École Polytechnique

Élèves voie 2

Semestre d'accueil

Introduction au calcul des probabilités

Résumé de cours

Alexandre POPIER, Olivier WINTENBERGER

Année : 2006-2007

La théorie des probabilités concerne la modélisation du hasard et le calcul des probabilités, son évaluation. Ces techniques doivent faire partie des connaissances de base de tous les ingénieurs, très souvent confrontés à des systèmes ou des situations de plus en plus complexes et parfois gouvernées en partie par le hasard. Elles jouent par exemple un rôle essentiel dans l'évaluation de la sûreté de fonctionnement des systèmes d'information et de communication, ou en finance quantitative (pricing d'options, couvertures, etc.).

Le texte qui suit constitue un résumé du cours d'initiation en probabilités du semestre d'accueil pour les élèves de la voie 2 en première année de l'École Polytechnique. Il ne saurait se substituer à un exposé complet et commenté et encore moins à la pratique d'exercices d'application. Il peut néanmoins servir de référence ou d'aide-mémoire en ce qui concerne les notions et outils de base de ces disciplines, exposés selon le cheminement naturel d'un cours.

Pour les étudiants, nous ajoutons qu'ils trouveront des informations sur les cours (contenu, modifications, lieux, horaires, etc.) sur la page Internet suivante :

- http://owintenb.free.fr/accueil.html

Pour avoir des complèments, nous renvoyons le lecteur aux différentes références à la fin du cours et à leurs bibliographies respectives.

Nous remercions Alexis Bienvenüe et Paul Doukhan pour nous avoir permis de nous inspirer largement de leurs cours de probabilité et statistique. Ces notes de cours sont évidemment une version préliminaire et nous serions reconnaissant à tout lecteur de nous faire part des fautes qu'il y aura détectées à l'adresse suivante :

- owintenb@univ-paris1.fr.

Table des matières

Chapitre 1

Première approche dans le cas fini, dénombrement

1.1 Motivation et introduction

La théorie des probabilités a pour objectif de modéliser des expériences où plusieurs résultats sont possibles, mais où leur réalisation n'est pas déterminée à l'avance : lancer de dés, prix d'une action, perturbations sur une ligne téléphonique, files d'attente, etc. Ceci pour évaluer les risques et mettre sur pied des stratégies pour faire face aux aléas. Ainsi un joueur au casino devra prendre en compte les chances de gain pour optimiser ses mises.

Dans ce chapitre, on va s'intéresser uniquement aux expériences pour lesquelles le nombre de résultats est fini. La première difficulté est donc de décrire correctement l'expérience, afin d'énumérer tous les résultats possibles. On regroupe ces issues possibles dans un ensemble, noté traditionnellement Ω et appelé *espace des observables* ou *espace des issues*. On appelle *événement* toute partie de Ω.

Le deuxième problème est alors d'attribuer à chaque événement $A \subseteq \Omega$ un nombre compris entre 0 et 1 qui estime les chances qu'a cet événement de se réaliser. On appelle ce nombre la *probabilité* de A.

Une fois ces deux premières étapes franchies, on peut effectuer tout un tas de calculs sur cet ensemble Ω, comme par exemple déterminer si un jeu de hasard est équitable, c'est-à-dire si en moyenne tous les joueurs ont autant de chances de gagner (au casino ou à la Française des Jeux, il n'y a aucun jeu équitable...).

1.1.1 Espace des observables

Dans chaque situation, l'espace des observables (ou des issues) est donc noté Ω, alors que les éléments de Ω, appelés observations, seront notés ω. Voyons sur des exemples comment on modélise une expérience aléatoire.

Exemple :
On considère un dé à 6 faces, numérotées de 1 à 6. On appelle Ω l'ensemble des résultats possibles d'un lancer, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si on suppose que le dé est équilibré, on associe

à chaque résultat la probabilité $1/6$. On pense donc à une probabilité comme à une fonction $\mathbb{P}$ sur Ω telle que :

$$\mathbb{P}(1) + \mathbb{P}(2) + \mathbb{P}(3) + \mathbb{P}(4) + \mathbb{P}(5) + \mathbb{P}(6) = 1. \tag{1.1}$$

Exemple :
On considère le même dé, sauf que sur la sixième face le nombre 6 est remplacé par le 5. Il y a donc deux faces où 5 est inscrit. L'ensemble des résultats est différent : $\widetilde{\Omega} = \{1, 2, 3, 4, 5\}$. Comme le dé est équilibré, les faces 1,2,3 et 4 ont chacune la probabilité $1/6$ de sortir, alors que $\mathbb{P}(5) = 2/6$.

Exemple :
Dans le premier exemple, quelle est la probabilité qu'un nombre pair sorte ? Il suffit de décomposer cet événement, disons A, en les différents résultats qu'il contient : $A = \{2, 4, 6\}$.

$$\mathbb{P}(A) = \mathbb{P}(2) + \mathbb{P}(4) + \mathbb{P}(6) = 1/2.$$

Il existe des cas où le choix de Ω n'est pas aussi immédiat que dans l'exemple du dé. Nous allons développer une expérience simple où choisir la modélisation permet d'évaluer rigoureusement le risque.

On a 3 cartes coloriées sur chaque face : rouge-rouge (RR), blanc-blanc (BB) et rouge-blanc (RB). Une personne tire une carte au hasard et vous en montre une face, elle aussi tirée au hasard. Si la face montrée est R, que pariez-vous pour la couleur de la face cachée ?

Intuitivement, il y a trois cas possibles qui se produisent avec la même probabilité : deux cas où la face montrée est R lorsque la carte est RR, un cas où la face montrée est R lorsque la carte est RB. Donc deux cas où la face cachée sera R, et un cas où elle sera B. On a donc deux fois plus de chance de gagner en pariant R.

Plus rigoureusement, l'expérience à modéliser consiste en deux étapes : on choisit une carte parmi $\{RR, BB, RB\}$, puis on choisit une des deux faces. On va supposer que dans la notation XY des cartes, X correspond au recto et Y au verso. On traduit que ces choix se font au hasard en disant que tous les résultats possibles ont même probabilité. On va décrire séparément les deux étapes. Le choix d'une carte est un élément de

$$\Omega_1 = \{RR, BB, RB\},$$

et on appelle $\mathbb{P}_1$ la probabilité sur Ω_1, avec

$$\mathbb{P}_1(RR) = \mathbb{P}_1(RB) = \mathbb{P}_1(BB) = 1/3.$$

Puis le choix d'une face correspond à un élément de $\Omega_2 = \{r, v\}$ (r pour recto, v pour verso) et $\mathbb{P}_2(r) = \mathbb{P}_2(v) = 1/2$. Pour l'expérience, l'ensemble des tirages est l'ensemble des couples (carte,face) :

$$\Omega = \{(RR, r), (RR, v), (RB, r), (RB, v), (BB, r), (BB, v)\} = \Omega_1 \times \Omega_2.$$

La probabilité associée aux résultats de Ω est $\mathbb{P}$:

$$\mathbb{P}(carte, face) = \mathbb{P}_1(carte)\mathbb{P}_2(face) = 1/6.$$

Le fait que $\mathbb{P}$ se factorise traduit *l'indépendance* des deux étapes (voir chapitre 2.2).

On décompose l'événement A « la face montrée est R et la face cachée est R » en éléments de Ω :

$$A = \{(RR, r), (RR, v)\}.$$

De même D est « la face montrée est R et la face cachée est B », soit $D = \{(RB, r)\}$. On voit donc que $\mathbb{P}(A) = 2/6$ alors que $\mathbb{P}(D) = 1/6$. On en déduit qu'il faut mieux parier que la face cachée est rouge.

1.1.2 Probabilité

On va définir la notion de *probabilité* sur un ensemble fini Ω. Avant, des rappels de théorie des ensembles : si $A \subseteq \Omega$ et $B \subseteq \Omega$, on note $A \cup B = \{\omega \in A \text{ ou } \omega \in B\}$, et $A \cap B = \{\omega \in A \text{ et } \omega \in B\}$. On interprète l'événement $A \cup B$ comme étant « A ou B se réalisent », et $A \cap B$ comme étant « A et B se réalisent ». On note $A \setminus B = \{\omega \in A \text{ et } \omega \notin B\}$ et $A^c (= \overline{A}) = \Omega \setminus A$. L'événement A^c est appelé *événement complémentaire* de A. On dit que deux événements A et B sont *disjoints* s'ils n'ont aucun élément en commum, c'est-à-dire si $A \cap B = \emptyset$, où $\emptyset$ est l'ensemble vide. Remarquons que A et A^c sont toujours disjoints, ainsi que $\{\omega\}$ et $\{\omega'\}$ si $\omega \neq \omega'$.

Les événements considérés sont des sous-ensembles de Ω : $A \subset \Omega$. Ce sont aussi des éléments de $\mathcal{P}(\Omega)$, l'ensemble de toutes les parties de Ω. Par exemple, si $\Omega = \{1, 2, 3\}$, alors

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

Plus généralement :

Proposition 1.1 *Si un ensemble Ω a un nombre fini N d'éléments, alors $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω a 2^N éléments.*

Preuve :

Soit $\Omega = \{\omega_1, \ldots, \omega_n\}$. On introduit pour une partie A de Ω la fonction caractéristique de A :

$$\mathbb{1}_A(\omega) = \begin{cases} 1 \text{ si } \omega \in A, \\ 0 \text{ sinon.} \end{cases}$$

Il y a une bijection entre $\mathcal{P}(\Omega)$, l'ensemble des partitions de Ω et $\{0; 1\}^\Omega$, l'ensemble des applications de Ω dans $\{0; 1\}$:

$$\begin{array}{rcl} \mathcal{P}(\Omega) & \rightarrow & \{0; 1\}^\Omega \\ A & \mapsto & \mathbb{1}_A. \end{array}$$

L'application réciproque est définie par $\phi \mapsto \{\omega \in \Omega; \phi(y) = 1\}$. Donc $\text{Card}(\mathcal{P}(\Omega)) = \text{Card}(\{0;1\}^\Omega) = 2^N$. $\square$

Dans l'introduction, on voulait associer à un événement (c'est-à-dire une partie de Ω) un nombre compris entre $[0,1]$. Donc une probabilité va être une fonction sur $\mathcal{P}(\Omega)$, qui doit vérifier certaines conditions du style de (1.1). Ceci se traduit ainsi :

Définition 1.2 (Probabilité (cas fini)) *Une* probabilité $\mathbb{P}$ *est une fonction de* $\mathcal{P}(\Omega)$ *dans* $[0,1]$ *telle que :*

1. $\mathbb{P}(\Omega) = 1$;
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *si* A *et* B *sont disjoints.*

Remarque 1.3 **Attention :** *cette définition n'est valable que si* Ω *est fini (voir chapitre 2 pour une définition rigoureuse pour* Ω *quelconque).*

Quelques résultats immédiats :

Propriétés 1.4 *De la définition 1.2, on obtient*

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;
2. $\mathbb{P}(\emptyset) = 0$;
3. $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B})$;
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;
5. $\mathbb{P}(\{\omega\} \cup \{\omega'\}) = \mathbb{P}(\{\omega\}) + \mathbb{P}(\{\omega'\})$ *si* $\omega \neq \omega'$.

Toutes ces propriétés de $\mathbb{P}$ se comprennent très bien si on garde à l'esprit qu'elles servent à modéliser une expérience aléatoire.

1.1.3 Le cas équiprobable

Considérons le cas particulier où tous les résultats possibles ont la même probabilité de se réaliser :

$$\forall (\omega, \omega') \in \Omega^2, \ \mathbb{P}(\{\omega\}) = \mathbb{P}(\{\omega'\}).$$

Cette hypothèse convient par exemple pour modéliser un dé équilibré et plus généralement les jeux de hasard (cartes, roulette, etc.). Alors, comme $\Omega = \cup_{\omega \in \Omega} \{\omega\}$,

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1,$$

d'où, d'après l'hypothèse d'équiprobabilité, pour tout $\omega_0 \in \Omega$,

$$1 = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega_0\}) = \text{Card}(\Omega)\, \mathbb{P}(\{\omega_0\}),$$

c'est-à-dire : $\mathbb{P}(\{\omega_0\}) = 1/\text{Card}(\Omega)$. On se reportera à la section suivante pour la définition de Card. Maintenant, si $A \in \mathcal{P}(\Omega)$,

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}. \tag{1.2}$$

La probabilité ainsi définie est appelée probabilité uniforme sur Ω. On voit donc que le calcul des probabilités dans ce cas très particulier revient à un calcul de cardinal d'ensemble. C'est l'objet de la section suivante.

1.2 Notions de dénombrement

1.2.1 Premières définitions

Définition 1.5 *On rappelle que le* cardinal *d'un ensemble fini Ω, noté $\mathrm{Card}(\Omega)$, représente son nombre d'éléments.*

Par exemple avec $\Omega = \{0, \ldots, 10\}$, on a :

$$\mathrm{Card}(\Omega) = 11.$$

Plus généralement, si

- $\mathcal{Y}$: ensemble à p éléments $\{y_1, \ldots, y_p\}$,
- $\mathcal{X}$: ensemble à n éléments $\{x_1, \ldots, x_n\}$.

alors l'ensemble des applications de $\mathcal{Y}$ dans $\mathcal{X}$ (noté $\mathcal{X}^{\mathcal{Y}}$) est tel que

$$\mathrm{Card}(\mathcal{X}^{\mathcal{Y}}) = n^p.$$

En effet, il y a n manières de choisir $f(y_1)$ parmi $\{x_1, \ldots, x_n\}$, n manières de choisir $f(y_2)$ parmi $\{x_1, \ldots, x_n\}$, etc.

Définition 1.6 *Une* partition *d'un ensemble Ω est une famille d'ensembles $\{A_i\}_{i=1,\ldots,p}$ telle que :*

$$\bigcup_{i=1,\ldots,p} A_i = \Omega, \quad A_i \bigcap A_j = \emptyset \quad \textit{dès que} \quad i \neq j.$$

Dans tout ce qui suit, nous noterons $n!$ le produit $1 \times 2 \times 3 \times \cdots \times n$, ce produit s'appelle factorielle n. On convient que $0! = 1$. Le produit cartésien de p ensemble $E_1, E_2, \ldots, E_p$, noté $E_1 \times E_2 \times \cdots \times E_p$, représente l'ensemble des p-uplets $(e_1, e_2, \ldots, e_p)$ où $e_1 \in E_1, e_2 \in E_2, \ldots, e_p \in E_p$.

1.2.2 Principes de la somme et du produit

On appelle le principe de la somme, la relation entre la somme des cardinaux d'une partition de Ω et le cardinal de Ω. Si des ensembles $A_1, A_2, \ldots, A_p$ constituent une partition d'un ensemble fini Ω, alors :

$$\mathrm{Card}(A_1) + \mathrm{Card}(A_2) + \cdots + \mathrm{Card}(A_p) = \mathrm{Card}(\Omega).$$

Soient A et B deux parties d'un ensemble fini Ω. On déduit immédiatement du principe de la somme les relations suivantes :

$$\begin{aligned} \mathrm{Card}(A \cup B) &= \mathrm{Card}(A) + \mathrm{Card}(B) - \mathrm{Card}(A \cap B), \\ \mathrm{Card}(A^c) &= \mathrm{Card}(\Omega) - \mathrm{Card}(A). \end{aligned}$$

L'axiome 2 de la notion de probabilité (cf. définition 1.2) est cohérent avec ce principe. On peut vérifier ainsi que la probabilité uniforme définie par (1.2) satisfait bien les axiomes d'une probabilité décrits dans la définition 1.2.

Si une situation comporte p étapes offrant respectivement $n_1, n_2, \ldots, n_p$ possibilités alors le nombre total d'issues est :

$$n_1 \times n_2 \times \cdots \times n_p.$$

Du principe du produit (ou multiplicatif) découle le cardinal du produit cartésien : si $\Omega_1, \Omega_2, \ldots, \Omega_p$ sont p ensembles de cardinal fini, alors :

$$\mathrm{Card}(\Omega_1 \times \Omega_2 \times \cdots \times \Omega_p) = \mathrm{Card}(\Omega_1) \times \mathrm{Card}(\Omega_2) \times \cdots \times \mathrm{Card}(\Omega_p).$$

1.2.3 Dénombrement des p-listes

Définition 1.7 *Soient $n \in \mathbb{N}^*$ et Ω un ensemble de cardinal n. Soit $p \in \mathbb{N}$. Une p-liste (ou liste de longueur p) est un p-uplet d'éléments de Ω.*

Une p-liste est donc un élément de $\Omega^p = \underbrace{\Omega \times \cdots \times \Omega}_{p \text{ facteurs}}$. On suppose que la 0-liste existe, c'est la liste qui ne comporte aucun élément.

Théorème 1.8 *Soit Ω un ensemble de cardinal fini n. Le cardinal de l'ensemble Ω^p des p-listes de Ω est n^p.*

La démonstration de ce théorème découle simplement du principe multiplicatif.

1.2.4 Dénombrement des Arrangements et des Permutations

Définition 1.9 *Soit Ω un ensemble de cardinal fini n et p un entier naturel tel que $0 \leq p \leq n$. Un p-arrangement (ou arrangement de p éléments) de Ω est une p-liste d'éléments distincts de Ω. Une permutation de Ω est un arrangement des n éléments de Ω.*

Théorème 1.10 *Soit Ω un ensemble fini de cardinal n et p un entier naturel tel que $0 \leq p \leq n$.*

- *Le nombre d'arrangements de p éléments de Ω est :*

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

- *Le nombre de permutations de Ω est : $A_n^n = n!$.*

Exemples :

- Le tiercé : une course de chevaux comporte 20 partants. Combien peut-il y avoir de résultats possibles de tiercés dans l'ordre?
  Soit Ω l'ensemble des numéros des chevaux. On a $\mathrm{Card}(\Omega) = 20$. Un tiercé correspond à un arrangement de 3 éléments de Ω, il y en a $A_{20}^3 = 6840$.

- Nombre de mots de 5 lettres, ayant un sens ou non, de notre alphabet : A_{26}^5.
- Tirage ordonné : Une urne contient 10 boules numérotées $0, 1, \ldots, 10$. On en tire successivement trois sans remise. Combien y a-t-il de tirages différents ?

1.2.5 Dénombrement des Combinaisons

Définition 1.11 *Soit Ω un ensemble de cardinal fini n et p un entier naturel tel que $0 \leq p \leq n$. Un p-*combinaison *(ou combinaison de p éléments) de Ω est une p-liste non ordonnée d'éléments distincts de Ω. La 2-combinaison $(1, 2)$ regroupe les 2-listes $(1, 2)$ et $(2, 1)$.*

Théorème 1.12 *Soit Ω un ensemble fini de cardinal n et p un entier naturel tel que $0 \leq p \leq n$. Le nombre de combinaisons de p éléments de Ω est :*

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

Les coefficients C_n^p sont aussi appelés coefficients binômiaux. Ils sont parfois notés $\begin{pmatrix} n \\ p \end{pmatrix}$.
Si p est strictement supérieur à n, on convient que dans ce cas $C_n^p = 0$.

Remarque 1.13 *Bien que les coefficients C_n^p soient définis sous la forme d'une fraction, ils sont bien entiers.*

Preuve :
Dénombrons les arrangements de p éléments d'un ensemble fini Ω de cardinal n. Un arrangement est caractérisé par :
- Le choix d'une partie de Ω a p éléments (notons qu'il y a C_n^p choix de telles parties).
- La façon d'ordonner les p éléments de la partie choisie ($p!$ façons).

Le principe multiplicatif donne alors $A_n^p = p! C_n^p$, d'où le théorème. □

Exemples :
- Le loto : on tire au hasard six boules parmi 49. Combien y a-t-il de tirages possibles ? C'est le nombre de façons de choisir six objets parmi 49, soit $C_{49}^6 = 13983816$.
- Nombres de dominos avec 7 numéros : $C_7^2 + 7 = 28$.
- Tirage simultané ou non ordonné : une urne contient 10 boules, numérotées de 1 à 10. On en tire simultanément 3. Combien y a-t-il de tirages différents ?

1.2.6 Propriétés des coefficients binômiaux

Propriétés 1.14 (Symétrie) *Pour tout entier n et tout entier p tel que $0 \leq p \leq n$, on a :*

$$C_n^p = C_n^{n-p}.$$

On en déduit :

$$C_n^0 = C_n^n = 1;\ C_n^1 = C_n^{n-1} = n.$$

Propriétés 1.15 (Relation de Pascal) *Pour tout entier $n \geq 2$ et tout entier p tel que $1 \leq p \leq n-1$, on a :*

$$C_n^p = C_{n-1}^{p-1} + C_{n-1}^p.$$

Preuve :

- **Démonstration ensembliste :** Soit Ω un ensemble de cardinal fini n avec $n \geq 2$. Soit a un élément fixé de Ω. Remarquons que les parties à p éléments de Ω se partagent en deux catégories :
  - celles ne contenant pas a, il y en a C_{n-1}^p,
  - celles contenant a, il y en a C_{n-1}^{p-1}.

