Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées
Bât. M2, F-59655 Villeneuve d'Ascq Cedex

Introduction au Calcul des Probabilités

Probabilités à Bac+2 et plus si affinités...

Charles SUQUET

DEUG MIAS 2 et MASS 2 **2002–2003**

Table des matières

Introduction

Issu du cours de Probabilités en DEUG MASS et MIAS, ce document s'adresse à un public varié. Les étudiants de DEUG pourront y trouver une rédaction détaillée de toutes les questions abordées en cours. Quelques développements vont au-delà du strict programme et sont susceptibles d'intéresser des lecteurs curieux ou plus avancés. Les outils mathématiques utilisés restent néanmoins strictement dans le cadre du DEUG.

Ce premier tome[1] est consacré à ce que l'on appelle les *probabilités discrètes*. Par rapport aux rudiments de calcul des probabilités enseignés au lycée, l'innovation est la prise en compte de l'infini. Cette notion s'introduit très naturellement en calcul des probabilités, par exemple dès qu'il s'agit de modéliser des temps d'attente. On ne peut pas étudier avec un espace Ω de cardinal fini une expérience aléatoire aussi simple que : « on lance un dé jusqu'à la première obtention d'un six ». Nous nous posons donc la question de la définition et de l'étude des probabilités sur des *univers* Ω infinis. Il est possible au niveau du DEUG de faire une théorie assez rigoureuse si l'on veut bien faire l'impasse sur les problèmes de construction (ou d'existence) de tels espaces probabilisés infinis capables de modéliser correctement les expériences aléatoires envisagées.

Le principal outil mathématique utilisé est celui des *séries*. Il permet une étude classique assez complète des variables aléatoires discrètes. Cette étude débouche sur deux grands théorèmes de convergence de la théorie des probabilités : la loi des grands nombres et la convergence vers une loi gaussienne qui sont discutés dans des cas simples dans les deux derniers chapitres. Nous avons choisi de donner autant que possible des démonstrations de ces théorèmes dans ces cas particuliers. Ces démonstrations sont instructives en elles-mêmes et peuvent être considérées comme une introduction au cours de Licence. Une autre particularité de ce document est la discussion sur les questions de vitesse de convergence à propos des approximations (par une loi de Poisson ou par une loi de Gauss). Trop souvent on trouve à ce sujet dans la littérature des recettes qui, données sans justification, ressemblent plus à de la cuisine[2] qu'à des mathématiques.

Chaque chapitre contient une section d'exercices qui suit autant que possible

[1] Y en aura-t-il un deuxième ?

[2] Il y a souvent de bonnes raisons cachées derrière une recette qui peut paraître arbitraire. . .

l'ordre d'exposition du cours[3]. Certains sont des applications directes du cours ou des sujets d'examen ou de D.S., d'autres des approfondissements. Leur niveau de difficulté n'a volontairement pas été indiqué a priori. De même, on ne trouvera pas dans cette introduction de plan de lecture détaillé pour chaque DEUG. De telles indications pourront être données en cours ou en TD, mais je n'ai pas souhaité cloisonner a priori une curiosité qui, pour un scientifique, est tout le contraire d'un vilain défaut…

Je remercie tous les collègues qui m'ont aidé directement ou indirectement à rédiger ce polycopié et plus particulièrement Maurice Chamontin, Sylvie Roelly et Marie-Claude Viano avec qui j'ai fait équipe en DEUG MASS et MIAS. Il va de soi qu'ils ne portent aucune responsabilité pour les quelques débordements auxquels j'ai pu me laisser aller ni pour les quelques fautes[4] que l'on ne manquera pas de trouver dans cette première édition [5](septembre 1996).

Comme prévu ci-dessus, le deuxième tome n'a toujours pas été écrit et un certain nombre d'erreurs ont été détectées dans la première édition et corrigées dans la deuxième [6] (septembre 1997). Je remercie tous ceux qui m'en ont signalé et plus particulièrement les étudiants de l'amphithéâtre de DEUG MASS 96–97 pour leur vigilance. Merci également à Michel Lifshits pour ses précisions sur l'historique du théorème de De Moivre-Laplace, à Youri Davydov et Myriam Fradon pour d'utiles discussions ainsi qu'à tous les chargés de TD de probabilités en DEUG MIAS pour leur participation active. *Last but not least*, merci à Daniel Flipo qui avec patience et disponibilité m'a fait bénéficier de ses compétences d'expert dans le traitement de texte scientifique LaTeX 2_{ε}.

Les troisième et quatrième éditions de ce polycopié (septembre 1998 et 1999), ont bénéficié des amendements et corrections suggérés par Myriam Fradon, Jeanne Devolder et Anne Philippe. C'est pour moi un plaisir de les en remercier ici.

La cinquième édition (septembre 2000) de ce polycopié s'est enrichie (alourdie ?) d'un chapitre sur les variables aléatoires réelles qui s'est substitué à la promesse électorale d'un deuxième tome. Le titre a changé en conséquence.

La sixième édition (septembre 2001) comprend quelques exercices supplémentaires. La septième est inchangée, sauf la correction d'un quarantaine (sic) de fautes de frappe ou d'orthographe. La plupart m'ont été signalées par Denis Bitouzé de l'Université du Littoral que je remercie pour sa lecture attentive. Ce document est disponible sur Internet, au format PDF, à l'adresse suivante `http ://www.univ-lille1.fr/labo-stat-proba/cs/`

Villeneuve d'Ascq, septembre 2002.

[3]Ces exercices ne se substituent pas aux séances de TD et à leurs fiches d'exercices mieux adaptées à chacun des publics concernés.

[4]Dont le nombre suit une loi de Poisson.

[5]Remerciements anticipés à tout lecteur qui m'aidera à réduire le paramètre de ladite loi pour la prochaine édition.

[6]Qui ne prétend pas en être exempte, voir exercice 5.7 pour une modélisation.

Chapitre 1

Espaces Probabilisés

1.1 Introduction

La théorie des probabilités fournit des modèles mathématiques permettant l'étude d'expériences dont le résultat ne peut être prévu avec une totale certitude. En voici quelques exemples :

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

Bien que le résultat précis de chacune de ces expériences soit imprévisible, l'observation et l'intuition nous amènent à penser que ces phénomènes obéissent à certaines lois. Par exemple si on jette 6000 fois le dé, on s'attend à ce que le nombre d'apparitions de la face « 3 » soit *voisin* de 1000. Si on met en service 100 ampoules, leurs durées de vie observées seront *concentrées* autour d'une certaine valeur moyenne.

La théorie des probabilités permet de donner un sens précis à ces considérations un peu vagues. La *statistique* permet de confronter les modèles probabilistes avec la réalité observée afin de les valider ou de les invalider. Par exemple si quelqu'un a 60 bonnes réponses sur 100 au questionnaire, est-il légitime de considérer qu'il a « mieux fait » que le hasard ? Sur les n objets prélevés en sortie de chaîne, k sont défectueux. Peut-on en déduire quelque chose sur la qualité de la production globale ?

1.2 Événements

La théorie moderne des probabilités utilise le langage des ensembles pour modéliser une expérience aléatoire. Nous noterons Ω un ensemble dont les éléments représentent tous les résultats possibles ou *événements élémentaires* d'une expérience aléatoire donnée. Les *événements* (ou événements composés) seront représentés par des parties (sous-ensembles) de Ω.

Il n'est pas toujours facile de trouver un ensemble Ω permettant de modéliser l'expérience aléatoire. Voici une règle pratique pour y arriver : les événements élémentaires sont ceux qui contiennent *l'information maximale* qu'il est possible d'obtenir de l'expérience. Par exemple si on jette un dé, l'événement A : « obtention d'un chiffre pair » n'est pas élémentaire. Il est composé des trois événements élémentaires 2, 4, 6 : $A = \{2, 4, 6\}$. Ici $\Omega = \{1, 2, 3, 4, 5, 6\}$. De même si on lance trois fois une pièce de monnaie, les événements élémentaires sont des triplets comme (p,f,p) indiquant le résultat précis de chacun des trois lancers. Ici $\Omega = \{\mathtt{f}, \mathtt{p}\}^3$. L'événement B « obtention de pile au deuxième des trois lancers » est composé : $B = \{(\mathtt{f}, \mathtt{p}, \mathtt{f}); (\mathtt{f}, \mathtt{p}, \mathtt{p}); (\mathtt{p}, \mathtt{p}, \mathtt{f}); (\mathtt{p}, \mathtt{p}, \mathtt{p})\}$.

Avec ce mode de représentation, les opérations logiques sur les événements : « et », « ou », « négation » se traduisent par des opérations ensemblistes : intersection, réunion, passage au complémentaire. Voici un tableau de correspondance entre les deux langages.

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

Les opérations logiques sur les événements peuvent bien sûr faire intervenir plus de deux événements. Ainsi, si $A_1, \dots, A_n$ sont des événements,

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

est l'ensemble des ω qui sont dans l'un au moins des A_i. C'est donc l'événement « réalisation de l'un au moins des A_i $(1 \leq i \leq n)$ ». De même :

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

est l'ensemble des ω qui sont dans tous les A_i. C'est donc l'événement « réalisation de chacun des A_i $(1 \leq i \leq n)$ ». Il est facile d'étendre ces définitions aux réunions et intersections d'une suite infinie d'événements :

$$\begin{aligned}
\bigcup_{i \in \mathbb{N}^*} A_i = \bigcup_{i=1}^{+\infty} A_i &= \{\text{réalisation de l'un au moins des } A_i,\ i \in \mathbb{N}^*\}, \\
\bigcap_{i \in \mathbb{N}^*} A_i = \bigcap_{i=1}^{+\infty} A_i &= \{\text{réalisation de tous les } A_i,\ i \in \mathbb{N}^*\}.
\end{aligned}$$

Ces opérations logiques sur des suites d'événements sont très utiles pour analyser des événements complexes à l'aide d'événements plus simples et, comme nous le verrons plus tard, calculer ainsi des probabilités. A titre d'illustration, examinons la situation suivante.

Exemple 1.1 *Alice et Bruno lancent le même dé à tour de rôle (Alice commence). Le gagnant est le premier à obtenir un « six ».*
On s'intéresse aux trois événements

$$\begin{aligned}
A &= \{\text{victoire d'Alice}\}, \\
B &= \{\text{victoire de Bruno}\}, \\
D &= \{\text{Il n'y a pas de vainqueur}\}.
\end{aligned}$$

La simplicité de leur définition est trompeuse. Leur structure compliquée peut être analysée à l'aide des événements plus simples suivants :

$$\begin{aligned}
F_n &= \{\text{fin de la partie au } n\text{-ième lancer}\}, \quad n \in \mathbb{N}^*, \\
S_j &= \{\text{le } j\text{-ème lancer donne un « six »}\}, \quad j \in \mathbb{N}^*.
\end{aligned}$$

Commençons par D. Il est clair que D se réalise si et seulement si le « six » n'apparaît à aucun lancer, autrement dit si et seulement si quel que soit $j \in \mathbb{N}^*$, l'événement S_j^c est réalisé. D'où :

$$D = \bigcap_{j \in \mathbb{N}^*} S_j^c.$$

Alice ne peut gagner la partie que lors d'un lancer de rang impair puisque les lancers de rang pair sont ceux de Bruno. Alice peut donc gagner à l'un des lancers $1, 3, 5, \ldots, 2k+1, \ldots$ Alice gagne si et seulement si la partie se termine par l'un de ces lancers. De même Bruno peut gagner aux lancers $2, 4, 6, \ldots, 2k, \ldots$ d'où :

$$A = \bigcup_{k\in\mathbb{N}} F_{2k+1} \qquad B = \bigcup_{k\in\mathbb{N}^*} F_{2k}.$$

Enfin, chaque F_n ($n \geq 2$) se réalise si et seulement si d'une part chacun des $n-1$ premiers lancers donne autre chose qu'un « six », autrement dit réalise S_j^c ($1 \leq j \leq n-1$) et d'autre part le n-ième lancer donne « six » (S_n se réalise). On a donc :

$$F_1 = S_1, \qquad F_n = \left(\bigcap_{j=1}^{n-1} S_j^c \right) \cap S_n, \qquad n > 1$$

et finalement :

$$A = \bigcup_{k\in\mathbb{N}} \left(\left(\bigcap_{j=1}^{2k} S_j^c \right) \cap S_{2k+1} \right), \qquad B = \bigcup_{k\in\mathbb{N}^*} \left(\left(\bigcap_{j=1}^{2k-1} S_j^c \right) \cap S_{2k} \right).$$

Remarquons que nous n'avons pas eu besoin de préciser dans quel ensemble Ω on travaillait pour effectuer les décompositions d'événements ci-dessus. Seule importait leur structure logique. Voici un choix possible (parmi d'autres) de Ω : on prend l'ensemble des suites de chiffres ω de l'un des deux types suivants. Soit ω est une suite *finie* de chiffres pris parmi $\{1, 2, 3, 4, 5\}$ et terminée par un 6. Soit ω est une suite *infinie* de chiffres pris parmi $\{1, 2, 3, 4, 5\}$ (et donc sans aucun 6). Remarquons qu'avec ce choix de Ω, D est l'ensemble des suites du deuxième type : $D = \{1, 2, 3, 4, 5\}^{\mathbb{N}^*}$.

Remarque[1] Si I est un ensemble d'indices *quelconque* non vide et $(A_i)_{i\in I}$ une famille d'événements indexée par I, on peut définir $\cup_{i\in I} A_i$ comme l'*ensemble* des ω appartenant à l'un au moins des A_i et $\cap_{i\in I} A_i$ comme l'*ensemble* des ω appartenant à chacun des A_i. Ces définitions sont globales et ne font appel à aucune structure d'ordre sur I ni, dans le cas où I est infini, à aucune notion de convergence.

1.3 La probabilité comme fonction d'ensembles

Ce titre appelle tout de suite une explication. La *probabilité* P telle que nous allons la définir ci-dessous est une fonction qui à un événement associe un nombre compris entre 0 et 1 et censé mesurer les chances de réalisation de cet événement. Pour des raisons sortant du cadre de ce cours, il n'est pas

[1] à passer en première lecture.

toujours possible d'attribuer ainsi de manière cohérente une probabilité à chaque partie de Ω. En d'autres termes, P ne peut pas être considérée comme une *application* de l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω dans $[0,1]$ mais comme une *fonction* ayant un domaine de définition $\mathcal{F}$ généralement plus petit que $\mathcal{P}(\Omega)$. Voici les propriétés qu'il est raisonnable d'exiger de $\mathcal{F}$:

– $\mathcal{F}$ contient Ω et tous les singletons $\{\omega\}$.
– $\mathcal{F}$ est stable par passage au complémentaire : si B est un événement de $\mathcal{F}$, B^c l'est aussi.
– $\mathcal{F}$ est stable par les opérations de réunion et d'intersection sur les *suites* d'événements[2]. Si $A_1, A_2, \dots$ est une suite finie ou infinie d'événements de $\mathcal{F}$, sa réunion et son intersection sont encore des événements de $\mathcal{F}$.

L'étude faite à l'exemple 1.1 montre que ces propriétés sont vraiment le moins que l'on puisse exiger de $\mathcal{F}$ dès que l'on travaille sur une expérience aléatoire ayant une infinité de résultats élémentaires possibles (i.e. Ω infini). Nous appellerons *famille d'événements observables* toute famille $\mathcal{F}$ de parties de Ω vérifiant les conditions ci-dessus[3].

Définition 1.1 *Soit Ω un ensemble et $\mathcal{F}$ une famille d'événements observables sur Ω. On appelle probabilité sur $(\Omega, \mathcal{F})$ toute application P de $\mathcal{F}$ dans $[0,1]$ vérifiant :*

(i) $P(\Omega) = 1$.

(ii) Pour toute suite $(A_j)_{j\geq 1}$ d'événements de $\mathcal{F}$ deux à deux disjoints (incompatibles) :

$$P\Big(\bigcup_{j\in\mathbb{N}^*} A_j\Big) = \sum_{j=1}^{+\infty} P(A_j).$$

Le triplet $(\Omega, \mathcal{F}, P)$ s'appelle espace probabilisé.

Définir une probabilité sur $(\Omega, \mathcal{F})$ c'est en quelque sorte attribuer une « masse » à chaque événement observable, avec par convention une masse totale égale à 1 pour l'événement certain Ω. La propriété (ii) s'appelle *σ-additivité*.

Exemple 1.2 (suite de l'exemple 1.1)
Revenons à la partie de dé entre Alice et Bruno . Admettons provisoirement que l'on ait construit un $(\Omega, \mathcal{F}, P)$ modélisant correctement cette expérience et que pour $n \geq 1$, $P(F_n) = (5/6)^{n-1}(1/6)$. Ce résultat peut se justifier par la règle du conditionnement en chaîne qui sera vue ultérieurement. On peut alors calculer la probabilité de victoire de chacun des joueurs.

[2] Bien sûr grâce à la stabilité par complémentaire, la stabilité par réunion équivaut à la stabilité par intersection.

[3] On dit aussi *tribu* ou σ-algèbre de parties de Ω, mais nous n'emploierons pas ce vocabulaire abstrait. La définition générale d'une tribu $\mathcal{F}$ ne suppose pas que tous les singletons $\{\omega\}$ soient des éléments de $\mathcal{F}$.

En effet on a

$$A = \bigcup_{k\in\mathbb{N}} F_{2k+1} \qquad B = \bigcup_{k\in\mathbb{N}^*} F_{2k}.$$

De plus chacune de ces réunions est *disjointe* : si $i \neq j$, $F_i \cap F_j = \emptyset$ car si un événement élémentaire ω était commun à F_i et F_j, cela signifierait que pour *la même suite*, le « six » apparaîtrait pour la *première* fois au i-ème lancer *et* au j-ème lancer ce qui est absurde. Donc les F_{2k+1} sont deux à deux disjoints et

$$P(A) = \sum_{k\in\mathbb{N}} P(F_{2k+1}) = \sum_{k=0}^{+\infty} \frac{1}{6}\Bigl(\frac{5}{6}\Bigr)^{2k} = \frac{1}{6}\sum_{j=0}^{+\infty}\Bigl(\frac{25}{36}\Bigr)^{j} = \frac{1}{6}\,\frac{1}{1-\frac{25}{36}} = \frac{6}{11},$$

en utilisant la convergence et le calcul de la somme de la série géométrique de raison 25/36. De même :

$$P(B) = \sum_{k\in\mathbb{N}^*} P(F_{2k}) = \sum_{k=1}^{+\infty} \frac{1}{6}\Bigl(\frac{5}{6}\Bigr)^{2k-1} = \frac{5}{36}\sum_{j=0}^{+\infty}\Bigl(\frac{25}{36}\Bigr)^{j} = \frac{5}{36}\,\frac{1}{1-\frac{25}{36}} = \frac{5}{11}.$$

On constate que Alice est légèrement avantagée par le fait de lancer la première, ce qui est conforme à l'intuition. De plus par la propriété d'additivité *2.(b)* ci-dessous, comme A, B et D sont trois événements disjoints dont la réunion est Ω, on en déduit que $P(D) = 0$. La probabilité qu'il n'y ait aucun vainqueur est donc nulle, ce qui là aussi est conforme à l'intuition. On remarquera cependant que dans le modèle choisi pour Ω, $D = \{1,2,3,4,5\}^{\mathbb{N}^*}$ est très loin d'être vide, c'est même un ensemble très « gros » du point de vue de la cardinalité : on peut démontrer qu'il est en bijection avec l'ensemble de tous les nombres réels[4]. ■

Proposition 1.2 (Propriétés générales d'une probabilité)
Toute probabilité P sur $(\Omega, \mathcal{F})$ vérifie les propriétés suivantes :

1. $P(\emptyset) = 0$.
2. *Additivité*
   (a) *Si $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.*
   (b) *Si les A_i $(1 \leq i \leq n)$ sont deux à deux disjoints :*

$$P\Bigl(\bigcup_{i=1}^{n} A_i\Bigr) = \sum_{i=1}^{n} P(A_i).$$

3. $\forall A \in \mathcal{F},\ P(A^c) = 1 - P(A)$.
4. $\forall A \in \mathcal{F},\ \forall B \in \mathcal{F},\ A \subset B \ \Rightarrow\ P(A) \leq P(B)$.
5. $\forall A \in \mathcal{F},\ \forall B \in \mathcal{F},\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
6. *Continuité monotone séquentielle*

[4] L'idée clé est de considérer le développement en base 5 des réels de $]0,1[$...

(a) *Si* $(B_n)_{n\geq 1}$ *est une suite croissante d'événements de* $\mathcal{F}$ *convergente*[5] *vers* $B \in \mathcal{F}$, *alors* $P(B) = \lim_{n\to+\infty} P(B_n)$. *Notation :*

$$B_n \uparrow B \Rightarrow P(B_n) \uparrow P(B) \quad (n \to +\infty).$$

(b) *Si* $(C_n)_{n\geq 1}$ *est une suite décroissante d'événements de* $\mathcal{F}$ *convergente*[6] *vers* $C \in \mathcal{F}$, *alors* $P(C) = \lim_{n\to+\infty} P(C_n)$. *Notation :*

$$C_n \downarrow C \Rightarrow P(C_n) \downarrow P(C) \quad (n \to +\infty).$$

7. (a) $\forall A \in \mathcal{F}, \forall B \in \mathcal{F} \qquad P(A \cup B) \leq P(A) + P(B)$.

(b) $\forall A_1, \dots, A_n \in \mathcal{F}, \qquad P\big(\bigcup_{i=1}^{n} A_i\big) \leq \sum_{i=1}^{n} P(A_i)$.

(c) $\forall A_1, \dots, A_n, \dots \in \mathcal{F}, \qquad P\big(\bigcup_{i\in\mathbb{N}^*} A_i\big) \leq \sum_{i=1}^{+\infty} P(A_i)$.

Preuve : Soit P une fonction d'ensemble $\mathcal{F} \to [0, 1]$ satisfaisant aux conditions (i) et (ii) de la définition 1.1, il s'agit de démontrer que P vérifie les propriétés *1* à *7*.

Preuve de 1. Comme $P(A_j) \geq 0$ pour tout $A_j \in \mathcal{F}$, on a toujours

$$\sum_{j\in\mathbb{N}^*} P(A_j) \geq P(A_1) + P(A_2),$$

le premier membre pouvant être égal à $+\infty$. En choisissant $A_j = \emptyset$ pour tout $j \in \mathbb{N}^*$ et en utilisant la σ-additivité (ii), on en déduit :

$$P(\emptyset) = P\big(\bigcup_{j\in\mathbb{N}^*} A_j\big) = \sum_{j=1}^{+\infty} P(A_j) \geq P(\emptyset) + P(\emptyset).$$

Par conséquent, $P(\emptyset) \geq 2P(\emptyset)$ et comme $P(\emptyset) \geq 0$, ceci entraîne $P(\emptyset) = 0$.

Preuve de 2. Soient $A_1, \dots, A_n$, n événements de $\mathcal{F}$ deux à deux disjoints. Pour $j > n$, posons $A_j = \emptyset$. On a ainsi une suite infinie $(A_j)_{j\geq 1}$ d'événements deux à deux disjoints. En utilisant la σ-additivité, on obtient alors :

$$P\big(\bigcup_{j=1}^{n} A_j\big) = P\big(\bigcup_{j\in\mathbb{N}^*} A_j\big) = \sum_{j=1}^{n} P(A_j) + \sum_{j=n+1}^{+\infty} P(A_j).$$

D'après *1*, la somme pour $j \geq n+1$ vaut 0, ceci prouve *2(b)*. Bien sûr *2(a)* n'est que le cas particulier $n = 2$.

[5] Ce qui signifie : $\forall n \geq 1$, $B_n \subset B_{n+1}$ et $B = \bigcup_{n\geq 1} B_n$.

[6] Ce qui signifie : $\forall n \geq 1$, $C_{n+1} \subset C_n$ et $C = \bigcap_{n\geq 1} C_n$.

Preuve de 3. Prendre $B = A^c$ dans *2 (a)* et utiliser (i).
Preuve de 4. Si $A \subset B$, alors $B = A \cup (B \cap A^c)$ et cette réunion est disjointe. D'après *2 (a)* on a $P(B) = P(A) + P(B \cap A^c)$ et comme $P(B \cap A^c) \geq 0$, on en déduit $P(B) \geq P(A)$.
Preuve de 5. On a les décompositions suivantes en unions disjointes :

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B), \\ A &= (A \cap B^c) \cup (A \cap B), \\ B &= (A \cap B) \cup (A^c \cap B). \end{aligned}$$

En utilisant l'additivité on en déduit :

$$\begin{aligned} P(A \cup B) &= P(A \cap B^c) + P(A \cap B) + P(A^c \cap B) \\ &= \big[P(A \cap B^c) + P(A \cap B)\big] + \big[P(A \cap B) + P(A^c \cap B)\big] \\ &\quad -P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Preuve de 6. Il suffit de prouver *6(a)*, la propriété *6(b)* s'en déduit en appliquant *6(a)* à la suite d'événements $B_n = C_n^c$. Admettons, pour l'instant, que pour tout $n \geq 1$, B_n vérifie la décomposition en union disjointe :

$$B_n = B_0 \cup \left(\bigcup_{i=1}^{n} (B_i \setminus B_{i-1}) \right).$$

En écrivant la réunion infinie des B_n à l'aide de cette décomposition et en « effaçant » toutes les répétitions des $B_i \setminus B_{i-1}$, on en déduit immédiatement que B vérifie la décomposition en union disjointe :

$$B = B_0 \cup \left(\bigcup_{i \in \mathbb{N}^*} (B_i \setminus B_{i-1}) \right).$$

Passant aux probabilités, ces deux décompositions nous donnent :

$$\begin{aligned} P(B_n) &= P(B_0) + \sum_{i=1}^{n} P(B_i \setminus B_{i-1}), \\ P(B) &= P(B_0) + \sum_{i=1}^{+\infty} P(B_i \setminus B_{i-1}). \end{aligned}$$

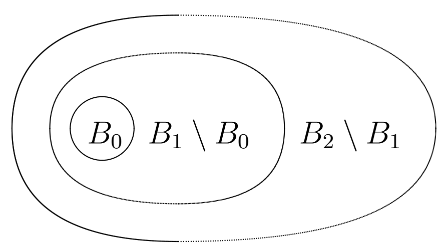
Comme cette série converge, sa somme est la limite de la suite de ses sommes partielles de rang n, ce qui s'écrit :

$$P(B) = \lim_{n \to +\infty} \left\{ P(B_0) + \sum_{i=1}^{n} P(B_i \setminus B_{i-1}) \right\} = \lim_{n \to +\infty} P(B_n).$$

Ainsi pour compléter la preuve, il ne reste plus qu'à justifier la décomposition de B_n. Posons :

$$D_n = B_0 \cup \Big(\mathop{\cup}_{i=1}^{n} (B_i \setminus B_{i-1}) \Big).$$

Pour montrer que $B_n = D_n$, il suffit de montrer que $D_n \subset B_n$ et $B_n \subset D_n$. La première inclusion est évidente puisque pour tout $i \leq n$, $B_i \setminus B_{i-1} \subset B_i \subset B_n$. Pour prouver l'inclusion inverse, on note ω un élément *quelconque* de B_n et on montre que ω appartient à D_n. Soit $i_0 = i_0(\omega)$ le plus petit des indices i tels que $\omega \in B_i$. Comme cet ensemble d'indices contient au moins n, on a $0 \leq i_0 \leq n$. Si $i_0 = 0$, $\omega \in B_0$ et comme $B_0 \subset D_n$, $\omega \in D_n$. Si $i_0 \geq 1$, par la définition même de i_0, on a $\omega \in B_{i_0}$ et $\omega \notin B_{i_0-1}$, donc $\omega \in B_{i_0} \setminus B_{i_0-1}$ et comme $i_0 \leq n$, $B_{i_0} \setminus B_{i_0-1} \subset D_n$ donc $\omega \in D_n$. Le raisonnement précédent étant valable pour tout ω de B_n, on en déduit $B_n \subset D_n$.



Preuve de 7(a). D'après *5* :

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B),$$

car $P(A \cap B) \geq 0$.
Preuve de 7(b). On remarque que pour tout $n \geq 1$ on a :

$$\mathop{\cup}_{i=0}^{n} A_i = \mathop{\cup}_{i=0}^{n} B_i,$$

où les B_i sont des événements deux à deux disjoints définis comme suit :

$$\begin{aligned} B_0 = A_0, \quad B_1 &= A_1 \cap B_0^c, \quad B_2 = A_2 \cap (B_0 \cup B_1)^c, \ldots \\ \ldots B_n &= A_n \cap (B_0 \cup B_1 \cup \ldots B_{n-1})^c, \ldots \end{aligned}$$

Par additivité :

$$P\Big(\mathop{\cup}_{i=0}^{n} A_i \Big) = P\Big(\mathop{\cup}_{i=0}^{n} B_i \Big) = \sum_{i=0}^{n} P(B_i).$$

Par construction pour tout i, $B_i \subset A_i$, d'où $P(B_i) \leq P(A_i)$ et :

$$P\Big(\overset{n}{\underset{i=0}{\cup}} A_i \Big) = \sum_{i=0}^{n} P(B_i) \leq \sum_{i=0}^{n} P(A_i).$$

<u>*Preuve de 7(c).*</u> Posons pour tout $n \geq 1$,

$$D_n = \overset{n}{\underset{i=0}{\cup}} A_i, \qquad D = \underset{n\geq 1}{\cup} D_n = \underset{i\in\mathbb{N}}{\cup} A_i.$$

La suite $(D_n)_{n\geq 1}$ est *croissante* et a pour limite D. Donc d'après *6(a)*, $P(D_n) \uparrow P(D)$ $(n \to +\infty)$. D'après *7(b)* on a :

$$\forall n \geq 1, \quad P(D_n) \leq \sum_{i=0}^{n} P(A_i).$$

Les deux membres de cette inégalité étant les termes généraux de deux suites croissantes de réels positifs, on obtient en passant à la limite quand n tend vers $+\infty$:

$$P(\underset{i\in\mathbb{N}}{\cup} A_i) = P(D) \leq \sum_{i=0}^{+\infty} P(A_i).$$

Ce qui prouve *7(c)*. Remarquons que les sommes partielles de la série convergent dans $\mathbb{R}^+ \cup \{+\infty\}$. Bien sûr l'inégalité obtenue n'a d'intérêt que lorsque la série de terme général $P(A_i)$ converge et a une somme strictement inférieure à 1. ∎

Le calcul de probabilités de réunions ou d'intersection est une question cruciale. La propriété 5 montre qu'en général on ne peut pas calculer $P(A \cup B)$ à partir de la *seule connaissance* de $P(A)$ et $P(B)$ et qu'on se heurte à la même difficulté pour $P(A \cap B)$ (voir l'exercice 1.6). Le calcul des probabilités d'intersections sera discuté au chapitre 2. Pour les probabilités de réunions, on peut se demander comment se généralise la propriété 5 lorsqu'on réunit plus de deux évènements. Il est facile de vérifier (faites-le !) que :

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

Le cas général est donné par la formule de Poincaré ci-dessous qui exprime $P(A_1 \cup \cdots \cup A_n)$ à l'aide des probabilités de *toutes les intersections* des A_i : 2 à 2, 3 à 3, etc. L'exercice 1.12 sur le problème des appariements présente une application de cette formule.

Proposition 1.3 (Formule de Poincaré)
Pour tout entier $n \geq 2$ et tous évènements $A_1, \ldots, A_n$:

$$P\Big(\overset{n}{\underset{i=1}{\cup}} A_i \Big) = \sum_{i=1}^{n} P(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1\leq i_1<i_2<\ldots<i_k\leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}). \quad (1.1)$$

Preuve : On raisonne par récurrence (voir aussi l'exercice 5.6 du chapitre 5 pour une autre méthode). La formule est vraie pour $n = 2$, car dans ce cas elle se réduit à

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.2}$$

Supposons la formule de Poincaré vraie au rang n (plus précisément on suppose que pour *toute* suite de n évènements $A_1, \dots, A_n$, l'égalité (1.1) est vérifiée). Pour en déduire qu'elle est alors vraie au rang $n + 1$, il nous faut calculer $P\Big(\overset{n+1}{\underset{i=1}{\cup}} A_i\Big)$. On commence par appliquer (1.2) avec $A = \overset{n}{\underset{i=1}{\cup}} A_i$ et $B = A_{n+1}$. On obtient ainsi :

$$\begin{aligned} P\Big(\overset{n+1}{\underset{i=1}{\cup}} A_i\Big) &= P\Big(\overset{n}{\underset{i=1}{\cup}} A_i\Big) + P(A_{n+1}) - P\Big(\Big[\overset{n}{\underset{i=1}{\cup}} A_i\Big] \cap A_{n+1}\Big) \\ &= P\Big(\overset{n}{\underset{i=1}{\cup}} A_i\Big) + P(A_{n+1}) - P\Big(\overset{n}{\underset{i=1}{\cup}} (A_i \cap A_{n+1})\Big). \end{aligned}$$

On applique maintenant l'hypothèse de récurrence (formule de Poincaré (1.1)) d'abord avec les n évènements $A_1, \dots, A_n$ puis avec les n évènements $A'_1, \dots, A'_n$, où l'on a posé $A'_i := A_i \cap A_{n+1}$. Il vient :

$$\begin{aligned} P\Big(\overset{n+1}{\underset{i=1}{\cup}} A_i\Big) = \; & \sum_{i=1}^{n} P(A_i) + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k}) \\ & + P(A_{n+1}) \\ & - \sum_{i=1}^{n} P(A'_i) - \sum_{j=2}^{n} (-1)^{j+1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} P(A'_{i_1} \cap \dots \cap A'_{i_j}) \\ = \; & \sum_{i=1}^{n+1} P(A_i) && (1.3) \\ & + \sum_{k=2}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k}) && (1.4) \\ & + (-1)^{2+1} \sum_{i=1}^{n} P(A_i \cap A_{n+1}) && (1.5) \\ & + \sum_{j=2}^{n} (-1)^{(j+1)+1} \sum_{1 \le i_1 < i_2 < \ldots < i_j \le n} P(A_{i_1} \cap \dots \cap A_{i_j} \cap A_{n+1}) && (1.6) \end{aligned}$$

Comparons ce résultat avec ce que l'on espère trouver, c'est-à-dire avec

$$\sum_{i=1}^{n+1} P(A_i) + \underbrace{\sum_{k=2}^{n+1} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \ldots < i_k \le n+1} P(A_{i_1} \cap \dots \cap A_{i_k})}_{=:T_{n+1}}.$$

Cela revient à vérifier que T_{n+1} est égal à la somme des lignes (1.4) à (1.6) ci-dessus. Partageons T_{n+1} en deux blocs comme suit. Le premier bloc regroupe tous les termes tels que $i_k < n+1$ (et donc $i_k \leq n$ et $k \leq n$). On le retrouve exactement à la ligne (1.4). Le deuxième bloc regroupe tous les termes pour lesquels $i_k = n+1$. Dans ce bloc, la somme des termes pour lesquels $k = 2$ se retrouve ligne (1.5). Il reste alors la somme des termes pour lesquels $3 \leq k \leq n+1$ et $i_k = n+1$ (donc $i_{k-1} \leq n$). Cette somme est exactement le contenu de la ligne (1.6), comme on peut le voir en faisant le changement d'indice $k = j+1$ dans (1.6). Ceci achève la récurrence. ■

1.4 Exemples

Nous examinons maintenant quelques exemples élémentaires.

Exemple 1.3 *On effectue une partie de pile ou face en trois coups. Quelle est la probabilité d'obtenir pile aux premier et troisième lancers ?*

On peut modéliser cette expérience en prenant $\Omega = \{\mathtt{f}, \mathtt{p}\}^3$ et pour famille d'événements observables $\mathcal{F} = \mathcal{P}(\Omega)$ l'ensemble de toutes les parties[7] de Ω. La pièce étant supposée symétrique, nous n'avons *a priori* pas de raison de supposer que l'un des 8 triplets de résultats possibles soit favorisé ou défavorisé par rapport aux autres. Nous choisirons donc P de sorte que tous les événements élémentaires aient même probabilité (hypothèse d'équiprobabilité), soit :

$$\forall \omega \in \Omega, \quad P(\{\omega\}) = \frac{1}{\operatorname{Card} \Omega} = \frac{1}{2^3}.$$

L'événement B dont on veut calculer la probabilité s'écrit :

$$B = \{(\mathtt{p},\mathtt{f},\mathtt{p}); (\mathtt{p},\mathtt{p},\mathtt{p})\}.$$

D'où :

$$P(B) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

■

Exemple 1.4 *On fait remplir un questionnaire à 20 questions binaires. Quelle est la probabilité qu'un candidat répondant au hasard obtienne au moins* 16 *bonnes réponses ?*

On choisit ici :

$$\Omega = \{\text{oui}, \text{non}\}^{20}, \qquad \mathcal{F} = \mathcal{P}(\Omega).$$

[7] Lorsque Ω est fini, il est toujours possible de faire ce choix.

Si le candidat répond complètement au hasard, on peut considérer que chacune des 2^{20} grilles de réponses possibles a la même probabilité d'apparaître (hypothèse d'équiprobabilité sur Ω). Pour tout $B \subset \Omega$, on a alors :

$$P(B) = \frac{\mathrm{Card} B}{\mathrm{Card} \Omega}.$$

En particulier pour $B = \{$obtention d'au moins 16 bonnes réponses$\}$,

$$P(B) = \frac{C_{20}^{16} + C_{20}^{17} + C_{20}^{18} + C_{20}^{19} + C_{20}^{20}}{2^{20}} = \frac{6196}{2^{20}} \simeq 0,006.$$

$\blacksquare$

Exemple 1.5 (Contrôle de production)
On prélève au hasard un échantillon de k pièces dans une production totale de N pièces comprenant en tout n pièces défectueuses. Quelle est la probabilité de :

$$A_j = \{\textit{il y a exactement } j \textit{ pièces défectueuses dans l'échantillon}\}?$$

On prend pour Ω l'ensemble de toutes les parties à k éléments d'un ensemble à N éléments (ensemble de tous les échantillons possibles de taille k), $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité sur Ω. Il suffit alors de dénombrer tous les échantillons ayant exactement j pièces défectueuses. Un tel échantillon se construit en prenant j pièces dans le sous-ensemble des défectueuses (C_n^j choix possibles) et en complétant par $k - j$ pièces prises dans le sous-ensemble des non-défectueuses (C_{N-n}^{k-j} choix possibles). On en déduit :

$$P(A_j) = \frac{C_n^j C_{N-n}^{k-j}}{C_N^k} \quad \text{si} \quad \left\{ \begin{array}{rcl} 0 \leq j & \leq & n, \\ 0 \leq j & \leq & k, \\ k - j & \leq & N - n. \end{array} \right.$$

Si l'une de ces trois conditions n'est pas vérifiée, $P(A_j) = 0$. $\blacksquare$

Ces trois premiers exemples concernent le cas où l'on peut choisir Ω fini. On peut caractériser complètement les probabilités sur les ensembles finis. C'est l'objet de l'énoncé suivant.

Proposition 1.4 *Soit $\Omega = \{\omega_1, \dots, \omega_n\}$ un ensemble fini à n éléments. La donnée d'une probabilité P sur $(\Omega, \mathcal{P}(\Omega))$ équivaut à celle d'une suite finie $(p_i)_{1 \leq i \leq n}$ de réels tels que :*

$$\sum_{i=1}^{n} p_i = 1, \qquad p_i \geq 0 \quad (1 \leq i \leq n)$$

et des n égalités

$$P(\{\omega_i\}) = p_i, \quad 1 \leq i \leq n.$$

Preuve : Si P est une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, on a $P(\Omega) = 1$ et comme Ω est la réunion *disjointe* des $\{\omega_i\}$ $(1 \leq i \leq n)$, on a :

$$\sum_{i=1}^{n} P(\{\omega_i\}) = 1.$$

Si on *définit* les réels p_i par $p_i := P(\{\omega_i\})$, il est clair qu'ils vérifient les conditions requises.

Réciproquement, *donnons nous* une suite finie $p_1, \dots, p_n$ de réels positifs de somme 1. *Définissons* la fonction d'ensemble Q sur $(\Omega, \mathcal{P}(\Omega))$ de la manière suivante :

(a) $Q(\emptyset) := 0$.

(b) $\forall i \in \{1, \dots, n\}$, $Q(\{\omega_i\}) := p_i$.

(c) $\forall B \in \mathcal{P}(\Omega)$ (c.à.d. $\forall B \subset \Omega$), $Q(B) := \displaystyle\sum_{\omega_i \in B} p_i$.

Remarquons que (a) et (b) sont deux cas particuliers de (c) si l'on convient qu'une somme indexée par l'ensemble vide (donc ayant 0 termes) vaut 0 et une somme indexée par un singleton donc ayant un seul terme vaut ce terme. Nous allons vérifier que la fonction d'ensembles Q ainsi définie est bien une probabilité sur $(\Omega, \mathcal{P}(\Omega))$, autrement dit que Q vérifie les conditions (i) $(Q(\Omega) = 1)$ et (ii) (σ-additivité) de la définition 1.1.

Vérification de (i) : En utilisant la définition de Q et l'hypothèse sur la somme des p_i :

$$Q(\Omega) = \sum_{\omega_i \in \Omega} p_i = \sum_{i=1}^{n} p_i = 1.$$

Vérification de (ii) : Soit $(A_j)_{j \geq 1}$ une suite de parties de Ω *deux à deux disjointes*. Comme Ω est fini, seul un nombre fini m de A_j sont non vides $(m \leq n)$. Notons les $A_{j_1}, \dots, A_{j_k} \dots, A_{j_m}$. Soit A leur réunion :

$$\bigcup_{j \in \mathbb{N}^*} A_j = \bigcup_{k=1}^{m} A_{j_k} = A.$$

D'après la définition de Q :

$$Q(A) = \sum_{\omega_i \in A} p_i.$$

Dans cette *somme finie*, regroupons en un même paquet tous les p_i indexés par des ω_i appartenant au même A_{j_k} :

$$Q(A) = \sum_{k=1}^{m} \sum_{\omega_i \in A_{j_k}} p_i = \sum_{k=1}^{m} Q(A_{j_k}).$$

Finalement

$$Q\Big(\bigcup_{j\in\mathbb{N}^*} A_j\Big) = Q(A) = \sum_{k=1}^{m} Q(A_{j_k}) = \sum_{k=1}^{+\infty} Q(A_j).$$

Ainsi la fonction d'ensemble Q vérifie la propriété de σ-additivité, c'est bien une probabilité sur $(\Omega, \mathcal{P}(\Omega))$. ■

Remarque : Lorsque Ω est fini, la façon la plus simple de construire une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est de choisir $p_i = 1/n$ pour tout i ($\operatorname{Card}\Omega = n$). On parle alors d'*équiprobabilité* ou de *probabilité uniforme* sur $(\Omega, \mathcal{P}(\Omega))$. C'est la modélisation qui s'impose naturellement lorsqu'on n'a pas de raison de penser *a priori* qu'un résultat élémentaire de l'expérience soit favorisé ou défavorisé par rapport aux autres. La situation est radicalement différente lorsque Ω est infini *dénombrable* (par exemple $\mathbb{N}$ ou une suite infinie d'éléments distincts). Sur un tel ensemble, *il ne peut pas y avoir d'équiprobabilité*. Imaginons que l'on veuille tirer une boule *au hasard* dans une urne contenant une infinité de boules ayant chacune un numéro entier distinct (et une boule par entier). Soit ω_i l'événement tirage de la boule numérotée i ($i \in \mathbb{N}$) et p_i sa probabilité. Par σ-additivité, on doit toujours avoir :

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

Mais si tous les p_i sont égaux, la série ci-dessus contient une infinité de termes tous égaux à p_0. Sa somme vaut alors $+\infty$ si $p_0 > 0$ ou 0 si $p_0 = 0$, il y a donc une contradiction.

Proposition 1.5 *Soit $\Omega = \{\omega_0, \omega_1, \dots, \omega_i, \dots\}$ un ensemble infini dénombrable (c'est-à-dire en bijection avec $\mathbb{N}$). La donnée d'une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ équivaut à la donnée d'une suite $(p_i)_{i\in\mathbb{N}}$ de réels tels que :*

$$\sum_{i=0}^{+\infty} p_i = 1, \qquad p_i \geq 0 \quad (i \in \mathbb{N})$$

et des égalités

$$P(\{\omega_i\}) = p_i, \quad i \in \mathbb{N}.$$

Preuve : La démonstration est analogue à celle de la proposition 1.4 en remplaçant les sommes finies par les séries. En particulier pour montrer la σ-additivité de Q, on utilise la propriété de sommation par paquets des séries qui est toujours vérifiée pour les séries *à termes positifs*. Les détails sont laissés au lecteur. ■

Exemple 1.6 (Une probabilité définie sur $\big(\mathbb{N}, \mathcal{P}(\mathbb{N})\big)$)
Soit a un réel strictement positif fixé. On pose :

$$\forall k \in \mathbb{N}, \qquad p_k = \frac{\mathrm{e}^{-a} a^k}{k!}.$$

On remarque que p_k est le terme général positif d'une série convergente :

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

Pour tout $A \subset \mathbb{N}$, on définit :

$$P(A) = \sum_{k \in A} p_k.$$

D'après la proposition 1.5, P est une probabilité sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. On l'appelle *loi de Poisson de paramètre a*. Calculons par exemple $P(2\mathbb{N})$ où $2\mathbb{N}$ désigne l'ensemble des entiers pairs.

$$P(2\mathbb{N}) = \sum_{k \in 2\mathbb{N}} p_k = \sum_{l=0}^{+\infty} \frac{\mathrm{e}^{-a} a^{2l}}{(2l)!} = \mathrm{e}^{-a} \operatorname{ch} a = \frac{1}{2}(1 + \mathrm{e}^{-2a}).$$

Une conséquence de ce résultat est : si l'on tire un nombre entier au hasard suivant une loi de Poisson, la probabilité qu'il soit pair est strictement supérieure à $\frac{1}{2}$.

1.5 Remarques sur le choix d'un modèle

Envisageons la situation suivante « on jette deux dés … ». Un événement élémentaire associé à cette expérience est la sortie de deux nombres entiers distincts ou non compris entre 1 et 6.

Une première modélisation consiste à choisir $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, à prendre comme ensemble d'événements observables $\mathcal{F} = \mathcal{P}(\Omega)$ et à attribuer à chaque événement élémentaire $\{(i, j)\}$ la probabilité $1/36$ (Modèle (1)).

Il est commode de représenter Ω sous la forme d'un tableau à 36 cases correspondant chacune à un événement élémentaire. On peut alors définir la probabilité comme un rapport d'aires (Modèle (1′)) :

$$P(B) = \text{Aire de } B / \text{Aire de } \Omega.$$

1	2	3	4	5	6	
						1
						2
						3
						4
						5
						6

Ce modèle (1) ou (1′) est accepté d'autant plus facilement qu'on aura précisé que les deux dés sont distinguables (par exemple un dé rouge et un vert, ou deux dés blancs lancés chacun sur une table différente). On peut ainsi distinguer l'événement $\{(2,3)\}$ de l'événement $\{(3,2)\}$ où la première composante désigne le résultat du dé rouge. A la question : quelle est la probabilité d'obtenir un 2 et un 3 et quelle est celle d'obtenir un double 6 ? On répondra *naturellement* :

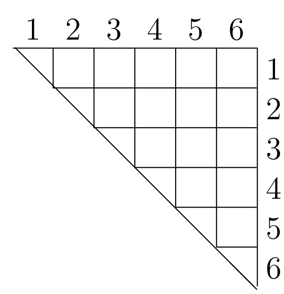
$$P(2 \text{ et } 3) = P(\{(2,3)\} \cup \{(3,2)\}) = 1/36 + 1/36 = 1/18,$$
$$P(6 \text{ et } 6) = P(\{(6,6)\}) = 1/36.$$

Supposons maintenant que les deux dés ne sont plus distinguables (par exemple jet de deux dés blancs sur une même table ou jet de deux dés de couleurs différentes, l'observation étant notée sans tenir compte de la couleur)

Voici une modélisation possible : on garde le même Ω mais on change d'ensemble d'événements observables. On ne considère comme événements observables que les parties symétriques de Ω (i.e. invariantes par $(i,j) \mapsto (j,i)$). Les événements élémentaires sont alors de deux types : les doubles et les paires de deux chiffres distincts. $\mathcal{F}$ est alors constituée de 2^{21} événements, ce modèle est moins riche en information que le précédent (2^{36} événements). On peut *raisonnablement* penser que la couleur n'influe pas sur les résultats et attribuer pour rester cohérent avec notre première modélisation la probabilité 1/36 aux *doubles* et 1/18 aux paires de deux chiffres distincts. On voit ainsi un exemple de situation pourtant très simple où *il n'y a pas* équiprobabilité des événements élémentaires. (Modèle (2))

Remarquons là aussi que l'on peut donner une représentation géométrique de ce modèle en repliant le carré le long de sa diagonale $i = j$.

Les événements élémentaires sont maintenant représentés par un carré ou un triangle et leur probabilité est définie comme un rapport d'aires (loi uniforme sur le triangle rectangle de côté 6). (Modèle (2′))

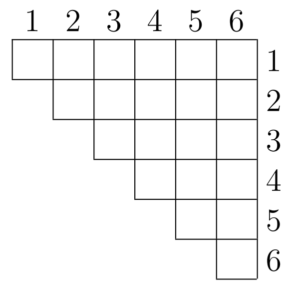


Un troisième modèle peut être proposé et il sera souvent adopté implicitement par des débutants à qui on aura dit : *on jette deux dés de même couleur. . .* On considère 21 événements élémentaires : les 6 doubles et les 15 paires à deux chiffres distincts.

$$\Omega = \{\underbrace{1,\ 2, \dots, 6}_{\text{doubles}},\ \underbrace{12, \dots, 16,\ 23, \dots 26,\ 34, \dots,\ 56}_{\text{distincts}}\}$$

On prend comme ensemble d'événements observables $\mathcal{F} = \mathcal{P}(\Omega)$ ensemble des parties de Ω. On définit la probabilité P par l'hypothèse d'équiprobabilité (Modèle (3)). On obtient une représentation géométrique équivalente à l'aide du schéma ci dessous :

Les événements élémentaires sont représentés par un carré et la probabilité est définie comme un rapport d'aires (loi uniforme sur la figure ci-contre). (Modèle (3′))



Avec ce modèle, la probabilité d'obtenir un double six est la même que celle d'obtenir un deux et un trois et vaut $\frac{1}{21}$.

On peut imaginer toute une liste d'excellents arguments qui militent en faveur des modèles (1) et (2) contre le modèle (3), par exemple : « on jette deux dés de couleurs différentes et on les filme avec deux caméras, une couleur et l'autre noir et blanc... », « il y a deux façons d'obtenir 2 et 3 et une seule de faire un double 6 ».

Tous ces arguments relèvent d'une analyse *a priori* de l'expérience et pas de la théorie mathématique des probabilités. Chacun des modèles présentés ci-dessus a sa cohérence comme objet mathématique. La question pertinente est : *parmi ces modèles, lequel (lesquels ?) représente(nt) le mieux la réalité ?*

Pour un physicien, ce qui fait la valeur d'un modèle est sa capacité à permettre la prévision du résultat d'expériences. Ainsi certaines expériences sont « expliquées » par la théorie de la relativité alors que leur résultat ne s'accorde pas avec la théorie de la mécanique classique.

Si l'on se place dans le cadre des modèles (1) et (2), la loi forte des grands nombres nous dit que si l'on jette les dés une infinité de fois, la fréquence d'apparition du double six va converger[8] vers $\frac{1}{36}$ tandis qu'avec le modèle (3) cette fréquence convergera vers $\frac{1}{21}$. La réalisation d'un grand nombre de lancers permet donc ici de constater que les modèles (1) et (2) rendent mieux compte de la réalité que le (3). Une question importante est alors : que faut-il entendre par *grand nombre* ? Cette question sera discutée ultérieurement.

1.6 Exercices

Ex 1.1. *Partie de dés entre Alice et Bruno* (cf. exemple 1.1)
D'après l'étude faite à l'exemple 1.1, on a :

$$A \cup B = \bigcup_{n \in \mathbb{N}^*} \Big(\Big(\bigcap_{j=1}^{n-1} S_j^c \Big) \cap S_n \Big).$$

[8] En un sens qui sera précisé dans le chapitre sur la loi des grands nombres.

La règle de calcul automatique du complémentaire nous donne alors :

$$\big(A \cup B\big)^c = A^c \cap B^c = \bigcap_{n\in\mathbb{N}^*} \Big(\Big(\mathop{\cup}_{j=1}^{n-1} S_j \Big) \cup S_n^c \Big).$$

D'autre part comme A, B et D forment une partition de Ω, on a

$$\big(A \cup B\big)^c = D = \bigcap_{j\in\mathbb{N}^*} S_j^c.$$

On en déduit :

$$\bigcap_{n\in\mathbb{N}^*} \Big(\Big(\mathop{\cup}_{j=1}^{n-1} S_j \Big) \cup S_n^c \Big) = \bigcap_{j\in\mathbb{N}^*} S_j^c.$$

Proposer une vérification directe de cette formule.

Ex 1.2. On lance un dé jusqu'à la première apparition du six. Notons :

$$S_i = \{\text{Le } i\text{-ème lancer donne un six}\}, \qquad i \in \mathbb{N}.$$

Écrire à l'aide des événements S_i et S_i^c l'événement

$$A = \{\text{La première apparition du six a lieu } \textit{après} \text{ le 5-ème lancer}\}.$$

Est-ce le même événement que

$$B = \{\text{Six n'apparaît pas au cours des 5 premiers lancers}\}?$$

Ex 1.3. On effectue une suite infinie de lancers d'un dé. Pour tout $i \in \mathbb{N}^*$, on note :

$$A_i = \{\text{Obtention de l'as au } i\text{-ème lancer}\}.$$

1) Définir par une phrase ne comportant aucun vocabulaire mathématique chacun des événements :

$$E_1 = \mathop{\cap}_{i=5}^{+\infty} A_i, \quad E_2 = \Big(\mathop{\cap}_{i=1}^{4} A_i^c \Big) \cap \Big(\mathop{\cap}_{i=5}^{+\infty} A_i \Big), \quad E_3 = \mathop{\cup}_{i>4} A_i.$$

2) Ecrire à l'aide des A_i l'événement « on obtient au moins une fois l'as au-delà du n-ième lancer ».

3) On pose $C_n = \mathop{\cup}_{i>n} A_i$. Montrer que la suite (C_n) est décroissante (i.e. que pour tout $n \geq 1$, C_{n+1} est inclus dans C_n). Caractériser d'une phrase ne comportant pas de vocabulaire mathématique l'événement $C = \mathop{\cap}_{n\geq 1} C_n$.

4) Ecrire à l'aide des A_i les événements :

$$\begin{aligned} B_n &= \{\text{On n'obtient plus que des as à partir du } n\text{-ième lancer}\} \\ B &= \{\text{On n'obtient plus que des as à partir d'un certain lancer}\} \end{aligned}$$

Ex 1.4. Soient f et g deux applications de Ω dans $\mathbb{R}$. Montrer que :

$$\{\omega \in \Omega,\ f(\omega) + g(\omega) \geq \varepsilon\} \subset \{\omega \in \Omega,\ f(\omega) \geq \frac{\varepsilon}{2}\} \cup \{\omega \in \Omega,\ g(\omega) \geq \frac{\varepsilon}{2}\}$$

Ex 1.5. On considère une suite de fonctions $(f_n)_{n\geq 1} : \Omega \to \mathbb{R}, \quad \omega \mapsto f_n(\omega)$. On pose pour tout $\varepsilon > 0$ et tout $k \in \mathbb{N}^*$:

$$B_{\varepsilon,k} = \{\omega \in \Omega,\ |f_k(\omega)| < \varepsilon\}.$$

1) On fixe $\varepsilon > 0$. Ecrire à l'aide d'opérations ensemblistes sur les $B_{\varepsilon,k}$, l'ensemble :

$$A_{\varepsilon,n} = \{\omega \in \Omega,\ \forall k \geq n,\ |f_k(\omega)| < \varepsilon\}.$$

2) Même question pour :

$$A_\varepsilon = \{\omega \in \Omega,\ \exists n(\omega),\ \forall k \geq n(\omega),\ |f_k(\omega)| < \varepsilon\}.$$

3) Montrer que l'ensemble :

$$A = \{\omega \in \Omega,\ \lim_{k\to+\infty} f_k(\omega) = 0\}$$

peut s'écrire à l'aide d'opérations ensemblistes sur *une suite* d'ensembles du type $B_{\varepsilon,k}$.

Ex 1.6. Deux événements A et B sont tels que : $P(A) = P(B) = 0,75$. Quel est le maximum de $P(A \cap B)$? Quel est son minimum ?

Ex 1.7. *Inégalité de Bonferroni*

1) Si $P(A) = 0.9$ et $P(B) = 0.8$, montrer que $P(A \cap B) \geq 0.7$.

2) Montrer que pour tous événements A et B,

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

3) Généraliser :

$$\forall A_1, \dots, A_n \in \mathcal{F}, \qquad P(A_1 \cap \cdots \cap A_n) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

Ex 1.8. On suppose que P et Q sont deux probabilités sur $(\Omega, \mathcal{F})$ vérifiant $P \leq Q$, ce qui signifie : $\forall A \in \mathcal{F}$, $P(A) \leq Q(A)$. Montrer qu'alors $P = Q$ (égalité de deux fonctions d'ensembles).

Ex 1.9. Soit $(A_n)_{n\geq 1}$ une suite d'événements ayant chacun une probabilité 1 (on rappelle que $P(A_n) = 1$ n'implique pas $A_n = \Omega$). On note A leur intersection. Que peut-on dire de $P(A)$?

Ex 1.10. *Dénombrement de sommes d'entiers*
Le but de cet exercice est d'établir les deux formules de dénombrement suivantes. Pour n, $r > 1$ entiers donnés, considérons l'équation à r inconnues :

$$x_1 + x_2 + \cdots + x_r = n. \tag{1.7}$$

Une solution de cette équation est un r-uplet $(x_1, \ldots, x_r)$ dont la somme des composantes x_i $(1 \leq i \leq r)$ vaut n. Lorsque les x_i sont des réels, (1.7) a une infinité de solutions. On s'intéresse aux solutions dont *toutes* les composantes sont des entiers.

Proposition 1.6
a) L'équation (1.7) admet exactement $C_{n+r-1}^{r-1} = C_{n+r-1}^{n}$ solutions à composantes entières positives ou nulles *;*
b) si $n \geq r > 1$, l'équation (1.7) admet exactement C_{n-1}^{r-1} solutions à composantes entières strictement positives.

Voici une traduction plus concrète du problème. On dispose de n jetons *identiques* (il n'est donc pas possible de les distinguer) et de r boîtes *numérotées* de 1 à r. Le cas a) donne le nombre de façons de répartir ces jetons dans les r boîtes (certaines pouvant rester vides). Le cas b) donne le nombre de répartitions pour lesquelles aucune boîte ne reste vide. Dans cette interprétation, x_i représente le nombre de jetons déposés dans la boîte n° i.

1) Démontrez la Proposition 1.6 b) en vous aidant du codage des répartitions possibles par des chaînes de caractères, illustré par l'exemple suivant avec $n = 9$ et $r = 4$. On représente d'abord chaque jeton par le caractère 'O', *avec un espace* entre deux caractères consécutifs :

O O O O O O O O O.

Pour représenter les 4 boîtes, il suffit d'insérer 3 caractères '|', chacun dans un espace libre. Par exemple, la chaîne

O O|O O O|O|O O O

code la répartition avec 2 jetons dans la première boîte, 3 dans la deuxième, 1 dans la troisième et 3 dans la quatrième, autrement dit la solution $(x_1, x_2, x_3, x_4) = (2, 3, 1, 3)$.

2) Prouvez la Proposition 1.6 a), en utilisant le codage suivant illustré à nouveau dans le cas $n = 9$ et $r = 4$. Le caractère '|' représente à la fois le début et la fin d'une boîte (sauf s'il est le premier ou le dernier de la chaîne). Il nous

en faut donc 5 pour représenter 4 boîtes. On représente encore chaque jeton par un 'O', mais cette fois on ne laisse plus d'espace entre les caractères de la chaîne. L'exemple de la question précédente sera alors codé par la chaîne de 14 caractères sans espaces :

$$|\mathsf{OO}|\mathsf{OOO}|\mathsf{O}|\mathsf{OOO}|.$$

Ceci permet de représenter une boîte vide par deux caractères '|' consécutifs. Par exemple les chaînes

$$|\mathsf{OO}||\mathsf{OOOO}|\mathsf{OOO}|, \qquad |||\mathsf{OOOO}|\mathsf{OOOOO}|,$$

représentent respectivement les solutions $(2, 0, 4, 3)$ et $(0, 0, 4, 5)$.

3) Montrer que l'on peut passer du cas a) au cas b) par un simple changement d'inconnue et utiliser cette remarque pour contrôler les résultats précédents.

Ex 1.11. *En attendant le métro* ↑ 1.10
Prologue. Dans une station de métro est disposée une rangée de 16 sièges. Un sociologue observe la répartition des voyageurs assis sur cette rangée. Il représente son observation par une chaîne de caractères 'O' (pour « occupé ») et 'L' (pour « libre ») :

$$\mathsf{OLOLLOLOLLLOLOLO}$$

Notre sociologue se demande si les voyageurs ont choisi leur place au hasard parmi les places disponibles, ou s'ils ont cherché à s'isoler. Si les voyageurs choisissent au hasard parmi les places disponibles, toutes les répartitions des 7 personnes sur les 16 places sont équiprobables. Pour se faire une idée, on compte les *séries*. Une série est ici un bloc de lettres identiques encadré (sauf s'il est en début ou en fin) par des lettres différentes. Dans l'observation réalisée, il y donc 13 séries :

$$\mathsf{O,\ L,\ O,\ LL,\ O,\ L,\ O,\ LLL,\ O,\ L,\ O,\ L,\ O.}$$

dont 7 séries de O, ce qui est le maximum possible pour cette lettre.

1) Une chaîne comportant m lettres L et n lettres O est construite au hasard. On se propose de calculer la probabilité qu'elle comporte exactement r séries de O. On pourra utiliser les résultats de l'exercice 1 comme suit. Pour $i = 1, \dots, r$, notons x_i le nombre de lettres O de la i-ème série de O (tous les x_i sont strictement positifs). Soit y_1 le nombre (éventuellement nul) de L avant la première série de O, y_{r+1} le nombre (éventuellement nul) de L après la dernière série de O. Pour $2 \leq i \leq r$, y_i désigne le nombre strictement positif de lettres L dans la série qui précède la i-ème série de O.

$$\underbrace{\mathsf{L\dots L}}_{y_1}\underbrace{\mathsf{O\dots O}}_{x_1}\underbrace{\mathsf{L\dots L}}_{y_2}\underbrace{\mathsf{O\dots O}}_{x_2}\dots\dots\underbrace{\mathsf{L\dots L}}_{y_r}\underbrace{\mathsf{O\dots O}}_{x_r}\underbrace{\mathsf{L\dots L}}_{y_{r+1}}$$

Les r-uplets $(x_1, \dots, x_r)$ possibles sont les solutions à composantes strictement positives de l'équation $x_1 + \dots x_r = n$. On se ramène à une condition du même type pour les y_i en faisant le changement de variable

$$z_1 = y_1 + 1, \quad z_{r+1} = y_r + 1, \qquad z_i = y_i \ (2 \leq i \leq r).$$

Montrer que pour $1 \leq r \leq n$ et $1 \leq r \leq m+1$, la probabilité p_r d'avoir r séries de O est

$$p_r = \frac{C_{m+1}^{r} C_{n-1}^{r-1}}{C_{m+n}^{n}}.$$

2) Pouvez vous aider le sociologue du prologue ?

