

MAT1720 – Introduction aux probabilités

Élise Davignon

Email address: elise.davignon@umontreal.ca

RÉSUMÉ. Ces notes de cours accompagnent le cours MAT1720 – Introduction aux probabilités, enseigné au département de Mathématiques et Statistiques de la faculté des arts et sciences de l'Université de Montréal.

Création le : 18 décembre 2019.

Dernière compilation le : 14 juillet 2023

Lien permanent vers la version la plus à jour :

https://www.dms.umontreal.ca/~davignon/files/notes_prob/notes_de_cours.pdf



Léa Stréiski @LeaStreliski · 20 min

Le lol > les mathématiques.

[Afficher cette discussion](#)

« Le lol > les mathématiques. »

– Léa Stréiski, 3 octobre 2022, via Twitter.

Capture d'écran faite le 3 octobre 2022.

Table des matières

Introduction	1
Chapitre 1. L'analyse combinatoire	3
1.1. Principes de base du dénombrement	3
1.2. Problèmes courants	6
1.3. Exercices	14
partie 1. Les espaces de probabilités.	19
Chapitre 2. Les axiomes des probabilités	21
2.1. L'ensemble fondamental et les événements	21
2.2. La mesure de probabilité	23
2.3. Propriétés de la mesure de probabilités	25
2.4. La mesure équiprobable	27
2.5. <i>Bonus</i> : Les limites de suites d'événements.	32
2.6. Le mot « presque »	36
2.7. Exercices	37
Chapitre 3. Probabilités conditionnelles et indépendance	41
3.1. Probabilités conditionnelles	41
3.2. La formule de probabilité totale	43
3.3. La formule de Bayes	47
3.4. Les événements indépendants	49
3.5. La probabilité conditionnelle comme nouvelle mesure	54
3.6. Exercices	58
partie 2. Les variables aléatoires	63
Chapitre 4. Les variables aléatoires	65
4.1. Les variables aléatoires	65
4.2. Les variables aléatoires discrètes	69
4.3. L'espérance	72
4.4. Les fonctions de variables aléatoires	74
4.5. La variance	77
4.6. Plusieurs distributions discrètes	78
4.7. Espérance de sommes de variables aléatoires	92
4.8. Exercices	95
Chapitre 5. Les variables aléatoires continues	99
5.1. Fonctions de répartition et densité	99

5.2. Espérance et Variance	104
5.3. Plusieurs distributions courantes	108
5.4. Fonctions d'une variable aléatoire continue	122
5.5. <i>Bonus</i> : les distributions mixtes.	124
5.6. Exercices	127
Chapitre 6. Variables aléatoires simultanées	131
6.1. Fonctions de répartition et densités jointes	131
6.2. Variables aléatoires indépendantes	138
6.3. Sommes de variables aléatoires indépendantes	144
6.4. Distributions conditionnelles	152
6.5. Statistiques d'ordre	157
6.6. Fonctions de vecteurs aléatoires et changements de variables	162
6.7. Variables aléatoires interchangeables	167
6.8. Exercices	172
partie 3. Les outils analytiques	179
Chapitre 7. L'espérance	181
7.1. L'espérance de fonctions de plusieurs variables aléatoires	181
7.2. Le nombre d'événements et les indicatrices	192
7.3. Espérances conditionnelles	195
7.4. Fonctions génératrices	209
7.5. Vecteurs gaussiens	222
7.6. Exercices	227
Chapitre 8. Théorèmes-limites	235
8.1. Bornes et inégalités	235
8.2. Les types de convergence	238
8.3. La loi faible des grands nombres	243
8.4. La loi forte des grands nombres	243
8.5. Le théorème de la limite centrale	246
8.6. Exercices	250
Annexes	253
Annexe A. Théorie des ensembles	255
A.1. Description d'ensembles	255
A.2. Opérations ensemblistes	256
A.3. Le produit cartésien	258
A.4. Propriétés des opérations	258
A.5. Ensembles et fonctions	259
A.6. <i>Bonus</i> : La cardinalité des ensembles	260
A.7. <i>Bonus</i> : Les relations d'ordre	260
A.8. <i>Bonus</i> : Limites d'ensembles	262
Annexe B. Rappels d'analyse	265
B.1. Structure des nombres réels	265
B.2. Un peu de topologie	266

B.3.	Suites et séries	267
B.4.	Fonctions	269
B.5.	Dérivation	270
B.6.	Intégration au sens de Riemann	271
B.7.	Symétries de fonctions.	272
B.8.	Les fonctions spéciales	273
B.9.	Les convolutions	274
B.10.	Les fonctions convexes	274
B.11.	Les notations asymptotiques	274
B.12.	<i>Bonus : L'intégrale de Stieltjes</i>	275

Introduction

Ce document contient les notes de cours pour le cours MAT1720 – Introduction aux probabilités. Il s'agit d'un cours de première année au Baccalauréat en Mathématiques au département de l'université de Montréal.

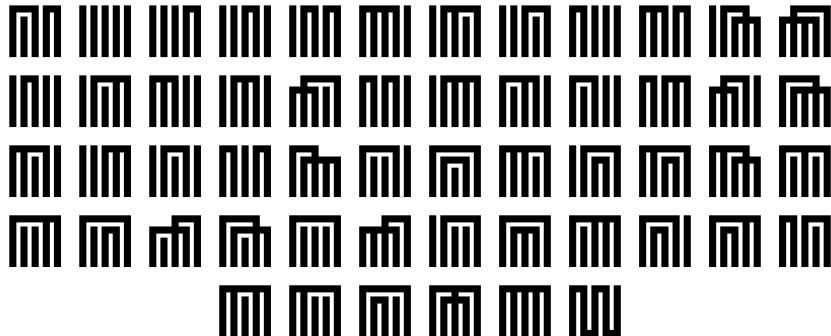
Le contenu est inspiré de celui présenté dans l'ouvrage *A First Course In Probability*, de Sheldon M. Ross. En particulier, les chapitres correspondent (à peu de choses près), et les exercices sont souvent tirés ou inspirés de ceux disponibles en fin de chapitre dans l'ouvrage de Ross.

REMARQUE. Veuillez noter que ce document est un chantier en cours, et sera probablement modifié au cours de la session ; la version la plus à jour sera toujours sur StudiUM, mais peut aussi être récupérée, en tout temps, à l'adresse https://www.dms.umontreal.ca/~davignon/files/notes_prob/notes_de_cours.pdf.

Date de compilation : 14 juillet 2023

Chapitre 1

L'analyse combinatoire



Combien de façon y a-t-il de partitionner un ensemble de cinq éléments ? *Le Dit du Genji* est une œuvre littéraire japonaise datant du XIe siècle, divisée en 54 *jō* (livres, ou chapitres). Chaque livre (sauf le dernier) porte un symbole représentant l'une des partitions possibles de 5 éléments. Mais attention ! Il y a un doublon...

Ce court chapitre porte sur l'analyse combinatoire. Il s'agit essentiellement d'un survol de quelques principes d'analyse combinatoire qui seront utiles dans le reste du cours.

1.1. Principes de base du dénombrement

L'analyse combinatoire vise à dénombrer les éléments d'un ensemble. Un problème typique d'analyse combinatoire serait comme suit :

EXEMPLE 1.1. Combien de mains différentes de cinq cartes à jouer ordinaires peut-on former à partir d'un paquet standard de 52 cartes ? Combien contiennent une paire ?

La solution la plus naïve consisterait à énumérer toutes les possibilités, une par une. Toutefois, il existe différentes techniques, toutes fondées sur des principes très simples, et qui permettent de simplifier de tels calculs.

1. Si il existe une bijection entre X et Y , $|X| = |Y|$.

Si X est l'ensemble des résultats possibles pour une expérience, et Y est l'ensemble des résultats possibles pour une seconde expérience, mais qu'il existe un façon d'associer les éléments de X et Y entre eux deux-à-deux, alors les deux expériencesont le même nombre de résultats possibles.

2. $|X \sqcup Y| = |X| + |Y|$.

Si pour une expérience les résultats possibles se divisent en deux catégories distinctes X et Y (disjoints), alors le nombre total de résultats possibles est simplement la somme du nombre de résultats possibles dans chaque cas.

(Ici, on utilise le symbole \sqcup pour symboliser le fait que l'on prend la réunion de X et Y , et que ces ensembles sont disjoints).

3. $|X \times Y| = |X| \cdot |Y|$.

Si X est l'ensemble des résultats possible pour une expérience, et Y est l'ensemble des résultats possibles pour une seconde expérience, alors le nombre de résultats possibles pour les deux expériences ensemble est simplement le produit du nombre de résultats possibles pour chacune des expériences.

EXEMPLE 1.2. Dans un tournoi par élimination avec 2^n participant.e.s, combien de parties sont jouées ?

SOLUTION. On sait qu'à chaque partie jouée, un.e participant.e sera éliminé.e. On sait que les participant.e.s ne seront pas éliminé.e.s deux fois. On sait aussi que tou.te.s les participant.e.s sauf un.e seront éliminé.e.s lors du tournoi. Puisqu'il y a $2^n - 1$ participant.e.s éliminé.e.s, il doit y avoir $2^n - 1$ parties jouées.

EXEMPLE 1.3. Montrer que nombre de façons de disposer de façon cohérente n parenthèses ouvrantes et n fermantes est égal au nombre de trajets constitués de $2n$ pas (vers l'avant ou l'arrière) qui ne passent jamais derrière leur point d'origine, et qui terminent à leur point d'origine.

SOLUTION. Tous les trajets de $2n$ pas ne passant jamais derrière leur point d'origine, et qui terminent à leur point d'origine, sont caractérisés par les faits suivant :

- On ne peut jamais faire plus de pas vers l'arrière qu'on n'a déjà fait de pas vers l'avant (sinon on reviendrait derrière notre point d'origine) ;
 - On doit avoir fait n pas vers l'arrière et n pas vers l'avant ;
- Dans le cas des parenthèses, pour les disposer de façon cohérente,
- On ne peut fermer plus de parenthèses qu'on en a déjà ouvertes ;
 - On doit avoir ouvert n parenthèses et fermé n parenthèses.

On remarque qu'il existe alors une bijection entre les trajets et les arrangements de parenthèses. Pour un arrangement de parenthèses donné, on trouve le trajet où

- Chaque parenthèse ouvrante est un pas vers l'avant.
- Chaque parenthèse fermante est un pas vers l'arrière.

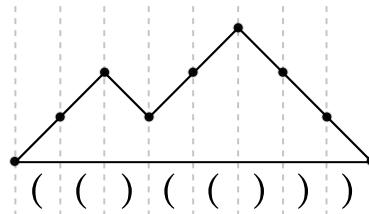


FIGURE 1.1. Correspondance entre une séquence d'ouvertures/fermetures de 8 parenthèses et un trajet de 8 pas. Chaque parenthèse ouvrante est un pas vers l'avant, chaque parenthèse fermante est un pas vers l'arrière.

C'est une bijection : tous les arrangements de parenthèses résultent en des trajets différents, et, bien sûr, si un trajet existe, on peut trouver l'arrangement de parenthèse qui lui est associé – on ouvre une parenthèse à chaque pas en avant, et on ferme une parenthèse à chaque pas en arrière.

Donc, il y a autant de trajets de $2n$ pas qui ne passent jamais derrière leur point d'origine et qui y finissent, qu'il y a d'arrangements cohérents de n parenthèses ouvrantes et n fermantes.

Pour les curieux/ses, le nombre C_n de tels arrangement est le *nième* nombre de Catalan.

EXEMPLE 1.4. Au restaurant, j'ai le choix entre 5 plats de pâtes, 10 pizzas différentes, et 3 salades. Combien de choix de repas ai-je ?

SOLUTION. J'ai $5 + 10 + 3 = 18$ choix de repas.

EXEMPLE 1.5. Au moment d'embarquer les animaux sur son arche, Noé compte 13 éléphants mâles et 7 femelles. Combien de couples d'éléphants Noé peut-il former ?

SOLUTION. Noé a $13 \times 7 = 91$ choix de couples d'éléphants possibles.

EXEMPLE 1.6. Noé compte 17 giraffes, dont 8 mâles. Mais parmi les 8 mâles, il y en a 1 qui est particulièrement capricieux, et qui refuse toutes les femelles sauf une.

Combien de couples de giraffes Noé peut-il former ?

SOLUTION. Soit Noé choisit le mâle capricieux. Dans ce cas, il ne peut former qu'un couple.

Ou alors Noé choisit l'un des 7 autres mâles, et il peut alors choisir l'une des $17 - 8 = 9$ femelles pour former un couple. En tout, il y a ainsi $7 \times 9 = 63$ couples possibles.

Donc, Noé peut former $63 + 1 = 64$ couples de giraffes.

EXEMPLE 1.7. Le Grand Prix d'Albert Park, en Australie, sera couru par 20 pilotes. Combien y a-t-il de podiums possibles ?

SOLUTION. Il y a 20 possibilités pour la première marche du podium, puis 19 possibilités pour la seconde marche, et 18 pour la troisième. Il y aura donc en tout $20 \times 19 \times 18 = 6840$ podiums possibles¹.

On compte aussi deux autres principes élémentaires de dénombrement, qui sont en fait des reformulations des principes énoncés plus haut, et qui sont aussi utiles fréquemment.