  Étant en présence d'une partition, le principe de la somme permet de conclure.
- **Démonstration algébrique :**

$$\begin{aligned} C_{n-1}^p + C_{n-1}^{p-1} &= \frac{(n-1)!}{p!(n-1-p)!} + \frac{(n-1)!}{(p-1)!(n-p)!} \\ &= \frac{(n-1)!(n-p) + (n-1)!p}{p!(n-p)!} = \frac{n!}{p!(n-p)!}. \end{aligned}$$

□

On en déduit que les coefficients C_n^p sont des entiers (pour tout entier naturel p compris entre 0 et n).

Le **triangle de Pascal**. La relation de Pascal permet de calculer les coefficients binômiaux de la façon suivante : pour trouver un certain coefficient, on additionne dans le tableau suivant les coefficients situés « juste au dessus » et « juste au dessus à gauche » entre eux :

n \ p	0	1	2	3	$\cdots$	p-1	p	$\cdots$	n-1	n
0	1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
1	1	1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
2	1	2	1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
3	1	3	3	1	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
$\cdots$	1				1	$\times$	$\times$	$\times$	$\times$	$\times$
p-1	1					1	$\times$	$\times$	$\times$	$\times$
p	1						1	$\times$	$\times$	$\times$
$\cdots$	1							1	$\times$	$\times$
n-1	1	n-1				C_{n-1}^{p-1}	C_{n-1}^p		1	$\times$
n	1	n-1					C_n^p			1

Le tableau est appelé **triangle de Pascal** en hommage à ce dernier qui écrivit en 1654 son « traité du triangle arithmétique » dans lequel il expose d'innombrables applications du triangle déjà connu de Tataglia (1556), Stiefel (1543) et des savants Chinois (1303). Pour finir, on rappelle un résultat majeur :

Théorème 1.16 (Formule du binôme de Newton) *Pour tous nombres complexes a et b et tout entier naturel n non nul :*

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

1.3 Autre exemple important : le schéma Succès-Échec fini.

Notre premier exemple important était la probabilité uniforme. Voyons maintenant un autre exemple tout aussi important.

Considérons le cas d'une expérience à deux issues, échec (noté 0) et succès (noté 1). On se fixe un nombre $0 < p < 1$, la probabilité d'obtenir un succès après une réalisation de l'expérience. Après n réalisations, on observe une suite de longueur n, composée de 1 et de 0. Pour modéliser cela, on introduit l'espace des observables $\Omega = \{0;1\}^n$ formé des 2^n suites $\omega_i = (\eta_1, \ldots, \eta_n)$ où les η_j sont égaux à 0 ou 1.

Introduisons alors la quantité $X(\omega_i)$ définie ainsi : si $\omega_i = (\eta_1, \ldots, \eta_n) \in \Omega$, alors $X(\omega_i)$ désigne le nombre de succès que comprend la suite ω_i. Par exemple, pour $n = 4$, $X(1101) = 3$, $X(0000) = 0$. Nous verrons dans le chapitre 3 que X est appelée variable aléatoire.

De plus, pour tout $\omega_i \in \Omega$ tel que $X(\omega_i) = k$, on pose $\mathbb{P}(\{\omega_i\}) = p^k(1-p)^{n-k}$. Comme tout événement $A \in \mathcal{P}(\Omega)$ est réunion de singletons $\{\omega_i\}$ deux à deux disjoints, cela suffit à définir $\mathbb{P}(A)$ et donc la probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{P}(\Omega))$.

Exercice 1.3.1 *Vérifier que $\mathbb{P}$ est bien une probabilité sur $(\Omega, \mathcal{P}(\Omega))$.*

Parmi ces événements, les plus importants sont les $\{X = k\}$ (ceci est une sténographie que nous utiliserons souvent pour écrire brièvement l'événement $\{\omega_i \in \Omega; X(\omega_i) = k\}$). Voici leur probabilité :

Proposition 1.17 *Pour le schéma Succès-Échec fini associé à la probabilité p d'un succès, si X est le nombre de succès en n expériences, alors $\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$.*

Preuve :

Notons $A = \{\omega_i \in \Omega; X(\omega_i) = k\}$. Définissons l'application de A dans $\mathcal{P}(\{1, 2, \ldots, n\})$ par $\omega_i \mapsto \{j; \eta_j = 1\}$. Il est clair que c'est une bijection entre A et l'ensemble des k-combinaisons de $\{1, 2, \ldots, n\}$; donc $\mathrm{Card}(A) = C_n^k$. Enfin puisque tous les $\{\omega_i\}$ contenus dans A ont la même probabilité $p^k(1-p)^{n-k}$, on obtient

$$\mathbb{P}(A) = \sum_{\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \mathrm{Card}(A) p^k (1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}.$$

□

Pour clore ce chapitre, remarquons que le cas où l'espace des observables est fini, est loin d'être satisfaisant dans bon nombre d'expériences aléatoires. Par exemple, considérons le schéma Succès-Échec (exemple type : jeu du *pile* ou *face*). Il peut être intéressant de chercher le premier instant où un Échec va se produire. L'espace des observables est alors infini : on lance la pièce jusqu'à tomber sur *face*, ce qui peut prendre un temps très grand (notamment si la pièce est très déséquilibrée). Ω est alors tous les suites possibles, composées de zéro (Échec) ou de un (Succès). Si on veut modéliser la distance électron-noyau dans un atome, la mécanique quantique nous apprend que cette distance est aléatoire et prend *a priori* tous les valeurs positives. Donc Ω est dans ce cas $\mathbb{R}_+^* =]0, +\infty[$. Que dire alors de $\mathcal{P}(\Omega)$? Simplement que c'est un ensemble « beaucoup plus gros » que Ω (ce qui n'est pas le cas si Ω est fini). La définition 1.2 n'est plus valable et on doit alors construire autrement les probabilités. C'est l'enjeu du chapitre suivant.

1.4 Exercices

Exercice 1.4.1 *A et B sont des évènements tels que* $P(A) = 0.6$, $P(B) = 0.3$ *et* $P(A \cap B) = 0.1$.

1. *Quelle est la probabilité que A ou B arrivent ?*
2. *Quelle est la probabilité qu'exactement un des deux évènements arrive ?*
3. *Quelle est la probabilité qu'au plus un des deux évènements arrive ?*
4. *Quelle est la probabilité que ni A ni B n'arrivent ?*

Exercice 1.4.2 *Sur* $\Omega = \{0, 1, 2, \ldots, 10\}$ *déterminer la probabilité* $\mathbb{P}$ *telle que* $\mathbb{P}(\{i\})$ *soit proportionnel à i pour tout i dans* Ω. *Calculer la probabilité pour qu'un événement soit pair puis la probabilité pour qu'il soit impair.*

Dans chacun des exercices suivants on commencera par décrire l'espace probabilisé $(\Omega, \mathbb{P})$ avant de passer aux calculs.

Exercice 1.4.3 *On tire 5 cartes d'un jeu de 52 cartes. Quelle est la probabilité d'obtenir un brelan (c'est-à-dire trois cartes d'une sorte, les deux autres étant différentes) ? Quelle est la probabilité d'obtenir un full (trois mêmes cartes et une paire) ?*

Exercice 1.4.4 (Trois dés) *On jette trois dés non pipés. Calculer :*

1. *la probabilité d'obtenir au moins un as;*
2. *la probabilité d'obtenir au moins deux faces portant le même chiffre;*
3. *la probabilité que la somme des points marqués sur les trois faces soit paire.*

Exercice 1.4.5 *En admettant que les naissances sont réparties uniformément dans l'année, déterminer la probabilité que parmi n individus deux au moins d'entre eux aient même jour d'anniversaire. On négligera le problème des années bissextiles.*

Exercice 1.4.6 (le paradoxe du Chevalier de Méré) *:*

Ce personnage marquant de la cour de Louis XIV était un joueur impénitent, toujours à la recherche de règles cachées lui permettant de réaliser un avantage sur ses adversaires. Voici deux de ses "régles".

1. *Il est avantageux de parier sur l'apparition d'au moins un "six" en lançant un dé quatre fois de suite.*
2. *Il est avantageux de parier sur l'apparition d'au moins un "double-six" en lançant deux dés vingt-quatre fois de suite.*

Ces règles sont-elles vraies ?

Exercice 1.4.7 *Dans une bibliothèque, n livres sont exposés sur une étagère rectiligne et répartis au hasard. Parmi ces n livres, k sont d'un même auteur A, les autres étant d'auteurs tous différents. Calculer la probabilité qu'au moins p livres de A se retrouvent côte à côte dans les cas suivants :*

1. $n = 20,\ k = 3,\ p = 3$;
2. $n = 20,\ k = 5,\ p = 2$.

Chapitre 2

Axiomes du calcul des probabilités

"La théorie des probabilités en tant que discipline mathématique peut et doit être développée à partir d'axiomes de la même manière qu'en Géométrie et en Algèbre."
A. N. Kolmogorov (1903-1987).

Le calcul des probabilités est la science qui modélise les expériences dont l'issue n'est pas prévisible *a priori*. Le jet d'un dé, le tirage du Loto sont des exemples classiques de ces expériences, dites aléatoires. Une modélisation implique une simplification des phénomènes observés dans ces expériences, mais cette simplification conduit à une quantification, donc à la possibilité de faire des calculs et de prédire. La modélisation du calcul des probabilités a été inventée par A. N. Kolmogorov dans un livre (intitulé « Les fondements de la théorie des probabilités ») paru en 1933. Cette modélisation mathématique est faite à partir d'un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, permettant de traiter les cas où l'espace des observables Ω n'est pas fini.

2.1 L'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$

2.1.1 L'espace des observables Ω

Nous conviendrons qu'effectuer une expérience aléatoire, c'est sélectionner par un procédé quelconque un élément ω dans un ensemble Ω. Par exemple, jeter un dé revient à sélectionner un élément de $\Omega = \{1, 2, 3, 4, 5, 6\}$. Cet ensemble Ω est appelé l'*espace des observables* (ou *espace des issues* ou espace des événements élémentaires). Ses points ω sont appelés observables (ou événements élémentaires). Il est très important qu'il soit clairement défini afin de regrouper au moins tous les résultats possibles de l'expérience aléatoire.

2.1.2 La tribu des événements $\mathcal{F}$

Le résultat d'une expérience aléatoire est par définition imprévisible. Toutefois, une fois fixé A un sous ensemble de l'espace des observables Ω, on s'intéresse à la possibilité qu'a le résultat ω de l'expérience de tomber dans A. Les parties de Ω pour lesquelles on se pose ce genre de question sont appelées des *événements*. Un des premiers points délicats de la théorie est qu'on ne va pas considérer tous les sous ensembles de Ω comme des événements. L'idée de Kolmogorov est que l'ensemble $\mathcal{F}$ des événements a une structure de tribu :

Définition 2.1 (Tribu) *Soit Ω un ensemble et soit $\mathcal{F}$ une partie de $\mathcal{P}(\Omega)$. $\mathcal{F}$ est une tribu s'il satisfait aux trois axiomes:*

1. *Si $A \in \mathcal{F}$, alors son complémentaire $A^c = \Omega \setminus A$ est aussi dans $\mathcal{F}$.*
2. *Si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{F}$, alors leur réunion $\bigcup_{n\geq 1} A_n$ est aussi dans $\mathcal{F}$.*
3. *L'ensemble vide $\emptyset$ est dans $\mathcal{F}$.*

Un élément de $\mathcal{F}$ est appelé un événement.

Rappelons qu'un ensemble Ω avec une infinité d'éléments est dit *dénombrable* s'il existe une bijection entre Ω et l'ensemble $\mathbb{N}$ des entiers positifs. L'ensemble $\mathbb{Q}$ des nombres rationnels est dénombrable, le segment $[0, 1]$ ne l'est pas.

Tirons quelques conséquences de ces axiomes.

Propriétés 2.2 *Soit $\mathcal{F}$ une tribu de parties de l'ensemble Ω. Alors $\Omega \in \mathcal{F}$.*
De plus, si on a une suite finie ou dénombrable $A_1, \ldots, A_n, \ldots$ d'éléments de $\mathcal{F}$, alors leur intersection $\bigcap_{n\geq 1} A_n$ est aussi dans $\mathcal{F}$.

Preuve :

En appliquant les axiomes 1 et 3, on a le premier résultat. Pour le second, il suffit de se rappeler que le complémentaire d'une réunion finie ou infinie d'ensembles est l'intersection des complémentaires (« Loi de Morgan »). Donc $\bigcap_{n\geq 1} A_n = (\bigcup_{n\geq 1} A_n^c)^c$, et le deuxième membre de cette égalité est donc dans $\mathcal{F}$: on applique successivement l'axiome 1, puis 2, puis 1 à nouveau. □

Le langage de la théorie des ensembles permet des calculs systématiques sur les événements. Toutefois, il faut savoir que le langage courant, que nous utilisons dans une première étape pour décrire des événements a sa traduction ensembliste :

- Ensemble Ω : *événement certain*,
- Ensemble vide $\emptyset$: *événement impossible*,
- $A \cup B$: A ou B sont réalisés (« ou » non exclusif),
- $A \cap B$: A et B sont réalisés,
- A et B sont disjoints, i.e. $A \cap B = \emptyset$: les événements A et B sont incompatibles,
- $A^c = \Omega \setminus A$: événement contraire de A.

D'après cette traduction ensembliste, le fait qu'on ne sorte pas de la famille des événements intéressants à considérer en prenant une intersection ou une réunion d'événements, est raisonnable, tant qu'on considère un nombre fini d'événements. C'est toujours le cas si l'ensemble Ω est fini, le nombre de parties de Ω étant aussi fini (voir la proposition 1.1). Le fait de se permettre ces opérations aussi pour un nombre infini d'événements est plus subtil. Le passage à l'infini est le passage de l'algèbre à l'analyse, donc à des approximations maniables et à de puissantes techniques issues du calcul différentiel et intégral. Quant au fait que dans ce passage à l'infini, on se limite à une infinité dénombrable d'événements (propriété 2 dans la définition 2.1), c'est un point technique qu'on ne justifiera que dans un cours de 3ème année d'université.

Finalement, ce point délicat : « on ne considère pas nécessairement tout sous-ensemble A de Ω comme un élément de la tribu $\mathcal{F}$ des événements » ne jouera pas un grand rôle dans la suite. Typiquement, nous envisagerons deux cas particuliers importants :

1. Le cas où Ω lui même est dénombrable, et nous prendrons comme tribu $\mathcal{F}$ la famille $\mathcal{P}(\Omega)$ de tous les sous-ensembles de Ω. C'est ce que nous avons fait dans le cas Ω fini.
2. Le cas où Ω est la droite réelle $\mathbb{R}$. Nous prendrons alors pour tribu $\mathcal{F}$ la tribu $\mathcal{B}(\mathbb{R})$ (dite *tribu de Borel*, dont les éléments sont appelés des *boréliens*) qui est la plus petite tribu qui contient tous les intervalles de $\mathbb{R}$.

Définition 2.3 *La plus petite tribu qui contient les ouverts de $\mathbb{R}^d$ muni de sa topologie canonique est appelée* tribu de Borel. *Les éléments de cette tribu sont appelés les* boréliens *de $\mathbb{R}^d$.*

On peut laborieusement démontrer que $\mathcal{B}(\mathbb{R}^d) \neq \mathcal{P}(\mathbb{R}^d)$; toutefois, une description complète des éléments de $\mathcal{B}(\mathbb{R}^d)$ n'est pas possible, et en fait pas très utile en pratique. Les seuls boréliens que nous aurons à manipuler seront les intervalles (attention, $\mathbb{R}$ ou une demi droite sont aussi des intervalles) ou des réunions finies, ou plus rarement dénombrables, d'intervalles.

Plus généralement, étant donnés un ensemble de parties de Ω noté $\mathcal{C}$, on définit la notion de *tribu engendrée par* $\mathcal{C}$, notée $\sigma(\mathcal{C})$, comme étant la plus petite tribu contenant tous les éléments de $\mathcal{C}$.

2.1.3 La probabilité $\mathbb{P}$

On quantifie désormais la possibilité qu'a le résultat ω de l'expérience de tomber dans un événement $A \in \mathcal{A}$, sous-ensemble de l'espace des observables Ω. La puissance du calcul des probabilités réside dans le fait qu'il ne manipule pas directement le résultat ω de l'expérience aléatoire mais mesure les chances d'arrivée d'un événement A du au hasard.

Définition 2.4 *Étant donnés un espace d'observables Ω et une tribu d'événements $\mathcal{F}$ formée de certains sous-ensembles de Ω, une probabilité $\mathbb{P}$ est une application de $\mathcal{F}$ dans $[0, 1]$, donc un procédé qui associe à tout événement A un nombre $\mathbb{P}(A)$ compris entre 0 et 1 appelé* probabilité *de A, et qui satisfait aux axiomes suivants :*

1. *l'événement certain est de probabilité 1 : $\mathbb{P}(\Omega) = 1$.*
2. **Axiome d'additivité dénombrable :**
   pour toute suite $A_1, A_2, \ldots, A_n \ldots$ d'événements de $\mathcal{F}$ qui sont deux à deux disjoints, c'est-à-dire tels que $A_k \cap A_j = \emptyset$ si $k \neq j$, alors la série $\sum_{k=1}^{\infty} \mathbb{P}(A_k)$ converge et a pour somme $\mathbb{P}(\bigcup_{k \geq 1} A_k)$.

Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est alors appelé un espace de probabilité.

Remarque 2.5 Attention *: le deuxième axiome étend celui de la définition 1.2 donnée dans le cas où Ω est fini. Nous traitons d'ensembles d'événements Ω pas nécessairement finis.*

Voici quelques conséquences immédiates, mais néanmoins utiles, des axiomes.

Propriétés 2.6 *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Alors*

1. $\mathbb{P}(\emptyset) = 0$.
2. *Si $A_1, A_2, \ldots, A_n$ dans $\mathcal{F}$ sont deux à deux disjoints, alors*
$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n);$$
en particulier $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
3. *Si A et B sont dans $\mathcal{F}$ et si $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.*
4. *Si A et B sont dans $\mathcal{F}$, mais ne sont pas nécessairement disjoints, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Si les $A_1, A_2, \ldots, A_n$ dans $\mathcal{F}$ ne sont pas nécessairement deux à deux disjoints, alors $\mathbb{P}(A_1 \cup \cdots \cup A_n) \leq \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_n)$.*
5. **Continuités croissante et décroissante :**
Soit une suite $B_1, B_2, \ldots, B_n \ldots$ d'événements de $\mathcal{F}$ qui soit ou bien croissante (c'est-à-dire que pour tout $n \geq 1$ on a $B_n \subset B_{n+1}$) ou bien décroissante (c'est-à-dire que pour tout $n \geq 1$ on a $B_n \supset B_{n+1}$). Alors, dans le cas croissant :
$$\lim_{n \to +\infty} \mathbb{P}(B_n) = \mathbb{P}(\bigcup_{n \geq 1} B_n);$$
et dans le cas décroissant :
$$\lim_{n \to +\infty} \mathbb{P}(B_n) = \mathbb{P}(\bigcap_{n \geq 1} B_n).$$
6. **Sous additivité dénombrable :**
Soit une suite $B_1, B_2, \ldots, B_n \ldots$ d'événements de $\mathcal{F}$. Alors ou bien la série $\sum_{k=1}^{\infty} \mathbb{P}(B_k)$ diverge; ou bien elle converge et dans ce cas sa somme est supérieure à $\mathbb{P}(\bigcup_{n \geq 1} B_n)$.

Preuve :

1. L'axiome d'additivité dénombrable est applicable à la suite constante définie par $A_n = \emptyset$, qui est effectivement formée d'événements deux à deux disjoints. La série dont le terme général $\mathbb{P}(\emptyset)$, est constant ne peut converger que si ce terme général est 0.
2. Sa première partie se démontre en appliquant l'axiome d'additivité dénombrable à $A_1, A_2, \ldots, A_n$ continuée par $\emptyset = A_{n+1} = A_{n+2} = \cdots$, et en utilisant le 1. Appliquer ça à $n = 2$, $A_1 = A$ et $A_2 = A^c$ fournit $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$ en utilisant le premier axiome d'une probabilité.