Ex 1.12. *Le problème des appariements*
Dans le vestiaire d'une piscine, chaque nageur range ses vêtements sur un cintre. Il le dépose au guichet où un employé équipe le cintre d'un bracelet rouge numéroté et remet au nageur un bracelet jaune portant le même numéro. Ainsi, à la fin de la séance, le nageur peut récupérer ses affaires en échange du bracelet. Avant l'ouverture au public, les bracelets sont rangés sur un tableau à N crochets supportant chacun un bracelet rouge et un jaune de même numéro[9].

Deux gamins turbulents s'introduisent dans le vestiaire avant l'ouverture. En se battant ils renversent le tableau portant les bracelets. Pour ne pas être découverts, ils les remettent en place en prenant bien soin de placer sur chaque crochet un bracelet jaune et un rouge, mais sans tenir compte des numéros.

A l'ouverture, N nageurs se présentent et devant l'affluence, l'employé remet à chacun son bracelet jaune sans prendre le temps de vérifier les numéros. On se propose de calculer les probabilités des événements :

$$E_{N,k} = \{\text{exactement } k \text{ nageurs retrouvent leurs affaires}\}.$$

On choisit comme espace probabilisé Ω_N ensemble des toutes les permutations (i.e. bijections) sur $\{1, \dots, N\}$ muni de l'équiprobabilité P_N. On notera :

$$B_i = \{\text{le } i\text{-ème nageur retrouve ses affaires}\}.$$

1) Pour $j \leq N$ et $1 \leq i_1 < i_2 < \dots < i_j \leq N$, calculer $P_N(B_{i_1} \cap \dots \cap B_{i_j})$. En déduire :

$$\sum_{i_1 < i_2 < \dots < i_j} P_N(B_{i_1} \cap \dots \cap B_{i_j}) = \frac{1}{j!}.$$

2) Calculer $P_N(E_{N,0})$.

3) On fixe k nageurs retrouvant leurs affaires. Exprimer à l'aide de $P_{N-k}(E_{N-k,0})$ le *nombre* de permutations des $N-k$ autres nageurs telles qu'aucun d'eux ne retrouve ses affaires. En déduire la valeur de $P_N(E_{N,k})$.

[9]Toute ressemblance avec une piscine proche du campus ne serait pas forcément aléatoire.

4) Pour k fixé, calculer : $p_k = \lim_{N\to\infty} P_N(E_{N,k})$. Montrer que la suite $(p_k)_{k\in\mathbb{N}}$ définit une probabilité P sur $\mathbb{N}$ (il s'agit d'une loi de Poisson).

5) Montrer que :

$$\forall k \in \{0,\dots,N\}, \qquad |P_N(E_{N,k}) - p_k| < \frac{1}{k!\,(N+1-k)!}.$$

En déduire :

$$\forall n \in \{0,\dots,N\}, \qquad \sum_{j=0}^{n} |P_N(E_{N,j}) - p_j| \leq \frac{\mathrm{e}}{(N+1-n)!}.$$

6) *Application* : après avoir vérifié que : $0.9994 < \sum_{j=0}^{5} p_j < 0.9995$, donner pour $N \geq 12$ quelconque des valeurs numériques approchées à 10^{-4} près des $P_N(E_{N,j})$ pour $j = 0, 1, \dots 5$ (ces valeurs ne dépendent pas de N). Même question pour la probabilité qu'au moins 6 nageurs retrouvent leurs affaires.

Ex 1.13. *Loi* ζ $\downarrow$ 2.18
Soit $a \in]1, +\infty[$. On définit le réel $\zeta(a)$ par :

$$\zeta(a) = \sum_{k=1}^{+\infty} \frac{1}{k^a}.$$

On peut alors définir une probabilité P_a sur $\mathbb{N}^*$ en posant (cf. proposition 1.5) :

$$p_k = \frac{1}{\zeta(a)k^a}, \qquad k \in \mathbb{N}^*.$$

On dit que P_a est la loi zeta de paramètre a. Si $m \in \mathbb{N}^*$, on note $m\mathbb{N}^*$ l'ensemble $\{km,\ k \in \mathbb{N}^*\}$ de ses multiples non nuls. Calculer $P_a(2\mathbb{N}^*)$. Généraliser.

Ex 1.14. *Sur la probabilité qu'un entier tiré au hasard soit un multiple. . .*
On a vu qu'il n'existe pas de probabilité uniforme sur $\mathbb{N}$ (cf. page 15). L'objet de cet exercice est de donner une réponse négative à la question moins naïve suivante : existe-t-il une probabilité P sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ telle qu'un entier tiré au hasard suivant cette loi P soit pair avec une probabilité $1/2$, multiple de 3 avec une probabilité $1/3$, multiple de 4 avec probabilité $1/4$, etc. Plus formellement, notons $n\mathbb{N}$ l'ensemble des multiples de l'entier n (y compris 0). On suppose qu'il existe une probabilité P vérifiant :

$$\forall n \in \mathbb{N}^*, \quad P(n\mathbb{N}) = \frac{1}{n}.$$

et on souhaite réfuter cette conjecture.

1) Montrer que nécessairement $P(\{0\}) = 0$.

2) Prouver la relation

$$\Big(1-\frac{1}{2}\Big)^{-1}\Big(1-\frac{1}{3}\Big)^{-1} = \sum_{m=0}^{+\infty}\sum_{k_1+k_2=m}\frac{1}{2^{k_1}3^{k_2}}.$$

3) Soit p_i le i-ème *nombre premier* ($\mathsf{p}_1 = 2$, $\mathsf{p}_2 = 3$, $\mathsf{p}_3 = 5$, ...). On pose pour $1 \leq k \leq n$:

$$\pi_{k,n} = \prod_{i=k}^{n}\Big(1-\frac{1}{\mathsf{p}_i}\Big)$$

et on se propose d'en trouver la limite lorsque n tend vers $+\infty$, k restant fixé. Justifier les écritures suivantes :

$$\begin{aligned}\prod_{i=1}^{n}\Big(1-\frac{1}{\mathsf{p}_i}\Big)^{-1} &= \sum_{l=0}^{+\infty}\sum_{k_1+\cdots+k_n=l}\frac{1}{\mathsf{p}_1^{k_1}\cdots\mathsf{p}_n^{k_n}}\\ &\geq \sum_{j=1}^{n}\frac{1}{j}.\end{aligned}$$

4) En déduire $\lim\limits_{n\to+\infty}\pi_{1,n}$, puis $\lim\limits_{n\to+\infty}\pi_{k,n}$ (k fixé).

5) Montrer que si les entiers a et b sont premiers entre eux :

$$P(\mathsf{a}\mathbb{N}\cap\mathsf{b}\mathbb{N}) = P(\mathsf{a}\mathbb{N})P(\mathsf{b}\mathbb{N}).$$

En déduire

$$P(\mathsf{a}\mathbb{N}^c\cap\mathsf{b}\mathbb{N}^c) = P(\mathsf{a}\mathbb{N}^c)P(\mathsf{b}\mathbb{N}^c).$$

6) On note pour alléger $E_i = \mathsf{p}_i\mathbb{N}$. Montrer que pour tout $n \geq 2$,

$$P\Big(\bigcap_{i=1}^{n} E_i^c\Big) = \prod_{i=1}^{n}P(E_i^c).$$

En déduire la valeur de $P\Big(\bigcap\limits_{i=k}^{+\infty} E_i^c\Big)$ pour tout $k \geq 1$.

7) Montrer que $P\Big(\bigcup\limits_{i=k}^{+\infty} E_i\Big) = 1$ pour tout k et en déduire que

$$P(\{0,1,\dots,k-1\}) = 0 \quad \text{pour tout } k,$$

ce qui est manifestement absurde (pourquoi ?).

Ex 1.15. Construire une probabilité sur P sur $(\mathbb{N}^*,\mathcal{P}(\mathbb{N}^*))$ vérifiant $P(\{n\}) > 0$ pour tout n et :

$$P(2\mathbb{N}) = \frac{1}{2},\quad P(3\mathbb{N}) = \frac{1}{3}.$$

Imaginer une expérience aléatoire correspondant à cette modélisation.

Chapitre 2

Conditionnement et indépendance

2.1 Probabilités conditionnelles

2.1.1 Introduction

Comment doit-on modifier la probabilité que l'on attribue à un événement lorsque l'on dispose d'une *information supplémentaire* ? Le concept de probabilité conditionnelle permet de répondre à cette question.

Par exemple, une fédération d'escrime regroupe N licenciés, dont N_H hommes et $N_F = N - N_H$ femmes. Il y a N_G gauchers (des deux sexes) parmi tous les licenciés. On choisit un individu au hasard. Notons :

$$\begin{aligned} G &= \{\text{l'individu choisi au hasard est gaucher}\} \\ H &= \{\text{l'individu choisi au hasard est un homme}\} \end{aligned}$$

On note $N_{G\cap H}$ le nombre d'escrimeurs hommes gauchers. On a bien sûr : $P(H) = N_H/N$ et $P(G \cap H) = N_{G\cap H}/N$. Quelle est la probabilité qu'un licencié *homme* choisi au hasard soit gaucher ? On dispose ici d'une information supplémentaire : on *sait* que l'individu choisi est un homme. En considérant les divers choix possibles d'un homme comme équiprobables, la probabilité cherchée est clairement : $N_{G\cap H}/N_H$. Ceci s'exprime aussi à l'aide de $P(H)$ et de $P(G \cap H)$ en remarquant que :

$$\frac{N_{G\cap H}}{N_H} = \frac{NP(G\cap H)}{NP(H)} = \frac{P(G\cap H)}{P(H)}.$$

Par analogie, nous donnons dans le cas général la définition formelle :

Définition 2.1 *Soit H un événement tel que $P(H) \neq 0$. Pour tout événement observable A, on définit :*

$$P(A \mid H) = \frac{P(A\cap H)}{P(H)},$$

appelée probabilité conditionnelle *de l'événement A* sous l'hypothèse H.

Remarquons que pour l'instant, il ne s'agit que d'un jeu d'écriture. On a simplement défini un réel $P(A \mid H)$ pour que :

$$P(A \cap H) = P(A \mid H)P(H).$$

Ce qui fait l'intérêt du concept de probabilité conditionnelle, c'est qu'il est souvent bien plus facile d'attribuer *directement* une valeur à $P(A \mid H)$ en tenant compte des conditions expérimentales (liées à l'*information H*) et d'en déduire ensuite la valeur de $P(A \cap H)$. Le raisonnement implicite alors utilisé est : tout espace probabilisé modélisant correctement la réalité expérimentale devrait fournir telle valeur pour $P(A \mid H)$...

Exemple 2.1 *Une urne contient r boules rouges et v boules vertes. On en tire deux l'une après l'autre, sans remise. Quelle est la probabilité d'obtenir deux rouges ?*

Notons H et A les événements :

$$H = \{\text{rouge au } 1^{\text{er}} \text{ tirage}\}, \quad A = \{\text{rouge au } 2^{\text{e}} \text{ tirage}\}.$$

Un espace probabilisé $(\Omega, \mathcal{F}, P)$ modélisant correctement cette expérience devrait vérifier :

$$\begin{aligned} P(H) &= \frac{r}{r+v}, \\ P(A \mid H) &= \frac{r-1}{r+v-1}. \end{aligned}$$

En effet, si H est réalisé, le deuxième tirage a lieu dans une urne contenant $r+v-1$ boules dont $r-1$ rouges. On en déduit :

$$P(\text{deux rouges}) = P(A \cap H) = P(A \mid H)P(H) = \frac{r-1}{r+v-1} \times \frac{r}{r+v}.$$

On aurait pu arriver au même résultat en prenant pour Ω l'ensemble des arrangements de deux boules parmi $r+v$, muni de l'équiprobabilité et en faisant du dénombrement :

$$\operatorname{card}\Omega = A_{r+v}^2 = (r+v)(r+v-1), \quad \operatorname{card}(A \cap H) = A_r^2 = r(r-1).$$

d'où :

$$P(A \cap H) = \frac{r(r-1)}{(r+v)(r+v-1)}.$$

Notons d'ailleurs que $\operatorname{card} H = r(r+v-1)$ d'où

$$P(H) = \frac{\operatorname{card} H}{\operatorname{card}\Omega} = \frac{r(r+v-1)}{(r+v)(r+v-1)} = \frac{r}{r+v}.$$

En appliquant la définition formelle de $P(A \mid H)$ on retrouve :

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)} = \frac{r(r-1)}{(r+v)(r+v-1)} \times \frac{r+v}{r} = \frac{r-1}{r+v-1},$$

ce qui est bien la valeur que nous avions attribuée *a priori* en analysant les conditions expérimentales.

Remarque : Il importe de bien comprendre que l'écriture $A \mid H$ ne désigne pas un nouvel événement[1] différent de A. Quand on écrit $P(A \mid H)$, ce que l'on a modifié, ce n'est pas l'événement A, mais la valeur numérique qui lui était attribuée par la fonction d'ensembles P. Il serait donc en fait plus correct d'écrire $P_H(A)$ que $P(A \mid H)$. On conservera néanmoins cette dernière notation essentiellement pour des raisons typographiques : $P(A \mid H_1 \cap H_2 \cap H_3)$ est plus lisible que $P_{H_1 \cap H_2 \cap H_3}(A)$.

2.1.2 Propriétés

Proposition 2.2 *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et H un événement* fixé *tel que $P(H) \neq 0$. Alors la fonction d'ensembles $P(\,.\mid H)$ définie par :*

$$P(\,.\mid H): \quad \mathcal{F} \to [0,1] \quad B \mapsto P(B \mid H)$$

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

La vérification est laissée en exercice. Une conséquence immédiate est que la fonction d'ensembles $P(\,.\mid H)$ vérifie toutes les propriétés de la proposition 1.2.

Corollaire 2.3 *La fonction d'ensemble $P(\,.\mid H)$ vérifie :*

1. *$P(\emptyset \mid H) = 0$, $P(\Omega \mid H) = 1$ et si $A \supset H$, $P(A \mid H) = 1$.*
2. *Si les A_i sont deux à deux disjoints :*

$$P\Big(\bigcup_{i=1}^{n} A_i \mid H\Big) = \sum_{i=1}^{n} P(A_i \mid H).$$

3. *Pour tout $B \in \mathcal{F}$, $P(B^c \mid H) = 1 - P(B \mid H)$.*
4. *Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$, si $A \subset B$, $P(A \mid H) \leq P(B \mid H)$.*
5. *Pour tous $A \in \mathcal{F}$ et $B \in \mathcal{F}$,*

$$P(A \cup B \mid H) = P(A \mid H) + P(B \mid H) - P(A \cap B \mid H).$$

[1] En fait cette écriture prise isolément (sans le P) *n'a pas de sens* et ne devrait *jamais* être utilisée. Le symbole $\mid$ ne représente pas une opération sur les événements qui l'entourent.

6. *Pour toute suite $(A_i)_{i\geq 1}$ d'événements :*

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

7. *Si $B_n \uparrow B$, $P(B_n \mid H) \uparrow P(B \mid H)$, $(n \to +\infty)$.*
8. *Si $C_n \downarrow C$, $P(C_n \mid H) \downarrow P(C \mid H)$, $(n \to +\infty)$.*

Nous n'avons vu jusqu'ici aucune formule permettant de calculer la probabilité d'une intersection d'événements à l'aide des probabilités de ces événements. Une telle formule *n'existe pas dans le cas général*. Les probabilités conditionnelles fournissent une méthode générale tout à fait naturelle pour calculer une probabilité d'intersection.

Proposition 2.4 (Règle des conditionnements successifs)
Si $A_1, \dots, A_n$ sont n événements tels que $P(A_1 \cap \ldots \cap A_{n-1}) \neq 0$, on a :

$$\begin{aligned} P(A_1 \cap \ldots \cap A_n) &= P(A_1)P(A_2 \,|\, A_1)P(A_3 \,|\, A_1 \cap A_2) \times \cdots \\ &\qquad \cdots \times P(A_n \mid A_1 \cap \ldots \cap A_{n-1}). \end{aligned}$$

Preuve : Pour $1 \leq i \leq n-1$, on a $\bigcap_{j=1}^{n-1} A_j \subset \bigcap_{j=1}^{i} A_j$ d'où :

$$0 < P\Big(\bigcap_{j=1}^{n-1} A_j\Big) \leq P\Big(\bigcap_{j=1}^{i} A_j\Big).$$

Donc $P\big(\bigcap_{j=1}^{i} A_j\big)$ n'est nul pour aucun $i \leq n-1$ et on peut conditionner par l'événement $\bigcap_{j=1}^{i} A_j$. Ceci légitime le calcul suivant :

$$\begin{aligned} & P(A_1)P(A_2 \,|\, A_1)P(A_3 \,|\, A_1 \cap A_2) \times \cdots \times P(A_n \mid A_1 \cap \ldots \cap A_{n-1}) \\ = \; & P(A_1) \times \frac{P(A_1 \cap A_2)}{P(A_1)} \times \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \times \cdots \times \frac{P(A_1 \cap \ldots \cap A_n)}{P(A_1 \cap \ldots \cap A_{n-1})} \\ = \; & P(A_1 \cap \ldots \cap A_n), \end{aligned}$$

après simplifications en chaîne de toutes ces fractions. ■

Les probabilités conditionnelles permettent aussi de calculer la probabilité d'un événement en conditionnant par *tous les cas possibles*. Du point de vue ensembliste, ces *cas possibles* correspondent à une *partition* de Ω. On dit qu'une famille $(H_i)_{i\in I}$ est une partition de Ω si elle vérifie les trois conditions :

- $\forall i \in I,\ H_i \neq \emptyset$.
- $\Omega = \bigcup_{i\in I} H_i$.

– Les H_i sont deux à deux disjoints ($i \neq j \Rightarrow H_i \cap H_j = \emptyset$).

Proposition 2.5 (Conditionnement par les cas possibles) [2]

(i) Si H est tel que $P(H) \neq 0$ et $P(H^c) \neq 0$, on a

$$\forall A \in \mathcal{F}, \quad P(A) = P(A \mid H)P(H) + P(A \mid H^c)P(H^c).$$

(ii) Si $H_1, \dots, H_n$ est une partition finie de Ω en événements de probabilités non nulles,

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=1}^{n} P(A \mid H_i)P(H_i).$$

(iii) Si $(H_i)_{i\in\mathbb{N}}$ est une partition de Ω telle que $\forall i \in \mathbb{N}$, $P(H_i) \neq 0$:

$$\forall A \in \mathcal{F}, \quad P(A) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

Preuve : Il suffit de vérifier (iii), les deux premières propriétés se démontrant de façon analogue. Comme $(H_i)_{i\in\mathbb{N}}$ est une partition de Ω,

$$A = A \cap \Omega = A \cap \Big(\bigcup_{i\in\mathbb{N}} H_i \Big) = \bigcup_{i\in\mathbb{N}} (A \cap H_i)$$

et cette réunion est disjointe car les H_i étant deux à deux disjoints, il en est de même pour les $(A \cap H_i)$. Par conséquent par σ-additivité :

$$P(A) = \sum_{i=0}^{+\infty} P(A \cap H_i) = \sum_{i=0}^{+\infty} P(A \mid H_i)P(H_i).$$

■

Lorsqu'on a une partition de Ω en n hypothèses ou cas possibles H_i et que l'on sait calculer les $P(H_i)$ et les $P(A \mid H_i)$, on peut se poser le problème inverse : calculer $P(H_j \mid A)$ à l'aide des quantités précédentes. La solution est donnée par la formule suivante quelquefois appelée (abusivement) formule des probabilités des causes.

Proposition 2.6 (Formule de Bayes)
Soit A un événement de probabilité non nulle. Si les événements H_i ($1 \leq i \leq n$) forment une partition de Ω et aucun $P(H_i)$ n'est nul, on a pour tout $j = 1, \dots, n$:

$$P(H_j \mid A) = \frac{P(A \mid H_j)P(H_j)}{\sum_{i=1}^{n} P(A \mid H_i)P(H_i)}.$$

[2] ou formule des probabilités totales.

Preuve : Par définition des probabilités conditionnelles on a :

$$P(H_j \mid A) = \frac{P(A \cap H_j)}{P(A)} = \frac{P(A \mid H_j)P(H_j)}{P(A)}.$$

Et il ne reste plus qu'à développer $P(A)$ en conditionnant par la partition $(H_i, 1 \leq i \leq n)$ comme à la proposition 17. ■

La même formule se généralise au cas d'une partition dénombrable. Ces formules sont plus faciles à retrouver qu'à mémoriser...

2.1.3 Quelques exemples

Exemple 2.2 *On considère deux urnes U_1 et U_2. L'urne U_1 contient r_1 boules rouges et v_1 boules vertes. L'urne U_2 contient contient r_2 boules rouges et v_2 boules vertes. On lance un dé. S'il indique le chiffre 1, on choisit l'urne U_1, sinon on choisit U_2. Dans chaque cas on effectue deux tirages avec remise dans l'urne choisie. Quelle est la probabilité d'obtenir une rouge au premier tirage ? deux rouges en tout ?*

Adoptons les notations d'événements suivantes :

$$R = \{\text{rouge au } 1^{\text{er}}\text{tirage}\}, \quad R' = \{\text{rouge au } 2^{\text{e}}\text{tirage}\},$$

$$H_1 = \{\text{choix de l'urne } U_1\}, \quad H_2 = H_1^c = \{\text{choix de l'urne } U_2\}.$$

Il est facile de calculer directement $P(R \mid H_i)$ et $P(R \cap R' \mid H_i)$ pour $i = 1, 2$. En effet, une fois l'urne U_i choisie, on a un problème classique de tirages avec remise dans la même urne que l'on peut traiter (par exemple) par le dénombrement. On a ainsi :

$$P(R \mid H_i) = \frac{r_i}{r_i + v_i}, \quad P(R \cap R' \mid H_i) = \Big(\frac{r_i}{r_i + v_i}\Big)^2.$$

La formule de conditionnement par la partition $\{H_1, H_2\}$ donne :

$$\begin{aligned} P(R) &= P(R \mid H_1)P(H_1) + P(R \mid H_2)P(H_2) \\ &= \frac{1}{6}\frac{r_1}{r_1 + v_1} + \frac{5}{6}\frac{r_2}{r_2 + v_2}. \end{aligned}$$

et

$$\begin{aligned} P(R \cap R') &= P(R \cap R' \mid H_1)P(H_1) + P(R \cap R' \mid H_2)P(H_2) \\ &= \frac{1}{6}\Big(\frac{r_1}{r_1 + v_1}\Big)^2 + \frac{5}{6}\Big(\frac{r_2}{r_2 + v_2}\Big)^2. \end{aligned}$$

■

Exemple 2.3 *Un questionnaire à choix multiples propose m réponses pour chaque question. Soit p la probabilité qu'un étudiant connaisse la réponse à une question donnée. S'il ignore la réponse, il choisit au hasard l'une des réponses proposées. Quelle est pour le correcteur la probabilité qu'un étudiant connaisse vraiment la bonne réponse lorsqu'il l'a donnée ?*

Notons :

$$\begin{aligned} B &= \{\text{l'étudiant donne la bonne réponse}\} \\ C &= \{\text{l'étudiant connaît la bonne réponse}\} \end{aligned}$$

On cherche $P(C \mid B)$. Avec ces notations, les données de l'énoncé peuvent se traduire par :

$$P(C) = p, \quad P(C^c) = 1 - p, \qquad P(B \mid C) = 1, \quad P(B \mid C^c) = \frac{1}{m}.$$

On en déduit :

$$\begin{aligned} P(C \mid B) &= \frac{P(B \cap C)}{P(B)} \\ &= \frac{P(B \mid C)P(C)}{P(B \mid C)P(C) + P(B \mid C^c)P(C^c)} \\ &= \frac{1 \times p}{1 \times p + \frac{1}{m}(1-p)} = \frac{mp}{1 + (m-1)p}. \end{aligned}$$

Pour p fixé, $P(C \mid B)$ est une fonction croissante de m, les deux bornes étant $P(C \mid B) = p$ (cas $m = 1$) et $P(C \mid B) \to 1$ ($m \to +\infty$). D'autre part pour m fixé, $P(C \mid B)$ est une fonction croissante de p. On a pour $p > 0$:

$$\frac{P(C \mid B)}{p} = \frac{m}{1 + (m-1)p} \geq 1,$$

l'égalité n'étant possible que pour $p = 1$. Tout ceci est conforme à l'intuition. ∎

Exemple 2.4 *Un test sanguin a une probabilité de* 0.95 *de détecter un certain virus lorsque celui ci est effectivement présent. Il donne néanmoins un faux résultat positif pour* 1% *des personnes non infectées. Si* 0.5% *de la population est porteuse du virus, quelle est la probabilité qu'une personne ait le virus sachant qu'elle a un test positif ?*

Notons

$$\begin{aligned} V &= \{\text{la personne testée a le virus}\}, \\ T &= \{\text{la personne testée a un test positif}\}. \end{aligned}$$

On cherche $P(V \mid T)$. Or on sait que :

$$P(V) = 0.005, \quad P(T \mid V) = 0.95, \quad P(T \mid V^c) = 0.01$$

On en déduit :

$$P(V \mid T) = \frac{P(T \cap V)}{P(T)} = \frac{P(T \mid V)P(V)}{P(T \mid V)P(V) + P(T \mid V^c)P(V^c)}$$
$$= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} \simeq 0.323$$

On voit ainsi que contrairement à ce que l'on aurait pu croire le test n'est pas fiable : si la personne présente un test positif, la probabilité qu'elle ne soit pas porteuse du virus est deux fois plus élevée que celle qu'elle le soit ! ∎

2.2 Indépendance

2.2.1 Indépendance de deux événements

Soient A et B deux événements de probabilité non nulle. Il arrive que la connaissance de la réalisation de A ne modifie pas notre information sur celle de B, autrement dit que $P(B \mid A) = P(B)$. C'est le cas par exemple lorsque l'on fait un tirage avec remise et que la réalisation de A ne dépend que du résultat du premier tirage, celle de B que du deuxième. Symétriquement on aura dans cet exemple $P(A \mid B) = P(A)$. Cette remarque se généralise :

Proposition 2.7 *Si A et B sont des événements de probabilité non nulle, les trois égalités suivantes sont équivalentes :*
(i) $P(B \mid A) = P(B)$,
(ii) $P(A \mid B) = P(A)$,
(iii) $P(A \cap B) = P(A)P(B)$.

Preuve : Comme $P(A) \neq 0$ et $P(B) \neq 0$, on a la chaîne d'équivalences :

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

∎

D'autre part la relation (iii) est toujours vérifiée dans le cas dégénéré où $P(A) = 0$ ou $P(B) = 0$. En effet, on a alors à la fois $P(A)P(B) = 0$ et $0 \leq P(A \cap B) \leq \min\big(P(A), P(B)\big) = 0$ d'où $P(A \cap B) = 0$. Ainsi la relation (iii) est un peu plus générale que (i) et (ii). Elle a aussi sur les deux autres l'avantage de la symétrie d'écriture. C'est elle que l'on retient pour définir l'indépendance.

Définition 2.8 *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. Deux événements A et B de cet espace sont dits indépendants lorsque :*

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

Exemple 2.5 *On jette deux fois le même dé. Les événements*

$$\begin{aligned} A &= \{\textit{obtention d'un chiffre pair au premier lancer}\}, \\ B &= \{\textit{obtention du 1 au deuxième lancer}\}, \end{aligned}$$

sont indépendants.

En effet, en prenant $\Omega = \{1, 2, \dots, 6\}^2$, $\mathcal{F} = \mathcal{P}(\Omega)$ et P l'équiprobabilité, on vérifie que :

$$P(A) = \frac{3 \times 6}{36} = \frac{1}{2}, \quad P(B) = \frac{6 \times 1}{36} = \frac{1}{6},$$
$$P(A \cap B) = \frac{3 \times 1}{36} = \frac{1}{12}, \quad P(A)P(B) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}.$$

Remarques :

- Si A est un événement tel que $P(A) = 0$ ou $P(A) = 1$, alors il est indépendant de tout événement, *y compris de lui même* (c'est le cas en particulier pour Ω et $\emptyset$).
- Deux événements *incompatibles* A et B avec $P(A) > 0$ et $P(B) > 0$ ne sont *jamais indépendants.* En effet $A \cap B = \emptyset$ implique $P(A \cap B) = 0$ or $P(A)P(B) \neq 0$.
- L'indépendance de deux événements A et B n'est pas une propriété intrinsèque aux événements, elle est toujours relative au modèle $(\Omega, \mathcal{F}, P)$ que l'on a choisi. Voici un exemple pour l'illustrer.

Exemple 2.6 *Une urne contient* 12 *boules numérotées de* 1 *à* 12*. On en tire une au hasard et on considère les événements :*

$$A = \{\textit{tirage d'un nombre pair}\}, \quad B = \{\textit{tirage d'un multiple de } 3\}.$$

L'espace probabilisé qui s'impose naturellement ici est $\Omega = \{1, \dots, 12\}$ muni de l'équiprobabilité P. Les événements A et B s'écrivent :

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

$$P(A) = \frac{6}{12} = \frac{1}{2}, \quad P(B) = \frac{4}{12} = \frac{1}{3}$$
$$P(A \cap B) = \frac{2}{12} = \frac{1}{6}, \quad \text{et} \quad P(A)P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6},$$

donc A et B sont indépendants.

On rajoute maintenant dans l'urne une boule numérotée treize et on recommence l'expérience.

Les événements A et B restent les mêmes, mais *le modèle a changé*. On a maintenant l'équiprobabilité P' sur $\Omega' = \{1, \dots, 13\}$ et :

$$P'(A) = \frac{6}{13}, \quad P'(B) = \frac{4}{13}, \quad P'(A \cap B) = \frac{2}{13}$$

mais

$$P'(A)P'(B) = \frac{6}{13} \times \frac{4}{13} = \frac{24}{169} \neq \frac{2}{13},$$

donc A et B ne sont plus indépendants. Un peu de réflexion permet de relier ces résultats calculatoires avec la notion intuitive d'indépendance présentée en introduction. Dans le premier cas, la proportion des multiples de trois parmi les pairs est la même que parmi les impairs. Le fait de savoir que la boule tirée est paire ne modifie en rien notre information sur B. Par contre dans le deuxième cas, l'ajout de la treizième boule modifie la proportion des multiples de trois : elle est plus élevée chez les pairs que chez les impairs. Donc le fait de savoir que la boule tirée est paire augmente un peu la probabilité que nous pouvons attribuer à B. ■

Proposition 2.9 *Si A et B sont indépendants, il en est de même pour les paires d'événements A et B^c, A^c et B, A^c et B^c.*

Preuve : Par hypothèse, $P(A \cap B) = P(A)P(B)$. En considérant la réunion disjointe $A = (A \cap B) \cup (A \cap B^c)$, nous avons : $P(A \cap B^c) = P(A) - P(A \cap B)$, d'où :

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)\bigl(1 - P(B)\bigr) = P(A)P(B^c).$$

Donc A et B^c sont indépendants. L'échange des rôles de A et B dans ce raisonnement donne l'indépendance de A^c et B. En réutilisant le premier résultat avec A^c à la place de A, on obtient alors celle de A^c et B^c. ■

2.2.2 Indépendance mutuelle

On se propose de généraliser la notion d'indépendance à plus de deux événements. Examinons d'abord la situation suivante.

Exemple 2.7 *Une urne contient quatre jetons : un bleu, un blanc, un rouge et un bleu-blanc-rouge. On en tire un au hasard. Considérons les trois événements*

$$\begin{aligned} A &= \{\textit{le jeton tiré contient du bleu}\}, \\ B &= \{\textit{le jeton tiré contient du blanc}\}, \\ C &= \{\textit{le jeton tiré contient du rouge}\}. \end{aligned}$$

Il est clair que $P(A) = P(B) = P(C) = 2/4 = 1/2$. D'autre part :

$$P(A \cap B) = P(\text{tricolore}) = \frac{1}{4} = P(A)P(B)$$

et de même $P(B \cap C) = 1/4 = P(B)P(C)$, $P(C \cap A) = 1/4 = P(C)P(A)$. Ainsi les événements A, B, C sont *deux à deux indépendants*.

D'autre part $P(A \mid B \cap C) = 1$ car $B \cap C = \{\text{tricolore}\}$. Donc la connaissance de la réalisation *simultanée* de B *et* C modifie notre information sur A. La notion d'indépendance deux à deux n'est donc pas suffisante pour traduire l'idée intuitive d'indépendance de plusieurs événements. Ceci motive la définition suivante.

Définition 2.10 *Trois événements A, B, C sont dits mutuellement indépendants lorsqu'ils vérifient les quatre conditions :*

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

Avec cette définition de l'indépendance des événements A, B et C on a bien[3] $P(A \mid B) = P(A)$, $P(A \mid B \cap C) = P(A)$, ainsi que toutes les égalités qui s'en déduisent par permutation sur les lettres A, B, C. On peut généraliser cette définition comme suit.

Définition 2.11 *Les n événements $A_1, \dots, A_n$ sont dits mutuellement indépendants si pour toute sous famille $A_{i_1}, \dots, A_{i_k}$ avec $1 \leq i_1 < \ldots < i_k \leq n$, on a :*

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \times \cdots \times P(A_{i_k}). \tag{2.1}$$

L'indépendance mutuelle implique évidemment l'indépendance deux à deux et la réciproque est fausse comme le montre l'exemple 2.7. Dans toute la suite, lorsque nous parlerons d'une famille de plusieurs événements indépendants sans autre précision , nous sous-entendrons systématiquement *mutuellement* indépendants.

Proposition 2.12 *Si $\{A_1, \dots, A_n\}$ est une famille de n événements indépendants, toute famille obtenue en remplaçant certains des A_i par leur complémentaire est encore indépendante.*

Preuve : Supposons la proposition démontrée dans le cas où l'on a remplacé *un seul* A_i par son complémentaire. Le cas général s'en déduit en utilisant cette propriété autant de fois qu'il y a de A_i changés en leur complémentaire. Dans le cas d'un seul A_i remplacé, on ne perd pas de généralité en supposant qu'il s'agit de A_1 (il suffit de changer l'indexation des événements, ce qui n'affecte pas leur indépendance mutuelle). Il nous reste alors à vérifier (2.1) dans le cas où $i_1 = 1$ avec $A_{i_1}^c$ à la place de A_{i_1} (dans le cas $i_1 > 1$, l'égalité ne fait intervenir que des éléments de la famille initiale et il n'y a donc rien à vérifier). Posons $B = A_{i_2} \cap \cdots \cap A_{i_k}$. L'hypothèse (2.1) appliquée à la famille $A_{i_2} \dots, A_{i_k}$

[3] Lorsque les probabilités conditionnelles existent.

nous donne $P(B) = P(A_{i_2}) \times \cdots \times P(A_{i_k})$. La même hypothèse appliquée à $A_{i_1} \ldots, A_{i_k}$ nous donne alors l'indépendance de A_{i_1} et de B. Par la proposition 2.9, on en déduit :

$$P(A_{i_1}^c \cap B) = P(A_{i_1}^c) \times P(B) = P(A_{i_1}^c) \times P(A_{i_2}) \times \cdots \times P(A_{i_k}),$$

ce qui achève la preuve. ■

Définition 2.13 *Une suite infinie d'événements est dite indépendante si toute sous suite finie est formée d'événements mutuellement indépendants.*

2.2.3 Épreuves répétées

Considérons une suite d'*épreuves* réalisées dans les mêmes conditions expérimentales. Par exemple tirages avec remise dans la même urne, lancers successifs d'un dé, ... Il est alors raisonnable de supposer que les résultats de tout sous ensemble fini d'épreuves n'ont aucune influence sur ceux des autres épreuves.

Définition 2.14 *On dit que les épreuves sont indépendantes si toute suite $(A_i)_{i\geq 1}$ telle que la réalisation de chaque A_i est déterminée uniquement par le résultat de la i-ème épreuve est une suite indépendante d'événements.*

Exemple 2.8 *On réalise une suite d'épreuves indépendantes. Chaque épreuve résulte en un succès avec probabilité $p \in]0,1[$ ou un échec avec probabilité $q = 1 - p$. Quelle est la probabilité des événements suivants :*
a) $A = \{$Au moins un succès au cours des n premières épreuves$\}$,
b) $B = \{$Exactement k succès au cours des n premières épreuves$\}$,
c) $C = \{$Toutes les épreuves donnent un succès$\}$?

Notons pour tout $i \geq 1$: $R_i = \{$succès à la i-ème épreuve$\}$, R_i^c est alors l'échec à la i-ème épreuve.
a) $A = \bigcup_{i=1}^{n} R_i$, d'où $A^c = \bigcap_{i=1}^{n} R_i^c$. Les R_i^c étant indépendants, on a :

$$P(A^c) = \prod_{i=1}^{n} P(R_i^c) = (1-p)^n = q^n.$$

On en déduit $P(A) = 1 - q^n$.
b)Traitons d'abord le cas $0 < k < n$. L'événement B est la réunion disjointe de tous les événements du type :

$$B_I = \Big(\bigcap_{i\in I} R_i\Big) \cap \Big(\bigcap_{j\in J} R_j^c\Big),$$

où I est une partie de cardinal k de $\{1, \ldots, n\}$ et J son complémentaire dans $\{1, \ldots, n\}$. L'ensemble d'indices I représente un choix possible des k épreuves donnant un succès, les autres épreuves indexées par J donnant alors un échec. En considérant tous les choix possibles de l'ensemble I (il y en a C_n^k), on obtient une partition de B par les B_I. Par indépendance des épreuves, pour tout I on a :

$$P(B_I) = \prod_{i\in I} P(R_i) \times \prod_{j\in J} P(R_j^c) = p^k q^{n-k}.$$

On voit ainsi que $P(B_I)$ ne dépend pas de I. On en déduit :

$$P(B) = \sum_{\substack{I\subset\{1,\dots,n\}\\ \mathrm{card}(I)=k}} P(B_I) = C_n^k p^k q^{n-k}.$$

La vérification de la validité de la formule $P(B) = C_n^k p^k q^{n-k}$ dans les cas $k = 0$ et $k = n$ est laissée au lecteur.
c) Pour $n \geq 1$, soit $C_n = \{$succès aux n premières épreuves$\}$. Clairement C est inclus dans C_n donc $0 \leq P(C) \leq P(C_n)$. En utilisant l'indépendance des R_i on obtient :

$$P(C_n) = P\Big(\bigcap_{i=1}^{n} R_i\Big) = \prod_{i=1}^{n} P(R_i) = p^n.$$

donc pour tout $n \geq 1$, $0 \leq P(C) \leq p^n$. En faisant tendre n vers $+\infty$, on en déduit $P(C) = 0$. ∎

2.3 Exercices

Ex 2.1. Donner une CNS pour que $P(A \mid H) = 1$.

Ex 2.2. Un avion a disparu et la région où il s'est écrasé est divisée pour sa recherche en trois zones de même probabilité. Pour $i = 1, 2, 3$, notons $1 - \alpha_i$ la probabilité que l'avion soit retrouvé par une recherche dans la zone i s'il est effectivement dans cette zone. Les constantes α_i représentent les probabilités de manquer l'avion et sont généralement attribuables à l'environnement de la zone (relief, végétation,...). On notera A_i l'événement *l'avion est dans la zone i*, et R_i l'événement *l'avion est retrouvé dans la zone i* ($i = 1, 2, 3$).

1) Pour $i = 1, 2, 3$, déterminer les probabilités que l'avion soit dans la zone i sachant que la recherche dans la zone 1 a été infructueuse.

2) Étudier brièvement les variations de ces trois probabilités conditionnelles considérées comme fonctions de α_1 et commenter les résultats obtenus.

Ex 2.3. Une urne contient 5 boules blanches et 7 rouges. On effectue 3 tirages d'une boule suivant la procédure suivante. A chaque tirage on prend une boule et on la remet dans l'urne *en y ajoutant une boule de même couleur*. Calculer les probabilités que l'échantillon de trois boules tiré contienne :

a) Aucune blanche.
b) Exactement une blanche.
c) Trois blanches.
d) Exactement deux blanches.

Ex 2.4. *L'erreur du Grand Inquisiteur*[4]
« Pour Tomás de Torquemada, le grand inquisiteur espagnol de funeste mémoire, une confession était une preuve indiscutable de culpabilité, même quand elle avait été arrachée sous la torture, cas fréquent à l'époque. Une des conclusions de R. Matthews dénommée *l'erreur de l'inquisiteur*, est surprenante : dans certaines circonstances, les aveux militent plus en faveur de l'innocence de l'accusé qu'ils ne prouvent sa culpabilité ! ». On note respectivement A, C et I les événements « l'accusé avoue », « l'accusé est coupable » et « l'accusé est innocent ». On pose $p = P(C)$ et on définit le rapport d'aveux r par :

$$r = \frac{P(A \mid I)}{P(A \mid C)}.$$

1) Calculer $P(C \mid A)$ en fonction de r et p.
2) Dans quel cas a-t-on $P(C \mid A) > P(C)$?
3) Proposer une interprétation du cas $r > 1$.

Ex 2.5. Une compagnie d'assurance répartit ses clients en trois classes R_1, R_2 et R_3 : les bons risques, les risques moyens et les mauvais risques. Les effectifs des ces trois classes représentent 20% de la population totale pour la classe R_1, 50% pour R_2 et 30% pour R_3. Les statistiques indiquent que les probabilités d'avoir un accident au cours de l'année pour une personne de l'une de ces trois classes sont respectivement de 0.05, 0.15 et 0.30.

1) Quelle est la probabilité qu'une personne choisie au hasard dans la population ait un accident dans l'année ?

2) Si M. Martin n'a pas eu d'accident cette année, quelle est la probabilité qu'il soit un bon risque ?

Ex 2.6. *Conditionnements multiples*
Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé et $H \in \mathcal{F}$ un évènement tel que $P(H) \neq 0$. On note P_H la fonction d'ensembles définie par

$$\forall A \in \mathcal{F}, \quad P_H(A) := P(A \mid H).$$

On sait (cf. Prop 2.2) que P_H est une nouvelle probabilité sur $(\Omega, \mathcal{F})$, on peut donc l'utiliser pour construire de nouvelles probabilités conditionnelles. On définit ainsi :

$$P(A \mid H_2 \mid H_1) := P_{H_1}(A \mid H_2).$$

[4] D'après l'article de Ian Stewart, *Pour la Science* 230, décembre 1996.

1) Quelle condition doivent vérifier H_1 et H_2 pour que cette définition soit légitime ?

2) Exprimer $P(A \mid H_2 \mid H_1)$ sous la forme d'une probabilité conditionnelle relative à un seul évènement.

3) En déduire que $P(A \mid H_2 \mid H_1) = P(A \mid H_1 \mid H_2)$.

4) Généraliser la question 2) au cas de n évènements $H_1, \dots, H_n$ et justifier à l'aide de ces nouvelles notations l'appellation règle des conditionnements successifs pour la Proposition 2.4.

Ex 2.7. On dispose de n urnes contenant chacune 3 boules. Parmi ces $3n$ boules, l'une est jaune et toutes les autres sont bleues. On ignore dans quelle urne se trouve la boule jaune.

1) On tire sans remise deux boules de l'urne 1. On note B_1 l'évènement « les deux boules tirées sont bleues » et J_1 l'évènement « la boule jaune est dans l'urne 1 ». Calculer $P(B_1 \mid J_1)$ et $P(B_1 \mid J_1^c)$. En déduire $P(B_1)$.

2) Quelle est la probabilité que la boule jaune soit dans l'urne 1 si le tirage a donné deux bleues ?

3) On reprend l'expérience avec cette fois n personnes chacune face à une urne où elles tirent simultanément et sans remise deux boules. On note B_i et J_i les événements définis de manière analogue à la première question.

a) Que vaut $P(B_i^c \mid J_k)$ pour $1 \leq i, k \leq n$? On distinguera les cas $i = k$ et $i \neq k$. En déduire $P(B_i)$.

b) Expliquer sans calcul pourquoi les évènements B_i et B_j $(i \neq j)$ ne sont pas indépendants.

c) Déterminer les valeurs des probabilités conditionnelles $P(B_i \cap B_j \mid J_k)$: on distinguera les deux cas $k \in \{i, j\}$ et $k \notin \{i, j\}$. En déduire $P(B_i \cap B_j)$.

d) Généraliser en calculant par la même méthode $P(B_{i_1} \cap B_{i_2} \cap \cdots \cap B_{i_r})$ où $1 \leq i_1 < i_2 < \cdots < i_r \leq n$.

e) Voyez vous une méthode plus rapide pour obtenir ce dernier résultat ?

Ex 2.8. Montrer que si A, B, C sont indépendants, $A \cup B$ et C sont indépendants. Généraliser.

Ex 2.9. On suppose que les événements $A_1, \dots, A_5$ sont indépendants. Donner une expression plus simple des probabilités : $P(A_1 \mid A_3)$, $P(A_3 \mid A_2 \cap A_4)$, $P(A_1 \cap A_3 \mid A_2)$, $P(A_1 \cap A_2 \mid A_3 \cap A_4 \cap A_5)$.

Ex 2.10. La définition de l'indépendance de trois événements s'écrit explicitement avec 4 égalités. Combien d'égalités faut-il pour écrire explicitement celle de l'indépendance de n événements ?

Ex 2.11. On s'intéresse à la répartition des sexes des enfants d'une famille de n enfants. On prend comme modélisation

$$\Omega_n = \{\mathtt{f},\mathtt{g}\}^n = \{\,(x_1,\dots,x_n),\ x_i \in \{\mathtt{f},\mathtt{g}\}, i=1,\dots,n\},$$

muni de l'équiprobabilité. On considère les évènements :

$$\begin{aligned} A &= \{\text{la famille a des enfants des deux sexes}\},\\ B &= \{\text{la famille a au plus une fille}\}. \end{aligned}$$

1) Montrer que pour $n \geq 2$, $P(A) = (2^n-2)/2^n$ et $P(B) = (n+1)/2^n$.

2) En déduire que A et B ne sont indépendants que si $n=3$.

Ex 2.12. On effectue des lancers répétés d'une paire de dés et on observe pour chaque lancer la *somme* des points indiqués par les deux dés. On se propose de calculer de deux façons la probabilité de l'événement E défini ainsi : *dans la suite des résultats observés, la première obtention d'un* 9 *a lieu avant la première obtention d'un* 7.

1) Quelle est la probabilité de n'obtenir ni 7 ni 9 au cours d'un lancer ?

2) *Première méthode :* On note $F_i = \{$*obtention d'un* 9 *au i-ème lancer*$\}$ et pour $n > 1$, $E_n = \{$ *ni* 7 *ni* 9 *ne sont obtenus au cours des* $n-1$ *premiers lancers et le n-ième lancer donne* 9$\}$. Dans le cas particulier $n=1$, on pose $E_1 = F_1$.

a) Exprimer E à l'aide d'opérations ensemblistes sur les E_n $(n \geq 1)$. Exprimer de même chaque E_n à l'aide des F_i et des $H_i = \{$ *ni* 7 *ni* 9 *au i-ème lancer*$\}$.

b) Calculer $P(E_n)$ en utilisant l'indépendance des lancers.

c) Calculer $P(E)$.

3) *Deuxième méthode :* On note $G_1 = \{$ *obtention d'un* 7 *au premier lancer*$\}$.

a) Donner une expression de $P(E)$ en utilisant le conditionnement par la partition $\{F_1, G_1, H_1\}$.

b) Donner sans calcul les valeurs de $P(E \mid F_1)$, $P(E \mid G_1)$ et expliquer pourquoi $P(E \mid H_1) = P(E)$.

c) En déduire la valeur de $P(E)$.

Ex 2.13. Trois personnes nommées A, B, C lancent à tour de rôle la même pièce de monnaie (ABCABCABC...). La probabilité d'obtenir pile lors d'un lancer est p $(0 < p < 1)$. Le gagnant est le premier qui obtient pile (la partie *s'arrête* alors).

1) On note A_n (resp. B_n, C_n) l'événement A *gagne la partie lors du n-ième lancer* (resp. B, C). Calculer $P(A_1)$, $P(B_2)$, $P(C_3)$. Les événements A_1 et B_2 sont-ils indépendants ?

2) En discutant suivant les valeurs de n, calculer $P(A_n)$, $P(B_n)$, $P(C_n)$.

3) Calculer la probabilité de gagner de chacun des trois joueurs.

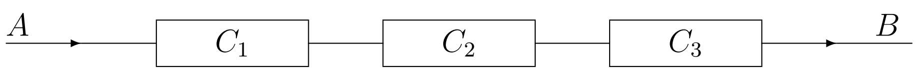
4) Quelle est la probabilité qu'il y ait un vainqueur ? Conclure.

Ex 2.14. On se propose de calculer la probabilité de fonctionnement de quelques circuits électriques simples à l'aide des probabilités de fonctionnement de leurs composants. Dans tout ce qui suit, les composants ont des fonctionnements *mutuellement indépendants*. On note p_i la probabilité de fonctionnement du composant C_i et F_i l'événement :

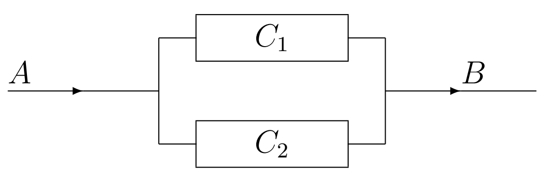
$$F_i = \{\text{le composant } C_i \text{ fonctionne}\}.$$

On dit que le circuit fonctionne si le courant peut passer du point A au point B.

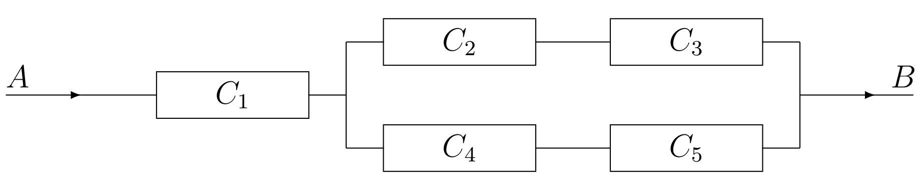
1) *Circuit série* : Calculer la probabilité de fonctionnement du circuit suivant :



2) *Circuit parallèle* : Calculer la probabilité de fonctionnement du circuit suivant :



3) *Circuit mixte* : Calculer la probabilité de fonctionnement du circuit suivant lorsque tous les p_i sont égaux à p.



Ex 2.15. Soit Ω un ensemble fini dont le cardinal est un *nombre premier*. On le munit de l'équiprobabilité P. Soient A et B deux parties de Ω distinctes chacune de $\emptyset$ et de Ω. Démontrer que A et B ne sont pas indépendantes.

Ex 2.16. *Épreuves répétées à trois issues.*
On considère une suite de n épreuves répétées indépendantes avec pour chaque épreuve trois résultats possibles : A avec probabilité a, B avec probabilité b ou C avec probabilité c ($a + b + c = 1$). On notera respectivement A_i, B_i et C_i les événements obtention du résultat A (respectivement, B, C) à la i-ème épreuve.

1) Dans cette question, $n = 5$. Quelle est la probabilité d'obtenir dans cet ordre 2 A suivis d'un B et de 2 C ? Quelle est celle d'obtenir (sans condition d'ordre) 2 A, 1 B et 2 C ?

2) Généraliser en montrant que la probabilité d'obtenir au cours des n épreuves (et sans condition d'ordre) i résultats A, j résultats B et k résultats C $(i + j + k = n)$ vaut :

$$\frac{n!}{i!\,j!\,k!} a^i b^j c^k.$$

Ex 2.17. On rappelle que si A est un événement, la variable aléatoire indicatrice de A est la fonction définie sur Ω par :

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

Autrement dit, $\mathbf{1}_A$ vaut 1 si A est réalisé et 0 sinon.

Soit $(A_i)_{i\geq 1}$ une suite d'événements *indépendants*. On note $p_i = P(A_i)$ et on suppose que :

$$a := \sum_{i=1}^{+\infty} p_i < +\infty.$$

Le but de cet exercice est de prouver l'inégalité

$$\forall n, k \in \mathbb{N}^*, \quad P\Big(\sum_{i=1}^{n} \mathbf{1}_{A_i} \geq k\Big) \leq \frac{a^k}{k!}.$$

La dernière question propose une application de cette inégalité.

1) Que peut-on dire du cas $k > n$? On suppose dans la suite $k \leq n$.

2) On note $B_{n,k} := \{\omega \in \Omega;\ \sum_{i=1}^{n} \mathbf{1}_{A_i}(\omega) \geq k\}$. Justifier l'inclusion

$$B_{n,k} \subset \bigcup_{\substack{F\subset\{1,\dots,n\} \\ \operatorname{card} F = k}} \bigcap_{i\in F} A_i.$$

3) En déduire que

$$P(B_{n,k}) \leq \sum_{\substack{F\subset\{1,\dots,n\} \\ \operatorname{card} F = k}} \prod_{i\in F} p_i.$$

4) On note $a_n = \sum_{i=1}^{n} p_i$. Montrer que

$$a_n^k \geq k! \sum_{\substack{F\subset\{1,\dots,n\} \\ \operatorname{card} F = k}} \prod_{i\in F} p_i.$$

Indication : On remarquera que

$$a_n^k = \sum_{(i_1,\dots,i_k)\in\{1,\dots,n\}^k} p_{i_1} \cdots p_{i_k}.$$

5) Conclure.

6) *Application à un problème de tir.* Dans un stand de tir, une cible mobile traverse le champ visuel d'un tireur (une traversée par épreuve). À chaque épreuve, le tireur tire un coup sur la cible. D'une épreuve à la suivante, la vitesse de la cible augmente de 20%. On suppose que pour un tireur donné, la probabilité de toucher la cible est *inversement proportionnelle* à la vitesse de la cible. Elle vaut ainsi $p \in]0,1[$ pour le premier tir, $\frac{5}{6}p$ pour le second (pourquoi ?), etc. Les tirs sont supposés indépendants, le tireur dispose d'autant de cartouches qu'il le souhaite et le défi qu'il doit relever est celui de toucher au moins 20 fois la cible. En utilisant le résultat démontré ci-dessus, majorer sa probabilité de réussir (indépendamment de p).

Ex 2.18. Loi ζ (suite)$\uparrow$ 1.13
On travaille avec l'espace probabilisé $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*), P_a)$ où P_a est la loi ζ de paramètre $a > 1$ (voir la définition exercice 1.13 page 24).

1) Donner une CNS sur les entiers $j > 1$ et $m > 1$ pour que les événements $A = j\mathbb{N}^*$ et $B = m\mathbb{N}^*$ soient P_a-indépendants.

2) Soit p_i le i-ème *nombre premier* ($\mathsf{p}_1 = 2$, $\mathsf{p}_2 = 3$, $\mathsf{p}_3 = 5$, $\ldots$). On note $A_i = \mathsf{p}_i\mathbb{N}^*$. Montrer que $(A_i)_{i\geq 1}$ est une suite d'événements indépendants.

3) Soit C_n l'ensemble des entiers de $\mathbb{N}^*$ divisibles par aucun des nombres premiers p_i pour $1 \leq i \leq n$. Calculer $P_a(C_n)$.

4) Déterminer $\bigcap_{n\geq 1} C_n$.

5) En déduire la formule d'Euler :

$$\forall a > 1, \quad \sum_{k=1}^{+\infty} \frac{1}{k^a} = \zeta(a) = \lim_{n\to\infty} \prod_{i=1}^{n} \Big(1 - \frac{1}{\mathsf{p}_i^a}\Big)^{-1} = \prod_{i=1}^{+\infty} \Big(1 - \frac{1}{\mathsf{p}_i^a}\Big)^{-1}.$$

6) On peut retrouver cette formule par une calcul direct sur la série définissant $\zeta(a)$. En voici le début :

$$\begin{aligned} \zeta(a) = \sum_{k=1}^{+\infty} \frac{1}{k^a} &= \sum_{2|k} \frac{1}{k^a} + \sum_{2\not| k} \frac{1}{k^a} \\ &= \frac{1}{2^a}\zeta(a) + \sum_{2\not| k} \frac{1}{k^a} \end{aligned}$$

On recommence avec la série $(1-2^{-a})\zeta(a)$ en séparant les termes dont l'indice est un multiple de 3 des autres...Expliciter cette méthode et résoudre le problème de convergence sous-jacent.

Ex 2.19. Un train se compose de 10 wagons citernes contenant un produit dangereux. Chacun des wagons peut avec une probabilité 0.1 (et indépendamment des autres) avoir un défaut. Avant le départ, les wagons sont examinés par deux

contrôleurs. Chacun d'eux vérifie tous les wagons, sans échanger d'information avec son collègue pendant le contrôle. On admet que chaque contrôleur peut déceler le défaut (s'il y en a un) d'un wagon donné avec une probabilité égale à 0.7. Un seul défaut suffit pour que le train soit retardé. Trouver les probabilités des événements suivants :
a) Le train est retardé.
b) Le train part avec au moins un wagon défectueux.

Chapitre 3

Variables aléatoires discrètes

3.1 Introduction

Dans de nombreux jeux, on fait intervenir le hasard en observant la somme des points marqués par deux dés. Considérons le jet d'un dé bleu et d'un dé rouge et notons S la somme des points obtenus. On modélise cette expérience en prenant l'équiprobabilité sur :

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

Un événement élémentaire ω est un couple (b, r) où b désigne le résultat du dé bleu et r celui du rouge. Pour tout événement élémentaire $\omega = (b, r)$, on a $S(\omega) = b + r$. Il est facile de représenter tous les cas possibles par un tableau à 36 cases.

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

On a ainsi défini une application S de Ω dans l'ensemble des sommes de points possibles : $\{2, 3, \ldots, 11, 12\}$. On dit que S est une *variable aléatoire* sur Ω. En fait, l'observation qui nous intéresse dans cette expérience, ce n'est pas ω, mais seulement $S(\omega)$. Ce que l'on aimerait connaître, c'est la probabilité que la somme des points prenne une valeur donnée, soit $P(S = k)$ pour k entier fixé entre 2 et 12. En utilisant l'équiprobabilité sur Ω et le tableau ci-dessus, nous obtenons immédiatement :

k	2	3	4	5	6	7	8	9	10	11	12
$P(S = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Cela revient à considérer un nouvel ensemble d'événements élémentaires :

$$\Omega' = S(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

et à munir cet ensemble de la probabilité P_S définie par le tableau des $P(S = k)$. Cette nouvelle probabilité s'appelle *loi de la variable aléatoire S*.

3.2 Généralités

3.2.1 Variable aléatoire discrète

Définition 3.1 *Soit $(\Omega, \mathcal{F}, P)$ un espace probabilisé. On appelle variable aléatoire discrète sur $(\Omega, \mathcal{F}, P)$ toute application X :*

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

vérifiant les deux conditions :

(i) L'ensemble des images $X(\Omega) = \{X(\omega),\ \omega \in \Omega\}$ est une partie au plus dénombrable de $\mathbb{R}$. On peut donc numéroter ses éléments par des indices entiers[1]

$$X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}.$$

(ii) Pour tout $x_k \in X(\Omega)$, $A_k = \{\omega \in \Omega, X(\omega) = x_k\}$ fait partie de la famille $\mathcal{F}$ d'événements auxquels on peut attribuer une probabilité par P.

L'événement A_k est aussi noté $X^{-1}(\{x_k\})$ (inverse *ensembliste*[2]) ou plus commodément $\{X = x_k\}$. Nous utiliserons l'abréviation v.a. pour variable aléatoire. Remarquons que la famille de tous les A_k forme une partition de Ω : on classe chaque élément de Ω selon son image par X. Il en résulte :

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

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

3.2.2 Loi d'une variable aléatoire discrète

L'application X permet de *transporter* la probabilité P de Ω sur $\mathbb{R}$: on considère les $P(X = x_k)$ comme des masses ponctuelles situées aux points x_k de la droite réelle. La probabilité d'une partie quelconque de $\mathbb{R}$ est alors définie comme la somme de ses masses ponctuelles.

[1] Pour tous les exemples classiques que nous rencontrerons, il est possible de les numéroter de manière croissante : $x_0 < x_1 < x_2 \dots$. Mais ce n'est pas toujours le cas, car l'ensemble des valeurs possibles peut être par exemple les décimaux (ou les rationnels) de $[0, 1]$ (voir exercice 3.15).

[2] Ceci ne suppose pas la bijectivité de X.

Définition 3.2 *Soit X une variable aléatoire discrète sur $(\Omega, \mathcal{F}, P)$. On lui associe la fonction d'ensemble P_X définie sur la famille de toutes les parties de $\mathbb{R}$ en posant :*

$$p_k = P_X(\{x_k\}) = P(A_k) = P(X = x_k),$$

puis pour tout $B \subset \mathbb{R}$:

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

La fonction d'ensembles P_X ainsi définie est une probabilité sur la famille $\mathcal{P}(\mathbb{R})$ de toutes les parties de $\mathbb{R}$. On l'appelle loi de la variable aléatoire X.

Remarquons que la définition de $P_X(B)$ a toujours un sens en convenant qu'une somme indexée par l'ensemble vide vaut 0 et en notant que s'il y a une infinité de x_k dans B, la somme sur B des p_k est une sous-série de la série à termes *positifs* convergente : $\displaystyle\sum_{x_k \in X(\Omega)} p_k = 1$. Le fait que P_X soit une probabilité résulte des propositions 1.4 et 1.5.

Remarque 1 : Deux variables aléatoires peuvent avoir même loi sans être égales. Par exemple considérons le jet de deux dés, l'un bleu et l'autre rouge. Notons X le nombre de points indiqué par le dé bleu et Y celui du rouge. Les variables aléatoires X et Y sont définies sur le même espace probabilisé $\Omega = \{1, 2, 3, 4, 5, 6\}^2$ muni de l'équiprobabilité. On a $X(\Omega) = Y(\Omega) = \{1, 2, 3, 4, 5, 6\}$ et :

$$\forall k \in \{1, 2, 3, 4, 5, 6\}, \quad P(X = k) = \frac{1}{6}, \quad P(Y = k) = \frac{1}{6}.$$

Donc X et Y ont même loi : $P_X = P_Y$. Pour autant on n'a pas l'égalité des variables aléatoires X et Y qui signifierait $X(\omega) = Y(\omega)$ *pour tout* $\omega \in \Omega$ (égalité de deux applications). Autrement dit, en lançant deux dés on obtiendrait *à coup sûr* un double. Par contre nous pouvons considérer *l'événement* $\{X = Y\}$ dont la réalisation n'est pas certaine et calculer sa probabilité :

$$P(X = Y) = P\Big(\bigcup_{k=1}^{6} \{(X, Y) = (k, k)\}\Big) = \frac{6}{36} = \frac{1}{6}.$$

On en déduit : $P(X \neq Y) = 5/6$.

Remarque 2 : Deux variables aléatoires discrètes X et Y peuvent avoir même loi sans que $X(\Omega) = Y(\Omega)$. Cela vient du fait que pour certaines valeurs x_k dans $X(\Omega)$ (ou $y_l \in Y(\Omega)$), on peut avoir $P(X = x_k) = 0$ (ou $P(Y = y_l) = 0$). Bien sûr, on pourrait redéfinir X et Y de façon à effacer ces valeurs particulières, ce qui reviendrait remplacer Ω par l'un de ses sous-ensembles Ω'. On ne fera pas cette convention[3], car on a parfois intérêt à laisser dans l'ensemble des valeurs possibles d'une v.a. des valeurs qui sont atteintes avec une probabilité nulle. C'est le cas lorsqu'on étudie des problèmes de convergence de suites de v.a. (voir la discussion sur la loi forte des grands nombres pour les fréquences).