4. Si $Y \subseteq X$, $|X \setminus Y| = |X| - |Y|$.

Si les résultats possibles de notre expérience sont tous les résultats possibles d'une autre expérience, sauf pour un certain nombre, d'un cas précis, alors on peut compter le nombre de résultats possibles pour notre expérience en soustrayant le nombre de résultats exclus du nombre de résultats total.

1. Sauf que en fait Hamilton gagne toujours, et il y a seulement 5 autres pilotes qui ont une voiture assez rapide pour avoir une chance de se ramasser sur le podium. Donc c'est plus genre $1 \times 5 \times 4 = 20$ podiums vraisemblables...

5. $|Y| = |X \times Y| / |X|$.

Si on sait que notre expérience est l'une de deux expériences, et qu'on connaît le nombre total de résultats possibles pour les deux expériences ensemble, on peut obtenir le nombre de résultats possibles pour notre expérience en divisant le nombre total par le nombre de résultats possibles pour l'autre expérience.

Tout ça c'est formulé de façon très... abstraite. Alors voici des exemples, pour mieux comprendre :

EXEMPLE 1.8. Noé a compté 59 alligators, dont 31 mâles. Parmi les 31 mâles, il y en a 2 qui ont une rancune envers leurs exes respectives. Combien de couples d'alligators Noé peut-il former ?

SOLUTION. Sans contraintes, Noé pourrait former $31 \times (59 - 31) = 868$ couples d'alligators. Mais il y a 2 alligators qui ne veulent pas être avec leur ex respectif. Donc ça fait deux couples qu'il faut exclure.

Donc, en tout, Noé peut former $868 - 2 = 866$ couples d'alligators.

EXEMPLE 1.9. Michael a instauré une nouvelle politique chez Dunder-Mifflin : chaque matin, chacun.e des 13 membres du personnel de bureau doit donner la main à tous les autres. Mais Dwight refuse de donner la main à Jim et à Pam, et Michael refuse de donner la main à Toby. Combien de poignées de main sont échangées chaque matin ?

SOLUTION. Il y aurait $13 \times 12 = 156$ poignées de main possible en tout – sauf qu'on a compté deux fois chaque poignée de main, comme ça (celle de Jim à Pam et celle de Pam à Jim, qui sont en fait la même poignée de main). On doit donc diviser par 2. Il y a donc $\frac{13 \times 12}{2} = 78$ poignées de main possibles en tout.

Mais il y a trois de ces poignées de main qui n'auront pas lieu : celle entre Michael et Toby, celle entre Dwight et Jim, et celle entre Dwight et Pam. Donc, en tout, il faut exclure 3 poignées de main, et on se retrouve avec un total de $78 - 3 = 75$ poignées de mains échangée chaque matin.

1.2. Problèmes courants

Tous – je répète, **tous** – les problèmes de combinatoire auxquels nous ferons face peuvent être résolus en utilisant les principes que nous venons d'énoncer. Toutefois, certains problèmes reviennent très souvent, et il est utile de les connaître pour éviter de réinventer la roue chaque fois.

1.2.1. Les permutations. On considère le nombre de façons d'ordonner n objets distinguables. Puisqu'on a n choix pour le premier, puis $n - 1$ pour le second, puis $n - 2$ pour le troisième, et ainsi de suite, on a au final

$$n \cdot (n - 1) \cdot (n - 2) \cdots 3 \cdot 2 \cdot 1 = n!$$

permutations possibles pour n objets distinguables.

EXEMPLE 1.10. Huit coureurs prennent part à la finale du 100 m. Combien de classements sont possibles ?

SOLUTION. Il y a $8! = 40320$ classements possibles.

EXEMPLE 1.11. Noé doit faire monter les animaux sur son arche un par un. Si il y a 10 espèces d'animaux (à cette époque-là ils en connaissaient pas beaucoup), combien d'ordres possibles y a-t-il si

- (a) tout le monde monte pêle-mêle ?
- (b) on embarque les femelles et les mâles séparément ?
- (c) on embarque les spécimens de même espèce l'un.e après l'autre ?

SOLUTION. (a) Si tout le monde embarque pêle-mêle, il y a 20 animaux, et donc $20! = 2432902008176640000$ ordres possibles. C'est de l'ordre de 2×10^{18} , soit 2 trillions.

- (b) Si on embarque les femelles et les mâles séparément, il y a 10 spécimen dans chaque catégorie, donc $10!$ possibilités pour les mâles, $10!$ possibilités pour les femelles, et 2 possibilités d'ordre : soit les mâles ou les femelles d'abord. En tout, ça fait donc $2 \times (10!)^2 = 26336378880000$. C'est de l'ordre de 10^{13} , soit environ 2 dizaines de billions.
- (c) Si on embarque les espèces l'une après l'autre, il y a 2 possibilités d'ordre pour chaque espèce, et $10!$ possibilités d'ordres entre les espèces. Donc en tout, ça fait $10! \times 2^{10} = 3715891200$. C'est de l'ordre de 10^9 , soit environ 3 milliards de possibilités ; encore plusieurs ordres de grandeur plus petit !

EXEMPLE 1.12. On considère un tirage sans remise de k boules dans une urne qui en contient n , numérotées, et on tient compte de l'ordre. On veut savoir combien il y a de tirages possibles.

SOLUTION. On peut voir la mise en ordre des n boules comme deux expériences simultanées : le tirage des k premières boules, dans l'ordre, puis la mise en ordre des $n - k$ boules restantes.

On sait qu'il y a $n!$ ordres possibles pour les n boules, au total. On sait qu'il y a $(n - k)!$ ordres possibles pour les $n - k$ boules restantes. Donc, le nombre de tirages ordonnés de k boules possibles est $n!/(n - k)!$.

1.2.2. Les combinaisons.

L'exemple typique d'un problème ici est le suivant :

EXEMPLE 1.13. Combien de façon y a-t-il d'ordonner m boules blanches et n boules noires ?

Ce qui rend ce problème plus particulièrement difficile, c'est que les boules blanches et les boules noires sont indistinguables respectivement entre elles.

Si on s'imagine que les boules, en plus d'être colorées, sont numérotées identiquement chacunes, alors, évidemment, en tout il y aurait $(m + n)!$ façons de les ordonner.

Mais on peut voir cela comme trois expériences : (a) l'ordre des couleurs, (b) l'ordre des boules blanches entre elles et (c) l'ordre des boules noires entre elles. On sait qu'il y a $m!$ possibilités pour l'ordre des boules blanches entre elles, et $n!$ possibilités pour l'ordre des boules noires entre elles.

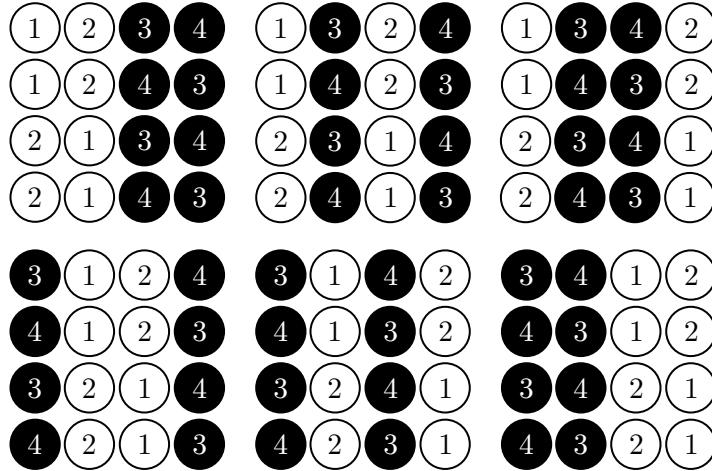


FIGURE 1.2. Les $4! = 24$ permutations de 4 boules numérotées, dont deux sont noires et deux sont blanches. On voit clairement qu'il doit y avoir $\binom{4}{2} = \frac{4!}{2! \cdot 2!} = 6$ façons de disposer 4 boules dont deux blanches et deux noires si on ignore les numéros.

Comme il y a $(m+n)!$ possibilités au total pour l'ordre des boules colorées et numérotées, alors il doit y avoir $(m+n)!/(m!n!)$ possibilités pour l'ordre des couleurs. La figure 1.2 illustre cette procédure pour 4 boules numérotées, dont deux sont noires et deux sont blanches.

De façon générale, le nombre de façons de tirer k objets parmi n , sans remises et sans se soucier de l'ordre, est donné par l'expression

$$\binom{n}{k} := \frac{n!}{k!(n-k)!},$$

qu'on lit simplement « k parmi n ». Ces coefficients sont appelés **coefficients binomiaux**.

EXEMPLE 1.14. Combien de chaînes peut on former avec m '0' et n '1'?

SOLUTION. On a $m+n$ caractères, et on doit en choisir n qui seront des '1'; le reste seront des '0'. On a donc $\binom{m+n}{m} = (m+n)!/(m!n!)$ possibilités.

EXEMPLE 1.15. Dans une ville découpée en quadrilatères, avec des avenues (nord-sud) et des rues (est-ouest) on doit se rendre de l'angle de la 3e avenue et de la 40e rue, à l'angle des 6e avenue et 47e rue. Combien de trajets sont possibles sans faire de détours ?

SOLUTION. On devra se déplacer de trois blocs vers l'est et de 7 blocs vers le nord. À chacune des 10 intersections que nous croiserons, il faudra choisir d'aller vers l'est ou vers le nord. Pour nous rendre à notre destination, il faudra avoir choisi exactement trois fois l'est.

On peut représenter ça par une chaîne de 10 caractères, dont 3 doivent être des '1', et le reste des '0'. Donc, il y a $\binom{10}{3} = 120$ trajets possibles qui nous mènent sans détour à notre destination.

EXEMPLE 1.16. Montrer le théorème binomial :

THÉORÈME 1.1 (Théorème binomial). *Soient $x, y \in \mathbb{R}$ (ou n'importe quelle algèbre commutative), et soit $n \in \mathbb{N}$.*

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

SOLUTION. On peut prouver ce résultat par induction (exercices), mais nous allons ici raisonner de façon combinatoire.

On commence par étendre le produit :

$$(x + y)^n = \underbrace{(x + y) \cdot (x + y) \cdots (x + y)}_{n \text{ fois}}.$$

En développant ce produit en à l'aide de la propriété de la distributivité de la multiplication, on obtiendra 2^n termes – un pour chaque choix de x ou de y dans chaque parenthèse.

Mais de ces 2^n termes, plusieurs seront identiques – en effet, puisque le produit est commutatif, tous les termes seront de degré n , et on pourra regrouper nos termes selon la puissance de x et la puissance de y . Et bien sûr, on aura autant de termes en $x^k y^{n-k}$ qu'il y avait de façons de choisir k fois un x et $n - k$ fois un y dans les parenthèses – c'est à dire $\binom{n}{k}$.

Au final, on se retrouve bien avec

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

□

1.2.3. Les coefficients multinomiaux. Disons qu'on reprenne notre problème de boules blanches et de boules noires, et qu'on y ajoute des boules rouges. Combien y a-t-il de façons d'agencer m boules blanches, n boules noires, et p boules rouges ?

La solution ? Procéder par étapes. Il y a $\binom{m+n+p}{n}$ façons de choisir quelles boules seront noires. Puis, il reste $m+p$ boules, d'entre lesquelles p doivent être rouges. Il y a $\binom{m+p}{p}$ façons de réaliser cela.

En tout, on a donc

$$\binom{m+n+p}{n} \binom{m+p}{p} = \frac{(m+n+p)!}{m!n!p!}$$

façons d'agencer m boules blanches, n boules noires et p boules rouges. Et si on rajoute une nouvelle couleur ? On refait le même procédé.

De façon générale, si on a n éléments de K catégories différentes, chacune de $n_i, i \leq K$ éléments, à agencer entre eux, il y aura

$$\binom{n}{n_1, n_2, n_3, \dots, n_K} := \frac{n!}{n_1!n_2!n_3!\cdots n_K!}.$$

Ces coefficients sont appelés **coefficients multinomiaux**, par analogie aux coefficients binomiaux.

On peut prouver le théorème suivant de façon analogue au théorème binomial :

THÉORÈME 1.2. *Si $x_1, x_2, \dots, x_K \in \mathbb{R}$ (ou n'importe quelle algèbre commutative), et que $n \in \mathbb{N}$, alors*

$$(x_1 + x_2 + x_3 + \cdots + x_K)^n = \sum_{n_1+\cdots+n_K=n} \binom{n}{n_1, \dots, n_K} x_1^{n_1} x_2^{n_2} \cdots x_K^{n_K}.$$

EXEMPLE 1.17. Au département il y a 60 étudiant.e.s gradué.e.s. Les cours d'algèbre linéaire et de calcul ont besoin de 4 auxiliaires chacun, le cours de probabilités, de maths discrètes, et d'analys ont besoin de 2 auxiliaires chacun, et le cours de mathématiques assistées par ordinateur a besoin d'un.e seul.e auxiliaire. Aucun autre cours n'a besoin d'auxiliaires.

Combien de répartition des tâches sont possibles ?

SOLUTION. D'abord on devra décider qui aura un emploi ! Il faut choisir 15 étudiant.e.s gradué.e.s parmi 60 – donc $\binom{60}{15}$ possibilités, soit 53194089192720 (de l'ordre de la dizaine de billions).