3. On écrit $B = A \cup (B \setminus A)$ comme réunion de deux ensembles disjoints (notez que $B \setminus A = B \cap A^c$ est bien dans $\mathcal{F}$), et on applique le 2):

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

4. On écrit comme dans la démonstration précédente : $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A), \mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B)$, puis on écrit $A \cup B = (A \cap B) \cup (B \setminus A) \cup (A \setminus B)$ comme réunion de trois ensembles deux à deux disjoints et on applique le 2) :

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

Pour terminer le 4, on démontre le résultat par récurrence sur n. C'est trivial pour $n = 1$. Si c'est démontré pour n, appliquons la première partie de 4 à $A = A_1 \cup \cdots \cup A_n$ et à $B = A_{n+1}$. On obtient, à l'aide de l'hypothèse de récurrence

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \leq \mathbb{P}(A) + \mathbb{P}(B) \leq (\sum_{k=1}^{n} \mathbb{P}(A_k)) + \mathbb{P}(A_{n+1}).$$

5. Dans le cas croissant, posons $A_1 = B_1$ et, pour $n \geq 2$, $A_n = B_n \setminus B_{n-1}$. Les $A_1, A_2, \ldots, A_n \ldots$ sont alors deux à deux disjoints. La série $\sum_{k=1}^{\infty} \mathbb{P}(A_k)$ est donc convergente. D'après la partie 2) de la proposition précédente, on a

$$\mathbb{P}(B_n) = \mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \mathbb{P}(A_k).$$

Passons à la limite dans l'égalité ci-dessus; on obtient

$$\lim_{n \to +\infty} \mathbb{P}(B_n) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Or d'après l'axiome d'additivité dénombrable, le second membre est $\mathbb{P}(\bigcup_{k \geq 1} A_k)$, qui est aussi par définition des A_n égal à $\mathbb{P}(\bigcup_{n \geq 1} B_n)$.
Dans le cas décroissant, on se ramène au cas précédent par passage au complémentaires, à l'aide de la loi de Morgan: le complémentaire d'une union est l'intersection des complémentaires :

$$\lim \mathbb{P}(B_n) = 1 - \lim \mathbb{P}(B_n^c) = 1 - \mathbb{P}(\cup_{n \geq 1} B_n^c) = 1 - (1 - \mathbb{P}(\cap_{n \geq 1} B_n)) = \mathbb{P}(\cap_{n \geq 1} B_n).$$

6. La suite d'événements définie par $C_n = B_1 \cup \cdots \cup B_n$ est croissante et on peut lui appliquer le 2). En utilisant aussi la sous additivité finie on a donc

$$\mathbb{P}(\bigcup_{n \geq 1} B_n) = \lim_{n \to +\infty} \mathbb{P}(C_n) \leq \lim_{n \to +\infty} (\mathbb{P}(B_1) + \cdots + \mathbb{P}(B_n)) = \sum_{k=1}^{\infty} \mathbb{P}(B_k).$$

□

Exercice 2.1.1 (Formule de Poincaré) *Cette formule est aussi appelée aussi principe d'exclusion-inclusion. H. Poincaré, 1854-1912). Montrer que :*

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \ldots \\ &+ (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n),\end{aligned}$$

ou encore :

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} S_k, \text{ où } S_k = \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} \mathbb{P}(A_{i_1} \cap \ldots A_{i_k}).$$

Vérifier de plus que les inégalités (de Bonferroni*) sont satisfaites :*

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) &\le S_1 \\ &\ge S_1 - S_2 \\ &\le S_1 - S_2 + S_3 \\ &\vdots\end{aligned}$$

2.2 Indépendance et probabilités conditionnelles

Soit un espace d'observables Ω, $\mathcal{F}$ une tribu sur Ω, et soit $\mathbb{P}$ une probabilité sur $\mathcal{F}$. On dit que $\mathcal{F}'$ est une sous-tribu de $\mathcal{F}$, si tout élément de $\mathcal{F}'$ est contenu dans $\mathcal{F}$ et si $\mathcal{F}'$ est une tribu.

Définition 2.7 (Indépendance de n tribus) *Soit $\mathcal{F}_1, \ldots, \mathcal{F}_n$ des sous-tribus de $\mathcal{F}$. On dit que ces tribus sont indépendantes si, quel que soit le choix de A_i dans $\mathcal{F}_i$:*

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

Il faut souligner le rôle essentiel de $\mathbb{P}$. Si on change de probabilité sur $\mathcal{F}$, il n'y a pas de raison pour que des tribus, qui étaient indépendantes, le restent sous la nouvelle probabilité.

Pour vérifier l'indépendance entre les tribus, il faut apparemment vérifier beaucoup de relations. Mais comme souvent dans les questions qui font intervenir les tribus, on peut se contenter de les vérifier pour des sous-classes.

Proposition 2.8 *Soient $\mathcal{C}_i$, $1 \le i \le n$, des sous-ensembles de $\mathcal{F}$, tels que pour tout i, $\Omega \in \mathcal{C}_i$ et pour tout A et B dans $\mathcal{C}_i$, $A \cap B \in \mathcal{C}_i$. Si*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}(A_i),$$

pour tout choix de A_i dans $\mathcal{C}_i$, alors les sous-tribus $\sigma(\mathcal{C}_i)$ sont indépendantes.

Ainsi pour vérifier que deux événements A et B sont indépendants, il suffit de vérifier $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Plus généralement, si $\{A_n\}_{n\in\mathbb{N}}$ est une suite d'événements :

Corollaire 2.9 *$\{A_n\}_{n\in\mathbb{N}}$ sont indépendants si pour tout ensemble fini d'indices $I \subset \mathbb{N}$:*

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

Par exemple pour trois événements $\{A, B, C\}$ il faut vérifier que

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

$$\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C),\ \text{et } \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C).$$

On peut généraliser la définition précédente à un ensemble quelconque de tribus.

Définition 2.10 (Indépendance de tribus) *Une famille $\mathcal{F}_i$, $i \in I$, de sous-tribus de $\mathcal{F}$ est dite indépendante si toute sous-famille finie extraite de la famille est indépendante.*

2.2.1 Probabilités conditionnelles

On s'intéresse au lancer successif de deux dés. On considère les deux événements suivants : A : « la somme des deux dés est égale à 11 » et B : « le premier dé donne 6 ». On vérifie que la probabilité de A est $1/18$.

Mais si on observe le premier dé et que l'événement B est réalisé, en quoi cela affecte la probabilité d'avoir une somme égale à 11 ? Autrement dit, sachant que B est réalisé, quelle est la probabilité de A ? Pour obtenir 11, le deuxième dé doit donner 5. Et ainsi la probabilité d'avoir A, sachant B, est de $1/6$.

Cet exemple motive la définition suivante.

Définition 2.11 *Soient deux événements A et B sur $(\Omega, \mathbb{P})$, avec $\mathbb{P}(B) > 0$. La* probabilité conditionnelle de A sachant B *est*

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

Si B ne se réalise jamais, i.e. $\mathbb{P}(B) = 0$, cette définition n'a aucun sens.

Proposition 2.12 *Soit deux événements A et B sur $(\Omega, \mathbb{P})$, avec $\mathbb{P}(B) > 0$. A et B sont indépendants si et seulement si*

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

ou encore avec $\mathbb{P}(A) > 0$, $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Dire que deux événements sont indépendants, c'est dire que la connaissance de l'un n'influence pas l'autre.

Exemple 1. On modélise le lancer de deux pièces semblables indépendantes par

$$\Omega = \{\omega = (\omega_1, \omega_2);\ \omega_i \in \{p, f\}\} \text{ et } \mathbb{P}(\omega) = \mathbb{P}_1(\omega_1)\mathbb{P}_1(\omega_2),$$

avec $\mathbb{P}_1(p) = 1 - \mathbb{P}_1(f) = \alpha$. On s'imagine qu'on a lancé la première pièce et que $\omega_1 = p$. On associe à chaque issue du lancer de la deuxième pièce une probabilité (sachant que $\omega_1 = p$) :

$$\mathbb{P}(\omega_2 = p|\omega_1 = p) = \frac{\mathbb{P}(\{p,p\})}{\mathbb{P}(\{\omega;\ \omega_1 = p\}} = \frac{\alpha\alpha}{\alpha(1-\alpha) + \alpha\alpha} = \alpha.$$

L'indépendance des deux pièces signifie que la connaissance du résultat du lancer de la première ne nous apprend rien sur le résultat du lancer de la seconde.

Exemple 2. On suppose qu'on a deux paires de pièces différentes : la 1ère paire consiste en deux pièces indépendantes et pile sort 99 fois sur 100 en moyenne; la deuxième paire consiste en deux pièces indépendantes mais pile sort 1 fois sur 10 en moyenne. Un jeu consiste en : (i) choisir au hasard une d'entre les deux paires, (ii) lancer les deux pièces correspondantes. On modélise l'espace des issues d'un jeu par :

$$\Omega = \{\omega = (\omega_1, \omega_2, n);\ \omega_i \in \{p, f\},\ n = 1, 2\},$$

et

$$\mathbb{P}(\omega_1, \omega_2, 1) = \frac{1}{2}\mathbb{P}_1(\omega_1)\mathbb{P}_1(\omega_2) \text{ et } \mathbb{P}_1(p) = \alpha = 0,99$$

et

$$\mathbb{P}(\omega_1, \omega_2, 2) = \frac{1}{2}\mathbb{P}_2(\omega_1)\mathbb{P}_2(\omega_2) \text{ et } \mathbb{P}_2(p) = \beta = 0,1.$$

On calcule ici

$$\begin{aligned}
\mathbb{P}(\omega_2 = p|\omega_1 = p) &= \frac{\mathbb{P}(\{\omega;\ \omega_2 = p, \omega_1 = p\})}{\mathbb{P}(\{\omega;\ \omega_1 = p\}} \\
&= \frac{\mathbb{P}(p,p,1) + \mathbb{P}(p,p,2)}{\mathbb{P}(p,p,1) + \mathbb{P}(p,f,1) + \mathbb{P}(p,p,2) + \mathbb{P}(p,f,2)} \\
&= \frac{\alpha^2/2 + \beta^2/2}{\alpha/2 + \beta/2} = \frac{\alpha^2 + \beta^2}{\alpha + \beta}
\end{aligned}$$

Par ailleurs si on ne savait rien sur la première pièce

$$\mathbb{P}(\omega_2 = p) = \mathbb{P}(\{p,p,1\}, \{f,p,1\}, \{p,p,2\}, \{f,p,2\}) = \frac{1}{2}(\alpha + \beta).$$

On note que $\mathbb{P}(\omega_2 = p|\omega_1 = p) > \mathbb{P}(\omega_2 = p)$ (resp. 0,91 et 0.54). Intuitivement, comme pile est sorti la 1ère fois, on a plus de chance que la paire choisie soit la paire 1.

Quelques règles de calcul sur les probabilités conditionnelles.

Proposition 2.13 *On déduit de la définition :*

1. *Si* $A \cap B = \emptyset$, $\mathbb{P}(A \cup B|C) = \mathbb{P}(A|C) + \mathbb{P}(B|C)$.

2. $\mathbb{P}(\overline{A}|B) = 1 - \mathbb{P}(A|B)$.
3. $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$.

Proposition 2.14 (Formule des probabilités totales) *Soit $\{A_i\}_{i=1,\ldots,n}$ une partition de Ω, telle que pour tout i, $\mathbb{P}(A_i) > 0$. Pour tout $B \in \mathcal{F}$,*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

Proposition 2.15 (Formule de Bayes) *De plus si $\mathbb{P}(B) > 0$, on a pour tout entier k :*

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\mathbb{P}(A_k)}{\sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i)}.$$

2.3 L'espace des observables est au plus dénombrable.

Lorsque l'espace des observables Ω est fini ou dénombrable, on suppose par convention que la tribu des événements $\mathcal{A}$ est $\mathcal{P}(\Omega)$, l'ensemble de toutes les parties de Ω. Dans ce contexte, les éléments de Ω sont indexés par des entiers naturels $\mathbb{N}$. On note dans la suite $\Omega = \{\omega_1, \ldots, \omega_n, \ldots\}$ (la liste est finie ou non). Les probabilités sont alors décrites par le résultat suivant :

Proposition 2.16 *Soit Ω un ensemble fini ou dénombrable. Soit $\omega \mapsto p_\omega$ une application de Ω dans $\mathbb{R}_+$ telle que $\sum_{w \in \Omega} p_w = 1$. Pour tout $A \subset \Omega$, notons alors $\mathbb{P}(A) = \sum_{w \in A} p_w$. Alors $(\Omega, \mathcal{P}(\Omega), P)$ est un espace de probabilité. Inversement, toute probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ est du type précédent, avec $p_\omega = P(\{\omega\})$.*

- Si Ω est fini, la proposition est évidente.
- Si Ω est dénombrable, les sommes ci dessus sont bien définies. En effet, on somme sur un sous ensemble de $\Omega = \{\omega_1, \ldots, \omega_n, \ldots\}$, en bijection avec un sous ensemble de $\mathbb{N}$ et le théorème suivant sur les séries permet de conclure à la validité de telles sommes.

Théorème 2.17 *Si la série $\sum_{n=0}^{\infty} u_n$ est absolument convergente de somme S, et si $n \mapsto \sigma(n)$ est une bijection de $\mathbb{N}$ sur lui même, alors $\sum_{n=0}^{\infty} u_{\sigma(n)}$ est aussi absolument convergente et de somme S.*

Une fois admis la proposition 2.16 et le théorème 2.17, le calcul des probabilités est rapporté à :

1. la détermination des probabilités $P(\{\omega\})$ des événements correspondant à la réalisation du résultat ω de l'expérience,
2. la détermination des résultats ω appartenant à l'événement A.

On ne revient pas sur le cas où Ω est fini. C'est l'objet du premier chapitre. Nous nous focalisons à présent sur le cas d'un espace des observables Ω infini.

Proposition 2.18 *Si Ω est infini dénombrable, alors $\mathcal{P}(\Omega)$ est infini non dénombrable.*

Preuve :
La démonstration est analogue à la démonstration de Cantor. Sans perte de généralité on suppose Ω égal à l'ensemble $\mathbb{N}$ des entiers positifs ou nuls. Si $X \subset \mathbb{N}$, on lui associe la fonction indicatrice $\mathbf{1}_X$ définie sur $\mathbb{N}$ et à valeurs 0 ou 1 par : $\mathbf{1}_X(k) = 1$ si $k \in X$ et $\mathbf{1}_X(k) = 0$ si $k \notin X$. Remarquons aussi qu'inversement, si une fonction f définie sur $\mathbb{N}$ est à valeurs 0 ou 1, alors c'est une indicatrice d'ensemble, c'est-à-dire qu'il existe X tel que $f = \mathbf{1}_X$: il s'agit de $X = \{k \in \mathbb{N}; f(k) = 1\}$.

Montrons alors la proposition par l'absurde en supposant que $\mathcal{P}(\mathbb{N})$ soit dénombrable, c'est-à-dire qu'il existe une application bijective $n \mapsto X_n$ de $\mathbb{N}$ sur $\mathcal{P}(\mathbb{N})$. Alors la fonction f définie sur $\mathbb{N}$ et à valeurs 0 ou 1 par $f(k) = 1 - \mathbf{1}_{X_k}(k)$ est l'indicateur de quelque sous-ensemble X_n de $\mathbb{N}$ et donc pour tout k de $\mathbb{N}$ on a $\mathbf{1}_{X_n}(k) = 1 - \mathbf{1}_{X_k}(k)$, ce qui est une contradiction si $k = n$. □

2.3.1 Un exemple : le schéma Succès-Échec infini.

Considérons de nouveau le cas d'une expérience à deux issues, échec (noté 0) et succès (noté 1). On se fixe un nombre $0 < p < 1$, la probabilité d'obtenir un succès après une réalisation de l'expérience. Nous souhaitons désormais modéliser des expériences ayant un nombre arbitraire de réalisations. Par exemple, on peut vouloir répéter les essais jusqu'à ce qu'apparaissent 4 succès consécutifs. Une telle modélisation est impossible avec le schéma fini (voir la section 1.3), et on prend alors pour espace Ω des observables l'ensemble $\{0,1\}^{\mathbf{N}^*}$ des suites infinies de 0 et de 1, en notant par $\mathbf{N}^*$ l'ensemble des entiers strictement positifs.

On peut démontrer que Ω est en bijection avec les parties de $\mathbf{N}^*$. Donc d'après la proposition 2.18, Ω n'est pas dénombrable. Cela cause une sérieuse difficulté en ce qui concerne la construction de l'espace de probabilité correspondant. On construit la tribu $\mathcal{F}$ et la probabilité $\mathbb{P}$ par un procédé d'approximation que nous décrivons maintenant.

Fixons $n \in \mathbb{N}^*$ et définissons $\Omega' = \{0,1\}^{\{1,\ldots,n\}}$ et $\Omega'' = \{0,1\}^{\{n+1,n+2,\ldots\}}$ tels que $\Omega = \Omega' \times \Omega''$. On définit la tribu suivante de parties de Ω : $\mathcal{A}_n = \{A \times \Omega''; A \in \mathcal{P}(\Omega')\}$. Intuitivement, les événements de $\mathcal{F}_n$ sont les événements ne dépendant que de ce qui s'est passé dans les n premières expériences. En particulier, nous avons $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Si $\omega' = (\omega_1, \ldots, \omega_n) \in \Omega'$ comprend k succès, définissons la probabilité $\mathbb{P}_n(\{\omega'\} \times \Omega'') = p^k(1-p)^{n-k}$. Cela permet donc de définir la probabilité $\mathbb{P}_n$ sur $\mathcal{A}_n$. L'espace de probabilité $(\Omega, \mathcal{A}_n, \mathbb{P}_n)$ est presque identique à l'espace du schéma Succès-Échec fini décrit dans la section 1.3.

Maintenant, notons $\mathcal{A}' = \cup_{n\geq 1}\mathcal{A}_n$. La famille $\mathcal{A}'$ n'est pas une tribu, car ce n'est pas fermé pour la réunion dénombrable. Voici un contre exemple. Soit A_n l'ensemble des suites ω infinies comprenant au moins un succès à l'instant n ou avant. Alors A_n est dans $\mathcal{A}_n$ et donc dans $\mathcal{A}'$. Pourtant $A = \cup_{n\geq 1}A_n$ n'est pas dans $\mathcal{A}'$. En effet A est l'ensemble des suites ω infinies comprenant au moins un succès. Mais il n'existe pourtant aucun n tel que $A \in \mathcal{A}_n$, et donc $A \notin \mathcal{A}'$. Réaliser cette chose subtile fait progresser dans la compréhension de la théorie. On définit alors la tribu $\mathcal{A}$ sur Ω comme la plus petite tribu contenant $\mathcal{A}'$.

Pour définir enfin la probabilité $\mathbb{P}$ sur $\mathcal{A}$, on fait l'observation essentielle suivante: on a non seulement $\mathcal{A}_n \subset \mathcal{A}_{n+1}$, mais de plus la restriction de $\mathbb{P}_{n+1}$ au sous ensemble $\mathcal{A}_n$ de $\mathcal{A}_{n+1}$, qui était le domaine de définition de P_{n+1}, coincide avec $\mathbb{P}_n$. Par conséquent, il existe une fonction universelle $\mathbb{P}'$ définie sur $\mathcal{A}'$ telle que pour tout $A \in \mathcal{A}'$ on ait $\mathbb{P}'(A) = \mathbb{P}_n(A)$ pour tous les n tels que $A \in \mathcal{A}_n$. A partir de ce point, les choses cessent d'être élémentaires, et nous admettons le théorème suivant :

Théorème 2.19 *Il existe une et une seule probabilité $\mathbb{P}$ sur $\mathcal{A}$ telle que pour tout $A \in \mathcal{A}'$ on ait $\mathbb{P}(A) = \mathbb{P}'(A)$.*

On peut ainsi démontrer l'idée intuitive qu'un événement (ici un succès) de probabilité strictement positive $p > 0$, même petite, finit toujours par arriver. Plus précisément, si A est l'ensemble des $\omega \in \Omega$ comprenant au moins un succès, alors $\mathbb{P}(A) = 1$. En effet, si B_n est l'ensemble des $\omega \in \Omega$ comprenant au moins un succès avant l'instant n ou à l'instant n, alors $A = \cup_{n\geq 1}B_n$ et $B_n \subset B_{n+1}$. Par continuité monotone on a donc $\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(A)$. Comme $\mathbb{P}(B^c) = (1-p)^n$ tend vers 0, on a le résultat.

Plus généralement on peut montrer que toute séquence a finie donnée à l'avance (par exemple 00011100010110001, ou le codage en binaire d'une fable de La Fontaine) finira par arriver. Plus précisément :

Théorème 2.20 *Soit $a = (a_1, \ldots, a_n) \in \{0,1\}^n$ une suite fixée de longueur n de succès et d'échecs, et soit l'événement*

$$A = \{\omega \in \Omega; \exists N \geq 0 \ \omega_{N+1} = a_1, \ldots, \omega_{N+n} = a_n\}.$$

Alors $\mathbb{P}(A) = 1$.