[3] Sauf dans le cas de l'exemple 3.2 ci-dessous.

3.2.3 Fonction de répartition

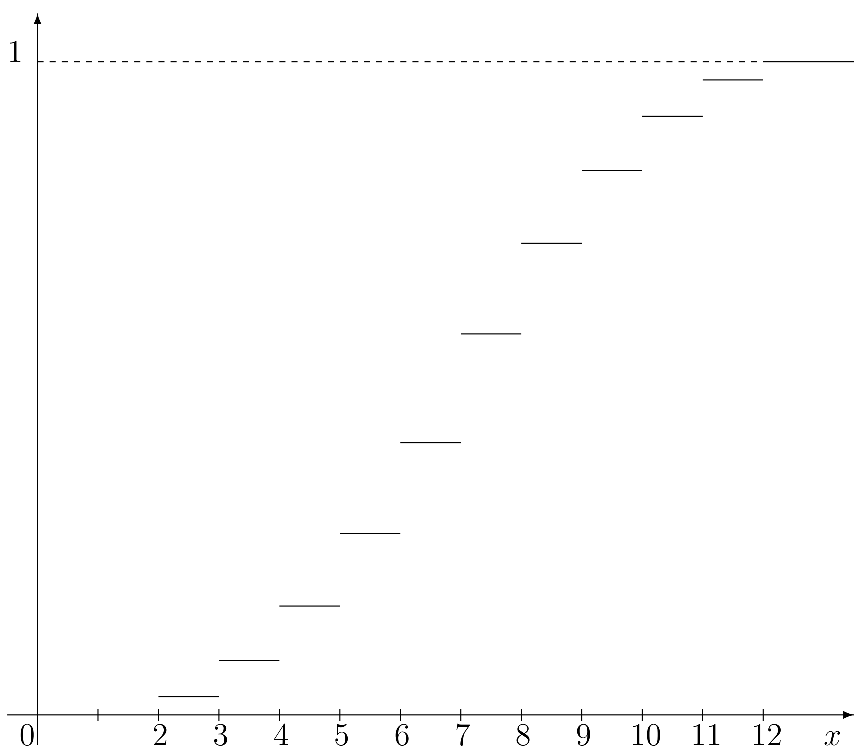
Définition 3.3 *On appelle fonction de répartition de la variable aléatoire X, la fonction F_X définie sur $\mathbb{R}$ par :*

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

On a aussi :

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

Voici à titre d'exemple la représentation graphique de F_S où S est la variable aléatoire somme des points de deux dés.



Théorème 3.4 *La fonction de répartition F_X d'une variable aléatoire discrète X est croissante sur $\mathbb{R}$, continue à droite et limitée à gauche en tout point. Elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Elle caractérise la loi de X, autrement dit : $F_X = F_Y$ si et seulement si les variables aléatoires X et Y ont même loi.*

Preuve : Vérifions d'abord la croissance de F_X sur $\mathbb{R}$. Soient s et t deux réels quelconques tels que $s \leq t$. Tout ω vérifiant $X(\omega) \leq s$ vérifie a fortiori $X(\omega) \leq t$. Cette implication se traduit par l'inclusion d'événements $\{X \leq s\} \subset \{X \leq t\}$ d'où $P(X \leq s) \leq P(X \leq t)$, autrement dit $F_X(s) \leq F_X(t)$. Ainsi F_X est croissante. Il en résulte qu'elle possède une limite à gauche et une limite à droite en tout point x de $\mathbb{R}$ (cf. cours d'analyse).

Le reste de la preuve repose sur la propriété de *continuité monotone séquentielle* de la probabilité (cf. proposition 1.2) que nous rappelons maintenant :

Soit P une probabilité sur l'espace $(\Omega, \mathcal{F})$.

(i) Si $(A_n)_{n\in\mathbb{N}^*}$ est une suite croissante d'événements (c'est-à-dire $\forall n \in \mathbb{N}^*$, $A_n \subset A_{n+1}$) alors

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

(ii) Si $(B_n)_{n\in\mathbb{N}^*}$ est une suite décroissante d'événements (c'est-à-dire $\forall n \in \mathbb{N}^*$, $B_{n+1} \subset B_n$) alors

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

Soit $x \in \mathbb{R}$ fixé. Comme on est assuré de l'existence de la limite à droite de F_X en ce point, pour montrer que cette limite vaut $F_X(x)$ et établir la continuité à droite de F_X en x, il suffit de vérifier que : $F_X(x) = \lim_{n\to+\infty} F_X(x + 1/n)$. Comme $F_X(x + 1/n) = P(X \leq x + 1/n)$, ceci résulte de la propriété (ii) ci-dessus appliquée aux événements $B_n = \{X \leq x + 1/n\}$ en remarquant que :

$$\bigcap_{n\in\mathbb{N}^*} \left\{X \leq x + \frac{1}{n}\right\} = \{X \leq x\}. \tag{3.1}$$

En effet, pour tout $\omega \in \{X \leq x\}$, on a $X(\omega) \leq x \leq x + 1/n$ pour tout $n \in \mathbb{N}^*$ et donc $\omega \in \bigcap_{n\in\mathbb{N}^*} B_n$ d'où $\{X \leq x\}$ est inclus dans l'intersection des B_n $(n \in \mathbb{N}^*)$. Réciproquement tout ω dans cette intersection vérifie : $\forall n \in \mathbb{N}^*$, $X(\omega) \leq x + 1/n$. Le passage à la limite quand n tend vers l'infini conservant cette inégalité large, on en déduit $X(\omega) \leq x$ soit encore $\omega \in \{X \leq x\}$. Ainsi l'intersection des B_n, $(n \in \mathbb{N}^*)$ est incluse dans $\{X \leq x\}$, ce qui achève la vérification de (3.1).

Comme F_X est croissante, elle admet des limites en $-\infty$ et $+\infty$. Si $X(\Omega)$ est borné inférieurement[4], il est clair que $F_X(x) = 0$ pour tout x assez petit et donc $\lim_{x\to-\infty} F_X(x) = 0$. Dans le cas général, on identifie la limite en $-\infty$ (dont on connaît l'existence) grâce à une suite particulière :

$$\lim_{x\to-\infty} F_X(x) = \lim_{n\to+\infty} F_X(-n) = \lim_{n\to+\infty} P(X \leq -n).$$

On utilise à nouveau la propriété (ii) avec $B_n = \{X \leq -n\}$ en montrant que :

$$\bigcap_{n\in\mathbb{N}^*} \{X \leq -n\} = \emptyset. \tag{3.2}$$

[4]Attention, dans le cas général x_0 n'est pas nécessairement le plus petit élément de $X(\Omega)$, on peut très bien avoir par exemple $X(\Omega) = \mathbb{Z}$.

En effet, soit ω un élément de cette intersection. Alors $X(\omega) \leq -n$ pour tout $n \in \mathbb{N}^*$ donc en passant à la limite, $X(\omega) = -\infty$, ce qui est impossible puisque X ne prend que des valeurs finies[5]. Donc cette intersection est vide et par (ii), $\lim_{n\to+\infty} P(X \leq -n) = P(\emptyset) = 0$. La preuve de $\lim_{x\to+\infty} F_X(x) = 1$ est analogue et est laissée au lecteur.

Supposons que les variables aléatoires X et Y aient même loi. Cela signifie que pour tout $B \subset \mathbb{R}$, $P_X(B) = P_Y(B)$. En choisissant B de la forme $]-\infty, t]$, $t \in \mathbb{R}$, on en déduit que pour tout réel t, $P(X \leq t) = P(Y \leq t)$ autrement dit $F_X(t) = F_Y(t)$. Donc les *fonctions* F_X et F_Y sont égales.

Réciproquement, supposons que les deux variables aléatoires discrètes X et Y aient même fonction de répartition, ce qui signifie :

$$\forall x \in \mathbb{R}, \quad F_X(x) = F_Y(x). \tag{3.3}$$

Admettons provisoirement le lemme suivant :

Lemme 3.5 *Si X est une v. a. discrète de fonction de répartition F, on a :*

$$\forall x \in \mathbb{R}, \quad P(X = x) = F(x) - F(x^-) = F(x) - \lim_{\varepsilon\to 0^+} F(x-\varepsilon). \tag{3.4}$$

D'après (3.3) et (3.4) on a donc $P(X = x) = P(Y = x)$ pour tout $x \in \mathbb{R}$. Il en résulte que X et Y ont même loi. ■

Preuve du lemme 3.5 : On remarque que pour tout $x \in \mathbb{R}$ et tout $n \in \mathbb{N}^*$,

$$P(x - \frac{1}{n} < X \leq x) = P(X \leq x) - P(X \leq x - \frac{1}{n}) = F_X(x) - F_X(x - \frac{1}{n}).$$

On utilise une nouvelle fois (ii) en vérifiant que :

$$\{X = x\} = \bigcap_{n\in\mathbb{N}^*} \Big\{x - \frac{1}{n} < X \leq x\Big\} \tag{3.5}$$

En effet, le premier membre de (3.5) est clairement inclus dans le second. Réciproquement, si ω est un élément quelconque de cette intersection, pour tout $n \in \mathbb{N}^*$, $x - 1/n < X(\omega) \leq x$, donc en passant à la limite quand n tend vers l'infini, $x \leq X(\omega) \leq x$ d'où $X(\omega) = x$. Ceci termine la justification de (3.5) et comme on sait que F_X a une limite à gauche au point x, on conclut en appliquant (ii) à la suite décroissante d'événements $B_n = \{x - 1/n < X \leq x\}$:

$$P(X = x) = \lim_{n\to+\infty} \big(F_X(x) - F_X(x - 1/n)\big) = F_X(x) - F_X(x^-).$$

■

[5] Ce qui ne signifie pas que $X(\Omega)$ soit borné...

3.3 Lois discrètes classiques

3.3.1 Lois de Bernoulli

Définition 3.6 *La variable aléatoire X suit la loi de Bernoulli de paramètre p ($p \in [0,1]$) si elle ne prend que deux valeurs 0 et 1 avec :*

$$P(X=1)=p \qquad P(X=0)=1-p=q.$$

On notera $X \sim \mathcal{B}(p)$.

Si A est un événement de probabilité p, son indicatrice définie par

$$\mathbf{1}_A(\omega)=\begin{cases}1 & \text{si} \quad \omega \in A,\\ 0 & \text{si} \quad \omega \notin A\end{cases} = \begin{cases}1 & \text{si} \quad A \text{ est réalisé}\\ 0 & \text{si} \quad A \text{ n'est pas réalisé}\end{cases}$$

est une variable aléatoire suivant la loi de Bernoulli de paramètre p. Réciproquement, si X est une v.a. de Bernoulli, on peut toujours écrire $X=\mathbf{1}_A$ en définissant $A=\{\omega\in\Omega,\ X(\omega)=1\}$.

3.3.2 Loi uniforme sur un ensemble fini de réels

Définition 3.7 *La variable aléatoire X suit la loi uniforme sur l'ensemble de réels $\{x_1,\dots,x_n\}$ si P_X est l'équiprobabilité sur cet ensemble.*

Autrement dit, l'ensemble des valeurs possibles de X est $X(\Omega)=\{x_1,\dots,x_n\}$ et :

$$\forall k=1,\dots,n, \qquad P(X=x_k)=\frac{1}{n}.$$

Par exemple le nombre de points indiqué par un dé suit la loi uniforme sur $\{1,2,3,4,5,6\}$.

3.3.3 Lois binomiales

Définition 3.8 *La variable aléatoire X suit la loi binomiale de paramètres n et p ($n \in \mathbb{N}^*$ et $p \in [0,1]$) si l'ensemble des valeurs possibles est $X(\Omega)=\{0,1,\dots,n\}$ et*

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

Notation : $X \sim \mathcal{B}(n,p)$.

La formule ci-dessus définit bien une loi de probabilité puisque les $C_n^k p^k(1-p)^{n-k}$ sont positifs et :

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

en appliquant la formule du binôme de Newton (d'où le nom de la loi). La loi binomiale $\mathcal{B}(n,p)$ est la loi du nombre de succès obtenus en une suite de n épreuves répétées indépendantes avec pour chaque épreuve une probabilité de succès p. Ceci a été démontré dans l'exemple 2.8.

De même, soit $A_1, \ldots, A_n$ une famille de n événements mutuellement indépendants ayant tous même probabilité p et notons X_i la variable de Bernoulli indicatrice de A_i :

$$X_i(\omega) = \begin{cases} 1 & \text{si} \quad \omega \in A_i, \\ 0 & \text{si} \quad \omega \in A_i^c. \end{cases}$$

Alors la variable aléatoire $S_n = \displaystyle\sum_{i=1}^{n} X_i$ suit la loi binomiale $\mathcal{B}(n,p)$.

3.3.4 Lois hypergéométriques

Alors que la loi binomiale intervient dans les tirages avec remise, la loi hypergéométrique correspond aux tirages sans remise.

Exemple 3.1 *Dans une production totale de N objets dont M sont défectueux, on prélève au hasard un échantillon de n objets (tirage sans remise). Soit X le nombre aléatoire d'objets défectueux dans l'échantillon. Quelle est sa loi ?*

On peut prendre comme espace Ω l'ensemble de tous les échantillons possibles (toutes les parties à n éléments d'un ensemble de cardinal N) muni de l'équiprobabilité. Chaque échantillon a ainsi une probabilité $1/C_N^n$ d'être choisi. Les échantillons (événements élémentaires) réalisant l'événement $\{X = k\}$ sont ceux qui contiennent k objets défectueux et $n-k$ objets défectueux. Ceci n'est réalisable que si $0 \leq k \leq M$ et $0 \leq n-k \leq N-M$. Dénombrons ces échantillons. On les forme en choisissant k objets défectueux dans une sous-population de M et en complétant par $n-k$ objets non défectueux choisis dans une sous population de $N-M$. Il y en a donc $C_M^k \times C_{N-M}^{n-k}$. Finalement :

$$P(X = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \quad \text{si} \quad \begin{cases} 0 \leq k \leq M, \\ 0 \leq n-k \leq N-M. \end{cases} \tag{3.6}$$

Définition 3.9 *La loi définie par (3.6) s'appelle loi hypergéométrique de paramètres N, M et n. Notation : $X \sim \mathcal{H}(N,M,n)$. Le paramètre N est l'effectif de la population totale, M celui de la sous-population à laquelle on s'intéresse et n la taille de l'échantillon observé.*

Pour une taille d'échantillon n *fixée*, plus N et M sont grands, moins les tirages sans remise diffèrent des tirages avec remise. Plus précisément, la loi hypergéométrique converge vers la loi binomiale au sens suivant.

Théorème 3.10 *On suppose que quand N tend vers $+\infty$, $M = M(N)$ tend vers $+\infty$ en vérifiant la condition :*

$$\lim_{N\to+\infty} \frac{M}{N} = p \quad \textit{avec} \quad 0 < p < 1. \tag{3.7}$$

Alors, n restant fixé, la loi hypergéométrique $\mathcal{H}(N, M, n)$ converge vers la loi binomiale $\mathcal{B}(n, p)$, ce qui signifie que si $(X_N)_{N\geq 1}$ est une suite de v.a. avec $X_N \sim \mathcal{H}(N, M, n)$ et Y est une v.a. de loi binomiale $\mathcal{B}(n, p)$, alors :

$$\forall k = 0, 1, \dots, n, \qquad \lim_{N\to+\infty} P(X_N = k) = P(Y = k), \tag{3.8}$$

autrement dit :

$$\forall k = 0, 1, \dots, n, \qquad \lim_{N\to+\infty} \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}. \tag{3.9}$$

Preuve : Remarquons d'abord que comme p est *strictement* positif, l'hypothèse (3.7) implique que M tend vers $+\infty$ avec N ; il en va de même pour $N - M$ puisque $p < 1$.

Pour n et k fixés, posons :

$$\begin{aligned}
p_N &= \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n} \\
&= \frac{M!}{k!(M-k)!} \times \frac{(N-M)!}{(n-k)!\big((N-M)-(n-k)\big)!} \times \frac{n!(N-n)!}{N!} \\
&= C_n^k \frac{M!}{(M-k)!} \times \frac{(N-M)!}{\big((N-M)-(n-k)\big)!} \times \frac{(N-n)!}{N!}.
\end{aligned} \tag{3.10}$$

Comme k est *fixé* et M tend vers $+\infty$, la première fraction dans (3.10) est le produit de k facteurs M, $(M-1)$, ..., $(M-k+1)$ tous équivalents[6] à M d'où :

$$\frac{M!}{(M-k)!} \sim M^k, \qquad N \to +\infty. \tag{3.11}$$

Par le même argument avec $n - k$ et $N - M$ au lieu de k et M :

$$\frac{(N-M)!}{\big((N-M)-(n-k)\big)!} \sim (N-M)^{n-k}, \qquad N \to +\infty. \tag{3.12}$$

Enfin :

$$\frac{(N-n)!}{N!} \sim \frac{1}{N^n}, \qquad N \to +\infty. \tag{3.13}$$

En reportant ces équivalents dans (3.10), on voit que lorsque N tend vers $+\infty$:

$$p_N \sim C_n^k \frac{M^k (N-M)^{n-k}}{N^n} = C_n^k \Big(\frac{M}{N}\Big)^k \Big(\frac{N-M}{N}\Big)^{n-k}, \tag{3.14}$$

d'où : $\displaystyle\lim_{N\to+\infty} p_N = C_n^k p^k (1-p)^{n-k}$. ■

[6] Rappelons que deux suites (u_N) et (v_N) sont dites équivalentes lorsque $u_N = v_N(1+\varepsilon_N)$ avec ε_N tendant vers 0 quand N tend vers $+\infty$ (notation : $u_N \sim v_N$).

3.3.5 Lois géométriques

Exemple 3.2 (Un problème de temps d'attente)
Considérons une suite infinie d'épreuves répétées indépendantes avec même probabilité de succès $p \in]0,1[$. *Soit* X *le numéro (aléatoire) de la première épreuve où l'on obtient un succès. Si l'on n'obtient jamais de succès, on conviendra que* $X = +\infty$. *Calculer* $P(X = k)$ *pour tout* $k \in \mathbb{N}^*$. *En déduire les valeurs de* $P(X \in \mathbb{N}^*)$ *et* $P(X = +\infty)$.
En notant $R_i = \{\text{succès à la } i\text{-ème épreuve}\}$, on a :

$$\begin{aligned} \{X = k\} &= \{\text{échec aux } (k-1) \text{ premières et succès à la } k\text{-ième}\} \\ &= \Big(\bigcap_{i=1}^{k-1} R_i^c \Big) \cap R_k. \end{aligned}$$

D'où par indépendance des épreuves :

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

Posons $q = 1 - p$ et notons que $q \in]0,1[$. La décomposition de l'événement $\{X \in \mathbb{N}^*\}$ en la réunion disjointe des $\{X = k\}$ $(k \in \mathbb{N}^*)$ nous donne par σ-additivité :

$$\begin{aligned} P(X \in \mathbb{N}^*) = \sum_{k \in \mathbb{N}^*} P(X = k) &= \sum_{k \in \mathbb{N}^*} q^{k-1} p \\ &= p \sum_{l \in \mathbb{N}} q^l \qquad (l = k-1) \\ &= p \frac{1}{1-q} = 1. \end{aligned}$$

Ainsi avec probabilité 1, le premier succès apparaît au bout d'un nombre *fini* d'épreuves[7]. Remarquons qu'on aurait pu arriver au même résultat en montrant que $P(X = +\infty) = 0$ par la méthode utilisée à l'exemple 2.8 c) en échangeant les rôles de succès et échec. Finalement on peut considérer X comme une variable aléatoire ayant comme ensemble de valeurs possibles $X(\Omega) = \mathbb{N}^*$ au lieu de $\mathbb{N}^* \cup \{+\infty\}$. ■

Définition 3.11 *Une variable aléatoire* X *suit la loi géométrique de paramètre* $p \in]0,1[$, *si* $X(\Omega) = \mathbb{N}^*$ *et :*

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

Notation : $X \sim \mathcal{G}(p)$.

[7] Mais pas borné par un nombre fixé choisi avant le début des épreuves…

Lorsque X suit une loi géométrique, les probabilités $P(X > n)$ ont une expression particulièrement simple en fonction de $q = 1 - p$. Calculons les de deux façons.
Première méthode : On calcule le reste d'une série géométrique :

$$\begin{aligned}
P(X > n) &= \sum_{k=n+1}^{+\infty} q^{k-1}p = \sum_{l=n}^{+\infty} q^l p \\
&= pq^n \sum_{l=n}^{+\infty} q^{l-n} = pq^n \sum_{j=0}^{+\infty} q^j \\
&= \frac{pq^n}{1-q} = q^n.
\end{aligned}$$

Deuxième méthode : On se place dans la situation de l'exemple 3.2. L'événement $\{X > n\}$ se réalise si et seulement si les n premières épreuves donnent un échec.

$$\{X > n\} = \mathop{\cap}_{i=1}^{n} R_i^c.$$

En utilisant l'indépendance des R_i on en déduit :

$$P(X > n) = \prod_{i=1}^{n} P(R_i^c) = q^n.$$

3.3.6 Lois de Poisson

Définition 3.12 *On dit que la variable aléatoire discrète X suit la loi de Poisson de paramètre $\lambda > 0$ si l'ensemble des valeurs possibles est $X(\Omega) = \mathbb{N}$ et*

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

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

On sait (cf. cours d'analyse) que la fonction exponentielle a un développement en série entière avec rayon de convergence infini. En particulier :

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

On a donc bien :

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

Une des raisons de l'importance de cette loi est le théorème de convergence de la loi binomiale vers la loi de Poisson.

Théorème 3.13 *Si* $(p_n)_{n\geq 1}$ *est une suite de réels de* $[0,1]$ *vérifiant*

$$np_n \to \lambda \in]0,+\infty[, \quad \textit{quand} \quad n \to +\infty, \tag{3.15}$$

alors :

$$\forall k \in \mathbb{N}, \quad C_n^k p_n^k (1-p_n)^{n-k} \longrightarrow \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad \textit{quand} \quad n \to +\infty.$$

Preuve : L'hypothèse (3.15) peut s'écrire sous la forme plus maniable : $np_n = \lambda u_n$ avec u_n tendant vers 1 quand n tend vers $+\infty$. Ainsi $p_n = \lambda u_n / n$ et

$$C_n^k p_n^k (1-p_n)^{n-k} = \frac{n!}{(n-k)!}\frac{1}{k!}\Big(\frac{\lambda}{n}\Big)^k u_n^k \Big(1-\frac{\lambda u_n}{n}\Big)^{n-k}. \tag{3.16}$$

Pour obtenir la limite de cette expression lorsque n tend vers $+\infty$, k restant *fixé*, on remarque successivement que :

$$\lim_{n\to+\infty} \frac{n!}{(n-k)!}\frac{1}{n^k} = 1, \tag{3.17}$$

$$\lim_{n\to+\infty} u_n^k = 1, \tag{3.18}$$

$$\lim_{n\to+\infty} \Big(1-\frac{\lambda u_n}{n}\Big)^{n-k} = \mathrm{e}^{-\lambda}. \tag{3.19}$$

Pour justifier (3.19), on écrit :

$$\Big(1-\frac{\lambda u_n}{n}\Big)^{n-k} = \exp\Big[(n-k)\ln\Big(1-\frac{\lambda u_n}{n}\Big)\Big], \tag{3.20}$$

puis comme $\lambda u_n / n$ tend vers 0 :

$$(n-k)\ln\Big(1-\frac{\lambda u_n}{n}\Big) \sim n\Big(-\frac{\lambda u_n}{n}\Big) \sim -\lambda, \quad (n \to +\infty).$$

Par continuité de la fonction exponentielle, la limite du second membre de (3.20) est donc bien $\mathrm{e}^{-\lambda}$, ce qui prouve (3.19). On obtient alors la conclusion du théorème en passant à la limite dans (3.16). ■

Application pratique : Le théorème 3.13 sert de justification théorique à la règle pratique suivante : lorsque n est « grand » et np « petit », on peut remplacer la loi binomiale $\mathcal{B}(n,p)$ par la loi de Poisson $\mathcal{P}(\lambda)$ où $\lambda = np$. En général on considère que n de l'ordre de quelques centaines et np de l'ordre de quelques unités donnent une bonne approximation. Sous cette forme, cette règle relève plus de la cuisine que des mathématiques. Il est possible par des techniques élémentaires (voir exercice 3.19) de contrôler l'erreur commise en utilisant cette approximation. Nous nous contenterons ici d'un exemple classique et d'une comparaison graphique pour illustrer la qualité de cette approximation.

Exemple 3.3 *Le président d'un bureau de vote est né un 1er avril. Il décide de noter le nombre X de personnes ayant leur anniversaire le même jour que lui parmi les 500 premiers électeurs qui se présentent.*

La situation peut être assimilée à une suite d'épreuves répétées indépendantes et X est une variable aléatoire suivant la loi binomiale de paramètres $n = 500$ et $p = 1/365$ (en négligeant la question des années bissextiles sinon on prendrait $p = 4/(3 \times 365 + 366)$, ce qui ne changerait pas grand chose numériquement). Ainsi :

$$P(X = k) = C_{500}^k \Big(\frac{1}{365}\Big)^k \Big(\frac{364}{365}\Big)^{500-k}.$$

La règle énoncée ci-dessus nous conduit à approximer la loi de X par une loi de Poisson de paramètre :

$$\lambda = np = 500 \times \frac{1}{365}.$$

Voici une comparaison numérique pour les petites valeurs de k :

k	0	1	2	3	4	5
$P(X = k)$	0.2537	0.3484	0.2388	0.1089	0.0372	0.0101
$\dfrac{e^{-\lambda}\lambda^k}{k!}$	0.2541	0.3481	0.2385	0.1089	0.0373	0.0102

Remarquons que la probabilité d'observer plus de 5 anniversaires un 1er avril, calculée par la loi exacte de X ou par son approximation poissonienne est inférieure à 0.003.

Comparaison graphique :

Les *diagrammes en bâtons* ci-dessous représentent la loi binomiale $\mathcal{B}(n, p)$ et la loi de Poisson approximante $\mathcal{P}(\lambda)$ avec $\lambda = np$. Les segments verticaux (les bâtons) du diagramme représentant la loi d'une variable discrète X (à valeurs dans $\mathbb{N}$) ont une hauteur égale à $P(X = k)$ avec une extrémité inférieure au point d'abscisse k de l'axe horizontal. Pour la lisibilité, on a légèrement décalé vers la gauche les bâtons de la loi de Poisson (en bleu) et vers la droite ceux de la loi binomiale(en rouge). Bien que le diagramme en bâtons de la loi binomiale $\mathcal{B}(n, p)$ soit constitué théoriquement de $n+1$ bâtons (et que celui de la loi de Poisson en ait une infinité), seul un petit nombre de bâtons est visible sur les graphiques, les autres correspondant à des probabilités trop petites[8]. L'échelle verticale de chaque figure a été choisie de façon adaptative de façon que l'avant dernière graduation verticale donne la valeur de la plus grande probabilité binomiale. On constate que pour $n = 200$ (figure 3.4), la différence entre les deux diagrammes n'est pratiquement plus discernable *visuellement*.

[8]En fait, on s'est contenté d'afficher les probabilités correspondant à k inférieur ou égal à la partie entière supérieure de $2\lambda + 4$. Il est facile de vérifier (cf. exercices 3.18 et 3.19) que la somme des probabilités ainsi négligées est inférieure à 1%, pour chacune des deux lois.

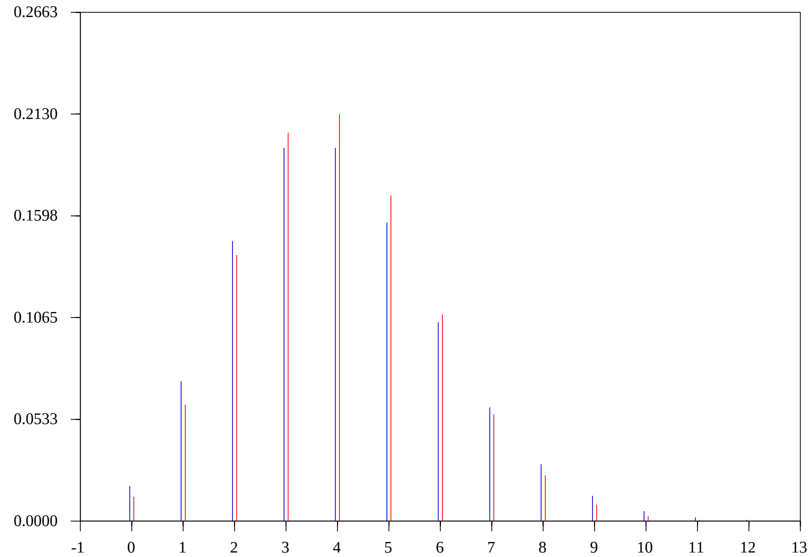


FIG. 3.1 – Lois $\mathcal{B}(25, 0.16)$ et $\mathcal{P}(4)$

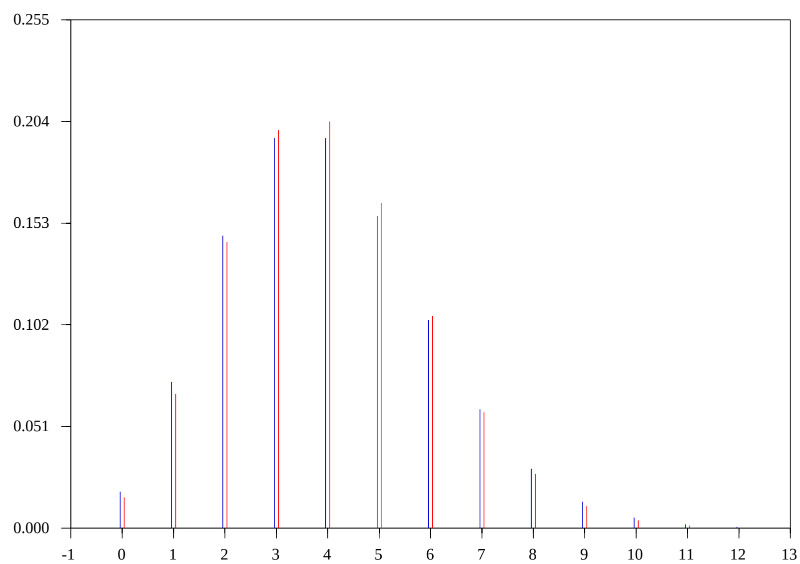


FIG. 3.2 – Lois $\mathcal{B}(50, 0.08)$ et $\mathcal{P}(4)$

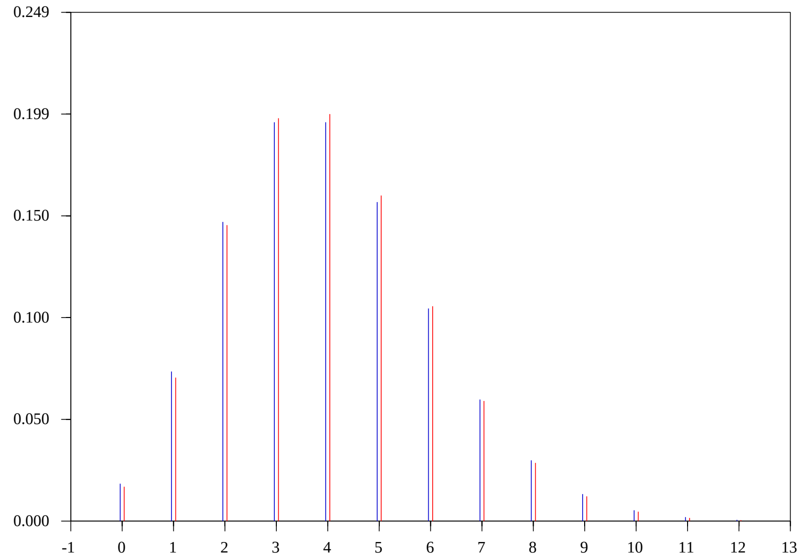


FIG. 3.3 – Lois $\mathcal{B}(100, 0.04)$ et $\mathcal{P}(4)$

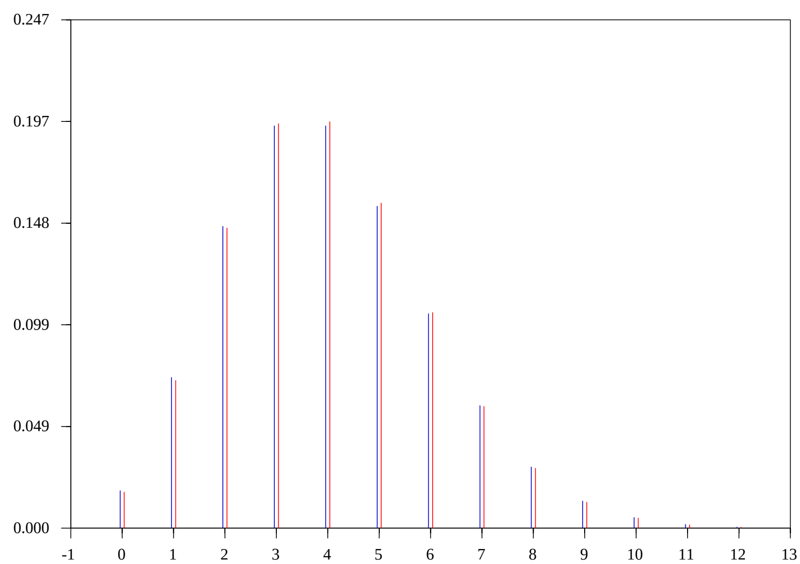


FIG. 3.4 – Lois $\mathcal{B}(200, 0.02)$ et $\mathcal{P}(4)$

3.3.7 Sur le caractère universel de la loi de Poisson

L'étude qui suit a pour but de mieux faire saisir l'importance de la loi de Poisson, en justifiant au passage le bien fondé de l'hypothèse (3.15) du théorème de convergence de la loi binomiale vers la loi de Poisson.

Considérons un phénomène se traduisant par des observations (ou réalisations) aléatoires pendant un intervalle de temps $[0,1[$ (exemples : désintégrations d'atomes, accidents d'avion, faux numéros de téléphone sur un standard, éruptions volcaniques, naissances de triplés,...). On suppose que le phénomène vérifie les hypothèses suivantes :

(a) Les observations dans des intervalles de temps disjoints sont indépendantes ;

(b) Pour tout réel t tel que $0 \leq t < t+T \leq 1$ la loi du nombre (aléatoire) d'observations dans l'intervalle $[t, t+T[$ ne dépend que de la durée T de cet intervalle.

Partageons l'intervalle de temps $[0,1[$ en n intervalles disjoints

$$I_{n,k} = \Big[\frac{k}{n}, \frac{k+1}{n}\Big[, \quad 0 \leq k < n.$$

Notons :

$$\begin{aligned} p_n &= P\big(\text{avoir } \textit{exactement} \text{ une observation dans } I_{n,k}\,\big)\\ r_n &= P\big(\text{avoir } \textit{au moins} \text{ une observation dans } I_{n,k}\,\big). \end{aligned}$$

D'après (b), p_n et r_n ne dépendent pas de k. En écrivant de deux façons la probabilité de n'avoir aucune observation dans $[0,1[$ on obtient :

$$\begin{aligned} 1 - r_1 &= P\Big(\bigcap_{k=0}^{n-1}\Big\{\text{aucune observation dans } \Big[\frac{k}{n}, \frac{k+1}{n}\Big[\Big\}\Big)\\ &= (1-r_n)^n \qquad \text{en utilisant } (a) \text{ et } (b). \end{aligned}$$

D'où $1-r_n = (1-r_1)^{1/n}$. Un développement limité à l'ordre 1 de cette expression nous permet d'écrire :

$$(1-r_1)^{1/n} = \exp\Big(\frac{1}{n}\ln(1-r_1)\Big) = 1 + \frac{1}{n}\ln(1-r_1) + \frac{\delta_n}{n}, \quad \text{où } \lim_{n\to+\infty}\delta_n = 0.$$

Nous noterons désormais :

$$-\ln(1-r_1) = \lambda, \qquad \lambda \in]0, +\infty[. \tag{3.21}$$

Il vient $r_n = \frac{\lambda}{n} - \frac{\delta_n}{n}$ d'où $\lim_{n\to+\infty} nr_n = \lambda$.

Pour le type de phénomène que nous envisageons, il est vraisemblable que l'on arrive à isoler les observations lorsque les intervalles de la subdivision sont assez petits : asymptotiquement, la probabilité d'avoir plus d'une observation dans $[k/n, (k+1)/n[$ est négligeable devant celle d'en avoir exactement une. Plus précisément, nous rajoutons à notre modèle l'hypothèse :

(c) $\varepsilon_n = \dfrac{r_n - p_n}{p_n} \longrightarrow 0$, quand $n \to +\infty$.

D'après (c), r_n/p_n converge vers 1, d'où $\lim_{n\to+\infty} np_n = \lambda$.

Cherchons maintenant la probabilité d'avoir exactement l observations dans $[0,1[$. Cette probabilité peut se décomposer en :

$$P(l \text{ observations dans } [0,1[) = P(A_n) + P(B_n), \tag{3.22}$$

où

$$\begin{aligned} A_n &= \big\{ l \text{ observations avec au plus une dans chaque } \ I_{n,k} \big\}, \\ B_n &= \big\{ l \text{ observations avec au moins un } I_{n,k} \text{ en contenant plusieurs} \big\}. \end{aligned}$$

Calcul de $P(A_n)$: Notons

$$\begin{aligned} D_i &= \Big\{ \text{exactement une observation dans} \Big[\frac{i}{n}, \frac{i+1}{n}\Big[\Big\}, \\ E_i &= \Big\{ \text{aucune observation dans} \Big[\frac{i}{n}, \frac{i+1}{n}\Big[\Big\}, \quad 0 \le i < n. \end{aligned}$$

L'événement A_n est la réunion disjointe de tous les événements du type :

$$\Big(\bigcap_{i\in I} D_i \Big) \cap \Big(\bigcap_{j\in J} E_j \Big),$$

où $I \subset \{1,\dots,n\}$, $\operatorname{card} I = l$ et $J = \{1,\dots,n\} \setminus I$. D'après l'hypothèse d'indépendance (a), la probabilité de chacun de ces événements est $p_n^l(1-r_n)^{n-l}$ d'où :

$$P(A_n) = C_n^l p_n^l (1-r_n)^{n-l}.$$

Pour trouver la limite de $P(A_n)$ lorsque n tend vers l'infini, l restant fixé, il suffit d'adapter la preuve du théorème 3.13 : ici nous avons à trouver la limite de $(1-r_n)^{n-l}$ au lieu de $(1-p_n)^{n-l}$. Or

$$(n-l)\ln(1-r_n) \sim -nr_n \sim -\lambda,$$

d'où $\lim_{n\to+\infty}(1-r_n)^{n-l} = \mathrm{e}^{-\lambda}$. On en déduit :

$$\lim_{n\to+\infty} P(A_n) = \frac{\mathrm{e}^{-\lambda}\lambda^l}{l!}. \tag{3.23}$$

Majoration de $P(B_n)$: Le calcul de $P(B_n)$ étant trop compliqué, nous nous contenterons d'une majoration. La réalisation de l'événement B_n implique l'existence d'au moins deux observations dans au moins l'un des intervalles de longueur $1/n$. Autrement dit :

$$B_n \subset \bigcup_{k=0}^{n-1} \Big\{ \text{au moins deux observations dans} \Big[\frac{k}{n}, \frac{k+1}{n}\Big[\Big\}.$$

Par conséquent

$$\begin{aligned} P(B_n) &\leq P\Big(\bigcup_{k=0}^{n-1}\Big\{\text{au moins deux observations dans}\Big[\frac{k}{n},\frac{k+1}{n}\Big[\Big\}\Big) \\ &\leq \sum_{k=0}^{n-1}(r_n-p_n) = n(r_n-p_n) = np_n\varepsilon_n. \end{aligned}$$

D'après (c) et la convergence de np_n vers λ, $np_n\varepsilon_n$ tend vers 0 quand n tend vers $+\infty$. Il en est donc de même pour $P(B_n)$.

Pour conclure, on remarque que (3.22) est vérifiée pour tout entier $n \geq 1$ et que le premier membre de cette égalité ne dépend pas de n. Cette égalité reste donc vraie à la limite :

$$P(l \text{ observations dans } [0,1[) = \lim_{n\to+\infty}\big(P(A_n)+P(B_n)\big) = \frac{\mathrm{e}^{-\lambda}\lambda^l}{l!},$$

d'après (3.23) et la majoration de $P(B_n)$. Ce résultat étant valable pour tout entier l, nous avons donc démontré :

Théorème 3.14 *Soit un phénomène donnant lieu à des observations aléatoires vérifiant les hypothèses :*

(a) *Les observations dans des intervalles de temps disjoints sont indépendantes ;*

(b) *Pour tout réel t tel que $0 \leq t < t+T \leq 1$ la loi du nombre (aléatoire) d'observations dans l'intervalle $[t, t+T[$ ne dépend que de la durée T de cet intervalle.*

(c) *En notant p_n la probabilité d'avoir exactement une observation dans un intervalle de temps de durée $1/n$ et r_n celle d'en avoir au moins une, $\varepsilon_n = \dfrac{r_n-p_n}{p_n} \longrightarrow 0$, quand $n \to +\infty$.*

Alors le nombre aléatoire d'observations dans l'intervalle $[0,1[$ suit la loi de Poisson de paramètre λ défini par

$$\lambda = -\ln(1-r_1).$$

Remarque : L'examen attentif de la démonstration ci-dessus montre que la structure d'ordre de l'intervalle $[0,1[$ n'y joue aucun rôle. L'important est la possibilité de réaliser une partition de $[0,1[$ en intervalles de même longueur tendant vers 0. Par conséquent en remplaçant la longueur par l'aire ou le volume, il est possible d'obtenir une version spatiale en dimension 2 ou 3 du théorème 3.14. Ceci permet de comprendre pourquoi la loi de Poisson fournit une bonne modélisation par exemple du nombre d'erreurs typographiques dans une page imprimée, du nombre d'impacts de météorites sur un territoire donné, du nombre d'accidents sur une portion d'autoroute pendant une période donnée, du nombre de raisins dans une portion de cake, du nombre d'étoiles dans une région de l'univers, ...

3.4 Exercices

Ex 3.1. *La propriété d'absence de mémoire*

1) Montrer que si X est une v. a. de loi géométrique, elle vérifie la propriété d'absence de mémoire suivante :

$$\forall k \in \mathbb{N},\ \forall n \in \mathbb{N}, \qquad P(X > n + k \mid X > n) = P(X > k). \tag{3.24}$$

Interpréter ce résultat en considérant une suite d'épreuves répétées.

2) Trouver toutes les lois qui vérifient la propriété (3.24).
Indication : On notera $G(n) = P(X > n)$ et on montrera que (3.24) se traduit par une relation simple entre $G(n + k)$, $G(n)$ et $G(k)$.

Ex 3.2.

1) Proposer des formules permettant de simplifier les expressions :

$$F(x, y) = \sum_{0\leq 2k\leq n} C_n^{2k} x^{2k} y^{n-2k}, \qquad G(x, y) = \sum_{0<2k+1\leq n} C_n^{2k+1} x^{2k+1} y^{n-2k-1}.$$

2) Soit X une variable aléatoire suivant la loi binomiale $\mathcal{B}(n, p)$. Calculer : $P(X \text{ pair})$.

Ex 3.3. *Code de la Route I.*

Pour l'examen du Code de la Route, les candidats doivent remplir un questionnaire de 40 questions en choisissant pour chacune d'elles l'une des 4 réponses proposées, dont une seule est exacte. Un candidat totalement ignorant décide de tenter sa chance en cochant complètement au hasard une réponse pour chaque question.

1) Quelle est la loi du nombre S de bonnes réponses du candidat ? Justifiez votre réponse.

2) Calculer $P(S \geq 36)$.

Ex 3.4. *Code de la Route II.*

Le modèle précédent est trop simpliste, en voici un plus réaliste. Le candidat n'est pas complètement ignorant et il arrive qu'il connaîsse la réponse à certaines questions. Il répond alors à coup sûr. S'il ignore la réponse, il choisit au hasard entre les 4 réponses proposées. On suppose toutes les questions indépendantes et que pour chacune de ces questions, la probabilité que le candidat connaisse la vraie réponse est p. Ce paramètre p mesure donc le vrai niveau du candidat. Malheureusement il n'est pas connu de l'examinateur. On définit les variables aléatoires suivantes :

S nombre de réponses connues du candidat ;
T nombre de bonnes réponses obtenues au hasard ;
U nombre total de bonnes réponses ($U = S + T$) ;

V nombre de mauvaises réponses ($S + T + V = 40$).

1) Quelle est la loi de S ?

2) On note B_i l'évènement « le candidat donne la bonne réponse à la $i^{\text{ième}}$ question ». Calculer $P(B_i)$ en fonction de p (on conditionnera par les deux cas possibles).

3) En déduire la loi de U, puis celle de V.

4) Utiliser une méthode analogue pour trouver la loi de T.

Ex 3.5. *Code de la Route III.*
Ce que peut réellement observer l'examinateur, c'est la valeur prise par U (ou ce qui revient au même par V). Les quantités intéressantes pour tirer des conclusions sur le niveau réel du candidat sont les $P(S = i \mid U = m)$ pour $i \leq m \leq 40$. Par exemple si le candidat a obtenu 38 bonnes réponses, on aimerait évaluer $P(S \geq 36 \mid U = 38)$.

1) Expliquer pourquoi pour $i \leq m \leq 40$,

$$P(S = i, U = m) = P(S = i, T = m - i, V = 40 - m).$$

2) En utilisant le résultat de l'exercice 2.16, en déduire en fonction de i, m et p, l'expression de $P(S = i \mid U = m)$.

3) En déduire l'expression de $P(S \geq 36 \mid U = 38)$.

Ex 3.6. *Contrôleur contre fraudeur*

Une compagnie de métro pratique les tarifs suivants. Le ticket donnant droit à un trajet coûte 5 F ; les amendes sont fixées à 100 F pour la première infraction constatée, 200 F pour la deuxième et 2 000 F pour la troisième. La probabilité p pour un voyageur d'être contrôlé au cours d'un trajet est supposée constante et connue de la seule compagnie. Un fraudeur décide de prendre systématiquement le métro sans payer jusqu'à la deuxième amende et d'arrêter alors de frauder. On note T le nombre de trajets effectués jusqu'à la deuxième amende (T est le numéro du trajet où le fraudeur est contrôlé pour la deuxième fois). On note $q = 1 - p$ la probabilité de faire un trajet sans contrôle.

1) Montrer que la loi de T est donnée par

$$P(T = k) = (k - 1)p^2 q^{k-2}, \quad k \geq 2.$$

2) Pour $n \in \mathbb{N}^*$, calculer $P(T > n)$. *Indication* : on pourra commencer par chercher une formule explicite pour la somme de la série entière

$$f(x) := \sum_{k=n+1}^{+\infty} x^{k-1},$$

puis pour sa dérivée terme à terme.

3) Calculer numériquement $P(T > 60)$ (pourquoi s'intéresse-t-on à cette quantité ?) lorsque $p = 1/10$ et lorsque $p = 1/20$.

4) D'un point de vue purement financier (et donc hors de toute considération de moralité), quel conseil donneriez vous au fraudeur ?

Ex 3.7. *Loi binomiale négative*

1) Soit $n \in \mathbb{N}^*$ fixé. Donner le développement en série entière de la variable q de :

$$f(q) = (1-q)^{-n}, \qquad q \in [0, 1[.$$

Dans la suite, on note $p = 1 - q$.

2) En déduire qu'en posant :

$$\forall k \in \mathbb{N}, \qquad P(X = k) = C_{n+k-1}^{k} p^n q^k,$$

on définit une loi de probabilité sur $\mathbb{N}$. Cette loi s'appelle *loi binomiale négative* de paramètres n et p.

3) On considère une urne contenant n_1 boules vertes et n_2 boules rouges. On note $p = n_1/(n_1 + n_2)$. On effectue des tirages avec remise d'une boule dans l'urne jusqu'à l'obtention de la n-ième boule verte. Soit Y la variable aléatoire égale au nombre de boules rouges ainsi tirées. Quelle est la loi de Y ?

Ex 3.8. On considère une suite infinie d'épreuves répétées indépendantes. La i-ème épreuve peut donner une réussite (évènement R_i) avec probabilité p $(0 < p < 1)$ ou un échec (évènement R_i^c) avec probabilité $q = 1 - p$. On note X le numéro (aléatoire) de l'épreuve où l'on observe la *deuxième* réussite.

1) Écrire les évènements $\{X = 2\}$, $\{X = 3\}$, $\{X = 4\}$ à l'aide des R_i et R_i^c et calculer leurs probabilités respectives.

2) Calculer $P(X = k)$ pour tout entier k.

3) Écrire à l'aide des R_i et R_i^c l'événement

$$\{\text{ on n'observe jamais de deuxième succès}\}$$

et montrer que sa probabilité est nulle.

4) Calculer l'espérance de X (voir chapitre 5).

Ex 3.9. On jette deux dés dont l'un est équilibré, l'autre est truqué de façon inconnue. On note X, Y les nombres de points indiqués respectivement par le dé équilibré et le dé truqué. La variable aléatoire X suit ainsi la loi uniforme sur $\{1, \dots, 6\}$, tandis que la loi de Y nous est inconnue : on sait seulement que l'ensemble des valeurs possibles de Y est $\{1, \dots, 6\}$, mais on ignore les valeurs des $P(Y = j)$ pour $j \in \{1, \dots, 6\}$. On suppose que le truquage *n'affecte pas l'indépendance des deux dés*. On note R la variable aléatoire égale au représentant dans $\{0, \dots, 5\}$ de la classe d'équivalence de $X + Y$ modulo 6.

1) Montrer sans faire de calcul que pour tout $r \in \{0, \dots, 5\}$ et tout $j \in \{1, \dots, 6\}$, il existe *un unique* $i \in \{1, \dots, 6\}$ tel que $i + j = r$ modulo 6.

2) Expliquer pourquoi l'événement $\{R = r\}$ est réunion de 6 événements *deux à deux disjoints* du type $\{X = i, Y = j\}$. Expliciter cette décomposition pour $\{R = 3\}$.

3) Calculer $P(R = r)$ pour $r \in \{0, \dots, 5\}$ et en déduire que R suit la loi uniforme sur $\{0, \dots, 5\}$ et ceci quelle que soit la loi de Y, c'est-à-dire *quel que soit le truquage* du deuxième dé.

Ex 3.10. *Groin chercheur*
Le nombre de truffes trouvées par le cochon Sherlock durant une période de recherche T (en heures) suit une loi de Poisson de paramètre $\lambda(T)$. On sait que $\lambda(1) = 1,7$.

1) Calculer la probabilité que Sherlock ne trouve aucune truffe en une heure.

2) Calculer la probabilité que Sherlock trouve au moins 2 truffes en une heure.

On note maintenant X_1 le nombre de truffes trouvées par Sherlock un jour donné entre 6h et 7h et X_2 le nombre de truffes trouvées par Sherlock le même jour entre 7h et 8h.

3) Quelle est l'espérance de $X_1 + X_2$?

4) En déduire $\lambda(2)$.

5) Les événements $\{X_1 = 0\}$ et $\{X_2 = 0\}$ sont-ils indépendants ?

Ex 3.11. Soit X une v.a. de Poisson de paramètre $\lambda > 0$. On définit la v.a. Y de la manière suivante :

– Si X prend une valeur nulle ou impaire alors Y prend la valeur 0.
– Si X prend une valeur paire alors Y prend la valeur $X/2$.

Trouver la loi de Y.

Ex 3.12. *Mélange de lois*
On suppose que le nombre N d'oeufs pondus par un insecte suit une loi de Poisson de paramètre α :

$$P(N = k) = \mathrm{e}^{-\alpha} \frac{\alpha^k}{k\,!} \qquad k \in \mathbb{N}.$$

On suppose également que la probabilité de développement d'un oeuf est p et que les oeufs sont mutuellement indépendants. On note S le nombre (aléatoire) de survivants. Montrer que S suit une loi de Poisson de paramètre $p\alpha$.

Ex 3.13. Le nombre de fois où un individu attrape un rhume en une année donnée est une variable de Poisson de paramètre $\lambda = 5$. Un nouveau médicament vient d'être mis sur le marché. Il réduit le paramètre de Poisson à $\lambda = 3$ pour 75% de la population. Pour le reste de la population, le médicament est sans effet notable sur les rhumes. Si un individu essaie le médicament pendant un an et attrape 2 rhumes au cours de cette période, quelle est la probabilité que le médicament lui ait été bénéfique ?

Ex 3.14. Un boulanger mélange 1000 raisins dans de la pâte pour fabriquer 100 brioches de même masse.

1) Quelle est la loi exacte du nombre X de raisins contenus dans une brioche achetée chez ce boulanger ? Précisez quelles hypothèses vous faites.

2) En utilisant une *approximation* classique de la loi de X, évaluer la probabilité que la brioche achetée contienne 10 raisins à deux unités près (i.e. $8 \leq X \leq 12$).

Ex 3.15. *Une loi discrète pathologique*[9]
Le but de cet exercice est de construire et d'étudier une v.a. discrète ayant pour ensemble $X(\Omega)$ de valeurs possibles l'ensemble $\mathbb{D}$ des nombres *décimaux* de $[0, 1[$. Un nombre décimal peut se représenter par une suite *finie* de *chiffres décimaux.* Cette représentation n'est pas unique, par exemple :

$$0.375 = 0.3750 = 0.37500 = 0.375000 = \dots$$

On peut aussi l'écrire sous la forme $k10^{-n}$ avec $k \in \mathbb{N}$ et $n \in \mathbb{N}^*$. Si k n'est pas divisible par 10, nous dirons que l'on a la forme réduite (pour l'exemple précédent, $k = 375$ et $n = 3$). Nous construisons X par la procédure suivante.

On dispose de deux urnes. La première contient des boules rouges et des vertes. On note p la proportion de boules rouges ($0 < p < 1$) et q celle des vertes. La deuxième urne contient 10 boules numérotées de 0 à 9. On effectue des tirages avec remise dans la première urne jusqu'à la première apparition d'une rouge. On note N le nombre (aléatoire) de tirages nécessaires. Une fois connue la valeur de N, on effectue N tirages avec remise d'une boule dans la deuxième urne. En notant Y_j le *chiffre* sorti lors du j-ème tirage dans la deuxième urne ($i \leq N$), on forme le nombre décimal :

$$X(\omega) = 0, Y_1(\omega)Y_2(\omega)\dots Y_{N(\omega)}(\omega) = \sum_{j=1}^{N(\omega)} \frac{Y_j(\omega)}{10^j}.$$

1) Quelle est la loi de N ?

2) Soit n fixé. Lorsque l'on effectue n tirages dans la deuxième urne, quelle est la probabilité d'obtenir une suite de n chiffres décimaux choisie à l'avance ?

[9] mais presque…

3) Calculer $P(X = 0.375)$ (attention, il y a une infinité de façons d'obtenir cette valeur !).

4) Soit $d \in D$ et $k10^{-n}$ sa forme réduite. Calculer $P(X = d)$.

5) Vérifier que D est dénombrable en décrivant un procédé permettant de *numéroter* bijectivement les éléments de D par ceux de $\mathbb{N}$.

6) Montrer qu'il n'existe pas de numérotation *croissante* des éléments de $X(\Omega)$ (i.e. vérifiant $\forall k \in \mathbb{N}$, $x_k < x_{k+1}$).

7) Calculer l'espérance de X (voir chapitre 5). Le résultat dépend de p. On vérifiera que dans tous les cas :

$$\frac{9}{20} < \mathbb{E}\,X < \frac{1}{2}.$$

Commenter ce résultat. Comment interpréter les cas limite $p = 1$ et $p = 0$?

8) Soit F la fonction de répartition de X. Montrer que F n'est constante dans aucun intervalle de $[0, 1[$ (si petit soit-il). Ce n'est donc pas une fonction en escaliers (puisqu'elle est constante sur $]-\infty, 0[$ et sur $[1, +\infty[$).

9) Montrer que F n'est continue sur aucun intervalle de $[0, 1[$ (aussi petit soit-il). Cependant F est continue en tout réel non décimal de $[0, 1[$ (il y en a une infinité non dénombrable).

10) Détailler le calcul de $P(X \leq 0.375 \mid N = j)$ en distinguant les deux cas $j < 3$ et $j \geq 3$.

11) Généraliser en montrant que :

$$\forall d \in D, \qquad P(X \leq d \mid N = j) = \frac{[10^j d] + 1}{10^j}.$$

12) Calculer $F(k/10)$ pour $0 \leq k \leq 9$ et $F(l/100)$ pour $0 \leq l \leq 99$.

13) Peut-on donner une formule générale pour $F(d)$, $d \in D$?

Ex 3.16. *Le terme central de la loi binomiale*[10]

Soit $p \in]0, 1[$ et $q = 1 - p$. Dans tout ce qui suit on note :

$$b(k, n, p) = C_n^k p^k q^{n-k} \quad \text{pour} \quad 0 \leq k \leq n,$$

k et n étant entiers. Ainsi $b(k, n, p)$ est la probabilité qu'une variable binomiale de paramètres n et p prenne la valeur k.

1) Vérifier que pour $k \geq 1$, on a :

$$\frac{b(k, n, p)}{b(k-1, n, p)} = 1 + \frac{(n+1)p - k}{kq}.$$

[10] Cet exercice est purement analytique. Il permet de démontrer un résultat important qui sera utilisé lors de la preuve du théorème de De Moivre Laplace.

En déduire que :

$$\begin{aligned} &\text{si } k \leq (n+1)p, \qquad b(k,n,p) \geq b(k-1,n,p), \\ &\text{si } k > (n+1)p, \qquad b(k,n,p) < b(k-1,n,p). \end{aligned}$$

Pour n et p fixés, le maximum de $b(k,n,p)$ est donc atteint en $k = m$, m étant défini comme l'unique entier tel que :

$$(n+1)p - 1 < m \leq (n+1)p.$$

Si $(n+1)p$ n'est pas entier, l'inégalité de droite est stricte et il y a un seul maximum. Si $(n+1)p$ est un entier on a $m = (n+1)p$ et $b(m-1,n,p) = b(m,n,p)$ réalisent tous deux le maximum. Dans tous les cas, le nombre $b(m,n,p)$ est appelé *terme central de la loi binomiale*. On se propose d'en donner un équivalent simple lorsque n tend vers $+\infty$.

2) Étudier les variations de la fonction :

$$f : [0,1] \to [0,1], \quad x :\mapsto x^m(1-x)^{n-m}.$$

3) Vérifier que $np \in \,](n+1)p - 1, (n+1)p[$. Montrer que :

$$\text{si } np \leq m \leq (n+1)p, \qquad b\Big(m, n, \frac{m}{n+1}\Big) \leq b(m,n,p) \leq b\Big(m,n,\frac{m}{n}\Big). \tag{3.25}$$

4) Montrer de même que :

$$\text{si } (n+1)p - 1 < m < np, \qquad b\Big(m,n,\frac{m+1}{n+1}\Big) \leq b(m,n,p) \leq b\Big(m,n,\frac{m}{n}\Big). \tag{3.26}$$

5) On rappelle la formule de Stirling : pour tout $j \in \mathbb{N}^*$,

$$j! = \sqrt{2\pi} j^{j+1/2} \mathrm{e}^{-j} \mathrm{e}^{\theta_j} \quad \text{où} \quad \frac{1}{12j+1} < \theta_j < \frac{1}{12j}.$$

En utilisant cette formule, donner un équivalent des bornes des encadrements (3.25) et (3.26). En déduire que :

$$b(m,n,p) \sim \frac{1}{\sqrt{2\pi npq}}, \quad (n \to +\infty). \tag{3.27}$$

En remplaçant les équivalents par des encadrements, et en se fatiguant un peu plus, on pourrait montrer qu'il existe une constante C indépendante de p et un entier n_0 dépendant de p tels que :

$$\forall n \geq n_0, \quad b(m,n,p) = \frac{1}{\sqrt{2\pi npq}}(1+\varepsilon_n) \quad \text{avec} \quad |\varepsilon_n| \leq \frac{C}{npq}. \tag{3.28}$$

Ex 3.17. Que pensez vous de l'affirmation suivante : « Si on lance un grand nombre (pair) de fois une pièce équilibrée, il y a une forte probabilité d'obtenir exactement autant de piles que de faces » ?

Ex 3.18. *La queue de la loi de Poisson*[11]

1) Si X suit la loi de Poisson de paramètre $\lambda > 0$, montrer que :

$$\forall k > \lambda - 1, \qquad P(X \geq k) < P(X = k)\frac{k+1}{k+1-\lambda}.$$

Indication : On part de :

$$\sum_{j=k}^{+\infty} \frac{\lambda^j}{j!} = \frac{\lambda^k}{k!}\Big(1 + \frac{\lambda}{k+1} + \frac{\lambda^2}{(k+1)(k+2)} + \frac{\lambda^3}{(k+1)(k+2)(k+3)} + \cdots\cdots\Big)$$

et on majore la parenthèse par la somme d'une série géométrique...

2) En déduire que :

$$\forall k \geq 2\lambda - 1, \qquad P(X > k) < P(X = k).$$

Ex 3.19. *Contrôle de l'erreur dans l'approximation poissonienne*
On se propose de donner des résultats quantitatifs sur l'approximation de la probabilité binomiale $b(k, n, p)$ par $\mathrm{e}^{-\lambda}\lambda^k/k!$ où $\lambda = np$.

1) Justifier l'encadrement suivant :

$$\forall u \in [0, 1[, \qquad \exp\Big(-u - \frac{u^2}{2(1-u)}\Big) \leq 1 - u \leq \exp(-u). \tag{3.29}$$

Indication : Dans le développement en série entière de $\ln(1-u)$, contrôler le reste de rang 2 par une série géométrique.

2) En déduire que si $0 \leq k \leq n$,

$$C_n^k = \frac{n^k}{k!}\prod_{j=1}^{k-1}\Big(1 - \frac{j}{n}\Big) \leq \frac{n^k}{k!}\exp\Big(-\frac{(k-1)k}{2n}\Big).$$

3) En déduire que si $n \geq 2$ et $0 \leq k \leq n$:

$$b(k, n, p) \leq \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}\exp\Big(\frac{k}{2n}(2\lambda + 1 - k)\Big).$$

En particulier :

$$\forall k \geq 2\lambda + 1, \qquad b(k, n, p) \leq \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}.$$

En combinant cette inégalité avec le résultat de l'exercice 3.18, on en déduit la majoration suivante de la queue de la loi binomiale :

$$\forall k \geq 2\lambda + 1, \qquad \sum_{j=k+1}^{n} b(j, n, p) \leq \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}. \tag{3.30}$$

[11] La « queue » de la loi d'une v.a. X est la fonction $t \mapsto P(X > t)$.

4) Montrer en utilisant l'encadrement de la question 1 et la formule de Stirling (cf. exercice 3.16) que l'on a pour $0 \leq k < n$ la minoration :

$$b(k,n,p) \geq \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \exp\left\{\frac{1}{2n}\left[(2\lambda+1)k - 2k^2 - \lambda^2\frac{n-k}{n-\lambda} - \frac{12k+1}{72(n-k)}\right]\right\}.$$

Ex 3.20. *Assurances maritimes et concentration du capital*
Une compagnie d'assurances assure une flotte de 500 navires de pêche valant chacun 6 millions de francs. Le risque assuré est la perte totale du navire qui est un événement de probabilité 0.001 pour une année. Les naufrages des différents navires sont considérés comme indépendants. On note X la variable aléatoire égale au nombre de navires perdus en une année.