Puis, parmi les étudiant.e.s qui ont eu un emploi, on doit attribuer les tâches. Il faut choisir 4 démonstrateurs pour algèbre linéaire, 4 pour calcul, 2 pour proba, 2 pour maths discrètes, 2 pour analyse et 1 pour maths assistées par ordinateur. Il y a donc

$$\binom{15}{4, 4, 2, 2, 2, 1} = \frac{15!}{(4!)^2(2!)^31!} = 283783500$$

possibilités d'attribuer les tâches parmi celles et ceux qui ont eu un emploi (ordre de la centaine de millions).

Au total, il y a donc $\binom{60}{15} \binom{15}{4, 4, 2, 2, 2, 1} = 15095604810422256120000$ possibles attributions de tâches (de l'ordre de la dizaine de trilliards)

EXEMPLE 1.18. Douze enfants doivent être répartis en trois équipes de 4 pour une partie enlevante de kinball au service de garde. Combien de répartitions possibles y a-t-il ?

SOLUTION. On doit créer une équipe avec des dossards noirs, une avec des dossards gris, et une avec des dossards roses (qui sont, comme chacun.e sait, les trois couleurs du Homnikin).

Évidemment, le nombre de telles répartitions est simplement $\binom{12}{4, 4, 4} = 34650$ – mais attention : ce n'est pas notre réponse !

En effet, ce calcul compte non seulement les répartitions des équipes possibles, mais aussi les différents choix de couleurs. Mais nous, on s'en fout, des couleurs. Tout ce qu'on veut, c'est diviser le groupe en trois. On va donc diviser par $3! = 6$, soit le nombre de permutations de couleurs.

En bout de compte, il y a donc

$$\frac{1}{3!} \binom{12}{4, 4, 4} = 5775$$

possibilités de répartitions différentes.

1.2.4. Le nombre de solutions naturelles à une équation simple. Ici, la question-type, c'est :

EXEMPLE 1.19. Combien l'équation

$$x_1 + x_2 + x_3 + \cdots + x_k = n$$

a-t-elle de solutions entières non-négatives ?

SOLUTION. Pour répondre à cette question, on emploie l'astuce suivante : on s'imagine qu'on a k « boîtes » – nos variables – et qu'on veut mettre n boules indistinguables dedans. Pour ce faire, on a une machine, qui commence au dessus de la première boîte, et on lui envoie une séquence de ‘0’ et de ‘1’. Chaque ‘0’ signifie « mets une boule dans la boîte », et chaque ‘1’ signifie « passe à la prochaine boîte ».

Puisqu'il y a k boîtes, il faudra changer de bête $k - 1$ fois – donc il faut envoyer un ‘1’ $k - 1$ fois. Mais puisqu'il y a n boules à mettre, il faut envoyer un ‘0’ n fois.

Au final, on voit qu'il y a une correspondance entre les chaînes binaires (de ‘0’ et de ‘1’) de longueur $n + k - 1$ qui contiennent n ‘0’ et $k - 1$ ‘1’, et les répartitions de boules dans des boîtes. Donc, il y a $\binom{n+k-1}{n}$ façons de répartir n boules dans nos k boîtes, ou $\binom{n+k-1}{n}$ solutions entières positives à notre équation.

EXEMPLE 1.20. Combien cette équation a-t-elle de solutions entières strictement positives ?

SOLUTION. Dans l'exemple précédent, on comptait le nombre de solutions entières non-négatives. En effet, remarquons que, en comptant toutes les chaînes binaires qui comptent $k - 1$ ‘1’ et n ‘0’, on compte des chaînes qui comportent, quelque part, deux ‘1’ collés :

$$\dots 01100\dots,$$

ce qui, pour nous, signifierait que notre machine « sauterait par-dessus une boîte sans mettre de boules dedans » – donc on pourrait avoir des solutions où certaines des variables sont nulles.

L'astuce, ici, c'est de réaliser que pour que toutes les solutions soient non-nulles, il faut que chaque boîte reçoive au moins une boule. Donc, on va seulement déposer une boule dans chaque boîte avant de partir notre machine ! Une fois que c'est fait, il nous reste $n - k$ boules à placer – et là, on peut faire ce qu'on veut.

Autrement dit, le nombre de solutions strictement positive à l'équation

$$x_1 + \cdots + x_k = n$$

est exactement le même que le nombre de solutions non-négatives à l'équation

$$x_1 + \cdots + x_k = n - k.$$

Donc, en raisonnant comme dans l'exemple précédent, ce nombre sera donné par $\binom{n-1}{n-k}$.

1.2.5. La formule du crible, ou le principe d'inclusion-exclusion. On a déjà vu que lorsque les résultats possibles d'une expérience se divisent en sous-ensembles disjoints, il suffit d'additionner le nombre de résultats dans chaque sous-ensemble pour obtenir le nombre de résultats possibles total.

Mais si les sous-ensembles ne sont pas disjoints ? On fait alors appel à la formule du crible :

THÉORÈME 1.3. *Soient $X_1, X_2, X_3, \dots, X_N$ des ensembles finis. Alors,*

$$\left| \bigcup_{k=1}^N X_k \right| = \sum_{n=1}^N (-1)^{n-1} \sum_{i_1 < \dots < i_n} \left| \bigcap_{j=1}^n X_{i_j} \right|.$$

Par exemple, pour le cas $N = 2$, on a

$$|X \cup Y| = |X| + |Y| - |X \cap Y|.$$

Pour le cas $N = 3$, on a

$$\begin{aligned} |X \cup Y \cup Z| &= |X| + |Y| + |Z| \\ &\quad - |X \cap Y| - |X \cap Z| - |Y \cap Z| \\ &\quad + |X \cap Y \cap Z|. \end{aligned}$$

PREUVE DE LA FORMULE DU CRIBLE. On fait la preuve par induction. On vérifie d'abord que c'est vrai pour $N = 2$. Dans ce cas, il suffit de constater que $X \cup Y = X \sqcup (Y \setminus X)$, et d'utiliser les principes qu'on a vus auparavant.

Supposons que ce soit vrai pour tout $N' \leq N$. Alors,

$$\begin{aligned} \left| \bigcup_{k=1}^{N+1} X_k \right| &= \left| \bigcup_{k=1}^N X_k \cup X_{N+1} \right| \\ &= \left| \bigcup_{k=1}^N X_k \right| + |X_{N+1}| - \left| X_{N+1} \cap \bigcup_{k=1}^N X_k \right| \end{aligned}$$

Et $X_{N+1} \cap \bigcup_{k=1}^N X_k = \bigcup_{k=1}^N (X_k \cap X_{N+1})$. Donc,

$$\begin{aligned} \left| \bigcup_{k=1}^{N+1} X_k \right| &= \left| \bigcup_{k=1}^N X_k \right| + |X_{N+1}| - \left| X_{N+1} \cap \bigcup_{k=1}^N X_k \right| \\ &= \sum_{n=1}^N (-1)^{n-1} \sum_{i_1 < \dots < i_n} \left| \bigcap_{j=1}^n X_{i_j} \right| + |X_{N+1}| - \sum_{n=1}^N (-1)^{n-1} \sum_{i_1 < \dots < i_n} \left| \bigcap_{j=1}^n (X_j \cap X_{N+1}) \right| \\ &= \sum_{n=1}^N (-1)^{n-1} \sum_{i_1 < \dots < i_n} \left| \bigcap_{j=1}^n X_j \right| + |X_{N+1}| + \sum_{n=2}^{N+1} (-1)^{n-1} \sum_{\substack{i_1 < \dots < i_n \\ i_n = N+1}} \left| \bigcap_{j=1}^n X_{i_j} \right| \\ &= \sum_{n=1}^{N+1} (-1)^{n-1} \sum_{i_1 < \dots < i_n} \left| \bigcap_{j=1}^n X_{i_j} \right| \end{aligned}$$

Le passage de la première à la seconde ligne est par hypothèse d'induction. Le passage de la seconde à la troisième est un simple changement d'indices, et le passage à la dernière se

fait en remarquant qu'on ajoute précisément les termes qui contiennent l'intersection avec X_{N+1} . \square

La notation dans cette preuve est très lourde, mais l'important, c'est surtout de comprendre comment appliquer la formule – revoyez les formules explicites pour 2 ou 3 ensembles.

EXEMPLE 1.21. Si 4 232 262 québécois.e.s ont voté à l'élection générale de 2014, que 4 099 623 québécois.e.s ont voté à l'élection générale de 2018, et qu'exactement 5 millions de québécois.e.s disent avoir voté à au moins une de ces deux élections, combien ont voté aux deux ?

SOLUTION. Si N est le nombre qu'on cherche, on a

$$5000000 = 4232262 + 4099623 - N,$$

d'où l'on déduit que

$$N = 4232262 + 4099623 - 5000000 = 3331885.$$

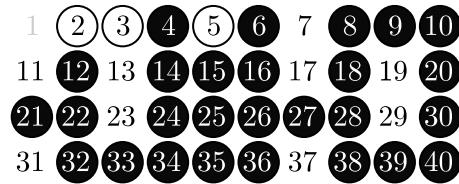


FIGURE 1.3. Le crible d'Ératosthène. Sachant que les nombres 2, 3, 5 et 7 sont premiers, on peut retirer les multiples de 2, 3 ou 5 pour tous les nombres plus petits ou égal à 40.

EXEMPLE 1.22. La figure 1.3 illustre la célèbre méthode du « crible d'Ératosthène » permettant de trouver de nouveaux nombres premiers. Il s'agit d'une méthode itérative : si on connaît les $n + 1$ premiers nombres premiers, p_1, \dots, p_{n+1} , alors on sait que tous les nombres entiers qui ne sont pas multiples des n premiers nombres premiers et qui sont inférieurs à p_{n+1}^2 sont aussi des nombres premiers.

Sans les énumérer, utiliser la formule du crible pour déterminer combien il y a de nombres premiers plus petits que 40.

SOLUTION. Tous les nombres composites (non-premiers) inférieurs à 40 sont des multiples de 2, 3 ou 5. Autrement, puisque ce sont des nombres composites, ils doivent être des multiples de deux nombres premiers supérieurs ou égaux à 7, mais $7^2 = 49 > 40$.

Donc, on peut compter le nombre de composites inférieurs ou égaux à 40 – il suffit de compter combien il y a de multiples de 2 ou 3 ou 5.

Il y a 20 multiples de 2, 13 multiples de 3 et 8 multiples de 5. Il y a 6 multiples de $6 = 2 \times 3$, 4 multiples de $10 = 2 \times 5$ et 2 multiples de $15 = 3 \times 5$. Il y a 1 multiple de $30 = 2 \times 3 \times 5$.

Donc, il y a $20 + 13 + 8 - 6 - 4 - 2 + 1 = 30$ nombres inférieurs ou égaux à 40 qui sont multiples de 2, 3 ou 5. Donc ça fait 10 nombres qui ne sont pas multiples, ni de 2, ni de 3, ni de 5. De ces 10, il y a évidemment 1, qui n'est pas premier. Donc on a 9 nombres premiers, plus 3, car 2, 3 et 5 sont aussi premiers. Ça donne donc 12 nombres qui sont premiers et inférieurs ou égaux à 40.

1.3. Exercices

EXERCICE 1.1. Axel, Bénédicte, Claude et Dominique veulent créer un groupe de musique ; chacun.e choisira un instrument parmi les trois suivants :

- la batterie ;
- la guitare ;
- la flûte à coulisse.

- (a) Combien y a-t-il de groupes possibles si chacun.e est libre de choisir n'importe lequel des trois instruments ?
- (b) Combien y a-t-il de groupes possibles si chaque instrument doit être joué par au moins un.e membre du groupe ?
- (c) Si Axel et Bénédicte savent seulement jouer de la guitare et de la batterie, combien peut-on former de groupes où tous les instruments sont joués par au moins un.e membre du groupe ?

EXERCICE 1.2. L'algèbre des octonions \mathbb{O} est non-associative – c'est à dire que si on a $x, y, z \in \mathbb{O}$, alors en général, on ne peut pas assumer que $(xy)z = x(yz)$.

Supposons qu'on a les facteurs $x_1, x_2, \dots, x_{n+1} \in \mathbb{O}$. On va noter C_n le nombre de façons possibles de placer des parenthèses pour prendre le produit de ces facteurs dans l'ordre. Par exemple, pour $n = 0$, on a $C_0 = 1$. Pour $n = 1$, on a aussi $C_1 = 1$ – puisque le seul produit possible est x_1x_2 . Pour $n = 2$, on a $C_2 = 2$ – en effet, les produits possibles sont $x_1(x_2x_3)$ et $(x_1x_2)x_3$.

- (a) Énumérer les possibilités pour $n = 3$, et donner la valeur pour C_3 .
- (b) Montrer que les C_n satisfont la récurrence suivante :

$$C_{n+1} = \sum_{k=0}^n C_k C_{n-k}.$$

- (c) Calculer C_4 et C_5 .

REMARQUE. Les C_n sont appelés **nombres de Catalan**, et sont une suite de nombres entiers qui apparaît dans plusieurs problèmes fameux en combinatoire. Ils ont une importance très particulière dans la théorie des matrices aléatoires, par exemple.