Preuve :

Soit k le nombre de S dans la suite a. Notons $A_N = \{\omega \in \Omega; \omega_{N+1} = a_1, \ldots, \omega_{N+n} = a_n\}$. Alors $\mathbb{P}(A_N) = p^k(1-p)^{n-k}$ par définition de $\mathbb{P}$. Introduisons $B_m = \cup_{j=0}^{m-1} A_{jn}$. Alors $B_m \subset B_{m+1}$ et $A = \cup_{N\geq 0}A_N \supset B = \cup_{m\geq 0}B_m$. On a de plus $\mathbb{P}(B_m^c) = \mathbb{P}(\cap_{j=0}^{m-1} A_{jn}^c) = (1-p^k(1-p)^{n-k})^m \to_{m\to\infty} 0$. Par continuité monotone, on a donc $\mathbb{P}(B^c) = 0$. D'où $1 = \mathbb{P}(B) \leq \mathbb{P}(A) = 1$. $\square$

Remarque 2.21 *Une conséquence fameuse de ce théorème est le paradoxe du singe savant : un singe tapant aléatoirement sur un clavier d'ordinateur, est sûr de reproduire l'intégralité des fables de La Fontaine.*

2.4 Exercices

Indépendance

Exercice 2.4.1 *Deux cartes sont tirées successivement d'un paquet de 52 cartes sans remise. Est-ce que les évènements « la première carte est un valet » et « la deuxième carte est un as » sont indépendants? Qu'en est-il si c'est un tirage avec remise?*

Exercice 2.4.2 *On jette deux dés. Est ce que les évènements « il y a au moins un 6 » et « la somme est 7 » sont indépendants?*

Exercice 2.4.3 *On jette deux dés non pipés, un dé moir et un dé blanc. On note A l'événement « le chiffre du dé noir est pair », et B l'événement « le chiffre du dé blanc est impair », C l'événement « les 2 chiffres ont même parité ». Montrer que A et C, A et B, B et C sont indépendants, mais que les trois événements ne le sont pas.*

Exercice 2.4.4 *Une personne achète un billet de lotterie par semaine. Chaque semaine la probabilité qu'il gagne est* $1/N$. *Montrer qu'il faut jouer à peu près* $2N/3$ *semaines pour que la probabilité de gagner au moins une fois soit de 0.5. (Approcher* $\ln 2$ *par 2/3 et* $\ln(1-1/N)$ *par* $-1/N$*).*

Exercice 2.4.5 *Dans une classe de 50 étudiants, quelle est la probabilité qu'au moins un étudiant soit né le 25 décembre? Combien doit-on avoir d'étudiants dans une classe pour qu'au moins l'un d'entre eux soit né le 25 décembre avec une probabilité supérieure à 0,5?*

Exercice 2.4.6 *Dans la situation la plus simple de transmission héréditaire, un « gène » se présente sous deux formes A et a, ce qui donne trois « génotypes » possibles: AA, Aa, aa. Chaque individu reçoit de ses parents un gène tiré au hasard dans la paire de chacun des deux. En supposant que les génotypes des parents sont indépendants et que leurs probabilités respectives sont p,q et r, trouver les probabilités P,Q et R qu'un enfant ait respectivement les génotypes AA, Aa et aa. En conclure qu'à partir de la deuxième génération ces probabilités sont constantes et vérifient* $Q^2 = 4PR$. *Ceci constitue la loi de Hardy-Weinberg.*

Probabilités conditionnelles

Exercice 2.4.7 *Une compagnie a deux usines A et B. L'usine A fabrique 80% des produits et la B les 20% restant. La proportion de produits avec défaut est de 0.05 pour A et de 0.01 pour B.*

1. *Quelle est la probabilité qu'un produit choisi au hasard soit sans défaut et provienne de A?*
2. *Quelle est la probabilité qu'un produit choisi au hasard ait un défaut?*

Exercice 2.4.8 *On a trois boites numérotées 1,2 et 3. La première contient 1 boule blanche et 2 boules noires, la deuxième 2 blanches et 1 noire, enfin la troisième 3 blanches. On choisit au hasard une des trois boites, puis on tire une boule de cette boite.Quelle est* *la probabilité d'avoir une boule blanche ? Sachant que la boule est blanche, quelle est la probabililté qu'elle provienne de la première boite ?*

Exercice 2.4.9 *Un étudiant de l'université va en cours avec une probabilité 0.8 quand il fait beau et 0.5 quand il pleut. Sachant qu'il pleut 20% des jours en janvier et que l'étudiant était en classe le 28 janvier, quelle est la probabilité qu'il pleuvait ce jour là? (On considèrera qu'il fait beau lorsqu'il ne pleut pas!)*

Exercice 2.4.10 *On estime que 10% de la population a une certaine maladie. Un test de détection existe mais n'est point parfait. La probabilité que le test soit positif pour une personne non atteinte par cette maladie est 0.05. La probabilité que le test soit négatif pour une personne qui en fait a la maladie est 0.01.*

Sachant que son test est positif, quelle est la probabilité pour une personne d'avoir cette maladie ?

Chapitre 3

Élément aléatoire

Dans le cas d'univers dénombrables, si une éventualité a une probabilité nulle, on peut l'ôter de l'univers (cf. la proposition 2.16) : ainsi, on peut toujours se ramener à un ensemble probabilisé pour lequel toutes les éventualités sont de probabilité non nulle. Par contre, dans le cas d'univers non dénombrables, ce n'est pas toujours le cas, d'où l'intérêt de la définition suivante :

Définition 3.1 *Pour* $(\Omega, \mathcal{A}, \mathbb{P})$, *un événement* $A \in \mathcal{A}$ *est dit* presque-impossible *ou* $\mathbb{P}$-négligeable *si* $\mathbb{P}(A) = 0$. *Son complémentaire est dit* presque-sûr.

Exemple :
Si on joue à pile ou face une infinité de fois, « obtenir uniquement des pile » est négligeable.

3.1 Élément aléatoire

3.1.1 Définition

Dans l'exemple du lancer de deux dés, avec $\Omega = \{1, \ldots, 6\}^2$, l'éventualité $\omega = (i, j)$ représente l'éventualité « le premier dé donne i et le deuxième j ». Pour un ω donné, le résultat du premier dé est donc la première coordonnée de ω. Soit $X : \Omega \longrightarrow \{1, \ldots, 6\}$ l'application définie par

$$X\Big((i, j)\Big) = i.$$

Le résultat du premier dé pour l'éventualité ω est alors donné par $X(\omega)$. L'ensemble A de toutes les éventualités pour lesquelles le premier dé a donné 2, c'est-à-dire l'événement « le premier dé a donné 2 », est :

$$A = \{\omega \in \Omega,\ X(\omega) = 2\} = \{X = 2\} = X^{-1}(\{2\}) = \{(2, 1), \ldots, (2, 6)\}.$$

Ainsi, si l'on ne s'intéresse qu'à un caractère de l'éventualité ω, cela revient à considérer une fonction $X(\omega)$, c'est-à-dire *une fonction du hasard.*

Définition 3.2 *Soit* T *un ensemble quelconque et* $\mathcal{T}$ *une tribu d'événements de* T. *Si*

$(\Omega, \mathcal{F}, \mathbb{P})$ *est un espace de probabilité, on appelle* élément aléatoire *défini sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *et à valeurs dans* T*, toute application* $X : \Omega \to T$ *mesurable, i.e. telle que pour tout* $B \in \mathcal{T}$*, la partie de* Ω*,* $X^{-1}(B) = \{\omega \in \Omega,\ X(\omega) \in B\}$*, est contenue dans* $\mathcal{F}$*.*

Pour $B \in \mathcal{T}$, l'événement $X^{-1}(B) = \{\omega \in \Omega,\ X(\omega) \in B\} \in \mathcal{F}$ pourra être noté $\{X \in B\}$. On définit ainsi simplement une nouvelle tribu constituée d'événements des événements $\{X^{-1}(B), B \in \mathcal{T}\}$ de Ωet appelée tribu engendrée par X. Cette sous-tribu de $\mathcal{F}$ est notée $\mathcal{F}_X$.

Exercice 3.1.1 *Montrer que* $\mathcal{F}_X$ *est bien une tribu et que c'est la plus petite sous-tribu de* $\mathcal{F}$ *qui rende* X *mesurable.*

Les ω sont souvent omis dans les notations probabilistes. Comme nous venons de la voir, $\{\omega \in \Omega,\ X(\omega) \in B\}$ sera noté $\{X \in B\}$.

Cas particuliers :
- Si $T = \mathbb{R}$ muni de la tribu des boréliens $\mathcal{B}(\mathbb{R})$ (cf. définition 2.3), on dit que X est une variable aléatoire réelle (v.a.r.).
- Si $T = \mathbb{R}^k$, on dit que X est un vecteur aléatoire, et dans ce cas, les composantes X_i de $X = (X_1, \dots, X_k)$ sont des v.a.r.
- De même, on peut définir une *fonction aléatoire* ou tout autre objet mathématique aléatoire (graphes, réseaux, formes, etc.).
- Si T est fini ou dénombrable, on dit que X est un élément aléatoire discret.

Exemples :
- Une fonction $f : \Omega \to \mathbb{R}$ constante $f \equiv a$ est un élément aléatoire, même si dans ce cas il n'y a plus d'aléatoire dans la valeur de f.
- Si $A \subset \Omega$ est un événement, alors la fonction indicatrice $\mathbf{1}_A : \mathcal{P}(\Omega) \to \{0, 1\}$ définie par
$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon} \end{cases}$$
est une variable aléatoire réelle.

3.1.2 Loi de probabilité d'un élément aléatoire

Définition 3.3 *Soit* $X : \Omega \longrightarrow T$ *un élément aléatoire de l'espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *à valeurs dans un espace* $(T, \mathcal{T})$*. On définit l'application* $\mathbb{P}_X$ *de* $\mathcal{T}$ *dans* $[0, 1]$ *par :*

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) \qquad \forall B \subset \mathcal{T}.$$

Alors $\mathbb{P}_X$ *est une probabilité sur* $\mathcal{T}$ *appelée* loi de probabilité *de* X*.*

Quand dans une expérience aléatoire on ne s'intéresse qu'à l'élément aléatoire X, cela revient à se placer dans le nouvel espace fondamental $(T, \mathcal{T}, \mathbb{P}_X)$. Dans bien des cas, on ne précisera pas l'espace Ω choisi.

Exemples :

- On lance deux dés. L'espace fondamental est $\Omega = \{\omega = (i,j) \ : \ 1 \leq i,j \leq 6\}$. Si on ne s'intéresse qu'à la somme $X(i,j) = X(\omega) = i + j$, on définit un élément aléatoire à valeurs dans $T = \{2, 3, \ldots, 12\}$ et $\mathbb{P}_X$ est définie par : $p_2 = 1/36$, $p_3 = 2/36, \ldots, p_{11} = 2/36$, $p_{12} = 1/36$.
- On effectue n expériences Succès-Échec de probabilité de succès p. On pose X le nombre de succès obtenus. Alors $X \in \{0, 1, 2, \ldots, n\}$ et $\forall k \in \{0, \ldots, n\}$,

$$\mathbb{P}(X = k) = \mathbb{P}_X(\{k\}) = \mathrm{C}_n^k p^k (1-p)^{n-k}.$$

Définition 3.4 *On dira que deux éléments aléatoires X et Y* ont même loi, *ou* sont identiquement distribués, *si leurs lois $\mathbb{P}_X$ et $\mathbb{P}_Y$ sont des probabilités identiques. Cette relation sera notée*

$$X \sim Y.$$

L'élément aléatoire X est entièrement déterminée par sa loi $\mathbb{P}_X$.

3.1.3 Variables indépendantes

Nous allons nous restreindre aux variables aléatoires, même si les définitions suivantes sont valables si T est différent de $\mathbb{R}^d$. Rappelons que si X est une variable aléatoire définie sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$, alors $\mathcal{F}_X$ est la plus petite sous-tribu de $\mathcal{F}$ qui rende X mesurable, i.e.

$$\mathcal{F}_X = \left\{ X^{-1}(A);\ A \in \mathcal{B}(\mathbb{R}^d) \right\}.$$

On note $\mathbb{P}_X$ la loi de X, i.e.

$$\forall A \in \mathcal{B}(\mathbb{R}^d),\ \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)).$$

Définition 3.5 (Variables indépendantes) *Des variables aléatoires $X_1, \ldots, X_n$ sont indépendantes si les tribus $\mathcal{F}_{X_i}$ sont indépendantes.*

Proposition 3.6 *Les v.a.r. X_i, $1 \leq i \leq n$ sont indépendantes si et seulement si pour tout $(A_1, \ldots, A_n) \in \mathcal{B}(\mathbb{R})^n$:*

$$\mathbb{P}_{(X_1,\ldots,X_n)}((A_1, \ldots, A_n)) = \mathbb{P}_{X_1}(A_1) \times \ldots \times \mathbb{P}_{X_n}(A_n).$$

3.1.4 Fonction de répartition

Ici nous allons nous restreindre au cas où les éléments aléatoires sont des variables réelles. Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle de l'espace de probabilité $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ et $\mathbb{P}_X$ sa loi de probabilité. Nous nous sommes ainsi rapporté à l'étude de l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$.

Définition 3.7 *On appelle* fonction de répartition *de X la fonction F_X définie sur $\mathbb{R}$, et à valeurs dans $[0, 1]$, par :*

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

On peut remarquer que la valeur de $\mathbb{P}_X$ sur tout intervalle $]a,b]$ s'obtient simplement à partir de la fonction de répartition :

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

La valeur de $\mathbb{P}_X$ sur n'importe quel sous-ensemble de $\mathbb{R}$ peut en fait être déduite de la fonction de répartition de X :

Lemme 3.8 *La fonction de répartition caractérise la loi : si deux variables aléatoires X et Y ont même fonction de répartition, alors elles ont même loi.*

De ce lemme et de la proposition 3.6 on en déduit :

Corollaire 3.9 *Les v.a. réelles X_i, $1 \leq i \leq n$, sont indépendantes si et seulement si pour tout $x_i \in \mathbb{R}$,*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \leq x_i\}\right) = \prod_{i=1}^{n} F_{X_i}(x_i);$$

F_{X_i} étant la fonction de répartition de X_i.

La fonction de répartition d'une variable aléatoire a par ailleurs les propriétés suivantes :

Propriétés 3.10 *Soit F_X la fonction de répartition d'une variable aléatoire réelle X. Alors :*

1. *F_X est croissante.*
2. *$\lim_{x\to-\infty} F_X(x) = 0$ et $\lim_{x\to+\infty} F_X(x) = 1$.*
3. *F_X est continue à droite.*
4. *F_X a une limite à gauche en tout point.*
5. *Si F_X admet une discontinuité en x_0, $F_X(x_0) - \lim_{x_0^-} F_X(\cdot) = \mathbb{P}(X = x_0)$.*

Preuve :

1. Si $x \leq y$, alors $X \leq x \Rightarrow X \leq y$ donc $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$.
2. En effet, comme F_X est croissante,

$$\begin{aligned} \lim_{x\to-\infty} F_X(x) &= \lim_{n\to+\infty} F_X(-n) = \lim_{n\to+\infty} \mathbb{P}_X(]-\infty,-n]) \\ &= \mathbb{P}_X\left(\bigcap_{n=1}^{+\infty}]-\infty,-n]\right) = \mathbb{P}_X(\emptyset) = 0. \end{aligned}$$

De même,

$$\begin{aligned} \lim_{x\to+\infty} F_X(x) &= \lim_{n\to+\infty} \mathbb{P}_X(]-\infty,n]) = \mathbb{P}_X\left(\bigcup_{n=1}^{n}]-\infty,n]\right) \\ &= \mathbb{P}_X(\mathbb{R}) = 1. \end{aligned}$$

3. Comme F_X est croissante,

$$\begin{aligned}\lim_{x^+} F_X &= \lim_{n\to+\infty} F_X(x+\frac{1}{n}) = \lim_{n\to+\infty} \mathbb{P}_X\left(\left]-\infty, x+\frac{1}{n}\right]\right) \\ &= \mathbb{P}_X\left(\bigcap_{n\geq 1}\left]-\infty, x+\frac{1}{n}\right]\right) = \mathbb{P}_X(]-\infty,x]) = F_X(x).\end{aligned}$$

4. Comme F_X est croissante et bornée, elle admet une limite à gauche en tout point.
5. On a :

$$\begin{aligned}F_X(x_0^-) &= \lim_{x\uparrow x_0} \mathbb{P}_X\left(]-\infty,x]\right) = \lim_{n\to+\infty} \mathbb{P}_X\left(\left]-\infty, x_0-\frac{1}{n}\right]\right) \\ &= \mathbb{P}_X\left(\bigcup_{n=1}^{n}\left]-\infty, x_0-\frac{1}{n}\right]\right) = \mathbb{P}_X(]-\infty,x_0[) \\ &= \mathbb{P}(X<x_0).\end{aligned}$$

Or, $\mathbb{P}(X=x_0) = \mathbb{P}(X\leq x_0) - \mathbb{P}(X<x_0)$.

□

3.2 Variables aléatoires discrètes

Une variable réelle X est dite discrète si elle ne peut prendre qu'un nombre fini ou dénombrable de valeurs réelles $\{x_i\}_{i\in I}$.

Propriétés 3.11 *Soit X une variable aléatoire discrète telle que $X(\Omega)=\{x_i\}_{i\in I}$. D'après la proposition 2.16, sa loi de probabilité $\mathbb{P}_X$ est définie par les probabilités $p_i=\mathbb{P}(X=x_i)$, qui vérifient $\sum_{i\in I} p_i = 1$. On peut écrire (la fonction indicatrice $\mathbf{1}_{\{X=x_i\}}$ est définie page 28)*

$$X(\omega)=\sum_{i\in I} x_i \mathbf{1}_{\{X=x_i\}}(\omega).$$

Sa fonction de répartition F_X est constante par paliers ; pour tout i elle a un saut en x_i de taille p_i (cf. propriété 3.10.5).

Autrement dit, la loi de X est caractérisé par la donnée des (x_i, p_i).

Proposition 3.12 *Si X et Y sont deux v.a.r. discrètes, X et Y sont indépendantes si, pour tout $i\in I$ et tout $j\in J$, on a :*

$$\mathbb{P}(\{X=x_i\}\cap\{Y=y_j\}) = \mathbb{P}(\{X=x_i\})\mathbb{P}(\{Y=y_j\}).$$

Exemple :
Soit X la variable aléatoire indiquant le résultat du lancé d'un dé équilibré. Sa fonction de répartition f saute de $1/6$ aux points $1,\ldots,6$ (voir figure 3.1).

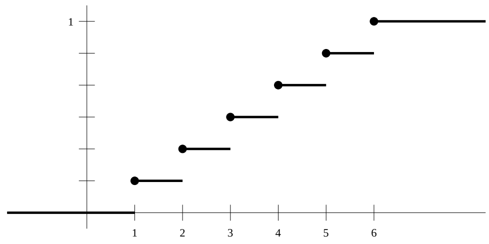


FIG. 3.1 – fonction de répartition de X

3.2.1 Moments

Définition 3.13 (Espérance) *L'espérance d'une v.a.r. discrète X de la loi $\mathbb{P}_X$ définie par (x_i, p_i) est définie par*

$$\mathbb{E}(X) = \sum_{i \in I} x_i p_i,$$

si le second membre converge absolument.

Propriétés 3.14 *Soient X et Y deux variables aléatoires. On a les propriétés :*

1. $\mathbb{E}(X)$ *est finie si et seulement si* $\mathbb{E}(|X|)$ *est finie;*
2. $|X| \leq Y$ *et* $\mathbb{E}(Y)$ *finie entraînent* $\mathbb{E}(X)$ *finie;*
3. $-\infty \leq a \leq X \leq b < +\infty \Longrightarrow a \leq \mathbb{E}(X) \leq b$*;*
4. $X = a$ *p.s. implique* $\mathbb{E}(X) = a$*;*
5. $\mathbb{E}(X)$ *finie implique* $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$*.*

Le théorème suivant consigne les propriétés usuelles de l'espérance mathématique

Théorème 3.15 *Soient X et Y deux variables aléatoires discrètes. Si $\mathbb{E}|X| < +\infty$ et $\mathbb{E}|Y| < +\infty$, alors on a les propriétés :*

A. *Linéarité*
   - *(A1.)* $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$*;*
   - *(A2.)* $\mathbb{E}(\lambda X) = \lambda \mathbb{E}X$ *pour tout* $\lambda \in \mathbb{R}$*.*

B. *Monotonie*
   - *(B1.)* $X \geq 0 \Rightarrow \mathbb{E}X \geq 0$*;*
   - *(B2.)* $X \geq Y \Rightarrow \mathbb{E}X \geq \mathbb{E}Y$*;*
   - *(B3.)* $X = Y$ *p.s.* $\Rightarrow \mathbb{E}X = \mathbb{E}Y$*.*

C. *Indépendance : si X et Y sont indépendantes, alors $\mathbb{E}(XY)$ est finie et l'on a*

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

Plus généralement pour toute fonction $g : \mathbb{R} \to \mathbb{R}$, on peut définir :

$$\mathbb{E}(g(X)) = \sum_{i \in I} g(x_i) p_i,$$

si le second membre converge absolument. L'espérance mathématique d'une variable aléatoire X ne dépend que de la loi de X et indique la *valeur moyenne* autour de laquelle

X prend ses valeurs. On introduit d'autres caractéristiques de la loi de X qui rendent compte de la *dispersion* de cette loi, par exemple les *moments*. Débutons par un lemme qui permet de comparer les moments de différents ordres.

Lemme 3.16 *Soient r et r' deux nombres réels tels que $0 < r < r'$ et X une variable aléatoire réelle. Si $\mathbb{E}|X|^{r'}$ est fini, $\mathbb{E}|X|^r$ est aussi fini.*

Définition 3.17 (Moments) *Soit X une v.a.r. discrète, de loi $\mathbb{P}_X$ donnée par (x_i, p_i). Soit a et r deux nombres réels. Si $\mathbb{E}|X-a|^r$ est fini, alors le* moment d'ordre r de X *centré en a est défini par :*

$$m_r^a = \mathbb{E}(X-a)^r = \sum_{i\in I} p_i(x_i - a)^r.$$

Le moment d'ordre r (centré en 0) est défini par :

$$m_r = \mathbb{E}(X^r).$$

De même, si $\mathbb{E}|X|$ est fini, ainsi que $\mathbb{E}\left(|X-\mathbb{E}X|^r\right)$, le moment centré d'ordre r (à la moyenne) est défini par :

$$\mu_r = \mathbb{E}\left[(X-\mathbb{E}X)^r\right].$$

Si $r=1$, alors $m_1 = \mathbb{E}X$ et $\mu_1 = 0$. Pour $r=2$, le moment centré μ_2 est encore appelé variance *de X et noté*

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

La variance est donc une quantité toujours positive. Sa racine carrée positive est appelée *écart-type* et notée $\sigma(X)$. Les variables aléatoires $X - \mathbb{E}X$ et $(X-\mathbb{E}X)/\sigma(X)$ sont respectivement *centrée* et *centrée réduite*.

