probabilité 1. Ce résultat est suffisant pour conclure que:

$$X_n \xrightarrow{p.s.} X$$

8.2 La loi des grands nombres (LGN)

Il s'agit d'un principe en statistique qui énonce que, à mesure que la taille de l'échantillon aléatoire augmente, la moyenne empirique converge vers la moyenne de la population. Formellement:

8.2.1 Définition

Forme faible de la loi des grands nombres

Soit une ensemble de variables aléatoires X_i qui sont iid avec $E(X_i) = \mu$ et $Var(X_i) = \sigma^2 < \infty$. La moyenne de ces variables est $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. On a

$$\bar{X}_n \xrightarrow{p} \mu$$

Forme forte de la loi des grands nombres

Soit une ensemble de variables aléatoires X_i qui sont iid avec $E(X_i) = \mu$ et $Var(X_i) = \sigma^2 < \infty$. La moyenne de ces variables est $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. On a

$$\bar{X}_n \xrightarrow{p.s.} \mu$$

8.3 Le Théorème Central Limite (TCL)

En sélectionnant des échantillons aléatoires simples de taille n d'une population, la distribution d'échantillonnage de la moyenne d'échantillon $\bar{x}$ peut être approchée par une distribution de probabilité normale, lorsque la taille de l'échantillon devient importante. L'implication intuitive est que la somme d'un grand nombre de variables aléatoires indépendantes doit avoir une distribution presque normale, quelle que soit la distribution des variables individuelles. En d'autres termes, la distribution normale sert d'"attracteur" pour les sommes aléatoires. Formellement

8.3.1 Définition

Définition

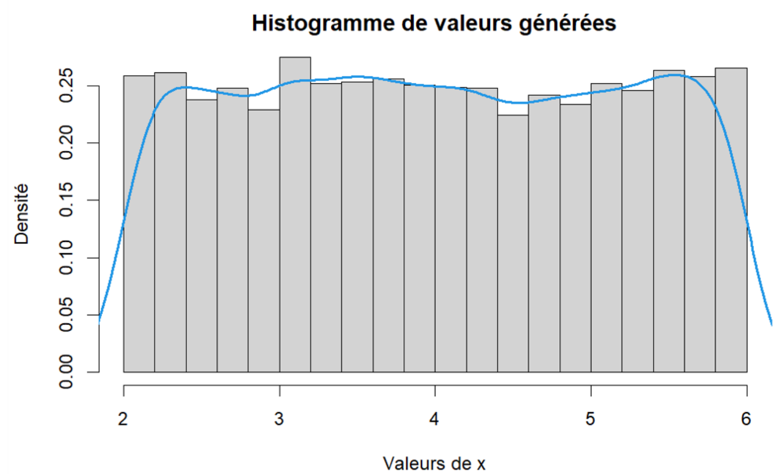
Soit une ensemble de variables aléatoires X_i qui sont iid avec $E(X_i) = \mu$ et $Var(X_i) = \sigma^2 < \infty$. La moyenne de ces variables est $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. On a

$$\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \rightarrow^{d} N(0,1)$$

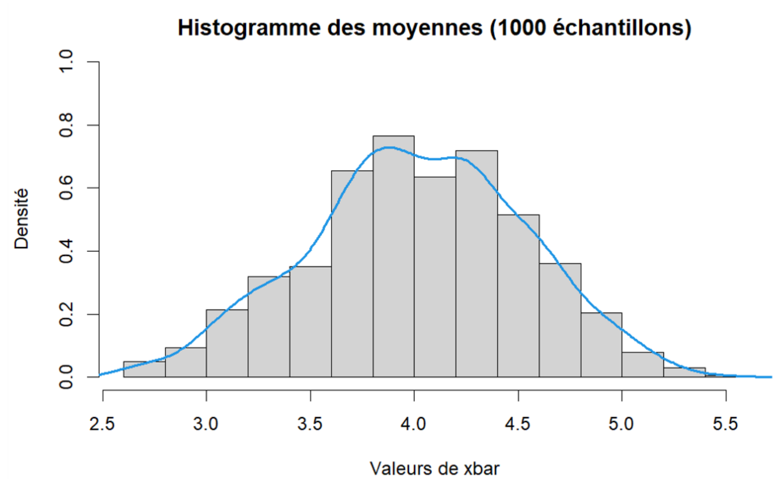
Remarque Une illustration connue du TCL est le Galton Board[1] qui montre qu'en rajoutant des variables aléatoires indépendantes, la somme tend vers la loi normale, voir lien suivant: `https://www.youtube.com/shorts/TwctT3Ncm1w`

Exemple: voilà ci-après une illustration de l'idée derrière le théorème:

1. On génére à l'aide d'un logiciel statistique 10000 valeurs aléatoires d'une distribution uniforme. L'histogramme des valeurs générées est comme suit



2. A partir des 10000 valeurs, on prend des échantillons de 30 valeurs et on calcule la moyenne de chaque échantillon prélevé. Après avoir calculé la moyenne pour 1000 échantillons, on note que l'histogramme des moyennes calculées se rapproche de la forme du courbe en cloche (loi normale), même si la distribution des valeurs à partir desquelles les moyenens sont calculées est uniforme.



[1] En référence à Francis Galton