1) Trouver la loi exacte de X.

2) Évaluer $P(X = 10)$ en utilisant une approximation de la loi de X.

3) La compagnie rembourse le 31 décembre, sur ses réserves, la valeur des bateaux assurés ayant fait naufrage dans l'année. A combien doivent s'élever ces réserves financières pour qu'elle puisse effectuer la totalité de ces remboursements avec une probabilité supérieure à 0.999 ? *Indication :* En utilisant l'inégalité (3.30) de l' exercice 3.19, montrer qu'il suffit que ces réserves représentent la valeur d'un très petit nombre de navires et faire un calcul basé sur la loi exacte de X.

4) Comparer avec ce que donnerait l'approximation poissonnienne.

5) La compagnie fusionne avec une autre compagnie identique (500 navires assurés). Reprendre brièvement les questions 1) et 3) et commenter le résultat obtenu.

Chapitre 4

Vecteurs aléatoires discrets

4.1 Introduction

Dans des situations où interviennent plusieurs variables aléatoires, le calcul de la probabilité d'un événement dont la réalisation dépend des valeurs de ces variables doit faire intervenir ces variables considérées *dans leur ensemble* et non chacune isolément. On est amené à étudier ainsi une nouvelle notion, celle de *vecteur aléatoire*. Commençons par préciser cette idée sur un exemple élémentaire.

Exemple 4.1 *Une urne contient* 7 *boules :* 2 *bleues,* 3 *blanches et* 2 *rouges. On en prélève* 3 *d'un coup. On note respectivement* X *et* Y *les nombres de boules bleues et blanches dans l'échantillon tiré. Calculer les probabilités suivantes :* $P(X > Y)$*,* $P(X = Y)$*,* $P(2$ *rouges*$)$*.*

Il est clair que ces deux variables suffisent à décrire complètement l'expérience puisque la composition de l'échantillon est déterminée par les valeurs de X et Y : le nombre de rouges étant $(3 - X - Y)$. L'espace probabilisé associé naturellement à cette expérience est l'ensemble de tous les échantillons possibles (il y en a $C_7^3 = 35$) muni de l'équiprobabilité. Les valeurs possibles du *couple aléatoire* (X, Y) sont dans l'ensemble $\{0, 1, 2\} \times \{0, 1, 2, 3\}$. Les probabilités d'observation de ces valeurs $P(X = i, Y = j)$ se calculent facilement en faisant du dénombrement.

$$\begin{aligned} P(X = i, Y = j) &= P(i \text{ bleues, } j \text{ blanches, } (3 - i - j) \text{ rouges}) \\ &= \frac{C_2^i C_3^j C_2^{3-i-j}}{C_7^3}, \end{aligned}$$

en faisant la convention $C_n^k = 0$ lorsque $k < 0$ ou $k > n$.

$\downarrow i \backslash j \rightarrow$	0	1	2	3	$P(X=i)$
0	0	$\frac{3}{35}$	$\frac{6}{35}$	$\frac{1}{35}$	$\frac{10}{35}$
1	$\frac{2}{35}$	$\frac{12}{35}$	$\frac{6}{35}$	0	$\frac{20}{35}$
2	$\frac{2}{35}$	$\frac{3}{35}$	0	0	$\frac{5}{35}$
$P(Y=j)$	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{12}{35}$	$\frac{1}{35}$	1

Les valeurs de $P(X=i, Y=j)$.

Le tableau des $P(X=i, Y=j)$ permet le calcul des *lois marginales* P_X et P_Y. Plus généralement, on peut à l'aide de ce tableau calculer la probabilité de tout événement dont la réalisation ne dépend que de la valeur du couple (X, Y). Ainsi :

$$\begin{aligned}
P(X > Y) &= \frac{2}{35} + \frac{2}{35} + \frac{3}{35} = \frac{1}{5}, \\
P(X = Y) &= 0 + \frac{12}{35} + 0 = \frac{12}{35}, \\
P(2 \text{ rouges}) &= P(X + Y = 1) = \frac{2}{35} + \frac{3}{35} = \frac{1}{7}.
\end{aligned}$$

4.2 Vecteurs aléatoires

Définition 4.1 *Soient X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. L'application :*

$$\Omega \rightarrow \mathbb{R}^2, \quad \omega \mapsto \big(X(\omega), Y(\omega)\big)$$

est appelée couple aléatoire discret de marginales X et Y et notée (X, Y).

Définition 4.2 *De même si $X_1, \dots, X_m$ sont des v.a. discrètes sur le même $(\Omega, \mathcal{F}, P)$, on définit le vecteur aléatoire $(X_1, \dots, X_m)$ comme l'application :*

$$\Omega \rightarrow \mathbb{R}^m, \quad \omega \mapsto \big(X_1(\omega), \dots, X_m(\omega)\big)$$

La v.a. X_i est la i-ème marginale du vecteur.

Le couple aléatoire (X, Y) permet de transporter la probabilité P de $\mathcal{F}$ sur l'ensemble des parties de $\mathbb{R}^2$.

Définition 4.3 *La loi $P_{X,Y}$ du couple (X, Y) est la probabilité définie sur l'ensemble des parties de $\mathbb{R}^2$ par :*

$$\forall B \subset \mathbb{R}^2, \qquad P_{X,Y}(B) = P\Big(\{\omega \in \Omega,\ \big(X(\omega), Y(\omega)\big) \in B\}\Big).$$

Les lois P_X et P_Y des v.a. X et Y sont appelées lois marginales du couple.

Dans la suite les ensembles de valeurs possibles pour les v.a. marginales X et Y seront notés :

$$X(\Omega) = \{x_0, x_1, \dots, x_i, \dots\} \quad \text{et} \quad Y(\Omega) = \{y_0, y_1, \dots, y_j, \dots\}.$$

Il est facile de voir que la loi du couple (X, Y) est caractérisée par les probabilités $P_{X,Y}\big(\{(x_i, y_j\}\big) = P(X = x_i, Y = y_j)$, pour $x_i \in X(\Omega)$, $y_j \in Y(\Omega)$.

Proposition 4.4 *Si (X, Y) est un couple aléatoire, ses lois marginales P_X et P_Y peuvent se calculer par :*

$$\forall x_i \in X(\Omega), \quad P(X = x_i) = \sum_{y_j \in Y(\Omega)} P(X = x_i, Y = y_j), \tag{4.1}$$

$$\forall y_j \in Y(\Omega), \quad P(Y = y_j) = \sum_{x_i \in X(\Omega)} P(X = x_i, Y = y_j). \tag{4.2}$$

Preuve : Il suffit de faire la vérification pour (4.1), celle de (4.2) est analogue en échangeant les rôles de X et Y. Pour i fixé, l'événement $\{X = x_i\}$ est la réunion de la famille dénombrable d'événements 2 à 2 disjoints $\{X = x_i, Y = y_j\}$ (pour tous les j tels que $y_j \in Y(\Omega)$). La relation (4.1) s'en déduit par σ-additivité. ■

Remarque : La connaissance de la loi du couple (X, Y) permet de calculer les lois marginales. Il importe de bien comprendre que la réciproque est fausse. *Il n'est généralement pas possible de calculer la loi $P_{X,Y}$ du couple aléatoire (X, Y) à partir de la seule connaissance de ses lois marginales P_X et P_Y.* Voici un exemple élémentaire de deux couples aléatoires ayant mêmes lois marginales sans avoir même loi (voir aussi l'exercice 4.1).

Exemple 4.2 *On jette un dé bleu et un rouge. On note X les points indiqués par le dé bleu, Y ceux du dé rouge et on pose $Z = 7 - X$. Alors les couples aléatoires (X, Y) et (X, Z) ont mêmes lois marginales mais pas même loi.*

En effet il est clair que X, Y et Z suivent chacune la loi uniforme sur $\{1, 2, 3, 4, 5, 6\}$. Les lois des couples sont données par les tableaux suivants

	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Loi de (X, Y)

	1	2	3	4	5	6
1	0	0	0	0	0	$\frac{1}{6}$
2	0	0	0	0	$\frac{1}{6}$	0
3	0	0	0	$\frac{1}{6}$	0	0
4	0	0	$\frac{1}{6}$	0	0	0
5	0	$\frac{1}{6}$	0	0	0	0
6	$\frac{1}{6}$	0	0	0	0	0

Loi de (X, Z)

■

L'extension au cas des vecteurs aléatoires de la définition 4.3 et de la proposition 4.4 ne pose pas d'autre problème que l'alourdissement des notations et est laissée au lecteur.

4.3 Variables aléatoires indépendantes

Définition 4.5 (Indépendance de deux v.a.)
Deux variables aléatoires discrètes X et Y sont dites indépendantes si pour tous sous-ensembles A et B de $\mathbb{R}$, les événements $\{X \in A\}$ et $\{Y \in B\}$ sont indépendants :

$$\forall A \subset \mathbb{R},\ \forall B \subset \mathbb{R} \quad P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \tag{4.3}$$

ce qui s'écrit aussi : $P_{X,Y}(A \times B) = P_X(A)P_Y(B)$.

Dans l'exemple 4.2 ci-dessus, les variables aléatoires X et Y sont indépendantes, les v.a. X et Z ne le sont pas. Il s'agit là de deux situations extrêmes. D'un côté la connaissance de la valeur prise par X ne nous apporte aucune information sur la valeur prise par Y (indépendance). De l'autre, la connaissance de la valeur de X détermine complètement la valeur de Z (dépendance fonctionnelle). La situation générale est intermédiaire entre ces deux cas extrêmes (voir un exemple à l'exercice 4.1).

Définition 4.6 (Indépendance d'une famille finie de v.a.)
Les m variables aléatoires discrètes $X_1, \dots, X_m$ sont dites indépendantes si pour toutes parties $A_1, \dots, A_m$ de $\mathbb{R}$, les événements $\{X_1 \in A_1\}, \dots, \{X_m \in A_m\}$ sont mutuellement indépendants.

Définition 4.7 (Indépendance d'une suite de v.a.)
Une suite $(X_n)_{n\in\mathbb{N}}$ de variable aléatoires discrètes est dite indépendante si toute sous-suite finie est indépendante au sens de la définition 4.6.

Dans une suite d'épreuves répétées indépendantes, si pour tout $i \geq 1$, X_i est une variable aléatoire dont la valeur ne dépend que du résultat de la i-ème épreuve (par exemple points obtenus lors du i-ème lancer d'un dé), la suite $(X_i)_{i\geq 1}$ est indépendante.

Pour vérifier l'indépendance de deux variables aléatoires discrètes, on dispose du critère calculatoire suivant :

Proposition 4.8 *Les variables aléatoire discrètes X et Y sont indépendantes si et seulement si :*

$$\forall x_i \in X(\Omega),\ \forall y_j \in Y(\Omega),\ P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j). \tag{4.4}$$

Preuve : D'abord il est clair que l'indépendance implique (4.4), il suffit de choisir $A = \{x_i\}$ et $B = \{y_j\}$ dans (4.3).

Réciproquement, supposons (4.4) vérifiée. Pour montrer l'indépendance de X et Y, on cherche à vérifier (4.3) pour deux parties A et B quelconques de $\mathbb{R}$.

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{(x_i,y_j)\in A\times B} P(X = x_i, Y = y_j) \\ &= \sum_{(x_i,y_j)\in A\times B} P(X = x_i)P(Y = y_j) \\ &= \sum_{x_i\in A} P(X = x_i) \sum_{y_j\in B} P(Y = y_j) \qquad (4.5) \\ &= P(X \in A)P(Y \in B). \end{aligned}$$

Dans (4.5) nous avons utilisé les propriétés des séries à termes positifs (sommation par paquets, produit de deux séries). ■

Remarque : Si l'on connaît la loi de X et celle de Y *et si l'on sait que* X *et* Y *sont indépendantes*, alors on peut reconstruire la loi du couple (X, Y) à partir des lois marginales. Il suffit d'utiliser (4.4). Au risque d'insister lourdement, rappelons que cette reconstruction n'est pas possible en général à partir de la *seule* connaissance des lois de X et Y.

La proposition 4.8 peut se généraliser au cas des vecteurs aléatoires (énoncé et démonstration laissés en exercice).

Proposition 4.9 *Soient X et Y deux variables aléatoires discrètes indépendantes, f et g deux fonctions dont les domaines de définition contiennent respectivement $X(\Omega)$ et $Y(\Omega)$. Alors les variables aléatoires discrètes $f(X)$ et $g(Y)$ sont indépendantes.*

Preuve : D'abord précisons la notation $f(X)$. Il s'agit en fait de l'application $f \circ X$:

$$f \circ X : \ \Omega \to \mathbb{R} \quad \omega \mapsto f\big(X(\omega)\big).$$

L'ensemble des valeurs possibles $f\big(X(\Omega)\big) = \{f(x_0), f(x_1), \dots\}$ est au plus dénombrable [1]. $f(X)$ est donc bien une variable aléatoire discrète (et de même pour $g(Y)$).

Pour prouver l'indépendance de $f(X)$ et $g(Y)$, il suffit d'après la proposition 4.8 de vérifier que pour tous $s \in f\big(X(\Omega)\big)$, et $t \in g\big(Y(\Omega)\big)$,

$$P\big(f(X) = s, g(Y) = t\big) = P\big(f(X) = s\big)P\big(g(Y) = t\big).$$

[1] On ne suppose pas f injective, on peut donc très bien avoir $X(\Omega)$ infini dénombrable et $f\big(X(\Omega)\big)$ fini. Dans tous les cas, $f\big(X(\Omega)\big)$ est en bijection avec une partie D de $X(\Omega)$ obtenue en regroupant dans la même classe tous les éléments de $X(\Omega)$ ayant même image par f et en choisissant un seul représentant dans chaque classe. Comme $X(\Omega)$ est au plus dénombrable, il en est de même pour $D \subset X(\Omega)$.

En utilisant (4.4) et le produit de deux séries à termes positifs, on obtient :

$$\begin{aligned} P\big(f(X)=s, g(Y)=t\big) &= \sum_{\substack{x_i:f(x_i)=s\\ y_j:g(y_j)=t}} P(X=x_i, Y=y_j) \\ &= \sum_{\substack{x_i:f(x_i)=s\\ y_j:g(y_j)=t}} P(X=x_i)P(Y=y_j) \\ &= \sum_{x_i:f(x_i)=s} P(X=x_i) \sum_{y_j:g(y_j)=t} P(Y=y_j) \\ &= P\big(f(X)=s\big)P\big(g(Y)=t\big). \end{aligned}$$

■

Proposition 4.10 (Somme de deux v.a. indépendantes)
Si X et Y sont deux v. a. indépendantes avec $X(\Omega) \subset \mathbb{N}$ et $Y(\Omega) \subset \mathbb{N}$, la loi de la variable aléatoire $X+Y$ est donnée par :

$$\begin{aligned} \forall n \in \mathbb{N}, \quad P(X+Y=n) &= \sum_{i+j=n} P(X=i)P(Y=j) \\ &= \sum_{i=0}^{n} P(X=i)P(Y=n-i). \end{aligned}$$

Preuve : Comme X et Y sont à valeurs entières, il en est de même pour $X+Y$ dont la loi sera caractérisée par les $P(X+Y=n)$. Pour les calculer, il suffit de décomposer l'événement $\{X+Y=n\}$ en la réunion disjointe de tous les événements $\{X=i, Y=j\}$ tels que $i+j=n$. On en déduit :

$$P(X+Y=n) = \sum_{i+j=n} P(X=i, Y=j) \tag{4.6}$$

$$ = \sum_{i+j=n} P(X=i)P(Y=j), \tag{4.7}$$

ce qui prouve la proposition. ■

Remarque : Si X et Y ne sont pas indépendantes, (4.7) n'est plus valable, on peut seulement écrire (4.6). On voit ainsi que l'on peut toujours calculer la loi de $X+Y$ si l'on connaît la loi du couple (X,Y). Par contre le calcul de la loi de $X+Y$ à partir des lois de X et Y n'est possible que sous l'indépendance. Par ailleurs la méthode utilisée se généralise au cas de variables aléatoires discrètes quelconques (pas forcément à valeurs entières) au prix de quelques complications d'écriture.

Exemple 4.3 *Soient X et Y deux variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs α et β. Déterminer la loi de leur somme S.*
Les lois de X et Y sont données par :

$$P(X=i)=\frac{\mathrm{e}^{-\alpha}\alpha^i}{i!},\quad P(Y=j)=\frac{\mathrm{e}^{-\beta}\beta^j}{j!}.$$

Comme X et Y sont indépendantes :

$$\begin{aligned}P(S=n)=\sum_{i=0}^{n}P(X=i)P(Y=n-i) &= \sum_{i=0}^{n}\frac{\mathrm{e}^{-\alpha}\alpha^i}{i!}\frac{\mathrm{e}^{-\beta}\beta^{n-i}}{(n-i)!}\\ &= \frac{\mathrm{e}^{-(\alpha+\beta)}}{n!}\sum_{i=0}^{n}C_n^i\alpha^i\beta^{n-i}\\ &= \frac{\mathrm{e}^{-(\alpha+\beta)}(\alpha+\beta)^n}{n!}.\end{aligned}$$

Ainsi S suit la loi de Poisson de paramètre $\lambda=\alpha+\beta$. ■

4.4 Exercices

Ex 4.1. On jette un dé bleu et un rouge. On note X les points indiqués par le dé bleu, Y ceux du dé rouge et on définit la variable aléatoire Z de la manière suivante :

$$Z(\omega)=\begin{cases} X(\omega) & \text{si} \quad X(\omega)\leq 3,\\ Y(\omega) & \text{si} \quad X(\omega)>3 \text{ et } Y(\omega)>3,\\ Y(\omega)+3 & \text{si} \quad X(\omega)>3 \text{ et } Y(\omega)\leq 3.\end{cases}$$

Déterminer les lois des couples (X,Y) et (X,Z). Vérifier que ces couples ont mêmes lois marginales mais pas même loi.

Ex 4.2. Montrer que deux variables aléatoires de Bernoulli X et Y sont indépendantes si et seulement si $P(X=1,Y=1)=P(X=1)P(Y=1)$.

Ex 4.3. Soient X et Y indépendantes et de même loi géométrique de paramètre p. Déterminer la loi de $X+Y$.

Ex 4.4. On lance indéfiniment le même dé. Soit X le numéro du premier lancer où l'on obtient un « trois » et Y celui du premier lancer où l'on obtient un « quatre ».

1) Quelle est la loi de X, celle de Y ? Ces deux variables aléatoires sont-elles indépendantes (répondre sans calculer la loi du couple) ?

2) Pour le k-ème lancer ($k \in \mathbb{N}^*$), on note A_k l'obtention d'un « trois », B_k celle d'un « quatre » et C_k celle d'aucun de ces deux chiffres. Exprimer à l'aide d'évènements de ces types, l'évènement $\{X = i, Y = j\}$. En déduire la loi du couple (X, Y).

3) Donner sans calcul (mais en justifiant votre réponse) les valeurs de $P(X < Y)$ et $P(X > Y)$.

4) On définit la variable aléatoire $Z = 3^X 4^Y$. Quelle est la probabilité que Z soit une puissance de 36 ?

Ex 4.5. On considère une suite d'épreuves répétées indépendantes. On note S_i l'événement *succès à la i-ème épreuve* et on pose $p = P(S_i)$, $q = 1 - p$. On définit les variables aléatoires X_1 et X_2 par : X_1 est le numéro de l'épreuve où apparaît le premier succès, $X_1 + X_2$ est le numéro d'apparition du deuxième succès.

1) Pour j et k dans $\mathbb{N}^*$, exprimer l'événement $\{X_1 = j, X_2 = k\}$ à l'aide des S_i. En déduire la loi de (X_1, X_2) puis celle de X_2. Vérifier que X_1 et X_2 sont indépendantes.

2) Calculer $P(X_1 = k \mid X_1 + X_2 = n)$ pour $k = 1, \dots, n - 1$. Commenter le résultat.

Ex 4.6. ↑ 3.7

1) On considère une suite d'épreuves de Bernoulli indépendantes avec pour chacune même probabilité de succès p. Soit Y le nombre aléatoire d'épreuves *avant* le premier succès (si celui ci se produit lors de la première épreuve, $Y = 0$). Quelle est la loi de Y, celle de $Y + 1$?

2) Donner une formule générale permettant de calculer la loi de

$$T_n = \sum_{i=1}^{n} Y_i,$$

où les i sont des v.a. indépendantes de même loi que Y.

3) Cette formule fait intervenir la quantité :

$$N_{n,k} = \sum_{i_1+\cdots+i_n=k} 1 = \operatorname{Card}\{(i_1, \dots, i_n) \in \mathbb{N}^n,\ i_1 + \cdots + i_n = k\},$$

qui peut aussi s'interpréter comme le nombre de façons de disposer k objets identiques dans n cases. Montrer que ce nombre vaut C_{n+k-1}^{k}. *Indication :* pour dénombrer ces dispositions, il est commode d'employer le *codage* suivant qui utilise deux caractères : | et $*$. Le premier sert à délimiter les boites, le deuxième représente l'un des objets. Par exemple si $n = 5$ et $k = 7$, la disposition comprenant deux objets dans la première boite, trois dans la deuxième, aucun dans la troisième, un dans la quatrième et un dans la cinquième est représentée par :

$$| ** | *** || * | * |$$

4) Reconnaître la loi de T_n. En déduire la loi de la somme de n variables indépendantes de même loi géométrique de paramètre p.

5) Soient U et V deux variables aléatoires indépendantes suivant des lois binomiales négatives de paramètres respectifs (n, p) et (m, p). Donner sans calcul la loi de $U + V$.

Ex 4.7. Soit (X, Y, Z) un triplet de variables aléatoires discrètes indépendantes. Soit f une fonction deux variables réelles dont le domaine de définition contient l'ensemble de toutes les valeurs possibles du couple aléatoire (X, Y).

1) Démontrer que les variables aléatoires $f(X, Y)$ et Z sont indépendantes (on pourra s'inspirer de la preuve de la proposition 4.9).

2) En déduire sans calcul la loi de $X + Y + Z$ lorsque X, Y et Z sont indépendantes et suivent des lois de Poisson de paramètres respectifs α, β, γ. Généraliser au cas de n variables de Poisson indépendantes.

3) Déduire de la première question que si $(X_1, \dots, X_n)$ est un vecteurs aléatoire à composantes indépendantes et si I et J sont deux sous ensembles disjoints de $\{1, \dots, n\}$, toute fonction des X_i indexés par I est indépendante de toute fonction des X_j indexés par J. *Indication* : traiter d'abord le cas où $\operatorname{card} J = 1$ en faisant une récurrence sur $\operatorname{card} I$.

Ex 4.8. Un militant entreprend de faire signer une pétition à l'entrée d'un supermarché. Le nombre de personnes X qu'il peut ainsi contacter est une variable aléatoire de Poisson de paramètre α. Soit p la probabilité qu'une personne ainsi sollicitée signe la pétition. On note Y le nombre total de signatures et Z le nombre total de refus de signature ($X = Y + Z$).

1) Soient j et k deux entiers. En distinguant les cas $j > k$ et $j \leq k$, calculer $P(Y = j \mid X = k)$.

2) En déduire $P(X = k, Y = j)$.

3) Déterminer la loi de Y. Les variables aléatoires X et Y sont elles indépendantes ?

4) En utilisant le résultat de la question 2), déterminer la loi du couple (Y, Z).

5) Ces deux variables aléatoires sont elles indépendantes ? Commenter.

Ex 4.9. On considère que le nombre de lapins (des deux sexes) engendrés par une lapine issue d'un élevage de laboratoire est une variable aléatoire S dont la loi est donnée par

$$P(S = n) = pq^n, \qquad n \in \mathbb{N}.$$

1) Quelle(s) relation(s) doivent vérifier les paramètres p et q pour qu'il s'agisse bien d'une loi de probabilité (attention, ce n'est pas exactement une loi géométrique).

2) Jane est une lapine issue de cet élevage. Quelle est la probabilité qu'elle n'engendre aucun lapin ? Qu'elle en engendre au moins 20 ?

3) Soit X le nombre de lapins femelles et Y le nombre de lapins mâles engendrés par Jane ($S = X + Y$). On suppose que la loi de X est telle que :

$$P(X = k \mid S = n) = C_n^k a^k b^{n-k} \quad 0 \leq k \leq n.$$

où les paramètres positifs a et b ne dépendent pas de n et vérifient $a + b = 1$. Interpréter cette relation et la signification des paramètres. Que vaut $P(X = k \mid S = n)$ pour $n < k$?

4) En utilisant la théorie des séries entières, montrer que pour tout k fixé et tout $x \in]-1, +1[$,

$$\sum_{n=k}^{+\infty} n(n-1)\dots(n-k+1)x^{n-k} = \frac{k!}{(1-x)^{k+1}}.$$

5) Calculer $P(X = k)$ pour tout $k \in \mathbb{N}$. En déduire que la loi de X est du même type que celle de S avec des paramètres différents.

6) Donner sans calcul la loi de Y.

7) Pour $i, j \in \mathbb{N}$, discuter les valeurs de :

$$P(X = i \text{ et } Y = j \mid S = n)$$

selon que $n = i+j$ ou $n \neq i+j$. En déduire $P(X = i, Y = j)$. Les deux variables aléatoires X et Y sont elles indépendantes ?

Ex 4.10. *La solitude du gardien de but*
On considère une suite de n épreuves répétées indépendantes avec pour chaque épreuve trois issues possibles : *succès* avec probabilité p, *échec* avec probabilité q ou *nul* avec probabilité r ($p + q + r = 1$). On notera respectivement S_i, E_i et N_i les événements *succès*, *échec*, *nul* à la i-ème épreuve.

1) Dans cette question, $n = 5$. Quelle est la probabilité d'obtenir dans cet ordre 2 succès suivis d'un échec et de 2 nuls ? Quelle est celle d'obtenir (sans condition d'ordre) 2 succès, 1 échec et 2 nuls ?

2) Généraliser en montrant que la probabilité d'obtenir au cours des n épreuves (et sans condition d'ordre) i succès, j échecs et k nuls ($i + j + k = n$) vaut :

$$\frac{n!}{i!\,j!\,k!} p^i q^j r^k.$$

3) On revient au cas $n = 5$ et on note X_ℓ la variable aléatoire codant le résultat de la ℓ-ième épreuve par $X_\ell = 1$ pour un succès, $X_\ell = -1$ pour un échec et $X_\ell = 0$ pour un nul. On définit la variable aléatoire

$$Z = \sum_{\ell=1}^{5} X_\ell.$$

Calculer $P(Z = 0)$.

4) *Application* : Un match de coupe entre deux équipes de football s'étant terminé sur un score nul, l'équipe qualifiée est désignée par la séance des penaltys. Un joueur de l'équipe A tire un penalty face au gardien de l'équipe B, puis un joueur de l'équipe B tire un penalty face à celui de l'équipe A et ainsi de suite jusqu'à ce que chaque équipe ait tiré 5 penaltys. On admet que la probabilité de réussir un penalty est dans chaque cas de $0,7$ et que tous les tirs sont indépendants. Calculer la probabilité que les deux équipes soient encore à égalité après avoir tiré chacune ses 5 penaltys. Calculer la probabilité de qualification de A au bout de ses 5 penaltys.

Ex 4.11. *La loi multinomiale*

1) Montrer que le nombre N de façons de répartir une population de n individus en k groupes d'effectifs respectifs $n_1, \dots, n_k$ où $n_1 + \dots + n_k = n$ est :

$$N = \frac{n!}{n_1! \dots n_k!}$$

2) En déduire la formule du multinôme : $\forall (a_1, \dots, a_k) \in \mathbb{R}^k$:

$$(a_1 + \dots + a_k)^n = \sum_{n_1 + \dots + n_k = n} \frac{n!}{n_1! \dots n_k!} a_1^{n_1} \dots a_k^{n_k}.$$

Indication : Considérer le premier membre comme un produit de n facteurs et examiner la forme générale d'un terme du développement de ce produit.

3) Soient $p_1, \dots, p_k$ des réels positifs tels que $p_1 + \dots + p_k = 1$.

a) Montrer que l'on définit bien une loi de probabilité sur $\mathbf{N}^k$ en posant :

$$P\left(\{(n_1, \dots, n_k)\}\right) = \begin{cases} 0 & \text{si } n_1 + \dots + n_k \neq n, \\ \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k} & \text{si } n_1 + \dots + n_k = n. \end{cases}$$

On dit que P est la loi multinomiale de paramètres $(n, p_1, \dots, p_k)$.

b) Soit $X = (X_1, \dots, X_k)$ une variable aléatoire à valeurs dans $\mathbf{N}^k$ suivant la loi multinomiale P définie ci-dessus. Montrer que X_i suit la loi binomiale de paramètres n, p_i.

c) Montrer que $X_1 + X_2$ suit la loi binomiale de paramètres $n, (p_1 + p_2)$. Généraliser au cas de : $Y = \sum_{i \in I} X_i$ où I est une partie de $\{1, \dots, k\}$.

d) On effectue une série de n épreuves répétées dans des conditions identiques avec pour chacune d'elles k résultats possibles de probabilités respectives $p_1, \dots, p_k$. On note X_l le nombre total de résultats du type $l, (1 \leq l \leq k)$. Quelle est la loi de $(X_1, \dots, X_k)$?

4) On lance $6n$ fois un dé, quelle est la probabilité d'obtenir exactement n fois chacune des faces ? Majorer cette probabilité à l'aide de la formule de Stirling :

$$\forall n \geq 1, \qquad n! = \sqrt{2\pi n}\, n^n e^{-n} e^{\theta_n}, \qquad \frac{1}{12n+1} < \theta_n < \frac{1}{12n}.$$

Application numérique : $n = 100$.

Ex 4.12. *Sondage téléphonique.*
Un enquêteur d'un institut de sondage dispose d'une liste de n personnes à interroger par téléphone. On suppose qu'à chaque appel d'un correspondant la probabilité de le contacter est p. L'enquêteur compose successivement les n numéros (première vague d'appels). On note X_1 la variable aléatoire nombre de correspondants contactés au cours de cette première vague. Lors de la deuxième vague d'appels l'enquêteur compose les $n - X_1$ numéros non obtenus et parvient à contacter X_2 correspondants...On demande de trouver les lois des variables aléatoires :

a) $X_l \quad (l \in \mathbf{N}^*)$;

b) $S_k = X_1 + \cdots + X_k \quad (k \geq 2)$;

c) V nombre de vagues nécessaires à l'obtention de tous les correspondants ;

d) Y nombre d'appels nécessaires à l'obtention d'un correspondant donné ;

e) T nombre total d'appels nécessaires à l'obtention de tous les correspondants.

Indications : On fera une hypothèse raisonnable d'indépendance entre les correspondants. Pour $k \geq 2$, on notera $E_{i,l}$ l'événement *le* i^e *correspondant est contacté lors de la* l^e *vague d'appels* $(1 \leq i \leq n, 1 \leq l \leq k-1)$ et par $F_{i,k}$ l'événement *le* i^e *correspondant n'est pas contacté lors des* $k-1$ *premières vagues d'appels.* Pour k fixé, on peut alors considérer une série de n épreuves répétées indépendantes de résultats possibles $(E_1, \ldots, E_{k-1}, F_k)$ et utiliser l'exercice 4.11 pour trouver la loi de $(X_1, \ldots, X_{k-1}, R_k)$ avec $R_k = n - S_{k-1}$.

Chapitre 5

Moments des variables aléatoires discrètes

Les moments d'une variable aléatoire sont certaines quantités numériques associées à sa loi et qui apportent une information sur cette loi. Leur étude nous permettra d'obtenir l'inégalité de Tchebycheff et la loi faible des grands nombres.

5.1 Espérance

Définition 5.1 *Soit X une variable aléatoire discrète vérifiant :*

$$\sum_{x_k \in X(\Omega)} |x_k| P(X = x_k) < +\infty. \tag{5.1}$$

On appelle espérance mathématique de X le réel $\mathbb{E}\, X$ défini par :

$$\mathbb{E}\, X = \sum_{x_k \in X(\Omega)} x_k P(X = x_k). \tag{5.2}$$

L'espérance de X apparaît ainsi comme le barycentre des valeurs possibles de X pondérées par leurs probabilités de réalisation. Notons que l'hypothèse (5.1) garantit l'existence de $\mathbb{E}\, X$ (convergence absolue de la série (5.2)). Remarquons aussi que l'espérance ne dépend que de la loi de X. Voici quelques calculs d'espérances de lois classiques.

Exemple 5.1 (v.a. constante)
Si X est constante ($\exists c \in \mathbb{R}$, $\forall \omega \in \Omega$, $X(\omega) = c$), son espérance est $\mathbb{E}\, X = c$.
Preuve : C'est une conséquence immédiate de la formule (5.2). ■

Exemple 5.2 (Loi de Bernoulli)
Si X suit la loi de Bernoulli de paramètre p, $\mathbb{E}\, X = p$.

En effet :

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

On a vu qu'une variable de Bernoulli de paramètre p peut toujours être considérée comme l'indicatrice d'un événement A de probabilité p. On a donc :

$$\text{Pour tout événement } A, \qquad \mathbb{E}\, \mathbf{1}_A = P(A). \tag{5.3}$$

■

Exemple 5.3 (Loi uniforme)
Si X suit la loi uniforme sur l'ensemble fini $\{x_1, \dots, x_n\}$, $\mathbb{E}\, X$ est égale à la moyenne arithmétique des x_i.
En effet : $X(\Omega) = \{x_1, \dots, x_n\}$ et $P(X = x_i) = 1/n$ pour tout i compris entre 1 et n. D'où :

$$\mathbb{E}\, X = \sum_{i=1}^{n} x_i P(X = x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

■

Le calcul de l'espérance des lois binomiale et hypergéométrique est moins immédiat et sera vu ultérieurement. Dans les trois exemples précédents, l'existence de $\mathbb{E}\, X$ ne pose aucun problème puisque $X(\Omega)$ étant fini, le second membre de (5.2) est la somme d'un nombre fini de termes.

Exemple 5.4 (Loi géométrique)
Si X suit la loi géométrique de paramètre $p > 0$, $\mathbb{E}\, X = \dfrac{1}{p}$.
On a ici $X(\Omega) = \mathbb{N}^*$, $P(X = k) = q^{k-1}p$ où $q = 1 - p \in]0, 1[$. La série à termes *positifs* :

$$\sum_{k=1}^{+\infty} kP(X = k) = \sum_{k=1}^{+\infty} kq^{k-1}p$$

est clairement convergente, ce qui justifie l'existence de $\mathbb{E}\, X$. On a alors sous réserve de justification :

$$\begin{aligned}
\mathbb{E}\, X = \sum_{k=1}^{+\infty} kq^{k-1}p &= p \sum_{k=1}^{+\infty} \left[\frac{d}{dx}(x^k) \right]_{x=q} && (5.4)\\
&= p \left[\frac{d}{dx} \left(\sum_{k=1}^{+\infty} x^k \right) \right]_{x=q} && (5.5)\\
&= p \frac{1}{(1-q)^2} = \frac{1}{p}. && (5.6)
\end{aligned}$$

Justification : La série entière de terme général x^k a pour rayon de convergence 1. On sait (cf. cours d'analyse) que sa somme est alors dérivable terme à terme en tout point de $]-1,+1[$. Ceci légitime le passage de (5.4) à (5.5). On remarque alors que :

$$\frac{d}{dx}\left(\sum_{k=1}^{+\infty} x^k\right) = \frac{d}{dx}\left(\sum_{k=0}^{+\infty} x^k\right) = \frac{d}{dx}\frac{1}{1-x} = \frac{1}{(1-x)^2},$$

ce qui justifie (5.6). ∎

Exemple 5.5 (Loi de Poisson)
Si X suit la loi de Poisson de paramètre λ, $\mathbb{E}\,X = \lambda$.

En effet on a ici : $X(\Omega) = \mathbb{N}$, $P(x = k) = \mathrm{e}^{-\lambda}\lambda^k/k!$. La série de terme général *positif* $u_k = kP(X = k)$ est convergente par application du critère de d'Alembert car $u_{k+1}/u_k = \lambda/k$ est inférieur à $1 - \varepsilon$ pour $k \geq k_0$. Donc $\mathbb{E}\,X$ existe. La somme de cette série se calcule ainsi :

$$\mathbb{E}\,X = \sum_{k=0}^{+\infty} kP(X = k) = \sum_{k=1}^{+\infty} \frac{\mathrm{e}^{-\lambda}\lambda^k}{(k-1)!} = \lambda\mathrm{e}^{-\lambda}\sum_{l=0}^{+\infty}\frac{\lambda^l}{l!} = \lambda.$$

∎

Remarques :

1. Si X et Y ont même loi, il est clair que $\mathbb{E}\,X = \mathbb{E}\,Y$. La réciproque est fausse. Voici un contre exemple. On choisit X suivant la loi uniforme sur $\{-1, +1\}$ et Y suivant la loi uniforme sur $\{-2, 0, +2\}$. Les lois P_X et P_Y sont différentes, mais $\mathbb{E}\,X = \mathbb{E}\,Y$:

$$\begin{aligned} \mathbb{E}\,X &= -1 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0, \\ \mathbb{E}\,Y &= -2 \times \frac{1}{3} + 0 \times \frac{1}{3} + 2 \times \frac{1}{3} = 0. \end{aligned}$$

2. Il y a des lois sans espérance. Par exemple celle définie sur $\mathbb{N}^*$ par :

$$P(X = k) = \frac{c}{k^2}, \quad \text{avec} \quad c\sum_{k=1}^{+\infty}\frac{1}{k^2} = 1 \quad \text{soit} \quad c = \frac{6}{\pi^2}.$$

   Il est clair que pour cette loi, la série (5.1) diverge.

Proposition 5.2 (Linéarité de l'espérance)
Pour toutes variables aléatoires X et Y ayant une espérance et toute constante a, les v.a. $X + Y$ et aX ont une espérance et :

$$\mathbb{E}(X + Y) = \mathbb{E}\,X + \mathbb{E}\,Y, \tag{5.7}$$

$$\mathbb{E}(aX) = a\,\mathbb{E}\,X. \tag{5.8}$$

Preuve : La vérification de (5.8) est directe. Pour prouver (5.7), on pose $Z = X + Y$ et on étudie dans un premier temps le cas particulier où $X(\Omega)$ et $Y(\Omega)$ sont finis. Il en est alors de même pour $Z(\Omega)$, ce qui règle la question de l'existence de $\mathbb{E}\, Z$. La loi de Z est donnée par :

$$\forall z_k \in Z(\Omega), \quad P(Z = z_k) = \sum_{x_i + y_j = z_k} P(X = x_i, Y = y_j), \tag{5.9}$$

la sommation étant étendue à tous les couples (x_i, y_j) de $X(\Omega) \times Y(\Omega)$ vérifiant la condition $x_i + y_j = z_k$. Pour alléger les écritures, dans les formules suivantes, nous abrégeons l'indexation par $x_i \in X(\Omega)$ en indexation par i et de même pour Y, j et Z, k.

$$\begin{aligned}
\mathbb{E}\, Z &= \sum_k z_k P(Z = z_k) && (5.10)\\
&= \sum_k z_k \sum_{x_i + y_j = z_k} P(X = x_i, Y = y_j) \\
&= \sum_k \sum_{x_i + y_j = z_k} (x_i + y_j) P(X = x_i, Y = y_j) && (5.11)\\
&= \sum_{i,j} (x_i + y_j) P(X = x_i, Y = y_j) && (5.12)\\
&= \sum_i \sum_j x_i P(X = x_i, Y = y_j) + \sum_i \sum_j y_j P(X = x_i, Y = y_j) && (5.13)\\
&= \sum_i x_i \Big[\sum_j P(X = x_i, Y = y_j) \Big] \\
&\qquad + \sum_j y_j \Big[\sum_i P(X = x_i, Y = y_j) \Big] && (5.14)\\
&= \sum_i x_i P(X = x_i) + \sum_j y_j P(Y = y_j) && (5.15)\\
&= \mathbb{E}\, X + \mathbb{E}\, Y. && (5.16)
\end{aligned}$$

Le passage de (5.11) à (5.12) se justifie en remarquant que les « droites » d'équation $x + y = z_k$ réalisent une partition du « rectangle » $X(\Omega) \times Y(\Omega)$.

Dans le cas général, $X(\Omega) \times Y(\Omega)$ pouvant être infini, il faut justifier l'existence de $\mathbb{E}\, Z$. Pour cela, on remplace dans les égalités (5.10) à (5.16) z_k par $|z_k|$ puis $x_i + y_j$ par $|x_i + y_j|$ (les probabilités restant inchangées). On passe de (5.12) à (5.13) en utilisant l'inégalité triangulaire. Ceci nous donne finalement :

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

d'après l'hypothèse d'existence de $\mathbb{E}\, X$ et $\mathbb{E}\, Y$. L'existence de $\mathbb{E}\, Z$ étant acquise, les égalités (5.10) à (5.16) sans valeurs absolues restent vraies par les propriétés des séries doubles absolument convergentes (sommation par paquets). ■

Voici deux applications de (5.7) à un calcul d'espérance.

Exemple 5.6 (Espérance d'une loi binomiale)
Si X suit la loi binomiale $\mathcal{B}(n,p)$, $\mathbb{E}\, X = np$.
On sait en effet que X a *même loi* que la somme

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

où les X_i sont des variables de Bernoulli indépendantes de paramètre p. Comme l'espérance d'une v. a. ne dépend que de sa loi, $\mathbb{E}\, X = \mathbb{E}\, S$ et par linéarité de l'espérance :

$$\mathbb{E}\, X = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}\, X_i = np,$$

en utilisant l'exemple 5.2. ■

Exemple 5.7 (Espérance d'une loi hypergéométrique)
Si X suit la loi hypergéométrique $\mathcal{H}(N,M,n)$, $\mathbb{E}\, X = n\dfrac{M}{N}$.
L'espérance d'une v. a. ne dépendant que de la loi, on ne perd pas de généralité en supposant que X est le nombre d'objets défectueux observé dans un échantillon de taille n prélevé sans remise dans une population totale de N objets dont M sont défectueux. En numérotant de 1 à M tous les objets défectueux, on a alors

$$X = X_1 + \cdots + X_M,$$

où

$$X_i = \begin{cases} 1 & \text{si le } i\text{-ème objet défectueux est prélevé,} \\ 0 & \text{sinon.} \end{cases}$$

Chaque X_i est une variable de Bernoulli de paramètre $p_i = P(X_i = 1)$. Le calcul de p_i se ramène au dénombrement de tous les échantillons possibles contenant le i-ème objet défectueux. Un tel échantillon est constitué en prenant le i-ème défectueux et en complétant par $n-1$ objets (défectueux ou non) choisis dans le reste de la population. D'où :

$$p_i = P(X_i = 1) = \frac{1 \times C_{N-1}^{n-1}}{C_N^n} = \frac{n}{N}.$$

Ainsi les X_i ont même loi et même espérance $\mathbb{E}\, X_i = n/N$. Par linéarité on en déduit :

$$\mathbb{E}\, X = \sum_{i=1}^{M} \mathbb{E}\, X_i = M \frac{n}{N}.$$

Remarquons que le résultat obtenu est le même que pour un prélèvement de n objets *avec remise* : dans ce cas, le nombre Y d'objets défectueux dans l'échantillon suit la loi binomiale $\mathcal{B}(n, M/N)$ et $\mathbb{E}\, Y = nM/N$. ■

Nous examinons maintenant le problème du calcul de l'espérance de $f(X)$ où X est une variable aléatoire discrète et f une fonction (déterministe) définie au moins sur $X(\Omega)$. Commençons par un exemple élémentaire. On suppose que X suit la loi uniforme sur $\{-2, -1, 0, 1, 2\}$ et que f est la fonction polynôme $f(x) = x^4 - x^2 + 2$. Posons $Y = f(X)$ et déterminons l'ensemble $Y(\Omega)$. On a $f(-1) = f(0) = f(+1) = 2$ et $f(-2) = f(+2) = 14$. Donc $Y(\Omega) = \{2, 14\}$ et en utilisant la définition de l'espérance :

$$\mathbb{E}\, Y = 2P(Y = 2) + 14P(Y = 14).$$

Le premier terme de cette somme peut se réécrire :

$$\begin{aligned} 2P(Y = 2) &= 2\big[P(X = -1) + P(X = 0) + P(X = 1)\big] \\ &= f(-1)P(X = -1) + f(0)P(X = 0) + f(1)P(X = 1). \end{aligned}$$

On obtient de même :

$$14P(Y = 14) = f(-2)P(X = -2) + f(2)P(X = 2).$$

En reportant ces résultats dans $\mathbb{E}\, Y$, on voit que :

$$\mathbb{E}\, f(X) = \sum_{k=-2}^{+2} f(k)P(X = k) = \sum_{x \in X(\Omega)} f(x)P(X = x).$$

L'intérêt de cette formule est qu'elle permet un calcul direct de $\mathbb{E}\, f(X)$ à partir de la loi de X, sans passer par la loi de Y.

Proposition 5.3 (Espérance d'une fonction d'une v.a.)
Soient X une v.a. discrète et f une fonction numérique dont l'ensemble de définition contient $X(\Omega)$. Si $\mathbb{E}\, f(X)$ existe,

$$\mathbb{E}\, f(X) = \sum_{x \in X(\Omega)} f(x)P(X = x). \tag{5.17}$$

Preuve : Posons $Y = f(X)$. L'ensemble $X(\Omega) = \{x_0, x_1, \dots\}$ est au plus dénombrable. L'application f n'étant pas supposée injective, il peut y avoir des répétitions dans la suite des $f(x_k)$. L'ensemble $Y(\Omega)$ qui peut s'écrire en effaçant toutes les répétitions de cette suite est lui même au plus dénombrable. Par hypothèse l'espérance de Y existe et est donc définie par la série absolument convergente (ou somme finie si $Y(\Omega)$ est fini) :

$$\mathbb{E}\, f(X) = \mathbb{E}\, Y = \sum_{y \in Y(\Omega)} yP(Y = y). \tag{5.18}$$

Pour chaque réel y de $Y(\Omega)$, notons B_y l'ensemble de ses antécédents :

$$B_y = \{x \in X(\Omega),\ f(x) = y\}, \qquad y \in Y(\Omega).$$

Ce sous-ensemble de $X(\Omega)$ contient au moins un élément et peut être fini ou infini dénombrable. On peut alors décomposer l'événement $\{Y = y\}$ en réunion disjointe (d'une famille au plus dénombrable) :

$$\{Y = y\} = \bigcup_{x \in B_y} \{X = x\}.$$

Le terme général de la série (5.18) peut donc s'écrire :

$$\begin{aligned} yP(Y = y) &= y \sum_{x \in B_y} P(X = x) && (5.19) \\ &= \sum_{x \in B_y} f(x)P(X = x) && (5.20) \end{aligned}$$

La série (5.20) est *absolument convergente* car f étant constante sur B_y, tous ses termes sont de même signe et la série à termes positifs (5.19) est convergente. Comme les B_y forment une partition de $X(\Omega)$, on a en utilisant la propriété de sommation par paquets des séries à termes *positifs* :

$$\begin{aligned} \sum_{x \in X(\Omega)} |f(x)|P(X = x) &= \sum_{y \in Y(\Omega)} \sum_{x \in B_y} |f(x)|P(X = x) \\ &= \sum_{y \in Y(\Omega)} |y|P(Y = y) < +\infty, \end{aligned}$$

d'après l'hypothèse d'existence de $\mathbb{E}\, Y$. Ceci légitime la même sommation *sans les valeurs absolues*, laquelle prouve (5.17). ■

Remarque : En appliquant la proposition 5.3 avec $f(x) = |x|$, on obtient :

$$\mathbb{E}\, |X| = \sum_{x_k \in X(\Omega)} |x_k|P(X = x_k).$$

Ceci est exactement la série (5.1) dont la convergence garantit l'existence de $\mathbb{E}\, X$. On en déduit :

Proposition 5.4 (Espérance et valeur absolue)
Si $\mathbb{E}\, X$ *existe,* $|\mathbb{E}\, X| \leq \mathbb{E}\,|X| < +\infty$.

Proposition 5.5 (Positivité de l'espérance)

(*i*) *Si* X *a une espérance et si* $X \geq 0$*, alors* $\mathbb{E}\, X \geq 0$.

(*ii*) *Si* X *et* Y *ont une espérance et si* $X \leq Y$*, alors* $\mathbb{E}\, X \leq \mathbb{E}\, Y$.

Preuve : Les notations $X \geq 0$ et $X \leq Y$ signifient respectivement : $\forall \omega \in \Omega$, $X(\omega) \geq 0$ (resp $X(\omega) \leq Y(\omega)$). Supposons d'abord que X est positive. Alors tous les réels x_k de $X(\Omega)$ sont positifs et si l'espérance de X existe, la série de terme général $x_k P(X = x_k)$ est une série convergente à termes positifs. Sa somme $\mathbb{E}\, X$ est donc un réel positif, ce qui prouve (i). Pour prouver (ii), on remarque que l'hypothèse $X \leq Y$ équivaut à la positivité de la v.a. $Z = Y - X$. En appliquant (i) à Z on en déduit : $\mathbb{E}(Y - X) \geq 0$, puis par linéarité $\mathbb{E}\, Y - \mathbb{E}\, X \geq 0$ d'où $\mathbb{E}\, X \leq \mathbb{E}\, Y$. ■

Pour terminer cette section, nous donnons une condition suffisante d'existence de l'espérance. La démonstration pourra être passée en première lecture.

Proposition 5.6 *Si* Z *est une v.a. discrète ayant une espérance et si pour tout* $\omega \in \Omega$, $|X(\omega)| \leq |Z(\omega)|$*, alors* X *possède une espérance.*

Preuve : Cet énoncé fait penser à la propriété suivante des séries : si pour tout $k \in \mathbb{N}$, $|u_k| \leq |v_k|$ et si v_k est le terme général d'une série absolument convergente, alors la série de terme général u_k est aussi absolument convergente. Malheureusement, la situation est ici plus compliquée pour la raison suivante. Il peut y avoir plusieurs valeurs différentes[1] z_l de $Z(\omega)$ lorsque ω décrit l'événement fixé $\{X = x_k\}$. Tout ce que l'on peut en dire est qu'elles vérifient toutes $|z_l| \geq |x_k|$. S'il n'y avait qu'une seule valeur $z_l = z_k$ commune à tous les $Z(\omega)$ pour $\omega \in \{X = x_k\}$, on serait ramené à l'énoncé sur les séries en posant $u_k = x_k P(X = x_k)$ et $v_k = z_k P(Z = z_k)$.

Le point clé de la démonstration est la remarque suivante :

$$\begin{aligned} P(X = x_k) &= \sum_{z_l \in Z(\Omega)} P(X = x_k, Z = z_l) \\ &= \sum_{\substack{z_l \in Z(\Omega) \\ |z_l| \geq |x_k|}} P(X = x_k, Z = z_l). \end{aligned} \tag{5.21}$$

En effet si $\omega \in \{X = x_k, Z = z_l\}$, on a simultanément $X(\omega) = x_k$ et $Z(\omega) = z_l$ et par hypothèse $|X(\omega)| \leq |Z(\omega)|$. Il en résulte que pour tous les z_l vérifiant $|z_l| < |x_k|$, l'événement $\{X = x_k, Z = z_l\}$ est impossible et par conséquent $P(X = x_k, Z = z_l) = 0$.

[1] Voire même une infinité.

Par hypothèse, la série de terme général positif $|z_l|P(Z = z_l)$ a une somme finie et nous voulons montrer qu'il en est de même pour la série de terme général $|x_k|P(X = x_k)$. Le calcul qui suit est légitimé par les propriétés des séries à termes *positifs* (dont les sommes ne sont pas nécessairement finies a priori) :

$$\begin{aligned}
\sum_{x_k \in X(\Omega)} |x_k| P(X = x_k) &= \sum_{x_k \in X(\Omega)} |x_k| \sum_{\substack{z_l \in Z(\Omega) \\ |z_l| \geq |x_k|}} P(X = x_k, Z = z_l) \\
&= \sum_{x_k \in X(\Omega)} \sum_{\substack{z_l \in Z(\Omega) \\ |z_l| \geq |x_k|}} |x_k| P(X = x_k, Z = z_l) \\
&\leq \sum_{x_k \in X(\Omega)} \sum_{\substack{z_l \in Z(\Omega) \\ |z_l| \geq |x_k|}} |z_l| P(X = x_k, Z = z_l) \\
&\leq \sum_{x_k \in X(\Omega)} \sum_{z_l \in Z(\Omega)} |z_l| P(X = x_k, Z = z_l) \\
&= \sum_{z_l \in Z(\Omega)} |z_l| \sum_{x_k \in X(\Omega)} P(X = x_k, Z = z_l) \\
&= \sum_{z_l \in Z(\Omega)} |z_l| P(Z = z_l) < +\infty,
\end{aligned}$$

d'après l'hypothèse d'existence de $\mathbb{E}\, Z$. ■

5.2 Moments d'ordre r

Définition 5.7 *Soit $r \in \mathbb{N}^*$. On appelle moment d'ordre r de la v. a. X la quantité*

$$\mathbb{E}(X^r) = \sum_{x_k \in X(\Omega)} x_k^r P(X = x_k), \tag{5.22}$$

lorsqu'elle est définie, c'est-à-dire si :

$$\mathbb{E}(|X|^r) = \sum_{x_k \in X(\Omega)} |x_k|^r P(X = x_k) < +\infty. \tag{5.23}$$

Lorsqu'il existe, ce moment d'ordre r sera noté $\mathbb{E}\, X^r$ (ne pas confondre avec $(\mathbb{E}\, X)^r$).

Lorsque $X(\Omega)$ est borné, (5.23) est vérifiée pour toute valeur de r (exercice) et X possède des moments de tout ordre. Lorsque $X(\Omega)$ n'est pas borné, la condition est d'autant plus contraignante que r est plus grand.

Proposition 5.8 *Si X possède un moment d'ordre r, elle possède aussi des moments de tout ordre $n \leq r$.*

Preuve : Par hypothèse (5.23) est vérifiée avec l'exposant r et il s'agit de prouver la même inégalité avec n à la place de r. La série de terme général *positif* $|x_k|^n P(X = x_k)$ a une somme finie ou infinie et on ne change pas cette somme en regroupant ses termes en deux paquets comme suit :

$$\mathbb{E}(|X|^n) = \sum_{x_k \colon |x_k| \leq 1} |x_k|^n P(X = x_k) + \sum_{x_k \colon |x_k| > 1} |x_k|^n P(X = x_k).$$

Dans la première somme, on majore $|x_k|^n$ par 1. Dans la deuxième on le majore par $|x_k|^r$ puisque $n \leq r$ et si $x > 1$, $x^n \leq x^r$.

$$\begin{aligned}
\mathbb{E}(|X|^n) &\leq \sum_{x_k \colon |x_k| \leq 1} P(X = x_k) + \sum_{x_k \colon |x_k| > 1} |x_k|^r P(X = x_k) \\
&\leq P(|X| \leq 1) + \sum_{x_k \in X(\Omega)} |x_k|^r P(X = x_k) \\
&= P(|X| \leq 1) + \mathbb{E}(|X|^r) < +\infty.
\end{aligned}$$

Ainsi $\mathbb{E}(|X|^n)$ est fini et X possède un moment d'ordre n. ■

Pour toute variable aléatoire X, $P(|X| \geq t)$ tend vers 0 quand t tend vers $+\infty$, c'est une conséquence du théorème 3.4. L'existence de moments nous donne des renseignements sur la vitesse de cette convergence[2]. C'est le sens du théorème suivant dû à Markov.

Théorème 5.9 (Inégalité de Markov)
Si X est une variable aléatoire positive ayant une espérance, on a :

$$\forall t > 0, \qquad P(X \geq t) \leq \frac{\mathbb{E}\, X}{t}. \tag{5.24}$$

Corollaire 5.10 *Si X est une variable aléatoire ayant un moment d'ordre r :*

$$\forall t > 0, \qquad P(|X| \geq t) \leq \frac{\mathbb{E}\big(|X|^r\big)}{t^r}. \tag{5.25}$$

Preuve : Dans la série à termes *positifs* définissant $\mathbb{E}\, X$, on peut effectuer un regroupement en deux paquets en classant les x_k comparativement à t :

$$\begin{aligned}
\mathbb{E}\, X &= \sum_{x_k \in X(\Omega)} x_k P(X = x_k) \\
&= \sum_{x_k \colon x_k \geq t} x_k P(X = x_k) + \sum_{x_k \colon x_k < t} x_k P(X = x_k).
\end{aligned}$$

[2]Inversement, on peut construire une variable aléatoire n'ayant pas d'espérance pour laquelle cette convergence est aussi lente que l'on veut.

Comme X est positive, tous les termes du deuxième paquet sont positifs ou nuls, par conséquent on obtient une minoration de $\mathbb{E}\, X$ en les effaçant :

$$\begin{aligned} \mathbb{E}\, X &\geq \sum_{x_k\colon x_k\geq t} x_k P(X = x_k) \\ &\geq \sum_{x_k\colon x_k\geq t} t P(X = x_k) = tP(X \geq t). \end{aligned}$$

Ainsi $\mathbb{E}\, X \geq tP(X \geq t)$, ce qui prouve (5.24). Pour le corollaire, X n'étant plus supposée positive, on applique l'inégalité de Markov à la v.a. positive $Y = |X|^r$ en notant que pour tout $t > 0$, $\{|X| \geq t\} = \{|X|^r \geq t^r\} = \{Y \geq t^r\}$. ■

5.3 Variance

Considérons un amphi de 100 étudiants venant de subir[3] un D.S. où la *moyenne* de l'amphi a été 10. On choisit un étudiant au hasard et on désigne par X sa note ; $X(\Omega)$ est l'ensemble des notes (toutes différentes) attribuées à ce D.S.. Comme tous les choix d'étudiants sont équiprobables, pour tout $x_k \in X(\Omega)$,

$$P(X = x_k) = \frac{\text{Nombre d'étudiants ayant obtenu la note } x_k}{100}.$$

D'où : $\mathbb{E}\, X = \displaystyle\sum_{x_k\in X(\Omega)} x_k P(X = x_k) = 10.$

Cette moyenne ne nous apporte pas une information très précise sur l'amphi. Elle peut être obtenue avec 100 étudiants ayant 10 aussi bien qu'avec 50 ayant 0 et 50 ayant 20 ou un grand nombre de situations intermédiaires entre ces deux configurations extrêmes. Il est important pour compléter l'information apportée par la moyenne de disposer d'une quantité permettant de mesurer la *dispersion* autour de cette moyenne. Le deuxième moment centré de X, à savoir $\mathbb{E}(X - \mathbb{E}\, X)^2$ joue ce rôle[4].

Définition 5.11 *Soit X une v.a. ayant un moment d'ordre 2. On appelle respectivement variance de X et écart type de X les quantités :*

$$\operatorname{Var} X = \mathbb{E}(X - \mathbb{E}\, X)^2, \qquad \sigma(X) = (\operatorname{Var} X)^{1/2}.$$

Remarquons que l'existence de $\mathbb{E}\, X^2$ entraîne celle de $\mathbb{E}\, X$ puis, en développant $(X - \mathbb{E}\, X)^2$ et en utilisant la linéarité, celle de $\operatorname{Var} X$. Lorsque X représente une grandeur physique, X, $\mathbb{E}\, X$ et $\sigma(X)$ ont la même unité, mais pas $\operatorname{Var} X$.

[3] C'est généralement le terme adéquat...

[4] On pourrait utiliser d'autres quantités comme $\mathbb{E}\,|X - \mathbb{E}\, X|$, mais leurs propriétés sont moins intéressantes.

Proposition 5.12 (Translation et changement d'échelle)
Si X a un moment d'ordre 2,

$$\forall a \in \mathbb{R},\ \forall b \in \mathbb{R}, \quad \operatorname{Var}(aX+b) = a^2 \operatorname{Var} X.$$

Preuve : On a en utilisant la définition 5.11, la linéarité de l'espérance et le fait que l'espérance de la constante b est b :

$$\begin{aligned}\operatorname{Var}(aX+b) &= \mathbb{E}[aX+b-\mathbb{E}(aX+b)]^2 = \mathbb{E}[aX+b-a\,\mathbb{E}\,X-b]^2\\ &= \mathbb{E}[a(X-\mathbb{E}\,X)]^2 = \mathbb{E}[a^2(X-\mathbb{E}\,X)^2]\\ &= a^2\,\mathbb{E}[(X-\mathbb{E}\,X)^2] = a^2 \operatorname{Var} X.\end{aligned}$$

■

En particulier si $a = 0$, ceci montre que la variance d'une constante est nulle. La réciproque est *presque* vraie :

Proposition 5.13

$$\operatorname{Var} X = 0 \Leftrightarrow X = \mathbb{E}\,X \text{ (p.s.)} \Leftrightarrow X \text{ est presque sûrement constante.}$$

Preuve : Notons $\mu = \mathbb{E}\,X$. L'égalité presque sûre (p.s.) signifie ici : $P(X = \mu) = 1$, de même X constante presque sûrement signifie : il existe une constante c telle que $P(X = c) = 1$. Rappelons que nous avons accepté de conserver dans l'ensemble $X(\Omega)$ des valeurs prises par X, des réels x_k tels que $P(X = x_k) = 0$. Ainsi, « X constante presque sûrement »ne signifie pas forcément $X(\Omega) = \{c\}$.

Les implications de droite à gauche sont évidentes. Dans l'autre sens, supposons que X a un moment d'ordre 2. On a alors :

$$\operatorname{Var} X = \mathbb{E}(X-\mu)^2 = \sum_{x_k \in X(\Omega)} (x_k-\mu)^2 P(X = x_k).$$

Cette somme ou série à termes positifs ne peut être nulle que si tous ses termes sont nuls, donc :

$$\forall x_k \in X(\Omega), \qquad (x_k - \mu) = 0 \quad \text{ou} \quad P(X = x_k) = 0.$$

Autrement dit, pour tous les $x_k \neq \mu$, $P(X = x_k) = 0$. On en déduit $P(X \neq \mu) = 0$ et $P(X = \mu) = 1$. La variable aléatoire X est donc égale presque sûrement à la constante $\mu = \mathbb{E}\,X$. ■

Pour le calcul de la variance, il est souvent plus commode d'utiliser la formule suivante que la définition 5.11.

Proposition 5.14 (Formule de Koenig)

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

Preuve : Rappelons que nous notons $\mathbb{E}\, X^2$ pour $\mathbb{E}(X^2)$ et que le second membre de la formule ci-dessus n'est donc généralement pas nul. On pose à nouveau $\mu = \mathbb{E}\, X$.

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

en utilisant la linéarité de l'espérance et l'espérance d'une constante. ■

Voyons maintenant quelques calculs de variance de lois classiques.

Exemple 5.8 (Variance d'une loi de Bernoulli)
La variance d'une loi de Bernoulli de paramètre p est $p(1-p) = pq$.
En effet si X suit la loi $\mathcal{B}(p)$, on a $\mathbb{E}\, X = p$ et $\mathbb{E}\, X^2 = 0^2 \times q + 1^2 \times p = p$ (avec $q = 1 - p$). D'où :

$$\operatorname{Var} X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = p - p^2 = p(1-p).$$

■

Exemple 5.9 (Variance de la loi uniforme sur $\{1, \dots, n\}$)
Si X suit la loi uniforme sur $\{1, \dots, n\}$, $\operatorname{Var} X = \dfrac{n^2-1}{12}$.
Dans ce cas, $X(\Omega) = \{1, \dots, n\}$ et pour tout entier k compris entre 1 et n, $P(x = k) = 1/n$. On calcule alors $\mathbb{E}\, X$ et $\mathbb{E}\, X^2$:

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

$$\begin{aligned}\mathbb{E}\, X^2 &= \sum_{k=1}^{n} k^2 P(X = k) = \frac{1}{n}\sum_{k=1}^{n} k^2 \\ &= \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.\end{aligned}$$

En appliquant la formule de Koenig, on obtient :

$$\operatorname{Var} X = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n+1)(n-1)}{12}.$$

■

Exemple 5.10 (Variance d'une loi binomiale)
Si X suit la loi $\mathcal{B}(n,p)$, $\operatorname{Var} X = np(1-p) = npq$.