Proposition 3.18 *Une v.a.r. X a un moment d'ordre deux $\mathbb{E}X^2$ fini si et seulement si son espérance mathématique $\mathbb{E}X$ et sa variance $Var(X)$ existent et sont finies, et*

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

Preuve :

En exercice. □

Corollaire 3.19 *Si X et Y sont deux v.a.r. indépendantes, alors*

$$Var(X+Y) = Var(X) + Var(Y).$$

3.2.2 Fonction génératrice

Ici on s'intéresse plus spécifiquement aux v.a.r. qui ne prennent que des valeurs entières positives ou nulles. Ainsi si X est une telle v.a., X est déterminée entièrement par la donnée de la suite $p_n = \mathbb{P}(X=n)$ pour $n \in \mathbb{N}$.

Définition 3.20 (Fonction génératrice) *Soit X une v.a.r. qui ne prend que des valeurs entières. Pour $s \in \mathbb{C}$, on définit la série entière*

$$\mathbb{E}(s^X) = \sum_{n=0}^{+\infty} p_n s^n.$$

La fonction $s \mapsto G_X(s) = \mathbb{E}(s^X)$ *est appelée* fonction génératrice *de* X.

Remarque 3.21 *Comme* $\sum_{n\in\mathbb{N}} p_n = 1$, *le rayon de convergence de la série entière* G_X *est au moins égal à 1.*

Théorème 3.22 *La fonction génératrice d'une v.a.* X *détermine entièrement la loi de cette variable. En d'autres termes, si deux v.a. (à valeurs entières positives) admettent la même fonction génératrice, alors elles ont la même loi.*

Preuve :
Vérifier que $G^{(n)}(0) = n!p_n$ pour tout $n \in \mathbb{N}$. □

Proposition 3.23 *La fonction génératrice* $G_X(s)$ *admet une dérivée à gauche* $G'_X(1)$ *en* $s = 1$, *si et seulement si* $\mathbb{E}(X)$ *existe et est finie, et l'on a :*

$$\mathbb{E}(X) = G'_X(1).$$

La preuve repose sur le lemme d'Abel

Lemme 3.24 1. *Si la série* $\sum_i \alpha_i$ $(i \geq 0)$ *est convergente de somme* α, *alors*

$$\lim_{s\to 1, s<1} \sum_{i=0}^{\infty} \alpha_i s^i = \sum_i \alpha_i = \alpha.$$

2. *Si les* α_i *sont positifs et si* $\lim_{s\to 1, s<1} \sum_{i=0}^{\infty} \alpha_i s^i = \alpha \leq +\infty$, *alors*

$$\sum_i \alpha_i = \alpha.$$

Proposition 3.25 *Si* X *et* Y *sont deux v.a. indépendantes, alors*

$$\forall |s| \leq 1, G_{X+Y}(s) = G_X(s)G_Y(s).$$

On caractérise ainsi la loi de la somme de X et Y.

3.2.3 Lois discrètes classiques

Schéma Succès-Échec simple, loi de Bernoulli

On reprend ici le schéma Succès-Échec du chapitre 1, pour $n = 1$. On réalise une fois une expérience dont les résultats possibles sont un succès noté 1, ou un échec noté 0. Le cas typique est le jeu *pile* ou *face*. L'espace des observables est

$$\Omega_1 = \{1; 0\}.$$

Si la pièce n'est pas truquée, pile et face ont la même chance de sortir. On associe la probabilité $\mathbb{P}_1$

$$\mathbb{P}_1(1) = \frac{1}{2}.$$

En général, on va supposer qu'une des faces est favorisée : il existe $p \in [0,1]$ tel que

$$\mathbb{P}_1(1) = p \text{ et } \mathbb{P}_1(0) = 1 - p.$$

Les cas $p = 0$ et $p = 1$ n'ont aucun intérêt et on suppose donc que $0 < p < 1$. On définit la variable aléatoire $X : \Omega_1 \to \mathbb{R}$, telle que $X(1) = 1$ et $X(0) = 0$.

Définition 3.26 (loi de Bernoulli) *X suit une* loi de Bernoulli de paramètre *p si :*

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

Proposition 3.27 *Si X suit une loi de Bernoulli de paramètre p, i.e. $X \sim Bern(p)$, alors $\mathbb{E}(X) = p$ et $Var(X) = p(1-p)$.*

Exercice 3.2.1 *Montrer que X a des moments de tout ordre et vérifier les résultats de la proposition précédente.*

Schéma Succès-Échec fini, loi binômiale

On reprend le schéma Succès-Échec où on réalise n fois la même expérience dont les résultats possibles sont un succès noté 1, ou un échec noté 0. X est la variable aléatoire qui compte le nombre de succès. Elle prend donc toutes les valeurs entières entre 0 et n, et suit comme loi

Définition 3.28 (loi binômiale) *Une v.a. X à valeurs dans $\{0, \dots, n\}$ suit une* loi binômiale de paramètres *(n, p), avec $n \geq 1$ et $p \in [0,1]$, si :*

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

On peut montrer que si $X_1, \dots, X_n$ sont n variables de Bernoulli de paramètre p, définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, alors la somme $S_n = \displaystyle\sum_{i=1}^{n} X_i$ suit une loi binômiale de paramètres n et p (utiliser par exemple la proposition 3.25, voir aussi l'exercice 3.4.15). La réciproque est également vraie :

Théorème 3.29 *B suit une loi binômiale de paramètres (n, p) si et seulement s'il existe n variables $(X_k)_{k=1,\dots,n}$ indépendantes et de Bernoulli de paramètre p telles que :*

$$B = \sum_{k=1}^{n} X_k.$$

Ceci permet de calculer très facilement l'espérance et la variance d'une loi binômiale, en utilisant les propriétés de l'espérance et de la variance de v.a. indépendantes.

Proposition 3.30 *Si X suit une loi binômiale de paramètres n et p, i.e. $X \sim Bin(n, p)$, alors $\mathbb{E}(X) = np$ et $Var(X) = np(1-p)$.*

Schéma Succès-Échec infini, loi géométrique

On est dans le cadre du schéma Succès-Échec infini du chapitre 2. On s'intéresse au nombre X de réalisations de la même expérience qu'il faudra répéter afin d'obtenir un premier succès. Si $p \in]0,1[$ est la probabilité d'obtenir un succès en une réalisation de l'expérience, la probabilité que $X = k$ est $p(1-p)^{k-1}$. Si p est égal à 0, on n'obtiendra jamais succès, et si p est égal à 1, on obtient toujours un succès dès la première réalisation. Ces deux cas n'ont donc plus rien d'aléatoire.

Définition 3.31 (loi géométrique) *X suit une* loi géométrique *de paramètre $p \in]0,1[$ dont la loi est donnée par :*

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

Pour calculer l'espérance et la variance d'une telle v.a., rappelons les formules suivantes :

Lemme 3.32 (rappel de quelques formules) *Soit q un réel tel que $0 < q < 1$.*

$$\sum_{i=0}^{n} q^i = \frac{1-q^{n+1}}{1-q}$$

$$\sum_{i=0}^{+\infty} q^i = \frac{1}{1-q}$$

$$\sum_{i=1}^{+\infty} i q^{i-1} = \frac{1}{(1-q)^2}$$

Ici la condition de convergence dans la définition de l'espérance prend tout son sens.

Proposition 3.33 *Si X suit une loi géométrique de paramètre $p \in]0,1[$, i.e. $X \sim Géo(p)$, alors $\mathbb{E}(X) = 1/p$ et $Var(X) = \frac{1-p}{p^2}$.*

Exercice 3.2.2 *Redémontrer le lemme précédent et vérifier les résultats sur l'espérance et la variance de X.*

Loi de Poisson

La principale utililté de la loi de Poisson se trouve dans les modélisations de files d'attente, de réseaux, de fiabilité de machines. C'est aussi une approximation de la loi binômiale quand n est grand.

Définition 3.34 *La* probabilité de Poisson *(S. Poisson, 1781-1840) sur $\mathbb{N}$ est définie par :*

$$\forall\, n \in \mathbb{N},\ p_n = \frac{e^{-\lambda}\lambda^n}{n!}.$$

où λ est un paramètre réel strictement positif. Une v.a.r. X suit une loi de Poisson de paramètre $\lambda > 0$, si X ne prend que des valeurs entières et

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

Exercice 3.2.3 *Vérifier que la probabilité définie précédemment satisfait tous les axiomes d'une probabilité.*

Proposition 3.35 *Si X suit une loi de Poisson, i.e. $X \sim Poisson(\lambda)$, alors $\mathbb{E}(X) = Var(X) = \lambda$.*

Théorème 3.36 (approximation de la loi binômiale) *Si n tend vers $+\infty$ et p_n est tel que np_n tend vers $\lambda \neq 0$, si $B_n \sim Bin(n, p_n)$, alors :*

$$\forall k \geq 0 \lim_{n \to +\infty} \mathbb{P}(B_n = k) = \mathbb{P}(P = k),$$

où P suit une loi de Poisson de paramètre λ.

On utilise ce théorème pour faire le calcul approché suivant : si B suit une loi binômiale de paramètres n et p, avec les conditions suivantes : $n > 30$ et $np \leq 5$, alors pour $\lambda = np$:

$$\mathbb{P}(B = k) \simeq \frac{\lambda^k}{k!} \exp(-\lambda).$$

Remarque 3.37 *Plus p est petit, meilleure est l'approximation.*

Plus généralement, pour B_i, pour $i = 1, \ldots, r$, des v.a. indépendantes, de loi binômiale de paramètres respectifs (n_i, p_i), si les p_i sont « assez petits », si on pose : $\lambda = n_1 p_1 + \ldots + n_r p_r$, et si P est une v.a. de loi de Poisson de paramètre λ, on a :

$$\forall k \geq 0, \ \mathbb{P}(B_1 + \ldots B_r = k) \simeq \mathbb{P}(P = k).$$

3.3 Variables aléatoires à densité

3.3.1 Probabilité absolument continue

Nous allons maintenant définir les probabilités absolument continues, qui constituent une classe importante de probabilités sur $\mathbb{R}$.

Définition 3.38 *Une probabilité $\mathbb{P}$ de $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ est dite* absolument continue *s'il existe une fonction $f : \mathbb{R} \to \mathbb{R}_+$ positive, telle que*

$$\mathbb{P}(A) = \int_A f(x)\, dx \qquad \forall A \subset \mathbb{R}. \tag{3.1}$$

La fonction f est alors appelée densité de probabilité *de $\mathbb{P}$, et on a :*

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1.$$

Lemme 3.39 *La relation (3.1) définit complètement la probabilité $\mathbb{P}$. On pourra ainsi parler de* la probabilité de densité *f, à partir du moment où f est une fonction positive dont l'intégrale sur $\mathbb{R}$ vaut 1.*

Exemple :
La probabilité uniforme sur $[a, b]$ est la probabilité absolument continue définie par une densité f constante sur $[a, b]$ (la constante de normalisation $1/(b-a)$ est là pour garantir que l'intégrale de f sur $\mathbb{R}$ vaut un) :

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x). \tag{3.2}$$

On remarquera que si on « tire au hasard » un nombre réel dans l'intervalle $[0, 1]$ (avec la loi uniforme $\mathbb{P}$), il n'est presque-sûrement pas rationnel : $\mathbb{P}([0, 1] \cap \mathbb{Q}) = 0$. En effet, comme $[0, 1] \cap \mathbb{Q}$ est dénombrable,

$$\begin{aligned}
\mathbb{P}([0,1] \cap \mathbb{Q}) &= \mathbb{P}\left(\bigcup_{x \in [0,1] \cap \mathbb{Q}} \{x\}\right) = \sum_{x \in [0,1] \cap \mathbb{Q}} \mathbb{P}(\{x\}) \\
&= \sum_x \int_x^x f(u)\, du = \sum_x 0 = 0.
\end{aligned}$$

Remarque 3.40 *En modifiant la densité d'une variable absolument continue en un point (ou en un nombre dénombrable de points), on ne change pas sa loi (regarder dans (3.1)). C'est pourquoi on dit que f est* une *densité de la probabilité $\mathbb{P}$, et non pas* la *densité de $\mathbb{P}$. Malgré tout, on a intérêt à ne pas modifier artificiellement la densité f, et la garder la plus régulière possible.*

Pourquoi f est-elle appelée la *densité de probabilité de* $\mathbb{P}$? C'est ce que nous explique le lemme suivant, où on reconnaît la définition des densités utilisées en physique : densité de masse, etc.

Lemme 3.41 *Soit $\mathbb{P}$ une probabilité absolument continue, de densité f. Alors, pour tout $x \in \mathbb{R}$ tel que f est continue en x,*

$$f(x) = \lim_{\varepsilon \to 0} \frac{\mathbb{P}([x-\varepsilon, x+\varepsilon])}{2\varepsilon}.$$

Preuve :
On peut encadrer l'intégrale de f sur l'intervalle $[x-\varepsilon, x+\varepsilon]$ à l'aide des bornes inférieures et supérieures de f sur cet intervalle :

$$2t \inf_{x-\varepsilon \leq t \leq x+\varepsilon} f(t) \;\leq\; \mathbb{P}([x-\varepsilon, x+\varepsilon]) = \int_{x-\varepsilon}^{x+\varepsilon} f(t)\, dt \;\leq\; 2t \sup_{x-\varepsilon \leq t \leq x+\varepsilon} f(t).$$

Il ne reste plus qu'à utiliser la continuité de f en x, qui entraîne :

$$\lim_{\varepsilon \to 0} \inf_{x-\varepsilon \leq t \leq x+\varepsilon} f(t) = \lim_{\varepsilon \to 0} \sup_{x-\varepsilon \leq t \leq x+\varepsilon} f(t) = f(x).$$

□

Le calcul des probabilités absolument continues de $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ revient donc à la donnée d'une densité $f : \mathbb{R} \to \mathbb{R}$. Dans ce contexte, l'analyse des fonctions réelles, que nous ne développerons pas ici, fournit de nombreux outils théoriques.

3.3.2 Variables aléatoires réelles absolument continues

Le vocabulaire relatif aux probabilités absolument continues sera aussi appliqué aux éléments aléatoires :

Définition 3.42 *Si X est une variable aléatoire (ou un vecteur aléatoire) réelle telle que $\mathbb{P}_X$ est une probabilité absolument continue de densité f_X, on dira que X est* absolument continue, *et a pour* densité *f_X.*

Dans le cas absolument continu, les propriétés suivantes sont des conséquences immédiates des définitions :

Propriétés 3.43 *La fonction de répartition F_X d'une variable aléatoire réelle absolument continue de densité f_X s'écrit :*

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du \qquad \forall x \in \mathbb{R}. \tag{3.3}$$

En conséquence, F_X est continue. Elle est aussi dérivable aux points de continuité de f_X avec $f_X(x) = dF_X(x)/dx$.

On rappelle que l'on peut définir la loi d'une variable aléatoire absolument continue par sa densité. Ainsi on dira que X suit une loi uniforme sur $[a, b]$ si elle admet pour densité la fonction f de (3.2).

Remarque 3.44 *Il existe des variables aléatoires réelles qui ne sont ni discrètes, ni absolument continues mais de loi de probabilité dite* mixte, *dont la fonction de répartition peut s'écrire :*

$$\forall\, x \in \mathbb{R}, \qquad F_X(x) = \alpha F_d(x) + \beta F_C(x),$$

où $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$ et où F_d est la fonction de répartition d'une variable discrète et F_c est celle d'une variable absolument continue.

Il existe aussi des variables aléatoires ni continues, ni discrètes, ni mixtes... mais nous n'en rencontrerons pas dans ce cours.

3.3.3 Moments

La loi d'une v.a. absolument continue X est donc caractérisée par la densité f_X.

Définition 3.45 (Espérance) *L'espérance d'une v.a. absolument continue X de loi $\mathbb{P}_X$ et de densité f_X est définie par*

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

si le second membre converge absolument.

De même pour g une fonction de $\mathbb{R}$ dans $\mathbb{R}$, on peut définir

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

Tous les résultats suivants, énoncés dans le cadre de variables discrètes, restent vrais dans le cadre de variables absolument continues : les propriétés 3.14, le théorème 3.15, le lemme 3.16, la définition 3.17, la proposition 3.18 et le corollaire 3.19.

3.3.4 Fonction caractéristique

Définition 3.46 *Soit X une v.a.r. On appelle* fonction caractéristique de X *(ou de la loi de X) la fonction de la variable réelle t définie par :*

$$\phi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_{\mathbb{R}} e^{itx} f_X(x) dx.$$

Si X est une variable discrète, cette définition est encore valable avec $\phi_X(t) = \sum_{k \in I} e^{itx_k} p_k$.

Théorème 3.47 *Désignons par ϕ la fonction caractéristique d'une v.a.r. X. Alors*

1. *ϕ est une fonction définie et continue pour tout $t \in \mathbb{R}$;*
2. *ϕ est bornée, et en fait, pour tout $t \in \mathbb{R}$, on a : $|\phi(t)| \leq \phi(0) = 1$;*
3. *pour tous a, b réels, on a, pour tout t réel, on a l'identité $\phi_{aX+b}(t) = e^{ibt}\phi_X(at)$.*

Théorème 3.48 *La fonction caractéristique d'une v.a.r. détermine la loi de cette variable. En d'autres termes, si deux v.a.r. admettent la même fonction caractéristique, elles ont même loi.*

Théorème 3.49 *Soit (X, Y) un couple de variables aléatoires indépendantes. Alors pour tout $t \in \mathbb{R}$:*

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

Proposition 3.50 *Soit X une v.a.r. de fonction caractéristique ϕ. Si $\mathbb{E}|X| < \infty$, alors ϕ est continument dérivable et $\phi'(0) = i\mathbb{E}(X)$.*

3.3.5 Lois absolument continues classiques

On a déjà évoqué la loi uniforme. Voici deux autres lois absolument continues classiques.

Loi exponentielle

Elle joue un rôle clé dans les processus de Markov et dans bon nombre de modèles.

Définition 3.51 *Une v.a.r. X suit une* loi exponentielle *de paramètre $\lambda > 0$ si X est à valeurs positives et si la densité de X est donnée par*

$$\forall x \in \mathbb{R}, \; f(x) = \lambda \exp(-\lambda x) \mathbf{1}_{\mathbb{R}_+}(x).$$

Propriétés 3.52 *Si X suit une loi exponentielle, i.e. $X \sim \mathcal{E}(\lambda)$, alors X admet des moments de tout ordre et $\mathbb{E}(X) = 1/\lambda$ et Var$(X) = 1/\lambda^2$.*

Preuve :

En exercice. □

Exemple d'application *Durée de bon fonctionnement d'un système*

On admet que cette durée est généralement imprévisible. On la représente donc par une variable aléatoire positive X. Comme nous l'avons déjà remarqué, il n'est pas nécessaire de décrire l'espace fondamental sur lequel elle peut être définie.

Définition 3.53 *On appelle* taux de défaillance *du système la fonction du temps définie par :*

$$\forall\, t \geq 0, \qquad \lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}\left[t < X \leq t + \Delta t \mid X > t\right].$$

Le taux de défaillance au temps t est la densité de probabilité pour le système de tomber en panne juste après le temps t, sachant qu'il est encore en marche au temps t.

Quelle est la loi de probabilité de X, si l'on suppose que cette fonction est constante ($\forall\, t \geq 0,\ \lambda(t) = \lambda > 0$; ce qui signifie que le système est sans usure) ?

En fait,

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta t \to 0} \frac{\mathbb{P} t < X \leq t + \Delta t \text{ et } X > t}{\Delta t . \mathbb{P} X > t} \\
&= \lim_{\Delta t \to 0} \frac{\mathbb{P} t < X \leq t + \Delta t}{\Delta t . \mathbb{P} X > t} \\
&= \lim_{\Delta t \to 0} \frac{F_X(t + \Delta t) - F_X(t)}{\Delta t (1 - F_X(t))}
\end{aligned}$$

Si on suppose que X admet une densité, alors :

$$\lambda(t) = \frac{f_X(t)}{1 - F_X(t)}.$$

La question posée revient à résoudre l'équation différentielle :

$$F'_X(x) = \lambda\left(1 - F_X(x)\right),$$

avec la condition initiale $F_X(0) = 0$ (car X est à valeurs positives), dont la solution est : $F_X(x) = 1 - e^{-\lambda x}$, $\forall\, x \geq 0$. D'où la densité de X :

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x).$$

Loi normale (ou de Gauss)

On dit que X est une variable aléatoire gaussienne (K. Gauss, 1777-1855) de moyenne $m \in \mathbb{R}$ et d'écart-type $\sigma > 0$ si elle est absolument continue et a pour densité

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

On note $\mathcal{N}(m, \sigma^2)$ cette loi. Pour cette loi, la fonction de répartition F_X n'est pas calculable de manière explicite. Mais elle joue un rôle central en probabilité (voir chapitre suivant et le théorème central limite).