- EXERCICE 1.3.** (a) Combien de manières y-a-t-il d'aligner 3 hommes et 3 femmes ?
- (b) Combien de manières y-a-t-il d'aligner 3 hommes et 3 femmes si les hommes restent ensemble et les femmes ensemble ?
 - (c) Combien de manières y-a-t-il d'aligner 3 hommes et 3 femmes si les hommes restent ensemble ?
 - (d) Combien de manières y-a-t-il d'aligner 3 hommes et 3 femmes si deux personnes de même genre ne peuvent pas être à côté ?

EXERCICE 1.4. Un enfant a 10 cubes, parmi eux 4 sont blancs, 3 sont noirs, 2 sont bleus et 1 bloc est vert. Combien y-a-t-il de manière d'aligner les cubes ?

EXERCICE 1.5. Combien y-a-t-il de manières d'asseoir 6 personnes en rangée si Alex et Bernard ne veulent pas être assis ensemble ?

EXERCICE 1.6. Combien de façons y a-t-il d'asseoir 6 personnes à une table ronde si l'orientation de la table n'a pas d'importance ?

EXERCICE 1.7. Supposons que 10 personnes, provenant de 5 couples, sont assises à une table (l'orientation est importante). Combien y-a-t-il de manières possibles de s'asseoir sachant que les couples veulent rester ensemble.

EXERCICE 1.8. Dans Manhattan, on chercher à partir du Starbucks au coin de la 5 ème avenue et de la 25 ème rue, jusqu'à notre Hôtel au coin de la 9 ème avenue et de la 32 ème rue.

- (a) Combien y-t-a-il de manières de faire cela sans faire de détour ? (en augmenter toujours d'une rue ou d'une avenue)
- (b) Combien y-t-a-il de manières de faire cela sans faire de détour en passant par la pharmacie au coin de la 7ème avenue et la 28ème rue ?

EXERCICE 1.9. Un ascenseur part du rez de chaussée (étage 0) avec 8 personnes (5 hommes et 3 femmes) à son bord, puis se vide étage par étage, jusqu'au 6e étage. De combien de façons cela peut-il se faire si :

- (a) on distingue les personnes ?
- (b) on ne distingue pas les personnes ?
- (c) on distingue entre les hommes et les femmes ?

EXERCICE 1.10. (a) Combien y a-t-il de mains possibles de 5 cartes au Poker (l'ordre des cartes n'est pas important ?

- (b) Combien y a-t-il de mains de 5 cartes qui comportent
 - i. une paire ?
 - ii. deux paires ?
 - iii. un brelan (trois pareilles) ?
 - iv. une quinte (une séquence de valeurs) ?
 - v. une flush (toutes les cartes sont de la même suite – par exemple, toutes les cartes sont des coeurs)
 - vi. une main pleine (une paire et trois pareilles) ?
 - vii. un carré (quatre pareilles) ?
 - viii. une quinte flush (une séquence de valeurs, toutes de la même suite), mais qui ne commence pas par un 10 ?
 - ix. une quinte flush Royale (la quinte flush qui commence par un 10).
- (c) Combien y a-t-il de mains de 5 cartes qui n'ont rien de spécial ?

EXERCICE 1.11. Combien y-a-t-il de k -uplets d'entiers (x_1, \dots, x_k) tels que $1 \leq x_i \leq n$ avec $x_1 < x_2 < \dots < x_k$?

EXERCICE 1.12. Montrer les propriétés suivantes pour les coefficients binomiaux :

- (a) La symétrie :

$$\binom{n}{k} = \binom{n}{n-k}.$$

- (b) La propriété du triangle de Pascal :

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

EXERCICE 1.13. Montrer le théorème binomial (Théorème 1.1) par induction.

EXERCICE 1.14. Dans un groupe de n personnes, supposons que nous voulons former un comité de k personnes ($k \leq n$) avec un président de comité.

- (a) En comptant d'abord le choix de comités puis le choix de président montrer que le nombre de choix est $k \binom{n}{k}$.
- (b) En comptant d'abord le choix du comité sans président puis le choix du président montrer que il y a $\binom{n}{k-1}(n-k+1)$ choix.
- (c) En comptant d'abord le choix du président puis le choix des autres membres du comité montrer qu'il y a $n \binom{n-1}{k-1}$ choix.
- (d) En déduire que $k \binom{n}{k} = \binom{n}{k-1}(n-k+1) = n \binom{n-1}{k-1}$.

EXERCICE 1.15. Combien y a-t-il de termes dans l'expansion multinomiale de

$$(x+y+z+w)^3?$$

Et dans l'expansion de

$$(x_1 + \dots + x_k)^n?$$

EXERCICE 1.16. Considérez l'équation

$$x_1 + \dots + x_k = n.$$

Montrez que le nombre de solutions entières non-négatives avec exactement r variables nulles est donné par

$$\binom{k}{r} \binom{n-1}{n-k+r}.$$

EXERCICE 1.17. (a) Dans un groupe de $m+n$ personnes, m portent des lunettes et n n'en portent pas. Montrer que le nombre de façons de former un comité de k personnes dont i ne portent pas de lunettes est

$$\binom{n}{i} \binom{m}{k-i}.$$

(b) Déduire l'identité de Vandermonde :

$$(1.3.1) \quad \binom{m+n}{k} = \sum_{i=0}^k \binom{n}{i} \binom{m}{k-i}.$$

EXERCICE 1.18. Déterminez (sans les énumérer) combien il y a de nombres premiers inférieurs ou égaux à 120.

Première partie

Les espaces de probabilités.

Chapitre 2

Les axiomes des probabilités

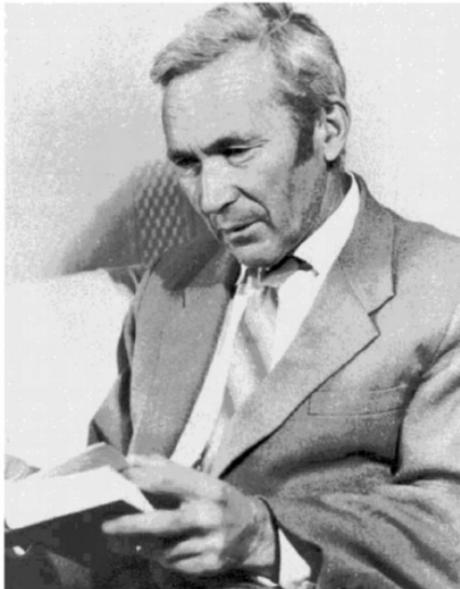


Photo : domaine public.

Andreï Nikolaevich Kolmogorov (1903 – 1987) est aujourd’hui connu comme l’un·e des mathématicien·ne·s à l’origine de la théorie moderne des probabilités. Notamment, on appelle souvent « axiomes de Kolmogorov » les axiomes définissant les mesures de probabilités. Lui sont également attribuées nombre de contributions aux domaines de la théorie des processus stochastiques, de la topologie, de l’étude des turbulences, de la logique mathématique et des théories des algorithmes et de l’information.

Parmi de nombreuses distinctions, notons en particulier le titre de *Héros du Travail Socialiste* qui lui fut décerné en 1963. Un sacré camarade.

Dans ce chapitre, nous verrons les définitions et les concepts qui sont la base de la théorie des probabilités. Nous abordons les notions d’ensemble fondamental, de tribu des événements, de mesure de probabilité, et d’espace de probabilités. Nous voyons également certaines propriétés importantes des probabilités.

Avant de débuter le chapitre, il est recommandé de se rafraîchir la mémoire au sujet de la théorie des ensembles. Consulter au moins les sections 1 à 4 de l’annexe A.

2.1. L’ensemble fondamental et les événements

Nous entrons maintenant dans le vif du sujet : les probabilités. Mais qu’est-ce au juste ?

La théorie des probabilités propose un modèle mathématique pour analyser le comportement de systèmes stochastiques, où les résultats sont (à toute fins pratiques) aléatoires.

2.1.1. L’ensemble fondamental. La base de notre modèle, ce sera une «expérience aléatoire», pouvant produire un ensemble de résultats, fixe, connu d’avance.

EXEMPLE 2.1. Le jeu de pile ou face peut être modélisé par une expérience aléatoire qui peut produire deux résultats : soit pile, soit face.

EXEMPLE 2.2. Un lancer de deux dés ordinaires à six faces peut être modélisé par une expérience aléatoire qui peut produire 36 résultats différents, correspondant à toutes les paires ordonnées d’entiers de 1 à 6.

EXEMPLE 2.3. Le nombre de particules radioactives détectées par un compteur Geiger en une minute peut être modélisé comme une expérience aléatoire dont le résultat serait un entier non-négatif.

EXEMPLE 2.4. La température qu'il fera demain peut-être modélisée par une expérience aléatoire dont le résultat sera un nombre réel.

L'ensemble fondamental peut être fini, infini dénombrable ou même indénombrable. On fait la définition suivante :

DÉFINITION 2.1 (Ensemble fondamental). L'ensemble des résultats possibles d'une expérience aléatoire est appelé **l'ensemble fondamental**, et on le note, plus souvent qu'autrement, Ω , mais parfois aussi S .¹ Il est toujours non-vide.

2.1.2. Les événements. Ce qui nous intéressera tout particulièrement, c'est de traduire des questions «naturelles» en questions mathématiques. Par exemple, on peut se demander : « quelles sont les chances qu'un tel parti remporte les élections ?» ou « quelles sont les chances que j'obtienne 'pile' quatre fois de suite en lançant une pièce dix fois ? », ou encore « quelles sont les chances qu'il pleuve demain ? »

Dans chacune de ces questions, la quantité que l'on cherche n'est pas associée directement à un résultat spécifique de notre expérience aléatoire – c'est plutôt une quantité associée à un ensemble de résultats qui partagent une caractéristique d'intérêt : l'ensemble des résultats d'élections qui permettraient d'obtenir une majorité relative de sièges, ou l'ensemble des séquences de pile ou face dans lesquelles on retrouve une suite de quatre fois pile, ou l'ensemble des configurations météorologiques qui occasionnent de la pluie.

C'est pour cette raison qu'on introduit la notion **d'événement**.

DÉFINITION 2.2 (Événement). Un **événement** est un sous-ensemble de l'ensemble fondamental Ω .

L'ensemble des événements, que nous noterons \mathcal{E} , est une partie de $\mathcal{P}(\Omega)$ (l'ensemble des parties de l'ensemble fondamental). Cependant, il y a une condition importante à respecter : \mathcal{E} doit être une *tribu* .

DÉFINITION 2.3 (Tribu). Soit Ω un ensemble non-vide. $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ est une **tribu** si elle respecte les axiomes suivants :

- i. L'ensemble Ω est un événement : $\Omega \in \mathcal{E}$.
- ii. Si $E \in \mathcal{E}$ est un événement, E^c est un événement : $E^c \in \mathcal{E}$.
- iii. Si $\{E_i \in \mathcal{E}\}_{i \in \mathbb{N}}$ est une famille dénombrable d'événements (une suite), alors

$$\bigcup_{i \in \mathbb{N}} E_i, \quad \bigcap_{i \in \mathbb{N}} E_i$$

sont des événements (ils sont dans \mathcal{E}).

Puisque les événements forment une tribu, cela signifie qu'on peut toujours construire de nouveaux événements à partir de ceux qu'on connaît déjà en les combinant au moyen d'opérations ensemblistes – les axiomes définissant les tribus garantissent que les parties

1. Le choix du symbole est arbitraire ; Ω et S sont des conventions fréquemment utilisées. S fait référence à *sample space*, et est souvent employé en statistiques. S est le symbole préféré par Ross dans *Initiation aux probabilités* .

de Ω ainsi construites seront aussi des événements. Les interprétations de ces opérations ensemblistes sont assez intuitives.

Imaginons d'abord que $\omega \in \Omega$ soit le résultat de notre expérience aléatoire (celui qui s'est finalement produit).

- Si $E, F \in \mathcal{E}$ sont des événements,
 - On dit que E est réalisé si $\omega \in E$.
 - $E \cup F$ est l'événement « E ou F est réalisé » ;
 - $E \cap F$ est l'événement « E et F sont réalisés » ;
 - $E \setminus F$ est l'événement « E est réalisé, mais pas F » ;
 - E^c est l'événement « E n'est pas réalisé.»
 - Si $E \subseteq F$, alors la réalisation de l'événement E entraîne automatiquement la réalisation de l'événement F .

REMARQUE. Très souvent, notre ensemble des événements, \mathcal{E} , sera simplement l'ensemble des parties de Ω (noté $\mathcal{P}(\Omega)$). Toutefois, pour des raisons dont nous reparlerons plus loin, ça ne sera pas toujours le cas ; ces exceptions sont de nature techniques. Nous en verrons, mais nous ne nous attarderons pas sur les éventuels problèmes dans ce cours.

Pour nos fins, vous pourrez toujours vous imaginer que $\mathcal{E} = \mathcal{P}(\Omega)$.

Nous avons donc deux ingrédients importants de notre modèle d'une expérience aléatoire : un ensemble fondamental Ω de résultats possibles pour notre expérience, et un ensemble d'événements $\mathcal{E} \subseteq \mathcal{P}(\Omega)$, qui sont des parties d' Ω .