On peut démontrer ce résultat par la formule de Koenig (exercice) ou le déduire d'une formule générale sur la variance d'une somme de v.a. indépendantes (cf. proposition 5.23). Nous en donnons une troisième preuve instructive en elle-même. On part du fait que X a même loi et donc même variance que :

$$S_n = \sum_{i=1}^{n} X_i,$$

où les X_i sont des v.a. de Bernoulli indépendantes de même paramètre p. Nous posons $Y_i = X_i - \mathbb{E}\, X_i = X_i - p$. On a alors :

$$S_n - \mathbb{E}\, S_n = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mathbb{E}\, X_i = \sum_{i=1}^{n} (X_i - \mathbb{E}\, X_i) = \sum_{i=1}^{n} Y_i.$$

D'où :

$$(S_n - \mathbb{E}\, S_n)^2 = \sum_{i=1}^{n} Y_i^2 + \sum_{\substack{1 \leq i,j \leq n \\ i \neq j}} Y_i Y_j.$$

Par linéarité de l'espérance, on en déduit :

$$\operatorname{Var} S_n = \mathbb{E}(S_n - \mathbb{E}\, S_n)^2 = \sum_{i=1}^{n} \mathbb{E}\, Y_i^2 + \sum_{\substack{1 \leq i,j \leq n \\ i \neq j}} \mathbb{E}(Y_i Y_j). \tag{5.26}$$

D'après l'exemple 5.8, on sait que $\mathbb{E}\, Y_i^2 = \operatorname{Var} X_i = pq$. D'autre part,

$$\begin{aligned} \mathbb{E}(Y_i Y_j) &= \mathbb{E}\big[(X_i - p)(X_j - p)\big] \\ &= \mathbb{E}(X_i X_j) - p\,\mathbb{E}\, X_j - p\,\mathbb{E}\, X_i + p^2 \\ &= \mathbb{E}(X_i X_j) - p^2. \end{aligned}$$

Il reste donc à calculer les $\mathbb{E}(X_i X_j)$ pour $i \neq j$. Comme les X_k ne peuvent prendre que les valeurs 0 ou 1, il en est de même pour $Z_{i,j} = X_i X_j$. C'est donc une v.a. de Bernoulli de paramètre p' donné par :

$$p' = P(X_i X_j = 1) = P(X_i = 1, X_j = 1) = P(X_i = 1) P(X_j = 1) = p^2,$$

en utilisant l'indépendance de X_i et X_j pour $i \neq j$. On en déduit $\mathbb{E}(X_i X_j) = p' = p^2$ et $\mathbb{E}(Y_i Y_j) = 0$ pour $i \neq j$. En reportant ce résultat dans (5.26) il vient :

$$\operatorname{Var} X = \operatorname{Var} S_n = \sum_{i=1}^{n} \mathbb{E}\, Y_i^2 = npq.$$

■

Exemple 5.11 (Variance d'une loi de Poisson)
Si X suit la loi de Poisson $\mathcal{P}(\lambda)$, $\operatorname{Var} X = \lambda$.

Il est facile de vérifier (exercice) qu'une loi de Poisson possède des moments de tout ordre donc a fortiori une variance. On sait déjà que $\mathbb{E}\,X = \lambda$, il nous reste à calculer $\mathbb{E}\,X^2$:

$$\begin{aligned}
\mathbb{E}\,X^2 &= \sum_{k=0}^{+\infty} k^2 \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{+\infty} k \frac{\mathrm{e}^{-\lambda}\lambda^k}{(k-1)!} \\
&= \mathrm{e}^{-\lambda} \sum_{k=1}^{+\infty} \big((k-1)+1\big) \frac{\lambda^k}{(k-1)!} \\
&= \lambda^2 \mathrm{e}^{-\lambda} \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \mathrm{e}^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2 + \lambda.
\end{aligned}$$

D'où $\operatorname{Var} X = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda$. ■

Exemple 5.12 (Variance d'une loi géométrique)
Si X suit la loi géométrique de paramètre p, $\operatorname{Var} X = \dfrac{q}{p^2}$

La vérification est laissée en exercice.

Une des raisons de l'importance de la variance est le théorème suivant.

Théorème 5.15 (Inégalité de Tchebycheff) *Si $\operatorname{Var} X$ existe,*

$$\forall t > 0, \quad P(|X - \mathbb{E}\,X| \geq t) \leq \frac{\operatorname{Var} X}{t^2}. \tag{5.27}$$

Preuve : Il suffit d'appliquer le corollaire de l'inégalité de Markov (corollaire 5.10) avec $r = 2$ à la v.a. $Y = X - \mathbb{E}\,X$. ■

Remarque : Si on pose $t = u\sigma(X)$ dans (5.27), on obtient une nouvelle forme de l'inégalité de Tchebycheff :

$$\forall u > 0, \quad P\big(|X - \mathbb{E}\,X| \geq u\sigma(X)\big) \leq \frac{1}{u^2}. \tag{5.28}$$

Sous cette forme, le majorant obtenu est indépendant de la loi de X. Ceci permet de comprendre pourquoi $\sigma(X)$ s'appelle *écart type* ou *unité d'écart*. Pour toute loi de probabilité ayant un moment d'ordre 2, la probabilité d'observer une déviation par rapport à l'espérance d'au moins u unités d'écart est majorée par u^{-2}.

Exemple 5.13 *On jette* 3600 *fois un dé. Minorer la probabilité que le nombre d'apparitions du 1 soit compris strictement entre* 480 *et* 720*.*
Notons S le nombre d'apparitions du 1. Cette variable aléatoire suit la loi binomiale $\mathcal{B}(3600, 1/6)$. La valeur exacte de la probabilité qui nous intéresse est :

$$P(480 < S < 720) = \sum_{k=481}^{719} C_{3600}^{k} \Big(\frac{1}{6}\Big)^{k} \Big(\frac{5}{6}\Big)^{3600-k}.$$

Le calcul de la valeur exacte de cette somme nécessiterait l'écriture d'un programme et une certaine puissance de calcul informatique. L'inégalité de Tchebycheff est une alternative pratique à un calcul aussi déraisonnable. En effet on a :

$$\mathbb{E}\, S = 3600 \times \frac{1}{6} = 600, \qquad \operatorname{Var} S = 3600 \times \frac{1}{6} \times \frac{5}{6} = 500$$

et on remarque que $480 - 600 = -120$, $720 - 600 = 120$ d'où :

$$480 < S < 720 \Leftrightarrow -120 < S - 600 < +120 \Leftrightarrow |S - 600| < 120.$$

L'inégalité de Tchebycheff s'écrit ici :

$$\forall t > 0, \quad P(|S - 600| \geq t) \leq \frac{500}{t^2}.$$

En particulier pour $t = 120$:

$$P(|S - 600| \geq 120) \leq \frac{500}{120^2} = \frac{5}{144}.$$

D'où en passant à l'événement complémentaire :

$$P(480 < S < 720) = P(|S - 600| < 120) \geq 1 - \frac{5}{144} \geq 0.965.$$

On peut aussi utiliser l'inégalité de Tchebycheff avec un intervalle non centré sur l'espérance. Il suffit alors d'utiliser le plus grand intervalle centré qu'il contient. Ainsi pour minorer $P(550 < S < 700)$, il suffit de remarquer que

$$550 < S < 700 \Leftarrow 550 < S < 650,$$

d'où :

$$\{550 < S < 700\} \supset \{550 < S < 650\}$$

et d'appliquer l'inégalité de Tchebycheff avec $t = 50$. Bien sûr le résultat obtenu sera moins bon. ■

On peut aussi utiliser l'inégalité de Tchebycheff avec un intervalle unilatéral (i.e. du type $[a, +\infty[$ ou $]-\infty, a]$).

Exemple 5.14 *Avec les notations de l'exemple précédent, proposer une valeur de u telle que $P(S \geq u) \leq 0.05$.*
Comme la v.a. S est bornée par 3600, la réponse n'a d'intérêt que pour $u < 3600$. Une première idée est d'essayer l'inégalité de Markov qui s'écrit ici :

$$P(S \geq u) \leq \frac{\mathbb{E}\, S}{u} = \frac{600}{u}.$$

Par cette inégalité, il *suffirait* de prendre u tel que $600/u \leq 0.05$. Malheureusement, la plus petite solution de cette inéquation est $u_0 = 600/0.05 = 12000$ et le résultat est sans intérêt. Pour utiliser l'inégalité de Tchebycheff, on commence par remarquer que pour tout t *positif*, l'inégalité $S - \mathbb{E}\, S \geq t$ implique $|S - \mathbb{E}\, S| \geq t$, ce qui se traduit par l'inclusion d'événements :

$$\{S - \mathbb{E}\, S \geq t\} \subset \{|S - \mathbb{E}\, S| \geq t\}.$$

On en déduit :

$$P(S \geq t + \mathbb{E}\, S) = P(S - \mathbb{E}\, S \geq t) \leq P(|S - \mathbb{E}\, S| \geq t) \leq \frac{500}{t^2}.$$

Il *suffit* alors de choisir la plus petite valeur de t telle que $500/t^2 \leq 0.05$ soit $t_1 = 100$. La valeur correspondante pour u étant $u_1 = t_1 + \mathbb{E}\, S = 700$. ■

5.4 Covariance

Le but de cette section est le calcul de la variance d'une somme de variables aléatoires. L'inégalité de Tchebycheff montre l'importance d'un tel calcul en vue des théorèmes de convergence. Le calcul de la variance d'une loi binomiale (exemple 5.10) illustre le rôle clé joué par des quantités comme $\mathbb{E}(X_i X_j)$ et $\mathbb{E}\big[(X_i - \mathbb{E}\, X_i)(X_j - \mathbb{E}\, X_j)\big]$. Nous allons généraliser l'étude de ces quantités.

Si X et Y sont des variables aléatoires discrètes, leur produit XY est aussi une v.a. discrète. En utilisant l'inégalité $|XY| \leq X^2 + Y^2$ et les propositions 5.6 et 5.2, on voit qu'une condition suffisante pour l'existence de $\mathbb{E}(XY)$ est que X et Y possèdent des moments d'ordre deux. Lorsqu'elle existe, l'espérance de XY se calcule par la formule suivante (justification laissée en exercice) :

$$\mathbb{E}(XY) = \sum_{\substack{x_i \in X(\Omega)\\ y_j \in Y(\Omega)}} x_i y_j P(X = x_i, Y = y_j). \tag{5.29}$$

Remarquons que son calcul nécessite en général la connaissance de la loi du couple (X, Y), celle des lois marginales ne suffit pas.

Théorème 5.16 (Inégalité de Cauchy Schwarz)
Si X et Y ont des moments d'ordre 2 :

$$|\,\mathbb{E}(XY)| \leq (\mathbb{E}\, X^2)^{1/2} (\mathbb{E}\, Y^2)^{1/2}. \tag{5.30}$$

Preuve : L'existence de $\mathbb{E}(XY)$ est assurée par celle de $\mathbb{E}\,X^2$ et de $\mathbb{E}\,Y^2$. De même *pour tout* $t \in \mathbb{R}$ le trinôme

$$t^2\,\mathbb{E}\,Y^2 + 2t\,\mathbb{E}(XY) + \mathbb{E}\,X^2 = \mathbb{E}(X+tY)^2$$

est défini et positif. Ceci n'est possible que si son discriminant est négatif. Ceci se traduit par :

$$\Delta' = \big(\mathbb{E}(XY)\big)^2 - \mathbb{E}\,X^2\,\mathbb{E}\,Y^2 \le 0,$$

ce qui fournit exactement l'inégalité cherchée. ■

L'inégalité de Cauchy Schwarz est une égalité si et seulement si $\Delta' = 0$, ce qui équivaut à l'existence d'un zéro double t_0 pour le trinôme, autrement dit à l'existence d'un réel t_0 tel que $\mathbb{E}(X + t_0Y)^2 = 0$. Par la proposition 5.13, ceci équivaut à $X + t_0Y = 0$ (p.s.). Ainsi une CNS d'égalité dans (5.30) est la colinéarité (presque sûre) de X et Y.

Contrairement à ce qui se passe pour la somme, l'espérance d'un produit n'est en général pas le produit des espérances (exercice 5.21). C'est néanmoins vrai dans le cas de l'indépendance.

Proposition 5.17 *Si X et Y sont* indépendantes *et ont une espérance, alors XY a une espérance et*

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

Preuve : Admettons provisoirement que XY possède une espérance. Alors en utilisant (5.29), l'indépendance et le produit de deux séries *absolument convergentes* on obtient :

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{\substack{x_i \in X(\Omega)\\ y_j \in Y(\Omega)}} x_i y_j P(X = x_i, Y = y_j)\\
&= \sum_{\substack{x_i \in X(\Omega)\\ y_j \in Y(\Omega)}} x_i y_j P(X = x_i)P(Y = y_j)\\
&= \sum_{x_i \in X(\Omega)} x_i P(X = x_i) \sum_{y_j \in Y(\Omega)} y_j P(Y = y_j)\\
&= \mathbb{E}\,X\,\mathbb{E}\,Y.
\end{aligned}$$

Pour compléter la preuve, il reste à montrer que les hypothèses $\mathbb{E}\,|X| < +\infty$ et $\mathbb{E}\,|Y| < +\infty$ entraînent $\mathbb{E}\,|XY| < +\infty$. En remplaçant les x et les y par leurs valeurs absolues dans les séries ci-dessus et en utilisant les propriétés des séries à termes positifs (à somme éventuellement infinie), on obtient $\mathbb{E}\,|XY| = \mathbb{E}\,|X|\,\mathbb{E}\,|Y|$ d'où $\mathbb{E}\,|XY| < +\infty$. ■

Remarque : La réciproque est fausse. Il peut arriver que $\mathbb{E}(XY) = \mathbb{E}\,X\,\mathbb{E}\,Y$ sans que X et Y soient indépendantes. Prenons par exemple X suivant la loi uniforme sur $\{-1, 0, 1\}$ et $Y = \mathbf{1}_{\{X=0\}}$. D'une part $\mathbb{E}\,X = 0$ et d'autre part XY est identiquement nulle donc $\mathbb{E}(XY) = 0 = \mathbb{E}\,X\,\mathbb{E}\,Y$. Clairement X et Y ne sont pas indépendantes. ■

Regardons maintenant comment calculer $\operatorname{Var}(X+Y)$ lorsque X et Y ont des moments d'ordre 2 (ce qui entraîne l'existence de cette variance...). En utilisant la définition de la variance et la linéarité de l'espérance on obtient :

$$\begin{aligned}
\operatorname{Var}(X+Y) &= \mathbb{E}\big[X + Y - \mathbb{E}(X+Y)\big]^2 \\
&= \mathbb{E}\big[(X - \mathbb{E}\,X) + (Y - \mathbb{E}\,Y)\big]^2 \\
&= \mathbb{E}\big\{(X - \mathbb{E}\,X)^2 + (Y - \mathbb{E}\,Y)^2 \\
&\qquad\qquad + 2(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y)\big\} \\
&= \operatorname{Var} X + \operatorname{Var} Y + 2\,\mathbb{E}\big[(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y)\big].
\end{aligned}$$

Ainsi on voit qu'en général $\operatorname{Var}(X+Y) \neq \operatorname{Var} X + \operatorname{Var} Y$ et qu'il y a un terme correctif. Nous le noterons $2\operatorname{Cov}(X, Y)$.

Définition 5.18 *Si les v.a. X et Y ont des moments d'ordre 2, on appelle covariance du couple aléatoire (X, Y) la quantité :*

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

Remarquons que $\operatorname{Cov}(X, X) = \operatorname{Var} X$.

Proposition 5.19 (Propriétés de la covariance)
Pour tout couple (X, Y) de v.a. ayant des moments d'ordre 2 :

(i) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$.

(ii) Pour tous réels a, b, c, d : $\operatorname{Cov}(aX + b, cY + d) = ac\operatorname{Cov}(X, Y)$.

(iii) $|\operatorname{Cov}(X, Y)| \leq \sigma(X)\sigma(Y)$.

(iv) Si X et Y sont indépendantes alors $\operatorname{Cov}(X, Y) = 0$ *(réciproque fausse).*

Preuve : La propriété de symétrie (i) découle immédiatement de la définition. Pour (ii) il suffit de remarquer que $\mathbb{E}(aX + b) = a\,\mathbb{E}\,X + b$ et d'utiliser la définition. La propriété (iii) est simplement l'inégalité de Cauchy-Schwarz appliquée aux variables $X' = X - \mathbb{E}\,X$ et $Y' = Y - \mathbb{E}\,Y$. Pour vérifier (iv), posons $m = \mathbb{E}\,X$ et $m' = \mathbb{E}\,Y$ et définissons les fonctions affines $f(x) = x - m$ et $g(y) = y - m'$. Il est clair que $\operatorname{Cov}(X, Y) = \mathbb{E}\big(f(X)g(Y)\big)$. Comme les variables X et Y sont indépendantes, il en est de même pour les v.a. $f(X)$ et $g(Y)$. Par la proposition 5.17 on a donc $\operatorname{Cov}(X, Y) = \mathbb{E}\big(f(X)\big)\,\mathbb{E}\big(g(Y)\big)$. Or $\mathbb{E}\big(f(X)\big) = \mathbb{E}(X - m) = \mathbb{E}\,X - m = 0$ d'où $\operatorname{Cov}(X, Y) = 0$. Pour voir que la réciproque est fausse, on peut utiliser le même contre exemple que pour la réciproque de la proposition 5.17. ■

Définition 5.20 (Coefficient de corrélation)
Si X et Y sont des v.a. non constantes ayant des moments d'ordre 2, on appelle coefficient de corrélation entre X et Y la quantité :

$$\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

D'après (*iii*) on a toujours $-1 \leq \rho(X,Y) \leq 1$. D'autre part il résulte facilement du cas d'égalité dans l'inégalité de Cauchy-Scharz que $|\rho|$ est maximal lorsque Y est une fonction affine de X : $Y = aX + b$. Quand $\rho = 0$ (ce qui arrive en particulier lorsque X et Y sont indépendantes), on dit que X et Y sont *non corrélées.*

Proposition 5.21 (Formule de Koenig)
Si la covariance de X et Y existe, elle peut se calculer par :

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

Preuve : La vérification est analogue à celle de la formule de Koenig pour la variance (qui n'est que le cas particulier $Y = X$) et est laissée en exercice. ■

Proposition 5.22 (Variance d'une somme, cas général)
Si les v.a. $X_1, \dots, X_n$ ont des moments d'ordre 2 :

$$\begin{aligned} \operatorname{Var}\Big(\sum_{i=1}^n X_i\Big) &= \sum_{i,j=1}^n \operatorname{Cov}(X_i, X_j) && (5.31)\\ &= \sum_{i=1}^n \operatorname{Var} X_i + \sum_{\substack{i,j=1\\ i\neq j}}^n \operatorname{Cov}(X_i, X_j) && (5.32)\end{aligned}$$

Preuve : Nous avons déjà rencontré le cas $n = 2$ pour lequel (5.32) s'écrit :

$$\operatorname{Var}(X+Y) = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}(X,Y). \tag{5.33}$$

Pour n quelconque, l'identité algébrique :

$$\Big(\sum_{i=1}^n Y_i\Big)^2 = \sum_{i,j=1}^n Y_i Y_j.$$

utilisée avec $Y_i = X_i - \mathbb{E}\,X_i$ et la linéarité de l'espérance nous donnent :

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^n X_i\right) &= \mathbb{E}\left\{\sum_{i=1}^n X_i - \mathbb{E}\left(\sum_{i=1}^n X_i\right)\right\}^2 \\
&= \mathbb{E}\left\{\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}\,X_i\right\}^2 \\
&= \mathbb{E}\left\{\sum_{i=1}^n Y_i\right\}^2 \\
&= \sum_{i,j=1}^n \mathbb{E}(Y_iY_j) = \sum_{i,j=1}^n \operatorname{Cov}(X_i, X_j).
\end{aligned}$$

■

Proposition 5.23 (Variance d'une somme et indépendance)
Si $X_1, \dots, X_n$ sont indépendantes et ont des moments d'ordre 2 :

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \operatorname{Var} X_i \tag{5.34}$$

En particulier, si les X_i sont indépendantes, de même loi ayant un moment d'ordre 2 :

$$\operatorname{Var}\left(\sum_{i=1}^n X_i\right) = n \operatorname{Var} X_1. \tag{5.35}$$

Preuve : Il suffit d'appliquer (5.32) en remarquant que l'indépendance implique la nullité des $\operatorname{Cov}(X_i, X_j)$ pour $i \neq j$. Si de plus les X_i ont même loi, elles ont même variance : $\operatorname{Var} X_i = \operatorname{Var} X_1$, ce qui donne (5.35). Remarquons qu'en fait nous avons seulement utilisé l'indépendance deux à deux et pas l'indépendance mutuelle. La proposition est donc vraie sous cette hypothèse plus générale[5]. ■

5.5 Exercices

Ex 5.1. Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}$ et d'espérance finie.

1) Vérifier que

$$P(X > 0) \leq \mathbb{E}\,X.$$

Donner un contre-exemple lorsque X n'est pas à valeurs entières.

2) On suppose que X a un moment d'ordre 2. Montrer que

$$P(X \geq 2) \leq \frac{1}{2}\,\mathbb{E}\,X(X-1).$$

Proposer et montrer une inégalité du même type s'il y a un moment d'ordre k.

[5] Et même dans le cas encore plus général de v.a. X_i deux à deux non corrélées...

Ex 5.2. Soit X une variable aléatoire discrète à valeurs dans $\mathbb{N}^*$. Pour tout $k \in \mathbb{N}^*$, on note $p_k = P(X = k)$. On suppose que X a une espérance mathématique $\mathbb{E}\, X$ finie et que la suite $(p_k)_{k\geq 1}$ est *décroissante* sur $\mathbb{N}^*$.

1) Démontrer l'inégalité :

$$\forall k \in \mathbb{N}^*, \quad P(X = k) < \frac{2\, \mathbb{E}\, X}{k^2}. \tag{5.36}$$

Indication : Considérer la somme partielle de rang k de la série définissant $\mathbb{E}\, X$.

2) L'inégalité (5.36) reste-t-elle vraie sans l'hypothèse de décroissance de $(p_k)_{k\geq 1}$?

3) Est-il possible qu'il existe une constante $c > 0$ et un entier k_0 tels que :

$$\forall k \geq k_0, \quad P(X = k) \geq \frac{c\, \mathbb{E}\, X}{k^2}? \tag{5.37}$$

Ex 5.3. On suppose que le couple de variables aléatoires discrètes (X, Y) a une loi donnée par :

$$\forall (i, j) \in \mathbb{N}^2, \quad P(X = i, Y = j) = \frac{\alpha}{(1 + i + j)!},$$

où α est une constante strictement positive qui sera précisée ultérieurement.

1) Expliquer sans calcul pourquoi les marginales X et Y ont même loi.

2) On pose $S = X + Y$. Montrer que :

$$\forall k \in \mathbb{N}, \quad P(S = k) = \frac{\alpha}{k!}.$$

3) En déduire la valeur de α et reconnaître la loi de S.

4) Calculer $P(X = 0)$. Les variables aléatoires X et Y sont-elles indépendantes ?

5) Donner sans calcul l'espérance de S et en déduire la valeur de $\mathbb{E}\, X$. La même méthode peut-elle servir pour obtenir $\operatorname{Var} X$?

6) Calculer $P(X = Y)$ et en déduire sans calcul $P(X > Y)$.

Ex 5.4. Une urne contient $2N$ jetons numérotés de 1 à N, chaque numéro étant présent deux fois. On tire m jetons sans remise. Soit S la variable aléatoire égale au nombre de paires restant dans l'urne. Calculer l'espérance de S. Ce modèle fut utilisé par Bernoulli au 18ème siècle pour étudier le nombre de couples survivants dans une population de N couples après m décès.
Indications : Pour $1 \leq i \leq N$, on considèrera la variable aléatoire X_i définie par :

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ème paire reste dans l'urne,} \\ 0 & \text{sinon.} \end{cases}$$

Calculer $P(X_i = 1)$ et $\mathbb{E}\, X_i$. En déduire la valeur de $\mathbb{E}\, S$.

Ex 5.5. Dans une urne contenant au départ une boule verte et une rouge on effectue une suite de tirages d'une boule selon la procédure suivante. Chaque fois que l'on tire une boule verte, on la remet dans l'urne *en y rajoutant une boule rouge*. Si l'on tire une boule rouge, on arrête les tirages. On désigne par X le nombre de tirages effectués par cette procédure. On notera V_i (resp. R_i) l'événement *obtention d'une boule verte au i-ème tirage* (resp. *rouge*).

1) Pour $k \in \mathbb{N}^*$, donner une expression de l'événement $\{X = k\}$ à l'aide des événements V_i $(1 \leq i \leq k-1)$ et R_k.

2) Que vaut $P(V_n \mid V_1 \cap \ldots \cap V_{n-1})$ pour $n \geq 2$?

3) Déterminer la loi de X.

4) Calculer $\mathbb{E}\, X$.

5) Calculer l'espérance de la variable aléatoire $\frac{1}{X}$.

6) On recommence l'expérience en changeant la procédure : à chaque tirage d'une boule verte on la remet dans l'urne *en y rajoutant une boule verte*. Comme précédemment, on interrompt les tirages à la première apparition d'une boule rouge. Soit Y le nombre de tirages effectués suivant cette nouvelle procédure. Trouver la loi de Y. Que peut-on dire de l'espérance de Y ? Interpréter.

Ex 5.6. *La formule de Poincaré*
En utilisant la relation $\mathbb{E}(\mathbf{1}_A) = P(A)$, démontrer la formule de Poincaré : Pour toute famille $A_1, \ldots, A_n$ d'événements :

$$\begin{aligned} P\Big(\bigcup_{i=1}^{n} A_i\Big) &= \sum_{i=1}^{n} P(A_i) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1}\cap A_{i_2}) + \cdots \\ &\quad \cdots + (-1)^{k+1} \sum_{1\leq i_1<i_2<\ldots<i_k\leq n} P(A_{i_1}\cap\cdots\cap A_{i_k}) + \cdots \\ &\quad \cdots + (-1)^{n+1} P(A_1\cap\cdots\cap A_n). \end{aligned}$$

Indication : Écrire l'indicatrice de l'intersection des A_i^c comme un produit d'indicatrices...

Ex 5.7. Un polycopié de 140 pages contient avant la première lecture un nombre $n \geq 100$ d'erreurs typographiques réparties au hasard (on suppose que n est inférieur au nombre total de caractères que peut contenir une page).

1) Quelle est la loi exacte de la variable aléatoire Y_0 égale au nombre d'erreurs de la page 13 avant la première lecture ? Par quelle loi peut-on l'approximer ? On supposera dans les deux questions suivantes que Y_0 suit cette loi approchée.

2) On effectue des lectures successives de cette page. A chaque lecture les erreurs subsistantes peuvent être détectées (et donc corrigées) indépendamment les unes des autres, chacune avec une probabilité 1/3. On posera $\lambda = n/140$.

On note Y_i le nombre d'erreurs restant sur la page 13 après la i-ème lecture. Calculer pour j et k entiers les probabilités conditionnelles $P(Y_1 = j \mid Y_0 = k)$ (distinguer selon que $j \leq k$ ou $j > k$).

3) En déduire la loi de Y_1, puis celle de Y_i.

4) En fait le nombre total d'erreurs dans le polycopié étant inconnu, on le modélise par une variable aléatoire N. Comme N est au plus égale au nombre maximum de caractères que peuvent contenir les 140 pages du polycopié, N est bornée et son espérance existe. On la note μ. Il sera commode d'écrire toutes les formules comme si l'ensemble des valeurs possibles de N était $N(\Omega) = \mathbb{N}$ (cela revient à écrire des séries dans lesquelles $P(N = n) = 0$ pour n assez grand...). On définit la variable aléatoire Z_i ($i \geq 1$) nombre total d'erreurs subsistant dans l'ensemble du polycopié après la i-ème lecture (l'indépendance des détections et leurs probabilités sont les mêmes que ci-dessus). Comme la loi de N est inconnue, nous ne pouvons plus calculer les lois des Z_i. Calculer $P(Z_1 = k \mid N = n)$. Montrer que $\mathbb{E}\, Z_1 = 2\mu/3$.

5) Exprimer $\mathbb{E}\, Z_i$ en fonction de μ.

Ex 5.8.

1) Soit Z une variable aléatoire de loi uniforme sur l'ensemble $\{1, 2, \dots, n\}$. Calculer explicitement $\mathbb{E}\, Z$ en fonction de n.

2) On considère l'expérience aléatoire suivante. On dispose au départ d'une urne vide et d'autant de boules numérotées que l'on veut. Par un procédé quelconque, on génère une variable aléatoire X à valeurs dans $\mathbb{N}^*$ (on suppose pour simplifier que $P(X = n) > 0$ pour tout $n \in \mathbb{N}^*$). Soit n la valeur effectivement obtenue par ce procédé. On rajoute alors dans l'urne n boules numérotées de 1 à n. On effectue enfin un tirage d'une boule dans cette urne. Soit Y son numéro. On a ainsi :

$$P(Y = k \mid X = n) = \begin{cases} \frac{1}{n} & \text{si} \quad 1 \leq k \leq n, \\ 0 & \text{sinon.} \end{cases}$$

Calculer $\displaystyle\sum_{k=1}^{+\infty} kP(Y = k \mid X = n)$.

3) On suppose que X a une espérance. Montrer que $\mathbb{E}\, Y = \dfrac{1}{2}\, \mathbb{E}\, X + \dfrac{1}{2}$.
Indication : Comme on ne connaît pas la loi de Y, il est commode de démarrer le calcul de la manière suivante :

$$\mathbb{E}\, Y = \sum_{k=1}^{+\infty} kP(Y = k) = \sum_{k=1}^{+\infty} k \sum_{n=1}^{+\infty} P(Y = k \mid X = n)P(X = n).$$

4) Pour générer la variable aléatoire X, on effectue des lancers répétés d'un même dé et on appelle X le nombre de lancers *avant* la *deuxième* obtention d'un « six ». Par exemple si la deuxième apparition du « six » a lieu lors du $10^{\text{ème}}$ lancer, on prend $X = 9$. Déterminer la loi de X puis celle de Y.

Ex 5.9. Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ et ayant une espérance.

1) Montrer que
$$\lim_{n\to\infty} nP(X \geq n) = 0.$$
Indication : On pourra utiliser, après l'avoir justifiée, l'inégalité :
$$nP(X \geq n) \leq \sum_{k=n}^{+\infty} kP(X = k).$$

2) Exprimer $\displaystyle\sum_{k=1}^{n} P(X \geq k)$ à l'aide des $P(X = k)$ $(1 \leq k \leq n-1)$ et de $P(X \geq n)$.

3) En déduire que :
$$\mathbb{E}\, X = \sum_{k=1}^{+\infty} P(X \geq k).$$

4) La valeur de X est fournie par un générateur de nombres aléatoires. Une fois connue effectivement cette valeur, disons n, on met dans une urne $n+1$ boules numérotées de 0 à n. On tire enfin une boule de cette urne et on note Y son numéro. Discuter en fonction de k et n des valeurs de la probabilité conditionnelle :
$$P(Y = k \mid X = n).$$

5) Le générateur est réglé de façon que X suive la loi de Poisson de paramètre λ. Trouver une relation simple entre $P(Y = k)$ et $P(X > k)$.

6) Expliquer comment la question 3) vous permet de contrôler ce résultat.

Ex 5.10. Généraliser la proposition 5.5 en remplaçant dans (i) $X \geq 0$ par $P(X \geq 0) = 1$ et dans (ii), $X \leq Y$ par $P(X \leq Y) = 1$.

Ex 5.11. *Un problème d'assurances.*

Une compagnie d'assurance étudie le coût annuel des sinistres « incendie de maison individuelle » pour ses clients d'un département donné. Le modèle mathématique utilisé est le suivant. On note X_i le coût (arrondi en nombre entier de francs) du $i^{\text{ème}}$ sinistre de l'année prochaine. On suppose que les X_i sont des variables aléatoires indépendantes de même loi d'espérance connue $\mathbb{E}\, X_i = \mu$. Le nombre de sinistres de l'année prochaine est une *variable aléatoire* N suivant la loi de Poisson de paramètre λ. On note $S_0 = 0$ et pour tout entier $n \geq 1$:
$$S_n = \sum_{i=1}^{n} X_i.$$

Le coût annuel des sinistres est la variable aléatoire T définie par

$$T = \begin{cases} \displaystyle\sum_{i=1}^{N} X_i, & \text{si } N \geq 1; \\ 0, & \text{si } N = 0. \end{cases}$$

On admettra que pour tout n, les variables aléatoires S_n et N sont indépendantes, ce qui implique que pour tous entiers i, j, les évènements $\{S_n = i\}$ et $\{N = j\}$ sont indépendants. Le but de l'exercice est d'exprimer $\mathbb{E}\,T$ en fonction de λ et μ. On remarquera que l'on ne peut pas utiliser directement la linéarité de l'espérance dans la somme définissant T car le nombre de termes N est *aléatoire*.

1) Que valent $\mathbb{E}\,N$ et $\mathbb{E}\,S_n$?

2) Montrer que pour tous entiers k, n,

$$P(T = k \mid N = n) = P(S_n = k).$$

3) Prouver que $\mathbb{E}\,T = \lambda\mu$. *Indication* : On pourra partir de la formule

$$\mathbb{E}\,T = \sum_{k=0}^{+\infty} kP(T = k)$$

et exprimer les $P(T = k)$ en conditionnant par rapport aux valeurs possibles de N.

Ex 5.12. L'avancement de certains jeux se fait selon la règle suivante : le joueur lance deux dés et avance son pion d'un nombre de cases donné par la somme des points obtenus. S'il a obtenu un double, il peut rejouer et avancer encore et ainsi de suite tant qu'il obtient des doubles. Après le premier lancer sans double, il passe la main au joueur suivant. On s'intéresse à la somme S des points obtenus par cette procédure par un joueur lors d'un tour. Pour $i \geq 1$, on note D_i l'événement *le i-ème lancer a lieu et donne un double* (on a donc $D_i \subset D_{i-1}$). On note X_i la variable aléatoire égale à 0 si le i-ème lancer n'a pas lieu et à la somme des points du i-ème lancer sinon.

1) Donner *brièvement* la loi de X_1 et son espérance.

2) Calculer $P(D_1)$, $P(D_i \mid D_{i-1})$ et trouver une relation de récurrence entre $P(D_i)$ et $P(D_{i-1})$. En déduire l'expression explicite de $P(D_i)$ en fonction de i.

3) Pour $i \geq 2$, donner les valeurs des probabilités conditionnelles

$$P(X_i = k \mid D_{i-1}) \quad \text{et} \quad P(X_i = k \mid D_{i-1}^c),$$

en distinguant les cas $k = 0$ et $k \in \{2, \dots, 12\}$.

4) En déduire une relation simple entre $P(X_i = k)$ et $P(X_1 = k)$ pour $k \in \{2, \dots, 12\}$ puis entre $\mathbb{E}\,X_i$ et $\mathbb{E}\,X_1$. Quelle est la loi de X_i ?

5) On pose $S_n = \sum_{i=1}^{n} X_i$. Montrer que $\mathbb{E}\, S_n$ converge en croissant vers $8,4$ lorsque n tend vers $+\infty$.

6) On note N le nombre aléatoire de lancers effectués selon la procédure ci-dessus. Calculer $P(N \geq n)$ et en déduire que $P(N = +\infty) = 0$. On admet alors que l'on peut remplacer Ω par $\Omega' = \{\omega \in \Omega;\ N(\omega) < +\infty\}$. C'est ce que nous ferons désormais. On peut alors considérer N comme une variable aléatoire à valeurs dans $\mathbb{N}^*$. Quelle est sa loi ?

7) On définit la variable aléatoire S sur Ω' par :

$$S(\omega) = \sum_{i=1}^{+\infty} X_i(\omega).$$

Notons que puisque ω est dans Ω', tous les termes de la série sont nuls à partir du rang (aléatoire) $N(\omega) + 1$, il n'y a donc pas de problème de convergence. S est le nombre total de points obtenus, sauf dans le cas où il y a une infinité de lancers. Comme celui-ci a une probabilité nulle, le fait de le laisser tomber n'affecte pas la loi du nombre total de points obtenus qui est donc celle de S.

Après avoir justifié l'inclusion :

$$\{S = k\} \subset \{12N \geq k\},$$

valable pour tout $k \in \mathbb{N}$, montrer que :

$$P(S = k) \leq 36q^k \quad \text{où} \quad q = 6^{-1/12}.$$

En déduire l'existence de $\mathbb{E}\, S$.

8) On définit sur Ω' la variable aléatoire $R_n = S - S_{n-1}$. Montrer qu'elle vérifie :

$$R_n \leq S\mathbf{1}_{\{S \geq 2n\}}.$$

En déduire que pour tout $n \in \mathbb{N}^*$:

$$0 \leq \mathbb{E}\, S - \mathbb{E}\, S_{n-1} \leq \mathbb{E}(S\mathbf{1}_{\{S \geq 2n\}}).$$

9) Exprimer $\mathbb{E}(S\mathbf{1}_{\{S \geq 2n\}})$ à l'aide des $P(S = k)$ et montrer qu'elle tend vers zéro quand n tend vers l'infini.

10) Conclure.

Ex 5.13. Montrer qu'une variable aléatoire bornée (i.e. $X(\Omega)$ est une partie bornée[6] de $\mathbb{R}$) a des moments de tout ordre.

Ex 5.14. Montrer qu'une loi de Poisson possède des moments de tout ordre. Si X suit la loi $\mathcal{P}(\lambda)$, calculer $\mathbb{E}(X - \lambda)^3$ et $\mathbb{E}(X - \lambda)^4$.

[6] Pas forcément finie !

Ex 5.15. Soit X une variable aléatoire. On suppose qu'il existe une constante $a > 0$ telle que $\mathbb{E}\exp(aX) < +\infty$. Que peut-on dire de la vitesse de convergence vers 0 de $P(X > t)$ quand t tend vers $+\infty$? Donner un exemple de v.a. non bornée vérifiant cette condition pour tout a.

Ex 5.16. En s'inspirant des exemples 5.7 et 5.10, calculer la variance d'une loi hypergéométrique.

Ex 5.17. ↑ 3.7, ↑ 4.6
Calculer l'espérance et la variance d'une loi binomiale négative de paramètres n et p.

Ex 5.18. Reprendre l'exemple 5.14 en calculant $\mathbb{E}\,S^2$ et en utilisant l'inégalité de Markov à l'ordre 2. Comparer avec le résultat fourni par l'inégalité de Tchebycheff.

Ex 5.19. Un hebdomadaire lance une campagne d'abonnement en envoyant par la poste n lettres proposant trois types d'abonnement : 1 an pour 200F, 2 ans pour 360 F et 3 ans pour 500 F. Les campagnes précédentes permettent de savoir que les probabilités p_i qu'une personne ainsi sollicitée s'abonne pour i années sont : $p_1 = 0.09$, $p_2 = 0.04$ et $p_3 = 0.02$. On admet que les différentes personnes sollicitées ont des comportements mutuellement indépendants.

1) Quelle est la probabilité qu'une personne sollicitée ne réponde pas ? Quelle est la loi du nombre S_n de lettres sans réponse ? Donner $\mathbb{E}\,S_n$ et $\operatorname{Var} S_n$.

2) L'éditeur se demande quel nombre minimal de lettres il doit envoyer pour qu'avec une probabilité d'au moins 0.95, le pourcentage de lettres sans réponses reste *au dessous* de 90%. En utilisant l'inégalité de Tchebycheff proposez lui une solution.

3) Pour $i = 1, 2, 3$, donner la loi de Y_i nombre d'abonnements pour i années reçus à la fin de la campagne. Les variables aléatoires Y_1 et Y_2 sont-elles indépendantes ?

4) Soit R_n la rentrée financière procurée par les abonnements. Que vaut $\mathbb{E}\,R_n$?

Ex 5.20. On se propose de démontrer l'inégalité de Cantelli : si X est une v.a. ayant une espérance m et une variance σ^2, on a :

$$\forall t > 0 \qquad P(X - m \geq t) \leq \frac{\sigma^2}{\sigma^2 + t^2}.$$

1) Vérifier que l'on peut se ramener au cas $m = 0$.

2) Montrer que pour tout $u \geq 0$:

$$P(X \geq t) \leq P\left((X+u)^2 \geq (t+u)^2\right) \leq \frac{\sigma^2 + u^2}{(t+u)^2}.$$

3) Conclure en choisissant u de de façon à minimiser le majorant ainsi obtenu.

Ex 5.21.

1) Donner un exemple très simple de couple (X, Y) de v.a. non indépendantes pour lequel on a $\mathbb{E}(XY) \neq \mathbb{E}\, X\, \mathbb{E}\, Y$.

2) Donner un exemple de couple (X, Y) de v.a. tel que le produit XY ait une espérance sans que X ni Y n'aient de moment d'ordre 2.

Ex 5.22. *Covariances d'une loi multinomiale*[7] $\uparrow$ 4.11
On considère une suite de n épreuves répétées indépendantes, chaque épreuve ayant k résultats possibles $r_1, \dots, r_k$. On note p_i la probabilité de réalisation du résultat r_i lors d'une épreuve donnée. Par exemple si on lance n fois un dé, $k = 6$ et $p_i = 1/6$ pour $1 \leq i \leq 6$. Si on lance n fois une paire de dés et que l'on s'intéresse à la somme des points, $k = 11$ et les p_i sont donnés par le tableau des $P(S = i + 1)$ page 47 : $p_1 = 1/36$, $p_2 = 2/36$,...

Pour $1 \leq i \leq k$, notons X_i le nombre de réalisations du résultat r_i au cours des n épreuves.

1) Expliquer sans calcul pourquoi $\operatorname{Var}(X_1 + \cdots + X_k) = 0$.

2) Quelle est la loi de X_i ? Que vaut sa variance ?

3) Pour $i \neq j$, donner la loi et la variance de $X_i + X_j$.

4) En déduire $\operatorname{Cov}(X_i, X_j)$.

5) Contrôler ce résultat en développant $\operatorname{Var}(X_1 + \cdots + X_k)$ et en utilisant la première question.

Ex 5.23. *Une preuve probabiliste d'un théorème d'analyse* $\downarrow$ 6.10
Le but de cet exercice est de présenter une démonstration probabiliste d'un célèbre théorème d'analyse (Bernstein-Weierstrass-Stone) : toute fonction continue sur $[0, 1]$ est limite *uniforme* sur cet intervalle d'une suite de polynômes. La méthode utilisée ici est due à Bernstein et donne une construction explicite de la suite de polynômes. Les trois dernières questions sont consacrées à la vitesse de convergence. On note $\mathcal{C}[0, 1]$ l'espace des fonctions continues sur $[0, 1]$ muni de la norme :

$$\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|.$$

La convergence suivant cette norme n'est autre que la convergence uniforme. Si $f \in \mathcal{C}[0, 1]$, on définit son polynôme de Bernstein d'ordre n par :

$$\mathrm{B}_n f(x) = \sum_{k=0}^{n} C_n^k f\Big(\frac{k}{n}\Big) x^k (1 - x)^{n-k}, \quad n \geq 1.$$

[7] Il n'est pas nécessaire de connaître la loi multinomiale (exercice 4.11) pour pouvoir faire cet exercice.

1) Justifier la relation :

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

2) Pour $x \in [0,1]$ fixé, considérons la variable aléatoire S_n de loi binomiale $\mathcal{B}(n,x)$. Vérifier que :

$$\mathbb{E}\, f\Big(\frac{S_n}{n}\Big) = \mathrm{B}_n f(x).$$

3) Justifier les inégalités :

$$\begin{aligned} |f(x) - \mathrm{B}_n f(x)| &\leq \sum_{k:|f(x)-f(k/n)|<\varepsilon} \varepsilon C_n^k x^k (1-x)^{n-k} \\ &\quad + \sum_{k:|f(x)-f(k/n)|\geq\varepsilon} 2\|f\|_\infty C_n^k x^k (1-x)^{n-k} \\ &\leq \varepsilon + 2\|f\|_\infty P\Big(\Big|f(x) - f\Big(\frac{S_n}{n}\Big)\Big| \geq \varepsilon\Big). \end{aligned} \tag{5.38}$$

4) La fonction f est uniformément continue sur $[0,1]$ (pourquoi ?). On a donc :

$$\forall\varepsilon > 0, \exists\delta > 0, \quad \text{tel que } |x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon,$$

δ ne dépendant que de f et ε, mais pas de x. En déduire que

$$P\Big(\Big|f(x) - f\Big(\frac{S_n}{n}\Big)\Big| \geq \varepsilon\Big) \leq P(|S_n - nx| \geq n\delta),$$

puis en appliquant l'inégalité de Tchebycheff :

$$P\Big(\Big|f(x) - f\Big(\frac{S_n}{n}\Big)\Big| \geq \varepsilon\Big) \leq \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2}.$$

5) En reportant cette majoration dans (5.38), on obtient finalement :

$$\forall n \geq 1, \forall x \in [0,1], \quad |f(x) - \mathrm{B}_n f(x)| \leq \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n} \tag{5.39}$$

Conclure.

6) On s'intéresse maintenant à la vitesse de convergence. Supposons d'abord que f est lipschitzienne : il existe une constante a telle que

$$\forall x, y \in [0,1], \quad |f(x) - f(y)| \leq a|x-y|.$$

On peut alors prendre $\delta = \varepsilon/a$ dans l'écriture de la continuité uniforme de f. En choisissant convenablement ε en fonction de n dans (5.39), en déduire que $\|f - \mathrm{B}_n f\|_\infty = O(n^{-1/3})$.

7) Plus généralement, on suppose f hölderienne d'exposant α : il existe des constantes $0 < \alpha \leq 1$ et $a > 0$ telles que :

$$\forall x, y \in [0,1], \quad |f(x) - f(y)| \leq a|x-y|^\alpha.$$

Montrer qu'alors $\|f - \mathrm{B}_n f\|_\infty = O\Big(n^{-\alpha/(2+\alpha)}\Big)$.

Chapitre 6

Loi des grands nombres

Les inégalités de moment (Markov, Tchebycheff) ont d'importantes applications à la convergence de la moyenne arithmétique :

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

des n premiers termes d'une suite de v.a. indépendantes et de même loi. Ce type de résultat est connu sous le nom de loi des grands nombres. Nous en donnons un premier aperçu[1].

6.1 Deux modes de convergence

Pour commencer, il convient de préciser ce que l'on entend par convergence d'une suite de variables aléatoires (X_n) vers une v.a. X. Comme les X_n sont des applications de Ω dans $\mathbb{R}$, le premier mode de convergence auquel on pense est la convergence *pour tout* $\omega \in \Omega$ de la suite de réels $X_n(\omega)$ vers le réel $X(\omega)$. Ceci correspond à la convergence simple d'une suite d'applications en analyse. Malheureusement pour le type de résultat que nous avons en vue, ce mode de convergence est trop restrictif. Pour la loi des grands nombres, même dans le cas le plus favorable[2], on ne peut empêcher que la suite étudiée diverge pour une infinité de ω. Ce qui sauve la situation est que l'ensemble de ces ω a une probabilité nulle. Ceci nous amène à définir la convergence *presque sûre* :

[1]Seuls sont au programme du DEUG dans ce chapitre, la convergence en probabilité et la loi faible des grands nombres avec ses applications. La convergence presque sûre et la loi forte des grands nombres sont destinés aux lecteurs plus curieux ou plus avancés. Ils pourront être considérés comme une introduction au cours de Licence. Néanmoins ils ont été rédigés en n'utilisant que des outils mathématiques du DEUG.

[2]Voir la discussion à propos de la loi forte des grands nombres pour les fréquences section 6.5.

Définition 6.1 (Convergence presque sûre)
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires et X une v.a. définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On dit que X_n converge presque sûrement vers X si l'ensemble des ω tels que $X_n(\omega)$ converge vers $X(\omega)$ a pour probabilité 1.
Notation : $X_n \xrightarrow[n\to+\infty]{p.s.} X$.

Rappelons qu'un événement de probabilité 1 n'est pas forcément égal à Ω, il peut même y avoir une infinité d'éléments dans son complémentaire (cf. page 6 la remarque dans l'exemple 1.2). Remarquons aussi que l'ensemble Ω' des ω tels que $X_n(\omega)$ converge vers $X(\omega)$ est bien un événement observable (cf. exercice 1.5 en remplaçant les fonctions par des variables aléatoires), c'est-à-dire un événement de la famille $\mathcal{F}$. Il est donc légitime de parler de sa probabilité.

Dans la convergence presque sûre, le rang n_0 à partir duquel on peut approximer $X_n(\omega)$ par $X(\omega)$ avec une erreur inférieure à ε dépend à la fois de ε et de $\omega \in \Omega'$: $n_0 = n_0(\varepsilon, \omega)$. On ne sait pas toujours expliciter la façon dont $n_0(\varepsilon, \omega)$ dépend de ω. D'autre part on peut très bien avoir $\sup\{n_0(\varepsilon, \omega),\ \omega \in \Omega'\} = +\infty$. Ceci fait de la convergence presque sûre en général un résultat essentiellement théorique[3]. Supposons que la valeur de X_n dépende du résultat de n épreuves répétées (ou de n observations). Savoir que X_n converge presque sûrement vers X ne permet pas de *prédire* le nombre non aléatoire n d'épreuves (ou d'observations) à partir duquel on aura $|X_n(\omega) - X(\omega)| < \varepsilon$ (sinon pour tous les $\omega \in \Omega'$, du moins avec une probabilité supérieure à un seuil fixé à l'avance par exemple 95%, 99%,...). Or cette question a une grande importance pratique pour le statisticien. C'est l'une des raisons de l'introduction de la *convergence en probabilité* qui permet de répondre à cette question lorsque l'on connaît la vitesse de convergence selon ce mode.

Définition 6.2 (Convergence en probabilité)
Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires et X une v.a. définies sur le même espace probabilisé $(\Omega, \mathcal{F}, P)$. On dit que X_n converge en probabilité vers X si :

$$\forall \varepsilon > 0, \qquad \lim_{n\to+\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Notation : $X_n \xrightarrow[n\to+\infty]{Pr} X$.

La convergence presque sûre implique la convergence en probabilité (exercice 6.3), la réciproque est fausse (exercice 6.4). Pour cette raison, la convergence en probabilité de la suite M_n définie en introduction s'appelle une loi *faible* des grands nombres, sa convergence presque sûre une loi *forte* des grands nombres.

[3]Sauf si l'on connaît la loi de la v.a. $\omega \mapsto n_0(\varepsilon, \omega)$, ou au moins si l'on sait majorer $P(n_0 > t)$...

6.2 Loi faible des grands nombres

Théorème 6.3 *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires deux à deux indépendantes, de même loi ayant un moment d'ordre* 2. *Alors :*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{Pr} \mathbb{E}\,X_1.$$

Preuve : Ici, la v.a. limite est la constante $\mathbb{E}\,X_1$ (ou n'importe quel $\mathbb{E}\,X_i$, puisque les X_i ayant même loi ont même espérance). Il s'agit donc de vérifier que :

$$\forall\varepsilon > 0, \quad \lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}\,X_1\right| \geq \varepsilon\right) = 0.$$

Posons $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. On a :

$$\mathbb{E}\,M_n = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\,X_i = \mathbb{E}\,X_1. \tag{6.1}$$

D'autre part, les X_i étant deux à deux indépendantes et de même loi on a d'après la proposition 5.23 :

$$\operatorname{Var} M_n = \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}(n\operatorname{Var} X_1) = \frac{1}{n}\operatorname{Var} X_1. \tag{6.2}$$

L'inégalité de Tchebycheff appliquée à chaque M_n nous dit que pour $\varepsilon > 0$ *fixé* :

$$\forall n \in \mathbb{N}^*, \quad P(|M_n - \mathbb{E}\,M_n| \geq \varepsilon) \leq \frac{\operatorname{Var} M_n}{\varepsilon^2}.$$

D'où compte tenu du calcul de $\mathbb{E}\,M_n$ et $\operatorname{Var} M_n$:

$$\forall n \in \mathbb{N}^*, \quad P(|M_n - \mathbb{E}\,X_1| \geq \varepsilon) \leq \frac{\operatorname{Var} X_1}{n\varepsilon^2}. \tag{6.3}$$

En faisant tendre n vers $+\infty$ (ε restant fixé) on en déduit :

$$\lim_{n\to+\infty} P(|M_n - \mathbb{E}\,X_1| \geq \varepsilon) = 0.$$

Ce raisonnement est valable pour tout $\varepsilon > 0$. ■

Remarque : Nous avons en fait démontré un peu plus que la seule convergence en probabilité. Nous avons d'après (6.3) une *vitesse de convergence* en $O(1/n)$. Si l'on connaît $\operatorname{Var} X_1$ ou si on sait le majorer, on peut donc répondre à la question posée page 118 lors de l'introduction de la convergence en probabilité.

Corollaire 6.4 (Loi faible des g. n. pour les fréquences)
Si $(X_n)_{n\geq 1}$ est une suite de v.a. de Bernoulli indépendantes de même paramètre p, alors :

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{Pr} p.$$

Preuve : Il suffit d'appliquer la loi faible des grands nombres en notant qu'ici $\mathbb{E}\, X_1 = p$. ■

Interprétation : Considérons une suite d'épreuves répétées indépendantes. Pour chaque épreuve la probabilité d'un « succès » est p. Notons X_i l'indicatrice de l'événement *succès à la i-ème épreuve*. Alors :
$S_n = \sum_{i=1}^{n} X_i$ est le nombre de succès en n épreuves et $M_n = n^{-1}S_n$ est la *fréquence* des succès au cours des n premières épreuves. Remarquons que pour tout ω, $0 \leq M_n(\omega) \leq 1$.

6.3 Estimation d'une proportion inconnue

On se propose d'estimer le paramètre p inconnu d'une loi de Bernoulli à partir des *observations* $X_i(\omega)$, $1 \leq i \leq n$, les X_i étant des v.a. de Bernoulli indépendantes de même paramètre p.

Exemple 6.1 *On a une urne comportant des boules rouges en proportion inconnue p et des boules vertes (en proportion $q = 1 - p$). On effectue n tirages d'une boule avec remise.*
Notons :

$$X_i = \mathbf{1}_{\{\text{rouge au } i\text{-ème tirage}\}}$$

et comme ci-dessus désignons par M_n la moyenne arithmétique des X_i ou fréquence d'apparition du rouge au cours des n premiers tirages. D'après la loi faible des grands nombres pour les fréquences, M_n converge en probabilité vers p. Comme on s'attend à ce que M_n soit proche de p pour les grandes valeurs de n, il est naturel d'estimer p par M_n. En fait on observe *une* valeur particulière $M_n(\omega)$ calculée à partir des résultats des n tirages *réellement* effectués. La question pratique qui se pose est de donner une « fourchette » pour l'approximation de p par la valeur observée $M_n(\omega)$. L'inégalité de Tchebycheff (6.3) pour M_n s'écrit ici :

$$P(|M_n - p| \geq t) \leq \frac{\operatorname{Var} X_1}{nt^2} = \frac{p(1-p)}{nt^2}. \tag{6.4}$$

Comme p est inconnu, on ne peut pas utiliser directement ce majorant. On remplace alors $p(1-p)$ par :

$$\sup_{x\in[0,1]} x(1-x) = \frac{1}{4}$$

(la parabole d'équation $y = x(1-x)$ a sa concavité tournée vers les y négatifs, les deux zéros du trinôme sont $x_1 = 0$ et $x_2 = 1$; par symétrie, le sommet a pour abscisse $(x_1 + x_2)/2 = 1/2$ et pour ordonnée $1/2(1-1/2) = 1/4$). En reportant dans (6.4), on obtient *quelle que soit la valeur inconnue* p :

$$P(|M_n - p| \geq t) \leq \frac{\operatorname{Var} X_1}{nt^2} = \frac{1}{4nt^2} \tag{6.5}$$

d'où en passant à l'événement complémentaire :

$$P(M_n - t < p < M_n + t) \geq 1 - \frac{1}{4nt^2}. \tag{6.6}$$

En pratique on remplace M_n par la valeur réellement observée $M_n(\omega)$ et on dit que $I =]M_n(\omega) - t, M_n(\omega) + t[$ est un intervalle de confiance (ou fourchette) pour p. Le deuxième membre de (6.5) peut s'interpréter comme un majorant de la probabilité de se tromper lorsque l'on déclare que p est dans I. On dit aussi que I est un intervalle de confiance au niveau $\alpha \geq 1 - 1/(4nt^2)$. ■

Exemple 6.2 (Sondage) *Avant le second tour d'une élection présidentielle opposant les candidats A et B, un institut de sondage interroge au hasard* 1 000 *personnes dans la rue*[4]. *On note p la proportion d'électeurs décidés à voter pour A dans la* population totale. *Dans l'échantillon sondé, cette proportion est égale à* 0.54. *Proposer un intervalle de confiance pour p au niveau* 0.95.

Le sondage peut être assimilé à un tirage avec remise (en admettant qu'une personne interrogée plusieurs fois accepte de répondre à chaque fois) et on est ramené à la situation de l'exemple précédent. Ici la fréquence *observée* réellement est $M_n(\omega) = 0.54$ et l'inégalité (6.6) nous dit que l'on peut prendre comme intervalle de confiance :

$$I =]0.54 - t, 0.54 + t[\quad \text{avec un niveau} \quad \alpha \geq 1 - \frac{1}{4nt^2}.$$

Comme on souhaite que α soit au moins égal à 0.95, il suffit de choisir la plus petite valeur de t telle que :

$$1 - \frac{1}{4\,000t^2} \geq 0.95 \Leftrightarrow t \geq \frac{1}{10\sqrt{2}} \simeq 0.0707.$$

En prenant $t = 0.071$, on obtient : $I =]0.469, 0.611[$. On remarque qu'une partie de cet intervalle correspond à $p < 1/2$. Ainsi, bien que le sondage donne 54% d'intentions de vote en faveur de A, l'inégalité (6.6) ne nous permet pas de pronostiquer sa victoire avec une probabilité d'erreur inférieure à 5%. ■

[4]Ceci est une simplification volontaire permettant d'assimiler la situation à un tirage avec remise : une même personne peut ainsi être interrogée plusieurs fois au cours du sondage. En pratique les méthodes utilisées par les instituts de sondage pour sélectionner un échantillon sont un peu plus compliquées…

Exemple 6.3 (Sondage, suite) *L'institut de sondage désire présenter à ses clients une fourchette à* $\pm 1\%$ *avec un niveau de confiance égal au moins à* 0.95%*. Combien de personnes doit-il interroger ?*
On repart de (6.6). Cette fois on impose $t = 0.01$ et on cherche n *minimal* tel que :

$$\frac{1}{4n \times 0.01^2} \leq 0.05$$

On trouve $n = 50\,000$, ce qui donne au sondage un coût prohibitif[5]. Nous reviendrons sur ce problème au chapitre suivant. ■

6.4 Convergence presque sûre des fréquences

On peut représenter graphiquement la suite $M_n(\omega)$ des fréquences de succès dans une suite d'épreuves de Bernoulli par la ligne brisée dont les sommets ont pour coordonnées $(n, M_n(\omega))$. A chaque ω correspond ainsi une ligne brisée infinie que nous appellerons *trajectoire*. La loi faible des grands nombres nous donne le comportement asymptotique de ces trajectoires *dans leur ensemble*. Elle signifie grosso modo que pour n grand fixé ($n \geq n_0(\varepsilon)$) la plupart des trajectoires vont traverser le segment vertical d' extrémités $(n, p-\varepsilon)$ et $(n, p+\varepsilon)$. Elle ne nous dit rien sur le comportement *individuel* de chaque trajectoire. Une trajectoire qui traverse $]p-\varepsilon, p+\varepsilon[$ à la verticale de n peut très bien sortir de la bande horizontale engendrée par ce segment au delà de n. Une question naturelle est alors : existe-t-il des trajectoires qui à partir d'un certain rang $n_0 = n_0(\omega, \varepsilon)$ restent dans la *bande* $\{(x, y) \in \mathbb{R}^2,\ x \geq n_0 \text{ et } p-\varepsilon < y < p+\varepsilon\}$? Nous allons montrer que l'ensemble des trajectoires qui vérifient cette propriété pour tout $\varepsilon > 0$ a pour probabilité 1, autrement dit que M_n converge presque sûrement vers p.

Théorème 6.5 (Loi forte des g. n. pour les fréquences)
Si $(X_n)_{n\geq 1}$ *est une suite de v.a. de Bernoulli indépendantes de même paramètre* p*, alors :*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{p.s.} p.$$

Preuve : Comme précédemment, nous notons :

$$S_n = \sum_{i=1}^{n} X_i \quad \text{et} \quad M_n = \frac{S_n}{n}.$$

[5] Les sondages ordinaires sont faits sur des échantillons de 500 ou 1 000 personnes. Pour les élections présidentielles, les instituts interrogent des échantillons de 5 000 personnes. La petite étude ci-dessus montre que pour gagner une décimale sur la précision du sondage (i.e. diviser par 10 la longueur de l'intervalle de confiance), il faut multiplier la taille de l'échantillon et donc le coût du sondage par 100...

Les deux ingrédients principaux de la démonstration sont :

- L'écriture de l'événement $\{M_n$ converge vers $p\}$ à l'aide d'opérations ensemblistes dénombrables sur les événements $\{|M_n - p| \geq \varepsilon\}$ dont on sait majorer les probabilités.
- L'obtention d'une vitesse de convergence vers 0 de ces mêmes probabilités suffisante pour que :

$$\sum_{n=1}^{+\infty} P(|M_n - p| \geq \varepsilon) < +\infty. \tag{6.7}$$

Remarquons que l'inégalité de Tchebycheff est ici trop faible puisqu'elle nous donne seulement une vitesse en $O(n^{-1})$. En fait, on peut obtenir une vitesse de convergence exponentielle grâce à l'inégalité suivante de Bernstein :

$$P\left(|M_n - p| \geq \varepsilon\right) \leq 2\exp(-2n\varepsilon^2). \tag{6.8}$$

Nous admettons provisoirement cette inégalité dont une preuve est proposée à l'exercice 6.7. A partir de maintenant, la démonstration se développe en 7 « pas » élémentaires.

1^er^ pas : On rappelle la traduction automatique des quantificateurs. Si I est un ensemble quelconque d'indices, $(\mathcal{P}_i)$ une propriété dépendant de l'indice i et A_i l'ensemble des $\omega \in \Omega$ vérifiant $(\mathcal{P}_i)$, on a :

$$\begin{aligned} \{\omega \in \Omega,\ \forall i \in I,\ \omega \text{ vérifie } (\mathcal{P}_i)\} &= \bigcap_{i\in I} A_i \\ \{\omega \in \Omega,\ \exists i = i(\omega) \in I,\ \omega \text{ vérifie } (\mathcal{P}_i)\} &= \bigcup_{i\in I} A_i \end{aligned}$$

Ainsi le quantificateur $\forall$ peut toujours se traduire par une intersection et le quantificateur $\exists$ par une réunion.