Propriétés 3.54 *Si X suit une loi normale de paramètres $m = 0$ et $\sigma^2 = 1$ (loi normale centrée réduite), alors*

1. $\mathbb{E}(X) = 0,\ Var(X) = 1$;
2. $\phi_X(t) = \phi_{0,1}(t) = \exp\left(-\frac{t^2}{2}\right)$;

Proposition 3.55 *Si $X \sim \mathcal{N}(0, 1)$, alors $Y = \sigma X + m \sim \mathcal{N}(m, \sigma^2)$.*

Comme F_Y n'est pas calculable explicitement, on utilisera la proposition précédente pour se ramener au cas $\mathcal{N}(0, 1)$ et on utilisera les tableaux donnés dans l'annexe B.

3.4 Exercices

Variables aléatoires discrètes

Exercice 3.4.1 *Quelle est l'espérance d'une variable aléatoire X uniformément distribuée sur $\{-1, 0, 3\}$? Et la variance ?*

Exercice 3.4.2 *On lance un dé. Vous me devez b euro si le dé tombe sur 5 ou 6. Je vous donne a euro sinon. Les probabilités de gagner ne sont clairement pas les mêmes. Néanmoins peut-on choisir b pour rendre le jeu équitable ?*

Exercice 3.4.3 *On lance trois pièces équilibrées. Soit X le nombre de « face » obtenu.*

1. *Trouver la distribution de X.*
2. *Calculer la probabilité $\mathbb{P}(X \geq 2)$.*
3. *Calculer l'espérance et la variance de X.*
4. *On effectue une série de lancers de trois pièces. Déterminer la probabilité d'avoir besoin d'au moins cinq lancers pour obtenir trois « pile ».*

Exercice 3.4.4 *On joue à pile ou face sept fois de suite. Soit B le nombre de « face ».*

1. *Quelle est la probabilité d'obtenir au moins une fois « face » ?*
2. *Quelle est la probabilité d'obtenir au moins trois fois « face » ?*
3. *Quelle est l'espérance de B ?*
4. *Quelle est la variance de B ?*

Exercice 3.4.5 *Soit X une variable aléatoire de loi géométrique.*

1. *Quelle est la probabilité que $X > t$ pour $t > 0$?*
2. *Sachant que $X > t$, quelle est la probabilité que $X > t + s$, pour s et t positifs ?*

Exercice 3.4.6 *Dans une usine, une pièce fabriquée a la probabilité 0.9 d'être utilisable. Soit X, le nombre minimum de pièces à fabriquer pour obtenir une pièce utilisable.*

1. *Déterminer la loi de probabilité de X.*

2. *Quel nombre minimum de pièces faut-il fabriquer pour avoir une pièce utilisable avec une probabilité supérieure ou égale à* 0.99 *?*

Exercice 3.4.7 *Deux joueurs jouent au dé. Le joueur A parie sur le* 6 *et le joueur B sur le* 1*. Le dé est jeté jusqu'à ce qu'un des deux joueurs gagne.*

1. *Quelle est la probabilité que le joueur A gagne?*
2. *Quelle est la probabilité que le joueur B gagne?*
3. *Soit T la variable aléatoire représentant le nombre de fois que le dé est jeté. Déterminer la loi de probabilité de T.*

Exercice 3.4.8 *On joue avec deux dés équilibrés; le premier dé est un tétraèdre régulier, le second un dé cubique. On associe à un jet les variables : X est la valeur du dé tétraédrique, Y celle du dé cubique, $S = X + Y$.*

1. *Calculer l'espérance et la variance de X et Y.*
2. *En déduire l'espérance et la variance de S (justifiez).*
3. *(a) Déterminer la loi de S.*
   (b) Calculer $\mathbb{P}(S > 7)$.
   (c) Calculer la probabilité que S prenne des valeurs paires.

Exercice 3.4.9 *Supposons que trois personnes entrent dans un ascenseur qui parcourt 5 étages. Quelle est l'espérance du nombre d'arrêts S que va devoir effectuer l'ascenseur? On suppose que les personnes choisissent leur arrêt indépendamment les unes des autres.*

Exercice 3.4.10 *Soit X le nombre d'anniversaire distincts dans une classe de 50 éléves. Quelle l'espérance de B ? Commenter, en se rappelant que la probabilité d'avoir au moins deux étudiants nés le même jour est de 0,96.*

Exercice 3.4.11 *Soit la suite $(\alpha_k)_{k\in\mathbb{N}}$ définie par*

$$\alpha_k = \frac{e^{-2}}{4} \times \frac{2^k}{k!} \times (1 + ak)\,,$$

avec $a \in \mathbb{R}$. On veut définir une variable aléatoire X sur $\mathbb{N}$ par $\mathbb{P}(X = k) = \alpha_k$ pour tout $k \in \mathbb{N}$.

a. *Trouver pour quelle valeur unique de a la loi de X est une loi de probabilité.*
b. *Dans ce cas, calculer son espérance et sa fonction génératice.*

Exercice 3.4.12 (Loi sans espérance) *Un exemple de v.a. discrète pour laquelle l'espérance n'est pas définie.*

1. *Utiliser le résultat suivant :*
$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}\ ,$$
   pour trouver c tel que $\mathbb{P}(X = k) = c/k^2$ soit une loi de v.a.
2. *Montrer que l'espérance de cette v.a. définie ainsi n'existe pas.*

Exercice 3.4.13 (Binomiale négative) *1. On jette un dé équilibré. Quelle est la probabilité que le second « six » apparaisse au bout de 10 lancers ?*

2. *Généralisation : dans un jeu de pile ou face, la probabilité d'avoir « pile » est p, $0 < p < 1$, celle d'avoir « face » est $1-p$. Soit S_r le nombre de jets nécessaires pour obtenir r fois « pile ». Calculer la loi de S_r, l'espérance et la variance.*
3. *Applications :*
   (a) *quelle est la probabilité que le cinquième enfant d'un couple soit leur seconde fille ?*
   (b) *On jette un dé équilibré. En moyenne, combien de jets sont nécessaires pour obtenir trois « six » ?*

Exercice 3.4.14 *Pierre a dans la main une pièce de monnaie équilibrée (probabilité $1/2$ de tomber sur face et probabilité $1/2$ de tomber sur pile). A chaque instant, il lance sa pièce et regarde le résultat.*

a. *Déterminer la loi du premier instant où Pierre obtient pile.*

Pierre et Paul ont maintenant chacun dans la main une pièce de monnaie équilibrée. Ils jouent au jeu suivant : à chaque instant, chacun lance sa pièce. Tant que les deux obtiennent face ils rejouent. Ils s'arrêtent donc lorsqu'au moins un des deux obtient pile. Alors celui qui a obtenu pile gagne et si les deux obtiennent pile en même temps, il y a égalité et le jeu s'arrête.

b. *Calculer la probabilité qu'il y ait égalité.*

c. *Calculer la probabilité que Pierre gagne.*

Exercice 3.4.15 *Calculer la fonction génératrice d'une loi de Bernoulli de paramètre p. En déduire la fonction génératrice de la loi $B(n,p)$, puis redémontrer que la somme de deux v.a. indépendantes de lois $B(n,p)$ et $B(m,p)$ suit la loi $B(n+m,p)$.*

Exercice 3.4.16 *Soient X et Y deux variables aléatoires indépendantes et à valeurs dans l'ensemble $\{1,\dots,6\}$. Montrer que, quelles que soient les lois de X et Y, la loi de $X+Y$ ne peut pas être la loi uniforme sur $\{2,\dots,12\}$. Cela montre qu'on ne peut pas truquer deux dés à six faces de sorte que la somme soit équirépartie sur $\{2,\dots,12\}$.*

Exercice 3.4.17 *Soient X et Y deux v.a. à valeurs dans $\mathbb{N}$. On définit la fonction génératrice du couple (X,Y) sur l'ensemble $[0,1]^2$ par*

$$g_{X,Y}(u,v) = \mathbb{E}(u^X v^Y) = \sum_{i,j\in\mathbb{N}} u^i v^j \mathbb{P}(X=i, Y=j).$$

1. *Montrer que $g_{X,Y}$ détermine $\mathbb{P}(X=i,Y=j)$ pour tout choix de i et j dans $\mathbb{N}$.*
2. *En déduire que si X et Y sont indépendantes si et seulement si*

$$g_{X,Y}(u,v) = g_X(u) g_Y(v) \quad \forall\ (u,v) \in [0,1]^2.$$

Exercice 3.4.18 *Soit X et Y deux variables aléatoires de Poisson de paramètre respectif λ et μ indépendantes.*

1. *Calculer $\mathbb{P}(X+Y=n)$, pour $n \in \mathbb{N}$. On pourra décomposer :*

$$\{X+Y=n\} = \bigcup_{k=0}^{n} \{X=k\} \cap \{Y=n-k\}.$$

2. *Quelle est la loi de $X+Y$?*

Exercice 3.4.19 *Calculer la fonction caractéristique d'une v.a. X qui suit une loi de Poisson de paramètre $\lambda > 0$. En déduire la loi d'une somme de n variables aléatoires indépendantes qui suivent des lois de Poisson de paramètres $\lambda_1, \ldots, \lambda_n$.*

Lois continues

Exercice 3.4.20 *Soit X une variable aléatoire de loi uniforme sur $[-2,4]$.*

1. *Déterminer la densité f de X. Représenter graphiquement f.*
2. *alculer $\mathbb{P}(-3 \leq X \leq 0)$, $\mathbb{P}(-2 \leq X \leq 0)$, $\mathbb{P}(0 \leq X \leq 2)$ et $\mathbb{P}(0 \leq X \leq 3)$.*

Exercice 3.4.21 *Soit X une v.a. ayant la densité f définie sur $[0,2]$ de la façon suivante :*

$$\begin{cases} f(x) &= x \quad \text{sur } [0,1], \\ f(x) &= 2-x \quad \text{sur } [1,2], \end{cases}$$

et f est nulle ailleurs.

1. *Vérifier que f est bien une densité et la représenter.*
2. *Calculer la fonction de répartition de X.*
3. *Calculer l'espérance et la variance de X.*

Exercice 3.4.22 *Soit X une variable aléatoire de densité f définie par $f(x) = 3x^2$ sur le segment $[0,1]$ et 0 ailleurs.*

1. *Calculer l'espérance de X.*
2. *Quelle est la valeur médiane de X ?*
3. *Calculer $Var(X)$. Quelle est la variance de la variable aléatoire $Y = -3X+1$?*

Exercice 3.4.23 *Si la variable X a pour densité f, quelle est la densité de $2X-1$? (penser à relier la fonction de répartition de X et celle de $2X-1$)*

Exercice 3.4.24 *Calculer les fonctions de répartition d'une loi de Bernoulli de paramètre p, de la loi uniforme sur (a,b) et de la loi exponentielle de paramètre λ, et les représenter.*

Exercice 3.4.25 *Soit X une v.a.r. et soit F_X sa fonction de répartition. Montrer que l'ensemble D des points de discontinuité de F_X est au plus dénombrable.*

Exercice 3.4.26 *On dit qu'une v.a.r. X est symétrique si X et $-X$ ont la même loi. Montrer que X est symétrique si et seulement si $\varphi_X(t) \in \mathbb{R}$ pour tout t.*

Exercice 3.4.27 (Médiane) *Soit X une variable aléatoire telle que $\mathbb{E}|X| < +\infty$. Si M est un réel telle que*

$$\mathbb{P}(X \leq M) \geq \frac{1}{2}, \quad \mathbb{P}(X \geq M) \geq \frac{1}{2},$$

montrer que pour tout nonbre réel a, on a l'inégalité :

$$\mathbb{E}\left[|X-a|\right] \geq \mathbb{E}\left[|X-M|\right].$$

Le nombre M est appelée une médiane de X. Donner un exemple de v.a.r. pour laquelle M n'est pas unique, puis un exemple où M est unique.

Exercice 3.4.28 *Soit X une variable aléatoire telle que $\mathbb{E}(X^2) < +\infty$. Montrer que pour tout nombre réel a, on a l'inégalité :*

$$\mathbb{E}\left[(X-a)^2\right] \geq \mathbb{E}\left[(X-\mathbb{E}X)^2\right] = \sigma^2.$$

Exercice 3.4.29 (Inégalité de Tchebychev) *Soit $r > 0$ un nombre réel et X une variable aléatoire réelle. Montrer que si $\mathbb{E}|X|^r = m_r < +\infty$, alors*

$$\mathbb{P}(|X| > t) \leq \frac{m_r}{t^r}.$$

Exercice 3.4.30 *Soit X_n une suite de variables aléatoires indépendantes et de même loi et soit N une v.a. de loi de Poisson $P(\lambda)$ indépendante des X_n. Montrer que la fonction caractéristique de $\sum_{i=1}^{N} X_i$ (valant 0 si $N = 0$) est : $\psi(t) = e^{\lambda(\varphi(t)-1)}$ où φ est la fonction caractéristique des X_i.*

Exercice 3.4.31 *Soit $X_1, \ldots, X_n$ des v.a.r. indépendantes et de même loi. On pose :*

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

et

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Montrer que si les variables X_i admettent des moments d'ordre 2, alors

$$E(S^2) = \frac{n-1}{n} Var(X_1).$$

Exercice 3.4.32 *Soient X et Y des variables aléatoires indépendantes de loi $\exp(\lambda)$. Calculer la densité du vecteur $(\min(X,Y), \max(X,Y))$.*

Exercice 3.4.33 *Soient $X_1, \ldots, X_n$ des v.a. indépendantes de lois exponentielles de paramètres $\alpha_1, \ldots, \alpha_n$ respectivement.*

1. *Montrer que la loi de $Y = \min(X_1, \ldots, X_n)$ est elle aussi exponentielle. Quel est son paramètre ?*

2. *Montrer que pour* $k = 1, \ldots, n$

$$P(X_k = \min(X_1, \ldots, X_n)) = \frac{\alpha_k}{\alpha_1 + \cdots + \alpha_n}.$$

Exercice 3.4.34 *Soit* X *une variable aléatoire de loi exponentielle de paramètre* $\lambda > 0$. *Soit* Y *une variable aléatoire de loi*

$$P(Y = 1) = \frac{1}{2} = P(Y = -1)\,.$$

On suppose que X *et* Y *sont indépendantes et on note* $Z = XY$.

a. *Calculer*
$$P(Z \geq 1 \,|\, Y = 1)\,.$$
b. *Calculer la fonction de répartition de* Z, *puis sa densité.*
c. *Calculer la fonction de répartition de* Z *dans le cas où la loi de* Y *est*
$$P(Y = 1) = \frac{1}{2} = P(Y = 0)\,.$$
La variable Z *admet-elle une densité ?*

Exercice 3.4.35 *Soit* T *la variable aléatoire représentant le temps d'attente d'un bus. On suppose que* T *est de loi exponentielle de paramètre* 3 *(3 étant le taux par heure), de densité*

$$f(x) = 3e^{-3x} \text{ pour } x \geq 0 \text{ et } 0 \text{ sinon.}$$

1. *Quelle est la probabilité d'attendre le bus au moins 20 minutes?*
2. *Sachant que nous avons déjà attendu le bus 20 minutes, quelle est la probabilité d'attendre 20 minutes supplémentaires avant que le bus n'arrive?*
3. *Sous quelles conditions la loi exponentielle est appropriée à ce problème ?*

Exercice 3.4.36 *Le nombre de demande d'accès à un site internet pendant une certaine heure est modélisé par une variable de Poisson, disons* X, *de paramètre* T *qui lui-même est une variable exponentielle de paramètre* λ. *Montrer que* X *suit une loi géométrique dont on donnera le paramètre. Quelle est l'espérance de* X *?*

Exercice 3.4.37 *Soit* X *une variable aléatoire de loi normale* $\mathcal{N}(0, 1)$.

1. *Quelle est la probabilité pour que* X *prenne la valeur* 1 *?*
2. *Calculer les probabilités des événements suivants à l'aide de la table de la loi normale centrée réduite* $\mathcal{N}(0, 1)$:
$$-0.7 < X < 2.4 \;;\; X > 1.5 \;;\; X \leq 0.85 \;;\; X > -0.84 \;;\; X < -0.06 \;;\; -0.51 < X \leq 0.62\;.$$
3. *Déterminer les réels* a, b *et* c *tels que*
$$P(|X| > a) = 0.73 \;;\; P(|X| < b) = 0.75 \;;\; P(X < c) = 0.80.$$

Exercice 3.4.38 *Montrer que si* X *suit la loi* $N(0, 1)$ *alors* $aX + b$ *suit la loi* $N(b, a^2)$. *Utiliser ceci pour calculer l'espérance et la variance d'une v.a. de loi* $N(\mu, \sigma^2)$.

Exercice 3.4.39 *Déduire de l'expression de la fonction caractéristique d'une variable de loi $N(0,1)$ celle d'une variable de loi $N(\mu,\sigma^2)$ où $\sigma^2 > 0$. En utilisant ce résultat, montrer que la somme de v.a. indépendantes de loi normale est elle aussi une v.a. de loi normale.*

Exercice 3.4.40 *Pour $\alpha > 0$ on pose*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

1. *Montrer que pour tout $\beta, \alpha > 0$ la fonction*

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{[0,\infty[}(x),$$

   est une densité. La loi qui lui est associée sera notée : $\Gamma(\alpha, \beta)$.
2. *Montrer par récurrence sur n que la loi χ^2_n coincide avec la loi $\Gamma(n/2, 1/2)$.*

Exercice 3.4.41 *Grâce aux courbes de croissance dans les carnets de santé, on estime que la moitié des filles de 16 ans mesurent plus de 161 cm, et que trois quarts mesurent moins de 165 cm. On suppose que la taille peut être modélisée par une variable aléatoire gaussienne de loi $\mathcal{N}(\mu, \sigma)$.*

1. *Quelle valeur doit-on prendre pour μ ?*
2. *Quelle valeur doit-on prendre pour σ ?*
3. *Quelle est la proportion de filles de 16 ans mesurant plus de 170 cm ?*

Exercice 3.4.42 *Un producteur prépare des lots de fruits pour la grande distribution. Les lots sont de trois types : ils peuvent contenir 4 pommes, ou bien 2 pommes et 2 poires, ou bien 4 poires. On admet que le poids d'une pomme est en moyenne de 250 gr, avec un écart-type de 20 gr. Le poids moyenne d'une poire est de 230 gr, avec un écart-type de 10 gr. On modélisera le poids d'une pomme et d'une poire par une v.a. gaussienne.*

1. *Quelle est la probabilité qu'un lot de 4 pommes pèse plus que 1kg 100gr ?*
2. *Quelle est la probabilité qu'un lot de 2 pommes et de 2 poires pèse moins que 1kg ?*
3. *Quelle est la probabilité qu'un lot de 4 pommes pèse plus qu'un lot de 4 poires ?*

Exercice 3.4.43 *Une usine fabrique des tiges d'acier dont la longueur est proche de 75 cm.*

1. *Selon le responsable de cette usine, environ 95% des tiges ont une longueur comprise entre 73 et 77 cm. Proposer une modélisation probabiliste de la longueur des tiges produites dans cette usine.*
2. *En réalité les tiges dont la longueur diffère de plus de 2,5 cm de 75 cm ne sont pas soumises à la vente. Déterminer la probabilité qu'une tige soit retirée de la vente.*
3. *Le contrôle de la production n'est pas parfait. En fait on estime qu'une tige a une probabilité de 0,97 d'être acceptée si elle est bonne, et une probabilité de 0,99 d'être rejetée si elle est défectueuse.*
   (a) *Calculer la probabilité qu'il y ait une erreur dans le contrôle de la production.*

(b) *En déduire la probabilité qu'une pièce ayant subi une erreur de contrôle soit bonne.*

Exercice 3.4.44 (Loi de Cauchy) *Une v.a. X suit une loi de Cauchy $C(0,1)$ si elle est absolument continue et admet pour densité*

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

1. *Vérifier que f est une densité.*
2. *Calculer l'espérance de X.*
3. *Quelle est la loi de $Y = \alpha X + \beta$? On dira que Y suit une loi de Cauchy de paramètres α et β.*
4. *Calculer la fonction caractéristique de Y.*
5. *Soit V une v.a. uniforme sur $]-\pi/2, \pi/2[$. Quelle est la loi de $\tan(V)$?*

Exercice 3.4.45 (Loi log-normale) *En linguistique on a trouvé que le nombre de mots par phrase (c'est-à-dire la longueur de la phrase mesurée en nombre de mots) suit approximativement une loi dite « Log-normale ». On dit qu'une variable aléatoire X à valeurs dans $]0,+\infty[$ suit une loi Log-normale de paramètres (μ,σ) avec μ réel et $\sigma > 0$, si $Y = \ln X$ suit la loi $\mathcal{N}(\mu,\sigma)$.*

1. *Déterminer la fonction de répartition de X.*
2. *Déterminer la densité de X.*
3. *Calculer $\mathbb{E}(X)$.*

Problème 3.4.1 (Loi de Pareto) *Soit Y une v.a. de loi exponentielle de paramètre λ. La variable aléatoire $X = e^Y$ est appelée variable aléatoire de Pareto, de loi de Pareto $\mathcal{P}(\lambda, 1)$.*

1. *On pose $r(x) = \mathbb{P}(X \geq x)$ (c'est la fonction de survie de X). Montrer que r est donnée par :*
$$r(x) = \begin{cases} x^{-\lambda}, & \text{si } x \geq 1; \\ 1, & \text{si } x < 1. \end{cases}$$
et la densité par :
$$f(x) = \begin{cases} \lambda/x^{\lambda+1}, & \text{si } x \geq 1; \\ 0, & \text{si } x < 1. \end{cases}$$
2. *Calculer $\mathbb{E}(X)$ (attention, il y a deux cas à distinguer).*
3. *Pour tout entier $k \geq 1$ évaluer $\mathbb{E}(X^k)$ en déterminant la fonction de survie de X^k.*
4. *On appelle fonction génératrice de X la fonction suivante : $g(u) = \mathbb{E}(e^{uX})$ pour tout $u \in \mathbb{R}$. Montrer que pour tout $u \in \mathbb{R}$:*
$$g(u) = \sum_{k=0}^{+\infty} \mathbb{E}(X^k)\frac{u^k}{k!}.$$
5. *En déduire que $g(u) < +\infty$ si et seulement si $u = 0$.*

Dans tout ce problème on pourra utiliser le résultat suivant : si X est une variable aléatoire continue,

$$\mathbb{E}(X) = \int_{\mathbb{R}^+} \mathbb{P}(X \geq x) dx.$$

Chapitre 4

Deux théorèmes limites et leurs applications

4.1 Loi des grands nombres

Considérons de nouveau le cas d'une expérience à deux issues, échec (noté 0) et succès (noté 1). On se fixe un nombre $0 < p < 1$, la probabilité d'obtenir un succès après une réalisation de l'expérience. On va modéliser le cas de cette expérience qui se répète infiniment dans des conditions identiques et indépendantes. On vérifie qu'on est bien dans le cas du schéma Succès-Échec infini décrit dans la section 2.3.1, où :

$$\Omega = \{\omega_i = (\eta_j,\ j \in \mathbb{N}) \ : \ \eta_j \in \{0,1\}\}.$$

On veut vérifier que p correspond à notre intuition d'être la fréquence asymptotique de réalisations des *succès* :

$$\lim_{n\to\infty} \frac{|\{i \leq n \ : \ \eta_i = \textit{succès}\}|}{n} = p. \tag{4.1}$$

On note $X_1, \ldots, X_n$ nos variables de Bernoulli modélisant chacune une réalisation de l'expérience (i.e. $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$ pour tout $0 \leq i \leq n$), et $S_n = X_1 + \ldots + X_n$. Le terme de gauche de l'équation précédente (4.1) correspond alors à une réalisation de la variable aléatoire S_n/n.