Pour répondre à nos questions, par contre, il manque un ingrédient crucial :

2.2. La mesure de probabilité

En mathématiques, une **mesure** est une fonction d'ensembles – c'est-à-dire une fonction dont l'argument est un ensemble – qui, en quelque sorte, « généralise la notion de cardinalité » de façon pratique pour les ensembles infinis.

L'idée, c'est qu'on voudrait que, dans un certain sens mathématique, l'intervalle $[0, 1]$ soit reconnu comme étant « plus petit » que l'intervalle $[0, 2]$. Mais au sens de la cardinalité, ça ne fonctionne pas : la bijection $x \mapsto 2x$ entre les deux intervalles signifie que $|[0, 1]| = |[0, 2]|$.

On se retourne donc vers une autre façon de « mesurer » des ensembles. Dans le cas des intervalles, on a la mesure de Lebesgue, qui correspond simplement à la « longueur » de l'intervalle (la différence entre sa borne supérieure et sa borne inférieure).

Vous connaissez déjà des mesures :

- la cardinalité est une mesure pour les ensembles finis ;
- la « longueur » de segments est une mesure pour les intervalles ;
- l'aire d'une région du plan est la mesure de Lebesgue de cet ensemble dans \mathbb{R}^2 ;
- le volume d'une région de l'espace est la mesure de Lebesgue de cet ensemble dans \mathbb{R}^3 ;
- la masse contenue dans une région de l'espace est aussi une mesure pour cette région de l'espace dans \mathbb{R}^3 .

Plus exotique que les autres, cette dernière fournit une analogie très pratique lorsqu'on parlera de mesures non uniformément réparties.

Malgré que ces notes ne sont pas un cours de théorie de la mesure, il est néanmoins pertinent d'introduire une définition formelle :

DÉFINITION 2.4 (Mesure positive). Soit Ω un ensemble non-vide et \mathcal{E} une tribu de parties de Ω . Une **mesure positive** $\mu : \mathcal{E} \rightarrow \mathbb{R}$ est une fonction d'ensembles qui respecte les axiomes suivants :

- i. Pour tout $E \in \mathcal{E}$, $\mu\{E\} \geq 0$.
- ii. Pour toute suite $(E_i \in \mathcal{E})_{i \in \mathbb{N}}$ de parties de Ω deux-à-deux disjointes (c'est-à-dire que pour tous $i \neq j$, on a que $E_i \cap E_j = \emptyset$), on doit avoir

$$\mu \left\{ \bigsqcup_{i=1}^{\infty} E_i \right\} = \sum_{i=1}^{\infty} \mu\{E_i\}.$$

Il est laissé à l'exercice 2.5 de vérifier que les exemples donnés plus haut vérifient cette définition.

En probabilités, nous nous servirons du même genre d'outils : nous allons définir une **mesure de probabilités**, qui servira non pas à déterminer «la taille» des ensembles, mais plutôt à déterminer «à quel point un événement est probable» .

DÉFINITION 2.5 (Mesure de probabilités). Soit Ω un ensemble fondamental pour une expérience aléatoire, et $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ l'ensemble des événements pour cette expérience aléatoire.

Alors, on définit la **mesure de probabilités** $\mathbb{P} : \mathcal{E} \rightarrow \mathbb{R}$, qui satisfait les axiomes suivants :²

- i. Pour tout événement $E \in \mathcal{E}$, on doit avoir $\mathbb{P}\{E\} \geq 0$.
- ii. Pour toute suite d'événements $(E_i \in \mathcal{E})_{i \in \mathbb{N}}$ deux-à-deux disjoints (c'est-à-dire que pour tous $i \neq j$, on a que $E_i \cap E_j = \emptyset$), on doit avoir

$$\mathbb{P} \left\{ \bigsqcup_{i=1}^{\infty} E_i \right\} = \sum_{i=1}^{\infty} \mathbb{P}\{E_i\}.$$

- iii. $\mathbb{P}\{\Omega\} = 1$.

REMARQUE. On remarque que la mesure de probabilité est simplement une mesure positive qui satisfait en plus la condition de normalisation $\mathbb{P}\{\Omega\} = 1$.

La mesure de probabilités est celle qui nous permet de répondre à nos questions. On peut l'interpréter de plusieurs façons

— la mesure de probabilité donne, dans la limite où l'on reproduirait la même expérience indépendamment un très grand nombre de fois, la proportion du nombre de fois où l'événement serait réalisé.

Nous allons voir que cette interprétation est cohérente avec la définition ; c'est un théorème qui se nomme «la loi des grands nombres.»

— La mesure de probabilités donne le «degré de vraisemblance» (c'est volontairement vague) qu'un événement se produise. Cette interprétation permet de donner un sens à la question « quelle est la probabilité que Donald Trump remporte l'élection présidentielle de 2020 ? » – une expérience qu'on ne peut pas répéter un grand nombre de fois (certain.e.s diraient : « ouf ! »).

Cette interprétation permet aussi de donner du sens à des notions plus subjectives. Par exemple, « après avoir regardé juste la première saison du téléroman *Rumeurs* , la probabilité qu'Esther finisse avec Benoît.»

Dans ce cas, la théorie fonctionne toujours, parce qu'en principe, c'est naturel d'exiger de quelqu'un.e (de parfaitement rationnel.le) que ses estimations des probabilités de divers événements satisfassent les axiomes de la définition.

2. Dans le livre de Ross, le premier axiome indique aussi que les probabilités sont bornées supérieurement par 1. Nous verrons plus loin que cela est superflu.

Pour finir, nous regroupons l'ensemble fondamental, la tribu des événements et la mesure de probabilités dans cette définition :

DÉFINITION 2.6 (Espace de probabilités). Soit Ω un ensemble non-vide, $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ une tribu d'événements de Ω , et $\mathbb{P} : \mathcal{E} \rightarrow \mathbb{R}$ une mesure de probabilités.

On appelle **espace de probabilités** le triplet $(\Omega, \mathcal{E}, \mathbb{P})$.

2.3. Propriétés de la mesure de probabilités

Les axiomes de la définition de mesure de probabilités sont assez minimaux. Toutefois, ils ont certaines conséquences qu'il est facile de démontrer rapidement :

PROPOSITION 2.1. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. La mesure de probabilités \mathbb{P} a les propriétés suivantes :*

- i. $\mathbb{P}\{\emptyset\} = 0$.
- ii. Pour tous événements $E, F \in \mathcal{E}$ tels que $E \subseteq F$, on doit avoir que $\mathbb{P}\{E\} \leq \mathbb{P}\{F\}$. En particulier, pour tout événement $E \in \mathcal{E}$, $\mathbb{P}\{E\} \leq 1$.
- iii. Pour tout événement $E \in \mathcal{E}$, $\mathbb{P}\{E^c\} = 1 - \mathbb{P}\{E\}$.

DÉMONSTRATION. i. Par l'axiome ii, les probabilités s'additionnent pour les événements disjoints. Bien sûr, on sait que

$$E \sqcup \emptyset = E.$$

Mais par l'axiome ii, on doit avoir, pour tout E ,

$$\mathbb{P}\{E\} = \mathbb{P}\{E \sqcup \emptyset\} = \mathbb{P}\{E\} + \mathbb{P}\{\emptyset\},$$

ce qui signifie immédiatement que $\mathbb{P}\{\emptyset\} = 0$.

- ii. Par l'axiome ii, on doit avoir que

$$\mathbb{P}\{F\} = \mathbb{P}\{E\} + \mathbb{P}\{F \setminus E\},$$

puisque $F \setminus E \subseteq E^c$ est disjoint de E .

Mais par l'axiome i, les probabilités sont positives. On a donc

$$\mathbb{P}\{F\} - \mathbb{P}\{E\} = \mathbb{P}\{F \setminus E\} \geq 0,$$

et on conclue que $\mathbb{P}\{F\} \geq \mathbb{P}\{E\}$. En prenant $F = \Omega$, on sait que tous les événements sont des parties de Ω , et par conséquent, $\mathbb{P}\{E\} \leq \mathbb{P}\{\Omega\} = 1$ par l'axiome iii.

- iii. Encore par l'axiome ii, clairement,

$$\mathbb{P}\{\Omega\} = \mathbb{P}\{E\} + \mathbb{P}\{E^c\},$$

puisque, par définition, $E \cap E^c = \emptyset$, et $E \sqcup E^c = \Omega$. Mais par l'axiome iii, $\mathbb{P}\{\Omega\} = 1$.

On a donc finalement

$$\mathbb{P}\{E^c\} = 1 - \mathbb{P}\{E\}.$$

□

Les dernières propriétés importantes concernent les réunions d'événements pas forcément disjoints.

Nous aurons d'abord besoin de ce lemme :

LEMME 2.1. Soit $(E_i \in \mathcal{E})_{i \in \mathbb{N}}$ une famille dénombrable d'événements.
Alors, si on définit

$$F_1 = E_1, \quad F_{n+1} = E_{n+1} \setminus \bigcup_{k=1}^n F_k,$$

on a les propriétés suivantes :

- i. Les F_i sont deux-à-deux disjoints ;
- ii. $F_n \subseteq E_n$ pour tout $n \in \mathbb{N}$.
- iii. Pour tout $n \in \mathbb{N}$, on a

$$\bigcup_{i=1}^n E_i = \bigcup_{i=1}^n F_i;$$

- iv. On a également

$$\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} F_i.$$

DÉMONSTRATION. La preuve est laissée en exercice (exercice 2.7). □

PROPOSITION 2.2 (Borne d'union). Soit $(E_i \in \mathcal{E})_{i \in \mathbb{N}}$ une famille dénombrable d'événements.

Alors,

$$\mathbb{P} \left\{ \bigcup_{i=1}^{\infty} E_i \right\} \leq \sum_{i=1}^{\infty} \mathbb{P} \{E_i\}.$$

DÉMONSTRATION. La preuve de cette proposition repose sur le lemme 2.1. Soit $(F_i \in \mathcal{E})_{i \in \mathbb{N}}$ une famille dénombrable définie comme dans ce lemme. Alors, puisque les F_i sont deux-à-deux disjoints

$$\begin{aligned} \mathbb{P} \left\{ \bigcup_{i=1}^{\infty} E_i \right\} &= \mathbb{P} \left\{ \bigcup_{i=1}^{\infty} F_i \right\} \\ &= \sum_{i=1}^{\infty} \mathbb{P} \{F_i\}. \end{aligned}$$

Mais puisque $F_i \subseteq E_i$ pour tout $i \in \mathbb{N}$, alors pour tout $i \in \mathbb{N}$, on a

$$\mathbb{P} \{F_i\} \leq \mathbb{P} \{E_i\},$$

et par conséquent, on a que

$$\mathbb{P} \left\{ \bigcup_{i=1}^{\infty} E_i \right\} = \sum_{i=1}^{\infty} \mathbb{P} \{F_i\} \leq \sum_{i=1}^{\infty} \mathbb{P} \{E_i\}.$$

□

Finalement, voici une généralisation de la formule du crible, pour les mesures de probabilités.

PROPOSITION 2.3 (Formule du crible). *Soient $E_1, E_2, \dots, E_N \in \mathcal{E}$ des événements. Alors, on a que*

$$\mathbb{P} \left\{ \bigcup_{i=1}^n E_i \right\} = \sum_{n=1}^N (-1)^{n-1} \sum_{i_1 < \dots < i_n} \mathbb{P} \left\{ \bigcap_{j=1}^n E_{i_j} \right\}.$$

DÉMONSTRATION. La preuve fonctionne exactement comme la preuve de la même formule pour les cardinalités, faite au chapitre 1.

En fait, cette preuve n'utilise que l'axiome **ii** de la définition de mesure de probabilités. Cet axiome étant commun à toutes les mesures, la formule du crible fonctionne aussi bien pour toutes les mesures. \square

Nous allons montrer un corollaire de la formule du crible qui sera très pratique.

Il s'agit d'une légère relaxation des hypothèses pour l'additivité. L'axiome **ii** dit qu'on peut additionner les probabilités d'événements disjoints deux-à-deux pour obtenir la probabilité de leur réunion.

Ce que nous allons faire, c'est montrer qu'on peut encore simplement additionner lorsque les événements sont « presque disjoints ».³

DÉFINITION 2.7. Dans un espace de probabilité $(\Omega, \mathcal{E}, \mathbb{P})$, deux événements $E, F \in \mathcal{E}$ sont **presque disjoints** si

$$\mathbb{P}\{E \cap F\} = 0.$$

COROLLAIRE 2.1. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. Si E_1, E_2, \dots, E_n sont des événements deux-à-deux presque disjoints (c'est-à-dire que pour tout $i \neq j$, on a $\mathbb{P}\{E_i \cap E_j\} = 0$, alors

$$\mathbb{P} \left\{ \bigcup_{i=1}^n E_i \right\} = \sum_{i=1}^n \mathbb{P}\{E_i\}.$$

DÉMONSTRATION. Si les événements sont deux-à-deux presque disjoints, alors toutes leurs intersections multiples ont probabilité 0, et elles ne contribuent donc pas à la somme. Les seuls termes qui restent sont les probabilités des événements individuels, et on retrouve

$$\mathbb{P} \left\{ \bigcup_{i=1}^n E_i \right\} = \sum_{i=1}^n \mathbb{P}\{E_i\}.$$

\square

Cette propriété est très utile car il arrive souvent que des événements ne soient pas totalement disjoints, mais que, néanmoins, leur intersection ne se réalise presque jamais (c'est-à-dire avec probabilité 0). Dans ces cas, on peut les traiter à toutes fins pratiques comme s'ils étaient disjoints.