2^e^ pas : Considérons l'ensemble :

$$C = \{\omega \in \Omega,\ \lim_{n\to+\infty} M_n(\omega) = p\}.$$

On peut exprimer C à l'aide des événements $\{|M_n - p| < \varepsilon\}$ en écrivant la définition de la limite :

$$\omega \in C \Leftrightarrow \forall \varepsilon > 0,\ \exists k = k(\omega, \varepsilon),\ \forall n \geq k, \quad |M_n(\omega) - p| < \varepsilon, \tag{6.9}$$

et en appliquant la règle de traduction automatique des quantificateurs :

$$C = \bigcap_{\varepsilon>0} \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} \{|M_n - p| < \varepsilon\}.$$

L'inconvénient de cette décomposition est que le « $\varepsilon > 0$ » dans la première intersection est une indexation par l'ensemble $I =]0, +\infty[$ qui n'est pas dénombrable. On ne peut donc pas appliquer les propriétés de σ-additivité ou de continuité monotone séquentielle à ce stade.
3^e pas : Il est facile de remédier à cet inconvénient : il suffit de *discrétiser* le ε dans la définition de la limite. On sait qu'on obtient une définition équivalente remplaçant dans (6.9) le « $\forall\varepsilon > 0$ » par « $\forall\varepsilon_j$ » où $(\varepsilon_j)_{j\in\mathbb{N}}$ est une suite strictement décroissante de réels tendant vers 0. On peut choisir par exemple $\varepsilon_j = 10^{-j}$. En appliquant à nouveau la traduction des quantificateurs, nous obtenons :

$$C = \bigcap_{j\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} \{|M_n - p| < \varepsilon_j\}.$$

Remarquons au passage que, sous cette forme, il est clair que l'*ensemble* C est en fait un *événement*, c'est-à-dire un membre de la famille $\mathcal{F}$ de parties de Ω sur laquelle est définie la fonction d'ensemble P. En effet, M_n étant une variable aléatoire, les $\{|M_n - p| < \varepsilon_j\}$ sont des événements et C s'obtient par des opérations ensemblistes dénombrables sur ces événements. Il est donc légitime de parler de la probabilité de C. Nous allons montrer que $P(C) = 1$.
4^e pas : Nous venons de passer d'une infinité non dénombrable de ε à une *suite* (ε_j). Le lemme suivant va nous permettre de travailler avec *une seule* valeur de ε.

Lemme 6.6 *Si $(A_j)_{j\in\mathbb{N}}$ est une suite d'événements ayant chacun une probabilité 1, alors leur intersection a aussi une probabilité 1.*

Preuve : Par passage au complémentaire, il suffit de prouver que la réunion des A_j^c a une probabilité nulle. Or :

$$0 \leq P\Big(\mathop{\cup}_{j\in\mathbb{N}} A_j^c\Big) \leq \sum_{j\in\mathbb{N}} P(A_j^c) = 0,$$

puisque chaque $P(A_j^c)$ est nul par hypothèse. Nous avons utilisé ici la propriété $7(c)$ de la proposition 1.2 pour *majorer* la probabilité d'une réunion dénombrable d'événements (pas forcément disjoints). ■

Si l'on prouve que pour chaque $\varepsilon > 0$ fixé, $P(C_\varepsilon) = 1$ où

$$C_\varepsilon = \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} \{|M_n - p| < \varepsilon\},$$

il suffira d'appliquer le lemme avec $A_j = C_{\varepsilon_j}$ pour obtenir $P(C) = 1$.
5^e pas : Soit donc $\varepsilon > 0$ fixé. Pour montrer que C_ε a une probabilité 1, on travaille sur son complémentaire que nous noterons B.

$$B = \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} \{|M_n - p| \geq \varepsilon\}.$$

On a :

$$B = \bigcap_{k \in \mathbb{N}} B_k \quad \text{avec} \quad B_k = \bigcup_{n \geq k} \{|M_n - p| \geq \varepsilon\}.$$

Donc B est inclus dans chaque B_k, d'où :

$$\forall k \in \mathbb{N}, \quad 0 \leq P(B) \leq P(B_k). \tag{6.10}$$

6^e^ pas : On majore $P(B_k)$ en utilisant à nouveau la proposition 1.2 7(c) :

$$0 \leq P(B_k) = P\Big(\underset{n \geq k}{\cup} \{|M_n - p| \geq \varepsilon\} \Big) \leq \sum_{n \geq k} P(|M_n - p| \geq \varepsilon).$$

D'après (6.8), ce majorant est le reste de rang k d'une série convergente. Il tend donc vers 0 quand k tend vers $+\infty$. Il en est donc de même pour $P(B_k)$.
7^e^ pas, conclusion : En passant à la limite quand k tend vers $+\infty$ dans (6.10), on en déduit $P(B) = 0$. En passant à l'événement complémentaire on a donc montré que $P(C_\varepsilon) = 1$. Comme la seule hypothèse faite sur ε pour obtenir ce résultat était $\varepsilon > 0$, on a donc $P(C_\varepsilon) = 1$ pour tout $\varepsilon > 0$. D'après le 4^e^ pas ceci entraîne $P(C) = 1$, autrement dit : M_n converge presque sûrement vers p. ■

Comme sous-produit de la démonstration que nous venons d'achever, nous avons montré au passage que la convergence en probabilité avec une vitesse suffisante implique la convergence presque sûre, plus précisément :

Théorème 6.7 (Condition suffisante de convergence p.s.)
Si $(Y_n)_{n \geq 1}$ et Y sont des variables aléatoires vérifiant :

$$\forall \varepsilon > 0, \qquad \sum_{n=1}^{+\infty} P(|Y_n - Y| > \varepsilon) < +\infty, \tag{6.11}$$

alors Y_n converge presque sûrement vers Y.

Preuve : Il suffit de remplacer $|M_n - p|$ par $|Y_n - Y|$ dans la démonstration ci-dessus. ■

6.5 Discussion

Considérons une urne contenant 10 boules numérotées de 0 à 9. La loi forte des grands nombres pour les fréquences nous dit que si l'on effectue une suite illimitée de tirages avec remise d'une boule, la fréquence d'apparition du chiffre 7 va converger vers 1/10 avec probabilité 1. Pour démontrer ce théorème, nous avons admis implicitement l'existence d'un espace probabilisé $(\Omega, \mathcal{F}, P)$ modélisant cette expérience (suite infinie de tirages avec remise). La construction mathématique rigoureuse d'un tel modèle présente une réelle difficulté qui est au coeur de la *théorie de la mesure* et relève du programme de la licence de mathématiques. Nous nous contenterons de quelques considérations élémentaires[6] sur cet espace probabilisé, utiles pour notre exploration de la loi forte des grands nombres.

L'espace Ω doit être assez « riche » pour « supporter » une suite infinie $(Y_i)_{i\geq 1}$ de v. a. indépendantes et de même loi uniforme sur $\{0,1,2,\dots,8,9\}$. La variable aléatoire Y_i s'interprète comme le numéro obtenu lors du i-ième tirage. On pose alors $X_i = \mathbf{1}_{\{Y_i=7\}}$ et $M_n = n^{-1}\sum_{i=1}^n X_i$ est la fréquence d'aparition du 7 en n tirages.

Nous allons examiner deux choix possibles pour Ω. Le premier et le plus naturel est de prendre :

$$\Omega = \{0,1,2,\dots,8,9\}^{\mathbb{N}^*}.$$

Autrement dit un élément quelconque ω de Ω est une suite $(c_i)_{i\geq 1}$ de chiffres décimaux. Le choix de la famille $\mathcal{F}$ d'événements observables est plus délicat. On ne peut pas prendre l'ensemble de toutes les parties de Ω car on ne pourrait pas attribuer une probabilité à chacune de ces parties de façon compatible avec ce que l'on sait déjà sur les tirages finis. Il est clair que $\mathcal{F}$ doit contenir les événements dont la réalisation ne dépend que d'un nombre fini de tirages (c'est bien le cas des événements du type $\{|M_n - p| > \varepsilon\}$ auxquels on sait attribuer une probabilité (au moins théoriquement puisque l'on sait écrire une formule donnant $P(n(p-\varepsilon) < S_n < n(p+\varepsilon))$ à l'aide de la loi binomiale). On prend pour $\mathcal{F}$ la plus petite famille d'événements observables[7] parmi celles qui contiennent les événements dont la réalisation ne dépend que d'un nombre fini d'épreuves. Pour définir la fonction d'ensemble P sur $\mathcal{F}$, on utilise un théorème de prolongement de la théorie de la mesure. On peut alors voir qu'avec ce modèle, *chaque événement élémentaire ω doit avoir une probabilité nulle*. En effet, fixons $\omega_0 = (u_1, u_2, \dots, u_n, \dots) \in \Omega$. On a :

$$\forall n \geq 1, \qquad \{\omega_0\} \subset \{Y_1 = u_1\} \cap \{Y_2 = u_2\} \cap \cdots \cap \{Y_n = u_n\},$$

d'où

$$P(\{\omega_0\}) \leq P(\bigcap_{i=1}^{n} \{Y_i = u_i\}) = \prod_{i=1}^{n} P(Y_i = u_i) = \Big(\frac{1}{10}\Big)^n,$$

en utilisant la nécessaire indépendance des Y_i. Ainsi :

$$\forall n \geq 1, \qquad 0 \leq P(\{\omega_0\}) \leq 10^{-n}.$$

En faisant tendre n vers l'infini on en déduit $P(\{\omega_0\}) = 0$. Ce raisonnement est valable pour tout ω_0 de Ω.

[6] Tout est relatif...

[7] i.e. vérifiant les trois conditions données page 5.

Notons que la nullité de $P(\{\omega\})$ pour tout $\omega \in \Omega$ ne contredit pas l'égalité $P(\Omega) = 1$. En effet on n'a pas le droit d'écrire ici « $P(\Omega) = \sum_{\omega\in\Omega} P(\{\omega\})$ » car l'ensemble d'indexation Ω *n'est pas* dénombrable (il est en bijection avec l'intervalle $[0,1]$ de $\mathbb{R}$).

Si E est un événement dénombrable, les événements élémentaires qui le composent peuvent être indexés par $\mathbb{N}$: $E = \{\omega_0, \omega_1, \dots, \omega_n, \dots\}$ et $P(E) = \sum_{n\in\mathbb{N}} P(\{\omega_n\}) = 0$. Ceci est valable a fortiori pour les événements finis.

Donc si un événement a une probabilité non nulle dans ce modèle, il est nécessairement composé d'une infinité non dénombrable d'événements élémentaires. La réciproque est fausse. Considérons en effet l'événement B défini comme l'obtention à chacun des tirages des seuls chiffres 0 ou 1. Dans notre modèle B est l'ensemble des suites de 0 et de 1, il n'est pas dénombrable (puisqu'en bijection avec $[0,1]$). Par ailleurs :

$$B = \bigcap_{i\in\mathbb{N}^*} \{Y_i = 0 \text{ ou } 1\}.$$

On a donc pour tout $n \geq 1$, $B \subset B_n = \bigcap_{i=1}^{n} \{Y_i = 0 \text{ ou } 1\}$, d'où

$$\forall n \geq 1, \quad 0 \leq P(B) \leq P(B_n) = \left(\frac{2}{10}\right)^n.$$

En faisant tendre n vers l'infini, on en déduit $P(B) = 0$. Notons d'autre part que si $\omega \in B$, ω ne contient aucun « 7 » parmi ses termes donc $M_n(\omega) = 0$ et B est inclus dans l'événement $\{M_n \to 0\}$ (ce qui prouve d'une autre façon que $P(B) = 0$ grâce à la loi forte des grands nombres). Ainsi le complémentaire de l'événement de probabilité 1 $\{M_n \to 1/10\}$ contient l'événement B et est donc lui même infini non dénombrable.

La situation est même encore plus surprenante : on peut faire converger $M_n(\omega)$ vers n'importe quel rationnel r fixé de $[0,1]$ et ce, pour tous les ω d'un événement C_r non dénombrable et de probabilité nulle (si $r \neq 1/10$). Voici comment faire. On pose $r = k/l$, $k \in \mathbb{N}$, $l \in \mathbb{N}^*$ et on définit C_r comme l'ensemble des suites de la forme :

$$\omega = (\underbrace{7,\dots,7}_{k}, \underbrace{u_{k+1},\dots,u_l}_{l-k}, \underbrace{7,\dots,7}_{k}, \underbrace{u_{l+k+1},\dots,u_{2l}}_{l-k}, \underbrace{7,\dots,7}_{k}, \dots\dots)$$

en répétant indéfiniment l'alternance de blocs de k chiffres 7 consécutifs et des blocs de $l-k$ chiffres u_i pouvant prendre seulement les valeurs 0 ou 1. Il est immédiat de vérifier que la fréquence des 7 dans une telle suite converge vers k/l, donc $C_r \subset \{M_n \to r\}$. Il est aussi clair que C_r est en bijection avec $\{0,1\}^{\mathbb{N}^*}$ (la bijection s'obtient en effaçant les 7 et sa réciproque en intercalant des blocs de k chiffres 7 consécutifs tous les $l-k$ chiffres binaires).

En adaptant ce procédé, on peut faire converger $M_n(\omega)$ vers *n'importe quel réel* x de $[0,1]$ sur un événement C_x non dénombrable et de probabilité nulle si $x \neq 1/10$ (exercice).

On peut aussi construire des événements non dénombrables et de probabilité nulle sur lesquels M_n ne converge vers aucune limite. A titre d'exemple voici comment construire un événement E tel que $\forall\omega \in E$:

$$\liminf_{n\to+\infty} M_n(\omega) = 0, \quad \text{et} \quad \limsup_{n\to+\infty} M_n(\omega) = 1. \tag{6.12}$$

Commençons par construire une suite particulière $\omega_0 = (c_i)_{i\geq 1}$ vérifiant (6.12) :

$$\omega_0 = (\underbrace{7,7}_{2},\underbrace{8,8,8,8}_{2^2},\underbrace{7,\dots,7}_{6^2},\underbrace{8,\dots,8}_{42^2},\underbrace{7,\dots\dots,7}_{(42+42^2)^2},\dots\dots).$$

et ainsi de suite en alternant indéfiniment des bloc de 7 consécutifs et de 8 consécutifs. La longueur de chaque bloc est le carré de la somme des longueurs de tous les blocs précédents. Avec cette construction, l'indice du dernier chiffre de chaque bloc est un entier de la forme $m + m^2$. A chaque étape, le dernier bloc placé écrase quasiment tout le passé et ainsi $M_n(\omega_0)$ va osciller indéfiniment entre 0 et 1. Plus précisément, si le bloc considéré se termine par un 8, il contient au moins m^2 chiffres 8 et donc au plus m chiffres 7 donc $M_{m^2+m}(\omega_0) \leq m/(m+m^2)$ et ce majorant tend vers 0 quand m tend vers $+\infty$. Si le bloc finit sur un 7, il contient au moins m^2 chiffres 7, donc $M_{m^2+m}(\omega_0) \geq (m^2/(m+m^2)$ et ce minorant tend vers 1 quand m tend vers $+\infty$. On a ainsi pu extraire de la suite $\big(M_n(\omega_0)\big)_{n\in\mathbb{N}^*}$ une sous suite convergeant vers 0 et une autre convergeant vers 1. Comme $0 \leq M_n(\omega_0) \leq 1$ pour tout n, on en déduit que ω_0 vérifie (6.12).

Pour obtenir une infinité non dénombrable de suites ω ayant la même propriété, il suffit de modifier légèrement la construction de ω_0 en :

$$\omega = (\underbrace{7,*}_{2},\underbrace{8,8,8,*}_{2^2},\underbrace{7,\dots,7,*}_{6^2},\underbrace{8,\dots,8,*}_{42^2},\underbrace{7,\dots\dots,7,*}_{(42+42^2)^2},\dots\dots)$$

où le dernier chiffre de chaque bloc de ω_0 est remplacé au choix par un 0 ou un 1 (représenté par l'astérisque ci-dessus).

En travaillant encore un peu, on pourrait de même montrer pour tout couple de réels (a,b) de $[0,1]$ tels que $a < b$, l'existence d'événements $E_{a,b}$ non dénombrables et de probabilité nulle sur lesquels M_n a pour limite inférieure a et pour limite supérieure b...

Tous ces exemples montrent que l'événement $\{M_n$ *ne converge pas vers* $1/10\}$ a une structure très complexe. Ainsi l'aspect naturel et intuitif de la loi forte des grands nombres pour les fréquences masque un résultat plus profond qu'il n'y paraît. Le *presque sûrement* qui figure dans l'énoncé de cette loi n'est pas une finasserie artificielle de puriste mais est bien inhérent au problème étudié.

On est naturellement tenté d'interpréter les résultats précédents du point de vue de la théorie des nombres en considérant les suites de chiffres décimaux sur lesquelles nous venons de travailler comme des développements décimaux illimités de nombres réels de $[0,1]$. Notre second modèle sera donc $(\Omega', \mathcal{F}', P')$ où :

$$\Omega' = [0,1]$$

et $\mathcal{F}'$ et P' restent à définir.

Cependant il se présente ici une difficulté qui fait que ce nouveau modèle ne se réduit pas à une traduction automatique du précédent. Si $(c_i)_{i\geq 1}$ est une suite de chiffres décimaux, la série :

$$\sum_{i=1}^{+\infty} \frac{c_i}{10^i} \tag{6.13}$$

converge et sa somme est un réel x de $[0,1]$ que l'on peut noter

$$x = 0.c_1c_2\dots c_i\dots = \sum_{i=1}^{+\infty} \frac{c_i}{10^i}.$$

Réciproquement, tout réel de $[0,1]$ admet un développement décimal du type (6.13). Ce développement est unique lorsque x n'est pas un *nombre décimal* (i.e. x n'est pas de la forme $k/10^n$, $k \in \mathbb{N}$, $n \in \mathbb{N}^*$). Par contre si x est décimal, il possède deux développements décimaux distincts[8]. Ceci provient de la sommation de série géométrique suivante :

$$\forall n \geq 1, \quad \sum_{i=n}^{+\infty} \frac{9}{10^i} = \frac{9}{10^n} \sum_{j=0}^{+\infty} \frac{1}{10^j} = \frac{9}{10^n} \frac{1}{\left(1 - \frac{1}{10}\right)} = \frac{1}{10^{n-1}}. \tag{6.14}$$

Cette relation permet de voir que si un développement décimal illimité ne comporte plus que des 9 à partir d'un certain rang n (le $(n-1)$-ème chiffre n'étant pas un 9), on ne change pas la somme de la série en remplaçant tous ces 9 par des 0 et en augmentant d'une unité le $(n-1)$-ème chiffre. On a ainsi la propagation d'une *retenue* depuis l'infini. Par exemple :

$$0.5972999999\dots = 0.5973000000\dots = \frac{5973}{10^4}$$

(il ne s'agit pas d'une égalité approchée, mais d'une égalité rigoureuse, les points de suspension représentant la répétition indéfinie du chiffre 9 ou 0 respectivement). Le développement ne comportant que des 9 à partir d'un certain rang est appelé *développement décimal impropre*, celui ne comportant que des 0 est appelé *développement décimal propre*.

[8] Voir exercice 6.11.

En revenant aux tirages illimités dans notre urne à dix boules, on voit que si l'on choisit $\Omega' = [0,1]$, les deux suites de résultats qui correspondent à un même réel *décimal* seront représentées par le même réel ω. Par exemple $(5,9,7,2,9,9,9,\dots)$ et $(5,9,7,3,0,0,0,\dots)$ seront représentées par l'événement élémentaire $\omega = 5973/1\,0000$.

Pour surmonter cette difficulté, nous « dédoublons » la suite $(Y_i)_{i\geq 1}$. Pour tout $i \geq 1$, on définit les deux variables aléatoires Y_i et Y_i' comme suit. Si $\omega \in [0,1]$ n'est pas décimal, $Y_i(\omega) = Y_i'(\omega)$ est le i-ème chiffre décimal de l'unique développement décimal de ω. Si ω est un décimal de $[0,1]$, $Y_i(\omega)$ est le i-ème chiffre de son développement propre, $Y_i'(\omega)$ le i-ème chiffre décimal de son développement impropre. On requiert, comme dans le premier modèle que chacune de ces deux suites soit indépendante et que chacune des variables Y_i et Y_i' suive la loi uniforme sur $\{0,1,\dots,8,9\}$. Ceci permet de montrer que chaque événement élémentaire ω doit avoir une probabilité P' nulle. D'autre part, Y_i et Y_i' diffèrent seulement sur l'ensemble D des décimaux de $[0,1]$ qui est dénombrable (voir exercice 3.15), donc de probabilité P' nulle. Ainsi les deux suites $(Y_i)_{i\geq 1}$ et $(Y_i')_{i\geq 1}$ sont égales P'-presque sûrement. Il est donc quand même possible d'interpréter la suite illimitée de tirages dans l'urne comme le choix aléatoire d'un réel ω de $[0,1]$ suivant la loi de probabilité P'.

On peut maintenant examiner les conséquences de notre cahier des charges (les conditions sur les suites de v.a. $(Y_i)_{i\geq 1}$ et $(Y_i')_{i\geq 1}$) sur la construction de $(\mathcal{F}', P')$. La condition d'indépendance de la suite $(Y_i)_{i\geq 1}$ avec même loi uniforme sur $\{0,1,\dots,8,9\}$ pour tout Y_i peut s'écrire comme suit. Pour tout $n \geq 1$, et tout n-uplet $(c_1,\dots,c_n)$ de chiffres décimaux,

$$P'(Y_1 = c_1, Y_2 = c_2, \dots, Y_n = c_n) = \prod_{i=1}^{n} P'(Y_i = c_i) = \frac{1}{10^n}.$$

En notant que l'on a exclu les développement impropres dans la définition des Y_i, on a l'équivalence :

$$Y_1(\omega) = c_1,\ Y_2(\omega) = c_2, \dots,\ Y_n(\omega) = c_n \ \Leftrightarrow\ \omega \in [\alpha_n, \alpha_n + 10^{-n}[,$$

où l'on a posé : $\alpha_n = c_1 10^{-1} + \cdots + c_n 10^{-n}$. Lorsque le n-uplet $(c_1,\dots,c_n)$ prend toutes les valeurs possibles (à n fixé), α_n décrit exactement l'ensemble des décimaux pouvant s'écrire sous la forme $k10^{-n}$. La condition sur la suite $(Y_i)_{i\geq 1}$ peut donc se traduire par :

$$\forall n \geq 1,\ \forall k = 0,1,\dots,10^n - 1, \qquad P'\Big(\Big[\frac{k}{10^n}, \frac{k+1}{10^n}\Big[\Big) = \frac{1}{10^n}.$$

L'utilisation de la suite (Y_i') à la place de (Y_i) dans le raisonnement ci-dessus nous aurait donné la même conclusion mais avec des intervalles ouverts à gauche et fermés à droite. Notons que dans les deux cas la probabilité P' de l'intervalle concerné est égale à sa *longueur*. On peut aussi utiliser chacun de ces deux résultats pour redémontrer que la probabilité d'un événement élémentaire ω est forcément nulle. Finalement, grâce à l'additivité de P' on en déduit facilement que la condition sur la suite (Y_i) équivaut à :

$$\forall a, b \in [0,1] \cap D \quad (a < b), \qquad P'([a,b]) = b - a \tag{6.15}$$

(ou à chacune des conditions obtenues avec $[a,b[$, $]a,b]$ ou $]a,b[$). Par continuité monotone de P', on en déduit que (6.15) s'étend au cas de réels a, $b > a$ quelconques de $[0,1]$: il suffit de considérer deux suites de décimaux $a_n \uparrow a$ et $b_n \downarrow b$ et de noter que $[a,b] = \cap_{n\geq 1}[a_n, b_n]$ (détails laissés en exercice).

Nous voyons maintenant que le problème de la construction de $(\mathcal{F}', P')$ est exactement celui de la construction d'une fonction d'ensemble $\sigma-$ *additive* prolongeant la fonction *longueur d'un intervalle*. Ce problème est celui de la construction de la *mesure* de Lebesgue. On peut le résoudre en prenant pour $\mathcal{F}'$ la plus petite famille d'événements observables (toujours au sens de la page 5) contenant les intervalles. On arrive ainsi à définir la longueur ou mesure de Lebesgue des sous ensembles de $[0,1]$ qui sont dans $\mathcal{F}'$. Si un tel sous ensemble est de la forme $B = \cup_{i\geq 1}]a_i, b_i[$ où les suites (a_i) et (b_i) vérifient pour tout n : $0 \leq a_n < b_n \leq a_{n+1} < b_{n+1} \leq 1$, alors B est une réunion disjointe d'intervalles et sa probabilité P' ou longueur est évidemment la série de terme général la longueur de $]a_i, b_i[$. Malheureusement, tous les éléments de la famille $\mathcal{F}'$ sont loin d'avoir une structure aussi simple et le calcul explicite de leur longueur n'est pas toujours possible (on sait qu'elle existe et on connaît ses propriétés). Nous connaissons déjà un exemple d'élément de $\mathcal{F}'$ qui ne peut pas s'écrire comme réunion dénombrable d'intervalles disjoints, c'est l'événement $C_7 = \{$*convergence de la fréquence du chiffre* 7 *vers* $1/10\}$. En effet par densité des décimaux, tout intervalle contient au moins un décimal (en fait une infinité) et si ω est décimal, $Y_i(\omega) = 0$ à partir d'un certain rang (de même $Y'_i(\omega) = 9$) par conséquent $M_n(\omega)$ converge vers 0 donc $\omega \notin C_7$. Ainsi C_7 ne peut s'écrire comme réunion dénombrable d'intervalles disjoints. Nous savons pourtant calculer sa *longueur* par la loi forte des grands nombres : elle vaut 1.

Dans toute cette section nous nous sommes intéressés à la fréquence d'apparition du 7. Bien sûr ce chiffre n'a été choisi que pour fixer les idées et n'importe quel autre chiffre décimal aurait tout aussi bien fait l'affaire. Pour généraliser un peu définissons $M_{n,j}$ comme la fréquence d'apparition du chiffre j $(j \in \{0, 1, \dots, 8, 9\})$ au cours des n premiers tirages. Notons de même C_j l'événement $\{M_{n,j}$ *converge vers* $1/10\}$. Par la loi forte des grands nombres, chaque C_j a une *longueur* (i.e. une probabilité P') égale à 1. Par le lemme 6.6, l'intersection de ces dix ensembles a aussi une longueur 1.

Convenons d'appeler *nombre normal* tout réel de $[0,1]$ tel que la fréquence de chacun des 10 chiffres décimaux $0, 1, \dots 9$ dans le développement décimal illimité de ce nombre converge vers $1/10$. Nous avons ainsi obtenu un résultat de théorie des nombres qui s'énonce ainsi : *l'ensemble de tous les nombres normaux* *de* $[0, 1]$ *a pour longueur* 1 (on dit aussi *presque tout* nombre de $[0, 1]$ est normal). Ce résultat est dû à Borel. On pourrait maintenant traduire tous les exemples étudiés dans le cadre du premier modèle et voir ainsi que l'ensemble de longueur nulle des nombres non normaux a une structure très complexe. Là encore, le théorème de Borel est plus profond qu'il n'y paraît à première vue...

6.6 Exercices

Ex 6.1. Montrer que si X_n converge presque sûrement vers X et Y_n converge presque sûrement vers Y, il existe un événement Ω_0 de probabilité 1 tel que pour tout $\omega \in \Omega_0$, le *couple* $(X_n(\omega), Y_n(\omega))$ converge vers $(X(\omega), Y(\omega))$. En déduire que pour toute fonction continue $g : \mathbb{R}^2 \to \mathbb{R}$, la suite de v.a. $g(X_n, Y_n)$ converge presque sûrement vers $g(X, Y)$. Applications : $g(x, y) = x + y$, $g(x, y) = xy$,...

Ex 6.2. *Convergence en probabilité et addition*

1) Soient U et V deux variables aléatoires positives, justifier l'inégalité :

$$\forall \varepsilon > 0, \quad P(U + V \geq \varepsilon) \leq P(U \geq \varepsilon/2) + P(V \geq \varepsilon/2).$$

2) En déduire que si X_n et Y_n convergent en probabilité vers respectivement X et Y, alors $X_n + Y_n$ converge en probabilité vers $X + Y$.

Ex 6.3. *La convergence p.s. implique la convergence en probabilité*
On suppose que la suite de v.a. $(X_n)_{n\geq 1}$ converge presque sûrement vers la v.a. X, ce qui signifie que l'événement :

$$\Omega' = \{\omega \in \Omega, \lim_{n\to+\infty} X_n(\omega) = X(\omega)\}$$

a pour probabilité 1. On fixe $\varepsilon > 0$ et on définit :

$$\Omega'_\varepsilon = \{\omega \in \Omega, \ \exists m_0 = m_0(\omega), \ \forall n \geq m_0, \ |X_n(\omega) - X(\omega)| < \varepsilon\}.$$

1) Vérifier que $P(\Omega'_\varepsilon) = 1$.

2) Exprimer Ω'_ε à l'aide d'opérations ensemblistes dénombrables sur les événements $\{|X_n - X| < \varepsilon\}$.

3) Pour tout $k \geq 1$, on définit les événements

$$A_k = \{\omega \in \Omega, \ \forall n \geq k, \ |X_n(\omega) - X(\omega)| < \varepsilon\}.$$

Vérifier que la suite $(A_k)_{k\geq 1}$ est croissante pour l'inclusion et identifier sa réunion. En déduire :

$$\forall \eta > 0, \ \exists k_0, \quad P(A_{k_0}) > 1 - \eta,$$

puis :

$$\forall n \geq k_0, \quad P(|X_n - X| < \varepsilon) > 1 - \eta.$$

4) Conclure.

Ex 6.4. *La convergence en probabilité n'implique pas la convergence p.s.* ↓6.12
Considérons le jeu de pile ou face infini avec $\Omega = \{0,1\}^{\mathbb{N}^*}$ où l'on code 0 pour face et 1 pour pile. On construit une suite de variables aléatoires $(T_n)_{n\geq 1}$ comme suit

$$\begin{array}{lcl} T_1 & = & 1 \\ T_2 & = & 1 \text{ si face } (\mathtt{f}) \text{ au premier coup, } 0 \text{ sinon} \\ T_3 & = & 1 \text{ si pile } (\mathtt{p}) \text{ au premier coup, } 0 \text{ sinon} \\ T_4 & = & 1 \text{ si } \mathtt{ff}, 0 \text{ sinon} \\ T_5 & = & 1 \text{ si } \mathtt{fp}, 0 \text{ sinon} \\ T_6 & = & 1 \text{ si } \mathtt{pf}, 0 \text{ sinon} \\ T_7 & = & 1 \text{ si } \mathtt{pp}, 0 \text{ sinon} \\ T_8 & = & 1 \text{ si } \mathtt{fff}, 0 \text{ sinon, etc.} \end{array}$$

Pour une définition plus formalisée, si :

$$n = 2^{k_n} + \sum_{i=1}^{k_n} a_i 2^{k_n - i} \qquad a_i \in \{0,1\}$$

est l'écriture de n en base 2, on a en notant X_i la variable indicatrice de l'événement « pile au i-ème coup » :

$$T_n = \prod_{i=1}^{k_n} \big(a_i X_i + (1-a_i)(1-X_i)\big)$$

1) En utilisant l'indépendance des X_i, montrer que :

$$P(T_n = 1) = \frac{1}{2^{k_n}}$$

En déduire que la suite (T_n) converge en probabilité vers 0.

2) Par ailleurs soit $\omega = (u_i)_{i\geq 1}$ *fixé* dans $\{0,1\}^{\mathbb{N}^*}$. Vérifier que la suite $(T_n(\omega))$ prend une infinité de fois la valeur 1. *Indication :* Pour $k \in \mathbb{N}^*$, poser

$$n_k = n_k(\omega) = 2^k + \sum_{i=1}^{k} u_i 2^{k-i}$$

et calculer $T_{n_k}(\omega)$.

3) Montrer que $T_n(\omega)$ prend aussi une infinité de fois la valeur 0.

4) En déduire que la suite de réels $(T_n(\omega))$ ne converge pour *aucun* $\omega \in \Omega$. A fortiori T_n ne converge pas presque sûrement.

Ex 6.5. Reprendre les calculs des exemples 6.2 et 6.3 avec un niveau de confiance de 99%.

Ex 6.6. *Une inégalité de Hoeffding*
Soit Z une variable aléatoire réelle vérifiant : $a \leq Z \leq b$ presque sûrement. On se propose de montrer l'inégalité suivante :

$$\mathbb{E}\left(\exp(Z - \mathbb{E}\, Z)\right) \leq \exp\left(\frac{(b-a)^2}{8}\right)$$

1) On rappelle qu'une fonction g est dite *convexe* sur un intervalle I, si pour tous $a, b \in I$ et tous $u, v \in [0,1]$ tels que $u + v = 1$:

$$g(ua + vb) \leq ug(a) + vg(b).$$

Une condition suffisante pour qu'une fonction soit convexe sur I est qu'elle ait une dérivée seconde positive sur I. Soit $t > 0$ et $d \in \mathbb{R}$. On définit la fonction g par $g(x) = \exp\bigl(t(x-d)\bigr)$. Vérifier sa convexité et en déduire :

$$\forall x \in [a,b], \mathrm{e}^{t(x-d)} \leq \frac{b-x}{b-a}\mathrm{e}^{t(a-d)} + \frac{x-a}{b-a}\mathrm{e}^{t(b-d)} \tag{6.16}$$

2) On pose $d = \mathbb{E}\, Z$. Déduire de la question précédente l'inégalité :

$$\mathbb{E}\left(\exp(t(Z-d))\right) \leq \frac{b-d}{b-a}\mathrm{e}^{t(a-d)} + \frac{d-a}{b-a}\mathrm{e}^{t(b-d)}$$

3) On pose :

$$f(t) = \ln\left[\frac{b-d}{b-a}\mathrm{e}^{ta} + \frac{d-a}{b-a}\mathrm{e}^{tb}\right] - td$$

Montrer que pour tout $t \in [0,1]$, $f''(t) \leq \frac{1}{4}(b-a)^2$. Conclure.

Ex 6.7. *Preuve de l'inégalité de Bernstein*
En utilisant l'inégalité de Hoeffding, on se propose de démontrer le théorème suivant

Théorème 6.8 (Bernstein) *Soient $X_1, \dots, X_n$ n variables aléatoires réelles indépendantes. On suppose que pour chaque $i = 1, \dots, n$, il existe des constantes a_i et b_i telles que presque sûrement, $a_i \leq X_i \leq b_i$. Alors on a :*

$$\forall t > 0 \quad P\left(\Bigl|\sum_{i=1}^n (X_i - \mathbb{E}\, X_i)\Bigr| \geq t\right) \leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{6.17}$$

1) Montrer que pour tout $u > 0$:

$$P(S_n - \mathbb{E}\, S_n \geq t) \leq \exp\left[-ut + \frac{u^2}{8}\sum_{i=1}^n (b_i - a_i)^2\right],$$

où l'on a noté $S_n = \sum_{i=1}^n X_i$.

2) Optimiser cette inégalité en choisissant u de façon à minimiser le second membre.

3) Démontrer une inégalité analogue pour $P(S_n - \mathbb{E}\, S_n \leq -t)$ et conclure.
Commentaires : Examinons ce que donne ce théorème dans le cas où les X_i sont indépendantes et de même loi. Alors on peut prendre tous les a_i égaux à un réel a et tous les b_i égaux à b. Tous les $\mathbb{E}\, X_i$ sont égaux à $\mathbb{E}\, X_1$. En posant $t = n\varepsilon$ dans (6.17), on obtient :

$$P\left(\left|\frac{S_n}{n} - \mathbb{E}\, X_1\right| \geq \varepsilon\right) \leq 2\exp\left(\frac{-2n\varepsilon^2}{(b-a)^2}\right). \tag{6.18}$$

En particulier si les X_i sont des v.a. de Bernoulli de paramètre p, on peut prendre $a = 0$ et $b = 1$ et on a :

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) \leq 2\exp(-2n\varepsilon^2). \tag{6.19}$$

Ex 6.8. On considère une suite infinie de jets d'un dé et on appelle S_n le nombre d'apparitions du 6 au cours des n premiers lancers.

1) En quels sens peut-on dire que S_n/n converge vers $1/6$?

2) En utilisant l'inégalité de Tchebycheff, puis celle de Bernstein, majorer :

$$P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| > 0.01\right).$$

3) On note u_n le majorant donné par l'inégalité de Tchebycheff, v_n celui donné par celle de Bernstein. Comparer ces deux majorants suivant les valeurs de n. *Indication :* On pourra par exemple étudier les variations de la fonction :

$$x \mapsto x\mathrm{e}^{-bx} \qquad x \in [0, +\infty[\qquad 0 < b < 1.$$

4) Montrer en utilisant l'inégalité de Bernstein que :

$$P\left(\forall k \geq 11 \quad \left|\frac{S_k}{k} - \frac{1}{6}\right| \leq \sqrt{\frac{2\ln k}{k}}\right) \geq 0.999.$$

Donner une interprétation graphique.

Ex 6.9. Énoncer une loi forte des grands nombres pour une suite de variables aléatoires indépendantes, de même espérance et uniformément bornées ($\exists a, b$ tels que $\forall i \in \mathbb{N}^*$, $a \leq X_i \leq b$ p.s.). Quelles sont les adaptations à faire dans la preuve du théorème 6.5 pour obtenir cette généralisation ?

Ex 6.10. *Vitesse de convergence des polynômes de Bernstein* ↑ 5.23
L'inégalité (6.19) permet d'améliorer les résultats obtenus à l'exercice 5.23 sur la vitesse de convergence uniforme des polynômes de Bernstein d'une fonction continue. L'utilisation de (6.19) à la place de l'inégalité de Tchebycheff nous donne en effet la majoration :

$$\|f - \mathrm{B}_n f\|_\infty \leq \varepsilon + 4\|f\|_\infty \exp(-2n\delta^2). \tag{6.20}$$

1) On suppose f lipschitzienne. Vérifier que le choix $\varepsilon = cn^{-\beta}$ dans (6.20) donne une vitesse de convergence en $O(n^{-\beta})$ pour tout $\beta < 1/2$, mais que la même méthode ne permet pas d'obtenir la vitesse $O(n^{-1/2})$.

2) Toujours avec f lipschitzienne, comment choisir c minimal pour obtenir avec $\varepsilon = c(\ln n/n)^{1/2}$ la vitesse $O\big((\ln n/n)^{1/2}\big)$?

3) On suppose maintenant f hölderienne d'exposant α. Montrer qu'avec un choix judicieux de ε, on obtient la vitesse $O\big((\ln n/n)^{\alpha/2}\big)$.

Ex 6.11. *Développement(s) décimaux illimités d'un réel*

1) Prouver l'existence du développement décimal illimité d'un réel x de $[0,1]$. On donnera un algorithme simple permettant d'obtenir à partir de x les sommes partielles de la série du type (6.13).

2) Étudier l'unicité de ce développement. *Indication :* Supposons que :

$$\sum_{i=1}^{+\infty} \frac{c_i}{10^i} = \sum_{i=1}^{+\infty} \frac{d_i}{10^i}$$

et notons n_0 le plus grand indice pour lequel les n_0 premiers chiffres décimaux c_i et d_i sont deux à deux égaux. Si $c_1 \neq d_1$, on prend $n_0 = 0$. On peut alors supposer que $d_{n_0+1} > c_{n_0+1}$. Examiner ce que cela implique pour les chiffres c_i et d_i , $i > n_0$ en utilisant la relation (6.14).

Ex 6.12. Reprendre l'exercice 6.4 avec $\Omega = [0,1]$, $\mathcal{F}'$ et P' étant définis à partir des intervalles et de leur longueur comme dans le deuxième modèle de la section 6.5. On définit $X_i(\omega)$ comme le i-ème chiffre du développement du réel ω en base 2. On conviendra d'exclure les développements binaires impropres (i.e. n'ayant que des 1 à partir d'un certain rang). Montrer qu'avec ce modèle les v.a. T_n s'interprètent simplement comme des indicatrices d'intervalles. Interpréter graphiquement la convergence en probabilité des T_n et leur non convergence p.s.

Ex 6.13. *Caractérisation des nombres rationnels normaux*

1) On considère le nombre réel $x \in [0,1]$ dont le développement décimal illimité s'écrit :

$$x = 0.35712712712\ldots712\ldots$$

On dit que ce développement est périodique, la période étant ici la séquence de chiffres 712 qui se répète indéfiniment (à partir d'un certain rang). Montrer par un calcul de somme de série que x est un rationnel.

2) Généraliser.

3) Réciproquement, montrer que tout rationnel de $[0, 1]$ a un développement décimal périodique.

4) Caractériser par une condition sur leur période les rationnels de $[0, 1]$ qui sont des nombres normaux au sens de Borel (voir page 131).

Chapitre 7

Approximation gaussienne de la loi binomiale

Nous connaissons déjà l'approximation d'une loi binomiale $\mathcal{B}(n, p)$ par une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = np$ lorsque n est « grand » et np « petit ». Nous étudions dans ce chapitre une approximation utilisable lorsque np ne peut être considéré comme « petit ». Le résultat théorique qui justifie cette approximation est le théorème de De Moivre-Laplace qui est lui même un cas particulier du théorème central limite. Ce dernier est, avec la loi des grands nombres, certainement le plus important théorème du calcul des probabilités. L'approximation qui nous intéresse fait intervenir une famille de fonctions appelées densités gaussiennes (ou normales) liées à la célèbre *courbe en cloche* de Gauss.

7.1 La courbe en cloche

Définition 7.1 *Soit $m \in \mathbb{R}$ et $\sigma \in]0, +\infty[$. On appelle densité gaussienne ou normale $f_{m,\sigma}$ sur $\mathbb{R}$ la fonction :*

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

La fonction $f_{0,1}$ est appelée densité normale ou gaussienne standard.

Dans cette famille de fonctions, m est un paramètre de localisation ou de position (c'est la valeur où $f_{m,\sigma}$ atteint son maximum). Le paramètre σ est un paramètre d'échelle, il caractérise l'aplatissement de la courbe. Les courbes représentatives $C_{m,\sigma}$ de ces fonctions se déduisent toutes de la courbe $C_{0,1}$ par translations et changements d'échelle. On passe de $C_{0,1}$ à $C_{m,\sigma}$ par la transformation : $(x, y) \mapsto (\sigma x + m, \frac{y}{\sigma})$ et de $C_{m,\sigma}$ à $C_{0,1}$ par $(x, y) \mapsto (\frac{x-m}{\sigma}, \sigma y)$. La courbe $C_{0,1}$ (figure 7.1) est très populaire sous le nom de courbe en cloche de Gauss.

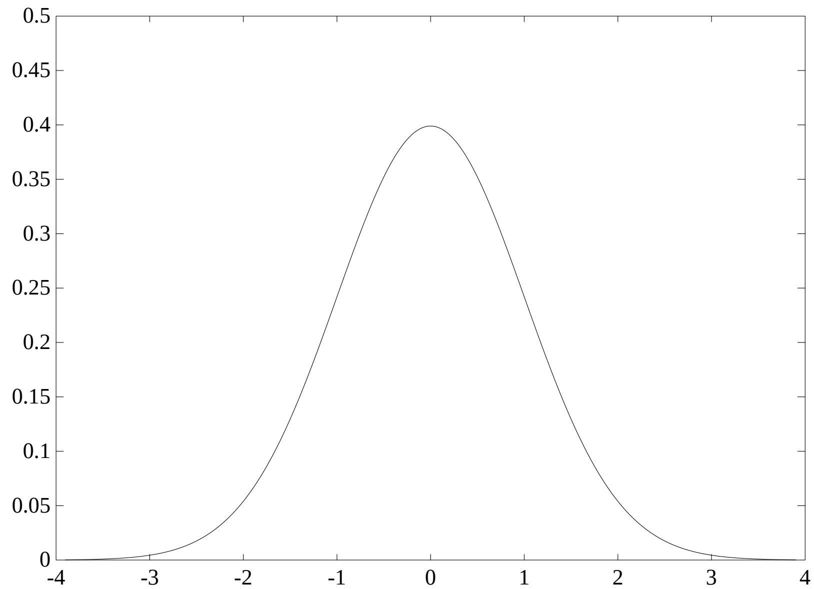


FIG. 7.1 – Densité $f_{0,1}$

Bien que $f_{0,1}(t)$ soit strictement positif pour tout réel t, la densité $f_{0,1}$ semble être à support dans $[-4, +4]$. Cela provient de la décroissance rapide de $\exp(-\frac{t^2}{2})$ quand t tend vers $+\infty$ et des limites de résolution du dessin.

L'influence des paramètres m et σ est illustrée par les figures 7.2 et 7.3.

Une propriété importante de $C_{0,1}$ est que l'aire qu'elle délimite avec l'axe des abscisses vaut 1 :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Bigl(-\frac{t^2}{2}\Bigr)\,\mathrm{d}t = 1. \tag{7.1}$$

L'exercice 7.1 présente une démonstration de cette relation. Les autres fonctions $f_{m,\sigma}$ ont aussi une intégrale généralisée sur $\mathbb{R}$ qui vaut 1. Pour le voir, soient $a < b$ deux réels quelconques. Par le changement de variable $u = (t - m)/\sigma$:

$$\int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right)\,\mathrm{d}t = \int_{a^*}^{b^*} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)\,\mathrm{d}u, \tag{7.2}$$

où les nouvelles bornes sont :

$$a^* = \frac{a-m}{\sigma} \quad \text{et} \quad b^* = \frac{b-m}{\sigma}.$$

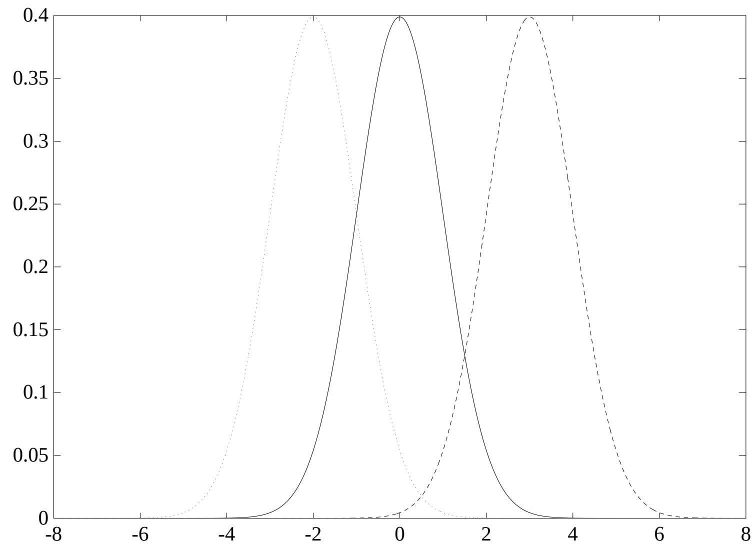


FIG. 7.2 – Densités $f_{-2,1}$, $f_{0,1}$, $f_{3,1}$

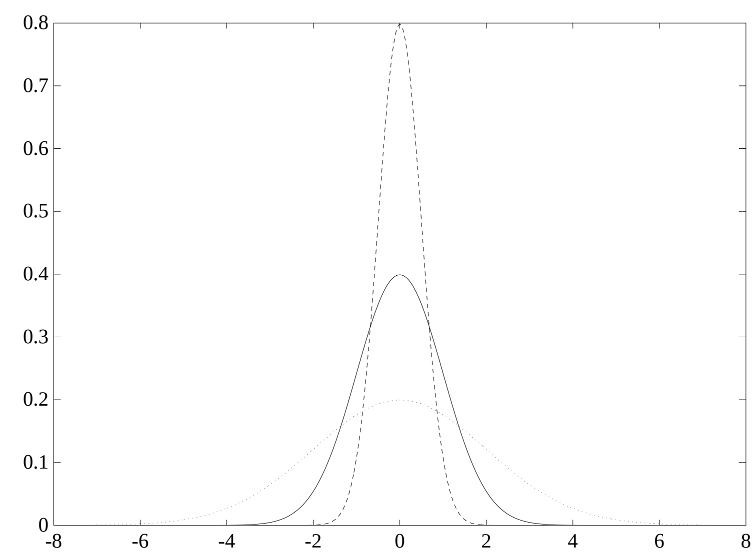


FIG. 7.3 – Densités $f_{0,1}$, $f_{0,0.5}$, $f_{0,2}$

Lorsque a et b tendent respectivement vers $-\infty$ et $+\infty$, il en est de même pour a^* et b^*. On en déduit :

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-m)^2}{2\sigma^2}\right) \mathrm{d}t = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u = 1.$$

L'aire délimitée par $C_{m,\sigma}$, l'axe des abscisses et les droites d'équation $t = a$, $t = b$ joue un rôle important dans l'approximation des probabilités binomiales $P(a < S_n \leq b)$. Par changement de variable, le calcul de cette aire se ramène à celui de l'aire correspondante pour $C_{0,1}$ avec a^* et b^* à la place de a et b. Elle peut donc s'écrire sous la forme $\Phi(b^*) - \Phi(a^*)$ où la fonction Φ est définie par :

$$\forall x \in \mathbb{R}, \qquad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t. \tag{7.3}$$

Cette intégrale ne peut s'exprimer à l'aide des fonctions usuelles. On peut en calculer une valeur approchée avec toute précision souhaitée (voir exercice 7.2). La figure 7.4 représente le graphe de Φ .

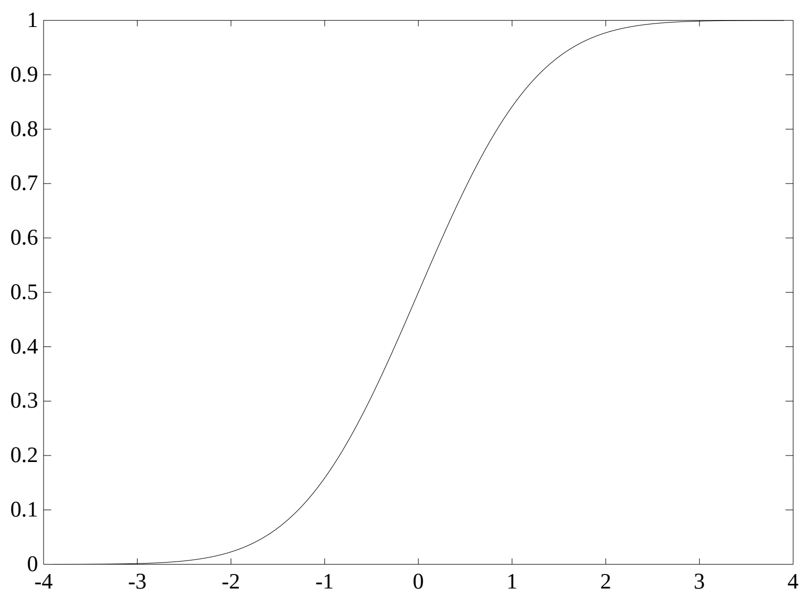


FIG. 7.4 – Fonction Φ : $x \mapsto \int_{-\infty}^{x} f_{0,1}(t)\,\mathrm{d}t$

La table en annexe donne les valeurs de $\Phi(x)$ par *pas* de 0.01 pour x compris entre 0 et 3 et quelques valeurs pour x compris entre 3 et 4.5. Pour x négatif, la parité de la densité entraîne la relation $\Phi(x) = 1 - \Phi(-x)$. Pour les « très grandes valeurs de x », (i.e. $|x| \geq 4$), on dispose du résultat suivant qui donne une évaluation de la « queue » de la loi normale :

Lemme 7.2 *Pour tout $x > 0$, on a l'encadrement :*

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) < 1 - \Phi(x) < \frac{1}{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right). \qquad (7.4)$$

Preuve : La dérivation de chacun des 3 membres A, B et C de (7.4) donne :

$$A'(x) = \left(-1 + \frac{3}{x^4}\right)\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} \qquad B'(x) = -\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} \qquad C'(x) = \left(-1 - \frac{1}{x^2}\right)\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}.$$

Il suffit alors d'intégrer sur $[x, +\infty[$ l'encadrement suivant vrai pour tout y :

$$\left(-1 - \frac{1}{y^2}\right)\frac{\mathrm{e}^{-y^2/2}}{\sqrt{2\pi}} < -\frac{\mathrm{e}^{-y^2/2}}{\sqrt{2\pi}} < \left(-1 + \frac{3}{y^4}\right)\frac{\mathrm{e}^{-y^2/2}}{\sqrt{2\pi}},$$

pour obtenir (7.4). ■

7.2 Étude graphique

Soient $X_1, \dots, X_n$ n v.a. de Bernoulli indépendantes de même paramètre p. On pose $S_n = X_1 + \cdots + X_n$. La loi de S_n est alors une binomiale $B(n,p)$:

$$P(S_n = k) = C_n^k p^k (1-p)^{n-k} \qquad (0 \leq k \leq n)$$

Les histogrammes[1] ci-dessous représentent cette loi pour différentes valeurs des paramètres n et p. Bien que l'histogramme de la loi $B(n,p)$ soit constitué théoriquement de $n+1$ rectangles, seule une partie d'entre eux est visible sur le dessin, les autres correspondant à des probabilités trop petites. Le nombre ε représente pour chaque figure un majorant de la probabilité correspondant à la réunion de ces rectangles « invisibles ». Sur chaque figure, la courbe en pointillés représente la densité $f_{m,\sigma}$ dont les paramètres sont donnés par : $m = \mathbb{E}\, S_n = np$ et $\sigma^2 = \operatorname{Var} S_n = npq$.

[1]Les rectangles de l'histogramme ont une aire proportionnelle aux nombres $P(S_n = k)$ et ont pour axes de symétrie les droites d'équation $x = k$ correspondantes.

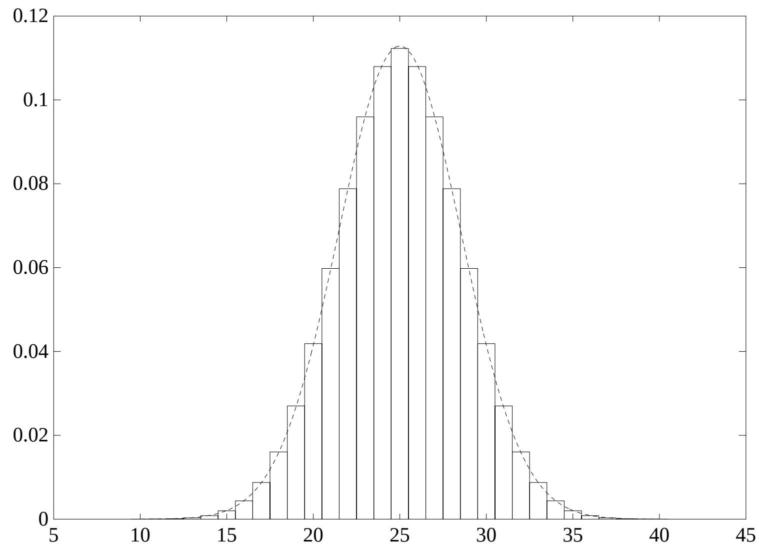


FIG. 7.5 – Binomiale $n = 50$, $p = 0.5$ ($\varepsilon = 3.1 \times 10^{-4}$)

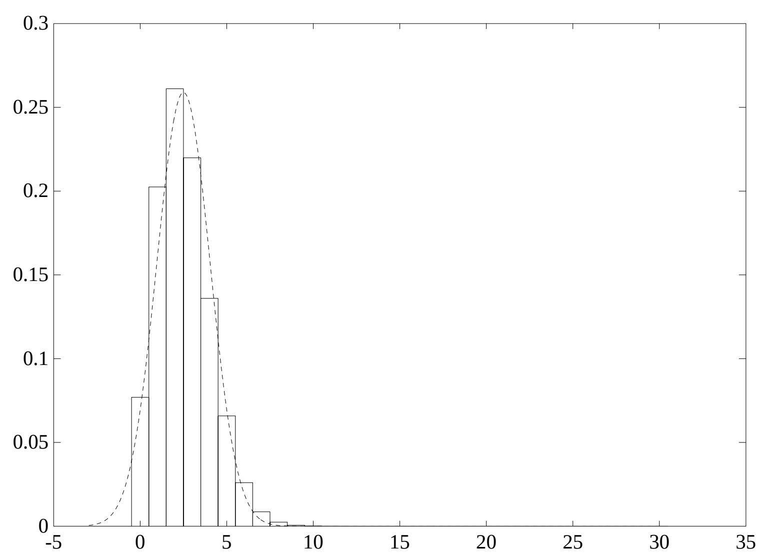


FIG. 7.6 – Binomiale $n = 50$, $p = 0.05$ ($\varepsilon = 1.6 \times 10^{-4}$)

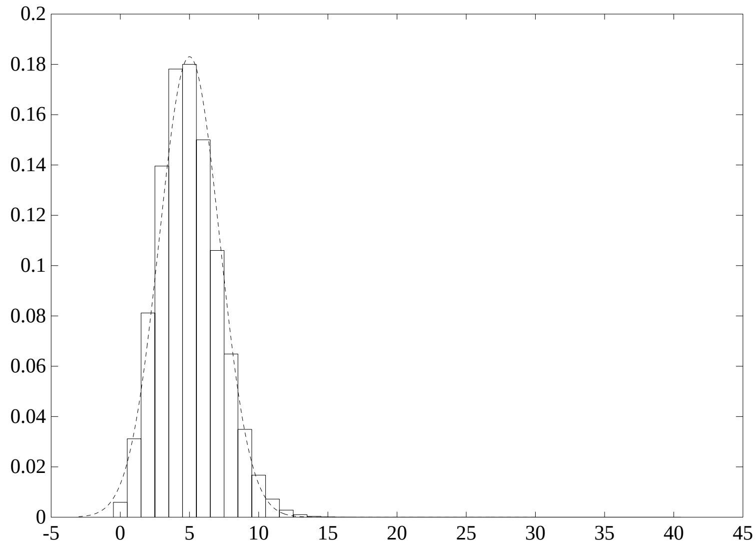


FIG. 7.7 – Binomiale $n = 100$, $p = 0.05$ ($\varepsilon = 1.4 \times 10^{-4}$)

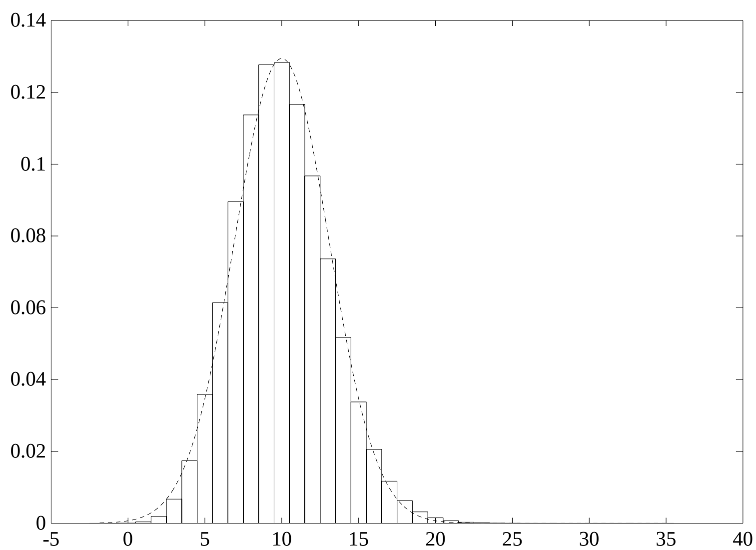


FIG. 7.8 – Binomiale $n = 200$, $p = 0.05$ ($\varepsilon = 2.3 \times 10^{-4}$)

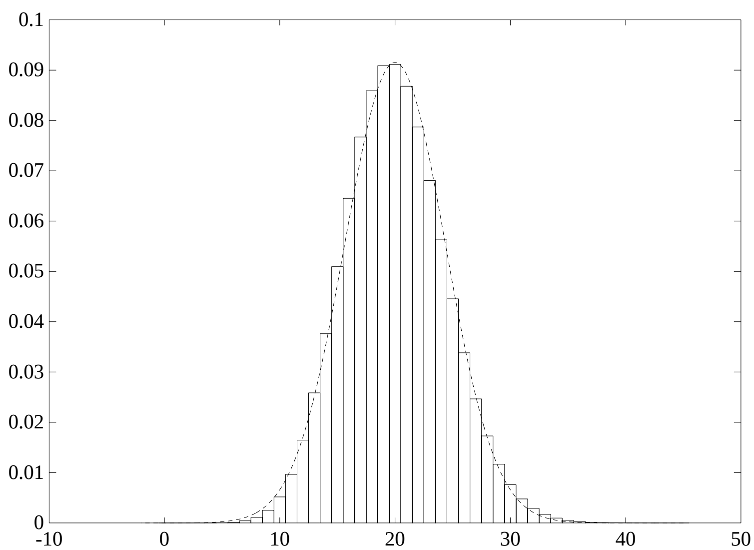


FIG. 7.9 – Binomiale $n = 400$, $p = 0.05$ ($\varepsilon = 4.9 \times 10^{-4}$)

7.3 Le théorème de De Moivre-Laplace

Ces figures laissent entrevoir la possibilité d'approximer une binomiale en un sens que nous précisons maintenant. Il est commode pour cela d'effectuer un changement d'origine de façon à centrer les variables aléatoires et un changement d'échelle de façon à ce qu'elles aient toutes pour variance 1. La nouvelle v.a. obtenue par ce procédé est :

$$S_n^* = \frac{S_n - \mathbb{E}(S_n)}{\sigma(S_n)} = \frac{S_n - np}{\sqrt{npq}} \tag{7.5}$$

Théorème 7.3 (De Moivre-Laplace) *Soit S_n une variable aléatoire de loi binomiale de paramètres n et p et S_n^* définie par (7.5). Pour tous réels $c < d$ fixés :*

$$\lim_{n\to+\infty} P(c < S_n^* \leq d) = \Phi(d) - \Phi(c) = \int_c^d \frac{1}{\sqrt{2\pi}} \exp\Bigl(-\frac{t^2}{2}\Bigr)\,\mathrm{d}t.$$

On a la même limite pour $P(c \leq S_n^* \leq d)$, $P(c \leq S_n^* < d)$ et $P(c < S_n^* < d)$. De Moivre a donné la première version de ce théorème[2] en 1733. Laplace (1812) prouva plus tard le même résultat par une méthode différente en obtenant une évaluation de la vitesse de convergence. La démonstration du théorème (cf. section 7.4) repose sur un encadrement adéquat des coefficients binomiaux et la formule de Stirling. Elle pourra sans inconvénient être passée en première lecture, l'important étant plutôt de comprendre la signification et l'utilisation pratique du théorème de De Moivre-Laplace. Comme dans le cas de l'approximation par une loi de Poisson, ce théorème permet l'approximation de $P(a < S_n \leq b)$ pour les grandes valeurs de n. Pour cela il suffit de remarquer que la fonction $x \mapsto (x - np)/\sqrt{npq}$ étant croissante, on a l'équivalence :

$$a < S_n(\omega) \leq b \;\Leftrightarrow\; \frac{a-np}{\sqrt{npq}} < \frac{S_n(\omega) - np}{\sqrt{npq}} = S_n^*(\omega) \leq \frac{b-np}{\sqrt{npq}}.$$

On en déduit que pour tout $n \geq 1$

$$P(a < S_n \leq b) = P\Bigl(\frac{a-np}{\sqrt{npq}} < S_n^* \leq \frac{b-np}{\sqrt{npq}}\Bigr)$$

Lorsque n est « grand » le théorème de De Moivre-Laplace nous dit que le second membre peut être approximé par :

$$P\Bigl(\frac{a-np}{\sqrt{npq}} < S_n^* \leq \frac{b-np}{\sqrt{npq}}\Bigr) \simeq \Phi\Bigl(\frac{b-np}{\sqrt{npq}}\Bigr) - \Phi\Bigl(\frac{a-np}{\sqrt{npq}}\Bigr).$$

Il suffit alors d'utiliser la table des valeurs de Φ pour calculer cette probabilité.