Remarque 4.1 Attention *La somme S_n est complètement décrite par les n premières réalisations $X_1, \ldots, X_n$, qui forment elles-même un schéma succès-échec fini, de longueur n. Pour tout n, ces modèles sont bien inclus dans notre modèle, le schéma succès-échec infini. Plutôt que de considérer des schéma fini de longueur croissantes, nous étudions seulement le schéma infini les englobant tous.*

Une autre façon de se poser cette question est de se demander s'il y a beaucoup de réalisations où $S_n/n \neq p$. On reformule alors la question

$$\forall \epsilon > 0,\ \lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \ : \left|\frac{S_n(\omega)}{n} - p\right| > \epsilon\right\}\right) = 0.$$

La réponse affirmative est la célèbre loi (faible) des grands nombres, qui relie les notions de fréquence et de probabilité :

Théorème 4.2 (Bernoulli, 1685)

$$\forall \epsilon > 0,\ \lim_{n\to\infty} \mathbb{P}\left(\left\{\omega\ :\ \left|\frac{S_n(\omega)}{n} - p\right| > \epsilon\right\}\right) = 0.$$

Ce premier résultat obtenu pour une suite de v.a. de Bernoulli a été généralisé à toute suite de v.a. $(X_n)_{n\in\mathbb{N}^*}$ indépendantes et identiquement distribuées (en abrégé i.i.d.), c'est-à-dire toutes les variables X_n ont la même loi pour tout $n \in \mathbb{N}^*$ sur un même espace Ω et sont indépendantes entre elles. La première difficulté est de prouver que de telles suites existent, difficulté résolue par Kolmogorov aux alentours de 1930.

Théorème 4.3 *Si $\mathbb{P}_0$ est une probabilité sur $\mathbb{R}^d$, il existe un espace $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel on peut définir une suite $(X_n)_n$ de v.a. à valeurs dans $\mathbb{R}^d$, indépendantes et identiquement distribuée, telles que $\mathbb{P}_{X_n} = \mathbb{P}_0$, pour tout n.*

On peut alors étendre la loi (faible) des grands nombres à la suite $(X_n)_{n\in\mathbb{N}^*}$, où on ajoute l'hypothèse que $\mathbb{E}(X_1^2) < \infty$, et à la v.a. « moyenne des (X_i) » :

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Par linéarité de l'espérance, on a immédiatement pour tout $n \geq 1$,

$$\mathbb{E}\left(\overline{X}_n\right) = \mu, \quad \mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma^2}{n}.$$

On en déduit alors le théorème suivant :

Théorème 4.4 (Loi (faible) des grands nombres)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. i.i.d. de moyenne μ et de variance σ^2. Alors pour tout $\varepsilon > 0$, on a

$$\lim_{n\to+\infty} \mathbb{P}\left(|\overline{X}_n - \mu| \geq \varepsilon\right) = 0.$$

On peut même préciser la vitesse de convergence :

$$\mathbb{P}\left(|\overline{X}_n - \mu| \geq b\right) \leq \frac{\sigma^2}{nb^2}.$$

Preuve :

C'est une conséquence immédiate de l'inégalité de Tchebychev à l'ordre 2 (voir exercice 3.4.29) appliquée à la variable $\overline{X}_n - \mu$. □

4.2 Théorème central limite

Une loi des grands nombres « lie » la fréquence et la probabilité. Une version « forte » du résultat précédent existe :

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \underset{n\to+\infty}{\longrightarrow} 0.$$

pour « presque tout $\omega \in \Omega$ ». Remarquons que quelque chose d'aléatoire converge vers un point complètement déterministe. On peut se demander ce qui se passe si au lieu de diviser par n, on divise par quelque chose qui tend aussi vers $+\infty$, mais moins vite que n. Est-ce qu'on peut obtenir à la limite un objet encore aléatoire ? Dans ce contexte, on va voir que la loi normale joue un rôle particulier, qui lui confère une grande importance en probabilité et en statistique.
Commençons donc par énoncer un résultat concernant les lois gaussiennes.

Proposition 4.5 *Si X et Y sont deux v.a. normales indépendantes de loi respective $\mathcal{N}(\mu,\sigma)$ et $\mathcal{N}(\nu,\theta)$, alors $X+Y$ est aussi une variable normale, de loi $\mathcal{N}(\mu+\nu,\sqrt{\sigma^2+\theta^2})$.*

Corollaire 4.6 *Soient $\{X_i,\ i=1,\dots,n\}$ des variables $\mathcal{N}(\mu,\sigma)$ indépendantes, alors*

$$\overline{X}_n \sim \mathcal{N}(\mu,\frac{\sigma}{\sqrt{n}}).$$

D'après la proposition 3.55, on peut représenter $\overline{X}_n$ comme

$$\overline{X}_n = \frac{\sigma}{\sqrt{n}}Z + \mu, \text{ avec } Z \sim \mathcal{N}(0,1),$$

d'où

$$Z = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}}.$$

Ce phénomène est en fait beaucoup plus général et fait l'objet du théorème suivant :

Théorème 4.7 (Théorème central limite) *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. i.i.d. de moyenne μ et de variance σ^2. Alors la distribution de $\overline{X}_n$ approche celle de la loi $\mathcal{N}(0,1)$ au sens suivant. Pour tout $a<b$ on a*

$$\lim_{n\to+\infty}\mathbb{P}\left(a < \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} < b\right) = \mathbb{P}(a<Z<b),$$

où $Z \sim \mathcal{N}(0,1)$. De façon équivalente, si on note $S_n = X_1 + \ldots + X_n$:

$$\lim_{n\to+\infty}\mathbb{P}\left(a < \frac{S_n - n\mu}{\sigma\sqrt{n}} < b\right) = \mathbb{P}(a<Z<b).$$

On note que $n\sigma^2 = \mathrm{Var}(S_n)$ et $n\mu = \mathbb{E}(S_n)$. Ainsi le théorème dit qu'en gros (voir l'équation 4.6 pour un énoncé plus rigoureux) :

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} \simeq Z.$$

4.3 Applications

La loi des grands nombres et le théorème central limite sont les deux théorèmes majeurs de la théorie des probabilités et il y en a un très grand nombre d'applications en mathématiques appliquées. Ici on va en citer deux, une numérique et une autre statistique.

4.3.1 Application numérique : méthode de Monte Carlo

Algorithme de calcul numérique d'intégrale : on veut calculer la valeur numérique d'une intégrale de la forme $\int_0^1 g(x)dx$, avec g continue. Une méthode dite de Monte-Carlo consiste à simuler une suite de v.a. $(U_n)_{n\geq 1}$ i.i.d. de loi uniforme sur $[0,1]$. La loi des grands nombres donne :

$$\lim_{n\to+\infty} \frac{g(U_1)+\ldots+g(U_n)}{n} = \mathbb{E}(g(U_1)) = \int_0^1 g(x)dx.$$

Donc si n est assez grand,

$$\frac{g(U_1)+\ldots+g(U_n)}{n} \simeq \int_0^1 g(x)dx.$$

Plus généralement si on sait simuler des v.a. $(X_n)_{n\geq 1}$ ayant une densité f, on va avoir une valeur numérique pour des intégrales de la forme :

$$\mathbb{E}(g(X_1)) = \int g(x)f(x)dx.$$

Un algorithme n'est pertinent que si on peut en donner la vitesse. Grâce au théorème central limite, on a une idée de la vitesse de convergence : elle est en $1/\sqrt{n}$. Et on ne pourra pas aller plus vite avec cette méthode. Les méthodes numériques déterministes sont plus rapides, mais leur rapidité dépend beaucoup de la régularité de la fonction à intégrer (au moins dérivable) et de la dimension de l'espace dans lequel la fonction prend ces valeurs. La méthode de Monte-Carlo a l'avantage de ne pas trop dépendre de la régularité, ni de la dimension. Remarquons pour finir que la variance joue un rôle important dans le théorème et on peut voir que plus cette variance est petite, plus le résultat numérique est proche du résultat théorique. Il existe quelques principes généraux qui permettent de réduire la variance.

4.3.2 Application statistique : modèle de Bernoulli

Dans la fabrication d'objets manufacturés, on suppose qu'une proportion inconnue d'objets défectueux $\theta \in [0,1]$ est produite. Cette proportion est évaluée par une fréquence d'apparition d'objets mal fabriqués dans des échantillons contrôlés au hasard.

- Lorsqu'elle dépasse un premier seuil θ_S, on renforce la surveillance de la production. Pour cela, on augmente la taille des échantillons utilisés pour évaluer θ.
- Lorsqu'elle dépasse un autre seuil $\theta_C > \theta_S$, on arrête la production pour réparer ou pour régler les machines.

Il faut bien sûr tenir compte du coût respectif de chaque type d'opération pour définir les seuils précédents. C'est l'objet du **contrôle de qualité**.

En pratique, on dispose de n observations de l'échantillon des objets contrôlés et (pour simplifier les choses), ils sont bons ou mauvais. On a ainsi une famille i.i.d. (indépendante et identiquement distribuée) $X_1, \ldots, X_n \sim Bern(\theta)$, de loi de Bernoulli de paramètre θ. Par suite X_i vaut 1 lorsque le i-ème objet contrôlé est de mauvaise qualité et 0 sinon, $X_i(\omega) \in \{0,1\}$ et $\mathbb{P}_\theta(X_i = 1) = 1 - \mathbb{P}_\theta(X_i = 0) = \theta$ lorsque le modèle est régi par la loi $\mathbb{P}_\theta$ (qui dépend donc de la valeur *inconnue* du paramètre θ). Nous considérons donc un schéma Succès-Echec fini de longueur n et de paramètre inconnu θ. L'*espace des paramètres* vaut ici $\Theta = [0,1]$.

Fondé sur cet exemple élémentaire, nous introduisons ici quelques notions fondamentales de statistique.

- On cherche d'abord à simplifier le jeu de données; la seule connaissance du nombre de pièces en bon état sur l'échantillon de taille n considéré sera suffisante pour résumer notre observation.

  Nous évoquons ici la question d'*exhaustivité*.
- On justifie ensuite le fait naturel que la probabilité empirique (proportion de pièces défectueuses) est une quantité approchant de manière raisonnable le paramètre de qualité θ de notre production. Cette quantité est aussi vue comme la plus vraisemblable, en un sens noté plus bas. D'autres manières de calculer empiriquement le paramètre sont ensuite envisagées, comme celle de *Bayes* qui probabilise l'espace des paramètres lui même et fournit ainsi des estimations convenables.

  C'est le problème d'*estimation* du paramètre θ.
- La fin du chapitre est dédiée à répondre à une question simple : *faut-il arrêter la production pour réparer les machines ?*

  Les problèmes envisagés sont ici ceux de la construction d'un *intervalle de confiance* et celui de *tests d'hypothèse*.
  - Une première réponse à cette question est en effet de situer de manière raisonnable le paramètre θ à partir de la seule donnée observée, c'est l'objet d'un *intervalle de confiance*. Avec une très grande probabilité, on doit pouvoir affirmer que $\theta \in IC$, IC est ainsi un intervalle dont les bornes sont aléatoires et fondées à partir de l'estimation de θ.
  - La construction d'un tel intervalle de confiance est délicate pour une taille d'échantillon fixée ; sa version asymptotique permet de mieux l'appréhender, au travers du théorème central limite.
  - Pour résoudre le problème de décision précédent, lié aux tests, on doit être en mesure de répondre à une question de type suivant :

    (Avec une grande probabilité) Le paramètre θ
    d'intérêt est-il suffisamment petit ?

  - Cette probabilité est un calcul difficile, on lui préfère une approximation obtenue via le théorème central limite.

– Une dernière question théorique se pose alors,

A partir de quelles valeurs (pour θ et pour la taille n de l'échantillon), l'asymptotique a-t-elle des caractéristiques acceptables ?

Résumer l'information

Si on cherche à évaluer θ à partir de l'observation $(X_1, \ldots, X_n)$, cela ne peut se faire qu'en considérant une fonction quelconque de l'observation disponible. On appelle *estimateur* $T_n = h(X_1, \ldots, X_n)$ toute fonction de l'observation.

Posons

$$S_n = X_1 + \cdots + X_n \tag{4.2}$$

alors $S_n \sim B(n, \theta)$ suit une loi binomiale :

$$\mathbb{P}_\theta(S_n = s) = C_n^s \theta^s (1-\theta)^{n-s}, \qquad s = 0, 1, \ldots, n.$$

On a alors,

$$\begin{aligned} \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n | S_n = s) &= \frac{\mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n, S_n = s)}{\mathbb{P}_\theta(S_n = s)} \\ &= \frac{\theta^s (1-\theta)^{n-s}}{C_n^s \theta^s (1-\theta)^{n-s}} \\ &= \frac{1}{C_n^s} \end{aligned} \tag{4.3}$$

car, lorsque $S_n = s$, exactement s des n variables aléatoires X_i prennent la valeur 1, ce qui justifie la valeur du numérateur.

La relation (4.3) signifie que, si la variable aléatoire S_n prend la valeur s ($S_n = s$), alors la configuration des X_i tels que $\sum_i X_i = s$ n'apporte aucune information sur le paramètre d'intérêt θ. Comme $\mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n | S_n = s)$ ne dépend pas de θ, on n'en apprendra pas plus de tout le résultat $(X_1, \ldots, X_n)$ de notre expérience aléatoire (qui peut prendre 2^n valeurs), que du résumé d'information constitué par S_n qui prend seulement $n+1$ valeurs. S_n est appelé résumé *exhaustif* de l'expérience $(X_1, \ldots, X_n)$, dans ce sens qu'il rapporte toute l'information relative à θ contenue dans notre expérience. Une définition plus rigoureuse de la notion d'*exhaustivité* existe mais dépasse le cadre de ce cours.

Moyenne empirique

La moyenne empirique est définie par la relation

$$\overline{X}_n = \frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n} \tag{4.4}$$

où S_n désigne le résumé exhaustif (4.2). Dans le contexte présent de variables de Bernoulli, $\overline{X}_n$ est la fréquence des pièces défectueuses dans l'échantillon examiné. En accord avec le sens commun, on dit que $\overline{X}_n$ *estime* θ. On a $\mathbb{E}_\theta \overline{X}_n = \theta$, et on dit que la variable aléatoire $\overline{X}_n$ *estime sans biais* le paramètre θ (le *biais* d'un estimateur T_n de θ est l'expression $\mathbb{E}_\theta T_n - \theta$).

Cet estimateur est naturel au sens que:

Proposition 4.8 *Soit $T_n = h(S_n)$ un estimateur sans biais de θ, fonction de la statistique exhaustive S_n, alors $T_n = \overline{X}_n$.*

Preuve :

Posons $g(x) = \frac{x}{n} - h(x)$, on doit prouver que

$$\{\forall \theta \in [0,1], \quad \mathbb{E}_\theta g(S_n) = 0\} \Rightarrow g \equiv 0$$

Cette relation s'écrit $\displaystyle\sum_{s=0}^{n} C_n^s \theta^s (1-\theta)^{n-s} g(s) = 0$. Le polynôme précédent en la variable $t = \theta/(1-\theta)$, identiquement nul si $t \in \mathbb{R}$ (ou si $\theta \in]0,1[$), a donc des coefficients nuls. □

Il est aussi *consistant*:

Proposition 4.9 *Pour tout $\varepsilon > 0$, on a*

$$\lim_{n\to+\infty} \mathbb{P}_\theta \left(|\overline{X}_n - \theta| \geq \varepsilon \right) = 0.$$

On a même :

$$\overline{X}_n \underset{n\to+\infty}{\longrightarrow} \theta.$$

pour « presque tout $\omega \in \Omega$ ».

Cet énoncé correspond aux lois « faible » et « forte » des grands nombres rappelées dans la section 4.1.

Maximum de vraisemblance

Rappelons que

$$\mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n) = \theta^s (1-\theta)^{n-s}, \qquad \mathbb{P}_\theta(S_n = s) = C_n^s \theta^s (1-\theta)^{n-s}$$

lorsque $s = x_1 + \cdots + x_n$. Par définition, la réalisation $x_1, \ldots, x_n$ de l'expérience est d'autant plus *vraisemblable* que $\theta^s(1-\theta)^{n-s}$ est grand.

Les expressions précédentes $V_\theta(X_1, \ldots, X_n) = \mathbb{P}_\theta(X_1 = x_1, \ldots, X_n = x_n)$ et $V_\theta^{(S_n)}(s) = \mathbb{P}_\theta(S_n = s)$ sont appelées *vraisemblances* de X et de S_n. La valeur du paramètre θ pour lequel le résultat de l'expérience obtenu est le plus vraisemblable s'obtient par maximisation de la vraisemblance $\theta \mapsto V_\theta(X_1, \ldots, X_n)$ (ou de manière équivalente de son logarithme $\theta \mapsto L_\theta(X_1, \ldots, X_n) = \log V_\theta(X_1, \ldots, X_n)$). Le maximum est atteint lorsque $\partial L_\theta / \partial \theta (X_1, \ldots, X_n) = S_n/\theta - (n - S_n)/(1-\theta) = 0$ soit $\theta = \overline{X}_n$.

La moyenne empirique est donc l'*estimateur du maximum de vraisemblance* de θ, noté $\hat{\theta}_n$. Notons que cette même expression maximise aussi la vraisemblance du paramètre exhaustif S_n. Cette méthode est très utile pour trouver l'expression d'estimateurs de paramètres inconnus.

Estimation bayésienne

Supposons maintenant que l'on dispose d'information *a priori* sur le paramètre θ. On appelle dans ce cas *la probabilité à priori* la probabilité ν de se trouver dans l'état θ.

Pour toute probabilité à priori ν sur $\Theta = [0, 1]$, on considère le *risque bayésien* d'un estimateur $T_n = h(X_1, \ldots, X_n)$

$$R_\nu(T_n) = \int_0^1 \mathbb{E}_\theta(T_n - \theta)^2 \, d\nu(\theta). \tag{4.5}$$

L'*estimateur bayésien* de θ est celui qui minimise $R_\nu(T_n)$. Posons

$$\mu_{k,\ell} = \int_0^1 \theta^k (1-\theta)^\ell d\nu(\theta), \qquad k, \ell = 1, 2, \ldots$$

Ce qui précède tend à privilégier un estimateur exhaustif de la forme $T_n = g(S_n)$ (fonction de S_n), et

$$\begin{aligned} R_\nu(T_n) &= \sum_{s=0}^n C_n^s \int_0^1 (g(s) - \theta)^2 \theta^s (1-\theta)^{n-s} d\nu(\theta), \\ &= \sum_{s=0}^n C_n^s \left(\mu_{s,n-s} g^2(s) - 2\mu_{s+1,n-s} g(s) + \mu_{s+2,n-s} \right). \end{aligned}$$

L'expression $R_\nu(T_n)$ est minimisée par $T_n = g(S_n)$ avec

$$g(s) = \frac{\mu_{s+1,n-s}}{\mu_{s,n-s}}, \ \forall s \in \{0, 1, \ldots, n\}.$$

Remarquons que l'estimateur $T_n = g(S_n)$ obtenu ici dépend de la probabilité à priori ν. Un critère différent du critère bayésien consiste à minimiser l'expression $T_n \mapsto \sup_\theta \mathbb{E}_\theta(T_n - \theta)^2$. C'est le critère *minimax*, qui minimise le *risque maximal*.