2.4. La mesure équiprobable

Nous avons posé les bases de la théorie, et donné des axiomes, et des propriétés importantes. Maintenant, il est temps de passer à l'étape pratique, en définissant pour nous-mêmes notre première mesure de probabilité.

3. Comme on le verra plus loin, en théorie des probabilités et, plus largement, en théorie de la mesure, le mot « presque » possède un sens très précis.

Pour commencer, nous allons nous concentrer sur des expériences pour lesquelles l'ensemble fondamental Ω est fini.

EXEMPLE 2.5. Voici des exemples d'expériences avec un nombre fini de résultats possibles :

- Jouer à pile ou face un nombre fini de fois.
- Lancer un nombre fini de dés un nombre fini de fois.
- Piger un nombre fini de boules d'un sac qui en contient un nombre fini.
- ...

Ces problèmes ne sont pas toujours les plus fascinants, mais ils ont le mérite d'être simples à aborder. En prime, on pourra toujours prendre $\mathcal{E} = \mathcal{P}(\Omega)$ dans ces cas.

Dans la plupart de ces expériences, il est naturel de s'imaginer que chacun des résultats possibles (chacun des éléments de l'ensemble fondamental) survient de façon équiprobable. C'est à dire que si $\Omega = \{1, 2, 3, \dots, N\}$, alors on devrait avoir

$$\mathbb{P}\{\{1\}\} = \mathbb{P}\{\{2\}\} = \dots = \mathbb{P}\{\{N\}\},$$

ce qui nous conduit à

$$\mathbb{P}\{\{i\}\} = \frac{1}{N}$$

pour tout $i \leq N$. On fait donc la définition suivante :

DÉFINITION 2.8. Soit Ω un ensemble fini non-vide, et $\mathcal{E} = \mathcal{P}(\Omega)$. On définit la **mesure équiprobable** $\mathbb{P} : \mathcal{E} \rightarrow [0, 1]$ pour tout $E \subseteq \Omega$ par

$$\mathbb{P}\{E\} = \frac{|E|}{|\Omega|}.$$

PROPOSITION 2.4. Soit Ω un ensemble fini non-vide et soit \mathbb{P} la mesure équiprobable sur Ω . \mathbb{P} est une mesure de probabilités.

DÉMONSTRATION. Il suffit de vérifier les axiomes des mesures de probabilités. La preuve est laissée à l'exercice 2.8. \square

EXEMPLE 2.6. On joue deux parties consécutives de pile ou face. Quelle est la probabilité

(a) d'obtenir au moins une fois face ?

(b) d'obtenir deux fois le même résultat, pile ou face ?

SOLUTION. (a) Ici, l'ensemble fondamental pourrait être $\Omega = \{PP, PF, FP, FF\}$, et l'événement «on obtient au moins une fois face» serait $E = \{PF, FP, FF\}$. Bien sûr, la probabilité serait donc

$$\mathbb{P}\{E\} = \frac{|E|}{|\Omega|} = \frac{3}{4}.$$

(b) L'événement « on obtient deux fois le même résultat » serait $M = \{PP, FF\}$, et il aurait évidemment probabilité $1/2$.

EXEMPLE 2.7. Si on lance en même temps cinq dés ordinaires à six faces,

(a) quelle est la probabilité que les dés affichent une suite ?

(b) quelle est la probabilité que les dés affichent cinq fois le même chiffre ?

- (c) quelle est la probabilité que les dés affichent au moins un 1 ou un 5 ?

SOLUTION. Ici, notre ensemble fondamental serait $\Omega = \{1, 2, 3, 4, 5, 6\}^5$, avec $|\Omega| = 6^5 = 7776$.

- (a) Si E_a est l'événement que les dés affichent une suite, alors E_a comprend exactement toutes les permutations possibles de $(1, 2, 3, 4, 5)$ ou $(2, 3, 4, 5, 6)$. En tout, il comprend donc $2 \times 5!$ résultats, et $|E_a| = 2 \times 5! = 240$.

La probabilité de l'événement E_a est donc

$$\mathbb{P}\{E_a\} = \frac{|E_a|}{|\Omega|} = \frac{2 \times 5!}{6^5} = \frac{240}{7776} = \frac{5}{162}.$$

- (b) Si E_b est l'événement que les dés affichent cinq fois le même chiffre, alors on peut l'écrire en extension : $E_b = \{(1, 1, 1, 1, 1), (2, 2, 2, 2, 2), \dots, (6, 6, 6, 6, 6)\}$, et $|E_b| = 6$. Donc, on a que

$$\mathbb{P}\{E_b\} = \frac{|E_b|}{|\Omega|} = \frac{1}{6^4} = \frac{1}{1296}.$$

- (c) Soit E_c l'événement que les dés affichent au moins un 1 ou un 5. Alors, l'événement complémentaire, E_c^c est l'événement qu'aucun des cinq dés n'affichent ni de 1, ni de 5. On a donc que $|E_c^c| = 4^5$, puisque chaque dé peut afficher l'un des 4 autres résultats.

On a donc que

$$\mathbb{P}\{E_c^c\} = \frac{4^5}{6^5} = \frac{32}{243}.$$

Maintenant, on voulait

$$\mathbb{P}\{E_c\} = 1 - \mathbb{P}\{E_c^c\} = \frac{211}{243},$$

soit environ 86%.

EXEMPLE 2.8. En assumant que, sur une période d'un an, les naissances sont équiprobablement réparties, dans un groupe de 130 étudiant.e.s, en ignorant les années bissextiles, quelle est la probabilité

- (a) qu'au moins un.e étudiant.e soit né.e un 17 janvier ?
 (b) qu'au moins une paire d'étudiant.e.s partagent le même anniversaire ?

SOLUTION. Pour ce problème, nous allons assumer que l'ensemble des anniversaires possibles est $\{1, 2, \dots, 365\}$; pour un groupe de N personnes, on a donc que notre ensemble fondamental sera $\Omega = \{1, 2, \dots, 365\}^N$. Pour $N = 130$, le cardinal de Ω est de l'ordre de 10^{333} .

- (a) On va noter A l'événement «Au moins un.e étudiant.e est né.e un 14 janvier». On cherche donc $\mathbb{P}\{A\}$.

On va d'abord considérer A^c , l'événement qu'aucun.e étudiant.e n'est né.e un 14 janvier. Pour calculer $\mathbb{P}\{A^c\}$, il suffit de trouver $|A^c|$. Ça n'est pas très difficile : on veut simplement savoir combien il y a de répartitions possibles des anniversaires, si tout le monde a un anniversaire qui n'est pas le 14 janvier.

Mais alors, pour tout le monde, on a le choix de 364 autres anniversaires ; Il y a donc $\underbrace{364 \times 364 \times \cdots \times 364}_{N \text{ fois}} = 364^N$ possibilités. Donc $|A^c| = 364^N$.

Finalement,

$$\mathbb{P}\{A\} = 1 - \mathbb{P}\{A^c\} = 1 - \frac{364^N}{365^N}.$$

Pour $N = 130$, ça donne une probabilité d'environ 29,985%.

- (b) On va noter B l'événement «Au moins une paire d'étudiant.e.s partagent le même anniversaire». Encore une fois, on s'intéresse d'abord à B^c , l'événement complémentaire qu'aucune paire d'étudiant.e.s ne partagent leur anniversaire – ou que tou.te.s les étudiant.e.s ont un anniversaire différent.

Il faut donc compter le nombre de répartition possibles des anniversaires tels que tout le monde a un anniversaire différent. On commence par choisir N anniversaires différents parmi 365 – il y a $\binom{365}{N}$ possibilités. Puis on les attribue aux N personnes, selon l'une de $N!$ permutations. Finalement, il y a $\binom{365}{N} \times N! = \frac{365!}{(365-N)!}$ possibilités.

Donc $|B^c| = \frac{365!}{(365-N)!}$, et

$$\mathbb{P}\{B^c\} = \frac{\frac{365!}{(365-N)!}}{365^N}.$$

Pour $N = 130$, $\mathbb{P}\{B^c\}$ est environ $3,76 \times 10^{-10}\%$.

Finalement,

$$\mathbb{P}\{B\} = 1 - \mathbb{P}\{B^c\} = 1 - \frac{\frac{365!}{(365-N)!}}{365^N};$$

pour $N = 130$, c'est environ 99,9999999996%.

EXEMPLE 2.9. Au brunch de fin de session de l'association des étudiant.e.s gradué.e.s, chacun.e des N étudiant.e.s arrive à 11h et empile son manteau dans un coin. Après le *last call* pour les mimosas à volonté, à 16h, tout le monde s'en va chaud raide en prenant au hasard un manteau de la pile (et sans conduire, car tout le monde est très responsable).

- (a) Quelle est la probabilité que tout le monde prenne le bon manteau ?
(b) Quelle est la probabilité que personne ne prenne le bon manteau ? Qu'arrive-t-il lorsque N tend vers l'infini ?

SOLUTION. On va considérer que chacune de nos N personnes sont numérotées de 1 à N . Alors, notre ensemble fondamental pourrait être l'ensemble des $N!$ permutations qui indiquent qui prend quel manteau.

- (a) Il n'y a qu'une seule permutation où chaque personne prend son manteau. Donc la probabilité ici est de $1/N!$.
(b) Si B est l'événement «personne ne prend le bon manteau», on va considérer plutôt l'événement B^c : « au moins une personne prend le bon manteau.»

Si E_i est l'événement «la personne i prend le bon manteau», alors on a le résultat suivant :

$$B^c = \bigcup_{i=1}^N E_i.$$

Par la formule du crible pour les probabilités (proposition 2.3), on a donc que

$$\mathbb{P}\{B^c\} = \sum_{n=1}^N (-1)^{n-1} \sum_{i_1 < \dots < i_n} \mathbb{P}\left\{\bigcap_{j=1}^n E_{i_j}\right\}.$$

Ici, $\mathbb{P}\left\{\bigcap_{j=1}^n E_{i_j}\right\}$ est simplement la probabilité que les personnes i_1, i_2, \dots, i_n reprennent toutes le bon manteau.

Bien entendu, si n personnes reprennent le bon manteau, alors il reste $(N - n)$ personnes qui peuvent reprendre leur manteau n'importe comment entre eux et elles. Ça fait donc $(N - n)!$ façons de distribuer les manteaux si n personnes «se retirent du jeu» en prenant leur manteau.

Donc, pour n'importe quelle combinaison $i_1 < i_2 < \dots < i_n \leq N$ d'indices, on a que

$$\mathbb{P}\left\{\bigcap_{j=1}^n E_{i_j}\right\} = \frac{(N - n)!}{N!}.$$

Mais il y a, dans notre somme, exactement $\binom{N}{n}$ tels termes. Finalement, notre formule du crible devient :

$$\mathbb{P}\{B^c\} = \sum_{n=1}^N (-1)^{n-1} \binom{N}{n} \frac{(N - n)!}{N!} = \sum_{n=1}^N (-1)^{n-1} \frac{1}{n!}.$$

Donc,

$$\mathbb{P}\{B\} = 1 - \mathbb{P}\{B^c\} = 1 - \sum_{n=1}^N (-1)^{n-1} \frac{1}{n!} = \sum_{n=0}^N \frac{(-1)^n}{n!}.$$

Et dans la limite où N tend vers l'infini, on a bien sûr que

$$\lim_{N \rightarrow \infty} \mathbb{P}\{B\} = e^{-1} = \frac{1}{e}.$$

ÉNIGME. Ce résultat est contre-intuitif. Ça vaut la peine de réfléchir à quelques questions :

1. Pourquoi la limite n'est pas 0 ? Plus il y a de monde, plus il devrait y avoir de chance que quelqu'un reparte avec son manteau, non ?
2. Pourquoi la convergence n'est pas monotone ? Remarquez que notre suite de probabilités pour N personnes est une série alternée, tronquée à N termes – ça signifie ici que pour un nombre N pair de personnes, c'est plus probable que personne ne reparte avec son manteau, que si on avait $N + 1$ ou $N - 1$ personnes.

EXEMPLE 2.10 (Adapté de l'exemple 5o). Si on joue à pile ou face $m + n$ fois, et qu'on est tombés m fois sur «pile» et n fois sur «face», quelle est la probabilité qu'il y ait eu k séquences contigues de «pile» ? Par exemple, dans la séquence «PFFPFPPPFPPFPPF», il y a 5 groupes contigus de «pile».

SOLUTION. Ici, notre ensemble fondamental Ω est l'ensemble de toutes les combinaisons possibles de m «pile» et n «face». On a donc $|\Omega| = \binom{m+n}{m}$.

On note E_k l'événement «le nombre total de groupes contigus de «pile» et de «face» est k ».

Supposons qu'on accepterait aussi les «groupes contigus de 0 «pile»». Alors, il y en aurait toujours $n+1$ – séparés par des F . Et on pourrait noter x_i le nombre de «pile» dans le i ème groupe contigu de «piles». Comme il y a m «pile» en tout, on aurait alors :

$$x_1 + x_2 + \cdots + x_{n+1} = m.$$

Notre problème, c'est qu'on veut qu'il y ait exactement k «vrais» groupes contigus – c'est à dire avec $x_i > 0$.