[2]Publiée en 1738 dans la seconde édition de son livre *The Doctrine of Chances*.

Pour que l'approximation soit légitime, il convient de savoir majorer l'erreur commise en fonction de n, p (et aussi de a et b). Une telle majoration sera donnée sans démonstration section 7.5. L'idée générale est que la vitesse de convergence dans le théorème de De Moivre-Laplace est, comme pour l'inégalité de Tchebycheff, en $O(n^{-1/2})$. La constante sous-entendue dans le O est d'autant meilleure que p est proche de $1/2$ (et se dégrade fortement quand p est proche de 0 ou 1). Cette dépendance par rapport à p est illustrée par les figures 7.5 à 7.9.

Exemple 7.1 *On lance* 3600 *fois un dé. Évaluer la probabilité que le nombre d'apparitions du* 1 *soit compris strictement entre* 540 *et* 660.
Soit S le nombre d'apparitions du 1. Cette variable aléatoire suit la loi binomiale $\mathcal{B}(3600, 1/6)$. On a $\mathbb{E}\,S = 600$ et $\operatorname{Var} S = 500$. En notant $S^* = (S - \mathbb{E}\,S)/\sigma(S)$ la variable centrée réduite associée à S :

$$P(540 < S < 660) = P\Big(\frac{540-600}{\sqrt{500}} < S^* < \frac{660-600}{\sqrt{500}}\Big).$$

Comme n est grand on peut utiliser l'approximation liée au théorème de De Moivre-Laplace :

$$P\Big(\frac{-60}{10\sqrt{5}} < S^* < \frac{60}{10\sqrt{5}}\Big) \simeq \Phi\Big(\frac{6}{\sqrt{5}}\Big) - \Phi\Big(\frac{-6}{\sqrt{5}}\Big).$$

En utilisant la parité de la densité $f_{0,1}$, on a pour tout $a > 0$: $\Phi(a) - \Phi(-a) = 2\Phi(a) - 1$. En utilisant la table des valeurs de Φ on obtient donc :

$$P(540 < S < 660) \simeq 2\Phi(2.68) - 1 \simeq 0.9926 \simeq 0.99.$$

(en raison de l'approximation de $6/\sqrt{5}$ par 2.68, il n'y a pas lieu de conserver les deux derniers chiffres).

Il est intéressant de comparer ce résultat avec ce qu'aurait donné l'inégalité de Tchebycheff. En reprenant les calculs de l'exemple 5.13, on obtient :

$$P(540 < S < 660) \geq 1 - \frac{500}{60^2} = \frac{31}{36} > 0.86.$$

Pour être honnête, il convient néanmoins de noter que les deux résultats sont de nature différente. Le deuxième est une minoration dont on est certain, pas une approximation. Pour pouvoir affirmer que le théorème de De Moivre-Laplace nous donne ici un meilleur résultat que l'inégalité de Tchebycheff, il faut donc vérifier que l'erreur d'approximation est inférieure à $0.99 - 0.86 = 0.13$. (À suivre...)

Exemple 7.2 *Une entreprise emploie 500 personnes qui déjeunent à la cantine à l'un ou l'autre des deux services avec une probabilité égale de manger au* *premier ou au second. Si le gérant veut avoir une probabilité supérieure à 95% de disposer d'assez de couverts, combien devra-t-il en prévoir pour chacun des deux services ?*

Voici une modélisation du problème. On numérote les 500 personnes par ordre alphabétique[3]. On note X_i la variable aléatoire valant 1 si la i-ème personne de la liste mange au premier service, 0 si elle mange au deuxième. Les X_i sont des variables de Bernoulli de même loi : $P(X_i = 0) = P(X_i = 1) = 1/2$. On suppose les X_i indépendantes, hypothèse raisonnable dans le cas d'une numérotation par ordre alphabétique[4]. Bien que l'énoncé ne le précise pas il s'agit évidemment de trouver le nombre *minimum* de couverts à disposer à chaque service, sinon avec 500 couverts à chaque service, le gérant ne prendrait aucun risque ! Notons k ce nombre minimum. Le nombre (aléatoire) de personnes mangeant au premier service est : $S_n = \sum_{i=1}^{500} X_i$ (avec $n = 500$). Le nombre de personnes mangeant au deuxième service est par conséquent $500 - S_n$ (on suppose que tout le monde mange). Le problème revient donc à trouver k minimal tel que :

$$P(S_n \leq k \text{ et } 500 - S_n \leq k) \geq 0.95.$$

Ce qui s'écrit encore :

$$P(500 - k \leq S_n \leq k) \geq 0.95. \tag{7.6}$$

La loi de S_n est la binomiale $\mathcal{B}(500, 1/2)$, mais comme n est grand, il est légitime d'utiliser l'approximation de cette loi par le théorème de De Moivre-Laplace. Pour ce faire, on normalise S_n en la centrant puis en divisant par l'écart type. L'espérance et la variance de S_n sont $\mathbb{E}\, S_n = n\frac{1}{2} = 250$ et $\operatorname{Var} S_n = n(\frac{1}{2})(1-\frac{1}{2}) = 125$. On peut écrire (7.6) sous la forme équivalente :

$$P\left(\frac{500 - k - 250}{\sqrt{125}} \leq S_n^* \leq \frac{k - 250}{\sqrt{125}}\right) \geq 0.95.$$

où $S_n^* = (S_n - 250)/\sqrt{125}$. En négligeant l'erreur d'approximation, le problème revient à trouver k minimal tel que :

$$\Phi\left(\frac{k - 250}{\sqrt{125}}\right) - \Phi\left(\frac{250 - k}{\sqrt{125}}\right) \geq 0.95.$$

En utilisant comme dans l'exemple précédent la relation $\Phi(a) - \Phi(-a) = 2\Phi(a) - 1$ avec ici $a = (k - 250)/\sqrt{125}$, on est ramené à la résolution de l'inéquation : $2\Phi(a) - 1 \geq 0.95$ équivalente à $\Phi(a) \geq 0.975$. La table nous donne $\Phi(1.96) = 0.9750$. On prendra donc pour k le plus petit entier tel que $(k - 250)/\sqrt{125} \geq 1.96$

[3] Ou tout autre ordre *fixé à l'avance*.

[4] Ce ne serait bien sûr pas le cas si on les avait numérotées d'après leur ordre d'arrivée à la cantine, mais cet ordre est lui même aléatoire et nous l'avons exclu.

d'où : $k \geq 250 + 1.96\sqrt{125} \approx 271.91$. Ainsi il suffit de prévoir 272 couverts par service pour avoir une probabilité supérieure à 0.95 que chacun puisse manger au service de son choix. On constate qu'en acceptant un risque faible de surcharge de la cantine, il y a moyen de réaliser une économie considérable en place et en mobilier. ■

7.4 Preuve du théorème de De Moivre-Laplace

Nous donnons dans cette section une démonstration du théorème de De Moivre-Laplace (cas général : p quelconque dans $]0,1[$). Les techniques utilisées relèvent uniquement du programme d'analyse du DEUG. Signalons qu'il existe une démonstration beaucoup plus rapide[5] basée sur la transformation de Fourier (cf. cours de Licence).

La preuve que nous proposons n'est pas la démonstration classique. Elle s'inspire de Feller[6]. En fait nous allons faire un peu plus que prouver le théorème de De Moivre-Laplace. Nous allons montrer l'existence d'un entier $n_0 = n_0(p)$ et d'une constante C tels que :

$$\forall n \geq n_0, \qquad \bigl|P(c \leq S_n^* \leq d) - \bigl(\Phi(d) - \Phi(c)\bigr)\bigr| \leq \frac{C}{\sqrt{npq}}, \tag{7.7}$$

ce qui prouvera le théorème tout en nous donnant une idée de la vitesse de convergence[7]. Nous notons dans la suite :

$$b(k,n,p) = C_n^k p^k q^{n-k}.$$

Il s'agit d'étudier le comportement asymptotique de :

$$P(c \leq S_n^* \leq d) = \sum_{c \leq x_k \leq d} b(k,n,p), \quad \text{où} \quad x_k = \frac{k-np}{\sqrt{npq}}. \tag{7.8}$$

Remarquons que dans cette somme, le nombre de termes (en gros $(d-c)\sqrt{npq}$) tend vers l'infini avec n, donc k peut très bien tendre vers l'infini avec n, mais pas n'importe comment. Il doit vérifier une condition du type :

$$\left|\frac{k-np}{\sqrt{npq}}\right| \leq c_0 = \max(|c|,|d|). \tag{7.9}$$

La démonstration se fait en deux temps :

[5]Mais moins instructive.

[6]W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I, ex. 19–21 p. 182 (2e édition).

[7]On ne cherchera pas à optimiser la valeur de C, ni même à l'expliciter.

– On cherche un encadrement de $b(k,n,p)$ donnant un équivalent de la forme $(2\pi npq)^{-1/2}\exp(-x_k^2/2)$ lorsque n tend vers $+\infty$ et k reste astreint à la condition (7.9).
– On traite alors le second membre de (7.8) comme une somme de Riemann que l'on approxime par l'intégrale $\Phi(d)-\Phi(c)$ en contrôlant l'erreur commise.

7.4.1 Évaluation asymptotique de $b(k,n,p)$

Notre point de départ est l'évaluation asymptotique du terme central de la loi binomiale pour laquelle nous renvoyons à l'exercice 3.16 p. 70 dont nous rappelons les résultats :

Le maximum de $b(k,n,p)$ à n et p fixés est atteint pour $k=m$ défini comme l'unique entier tel que :

$$(n+1)p-1<m\leq(n+1)p.$$

Il existe un entier $n_1=n_1(p)$ et une constante c_1 tels que

$$\forall n\geq n_1,\quad b(m,n,p)=\frac{1}{\sqrt{2\pi npq}}(1+\varepsilon_n)\quad\text{avec}\quad|\varepsilon_n|\leq\frac{c_1}{npq}.\tag{7.10}$$

On raccroche l'évaluation asymptotique de $b(k,n,p)$ à celle de $b(m,n,p)$ en écrivant :

$$\text{si }\ k>m,\qquad b(k,n,p)\ =\ b(m,n,p)\prod_{j=m+1}^{k}\frac{b(j,n,p)}{b(j-1,n,p)},\tag{7.11}$$

$$\text{si }\ k<m,\qquad b(k,n,p)\ =\ b(m,n,p)\prod_{j=k+1}^{m}\frac{b(j-1,n,p)}{b(j,n,p)}.\tag{7.12}$$

Nous sommes ainsi conduits à la recherche d'un encadrement du quotient :

$$\frac{b(j,n,p)}{b(j-1,n,p)}=\frac{(n+1-j)p}{jq}.\tag{7.13}$$

Pour des raisons de simplification calculatoire, on « centre » j par $(n+1)p$ et on « normalise » en divisant par $\sqrt{(n+1)pq}$. Pour cela, posons :

$$j=(n+1)p+\delta_j\qquad\text{et}\qquad(n+1)pq=\tau_n^2\quad(\tau>0).$$

Avec ces notations, on a :

$$(n+1-j)p=(n+1)p-(n+1)p^2-p\delta_j=(n+1)p(1-p)-p\delta_j=\tau_n^2-p\delta_j.$$

En reportant dans (7.13), il vient :

$$\frac{b(j,n,p)}{b(j-1,n,p)}=\frac{\tau_n^2-p\delta_j}{\tau_n^2+q\delta_j}=\frac{1-pu_j}{1+qu_j}\quad\text{avec}\quad u_j=\frac{\delta_j}{\tau_n^2}.\tag{7.14}$$

Examinons d'abord le cas où $k > m$: on a alors $j > m$ et $\delta_j > 0$ donc $u_j > 0$. Nous pouvons alors utiliser avec $x = qu_j$ l'encadrement suivant :

$$\forall x > 0, \qquad 1 - x < \frac{1}{1+x} < 1 - x + x^2.$$

En effet, pour $x \geq 1$, cet encadrement est trivial, pour $0 < x < 1$, on utilise le développement $(1+x)^{-1} = 1 - x + x^2 - x^3 + \cdots$ en série alternée dont le terme général décroît en valeur absolue. La somme de la série est donc encadrée par deux sommes partielles consécutives.

Minoration du quotient :

$$\frac{1 - pu_j}{1 + qu_j} > (1 - pu_j)(1 - qu_j) = 1 - (p+q)u_j + pqu_j^2 > 1 - u_j.$$

Majoration du quotient :

$$\begin{aligned}
\frac{1 - pu_j}{1 + qu_j} &< (1 - pu_j)(1 - qu_j + q^2u_j^2) \\
&= 1 - (p+q)u_j + (pq + q^2)u_j^2 - pq^2u_j^3 \\
&< 1 - u_j + qu_j^2 \\
&< 1 - u_j + u_j^2.
\end{aligned}$$

Nous obtenons ainsi l'encadrement :

$$\text{Si } j > m, \qquad 1 - u_j < \frac{b(j,n,p)}{b(j-1,n,p)} < 1 - u_j + u_j^2. \tag{7.15}$$

Dans le cas où $k < m$, on s'intéresse au quotient inverse :

$$\frac{b(j-1,n,p)}{b(j,n,p)} = \frac{1 + qu_j}{1 - pu_j}.$$

Cette fois u_j est négatif. En échangeant les rôles de p et q et en remplaçant u_j par $-u_j$, on est donc ramené au cas précédent, d'où :

$$\text{Si } j < m, \qquad 1 + u_j < \frac{b(j-1,n,p)}{b(j,n,p)} < 1 + u_j + u_j^2. \tag{7.16}$$

L'étape suivante est d'encadrer les produits figurant dans (7.11) et (7.12). Pour cela, il est commode d'utiliser le lemme suivant qui procure un encadrement de chaque facteur par des exponentielles.

Lemme 7.4

$$\forall t \in]0,1[, \qquad \exp\Big(-t - \frac{t^2}{2(1-t)}\Big) < 1 - t < \exp(-t).$$

Preuve : L'inégalité de droite provient de la convexité de la fonction $t \mapsto \mathrm{e}^{-t}$: la courbe représentative est toujours au dessus de sa tangente au point d'abscisse 0 (on peut aussi considérer le développement en série entière de e^{-t} et remarquer que pour $0 < t < 1$, il s'agit d'une série alternée dont le terme général décroît en valeur absolue...).

Pour l'inégalité de gauche, on utilise le développement en série entière :

$$\forall t \in]0,1[, \qquad \ln(1-t) = -t - \frac{t^2}{2} - \frac{t^3}{3} - \cdots - \frac{t^k}{k} - \cdots = -t - \sum_{k=2}^{+\infty} \frac{t^k}{k}.$$

En minorant chaque dénominateur par 2 dans la série, on exploite la formule explicite pour la somme d'une série géométrique pour obtenir :

$$\sum_{k=2}^{+\infty} \frac{t^k}{k} < \frac{t^2}{2}(1 + t + t^2 + \cdots) = \frac{t^2}{2(1-t)}.$$

On en déduit :

$$\forall t \in]0,1[, \qquad \ln(1-t) > -t - \frac{t^2}{2(1-t)},$$

et il ne reste plus qu'à prendre l'exponentielle des deux membres. ∎

Notons maintenant que d'après (7.9), u_j tend vers 0 comme $n^{-1/2}$ lorsque n tend vers $+\infty$ (et même uniformément par rapport à j) : on peut trouver un entier $n_2 = n_2(p,c,d)$ tel que :

$$\forall n \geq n_2, \qquad 0 \leq |u_j| \leq \frac{1}{2},$$

pour tous les j concernés par notre démonstration. Si $t \in [0, 1/2]$, $1/(1-t) \leq 2$. Combinée à cette remarque, l'application du lemme 7.4 avec $t = u_j$ nous fournit :

$$\forall n \geq n_2, \ \forall j > m, \qquad \exp(-u_j - u_j^2) < 1 - u_j.$$

Le même lemme avec $t = u_j - u_j^2$ nous donne :

$$\forall n \geq n_2, \ \forall j > m, \qquad 1 - u_j + u_j^2 < \exp(-u_j + u_j^2).$$

Ces deux inégalités nous permettent de remplacer les bornes de l'encadrement (7.15) par des exponentielles. On en déduit pour tout $n \geq n_2$ et tout $k > m$:

$$\exp\Bigl[-\sum_{j=m+1}^{k} u_j - \sum_{j=m+1}^{k} u_j^2\Bigr] < \prod_{j=m+1}^{k} \frac{b(j,n,p)}{b(j-1,n,p)} < \exp\Bigl[-\sum_{j=m+1}^{k} u_j + \sum_{j=m+1}^{k} u_j^2\Bigr].$$

Évaluation de la somme des u_j

La somme des termes d'une progression arithmétique est égale au nombre de termes multiplié par la moyenne du premier et du dernier. En appliquant cette règle, puis en centrant par rapport à $(n+1)p$, nous obtenons :

$$\begin{aligned}
\sum_{j=m+1}^{k} u_j &= \frac{1}{\tau_n^2} \sum_{j=m+1}^{k} \bigl(j-(n+1)p\bigr) \\
&= \frac{1}{2\tau_n^2}(k-m)\bigl(k+m+1-2(n+1)p\bigr) \\
&= \frac{1}{2\tau_n^2}\Bigl\{\bigl(k-(n+1)p\bigr)-\bigl(m-(n+1)p\bigr)\Bigr\} \\
&\qquad\qquad \times \Bigl\{\bigl(k-(n+1)p\bigr)+\bigl(m-(n+1)p\bigr)+1\Bigr\} \\
&= \frac{\bigl(k-(n+1)p\bigr)^2-\bigl(m-(n+1)p\bigr)^2}{2\tau_n^2}+\frac{k-m}{2\tau_n^2} \\
&= \frac{\bigl(k-(n+1)p\bigr)^2}{2\tau_n^2}-\frac{\bigl(m-(n+1)p\bigr)^2}{2\tau_n^2}+\frac{k-m}{2\tau_n^2}. \qquad (7.17)
\end{aligned}$$

Comme $\bigl(m-(n+1)p\bigr)^2$ est majoré par 1, la somme des deux derniers termes de (7.17) est comprise entre $(k-1-m)/(2\tau_n^2)$ et $(k-m)/(2\tau_n^2)$. Il existe donc une constante c_2 (par exemple $c_2=(1+c_0)/2$) telle que :

$$\frac{\bigl(k-(n+1)p\bigr)^2}{2\tau_n^2}-\frac{c_2}{\tau_n} \leq \sum_{j=m+1}^{k} u_j \leq \frac{\bigl(k-(n+1)p\bigr)^2}{2\tau_n^2}+\frac{c_2}{\tau_n}, \qquad (7.18)$$

pour tout $n \geq n_2$ et tous les $k>m$ vérifiant (7.9).

Majoration de la somme des u_j^2

A la différence de la somme des u_j qui est le terme principal, celle des u_j^2 est un terme d'erreur. Comme nous voulons seulement avoir une idée de la vitesse de convergence dans le théorème de De Moivre-Laplace, une majoration grossière nous suffit. On l'obtient en majorant chaque terme de la somme par le plus grand d'entre eux :

$$\sum_{j=m+1}^{k} u_j^2 = \frac{1}{\tau_n^4} \sum_{j=m+1}^{k} \bigl(j-(n+1)p\bigr)^2 \leq \frac{(k-m)^3}{\tau_n^4}.$$

Il existe une constante c_3 (par exemple $c_3=(1+c_0)^3$) telle que pour tout $n \geq n_2$ et tous les $k>m$ vérifiant (7.9) :

$$0 \leq \sum_{j=m+1}^{k} u_j^2 \leq \frac{c_3}{\tau_n}. \qquad (7.19)$$

Posons pour alléger les écritures :

$$t_k = \frac{\bigl(k-(n+1)p\bigr)^2}{\tau_n}.$$

De (7.18) et (7.19), on tire :

$$\exp\Bigl(-\frac{t_k^2}{2}-\frac{c_2+c_3}{\tau_n}\Bigr) \le \prod_{j=m+1}^{k} \frac{b(j,n,p)}{b(j-1,n,p)} \le \exp\Bigl(-\frac{t_k^2}{2}+\frac{c_2+c_3}{\tau_n}\Bigr). \qquad (7.20)$$

En combinant (7.10), (7.11) et (7.20), on vérifie facilement qu'il existe une constante c_4 et un entier $n_3 = n_3(p,c,d)$ tels que pour tout $n \ge n_3$ et tous les $k > m$ vérifiant (7.9) :

$$b(k,n,p) = \frac{\mathrm{e}^{-t_k^2/2}}{\sqrt{2\pi npq}}(1+\varepsilon'_{n,k}), \quad \text{avec} \quad |\varepsilon'_{n,k}| \le \frac{c_4}{\tau_n}. \qquad (7.21)$$

Pour le cas $j < m$, u_j est négatif et on a les mêmes estimations que pour $j > m$ en remplaçant u_j par $-u_j$ (en prenant garde que dans l'évaluation de la somme des u_j, le nombre de termes est $(m-k)$ au lieu de $(k-m)$). On vérifie ainsi facilement[8] que (7.21) reste valable pour $k < m$.

Avant de passer à la suite de notre programme, il convient de mettre le résultat obtenu sous une forme plus commode en revenant à la normalisation et au centrage traditionnels. Posons :

$$\sigma_n = \sqrt{npq} \quad \text{et} \quad x_k = \frac{k-np}{\sigma_n}. \qquad (7.22)$$

Notons que $1/\tau_n \le 1/\sigma_n$. On vérifiera à titre d'exercice que :

$$\exp(-t_k^2/2) = \exp(-x_k^2/2)(1+\varepsilon''_{n,k}) \quad \text{avec} \quad |\varepsilon''_{n,k}| \le \frac{c_5}{\sigma_n},$$

pour $n \ge n_4$ et tous les k vérifiant (7.9). Le résultat de cette partie peut donc s'énoncer :

Théorème 7.5
Avec les notations (7.22) ci-dessus, il existe une constante A et un entier $n_0 = n_0(p,c_0)$ tels que pour tout $n \ge n_0$ et tout k tel que $|k-np| \le c_0\sqrt{npq}$:

$$b(k,n,p) = \frac{\exp(-x_k^2/2)}{\sigma_n\sqrt{2\pi}}(1+\varepsilon_{n,k}) \quad \textit{avec} \quad |\varepsilon_{n,k}| \le \frac{A}{\sigma_n}. \qquad (7.23)$$

[8] Pour les courageux et les sceptiques pas encore épuisés. . .s'il en reste.

7.4.2 Sommes de Riemann

Disposant d'une évaluation asymptotique de chaque terme $b(k,n,p)$, nous pouvons passer à celle de leur somme indexée par la condition :

$$c \leq \frac{k-np}{\sqrt{npq}} \leq d.$$

Nous notons respectivement k_1 et k_2 le plus petit et le plus grand entier vérifiant cette condition. Nous nous proposons dans cette partie d'évaluer à l'aide d'une intégrale la somme :

$$P(c \leq S_n^* \leq d) = \sum_{k=k_1}^{k_2} b(k,n,p) = \frac{1}{\sigma_n} \sum_{c \leq x_k \leq d} f(x_k)(1+\varepsilon_{n,k}),$$

où l'on pose :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big).$$

Le nombre de termes dans cette somme est $k_2 - k_1 + 1$ qui est majoré par $(d-c)\sigma_n + 1$. Compte tenu de (7.23), on en déduit :

$$\frac{1}{\sigma_n} \sum_{k=k_1}^{k_2} |\varepsilon_{n,k}| \leq \frac{A\big((d-c)\sigma_n + 1\big)}{\sigma_n^2} \leq \frac{c_6}{\sigma_n}.$$

Par conséquent :

$$\Big| P(c \leq S_n^* \leq d) - \frac{1}{\sigma_n} \sum_{c \leq x_k \leq d} f(x_k) \Big| \leq \frac{c_6}{\sigma_n}. \tag{7.24}$$

On est ainsi ramené à l'évaluation de la somme de Riemann de terme général $\sigma_n^{-1} f(x_k)$. On considère pour cela $\sigma_n^{-1} f(x_k)$ comme l'aire d'un rectangle de base $[s_k, s_{k+1}]$ et de hauteur $f(x_k)$, où les s_k constituent un partage de l'intervalle $[c,d]$ tel que $s_{k+1} - s_k = \sigma_n^{-1}$ et $s_k \leq x_k \leq s_{k+1}$ pour tout k. A priori, x_k peut être placé n'importe où[9] sur $[s_k, s_{k+1}]$. Nous admettrons (cf. exercice 7.10) que le meilleur choix pour une fonction f de classe C^2, est de prendre x_k au milieu de $[s_k, s_{k+1}]$, ce qui revient à poser :

$$s_k = x_k - \frac{1}{2\sigma_n} \quad \text{et} \quad s_{k+1} = x_k + \frac{1}{2\sigma_n} \qquad (k_1 \leq k \leq k_2).$$

Lorsque f est monotone sur $[x_{k-1}, x_{k+1}]$ (ce qui est bien le cas ici sauf peut-être pour l'un de ces intervalles), il est facile de se convaincre que ce choix est meilleur que de prendre comme partage la suite des x_k elle même (figure 7.10).

[9] Le choix de la position de l'un des x_k détermine celle de tous les autres par translation.

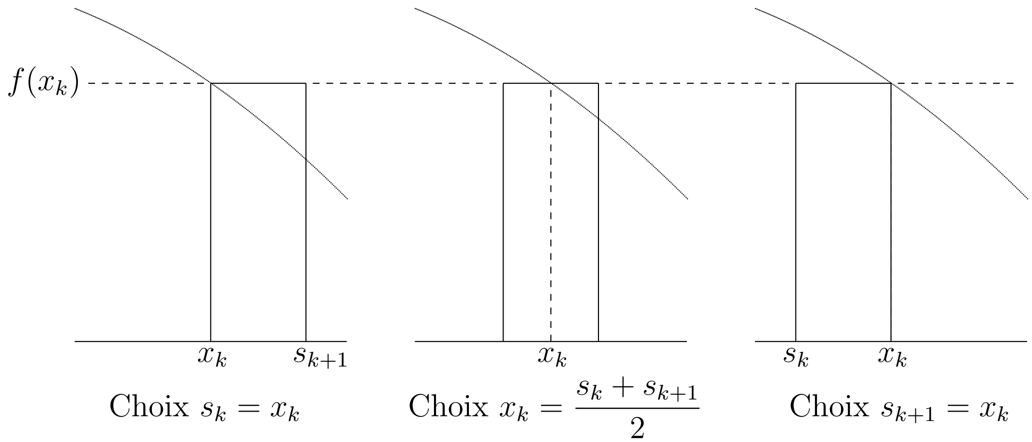


Fig. 7.10 – Divers choix de s_k

Avec notre choix des s_k, on a (cf. exercice 7.10) :

$$\Big|\frac{1}{\sigma_n} f(x_k) - \int_{s_k}^{s_{k+1}} f(x)\,\mathrm{d}x\Big| \leq \frac{1}{24\sigma_n^3} \sup_{[s_k,s_{k+1}]} |f''|.$$

La fonction f'' est donnée ici par $f''(x) = (2\pi)^{-1/2}(x^2-1)\exp(-x^2/2)$. Elle est bornée sur $\mathbb{R}$, le maximum de $|f''|$ est atteint en 0 et vaut $(2\pi)^{-1/2}$. Nous pouvons donc écrire :

$$\frac{1}{\sigma_n}\sum_{k=k_1}^{k_2} f(x_k) = \int_{x_{k_1}-1/(2\sigma_n)}^{x_{k_2}+1/(2\sigma_n)} f(x)\,\mathrm{d}x + \Delta_n, \tag{7.25}$$

où

$$|\Delta_n| \leq \frac{1}{\sqrt{2\pi}} \times \frac{k_2 - k_1 + 1}{24\sigma_n^3} \leq \frac{(d-c)\sigma_n + 1}{24\sqrt{2\pi}\,\sigma_n^3} \leq \frac{c_7}{\sigma_n^2}. \tag{7.26}$$

Enfin dans l'évaluation de notre somme de Riemann par $\Phi(d) - \Phi(c)$, nous devons tenir compte d'une dernière source d'erreur : les deux termes de bord générés par la non coïncidence (en général) des bornes c et $x_{k_1} - 1/(2\sigma_n)$ à gauche et d et $x_{k_2} + 1/(2\sigma_n)$ à droite (cf. figure 7.11). Comme à chaque fois la distance entre les deux bornes concernées n'excède pas $1/(2\sigma_n)$ et la fonction f est majorée sur $\mathbb{R}$ par $1/\sqrt{2\pi}$, on contrôle ces deux termes d'erreur par l'inégalité :

$$\Big|\int_{x_{k_1}-1/(2\sigma_n)}^{c} f(x)\,\mathrm{d}x\Big| + \Big|\int_{x_{k_2}+1/(2\sigma_n)}^{d} f(x)\,\mathrm{d}x\Big| \leq \frac{1}{\sqrt{2\pi}\,\sigma_n}. \tag{7.27}$$

Conclusion

L'inégalité triangulaire nous permet de déduire de (7.24), (7.25), (7.26) et (7.27)

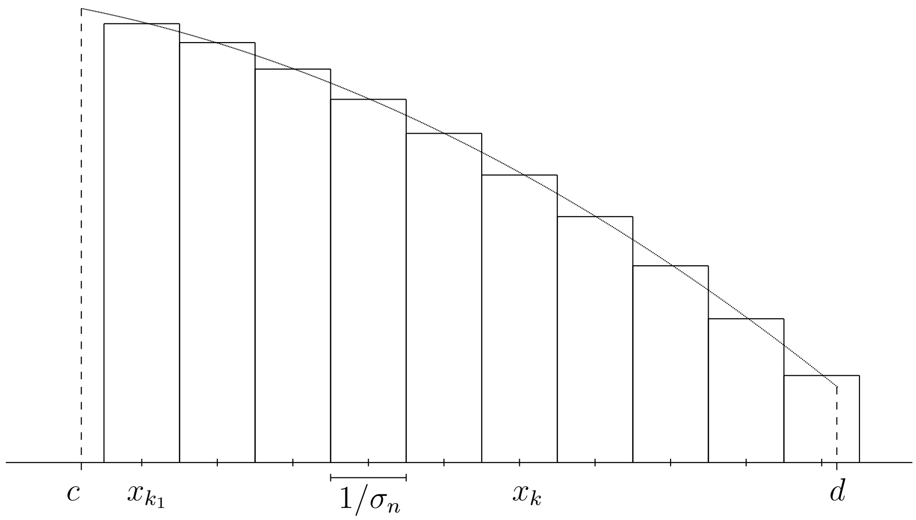


FIG. 7.11 – Approximation par $\Phi(d) - \Phi(c)$

l'existence d'une constante C et d'un entier $n_0 = n_0(p, c, d)$ tels que :

$$\forall n \geq n_0, \qquad \left| P(c \leq S_n^* \leq d) - \int_c^d \frac{\exp(-x^2/2)}{\sqrt{2\pi}} \, \mathrm{d}x \right| \leq \frac{C}{\sqrt{npq}}. \tag{7.28}$$

Ceci achève la preuve du théorème de De Moivre-Laplace. ■

Remarque :
Lorsqu'on applique ce théorème d'un point de vue pratique (i.e. avec n fixé), on peut éliminer l'erreur due aux termes de bord (7.27) en remplaçant c par $x_{k_1} - 1/(2\sigma_n)$ et d par $x_{k_2} + 1/(2\sigma_n)$. On peut voir que cela ne change rien pour S_n^*, en effet :

$$\begin{aligned} x_{k_1} - \frac{1}{2\sigma_n} \leq S_n^* \leq x_{k_2} + \frac{1}{2\sigma_n} \quad & \Leftrightarrow \quad \frac{k_1 - 1/2 - np}{\sqrt{npq}} \leq \frac{S_n - np}{\sqrt{npq}} \leq \frac{k_2 + 1/2 - np}{\sqrt{npq}} \\ & \Leftrightarrow \quad k_1 - \frac{1}{2} \leq S_n \leq k_2 + \frac{1}{2} \\ & \Leftrightarrow \quad k_1 \leq S_n \leq k_2, \end{aligned}$$

puisque S_n ne prend que des valeurs entières. Ce procédé d'approximation s'appelle *correction de continuité*.

7.5 Vitesse de convergence

La démonstration du théorème de De Moivre-Laplace nous a permis de voir[10] que la vitesse de convergence était en $O(n^{-1/2})$, plus précisément nous avons montré que l'erreur commise en utilisant l'approximation :

$$P(a \leq S_n \leq b) \simeq \Phi\Big(\frac{b-np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{a-np}{\sqrt{npq}}\Big)$$

est majorée par $C/\sqrt{npq}$. Par ailleurs l'évaluation du terme dominant de la loi binomiale par la formule de Stirling nous montre qu'on ne peut pas espérer en général avoir une vitesse de convergence meilleure que $O(n^{-1/2})$. De nombreux travaux de l'école russe du début du siècle ont été consacrés à la détermination d'une valeur explicite pour la constante C et plus généralement à la recherche d'un contrôle fin de l'erreur tenant compte des valeurs de p, a et b. Les calculs sont trop compliqués pour trouver leur place dans ce polycopié. Nous nous contenterons de présenter sans démonstration des résultats dûs à Uspensky[11].

Le résultat le plus fin concerne l'approximation obtenue en effectuant la « correction de continuité ». Dans cette approximation, chaque probabilité binomiale $P(S_n = k)$ est représentée graphiquement par le rectangle élémentaire de base $[k-1/2,\ k+1/2]$ et de hauteur $P(S_n = k)$. C'est exactement la convention adoptée sur les figures 7.5 à 7.9 pour représenter la loi de S_n. De plus les bornes de variation de Φ sont ajustées de façon à prendre en compte seulement un nombre entier de ces rectangles (plus de « chutes » aux bords). La réduction des bornes (en abscisse) du rectangle élémentaire par la transformation $t \mapsto (t-np)/\sqrt{npq}$ nous donne comme nouvelles bornes : $(k-1/2-np)/\sqrt{npq}$ et $(k+1/2-np)/\sqrt{npq}$. Si k_1 et k_2 sont des *entiers*, l'approximation avec correction de continuité s'écrit ainsi :

$$\begin{aligned} P(k_1 \leq S_n \leq k_2) &= P\Big(k_1 - \frac{1}{2} \leq S_n \leq k_2 + \frac{1}{2}\Big) \\ &= P\Big(\frac{k_1 - 0.5 - np}{\sqrt{npq}} \leq S_n^* \leq \frac{k_2 + 0.5 - np}{\sqrt{npq}}\Big) \\ &\simeq \Phi\Big(\frac{k_2 + 0.5 - np}{\sqrt{npq}}\Big) - \Phi\Big(\frac{k_1 - 0.5 - np}{\sqrt{npq}}\Big). \end{aligned}$$

Notons pour alléger :

$$z_1 = \frac{k_1 - 0.5 - np}{\sqrt{npq}}, \qquad z_2 = \frac{k_2 + 0.5 - np}{\sqrt{npq}}. \tag{7.29}$$

Voici maintenant une évaluation de l'erreur d'approximation pour la correction de continuité :

[10] La lecture de cette section ne présuppose pas celle de la précédente. Il y a donc forcément quelques redites…

[11] J. V. Uspensky, *Introduction to mathematical probability*, McGraw-Hill 1937.

Théorème 7.6 (Uspensky)
Si S_n suit la loi binomiale $\mathcal{B}(n,p)$ et si $npq \geq 25$, alors pour tous entiers k_1 et k_2 tels que $0 \leq k_1 \leq k_2 \leq n$, on a :

$$P(k_1 \leq S_n \leq k_2) = \Phi(z_2) - \Phi(z_1) + \frac{q-p}{6\sqrt{2\pi npq}} \Big[(1-t^2)\exp\Big(-\frac{t^2}{2}\Big)\Big]_{z_1}^{z_2} + \varepsilon,$$

où z_1 et z_2 sont définis par (7.29) et où le terme complémentaire ε vérifie :

$$|\varepsilon| < \frac{0.13 + 0.18|p-q|}{npq} + \exp\Big(-\frac{3}{2}\sqrt{npq}\Big). \tag{7.30}$$

Le terme d'erreur dominant est en $1/\sqrt{npq}$, il se dégrade quand p est proche de 0 ou de 1 ($q = 1-p$ est alors proche de 0).

Dans le cas particulier où $p = 1/2$, le terme en $1/\sqrt{npq}$ s'annule et l'erreur d'approximation est en $O(1/(npq))$. De même si p est quelconque mais $z_1 = -z_2$ (intervalle $[k_1, k_2]$ symétrique autour de np).

Lorsque l'on n'effectue pas la correction de continuité, les résultats sont moins bons, mais on a toujours en général un terme d'erreur dominant en $1/\sqrt{npq}$. Quelques notations supplémentaires sont nécessaires à leur exposition. Nous souhaitons évaluer $P(x_1 \leq S_n^* \leq x_2)$ pour x_1 et x_2 réels fixés. Pour $x \in \mathbb{R}$, on appelle partie fractionnaire de x le réel $\theta = x - [x]$ (où $[x]$ désigne la partie entière de x, plus grand entier inférieur ou égal à x). On note alors :

$$\theta_1 \quad \text{la partie fractionnaire de} \quad nq - x_1\sqrt{npq}, \tag{7.31}$$
$$\theta_2 \quad \text{la partie fractionnaire de} \quad np - x_2\sqrt{npq}. \tag{7.32}$$

Théorème 7.7 (Uspensky)
Si S_n suit la loi binomiale $\mathcal{B}(n,p)$ et si $npq \geq 25$, alors pour tous réels x_1 et x_2 avec $x_1 < x_2$, on a :

$$P(x_1 \leq S_n^* \leq x_2) = \Phi(x_2) - \Phi(x_1) + \frac{A}{\sqrt{2\pi npq}} + \frac{q-p}{6\sqrt{2\pi npq}} \Big[(1-t^2)e^{-t^2/2}\Big]_{x_1}^{x_2} + \varepsilon',$$

où θ_1 et θ_2 sont définis par (7.31) et (7.32), A par :

$$A = \Big(\frac{1}{2} - \theta_1\Big)\exp\Big(-\frac{x_1^2}{2}\Big) + \Big(\frac{1}{2} - \theta_2\Big)\exp\Big(-\frac{x_2^2}{2}\Big)$$

et le terme ε' est majoré par :

$$|\varepsilon'| < \frac{0.20 + 0.25|p-q|}{npq} + \exp\Big(-\frac{3}{2}\sqrt{npq}\Big). \tag{7.33}$$

On peut faire des commentaires analogues à ceux qui suivent le théorème 7.6. Il est intéressant de déduire du théorème 7.7 une majoration uniforme (par rapport à x_1 et x_2) de l'erreur d'approximation du type $C/\sqrt{npq}$. Ceci permet de se faire rapidement une première idée et de voir s'il y a lieu d'affiner l'approximation à l'aide de l'un des résultats ci-dessus.

Corollaire 7.8 (Majoration uniforme de l'erreur)
Si $npq \geq 25$, pour tous réels $x_1 < x_2$,

$$\bigl|P(x_1 \leq S_n^* \leq x_2) - \bigl(\Phi(x_2) - \Phi(x_1)\bigr)\bigr| \leq \frac{0.588}{\sqrt{npq}}.$$

Preuve : Comme θ_1 et θ_2 sont dans $[0,1[$, il est clair que $|A| \leq 1$. D'autre part l'étude des variations de la fonction $t \mapsto (1-t^2)\mathrm{e}^{-t^2/2}$ montre que sur $\mathbb{R}$ son maximum est 1 (atteint en $t=0$) et son minimum $-2\mathrm{e}^{-1.5} > -0.4463$ (atteint en $t=-\sqrt{3}$ et $t=\sqrt{3}$). On a donc :

$$\left|\frac{q-p}{6\sqrt{2\pi npq}}\Bigl[(1-t^2)\mathrm{e}^{-t^2/2}\Bigr]_{x_1}^{x_2}\right| \leq \frac{1}{\sqrt{npq}}\frac{1+0.4463}{6\sqrt{2\pi}}.$$

Par hypothèse, $\sqrt{npq} \geq 5$, d'où :

$$\frac{0.20+0.25|p-q|}{npq} \leq \frac{1}{\sqrt{npq}}\frac{0.45}{5} = \frac{0.09}{\sqrt{npq}}.$$

Enfin, pour le dernier terme, on écrit :

$$\exp\Bigl(-\frac{3}{2}\sqrt{npq}\Bigr) = \frac{1}{\sqrt{npq}}\Bigl(\sqrt{npq}\exp\Bigl(-\frac{3}{2}\sqrt{npq}\Bigr)\Bigr) \leq \frac{1}{\sqrt{npq}}\sup_{x\geq 5}\bigl(x\mathrm{e}^{-3x/2}\bigr).$$

L'étude des variations de la fonction $x \mapsto x\mathrm{e}^{-3x/2}$ montre qu'elle est décroissante et positive sur $[2/3,+\infty[$, le supremum ci-dessus est donc atteint en $x=5$ et vaut $5\mathrm{e}^{-7.5} < 0.0028$. Finalement en rassemblant toutes ces majorations partielles, on voit que l'on a bien un majorant du type $C/\sqrt{npq}$ et que l'on peut prendre :

$$C = \frac{1}{\sqrt{2\pi}}\Bigl(1+\frac{1.4463}{6}\Bigr) + 0.09 + 0.0028 < 0.5880.$$

■

Exemple 7.3 (suite de l'exemple 7.1)
On lance 3600 *fois un dé et on s'intéresse à la probabilité que le nombre d'apparitions du* 1 *soit compris strictement entre* 540 *et* 660. *D'après l'approximation gaussienne utilisée à l'exemple 7.1, cette probabilité vaut* environ 0.9926. *D'autre part, l'inégalité de Tchebycheff nous dit qu'elle est certainement supérieure à* 0.86. *On voudrait savoir si l'approximation gaussienne donne réellement un meilleur résultat.*

D'après le corollaire 7.8, l'erreur Δ commise en faisant l'approximation est majorée par :

$$\Delta \leq \frac{0.588}{\sqrt{npq}} = \frac{0.588}{\sqrt{500}} < 0.0263.$$

On peut donc affirmer que

$$P(540 < S_n < 660) \geq 0.9926 - 0.0263 > 0.9662$$

(en diminuant d'une unité le quatrième chiffre après la virgule pour prendre en compte la précision de 10^{-4} dans la table des valeurs de Φ). L'approximation gaussienne donne donc bien un meilleur résultat que l'inégalité de Tchebycheff.

■

7.6 Exercices

Ex 7.1. *Un calcul d'intégrale*

1) Montrer la convergence de l'intégrale généralisée :

$$I = \int_0^{+\infty} \exp\Bigl(-\frac{x^2}{2}\Bigr)\,\mathrm{d}x.$$

2) Vérifier que :

$$I^2 = \int_0^{+\infty}\int_0^{+\infty} \exp\Bigl(-\frac{(x^2+y^2)}{2}\Bigr)\,\mathrm{d}x\,\mathrm{d}y.$$

3) Calculer cette intégrale double en passant en coordonnées polaires.

4) En déduire que :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\Bigl(-\frac{x^2}{2}\Bigr)\,\mathrm{d}x = 1.$$

Ex 7.2. *Comment construire la table des valeurs de Φ*

Utiliser le lemme sur l'évaluation de la queue de la loi normale pour donner une méthode pratique de construction d'une table des valeurs de Φ avec une précision au moins égale à 10^{-d}. *Indication :* On considérera le développement en série de Taylor de la densité gaussienne standard sur un intervalle compact $[-a, +a]$ convenablement choisi et on remarquera qu'il s'agit d'une série alternée. Il est donc facile d'encadrer les sommes partielles et de majorer l'erreur commise en remplaçant la série par l'une de ses sommes partielles. On est ainsi ramené à un calcul de polynôme.

Ex 7.3. Trouver un entier k tel qu'avec une probabilité d'environ 0.5 le nombre de faces obtenues en 1000 lancers d'une pièce soit compris au sens large entre 490 et k.

Ex 7.4. Un central téléphonique dessert 5000 abonnés. A un instant donné, chaque abonné a une probabilité égale à 2 % d'utiliser son téléphone et les appels des abonnés sont supposés indépendants. Quel nombre minimal d'appels doit pouvoir traiter simultanément le central pour que sa probabilité d'être saturé à un instant donné soit inférieure à 2.5 % ?

Ex 7.5. *Surréservation aérienne*

Il arrive assez souvent que le nombre de réservations pour une liaison aérienne soit supérieur au nombre de passagers se présentant effectivement le jour du vol. Cela est dû à des empêchements imprévisibles de certains passagers et à une politique systématique de certains d'entre eux qui réservent des places sur plusieurs vols de façon à choisir au dernier moment celui qui leur convient le mieux (en raison de la concurrence, les compagnies ne pénalisent pas les clients qui se désistent et ne font payer effectivement que ceux qui embarquent). Pour compenser ce phénomène, une compagnie aérienne exploitant un avion de 300 places décide de faire de la surréservation (*surbooking*) en prenant pour chaque vol un nombre $n > 300$ de réservations. S'il se présente plus de 300 passagers à l'embarquement, les 300 premiers arrivés prennent leur vol et les autres sont dédommagés financièrement.

1) On considère que les passagers sont mutuellement indépendants et que la probabilité de désistement de chacun d'eux est 10%. On note n le nombre de réservations prises par la compagnie pour un vol donné et S_n le nombre (aléatoire) de passagers se présentant à l'embarquement pour ce vol. Donner la loi de S_n, $\mathbb{E}\, S_n$ et $\operatorname{Var} S_n$.

2) Le directeur commercial de la compagnie aimerait connaître la valeur maximale de n telle que : $P(S_n \leq 300) \geq 0.99$. En utilisant le théorème de De Moivre-Laplace, proposez une solution approchée de ce problème.

Ex 7.6. Une entreprise de vente par correspondance envoie un lundi 2100 colis. La probabilité qu'un colis soit distribué le lendemain est 0.7, celle qu'il soit retourné pour erreur d'adresse est de 0.001. On note X le nombre de colis distribués le lendemain et Y le nombre de colis retournés pour erreur d'adresse.

1) Donner la loi exacte, l'espérance et la variance de X. Même question pour Y.

2) Proposer une valeur approchée de la probabilité $P(1358 \leq X \leq 1442)$ en expliquant soigneusement la méthode utilisée.

3) Par quelle loi discrète classique peut-on approcher la loi de Y ? Utiliser cette approximation pour évaluer $P(Y \geq 4)$.

Ex 7.7. *Un modèle simple pour le mélange de deux gaz*

Un récipient hermétique de 2 litres est séparé en deux parties symétriques « gauche » et « droite » par une cloison hermétique munie d'une vanne à large ouverture. La vanne étant fermée, la partie gauche du récipient contient au départ 1 litre d'oxygène et sa partie droite 1 litre d'azote le tout à la pression atmosphérique. On ouvre la vanne de la cloison intermédiaire et on laisse s'effectuer le mélange, puis au bout d'un temps suffisant on ferme la vanne. On mesure alors la proportion d'azote et d'oxygène dans la partie gauche et on constate *expérimentalement* qu'elles sont égales. Le but de cet exercice est l'étude d'un modèle simple permettant de prévoir ce phénomène.

Les molécules étant agitées d'un mouvement incessant, on suppose qu'après fermeture de la vanne, chaque molécule a une probabilité 1/2 de se trouver dans la partie gauche du récipient. On dispose de n molécules d'oxygène et n d'azote. On indexe les molécules d'oxygène de 1 à n et celles d'azote de $n+1$ à $2n$. On note X_i la variable de Bernoulli indicatrice de l'événement *la i-ème molécule se trouve dans la partie gauche après fermeture de la vanne*. On suppose les X_i *indépendantes*. On pose :

$$S_n = \sum_{i=1}^{n} X_i, \qquad\qquad T_n = \sum_{i=n+1}^{2n} X_i.$$

La variable S_n est donc le nombre aléatoire de molécules d'oxygène et T_n celui des molécules d'azotes dans la partie gauche après fermeture.

1) Quelle est la loi exacte de S_n, son espérance et sa variance (en fonction de n) ? On a clairement les mêmes résultats pour T_n.

2) Soit $x > 0$ un nombre fixé. On considère l'événement

$$A = \left\{ \frac{n}{2} - \frac{x}{2}\sqrt{n} \leq S_n \leq \frac{n}{2} + \frac{x}{2}\sqrt{n} \right\}.$$

En utilisant le théorème de De Moivre-Laplace, montrer que l'on peut approximer $P(A)$ par $2\Phi(x) - 1$ où Φ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$. On a bien sûr le même résultat pour $P(B)$, avec

$$B = \left\{ \frac{n}{2} - \frac{x}{2}\sqrt{n} \leq T_n \leq \frac{n}{2} + \frac{x}{2}\sqrt{n} \right\}.$$

3) On suppose désormais que $n - x\sqrt{n} > 0$. On s'intéresse à l'événement

$$C = \left\{ \frac{n - x\sqrt{n}}{n + x\sqrt{n}} \leq \frac{T_n}{S_n} \leq \frac{n + x\sqrt{n}}{n - x\sqrt{n}} \right\}.$$

Montrer que $A \cap B \subset C$.

4) En négligeant l'erreur due à l'approximation gaussienne, proposer à l'aide de $\Phi(x)$ une majoration de $P(A^c \cup B^c)$. En déduire une *minoration* de $P(C)$. On exprimera simplement le résultat final en fonction de la quantité $R(x) = (1 - \Phi(x))$.

5) Application numérique : $n = 10^{22}$, $x = 10$. On utilisera la formule d'encadrement de $1 - \Phi(x)$ pour les « très grandes valeurs de x » fournie à la fin des tables. Commentez le résultat obtenu.

Ex 7.8. Une compagnie d'assurances assure sur une année n personnes contre un certain risque. On note X_i la somme qu'aura à verser cette année la compagnie au i-ème client. C'est une variable aléatoire qui prend la valeur 0 lorsque le client n'est pas sinistré. On peut en général considérer que les variables aléatoires $X_1, \dots X_n$ sont indépendantes. Supposons qu'elles obéissent toutes à une même loi d'espérance mathématique μ et de variance σ^2. Soit x la prime demandée à chacun des n clients. Comment la compagnie doit-elle fixer x pour qu'avec une probabilité supérieure à $1 - \varepsilon$ la différence entre l'encaissement des primes et les remboursements sur l'année reste supérieure à un montant b déterminé par ses frais de gestion et le bénéfice minimum qu'elle désire faire ?

Ex 7.9. Montrer que le théorème de De Moivre-Laplace implique la loi faible des grands nombres pour les fréquences. *Indication :* pour $\varepsilon > 0$ fixé, contrôler l'erreur commise écrivant l' approximation :

$$P\Big(\Big|\frac{S_n}{n} - p\Big| \leq \varepsilon\Big) \simeq \Phi\Big(\varepsilon\sqrt{\frac{n}{pq}}\Big) - \Phi\Big(-\varepsilon\sqrt{\frac{n}{pq}}\Big).$$

Ex 7.10. *Sur l'approximation d'une somme de Riemann*
Soit f une fonction deux fois continûment dérivable sur un intervalle $[a, b]$ et (s_k) un partage quelconque de $[a, b]$. Soit $x_k \in [s_k, s_{k+1}]$.

1) Montrer qu'en général on a :

$$\Big|\frac{1}{s_{k+1} - s_k} f(x_k) - \int_{s_k}^{s_{k+1}} f(x)\,\mathrm{d}x\Big| \leq \frac{c}{(s_{k+1} - s_k)^2}.$$

2) Lorsque x_k est le milieu du segment $[s_k, s_{k+1}]$:

$$\Big|\frac{1}{s_{k+1} - s_k} f(x_k) - \int_{s_k}^{s_{k+1}} f(x)\,\mathrm{d}x\Big| \leq \frac{m_k}{24(s_{k+1} - s_k)^3},$$

où m_k est le maximum de $|f''|$ sur $[s_k, s_{k+1}]$.
Indication : utiliser la formule de Taylor-Lagrange.

Ex 7.11. En reprenant l'exemple 7.3, donner une majoration plus fine de l'erreur d'approximation.

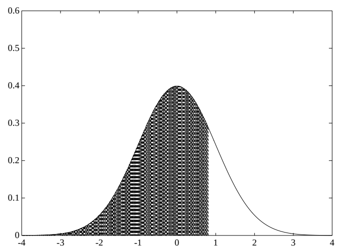
Ex 7.12. *Retour sur le problème des deux dés*

1) Combien de fois faut-il lancer une paire de dés pour qu'avec une probabilité supérieure à 0.999, la fréquence d'apparition du double six soit plus proche de 1/36 que de 1/21 ? (On supposera que la *probabilité* d'obtention d'un double six est effectivement 1/36...).

2) Reprendre la même question en supposant que la probabilité d'apparition du double six est 1/21.

3) Définir et discuter une procédure permettant de répondre expérimentalement à la question posée à la fin du chapitre 1.

Table des valeurs de Φ (loi normale standard)

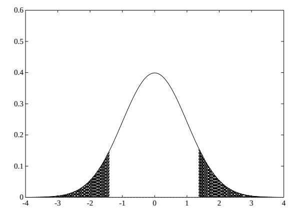


x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6627	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	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7356	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	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	0.8414	0.8438	0.8461	0,8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8622
1.1	0.8643	0.8665	0.8687	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	0.9083	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9193	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	0.9485	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	0.9648	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	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9895	0.9898	0.9901	0.9903	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9944	0.9946	0.9948	0.9949	0,9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	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

Table pour les grandes valeurs de x

x	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.8	4.0	4.5
$\Phi(x)$	0.99865	0.99904	0.99931	0.99952	0.99966	0.99976	0.999841	0.999928	0.999968	0.999997

La table donne les valeurs de $\Phi(x)$ pour x positif. Lorsque x est négatif, on utilise la relation $\Phi(x) = 1 - \Phi(-x)$ qui résulte de la parité de la densité gaussienne $N(0,1)$. Exemple : pour $x = -1,8$, on trouve : $\Phi(x) = 1 - 0,9641 = 0,0359$.



Chapitre 8

Variables aléatoires réelles

8.1 Sortie du cadre discret

Le théorème de De Moivre-Laplace nous a montré que pour la variable aléatoire discrète S_n^*, à la limite, les probabilités $P(a < S_n^* \leq b)$ devenaient des intégrales $\int_a^b f_{0,1}(t)\,\mathrm{d}t$. De même la fonction de répartition F_n^* de S_n^* tend vers la fonction continue Φ qui a toutes les propriétés d'une fonction de répartition (croissance de 0 à 1 sur $]-\infty, +\infty[$, continuité à droite et limites à gauche en tout point) sauf les discontinuités. Il est alors bien tentant d'élargir notre définition des variables aléatoires pour pouvoir dire que Φ est la fonction de répartition d'une variable aléatoire Z et que $P(a < Z \leq b) = \int_a^b f_{0,1}(t)\,\mathrm{d}t$. Avant de succomber définitivement à cette tentation, nous allons examiner d'autres exemples assez naturels de convergence de fonctions de répartition de variables discrètes vers des fonctions croissantes continues.

Convergence de lois uniformes confinées dans $[0, 1]$.

Pour $n \geq 2$, considérons une variable aléatoire U_n de loi uniforme sur l'ensemble $\{1/n, 2/n, \dots, n/n\}$. Soit F_n sa fonction de répartition. C'est la fonction en escalier nulle sur $]-\infty, 1/n[$, valant 1 sur $[1, +\infty[$ et avec sauts d'amplitude $1/n$ aux points k/n $(1 \leq k \leq n)$. On a clairement $F_n(k/n) = k/n$ et pour tout $x \in [0, 1[$, on a un unique k entre 1 et n tel que $(k-1)/n \leq x < k/n$ et $F_n(x) = (k-1)/n$. On a ainsi

$$\forall x \in [0, 1[, \quad F_n(x) \leq x < F_n(x) + \frac{1}{n},$$

d'où $0 \leq x - F_n(x) < 1/n$. Si nous faisons tendre n vers $+\infty$, ceci montre que $F_n(x)$ converge (uniformément) sur $[0, 1[$ vers x. Comme F_n est constante égale à 0 sur $]-\infty, 0[$ et constante égale à 1 sur $[1, +\infty[$, F_n converge sur tout $\mathbb{R}$ vers la fonction continue F définie par

$$F(x) = \begin{cases} 0 & \text{si} \quad -\infty < x < 0, \\ x & \text{si} \quad 0 \leq x \leq 1, \\ 1 & \text{si} \quad 1 < x < +\infty. \end{cases}$$

À nouveau F a toutes les propriétés d'une fonction de répartition, sauf les sauts.

L'exemple que nous venons d'étudier semble à première vue assez artificiel. Il est pourtant relié à un problème essentiel : comment choisir *au hasard* un nombre réel entre 0 et 1 ? ou encore un point d'un segment ? ou encore construire un générateur de nombres aléatoires pour les simulations informatiques ? En pratique, on écrit les nombres sous forme décimale (ou binaire). Choisir un réel au hasard revient ainsi à choisir la suite de ses chiffres décimaux. Et bien sûr comme il est impossible d'écrire une infinité de chiffres, on se contente d'une approximation en tronquant au-delà d'un certain rang m. On considèrera donc une variable aléatoire U_n comme ci-dessus avec $n = 10^m$. Voici une façon complètement élémentaire de générer une telle variable. Dans une urne contenant dix boules numérotées de 0 à 9, on effectue m tirages avec remise. Notons Y_i le chiffre sorti au i-ème tirage. Posons pour $n = 10^m$

$$U_n(\omega) = V_m(\omega) := \sum_{i=1}^{m} \frac{Y_i(\omega)}{10^i}.$$

Ce modèle a déjà été discuté au chapitre 6 (section 6.5). On y a déjà vu que V_m suit la loi uniforme sur l'ensemble des décimaux $\mathbb{D}_m := \{k/10^m;\ 0 \leq k \leq 10^m\}$. D'après ce qui précède, sa fonction de répartition converge vers F. Il n'est pas difficile ici d'imaginer ce que pourrait bien être une variable aléatoire (généralisée) ayant pour fonction de répartition F. Il suffit de remarquer que la série

$$U(\omega) := \sum_{i=1}^{+\infty} \frac{Y_i(\omega)}{10^i}$$

converge pour tout ω vers un réel de $[0, 1]$ puisque le terme général est compris entre 0 et $9/10^i$.

Convergence de lois géométriques par changement d'échelle de temps.

Nous avons vu au chapitre 3 que le temps d'attente du premier succès dans une suite d'épreuves répétées indépendantes suit une loi géométrique. L'expression « temps d'attente » est ici relative à l'horloge intrinsèque de l'expérience dont l'unité de temps est la durée d'une épreuve. Par exemple si l'on fait des tirages avec remise d'une boule dans une urne contenant des boules vertes et des rouges, le numéro du tirage où l'on sort pour la première fois une boule verte est assimilé à la durée de l'attente de ce premier succès[1].

Envisageons maintenant une situation où le temps d'attente a un sens plus physique. Par exemple l'attente du premier éclair d'un orage, celle du premier tremblement de terre de l'année, de la première désintégration d'atome dans un accélérateur de particules,. . .On dispose d'une horloge H_0 pour mesurer ce temps

[1] Ceci suppose implicitement que tous les tirages ont la même durée. D'ailleurs, même si ce n'est pas le cas, il suffit d'imaginer un observateur qui n'aurait pas d'autre moyen de mesurer le temps que de compter le nombre de tirages, d'où le nom d'horloge intrinsèque. . .

d'attente. La durée observée sera toujours un multiple entier de la plus petite durée u_0 mesurable par l'horloge H_0. On découpe alors le temps en intervalles $[0, u_0[$, $[u_0, 2u_0[$, $[2u_0, 3u_0[$, etc. Si l'évènement que l'on attend se produit pour la première fois dans l'intervalle de temps $[ku_0, (k+1)u_0[$, son temps d'attente mesuré par notre horloge aura été ku_0. Si l'on remplace H_0 par une horloge dix fois plus précise H_1, la plus petite unité de temps mesurable sera $u_1 = u_0/10$ et le temps d'attente observé sera l'un des 10 nombres $10ku_1$, $(10k+1)u_1$, ..., $(10k+9)u_1$...

Pour fixer les idées, imaginons que l'on attende le premier éclair lors d'un orage. Prenons comme unité de départ $u_0 = 1$ seconde et notons $A_{[s,t[}$, l'évènement « il n'y a pas d'éclair pendant l'intervalle $[s, t[$ ». On suppose ici que s et t sont des réels ($0 \leq s < t$) mesurant le temps en secondes avec une précision infinie, ce qui est évidemment inaccessible à l'expérience. Nous ferons les hypothèses suivantes (analogues aux hypothèses (a) et (b)) du théorème 3.14) :

(a) Les $A_{[s,t[}$ indexés par des intervalles de temps disjoints sont indépendants ;

(b) Pour tout réels $0 \leq s < t$, $P(A_{[s,t[})$ ne dépend que de la durée $t - s$. Autrement dit, il existe une fonction $h : \mathbb{R}^+ \to [0,1]$ telle que $P(A_{[s,t[}) = h(t-s)$.

Soit X_0 le temps d'attente du premier éclair mesuré par l'horloge H_0. La variable aléatoire discrète X_0 prend ses valeurs dans $\mathbb{N}$. Pour trouver sa loi, on remarque d'abord que $p_0 := P(X_0 = 0) = P(A^c_{[0,1[}) = 1 - h(1)$. Ensuite pour $k \in \mathbb{N}^*$, la décomposition

$$\{X_0 = k\} = \Big(\bigcap_{i=1}^{k} A_{[i-1,i[} \Big) \cap A^c_{[k,k+1[}$$

combinée avec les hypothèses d'indépendance (a) et de stationnarité (b) nous donne en posant $q_0 = h(1) = 1 - p_0$:

$$P(X_0 = k) = q_0^k p_0. \tag{8.1}$$

Cette loi ressemble à la loi géométrique (à une translation près : $1 + X_0$ suit la loi géométrique de paramètre p_0). On suppose désormais que q_0 est strictement inférieur à 1 et strictement positif. Comme pour la loi géométrique, il est facile dévaluer la queue de la loi de X_0. En effet, pour tout entier k, on a

$$P(X_0 > k) = \sum_{j=k+1}^{+\infty} q_0^j p_0 = q_0^{k+1} p_0 \sum_{i=0}^{+\infty} q_0^i = q_0^{k+1} \frac{p_0}{1 - q_0} = q_0^{k+1}.$$

On en déduit que pour tout $x \in \mathbb{R}^+$,

$$P(X_0 > x) = P(X_0 > [x]) = q_0^{[x]+1}, \tag{8.2}$$

où l'on a noté $[x]$ la partie entière de x, c'est-à-dire l'unique entier k tel que $k \leq x < k + 1$.

Maintenant, supposons que l'on dispose d'une suite d'horloges (ou chronomètres) H_n où la plus petite unité de temps mesurable par H_n est $u_n = 10^{-n}$ secondes. Notons X_n le temps d'attente du premier éclair mesuré par cette horloge dans l'unité u_n. Notons Y_n ce même temps d'attente converti en secondes. Par exemple avec $n = 3$, si l'on observe $X_3(\omega) = 5\,347$ millièmes de seconde, on aura $Y_3(\omega) = 5{,}347$, $X_2(\omega) = 534$, $Y_2(\omega) = 5{,}34$, $X_1(\omega) = 53$, $Y_1(\omega) = 5{,}3$, et $X_0(\omega) = Y_0(\omega) = 5$. Il est clair que pour trouver la loi de X_n, il suffit de remplacer dans (8.1), q_0 par $q_n = h(10^{-n})$ et p_0 par $p_n = 1 - q_n$. D'autre part la relation entre $h(1)$ et $h(10^{-n})$ découle immédiatement de la décomposition

$$A_{[0,1[} = \bigcap_{i=1}^{10^n} A_{[(i-1)10^{-n}, i10^{-n}[}.$$

On obtient donc $h(1) = h(10^{-n})^{10^n}$. Pour alléger les écritures, notons $h(1) = \exp(-a)$, ce qui est toujours possible avec un paramètre $a > 0$ puisque $0 < h(1) < 1$. Avec cette notation, $q_n = h(10^{-n}) = \exp(-a10^{-n})$.