Intervalles de confiance

La loi faible des grands nombres 4.4 nous dit que

$$\mathbb{P}_\theta \left(\theta \notin \left[\overline{X}_n - \delta, \overline{X}_n + \delta \right] \right) = \mathbb{P}_\theta \left(\left| \overline{X}_n - \theta \right| > \delta \right) \leq \frac{\theta(1-\theta)}{n\delta^2} \leq \frac{1}{4n\delta^2}$$

car $\theta(1-\theta) \leq \frac{1}{4}$ pour $\theta \in \Theta$. Ainsi la confiance que l'on peut mettre dans le fait que $\theta \in I(X)$ où $IC(X) = \left[\overline{X}_n - \delta, \overline{X}_n + \delta \right]$ est au moins égale à $1 - \alpha$ lorsque $\alpha = \frac{1}{4n\delta^2}$.

On dit que $IC(X)$ est un intervalle de confiance (exact) au niveau α. Notons que l'intervalle $IC(X)$ proposé a des extrémités aléatoires.

Intervalles de confiance asymptotiques

Les intervalles de confiance asymptotiques sont construits à partir des théorèmes 4.1 et 4.7 du début du chapitre. Le terme asymptotiques veut dire que ces intervalles sont construit à partir de résultats limites lorsque $n \to +\infty$, contrairement aux intervalles de confiances exacts. La loi des grands nombres 4.1, combinée au théorème centrale limite 4.7 conduit, grâce au lemme de Slutsky qui ne sera pas explicité ici, au résultat suivant :

$$\lim_{n\to+\infty} \mathbb{P}\left(a < \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} < b\right) = \mathbb{P}(a < Z < b), \text{ avec } Z \sim \mathcal{N}(0,1). \tag{4.6}$$

Remarquons que dans notre example, on a l'égalité :

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}} = \sqrt{n}\frac{\left(\overline{X}_n - \theta\right)}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}, \qquad \text{sous la loi } \mathbb{P}_\theta$$

L'intervalle de confiance approché

$$IC(X) = \left[\overline{X}_n - \varphi_{1-\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}, \overline{X}_n + \varphi_{1-\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\right]$$

admet alors le niveau asymptotique α. On note ici φ_a l'unique nombre réel tel que $\mathbb{P}(Z < \varphi_a) = a$, il est appelé *quantile d'ordre a* de la loi normale.

Lorsque l'intervalle $IC(X)$ contient de grandes valeurs pour le paramètre θ, on imagine aisément qu'il est temps de re-régler la machine qui produit des pièces trop déffectueuses.

Contrôle de qualité

Pour satisfaire aux exigences du contrôle de qualité, on doit obtenir une règle de décision pour tester une hypothèse du type $\theta \leq \theta_0$ contre $\theta > \theta_0$. Soit $\alpha > 0$. Notons, qu'il existe un plus petit entier $k_\alpha \in \{0, 1, \ldots, n\}$ tel que

$$\mathbb{P}_{\theta_0}(S_n \leq k_\alpha) = \sum_{s=0}^{k_\alpha} C_n^s \theta_0^s (1-\theta_0)^{n-s} \geq 1 - \alpha$$

On acceptera l'hypothèse $\theta \leq \theta_0$ lorsque l'observation est telle que $S_n < k_\alpha$, et on la rejette lorsque $S_n \geq k_\alpha$. Pour quantifier ce test, la proposition suivante est essentielle (la preuve est laissée en exercice).

Proposition 4.10 *Soit $k \in \{0, 1, \ldots, n\}$, quelconque, alors l'application $\theta \mapsto \mathbb{P}_\theta(S > k)$ est croissante de $[0,1]$ dans $[0,1]$.*

La probabilité de rejeter l'hypothèse $\theta \leq \theta_0$ à tort est le *niveau du test*. En utilisant la proposition 4.10, on obtient :

$$\sup_{\theta\leq\theta_0} \mathbb{P}_\theta(S_n > k_\alpha) = \mathbb{P}_{\theta_0}(S_n > k_\alpha) \leq \alpha.$$

Un autre caractère du test est de savoir si on a effectivement eu raison de refuser l'hypothèse $\theta \leq \theta_0$ lorsque $\theta \leq \theta_0$; soit $\theta > \theta_0$, on note $\beta_\theta = \mathbb{P}_\theta(S_n > k_\alpha)$, la *puissance du test*. Alors la proposition 4.10 prouve aussi que $\beta_\theta \geq \alpha$, on dit que le test est *sans biais*.

Tests asymptotiques

Dans la section qui précède, l'inconvénient de la manière de procéder réside dans ce que la valeur de k_α n'est pas toujours simple à obtenir, même si elle est tabulée pour les petites valeurs de n. On se pose donc ici la question naturelle de savoir ce qui se passe quand n est grand.

Par suite du théorème central limite 4.7, pour n grand, on a pour $Z \sim \mathcal{N}(0,1)$ normale standard de densité $\varphi(u) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$,

$$\mathbb{P}_\theta(S_n > k) \equiv \mathbb{P}\left(Z > \frac{k - n\theta}{\sqrt{n}}\right) = 1 - \Phi\left(\frac{k - n\theta}{\sqrt{n}}\right)$$

où $\Phi(x) = \int_{-\infty}^{x} \varphi(u)\, du$.

Pour déterminer un k asymptotiquement raisonnable lorsque n est grand, on posera ainsi

$$k = n\theta_0 + \sqrt{n}\varphi_{1-\alpha} \backsimeq k_\alpha.$$

Ici φ_a est le *quantile d'ordre* a de la loi normale, i.e. tel que $\mathbb{P}(Z < \varphi_a) = a$. Si $\alpha = 0,05$ on sait que $\varphi_{1-\alpha} \backsimeq 1,96$ et pour $\alpha = 0,001$, $\varphi_{1-\alpha} \backsimeq 3$ (voir la table « Loi normale : Φ^{-1} » de l'annexe B).

Validité de l'asymptotique

Une question importante reste posée:

Quelle taille effective des échantillons permet l'approximation précédente ?

Cette question est fondamentale pour comprendre la nature quotidienne des approximations faites par les statisticiens, lorsqu'ils passent de « n grand » à des résultats valables à la limite lorsque $n \to +\infty$ dans toute procédure dite « ascymptotique ».

Théorème 4.11 (Petrov) *Soit $\epsilon > 0$ fixé, alors,*

$$\sup_{u \in \mathbb{R}} \Delta_{n,\theta}(u) = \mathcal{O}\left((n\theta)^{-\frac{1}{2}}\right)$$

uniformément pour $\theta \in [\epsilon, 1-\epsilon]$, si on pose

$$\Delta_{n,\theta}(u) = \left|\mathbb{P}_\theta\left(\frac{S_n - n\theta}{\sqrt{n\theta(1-\theta)}} \leq u\right) - \Phi(u)\right|$$

Ce théorème permet de valider l'approximation gaussienne lorsque le produit $n\theta$ est grand. **En statistique, on se contente traditionnellement de supposer**

$$n\theta \geq 5.$$

4.4 Exercices

Exercice 4.4.1 *Montrer en utilisant le Théorème central limite que*

$$\lim_n \sum_{k=n}^{\infty} e^{-n}\frac{n^k}{k!} = \frac{1}{2}.$$

Exercice 4.4.2 *On suppose que pour chaque n, X_n est une varaible aléatoire de loi $B(n,p)$ où $0 < p < 1$. Exprimer*

$$\lim_n P(X_n < np + \sqrt{n}),$$

comme une intégrale de la densité de la loi $N(0,1)$.

Exercice 4.4.3 (Théorème de Weierstrass)
Le théorème de Bernoulli 4.2 fournit une démonstration remarquable, due à Bernstein, du théorème d'approximation de Weierstrass, qui affirme qu'une fonction continue sur un intervalle borné peut être approchée par des polynômes, et ceci uniformément sur cet intervalle.

Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d. de loi commune de Bernoulli de paramètre $p \in]0,1[$. On pose $Y_n = (1/n)\sum_{k=1}^{n} X_k$.

1. *Rappeler l'énoncé du théorème de Bernoulli.*
2. *Soit $h : [0,1] \to \mathbb{R}$ une fonction continue. Montrer que $\mathbb{E}(h(Y_n)) \to h(p)$ ($n \to +\infty$) et ceci uniformément en $p \in]0,1[$.*

Exercice 4.4.4 *Une machine a produit dans le passé des rondelles ayant une épaisseur de 0,05 cm. Pour déterminer si la machine est encore en état de marche, on choisit un échantillon de 10 rondelles dont les épaisseurs vérifient : $\overline{X}_n = 0,053$ cm et $\sigma_n^2 = 9(10^{-6})$ cm^2. Tester l'hypothèse que la machine est en état de marche au niveau $0,95$, puis au niveau $0,99$.*

Bibliographie

[1] M. Crouzeix, A.L. Mignot, *Analyse numérique des équations différentielles.* Masson, Paris.

[2] J.P. Demailly, *Analyse numérique et équations différentielles.* PUG, Grenoble.

[3] M. Benaïm, *Promenade aléatoire.* Cours polycopié de l'École Polytechnique, Paris.

[4] F. Comets, *Introduction aux probabilités et à la simulation aléatoire.* Cours polycopié de l'École Polytechnique, Paris.

[5] W. Feller**, *An Introduction to Probability Theory and Its Applications Volume 1.* Wiley Text Books, third edition (1968).

[6] D. Foata, A. Fuchs, *Calcul des probabilités.* cours, exercices et problèmes corrigés, Dunod, Paris (1998).

[7] G. Grimmett, D. Welsh, *Probability: An Introduction.* Oxford University Press, Oxford (1986).

[8] A. Monfort*, *Statistique.* Cours polycopié de l'École Polytechnique, Paris.

[9] D. Revuz*, *Probabilités.* Hermann, Paris (1997).

[10] A. N. Shiryaev**, *Probability.* Telos, 2nd edition (1995).

* : Pour aller plus loin.

** : Pour aller beaucoup plus loin.

Annexe A

Tables des lois

Lois discrètes

Loi	Probabilités	Espérance	Variance	Fonction génératrice
Bernoulli $\mathcal{B}(p)$ $p \in]0,1[$	$p_1 = p, p_0 = 1-p$	p	$p(1-p)$	$1-p+pz$
Binomiale $\mathcal{B}(n,p)$ $n \in \mathbb{N}^*,\ p \in]0,1[$	$C_n^k p^k (1-p)^{n-k}$ $k \in \{0,\ldots,n\}$	np	$np(1-p)$	$(1-p+pz)^n$
Géométrique $\mathcal{G}(p)$ $p \in]0,1[$	$p(1-p)^{k-1}$ $k \in \mathbb{N}^*$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pz}{1-(1-p)z}$
Poisson $\mathcal{P}(\lambda)$ $\lambda \in \mathbb{R}_+^*$	$e^{-\lambda}\frac{\lambda^k}{k!}$ $k \in \mathbb{N}$	λ	λ	$e^{\lambda(z-1)}$

Lois continues

Loi	Densité	Espérance	Variance	Fonction caractéristique
Uniforme $\mathcal{U}(a,b)$ $a<b$	$\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{ibt}-e^{iat}}{i(b-a)t}$
Exponentielle $\mathcal{E}(\lambda)$ $\lambda \in \mathbb{R}_+^*$	$\lambda e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{1}{1-it/\lambda}$
Gamma $G(\alpha,\lambda)$ $\alpha,\lambda \in \mathbb{R}_+^*$	$\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}\mathbf{1}_{\mathbb{R}_+^*}(x)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\left(1-\frac{it}{\lambda}\right)^{-\alpha}$
Normale $\mathcal{N}(m,\sigma^2)$ $m \in \mathbb{R},\ \sigma \in \mathbb{R}_+^*$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$e^{imt-\sigma^2t^2/2}$

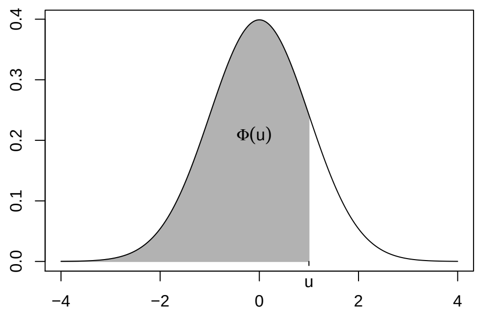
$$\Gamma(a) = \int_0^\infty e^{-x}x^{a-1}\,dx \qquad \text{pour } a \in \mathbb{R}_+^*,$$
$$\Gamma(1) = 1, \qquad \Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}^*.$$

Annexe B

Tables statistiques

Loi normale : Φ

Si Z est une variable aléatoire de loi $\mathcal{N}(0,1)$, la table donne la valeur de la fonction de répartition de Z : $\Phi(u) = \mathbb{P}(Z \leq u)$.



u	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

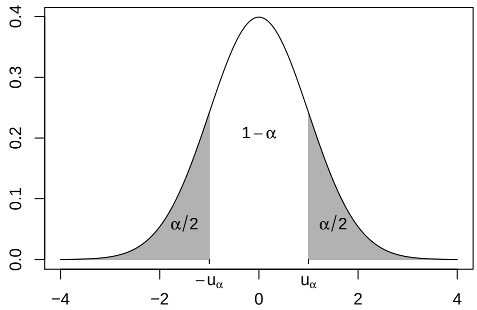
Grandes valeurs de u

u	3.0	3.5	4.0	4.5
$\Phi(u)$	0.998650	0.999767	0.999968	0.999997

Loi normale : Φ^{-1}

Si Z est une variable aléatoire de loi $\mathcal{N}(0,1)$, et α un réel de $[0,1]$, la table donne la valeur

$$\varphi_{1-\alpha} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \qquad \text{telle que} \quad \mathbb{P}(|Z| > \varphi_{1-\alpha}) = \alpha.$$



α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	$+\infty$	2.5758	2.3263	2.1701	2.0537	1.9600	1.8808	1.8119	1.7507	1.6954
0.1	1.6449	1.5982	1.5548	1.5141	1.4758	1.4395	1.4051	1.3722	1.3408	1.3106
0.2	1.2816	1.2536	1.2265	1.2004	1.1750	1.1503	1.1264	1.1031	1.0803	1.0581
0.3	1.0364	1.0152	0.9945	0.9741	0.9542	0.9346	0.9154	0.8965	0.8779	0.8596
0.4	0.8416	0.8239	0.8064	0.7892	0.7722	0.7554	0.7388	0.7225	0.7063	0.6903
0.5	0.6745	0.6588	0.6433	0.6280	0.6128	0.5978	0.5828	0.5681	0.5534	0.5388
0.6	0.5244	0.5101	0.4959	0.4817	0.4677	0.4538	0.4399	0.4261	0.4125	0.3989
0.7	0.3853	0.3719	0.3585	0.3451	0.3319	0.3186	0.3055	0.2924	0.2793	0.2663
0.8	0.2533	0.2404	0.2275	0.2147	0.2019	0.1891	0.1764	0.1637	0.1510	0.1383
0.9	0.1257	0.1130	0.1004	0.0878	0.0753	0.0627	0.0502	0.0376	0.0251	0.0125

Petites valeurs de α

α	0.002	0.001	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
u_α	3.0902	3.2905	3.8906	4.4172	4.8916	5.3267	5.7307	6.1094

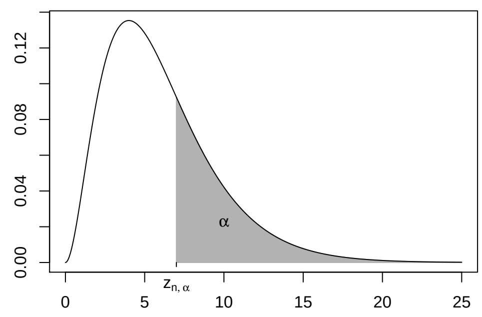
Pour $p < \frac{1}{2}$, $\Phi^{-1}(p) = -\varphi_{1-2p}$.

Pour $p \geq \frac{1}{2}$, $\Phi^{-1}(p) = \varphi_{1-2(1-p)}$.

Loi du χ^2

Si X est une variable aléatoire de loi du χ^2 à n degrés de libertés, F_n sa fonction de répartition, et α un réel de $[0,1]$, la table (en page suivante) donne la valeur $z_{n,\alpha} = F_n^{-1}(1-\alpha)$ telle que

$$\mathbb{P}(X > z_{n,\alpha}) = \alpha.$$



Pour $n > 30$, on admet que :

$$\begin{aligned} z_{n,\alpha} &\simeq \frac{1}{2}\left(u_{2\alpha} + \sqrt{2n-1}\right)^2 \qquad \text{si} \quad \alpha < 1/2, \\ z_{n,\alpha} &\simeq \frac{1}{2}\left(\sqrt{2n-1} - u_{2(1-\alpha)}\right)^2 \qquad \text{si} \quad \alpha \geq 1/2. \end{aligned}$$

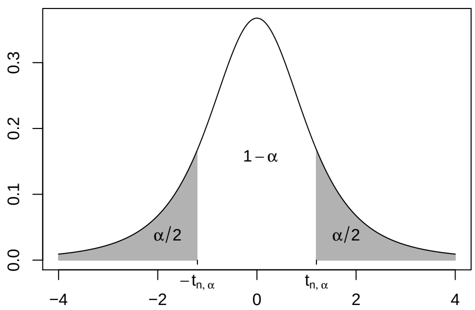
n \ α	0.995	0.99	0.975	0.95	0.9	0.8	0.7	0.5	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001
1	0.00004	0.0002	0.001	0.004	0.02	0.06	0.15	0.45	1.07	1.64	2.71	3.84	5.02	6.63	7.88	10.83
2	0.01	0.02	0.05	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.61	5.99	7.38	9.21	10.60	13.82
3	0.07	0.11	0.22	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.81	9.35	11.34	12.84	16.27
4	0.21	0.30	0.48	0.71	1.06	1.65	2.19	3.36	4.88	5.99	7.78	9.49	11.14	13.28	14.86	18.47
5	0.41	0.55	0.83	1.15	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	12.83	15.09	16.75	20.52
6	0.68	0.87	1.24	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	14.45	16.81	18.55	22.46
7	0.99	1.24	1.69	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.01	18.48	20.28	24.32
8	1.34	1.65	2.18	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	17.53	20.09	21.95	26.12
9	1.73	2.09	2.70	3.33	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.02	21.67	23.59	27.88
10	2.16	2.56	3.25	3.94	4.87	6.18	7.27	9.34	11.78	13.44	15.99	18.31	20.48	23.21	25.19	29.59
11	2.60	3.05	3.82	4.57	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	21.92	24.72	26.76	31.26
12	3.07	3.57	4.40	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	23.34	26.22	28.30	32.91
13	3.57	4.11	5.01	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	24.74	27.69	29.82	34.53
14	4.07	4.66	5.63	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.12	29.14	31.32	36.12
15	4.60	5.23	6.26	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	27.49	30.58	32.80	37.70
16	5.14	5.81	6.91	7.96	9.31	11.15	12.62	15.34	18.42	20.47	23.54	26.30	28.85	32.00	34.27	39.25
17	5.70	6.41	7.56	8.67	10.09	12.00	13.53	16.34	19.51	21.61	24.77	27.59	30.19	33.41	35.72	40.79
18	6.26	7.01	8.23	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	31.53	34.81	37.16	42.31
19	6.84	7.63	8.91	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	32.85	36.19	38.58	43.82
20	7.43	8.26	9.59	10.85	12.44	14.58	16.27	19.34	22.77	25.04	28.41	31.41	34.17	37.57	40.00	45.31
21	8.03	8.90	10.28	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	35.48	38.93	41.40	46.80
22	8.64	9.54	10.98	12.34	14.04	16.31	18.10	21.34	24.94	27.30	30.81	33.92	36.78	40.29	42.80	48.27
23	9.26	10.20	11.69	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.08	41.64	44.18	49.73
24	9.89	10.86	12.40	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	39.36	42.98	45.56	51.18
25	10.52	11.52	13.12	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	40.65	44.31	46.93	52.62
30	13.79	14.95	16.79	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	46.98	50.89	53.67	59.70

Loi de Student

Si X est une variable aléatoire de loi de Student $\mathcal{S}_n$, F_n sa fonction de répartition et α un réel de $[0,1]$, la table (en page suivante) donne la valeur de $t_{n,\alpha} = F_n^{-1}(1-\alpha/2)$ telle que

$$\mathbb{P}(|X| > t_{n,\alpha}) = \alpha.$$

On a $t_{+\infty,\alpha} = u_\alpha$.



n \ α	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
80	0.126	0.254	0.387	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.416
120	0.126	0.254	0.386	0.526	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373