Au final, ce qu'on cherche, c'est le nombre total de solutions de l'équation

$$x_1 + x_2 + \cdots + x_{n+1} = m,$$

où exactement $n+1-k$ variables sont nulles.

Par un résultat montré à l'exercice 1.16, ce nombre est donné par

$$\binom{n+1}{k} \binom{m-1}{m-k}.$$

On a finalement que la probabilité qu'on ait k groupes contigus de «pile» et de «face» est

$$\mathbb{P}\{E_k\} = \frac{\binom{n+1}{k} \binom{m-1}{m-k}}{\binom{m+n}{m}}.$$

Par exemple, pour $m = n = 5$, et $k = 5$, on a que le nombre de combinaisons à 5 «groupes contigus d'au moins 1 pile» est donné par $\binom{5+1}{5} \binom{5-1}{5-5} = \binom{6}{5} \binom{4}{0} = 6$, qui correspond aux 6 possibilités :

$$\begin{array}{ll} \text{FPFPFPFPFP} & \text{PFPFPFPFP} \\ \text{PFFPFPPFPFP} & \text{PFPFPFPFPFP} \\ \text{PFPFFPFPFP} & \text{PFPFPFPFPFP} \end{array}$$

Et la probabilité est donc

$$\mathbb{P}\{E_5\} = \frac{6}{252} = \frac{1}{42}.$$

2.5. Bonus : Les limites de suites d'événements.

L'une des choses qu'on voudra souvent faire, en probabilités, c'est de prendre des limites. Il sera important de savoir comment la mesure de probabilités interagit avec les limites si on veut pouvoir se débrouiller.

Il existe certaines suites d'événements qui admettent une notion de limite bien définie. Ce sujet est abordé dans la section A.8 de l'annexe sur la théorie des ensemble, à la fin de ces notes.

Toutefois, nous rappelons les notions importantes ici.

2.5.1. Les suites d'événements monotones. On dit qu'une suite d'événements $(E_n \in \mathcal{E})_{n \in \mathbb{N}}$ est **monotone croissante** (resp. **décroissante**) si on a que, pour tout $n \in \mathbb{N}$, $E_n \subseteq E_{n+1}$ (resp. \supseteq).

Ces suites sont particulières ; si $(E_n)_{n \in \mathbb{N}}$ est une suite croissante (resp. décroissante), nécessairement

$$\bigcup_{i=1}^n E_i = E_n \quad (\text{resp. } \bigcap),$$

et on peut alors définir

$$\lim_{n \rightarrow \infty} E_n = \bigcup_{i \in \mathbb{N}} E_i \quad (\text{resp. } \bigcap).$$

Moralement, la limite d'une suite d'événements croissants est l'événement qui se réalise si au moins un des événements de la suite est réalisé. La limite d'une suite d'événements décroissants est l'événement qui se réalise si *tous* les événements de la suite sont réalisés.

Pour trouver des exemples pertinents, il faut sortir du cadre fini.

EXEMPLE 2.11. Considérons $\Omega = \mathbb{N}$,

- Les événements $E_n = \{k \in \mathbb{N} : k \leq n\}$ ont Ω pour limite ;
 - Les événements $F_n = \{k \in \mathbb{N} : k \text{ n'est pas un multiple des } n \text{ premiers nombres premiers}\}$ on $F = \{k \in \mathbb{N} : k \text{ est premier}\}$ pour limite.
- Si on considère $\Omega = \mathbb{R}$,
- Les événements $G_n = [-n, n]$ convergent vers \mathbb{R} .
 - Les événements $H_n = (-1/n, 1/n)$ convergent vers $\{0\}$.

L'intérêt, c'est que si on accepte que certaines suites d'événements (principalement les suites d'événements monotones) admettent des limites, alors on peut prouver la proposition suivante :

PROPOSITION 2.5 (Continuité des probabilités). *Soit $(E_n)_{n \in \mathbb{N}}$ une suite d'événements monotone (croissante ou décroissante).⁴ Alors,*

$$\lim_{n \rightarrow \infty} \mathbb{P}\{E_n\} = \mathbb{P}\left\{\lim_{n \rightarrow \infty} E_n\right\}.$$

DÉMONSTRATION. Nous allons d'abord montrer que c'est vrai si E_n est croissante. On peut alors définir

$$F_n = E_n \setminus \bigcup_{k=1}^{n-1} E_k,$$

4. En fait il suffit que la suite d'événements admette une limite. Voir A.8.

une suite d'événements disjoints entre eux (lemme 2.1), et on calcule alors que

$$\begin{aligned}
 \mathbb{P} \left\{ \lim_{n \rightarrow \infty} E_n \right\} &= \mathbb{P} \left\{ \bigcup_{k \in \mathbb{N}} E_k \right\} \\
 &= \mathbb{P} \left\{ \bigcup_{k \in \mathbb{N}} F_k \right\} \\
 &= \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{P} \{F_k\} \\
 &= \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{k=1}^n F_k \right\} \\
 &= \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{k=1}^n E_k \right\} \\
 &= \lim_{n \rightarrow \infty} \mathbb{P} \{E_n\}.
 \end{aligned}$$

Si E_n est plutôt décroissante, alors E_n^c est croissante, et bien sûr, on a (par les identités de De Morgan) que $(\lim_{n \rightarrow \infty} E_n)^c = \lim_{n \rightarrow \infty} E_n^c$. Or,

$$\begin{aligned}
 \mathbb{P} \left\{ \lim_{n \rightarrow \infty} E_n \right\} &= 1 - \mathbb{P} \left\{ (\lim_{n \rightarrow \infty} E_n)^c \right\} \\
 &= 1 - \lim_{n \rightarrow \infty} \mathbb{P} \{E_n^c\} \\
 &= 1 - \lim_{n \rightarrow \infty} (1 - \mathbb{P} \{E_n\}) \\
 &= \lim_{n \rightarrow \infty} \mathbb{P} \{E_n\}.
 \end{aligned}$$

□

EXEMPLE 2.12 (Un paradoxe (exemple 7a)). Imaginez que vous avez une infinité de boules, numérotées par les nombres naturels. Vous avez également un grand seau qui peut accomoder autant de boules que vous ne le souhaitez.

Le processus est simple : Au temps $t_n = \sum_{k=1}^n 2^{-k} = 1 - 2^{-(n+1)}$, on mets les boules $\{10(n-1) + 1, \dots, 10n\}$ dans le seau. Puis on retire la boule $b(n)$.

À la fin du processus, au temps $t_\infty = 1$, on considère combien il y a de boules dans le seau. Par exemple, si $b(n) = 10n$, alors après que tout le processus ait été complété, au temps 1, on aura mis toutes les boules dans le seau, et retiré seulement les boules qui sont des multiples de 10. Il y aura donc encore une infinité de boules dans le seau.

Mais si $b(n) = n$, alors après que tout le processus ait été complété, au temps 1, on aura mis toutes les boules dans le seau, mais on les aura aussi toutes retirées, et le seau sera vide.

- (a) Supposons que je veux qu'il reste k boules dans le seau à la fin du processus. Est-ce que je peux trouver une fonction $b(n)$ de choix de boules qui laisse k boules dans le seau à la fin du processus, pour $k \in \mathbb{N}$?
- (b) Supposons que $b(n)$ est une boule choisie au hasard parmi celles qui sont dans l'urne au temps t_n . Quelle est la probabilité que le seau soit vide au temps 1 ?

SOLUTION.

Oui, il est possible de choisir une fonction $b(n)$ qui laisse k boules dans le seau. On choisit

$$b(n) = \begin{cases} 2n & \text{si } n \leq k \\ k+n & \text{si } n > k \end{cases},$$

qui va prendre toutes les boules sauf les boules $\{1, 3, 5, \dots, 2k-1\}$ – et donc laisser un nombre de k boules dans la boîte. De plus il est clair que $b(n) \leq 10n$ pour tout n .

Si $b(n)$ est une boule choisie au hasard parmi celles qui sont dans le seau au temps t_n , juste après avoir ajouté les boules $10(n-1)+1$ à $10n$, alors au total, il y a $9(n-1)+10$ boules dans le seau au temps n .

On va noter l'événement E_n que la boule #1 après le retrait de la boule au temps t_n .

Notre ensemble fondamental sera l'ensemble des séquences possibles pour retirer des boules.

Après avoir retiré n boules, il y a eu

$$\prod_{k=1}^n (9k+1) = 10 \times 19 \times \cdots \times 9n+1$$

tirages possibles.

Mais de ceux-là, seulement

$$\prod_{k=1}^n (9k) = 9 \times 18 \times \cdots \times 9n$$

tirages laissent chaque fois la boule #1 dans le seau.

Donc,

$$\mathbb{P}\{E_n\} = \frac{\prod_{k=1}^n 9k}{\prod_{k=1}^n (9k+1)}.$$

Puisque les événements E_n sont décroissants, il suit que l'événement $\lim_{n \rightarrow \infty} E_n = \bigcap_{k \in \mathbb{N}} E_k$ que la boule #1 soit encore dans le seau au temps 1, est simplement

$$\mathbb{P}\left\{\lim_{n \rightarrow \infty} E_n\right\} = \lim_{n \rightarrow \infty} \mathbb{P}\{E_n\} = \lim_{n \rightarrow \infty} \prod_{k=1}^n \frac{9k}{9k+1},$$

Bien sûr,

$$\begin{aligned} \prod_{k=1}^n \frac{9k+1}{9k} &= \prod_{k=1}^n \left(1 + \frac{1}{9k}\right) \\ &\geq \sum_{k=1}^n \frac{1}{9k} \end{aligned}$$

Puisque la série diverge, alors $\prod_{k=1}^n \frac{9k+1}{9k}$ tend vers $+\infty$ et $\prod_{k=1}^n \frac{9k}{9k+1}$ tend vers 0.

On a donc finalement que $\mathbb{P}\{\lim_{n \rightarrow \infty} E_n\} = 0$ – soit que la probabilité que la boule #1 soit restée dans le seau après tout le processus est égale à 0.

Soit $F_1 = \lim_{n \rightarrow \infty} E_n$ l'événement la boule #1 reste dans le seau. On définit de façon similaire F_i l'événement que la boule # i reste dans le seau, et on montre, de façon tout à fait analogue, que $\mathbb{P}\{F_i\} = 0$ pour tout i .

Mais la probabilité que le seau contienne au moins une boule est

$$\mathbb{P} \left\{ \bigcup_{i \in \mathbb{N}} F_i \right\} \leq \sum_{i=1}^{\infty} \mathbb{P} \{ F_i \} = 0,$$

et donc la probabilité que le seau soit vide est 1.

2.6. Le mot « presque »

À travers ces notes de cours, et plus largement à travers la littérature mathématique traitant de probabilités, on croisera très souvent le mot « presque ». Par exemple, on dira qu'un événement se produit « presque sûrement ».

Il vaut donc la peine, à la fin de ce chapitre, de prendre un peu de temps pour mentionner le sens précis de ce terme.

DÉFINITION 2.9 (Presque). En théorie de la mesure, le terme **presque** (ou **almost** en anglais) signifie toujours : « *sauf pour un ensemble de mesure nulle* ».

Lorsqu'une confusion est possible, on fait précéder le terme « presque » du symbole de la mesure pertinente.

Comme nous venons de le voir, le terme « presque » a un sens mathématique très précis dans le contexte général de la théorie de la mesure. Comme les mesures de probabilités sont des mesures, il suit inévitablement qu'on peut employer ce terme en probabilités.

EXEMPLE 2.13. Supposons qu'on lance une pièce de monnaie de manière répétée. On dira alors que l'événement $E = \{\text{La pièce tombera sur Face éventuellement}\}$ est réalisé presque sûrement. Cela signifie que cet événement se réalise avec probabilité 1 :

$$\mathbb{P} \{ E \} = 1.$$

Pour les espaces de probabilités où l'ensemble fondamental est fini, cela a peu d'importance – on peut facilement montrer que le seul événement de probabilité nulle est l'événement vide.

Par contre, ainsi qu'on peut le voir à l'exemple 2.13, dans le cas des espaces avec ensemble fondamental infini, la distinction s'impose.

En effet, il n'est pas évident *a priori* que la probabilité d'obtenir Face éventuellement soit 1. Si, par exemple, on considère l'ensemble fondamental constitué de toutes les suites possibles infinies de P et de F , cet ensemble comprend entre autres éléments la suite constituée uniquement de P à l'infini – c'est la possibilité où on lancerait Pile à chaque fois, pour toujours. C'est l'événement E^c , et il est non-vide.

Cet événement peut, techniquement, être réalisé. Par contre, (et c'est un bon exercice), on peut montrer qu'il a une probabilité nulle, et ce malgré le fait que l'événement soit non-vide.

C'est dans ces cas-là qu'on va employer le terme « presque » pour marquer la distinction : l'événement est réalisé *presque sûrement* – c'est à dire qu'il recouvre tout Ω , *sauf pour un événement de mesure nulle*.

On utilisera parfois l'abréviation p.s. pour « presque sûrement » ; c'est le cas de figure le plus courant, qui signifie : « avec probabilité 1. ».