Comment se comporte la fonction de répartition de Y_n lorsque n tend vers l'infini ? En utilisant (8.2) avec q_n à la place de q_0, on obtient :

$$P(Y_n > x) = P(X_n > 10^n x) = q_n^{[10^n x]+1} = \exp\bigl(-a10^{-n}([10^n x] + 1)\bigr).$$

Grâce à l'encadrement $y - 1 < [y] \leq y$ valable pour tout réel y, on voit que

$$\frac{10^n x}{10^n} < \frac{[10^n x] + 1}{10^n} \leq \frac{10^n x + 1}{10^n},$$

ce qui se simplifie en $x < 10^{-n}([10^n x] + 1) \leq x + 10^{-n}$ et assure la convergence de $P(Y_n > x)$ vers $\exp(-ax)$.

On en déduit immédiatement que la fonction de répartition de Y_n converge en tout point x vers

$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - \mathrm{e}^{-ax} & \text{si } x > 0. \end{cases}$$

À nouveau F a toutes les propriétés d'une fonction de répartition, sauf les sauts.

8.2 Notion de variable aléatoire réelle

Dans l'observation d'une grandeur physique X (longueur, aire, volume, temps, intensité de courant,...), il est impossible pratiquement d'avoir une précision infinie. Dans ces conditions, attribuer une valeur précise $X = x$ (x étant un réel) signifie implicitement que la vraie valeur de X est dans un petit *intervalle* contenant x et dont la longueur dépend de la précision de l'appareil de mesure dont on dispose. Dans le même ordre d'idées, on peut remarquer que dans les exemples de convergence de fonctions de répartition vus en introduction, les probabilités de valeurs ponctuelles s'annulent à la limite. Il n'en va pas de même pour les probabilités d'intervalles. Pour généraliser la théorie des variables aléatoires discrètes que nous avons étudiée jusqu'à présent, il apparaît donc pertinent de ne pas baser cette théorie sur les probabilités du type $P(X = x)$ mais plutôt sur les $P(X \in I)$ où I est un intervalle. La définition ci-dessous est donc motivée par la nécessité d'attribuer de façon cohérente une probabilité aux ensembles $\{\omega \in \Omega; X(\omega) \in I\}$.

Définition 8.1 *Soit* $(\Omega, \mathcal{F}, P)$ *un espace probabilisé. On appelle variable aléatoire réelle sur* $(\Omega, \mathcal{F}, P)$ *toute application* X *:*

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

vérifiant la condition :

(*i*) *Pour tout intervalle* I *de* $\mathbb{R}$, $A = \{\omega \in \Omega, X(\omega) \in I\}$ *fait partie de la famille* $\mathcal{F}$ *d'événements auxquels on peut attribuer une probabilité par* P.

Cette définition a un contenu plus riche qu'il n'y paraît. Par exemple si X est une variable aléatoire réelle, $\{X \notin [0, 1]\}$ est dans $\mathcal{F}$ puisque $\mathcal{F}$ est stable par passage au complémentaire. De même $\{X \in]-3, 1]$ ou $X \in]4, 5]\}$ est dans $\mathcal{F}$ puisque $\mathcal{F}$ est stable par union dénombrable, donc *a fortiori* par union finie. Plus généralement, on peut montrer que si $\mathcal{B}_{\mathbb{R}}$ est la plus petite famille *contenant tous les intervalles de* $\mathbb{R}$ et telle que [2]

(a) $\mathbb{R} \in \mathcal{B}_{\mathbb{R}}$;

(b) $\mathcal{B}_{\mathbb{R}}$ est stable par passage au complémentaire dans $\mathbb{R}$;

(c) $\mathcal{B}_{\mathbb{R}}$ est stable par union dénombrable ;

alors pour tout $B \in \mathcal{B}_{\mathbb{R}}$, $X^{-1}(B)$ est dans $\mathcal{F}$. On peut donc attribuer une probabilité à $\{X \in B\}$.

Cette famille $\mathcal{B}_{\mathbb{R}}$ s'appelle tribu borélienne de $\mathbb{R}$. Il se trouve qu'il existe des parties de $\mathbb{R}$ qui ne sont pas dans $\mathcal{B}_{\mathbb{R}}$. La situation est donc plus complexe que pour une variable aléatoire discrète Y où $Y^{-1}(B) \in \mathcal{F}$ *pour tout* $B \subset \mathbb{R}$. Cette propriété des variables aléatoires discrètes repose sur le fait que l'ensemble des valeurs possibles $Y(\Omega)$ est *dénombrable*. Pour une variable aléatoire réelle, en général on ne dispose plus de cette hypothèse. Bien sûr, les variables aléatoires discrètes sont aussi des variables aléatoires réelles.

Les propriétés (a)–(c) sont les mêmes que celles attribuées à la famille d'évènements observables $\mathcal{F}$ lors de la définition 1.1 de l'espace probabilisé $(\Omega, \mathcal{F}, P)$ mais avec $\mathbb{R}$ à la place de Ω. On peut alors définir la loi de X comme dans le cas d'une variable discrète (cf. la définition 3.2) mais avec $\mathcal{B}_{\mathbb{R}}$ au lieu de $\mathcal{P}(\mathbb{R})$.

Définition 8.2 *Soit* X *une variable aléatoire réelle sur* $(\Omega, \mathcal{F}, P)$. *On lui associe la fonction d'ensembles* P_X *définie sur la famille* $\mathcal{B}_{\mathbb{R}}$ *de parties de* $\mathbb{R}$ *en*

---

[2] Plus petite au sens de l'inclusion, ce qui signifie que $\mathcal{B}_{\mathbb{R}}$ est une sous-famille de toute autre famille $\mathcal{B}'$ contenant les intervalles et ayant les trois propriétés (a)–(c).

posant :

$$P_X(B) = P(X \in B) = P(\{\omega \in \Omega; X(\omega) \in B\} = P\bigl(X^{-1}(B)\bigr).$$

La fonction d'ensembles P_X ainsi définie est une probabilité sur $\mathcal{B}_\mathbb{R}$. On l'appelle loi de la variable aléatoire X.

Pour légitimer cette définition, il nous faut vérifier que la fonction d'ensembles P_X est bien une probabilité sur $\mathcal{B}_\mathbb{R}$. En effet, $P_X(\mathbb{R}) = P(\Omega) = 1$. De plus si $(B_i)_{i\in\mathbb{N}}$ est une suite de parties de $\mathbb{R}$ éléments de $\mathcal{B}_\mathbb{R}$ et deux à deux disjointes, on voit aisément que

$$X^{-1}\Bigl(\bigcup_{i\in\mathbb{N}} B_i\Bigr) = \{\omega \in \Omega; X(\omega) \in \bigcup_{i\in\mathbb{N}} B_i\} = \bigcup_{i\in\mathbb{N}} \{\omega \in \Omega; X(\omega) \in B_i\}.$$

Comme les évènements $\{X \in B_i\}$ sont deux à deux disjoints, la σ-additivité de P_X découle alors de celle de P :

$$\begin{aligned} P_X\Bigl(\bigcup_{i\in\mathbb{N}} B_i\Bigr) = P\Bigl(X^{-1}\Bigl(\bigcup_{i\in\mathbb{N}} B_i\Bigr)\Bigr) &= P\Bigl(\bigcup_{i\in\mathbb{N}} X^{-1}(B_i)\Bigr) \\ &= \sum_{i\in\mathbb{N}} P\bigl(X^{-1}(B_i)\bigr) \\ &= \sum_{i\in\mathbb{N}} P_X(B_i). \end{aligned}$$

Théorème 8.3 *Soit X une variable aléatoire réelle.*

(a) Sa loi P_X est caractérisée par les probabilités d'intervalles

$$P_X(]a,b]) = P\bigl(X \in]a,b]\bigr), \quad a, b \in \mathbb{R}.$$

(b) Elle est aussi caractérisée par la fonction de répartition $F_X : \mathbb{R} \to [0,1]$ définie par

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

La conséquence pratique de ce théorème est que pour montrer que deux variables aléatoires X et Y ont même loi, il suffit de montrer que pour tous réels $a \le b$, $P(X \in]a,b]) = P(Y \in]a,b])$ ou que X et Y ont même fonction de répartition. La caractérisation (a) revient à dire que deux probabilités (au sens fonctions d'ensembles) sur $\mathcal{B}_\mathbb{R}$ qui coïncident sur la famille des intervalles de $\mathbb{R}$ sont égales. C'est un résultat qui sera vu en Licence et que nous admettons. La caractérisation (b) découle alors facilement de (a) en remarquant que $P(X \in]a,b]) = F_X(b) - F_X(a)$.

Théorème 8.4 *La fonction de répartition F_X d'une variable aléatoire réelle X est croissante sur $\mathbb{R}$, continue à droite et limitée à gauche en tout point. Elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$.*

Preuve : La démonstration est essentiellement la même que celle du théorème 3.4. L'adaptation est laissée au lecteur. ■

Signalons enfin que le théorème 8.4 a la réciproque suivante.

Théorème 8.5 *Soit F une fonction définie et croissante sur $\mathbb{R}$. On suppose de plus que F est continue à droite et limitée à gauche en tout point et qu'elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Alors il existe une espace probabilisé $(\Omega, \mathcal{F}, P)$ et une variable aléatoire réelle X définie sur cet espace et ayant F pour fonction de répartition.*

La preuve complète de ce théorème relève elle aussi du programme de Licence. Le schéma de preuve est le suivant. On prend $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}_{\mathbb{R}}$ et pour X l'application identité de $\mathbb{R}$. On commence par définir P sur la famille $\mathcal{I}$ des intervalles de la forme $]a, b]$ en posant $P(]a, b]) := F(b) - F(a)$. La difficulté est de prouver que P se prolonge à toute la famille $\mathcal{B}_{\mathbb{R}}$...

8.3 Variables à densité

8.3.1 Densité

La loi d'une variable aléatoire X est à densité f si pour tout intervalle de $\mathbb{R}$, la probabilité d'appartenance de X à cet intervalle peut s'écrire comme l'intégrale de f sur cet intervalle. L'apparente simplicité de cette définition informelle est trompeuse. Dans le cadre du programme de DEUG, la seule notion d'intégrale que l'on connaisse est celle de Riemann et il se trouve qu'elle n'est pas totalement satisfaisante pour les besoins de la théorie des probabilités. Nous nous contenterons donc d'une définition susceptible de réinterprétation ultérieure.

Définition 8.6 *Une fonction $f : \mathbb{R} \to \mathbb{R}$ est appelée densité de probabilité si elle est positive (en tout point $t \in \mathbb{R}$ où elle est définie, $f(t) \geq 0$), intégrable sur $\mathbb{R}$ et si*

$$\int_{-\infty}^{+\infty} f(t)\, \mathrm{d}t = 1. \tag{8.3}$$

Définition 8.7 *La variable aléatoire réelle X suit la loi de densité f si :*

$$\forall a \in \mathbb{R},\ \forall b \geq a, \qquad P(X \in]a, b]) = \int_a^b f(t)\, \mathrm{d}t. \tag{8.4}$$

Il est clair d'après cette définition que si Y est une autre variable aléatoire ayant même loi que X (donc mêmes probabilités d'appartenance aux intervalles), elle a aussi la densité f. Il serait donc plus correct de parler de la densité de la loi de X.

Voyons maintenant à quoi peut ressembler une densité. D'abord si f est une fonction positive définie seulement sur un intervalle $]a, b[$ de $\mathbb{R}$ et telle que $\int_a^b f(t)\, \mathrm{d}t = 1$, on peut en faire une densité en la prolongeant à tout $\mathbb{R}$ en posant $f(t) := 0$ pour $t \notin]a, b[$.

Voici quatre exemples simples de densités :

$$f_1(t) := \frac{1}{b-a}\mathbf{1}_{[a,b]}(t); \qquad f_2(t) := \frac{1}{2\sqrt{t}}\mathbf{1}_{]0,1]}(t);$$

$$f_3(t) := \mathrm{e}^{-t}\mathbf{1}_{[0,+\infty[}(t); \qquad f_4(t) := \frac{1}{\pi(1+t^2)}.$$

Ces exemples entrent tous dans le cadre de ce que nous appellerons le *modèle courant*, lequel recouvre toutes les densités classiques.

Modèle courant de densité : f est positive sur son ensemble de définition et vérifie l'un des deux cas suivants

(i) f est définie et continue sur $\mathbb{R}$ et son intégrale de Riemann généralisée $\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t$ converge et vaut 1. C'est le cas de f_4 ci-dessus.

(ii) f est définie sur $\mathbb{R}$ privé d'un ensemble fini de points $a_1 < \ldots < a_n$. Sur chacun des intervalles ouverts $]-\infty, a_1[$, $]a_i, a_{i+1}[$ $(1 \leq i < n)$, $]a_n, +\infty[$, f est continue et a une intégrale de Riemann (ordinaire ou généralisée) convergente et la somme de toutes ces intégrales vaut 1. Les fonctions f_1, f_2 et f_3 sont dans ce cas.

Pour obtenir des densités d'une forme plus complexe, on peut considérer le modèle suivant.

Modèle plus sophistiqué de densité : f est définie et positive sur l'ensemble $\mathbb{R}\setminus D$ où D est une partie dénombrable de $\mathbb{R}$ (on peut donc considérer les éléments de D comme les termes d'une suite infinie de réels tous distincts[3]). De plus $\mathbb{R}\setminus D$ peut s'écrire comme réunion dénombrable $\cup_{i\in\mathbb{N}}]a_i, b_i[$ d'intervalles ouverts disjoints de $\bar{\mathbb{R}}$. Remarquons que l'on ne suppose pas nécessairement $b_n = a_{n+1}$. On suppose en outre que pour tout $i \in \mathbb{N}$ et tout intervalle $[\alpha, \beta] \subset]a_i, b_i[$, f est Riemann intégrable sur $[\alpha, \beta]$ et l'intégrale (ordinaire ou généralisée) $\int_{a_i}^{b_i} f(t)\,\mathrm{d}t$ converge. Finalement on suppose aussi que

$$\int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t := \sum_{i=0}^{+\infty} \int_{a_i}^{b_i} f(t)\,\mathrm{d}t = 1.$$

Proposition 8.8 *Si la variable aléatoire X a pour densité f, sa fonction de répartition F vérifie :*

(i) $\forall x \in \mathbb{R}$, $F(x) = \int_{-\infty}^{x} f(t)\ \mathrm{d}t$ *;*

(ii) F *est continue sur* $\mathbb{R}$.

(iii) Si f *est continue au point* x_0*, alors* F *est dérivable en ce point et* $F'(x_0) = f(x_0)$.

[3]Il n'est pas toujours possible de choisir la numérotation des éléments de D de façon à ce que cette suite soit monotone. C'est clairement impossible lorsque D a au moins deux points d'accumulation finis

Preuve : Puisque X a pour densité f, on a pour tous réels $a < b$,

$$P(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t)\,\mathrm{d}t. \tag{8.5}$$

Preuve de (i) : Il suffit d'appliquer (8.5) avec $b = x$ fixé et $a = -n$ pour chaque $n \in \mathbb{N}$ tel que $-n < x$. La suite d'événements

$$A_n := \{X \in]-n, x]\}, \quad n > -x,$$

est croissante pour l'inclusion et a pour réunion $A = \{X \in]-\infty, x]\}$. Par continuité monotone séquentielle (cf. proposition 1.2), on a $P(A_n) \uparrow P(A)$, d'où

$$F(x) = P(A) = \lim_{n\to+\infty} P(X \in]-n, x]) = \lim_{n\to+\infty} \int_{-n}^{x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} f(t)\,\mathrm{d}t,$$

en notant que l' intégrale généralisée de la densité f converge en $-\infty$.
Preuve de (ii) : Fixons $x_0 \in \mathbb{R}$ quelconque. On sait déjà que F est continue à droite en tout point comme *toute* fonction de répartition. Il suffit donc de montrer la continuité à gauche en x_0. Nous nous contenterons de le faire sous l'hypothèse additionnelle suivante : « il existe $a < x_0$ tel que f soit définie et Riemann intégrable sur tout intervalle $[a, a'] \subset [a, x_0[$ ». On a alors

$$\lim_{x\uparrow x_0} \int_a^x f(t)\,\mathrm{d}t = \int_a^{x_0} f(t)\,\mathrm{d}t,$$

où la deuxième intégrale est soit une intégrale de Riemann ordinaire soit une intégrale généralisée convergente. Cette relation peut aussi s'écrire à l'aide de F :

$$\lim_{x\uparrow x_0} \big(F(x) - F(a)\big) = F(x_0) - F(a).$$

On en déduit par addition de $F(a)$ que $F(x)$ tend vers $F(x_0)$ quand x tend vers x_0 par valeurs inférieures. L'hypothèse supplémentaire que nous avons introduite est vérifiée par toutes les densités du modèle courant ci-dessus et aussi par les densités du modèle plus général lorsque x_0 n'est pas un point d'accumulation de D.
Preuve de (iii) : Puisque f est continue en x_0, elle est définie sur tout un voisinage de x_0 et donc sur tout un intervalle $]a, b[$ contenant x_0. La continuité de f en x_0 peut alors s'écrire :

$$\forall \varepsilon > 0, \exists]x_0 - \delta, x_0 + \delta[\subset]a, b[; \quad \forall t \in]x_0 - \delta, x_0 + \delta[, \quad |f(t) - f(x_0)| < \varepsilon. \tag{8.6}$$

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

$$|F(x_0 + h) - F(x_0) - hf(x_0)| = \left| \int_{x_0}^{x_0+h} \Big(f(t) - f(x_0)\Big)\,\mathrm{d}t \right| \le h\varepsilon.$$

En divisant par h on voit que F a bien une dérivée en x_0 et que celle ci vaut $f(x_0)$. ■

Remarques :

- Pour toute densité f (au sens de la définition 8.6), il existe une variable aléatoire X ayant f pour densité : il suffit d'appliquer le théorème 8.5 en définissant F par *(i)*.
- D'après *(ii)* toute variable aléatoire à densité a une fonction de répartition continue. La réciproque *est fausse* : il existe des lois à fonction de répartition continue sans densité.
- Par ailleurs si X a une densité, sa fonction de répartition n'est pas forcément dérivable en tout point. Par exemple la densité f_2 ci-dessus a pour fonction de répartition associée $F_2(x) = \sqrt{x}\mathbf{1}_{]0,1]}(x) + \mathbf{1}_{]1,+\infty[}(x)$ (cette écriture condensée signifie que $F_2(x)$ est nul sur $\mathbb{R}^-$, vaut $\sqrt{x}$ entre 0 et 1 et reste constant égal à 1 sur $]1, +\infty[$). F_2 est dérivable en tout point sauf en 0 et en 1.

La proposition suivante donne une règle pratique permettant de trouver la densité (lorsqu'elle existe !) à partir de la fonction de répartition dans les cas les plus courants.

Proposition 8.9 *On suppose que la fonction de répartition F de X est C^1 par morceaux au sens suivant : F est continue sur $\mathbb{R}$ et dérivable sur $\mathbb{R}$ privé (éventuellement) d'un ensemble fini de points $a_1 < \ldots < a_n$. Sur chacun des intervalles ouverts $]-\infty, a_1[$, $]a_i, a_{i+1}[$ $(1 \leq i < n)$, $]a_n, +\infty[$, la dérivée f de F est continue. Alors X a pour densité f.*

Preuve : Il est commode de poser $a_0 := -\infty$ et $a_{n+1} = +\infty$. Sur chacun des intervalles ouverts I découpés par les a_i, F est dérivable et sa dérivée f est continue. On sait alors que f a une infinité de primitives sur I et que si l'on fixe un α dans I, toute primitive H de f sur I est de la forme $H(x) = \int_\alpha^x f(t)\,\mathrm{d}t + C$, avec C constante. Comme F est l'une des primitives de f sur I, en prenant $H = F$ et en faisant $x = \alpha$, on voit que la constante C vaut $F(\alpha)$. On a donc pour α et x quelconques dans I, $F(x) - F(\alpha) = \int_\alpha^x f(t)\,\mathrm{d}t$. Fixons α et prenons $x \geq \alpha$. Faisons tendre x vers la borne supérieure a_i de I. Comme F est continue (ou dans le cas $a_i = +\infty$, F a une limite 1), l'intégrale généralisée $\int_\alpha^{a_i} f(t)\,\mathrm{d}t$ converge et vaut $F(a_i) - F(\alpha)$ (ou $1 - F(\alpha)$ quand $a_i = +\infty$). De même en faisant tendre α vers a_{i-1} on voit que l'intégrale généralisée $\int_{a_{i-1}}^{a_i} f(t)\,\mathrm{d}t$ converge et vaut $F(a_i) - F(a_{i-1})$ (ou $F(a_i)$ quand $a_{i-1} = -\infty$). Finalement soient a et $b > a$ quelconques dans $\mathbb{R}$. Si a et b sont dans le même intervalle I on a directement $F(b) - F(a) = \int_a^b f(t)\,\mathrm{d}t$. Sinon on note $(a_i)_{i_0 \leq i \leq i_1}$ l'ensemble de tous les a_i qui sont dans $[a, b]$ et on écrit

$$F(b)-F(a) = F(a_{i_0})-F(a)+\sum_{i=i_0}^{i_1-1}\big(F(a_{i+1})-F(a_i)\big)+F(b)-F(a_{i_1}) = \int_a^b f(t)\,\mathrm{d}t,$$

en utilisant la relation de Chasles pour les intégrales généralisées. On a donc toujours $P(X \in]a, b]) = F(b) - F(a) = \int_a^b f(t)\,\mathrm{d}t$, ce qui montre que X a pour densité f. ∎

8.3.2 Moments des variables à densité

Définition 8.10 *Si la loi de la variable aléatoire réelle X a une densité f telle que l'intégrale $\int_{-\infty}^{+\infty} |x|f(x)\,\mathrm{d}x$ converge, on appelle espérance de X le réel*

$$\mathbb{E}\,X := \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x.$$

Lorsque les intégrales et les séries concernées sont *absolument* convergentes, on a le tableau comparatif suivant.

	Variable discrète	Variable à densité f
$X(\Omega)$	$\{x_0, x_1, x_2, \dots\}$	$\mathbb{R}$ ou intervalle
$P(a < X \leq b)$	$\displaystyle\sum_{a<x_k\leq b} P(X = x_k)$	$\displaystyle\int_a^b f(t)\,\mathrm{d}t$
$F(x) = P(X \leq x)$	$\displaystyle\sum_{x_k\leq x} P(X = x_k)$	$\displaystyle\int_{-\infty}^x f(t)\,\mathrm{d}t$
$\mathbb{E}\,X$	$\displaystyle\sum_{x_k\in X(\Omega)} x_k P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x$
$\mathbb{E}\,g(X)$	$\displaystyle\sum_{x_k\in X(\Omega)} g(x_k)P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} g(x)f(x)\,\mathrm{d}x$
$\mathbb{E}\,X^2$	$\displaystyle\sum_{x_k\in X(\Omega)} x_k^2 P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x$
$\operatorname{Var} X$	$\displaystyle\sum_{x_k\in X(\Omega)} (x_k - \mathbb{E}\,X)^2 P(X = x_k)$	$\displaystyle\int_{-\infty}^{+\infty} (x - \mathbb{E}\,X)^2 f(x)\,\mathrm{d}x$

Commentaires : Nous admettons les formules donnant $\mathbb{E}\,X^2$ et $\mathbb{E}\,g(X)$ lorsque X est à densité. Dans le cas « courant » où F est C^1 par morceaux, on peut les obtenir par changement de variable. On remarquera l'analogie formelle entre les formules de ce tableau. En gros, on passe des v.a. discrètes aux v.a. à densité en remplaçant les séries par des intégrales et les probabilités ponctuelles $P(X = x)$ par les « probabilités infinitésimales » $f(t)\,\mathrm{d}t$. Cette analogie s'explique par l'existence d'une théorie de l'intégration qui unifie ces deux cas. Cette théorie relève d'un cours de Licence. Elle est indispensable si on veut établir proprement une propriété aussi naturelle que la linéarité de l'espérance. En effet il est facile de trouver des exemples de variables aléatoires X et Y ayant chacune une densité et une espérance mais telles que $X+Y$ n'ait pas de densité (cf. exercice 8.4). Il est donc impossible de démontrer la formule $\mathbb{E}(X+Y) = \mathbb{E}\,X+\mathbb{E}\,Y$ avec la définition 8.10.

8.4 Lois à densité classiques

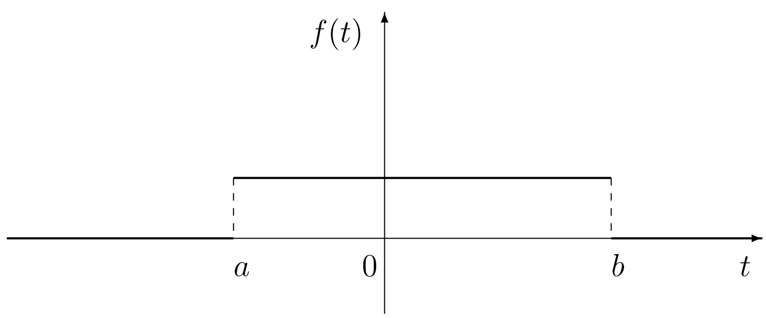
8.4.1 Lois uniformes

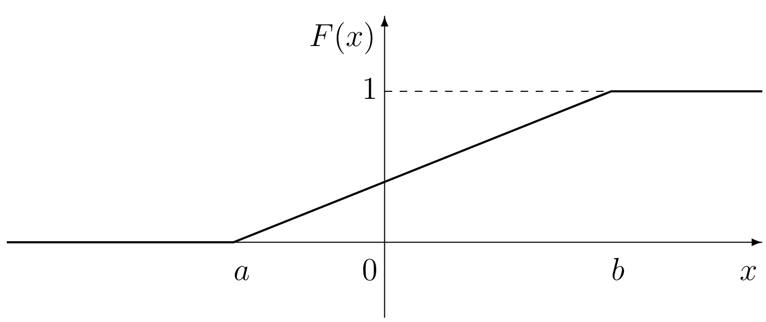
Définition 8.11 *La variable aléatoire réelle X suit la loi uniforme sur l'intervalle $[a, b]$ $(-\infty < a < b < +\infty)$ si elle a une densité f constante sur cet intervalle et nulle en dehors. On a alors*

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

La fonction de répartition F est affine par morceaux :

$$F(x) = \begin{cases} 0 & \text{si} \quad -\infty < x \leq a; \\ \dfrac{x-a}{b-a} & \text{si} \quad a < x \leq b; \\ 1 & \text{si} \quad b < x < +\infty. \end{cases}$$





La règle suivante permet en général d'éviter les calculs d'intégrales (élémentaires mais fastidieux) pour évaluer $P(X \in B)$ lorsque B est une réunion finie ou dénombrable d'intervalles (voir l'exercice 8.2). La preuve est laissée en exercice.

Proposition 8.12 *Si X suit la loi uniforme sur $[a, b]$, alors pour tout intervalle I de $\mathbb{R}$,*

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

où $\ell(J)$ désigne la longueur de l'intervalle J.

En particulier pour $I = [a,b]$ on voit que $P(X \in [a,b]) = 1$. Ainsi une variable aléatoire de loi uniforme est bornée. Elle a donc des moments de tout ordre. Calculons l'espérance et la variance.

Proposition 8.13 *Si X suit la loi uniforme sur $[a,b]$, son espérance et sa variance sont données par :*

$$\mathbb{E}\, X = \frac{a+b}{2}, \qquad \operatorname{Var} X = \frac{(b-a)^2}{12}.$$

La valeur de l'espérance est conforme à l'intuition si l'on se rappelle l'interprétation de l'espérance comme barycentre d'un système de masses : le centre de gravité d'un fil homogène correspondant au segment $[a,b]$ est bien le milieu de ce segment.
Preuve :

$$\mathbb{E}\, X = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \int_a^b \frac{x\,\mathrm{d}x}{b-a} = \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{1}{2(b-a)}(b^2-a^2) = \frac{a+b}{2}.$$

Le moment d'ordre deux se calcule de la même façon.

$$\mathbb{E}\, X^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,\mathrm{d}x = \int_a^b \frac{x^2\,\mathrm{d}x}{b-a} = \left[\frac{x^3}{3(b-a)}\right]_a^b = \frac{b^3-a^3}{b-a} = \frac{1}{3}(b^2+ab+a^2).$$

La variance s'en déduit grâce à la formule de Koenig.

$$\operatorname{Var} X = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = \frac{1}{3}(b^2+ab+a^2) - \frac{(a+b)^2}{4} = \frac{1}{12}(a^2+b^2-2ab).$$

■

Une des raisons de l'importance de la loi uniforme sur $[0,1]$ est le théorème suivant.

Théorème 8.14 *Si X est une variable aléatoire réelle de fonction de répartition continue strictement croissante F et si U est une variable aléatoire de loi uniforme sur $[0,1]$, alors la variable aléatoire $Y := F^{-1}(X)$ a même loi que X.*

Rappelons qu'avoir même loi que X ne signifie aucunement être égale à X. Ce théorème permet de réduire la simulation informatique de la loi de X à celle de U. Le résultat s'étend à toutes les fonctions de répartition, sans hypothèse de continuité ni de croissance stricte (voir l'exercice 8.3).
Preuve : Comme F est continue strictement croissante, c'est une *bijection* de $\mathbb{R}$ sur son image $]0,1[$ (en raison de la stricte monotonie de F, les bornes 0 et 1 ne sont pas atteintes). Par conséquent $F^{-1} :]0,1[\to \mathbb{R}$ est bien définie et vérifie :

$$\forall u \in]0,1[, \forall x \in \mathbb{R}, \quad F^{-1}(u) \leq x \text{ si et seulement si } u \leq F(x).$$

Comme $P(0 < U < 1) = 1$, on en déduit que les évènements $\{F^{-1}(U) \le x\}$ et $\{U \le F(x)\}$ ont même probabilité. Pour obtenir la fonction de répartition de Y, on remarque alors que pour tout $x \in \mathbb{R}$,

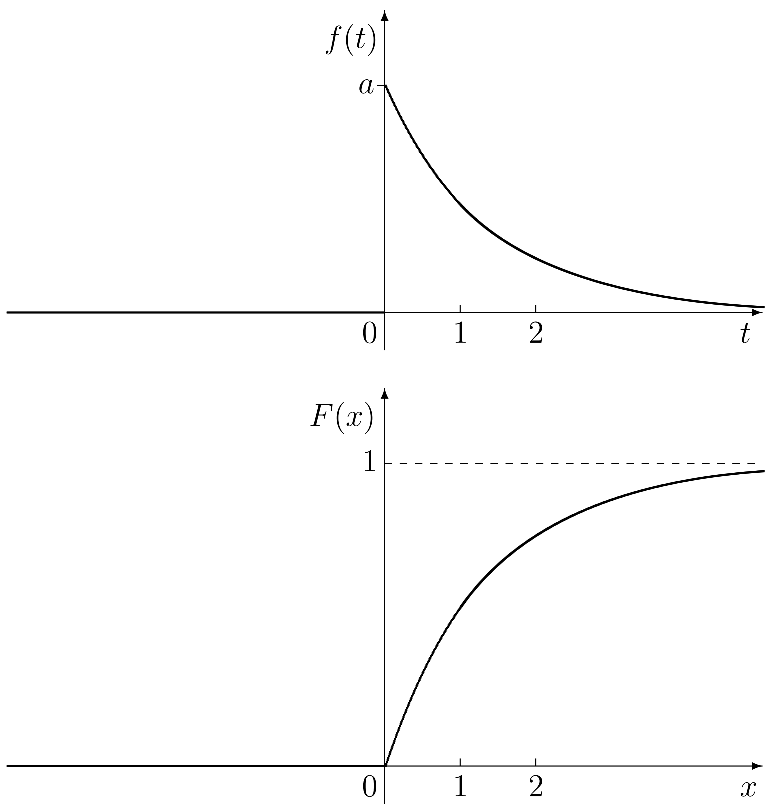
$$P(Y \le x) = P(F^{-1}(U) \le x) = P(U \le F(x)) = \frac{\ell([0, F(x)])}{\ell([0,1])} = F(x).$$

Ainsi Y a pour fonction de répartition F donc a même loi que X. ∎

8.4.2 Lois exponentielles

Définition 8.15 *Soit a un réel strictement positif. La variable aléatoire réelle X suit la loi exponentielle de paramètre a si elle admet pour densité*

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



En pratique, plutôt que de travailler avec la fonction de répartition d'une loi exponentielle, il est plus commode d'utiliser la *fonction de survie* G :

$$G(x) = P(X > x) = 1 - F(x) = \begin{cases} 1 & \text{si } x \le 0, \\ \mathrm{e}^{-ax} & \text{si } x > 0. \end{cases}$$

Les lois exponentielles sont souvent utilisées pour modéliser des temps d'attente. Par exemple, temps d'attente à partir de maintenant du prochain tremblement de terre, du prochain faux numéro sur une ligne téléphonique, de la prochaine désintégration d'un atome de radium, etc.

De même que l'absence de mémoire en temps discret était une propriété caractéristique des lois géométriques (cf. exercice 3.1), l'absence de mémoire en temps continu caractérise les lois exponentielles.

Théorème 8.16 *(i) Si la variable aléatoire X suit une loi exponentielle, alors elle vérifie la propriété d'absence de mémoire :*

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad P(X > t + s \mid X > t) = P(X > s). \tag{8.7}$$

(ii) Réciproquement si une variable aléatoire X vérifie (8.7), alors elle suit une loi exponentielle.

Preuve : Remarquons d'abord que la probabilité conditionnelle dans (8.7) s'exprime commodément à l'aide de la fonction de survie G de la variable aléatoire X (définie par $G(x) := P(X > x)$). En effet, s étant positif, on a $t + s \geq t$ d'où l'inclusion de $\{X > t + s\}$ dans $\{X > t\}$ et l'égalité d'évènements : $\{X > t + s\} \cap \{X > t\} = \{X > t + s\}$. On en déduit

$$P(X > t + s \mid X > t) = \frac{P(X > t + s)}{P(X > t)} = \frac{G(t + s)}{G(t)}. \tag{8.8}$$

Preuve de (i) : Si X suit la loi exponentielle de paramètre a, on a $G(x) = \mathrm{e}^{-ax}$ pour tout x positif et (8.8) se traduit alors par :

$$P(X > t + s \mid X > t) = \frac{\mathrm{e}^{-a(t+s)}}{\mathrm{e}^{-at}} = \mathrm{e}^{-as} = P(X > s).$$

Preuve de (ii) : Soit X une variable aléatoire dont la loi vérifie (8.7) et G sa fonction de survie. Comme $G = 1 - F$ (où F désigne la fonction de répartition de X), G est décroissante et continue à droite et tend vers 0 en $+\infty$. De plus l'écriture de (8.7) suppose implicitement que $G(t) > 0$ pour tout $t \geq 0$ car sinon $P(\,.\, \mid X > t)$ ne serait pas définie. Grâce à (8.8), on voit que la propriété d'absence de mémoire (8.7) équivaut à

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad \frac{G(t + s)}{G(t)} = G(s).$$

La fonction de survie G doit donc être une solution décroissante, continue à droite, tendant vers 0 en $+\infty$ et telle que $0 < G(t) \leq 1$ de l'équation fonctionnelle[4] :

$$\forall s \in \mathbb{R}^+, \forall t \in \mathbb{R}^+, \quad G(t + s) = G(t)G(s). \tag{8.9}$$

[4]Une équation fonctionnelle est une équation dont l'inconnue est. . .une fonction ! Les équations différentielles sont des exemples bien connus d'équations fonctionnelles.

En faisant $s = t = 0$ dans (8.9), on obtient $G(0) = G(0)^2$ et comme $G(0) > 0$, on a

$$G(0) = 1. \tag{8.10}$$

En faisant $s = t$ dans (8.9), on obtient $G(2t) = G(t)^2$, puis de proche en proche

$$\forall n \in \mathbb{N}^*, \forall t \geq 0, \quad G(nt) = G(t)^n. \tag{8.11}$$

En particulier pour $t = 1/d$, $d \in \mathbb{N}^*$:

$$\forall n \in \mathbb{N}^*, \forall d \in \mathbb{N}^*, \quad G\Big(\frac{n}{d}\Big) = G\Big(\frac{1}{d}\Big)^n. \tag{8.12}$$

Lorsque $n = d$, (8.12) donne $G(1) = G(1/d)^d$ d'où

$$\forall d \in \mathbb{N}^*, \quad G\Big(\frac{1}{d}\Big) = G(1)^{1/d}. \tag{8.13}$$

Nous connaissons maintenant G sur l'ensemble des rationnels positifs puisque (8.10), (8.11), (8.12) et (8.13) nous donnent

$$\forall r \in \mathbb{Q}^+, \quad G(r) = G(1)^r. \tag{8.14}$$

Soit $x \in \mathbb{R}^+ \setminus \mathbb{Q}^+$, x est limite d'une suite décroissante (r_n) de rationnels. Comme G est continue à droite, $G(r_n)$ converge vers $G(x)$. D'autre part l'application $y \mapsto G(1)^y$ est continue sur $\mathbb{R}$. Ainsi en appliquant (8.14) à r_n et en faisant tendre n vers l'infini on obtient

$$\forall x \in \mathbb{R}^+, \quad G(x) = G(1)^x. \tag{8.15}$$

A priori la constante $G(1)$ est dans $]0, 1]$. On peut écarter la valeur $G(1) = 1$ car sinon d'après (8.15), la limite en $+\infty$ de G serait 1 alors qu'elle vaut 0.

Finalement, puisque $0 < G(1) < 1$, on peut poser $G(1) = \mathrm{e}^{-a}$ pour un réel $a > 0$ (cela revient à prendre $a = -\ln G(1)$). On peut alors réécrire (8.15) sous la forme

$$\forall x \in \mathbb{R}^+, \quad G(x) = \mathrm{e}^{-ax}.$$

La fonction de survie G est donc la même que celle de la loi exponentielle de paramètre a, donc X suit cette loi (puisque la fonction de survie caractérise la loi au même titre que la fonction de répartition). ∎

8.4.3 Lois gaussiennes

Nous avons déjà rencontré ces lois lors de l'étude du théorème de De Moivre Laplace. Elles jouent un rôle capital dans l'étude des lois limites de sommes de variables aléatoires indépendantes.

Définition 8.17 *On dit que la variable aléatoire X suit la loi gaussienne ou normale $\mathcal{N}(m,\sigma)$ si elle a pour densité la fonction :*

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

La loi $\mathcal{N}(0,1)$ est appelée loi normale standard.

Tous les calculs de probabilités concernant une variable aléatoire de loi $\mathcal{N}(m,\sigma)$ peuvent se ramener à des calculs sur une variable de loi normale standard.

Proposition 8.18 *Si la variable aléatoire X suit la loi $\mathcal{N}(m,\sigma)$, alors $Y := (X-m)/\sigma$ suit la loi $\mathcal{N}(0,1)$.*

Preuve : On calcule $P(a < Y \leq b)$ pour a et b réels quelconques $(a < b)$.

$$\begin{aligned} P\Big(a < \frac{X-m}{\sigma} \leq b\Big) &= P(\sigma a + m < X \leq \sigma b + m) \\ &= \int_{\sigma a+m}^{\sigma b+m} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \mathrm{d}x. \end{aligned}$$

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

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

Donc Y a bien la densité $f_{0,1}$. ■

En raison de la décroissance rapide de l'exponentielle, il est clair que les variables gaussiennes ont des moments de tout ordre. L'interprétation des paramètres m et σ est très simple.

Proposition 8.19 *Si la variable aléatoire X suit la loi $\mathcal{N}(m,\sigma)$,*

$$\mathbb{E}\, X = m, \qquad \operatorname{Var} X = \sigma^2.$$

Preuve : La variable aléatoire $Y := (X-m)/\sigma$ suit la loi $\mathcal{N}(0,1)$ et l'on a $X = \sigma Y + m$. Il suffit donc de montrer que $\mathbb{E}\, Y = 0$ et $\operatorname{Var} Y = \mathbb{E}\, Y^2 = 1$. L'espérance de Y s'écrit

$$\mathbb{E}\, Y = \int_{-\infty}^{+\infty} \frac{y}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y.$$

La convergence de cette intégrale généralisée étant acquise, il est clair qu'elle vaut 0 en raison de l'imparité de l'intégrande (la fonction sous le signe somme). On a alors $\operatorname{Var} Y = \mathbb{E}\, Y^2$ et il nous faut vérifier que

$$\mathbb{E}\, Y^2 = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \mathrm{d}y = 1.$$

Une façon de le voir est d'intégrer par parties la densité $f_{0,1}$. Pour celà considérons pour $a > 0$ l'intégrale

$$I(a) := \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \, \mathrm{d}y$$

et posons $u = \exp(-y^2/2)$, $\mathrm{d}v = (2\pi)^{-1/2} \, \mathrm{d}y$, $\mathrm{d}u = -y \exp(-y^2/2) \, \mathrm{d}y$ et $v = (2\pi)^{-1/2} y$. Il vient

$$\begin{aligned} I(a) &= \left[\frac{y}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right)\right]_{-a}^{a} - \int_{-a}^{a} \frac{-y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \, \mathrm{d}y \\ &= \frac{2a}{\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right) + \int_{-a}^{a} \frac{y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \, \mathrm{d}y. \end{aligned}$$

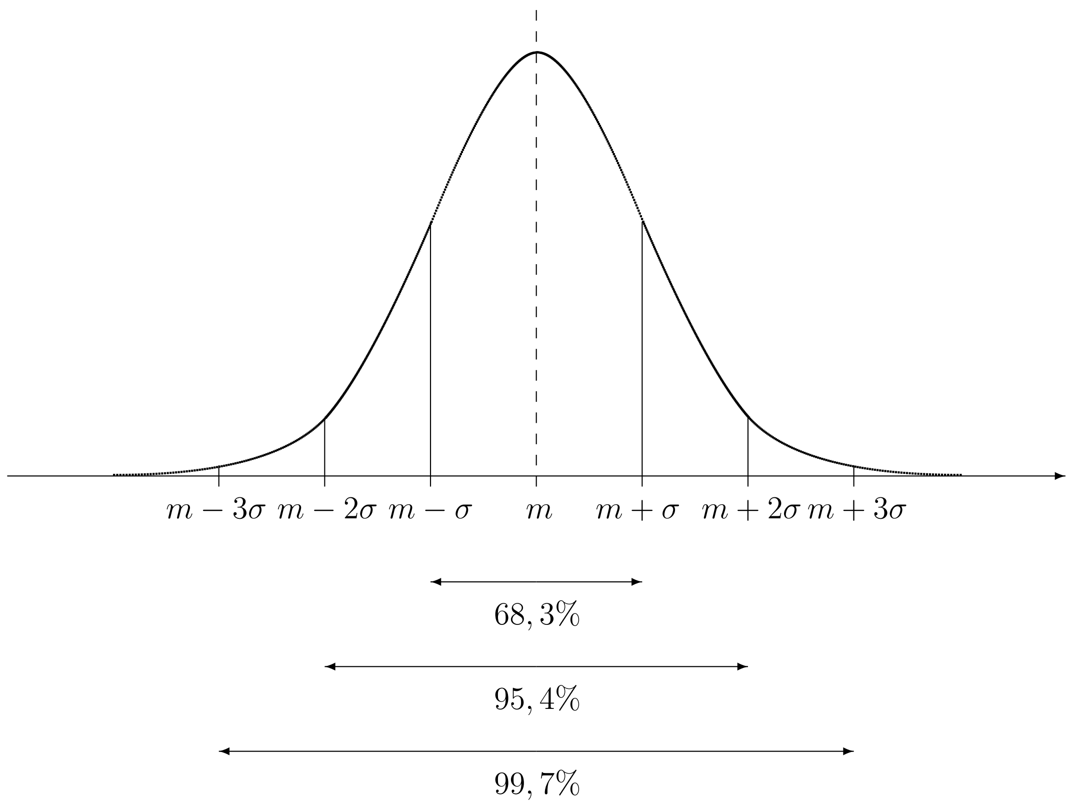
En faisant tendre a vers l'infini[5] dans cette égalité, on obtient

$$1 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \, \mathrm{d}y = \int_{-\infty}^{+\infty} \frac{y^2}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) \, \mathrm{d}y.$$

■

La figure suivante illustre la signification du paramètre de position m et du paramètre de dispersion σ pour la loi gaussienne $\mathcal{N}(m, \sigma)$.

[5]Pour aller plus vite, on a intégré sur un intervalle symétrique $[-a, +a]$ parce que l'on sait déjà que les intégrales généralisées concernées sont convergentes. Si l'on voulait se servir de ce calcul pour montrer leur convergence, il faudrait bien sûr intégrer sur un intervalle $[-a, b]$ et faire tendre a et b séparément vers $+\infty$.



Cette concentration de pratiquement toute la probabilité dans l'intervalle $[m - 3\sigma, m + 3\sigma]$ permet l'utilisation des lois gaussiennes pour modéliser des grandeurs aléatoires qui *a priori* prennent leurs valeurs seulement dans un petit intervalle de $\mathbb{R}^+$: taille, poids, ..., même si théoriquement une variable gaussienne peut prendre toute valeur entre $-\infty$ et $+\infty$.

8.5 Exercices

Ex 8.1. Le temps d'attente (en minutes) pour accéder à des données suit une loi uniforme $\mathcal{U}[1, 6]$.

1) Déterminer la probabilité d'attendre au moins 4 minutes.
2) Déterminer le temps d'attente moyen.

Ex 8.2. Un arrêt de bus est desservi tous les quart d'heures à partir de 7 h du matin (inclus). Un passager arrive à l'arrêt à un instant aléatoire de loi uniforme sur [7h ; 7h30]. Quelle est la probabilité qu'il attende moins de 5 mn pour un bus ? plus de 10mn ?

Ex 8.3. Soit F une fonction de répartition. On définit sur $]0,1[$ son *inverse généralisée* F^{-1} par

$$F^{-1}(u) := \inf\{x \in \mathbb{R};\ F(x) \geq u\}, \quad u \in]0,1[.$$

1) Pourquoi a-t-on exclu les cas particuliers $u = 0$ et $u = 1$ de la définition ci-dessus ? Que peut-on dire des limites de F^{-1} en 0 et en 1 ?

2) Pour $u \in]0,1[$, on pose $I_u := \{x \in \mathbb{R};\ F(x) = u\}$. Montrer que I_u est un intervalle fermé à gauche ou l'ensemble vide.

3) Donner la construction graphique de $F^{-1}(u)$ dans chacun des cas possibles pour I_u : $I_u = \emptyset$, $I_u = \{x_0\}$, $I_u = [x_0, x_1[$, $I_u = [x_0, x_1]$.

4) Représenter F^{-1} lorsque F est en escaliers.

5) Montrer que les inégalités $F^{-1}(u) \leq x$ et $u \leq F(x)$ sont équivalentes en justifiant chacune des implications ci-dessous. Dans un sens,

$$u \leq F(x) \Rightarrow F^{-1}(u) \leq x.$$

Dans l'autre sens,

$$F^{-1}(u) \leq x \Rightarrow \forall \varepsilon > 0, F(x+\varepsilon) \geq u \Rightarrow F(x) \geq u.$$

6) Soit X une variable aléatoire de fonction de répartition F et U une variable aléatoire de loi uniforme sur $[0,1]$. Montrer que X et $F^{-1}(U)$ ont même loi.

Ex 8.4. Soit X une variable aléatoire de loi uniforme sur $[0,1]$ et $Y := 1 - X$. Trouver la fonction de répartition de Y et en déduire sa loi. Ceci fournit un exemple élémentaire de deux variables à densité dont la somme est une variable discrète.

Ex 8.5. Soit X une variable aléatoire de loi uniforme sur $[0,1]$ et Y la variable aléatoire définie sur le même espace Ω par

$$Y(\omega) := \begin{cases} X(\omega) & \text{si } \ X(\omega) \in [0,1/4] \cup [3/4,1]; \\ 1 - X(\omega) & \text{si } \ X(\omega) \in]1/4, 3/4[. \end{cases}$$

Quelle est la loi de Y ? Trouver la fonction de répartition de la variable aléatoire $Z := X + Y$ et vérifier que Z n'est ni discrète ni à densité.

Ex 8.6. Soit X une variable aléatoire de densité :

$$f(x) = \begin{cases} 2x/\theta^2 & \text{si } x \in [0;\theta], \\ 0 & \text{sinon.} \end{cases}$$

1) Calculer $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ puis $\mathrm{Var}[X]$ en fonction de θ.

2) Calculer $P(X < \theta/4)$.

Ex 8.7. *Consommation d'eau*
La consommation journalière en eau d'une agglomération au cours du mois de juillet est une variable aléatoire X dont la densité f a la forme :

$$f(t) = c(t-a)(b-t)\mathbf{1}_{[a,b]}(t), \quad t \in \mathbb{R},$$

où a, b, c sont des constantes strictement positives ($a < b$) et $\mathbf{1}_{[a,b]}(t) = 1$ si $t \in [a,b]$ et 0 sinon.

1) Vérifier que l'on a pour tout $n \in \mathbb{N}$:

$$\int_a^b (t-a)^n (b-t)\,dt = \frac{(b-a)^{n+2}}{(n+1)(n+2)}.$$

2) Exprimer la constante c en fonction de a et b.

3) Calculer $\mathbb{E}(X-a)$ et $\mathbb{E}\big[(X-a)^2\big]$. En déduire $\mathbb{E}\,X$ et $\operatorname{Var} X$.

4) Donner la fonction de répartition F de la variable aléatoire X : on distinguera pour le calcul de $F(x)$ les cas $x < a$, $a \leq x \leq b$ et $x > b$ et, dans le deuxième cas, on écrira $F(x)$ en fonction de $(x-a)$ et $(b-x)$ sans développer ni réduire le polynôme obtenu. Donner l'allure des représentations graphiques de f et F. Proposer une interprétation physique des constantes a et b.

5) En notant X_i la consommation du i-ème jour et en supposant que les X_i sont indépendantes et de même loi que X, exprimer à l'aide de F et de n la fonction de répartition de la variable aléatoire $M_n = \max\limits_{1\leq i\leq n} X_i$. *Indication :* On commencera par écrire l'événement $\{M_n \leq x\}$ en fonction des événements $\{X_i \leq x\}$.

6) En fait la ville est alimentée en eau par un canal qui peut fournir au maximum une quantité journalière d'eau $x_0 = a + 0.9(b-a)$ et par un réservoir de sécurité dans lequel elle peut puiser en cas de trop forte demande. Calculer numériquement la probabilité qu'au cours des 31 jours du mois de juillet, on ne fasse jamais usage du réservoir de sécurité (le résultat ne dépend ni de a ni de b).

Ex 8.8. Soit X une variable aléatoire suivant la loi exponentielle de paramètre a. On lui associe la variable aléatoire discrète $Y = [X]$, où les crochets désignent la partie entière. Quelle est la loi de Y ? Quelle est la loi de la partie fractionnaire $Z := X - [X]$?

Ex 8.9. *Calcul des moments de la loi gaussienne standard*
Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$.

1) Que valent les $\mathbb{E}\,X^{2n+1}$ pour $n \in \mathbb{N}$?

2) On pose $c_n = \mathbb{E}\,X^{2n}$. Montrer en intégrant par parties que

$$\forall n \in \mathbb{N}, \quad c_n = \frac{1}{2n+1} c_{n+1}.$$

en déduire une formule explicite pour $\mathbb{E}\,X^{2n}$.

Ex 8.10. *Une formule simple pour la queue de la loi normale*
Soit X une variable aléatoire de loi $\mathcal{N}(0,1)$.

1) En utilisant le changement de variable $t = x + s$ dans la relation

$$P(X \geq x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \,\mathrm{d}t,$$

vérifier que l' on a pour tout $x \geq 0$,

$$P(X \geq x) \leq \frac{1}{2} \exp\bigl(-x^2/2\bigr), \qquad P(|X| \geq x) \leq \exp\bigl(-x^2/2\bigr).$$

Ces majorations sont moins bonnes que le lemme 7.2 pour $x \geq 2/\sqrt{2\pi} \simeq 0{,}798$. Elles ont l'avantage d'être plus faciles à mémoriser et de donner pour les petites valeurs de x un majorant inférieur à 1.

2) Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de même loi que X. Montrer que pour tout $r \geq 0$,

$$P\Bigl(\min_{1\leq i\leq n} |X_i| \geq \frac{r}{\sqrt{n}}\Bigr) \leq \exp\bigl(-r^2/2\bigr).$$

3) Montrer de même que pour tout $c > 1$,

$$\lim_{n\to+\infty} P\Bigl(\max_{1\leq i\leq n} |X_i| < c\sqrt{2\ln n}\Bigr) = 1.$$

Annexe A

Ensembles et dénombrements

A.1 Généralités

Soit Ω un ensemble ; A est un *sous-ensemble* (ou une *partie*) de Ω si tout élément de A est aussi un élément de Ω ($\forall \omega \in A$, $\omega \in \Omega$). On note $A \subset \Omega$. On appelle $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω ce que l'on peut noter[1]

$$\mathcal{P}(\Omega) = \{A; A \subset \Omega\}.$$

ainsi les écritures $A \subset \Omega$ et $A \in \mathcal{P}(\Omega)$ sont deux façons de dire la même chose[2].

Si A et B sont deux parties du même ensemble Ω, on dit que A est *inclus* dans B (notation $A \subset B$) si tout élément de A est aussi élément de B ($\forall \omega \in A$, $\omega \in B$), autrement dit, si l'appartenance à A *implique* l'appartenance à B :

$$A \subset B \quad \text{signifie} \quad \forall \omega \in \Omega, (\omega \in A) \Rightarrow (\omega \in B).$$

Soit I un ensemble quelconque d'indices (fini ou infini) et $(A_i)_{i \in I}$ une famille de parties de Ω. On définit son intersection $\underset{i \in I}{\cap} A_i$ et sa réunion, $\underset{i \in I}{\cup} A_i$ par :

$$\underset{i \in I}{\cap} A_i = \{\omega; \forall i \in I, \omega \in A_i\} \qquad \text{et} \qquad \underset{i \in I}{\cup} A_i = \{\omega; \exists i \in I, \omega \in A_i\}.$$

Remarque : La réunion et l'intersection d'une famille de parties de Ω sont définies de façon globale, elles s'obtiennent d'un coup, sans passage à la limite quand I est infini et sans qu'un ordre éventuel sur l'ensemble d'indices I n'ait d'importance. Dans la définition de la propriété de σ-additivité :

$$P\Big(\underset{i \in \mathbb{N}^*}{\cup} A_i\Big) = \sum_{i=1}^{+\infty} P(A_i),$$

[1]Dans toutes les écritures d'ensembles entre accolades, nous utilisons le point virgule au sens de « tel que ».

[2]Noter cependant la différence de statut de A : dans la première écriture, A est considéré comme un ensemble, dans la deuxième comme un élément d'un ensemble d'un type un peu particulier.

le premier membre s'obtient sans passage à la limite, tandis que le deuxième résulte d'un passage à la limite. Cette définition reste cohérente parce que pour une série à termes *positifs*, la valeur de la somme ne dépend pas de l'ordre de sommation. Si on voulait définir la σ-additivité pour une fonction d'ensemble pouvant prendre des valeurs positives et négatives, il faudrait imposer en plus que la série ci-dessus soit *absolument convergente*.

Réunion et intersection sont très utiles pour la la traduction automatique des quantificateurs. Si I est un ensemble quelconque d'indices, (π_i) une propriété dépendant de l'indice i et A_i l'ensemble des $\omega \in \Omega$ vérifiant (π_i), on a :

$$\begin{aligned} \{\omega \in \Omega;\ \forall i \in I,\ \omega \text{ vérifie } (\pi_i)\} &= \bigcap_{i\in I} A_i \\ \{\omega \in \Omega;\ \exists i = i(\omega) \in I,\ \omega \text{ vérifie } (\pi_i)\} &= \bigcup_{i\in I} A_i \end{aligned}$$

Ainsi le quantificateur $\forall$ peut toujours se traduire par une intersection et le quantificateur $\exists$ par une réunion. (Pour une illustration, voir le chapitre 6, p. 123).

L'intersection et l'union sont distributives l'une par rapport à l'autre, c'est à dire

$$B \cup \Big(\cap_{i\in I} A_i \Big) = \cap_{i\in I} (A_i \cup B) \qquad B \cap \Big(\cup_{i\in I} A_i \Big) = \cup_{i\in I} (A_i \cap B).$$

Le complémentaire de A (dans Ω) est l'ensemble $A^c := \{\omega \in \Omega; \omega \notin A\}$. L'opération passage au complémentaire (qui est une bijection de $\mathcal{P}(\Omega)$ dans lui même) vérifie $(A^c)^c = A$, $\Omega^c = \emptyset$, $\emptyset^c = \Omega$ et échange réunions et intersections grâce aux très utiles formules :

$$\Big(\cap_{i\in I} A_i \Big)^c = \cup_{i\in I} A_i^c \qquad \Big(\cup_{i\in I} A_i \Big)^c = \cap_{i\in I} A_i^c.$$

On définit le produit cartésien de deux ensembles E et F noté $E \times F$ par :

$$E \times F := \{(x, y); x \in E, y \in F\}.$$

Attention dans cette écriture (x, y) ne désigne en aucune façon un ensemble mais un couple d'éléments (l'ordre d'écriture a une importance). Pour éviter toute confusion, on utilise des accolades pour la description des ensembles et des parenthèses pour les couples , triplets, etc, d'éléments.

L'ensemble $E^2 = E \times E = \{(x_1, x_2); x_1 \in E, x_2 \in E\}$ peut être utilisé pour représenter l'ensemble de toutes les applications de $\{1, 2\}$ dans E (le couple (x_1, x_2) correspondant à l'application $f : \{1, 2\} \to E$ définie par $f(1) = x_1$ et $f(2) = x_2$). Il pourrait de la même façon représenter les applications d'un ensemble à deux éléments dans E (remplacer les chiffres 1 et 2 par n'importe quelle paire de symboles distincts : 0 et 1, a et b, etc.).

Plus généralement, pour $n \geq 2$, E^n est l'ensemble des n-uplets ou listes de longueur n d'éléments de E. Dans un n-uplet $(x_1, \dots, x_n)$, il peut y avoir des répétitions. On peut aussi utiliser E^n pour représenter toutes les applications de l'ensemble $\{1, \dots, n\}$ (ou de n'importe quel ensemble à n éléments) dans E.

Soit I un ensemble quelconque, fini ou infini. Par analogie avec ce qui précède, l'ensemble de toutes les applications $f : I \to E$ sera noté E^I. Par exemple avec $E = \{0, 1\}$ et $I = \mathbb{N}$, on obtient l'ensemble $\{0, 1\}^{\mathbb{N}}$ de toutes les suites de chiffres binaires indexées par $\mathbb{N}$: $\{0, 1\}^{\mathbb{N}} = \{u = (u_i)_{i \in \mathbb{N}}; u_i = 0 \text{ ou } 1\}$. Avec $E = \mathbb{R}$ et $I = [0, 1]$, on obtient l'ensemble $\mathbb{R}^{[0,1]}$ des fonctions définies sur l'intervalle $[0, 1]$ et à valeurs dans $\mathbb{R}$.

A.2 Ensembles finis

Soit E un ensemble fini. Le cardinal de E, noté $\operatorname{card} E$ est le nombre de ses éléments.

Proposition A.1 *Soient E et F deux ensembles finis. Leur produit cartésien a pour cardinal*

$$\operatorname{card}(E \times F) = \operatorname{card} E \times \operatorname{card} F.$$

Proposition A.2 *(a) Si $\operatorname{card} E = n$ et $\operatorname{card} F = p$, l'ensemble F^E des applications de E dans F est fini et a pour cardinal p^n, autrement dit :*

$$\operatorname{card}\bigl(F^E\bigr) = (\operatorname{card} F)^{\operatorname{card} E}.$$

(b) Comme $\mathcal{P}(E)$ est en bijection avec l'ensemble $\{0, 1\}^E$ des applications de E dans $\{0, 1\}$,

$$\operatorname{card} \mathcal{P}(E) = 2^n = 2^{\operatorname{card} E}.$$

Une bijection naturelle entre $\mathcal{P}(E)$ et $\{0, 1\}^E$ est l'application φ qui à toute partie A de E associe sa fonction indicatrice :

$$\varphi : \mathcal{P}(E) \to \{0, 1\}^E \qquad A \mapsto \varphi(A) := \mathbf{1}_A.$$

Rappelons que l'indicatrice d'une partie A de E est l'application

$$\mathbf{1}_A : E \to \{0, 1\} \qquad \omega \mapsto \mathbf{1}_A(\omega) := \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Si E est un ensemble de cardinal n et p un entier tel que $1 \leq p \leq n$, on appelle *arrangement* de p éléments de E tout p-uplet $(x_1, x_2, \dots, x_n)$ d'éléments *tous distincts* de E. Un tel arrangement représente une *injection* de $\{1, \dots, p\}$ dans E.

Proposition A.3 *Le nombre d'arrangements de p éléments de E ($1 \leq p \leq n = \operatorname{card} E$) est*

$$A_n^p = n(n-1)(n-2)\cdots(n-p+1) = \frac{n!}{(n-p)!}.$$

A_n^p est aussi le nombre d'injections d'un ensemble I de cardinal p (par exemple $\{1,\dots,p\}$) dans E. En particulier pour $I = E$ (et donc $p = n$), on obtient le nombre de bijections de E dans lui même (appelées aussi permutations *de E) :*

$$\text{nombre de permutations de } E = A_n^n = n!$$

On appelle *combinaison* de p éléments de E ($1 \leq p \leq n = \operatorname{card} E$) toute partie de cardinal p de E. Une telle combinaison, comme un arrangement a tous ses éléments distincts, mais l'odre d'écriture n'a pas d'importance.

Proposition A.4 *Le nombre de combinaisons de p éléments de E ($1 \leq p \leq n = \operatorname{card} E$) est*

$$C_n^p = \frac{n(n-1)(n-2)\cdots(n-p+1)}{p(p-1)\cdots 1} = \frac{n!}{p!(n-p)!}.$$

Bibliographie

[1] P. Barbe et M. Ledoux. *Probabilité*. Espaces 34, Belin, 1998.

[2] P. Billingsley. *Probability and measure*. Wiley, 1979.

[3] E. Borel, *Probabilité et certitude*. Que sais-je ? N° 445 P.U.F.

[4] W. Feller, *An Introduction to Probability Theory and its Applications*, Vol. I. Wiley.

[5] D. Foata et A. Fuchs, *Calcul des Probabilités*. Dunod, 1998.

[6] *Le hasard*, numéro hors-série de *Pour la Science*, avril 1996.

[7] S. M. Ross, *Initiation aux probabilités*. Presses polytechniques et universitaires romandes, Lausanne.

[8] D. Ruelle, *Hasard et Chaos*. Odile Jacob, 1991.

[9] J. V. Uspensky, *Introduction to mathematical probability*. McGraw-Hill, 1937.