EXEMPLE 2.14. Nous avons déjà vu le mot « presque » dans le contexte des ensembles « presque disjoints » – remarquez que deux ensembles presque disjoints ont une intersection vide *sauf pour un ensemble de mesure nulle*. L'emploi du mot « presque » dans ce contexte est donc parfaitement cohérent avec la définition que nous venons de donner.

2.7. Exercices

EXERCICE 2.1. Décrivez des ensembles fondamentaux pour chacune expériences aléatoires suivantes :

- (a) la pige initiale des lettres au Scrabble ;
- (b) le tirage des numéros au lotto 6/49 ;
- (c) l'attribution des notes des étudiant.e.s par la méthode de l'escalier (qui consiste à garrocher les copies d'examen dans l'escalier, et d'attribuer les notes en ordre de hauteur dans l'escalier) ;
- (d) un lancer de dart les yeux bandés ;
- (e) un singe qui tape sur une machine à écrire ;
- (f) un clic sur le bouton « article au hasard » de Wikipedia.

EXERCICE 2.2. Soient E , F et G trois événements. En notation ensembliste, trouvez des expressions pour les événements suivants :

- (a) seulement E est réalisé ;
- (b) seulement E et G sont réalisés ;
- (c) Au moins l'un des trois se réalise ;
- (d) Au moins deux des trois se réalisent ;
- (e) Les trois se réalisent ;
- (f) Aucun des trois ne se réalise ;
- (g) Au plus un se réalise ;
- (h) Au plus deux se réalisent ;
- (i) Exactement deux se réalisent ;
- (j) Au plus trois se réalisent ;

EXERCICE 2.3. Soit $(E_i)_{i \in \mathbb{N}}$ une famille dénombrable d'événements. En notation ensembliste, trouvez des expressions pour les événements suivants :

- (a) le premier événement à se réaliser est l'événement k ;
- (b) tous les événements sont réalisés à partir de l'événement k ;
- (c) tous les événements sont réalisés sauf pour un nombre fini ;
- (d) Une infinité d'événements est réalisée.

EXERCICE 2.4. On lance deux dés à six faces.

- (a) Décrire l'ensemble fondamental Ω .

Soit E l'événement «la somme des deux dés est impaire», F l'événement «l'un des deux dés tombe sur 1», et G , «La somme des deux dés est 5».

- (b) Écrire E , F et G en extension.

(c) Décrivez les événements suivants :

- i. $E \cap F$;
- ii. $E \cup F$;
- iii. $F \cap G$;
- iv. $E \cap F^c$;
- v. $E \cap F \cap G$;

EXERCICE 2.5. Montrer que les exemples du début de la section 2.2 sont bel et bien des mesures positives, en accordance avec la définition 2.4

EXERCICE 2.6. Nous allons montrer d'autres propriétés de la mesure de probabilités.

(a) Soient $E, F \in \mathcal{E}$ deux événements tels que $F \subseteq E$, et \mathbb{P} une mesure de probabilités. Alors, montrer que

$$\mathbb{P}\{E \setminus F\} = \mathbb{P}\{E\} - \mathbb{P}\{F\}.$$

(b) Soient deux événements $G, H \in \mathcal{E}$. On note $G \Delta H$ la *différence symétrique* entre G et H ; c'est l'événement où exactement l'un de G ou H est réalisé. Montrer que

$$\mathbb{P}\{G \Delta H\} = \mathbb{P}\{G\} + \mathbb{P}\{H\} - 2\mathbb{P}\{G \cap H\}.$$

EXERCICE 2.7. Prouver le lemme suivant (lemme 2.1) :

LEMME. Soit $(E_i \in \mathcal{E})_{i \in \mathbb{N}}$ une famille dénombrable d'événements.
Alors, si on définit

$$F_1 = E_1, \quad F_{n+1} = E_{n+1} \setminus \bigcup_{k=1}^n F_k,$$

on a les propriétés suivantes :

- i. Les F_i sont deux-à-deux disjoints ;
- ii. $F_n \subseteq E_n$ pour tout $n \in \mathbb{N}$.
- iii. Pour tout $n \in \mathbb{N}$, on a

$$\bigcup_{i=1}^n E_i = \bigcup_{i=1}^n F_i;$$

iv. On a également

$$\bigcup_{i=1}^{\infty} E_i = \bigcup_{i=1}^{\infty} F_i.$$

EXERCICE 2.8. Soit Ω non-vide tel que $|\Omega| < +\infty$, soit $\mathcal{E} = \mathcal{P}(\Omega)$ et soit $\mathbb{P} : \mathcal{E} \rightarrow \mathbb{R}$ donnée par

$$\mathbb{P}\{E\} = \frac{|E|}{|\Omega|}.$$

Montrer que \mathbb{P} est une mesure de probabilité. (C'est-à-dire, montrer qu'elle respecte les axiomes de la définition 2.5.)

EXERCICE 2.9. Noé abandonne l'assiduité ; il s'en remet à la grâce du SEIGNEUR : il choisira aléatoirement parmi un groupe de N éléphants (dont m mâles) les deux qui iront sur son arche.

- (a) Décrire l'ensemble fondamental si l'ordre de sélection est important. Quelle est la cardinalité de cet ensemble ?
- (b) Décrire l'événement « Le couple sélectionné est mixte » dans cet ensemble fondamental. En assumant une mesure de probabilité équiprobable, quelle est la probabilité que le couple sélectionné soit mixte ?

Maintenant, on va changer de représentation.

- (c) Décrire l'ensemble fondamental si l'ordre de sélection n'est pas important. Quelle est la cardinalité de cet ensemble ?
- (d) Décrire l'événement « le couple sélectionné est mixte » dans cet ensemble fondamental. En assumant une mesure de probabilité équiprobable, quelle est la probabilité que le couple sélectionné soit mixte ?

Remarquez que les résultats ne changent pas. C'est vrai en général : pour des expériences aléatoires qui consistent à choisir un certain nombre d'objets parmi N , on peut considérer que l'ordre importe, ou pas, selon nos besoins.

EXERCICE 2.10. On place dans un chapeau les lettres A, E, I, M, N, O, P. Puis, on les retire au hasard, une par une.

- (a) Quelle est la probabilité que les deux premières lettres forment un mot ?
- (b) Quelle est la probabilité que les trois premières lettres forment un mot ?
- (c) Quelle est la probabilité que les quatre premières lettres forment un mot ?
- (d) Quelle est la probabilité que le mot PIANO apparaisse éventuellement ?

Notez : J'utilise le dictionnaire officiel du Scrabble français.

EXERCICE 2.11. On assoit Caoimhe, Deidre, Finn, Gráinne, Róisín, Saoirse, Seamus, Sean, Sinead, Siobahn et Tadhg⁵ au hasard.

- (a) On les assoit autour d'une table ronde (l'orientation absolue n'importe pas). Quelle est la probabilité que tous les gens avec un prénom commençant par S soient assis.e.s ensemble ?
- (b) Et si on les avait plutôt assis.e.s en rangée, sur un banc ?
- (c) Si on les assoit sur un banc, quelle est la probabilité
 - i. qu'il n'y ait exactement une personne entre Caoimhe et Saoirse ?
 - ii. qu'il y ait deux personnes entre Caoimhe et Saoirse ?
 - iii. qu'il y ait trois personnes entre Caoimhe et Saoirse ?
- (d) Si on les assoit sur un banc, quelle est la probabilité que toutes les filles soient assises ensemble ?

5. Ce sont tous des prénoms traditionnels irlandais.

EXERCICE 2.12. En assumant que, sur une période de quatre ans, les naissances sont équiprobablement réparties, dans un groupe de 130 étudiant.e.s, quelle est la probabilité

- (a) qu'au moins un.e étudiant.e soit né.e un 29 février ?
- (b) qu'au moins une paire d'étudiant.e.s partagent le même anniversaire ?

EXERCICE 2.13. On brasse un paquet de 52 cartes à jouer standard. Quelle est la probabilité

- (a) que la 14^e carte soit un as ?
- (b) que le premier as soit la 14^e carte ?

EXERCICE 2.14. On lance une paire de dés jusqu'à ce que les dés affichent soit la somme de 5 ou la somme de 7. Trouver la probabilité que la somme de 5 apparaisse en premier.

Indice : noter E_n l'événement que ni le 5, ni le 7 sont apparus aux $n - 1$ premiers lancers, et que le *nième* lancer a donné une somme de 5. Puis, montrer que la probabilité recherchée est $\sum_{k=1}^{\infty} \mathbb{P}\{E_k\}$.

EXERCICE 2.15. On lance une paire de dés à répétition.

- (a) Quelle est la probabilité d'obtenir trois doubles de suite ?
- (b) Quelle est la probabilité d'obtenir au moins un double 6 en n lancers ?
- (c) Combien de lancers faudra-t-il faire pour que la probabilité d'obtenir au moins une paire de 6 soit supérieure à $1/2$?

EXERCICE 2.16. On place $2n + 1$ boules dans un grand sac ; n sont noires, et $n + 1$ sont blanches. Notre objectif, c'est d'utiliser ce sac pour choisir entre deux options, de façon la plus équitable possible, en faisant des tirages répétés.

Pour ce faire, on va considérer un événement A ; on choisira la première option si A est réalisé, et la seconde option si B est réalisé. On dira que le «biais» de notre méthode sera

$$b(A) = \frac{|\mathbb{P}\{A\} - \mathbb{P}\{A^c\}|}{2}.$$

Cette valeur vaut 1 lorsque notre méthode de choix est totalement biaisée, et 0 lorsqu'elle est parfaitement équitable.

L'objectif, c'est bien sûr de trouver un A tel que $b(A)$ soit le plus petit possible.

- (a) Si on pige une seule boule, et qu'on choisit l'événement A , « on pige une boule blanche », calculer $b(A)$.
- (b) Si on pige deux boules avec remise, et qu'on choisit l'événement A , « les deux boules sont de la même couleur », calculez $b(A)$. Cette méthode est-elle plus ou moins biaisée que la méthode précédente ?

Chapitre 3

Probabilités conditionnelles et indépendance



L'émission de télévision *Let's Make a Deal*, en ondes depuis 1963 sur le réseau américain CBS, a servi d'inspiration pour un célèbre problème de probabilités qui porte le nom de son premier et de son plus célèbre animateur...

Nous allons maintenant nous pencher sur les probabilités conditionnelles. Elles sont un outil conceptuel puissant, qui est directement lié à notre intuition, mais qui peut produire des résultats surprenants.

Nous parlerons aussi de la notion centrale *d'indépendance* en probabilités ; nous verrons sa définition et ses usages dans le contexte des événements.

3.1. Probabilités conditionnelles

Au cœur d'une expérience aléatoire, il y a cette idée que nous n'avons «aucune information» *a priori* sur le résultat de l'expérience aléatoire qui se déroulera.¹

Toutefois, il serait logique de se permettre de considérer l'expérience aléatoire qui résulterait si on disposait d'informations partielles sur le résultat aléatoire.

1. Nous pourrions discuter beaucoup plus longuement sur la notion – ici très vague – *d'information*. Elle a été formalisée d'abord par Claude Shannon pour les fins d'obtenir une théorie de l'information (et de l'informatique) plus solide. Sachons néanmoins qu'à ce sens, même rigoureux, l'affirmation reste « vraie » : on présume toujours qu'une expérience aléatoire «produira» de l'information lorsqu'elle se réalisera. La quantité exacte d'information produite (mesurée en bits) dépend d'à quel point il est probable que le résultat soit «surprenant» (improbable).

EXEMPLE 3.1. Quelqu'un lance deux dés derrière un paravent et vous demande de deviner quel est la somme indiquée par les dés.

- (a) Quelle prédition feriez-vous ?
- (b) Et si la personne vous disait que la somme n'est pas un nombre premier ?

SOLUTION. (a) Ici, notre ensemble fondamental est $\Omega = \{1, 2, 3, 4, 5, 6\}^2$, et bien sûr, la valeur la plus probable est 7, avec une probabilité de $1/6$. Ce serait clairement la meilleure prédition à faire.

- (b) Ici notre réponse change – on sait que le résultat n'est pas premier ! On peut donc éliminer 2, 3, 5, 7 et 11 de la liste des réponses possibles.

Parmi les possibilités restantes, il y a 4 (avec probabilité $1/12$), 6 ou 8 (chacunes avec probabilité $5/36$), 9 (avec probabilité $1/9$), 10 (avec probabilité $1/12$), et 12 (avec probabilité $1/36$).

On voudrait raisonner que, privés des options des nombres premiers, on devrait répondre 6 ou 8. Bien que la somme de ces probabilités n'est pas 1, si on les renormalisait en les gardant toutes proportionnelles on obtiendrait une «nouvelle» mesure de probabilités, recalibrée pour tenir compte de notre nouvelle information...

DÉFINITION 3.1 (Probabilité conditionnelle). Si $(\Omega, \mathcal{E}, \mathbb{P})$ est un espace de probabilités, et que $A, B \in \mathcal{E}$ sont des événements avec $\mathbb{P}\{B\} > 0$,² alors, on notera

$$\mathbb{P}\{A | B\}$$

la **probabilité conditionnelle** de A sachant B . Celle-ci est définie par :

$$(3.1.1) \quad \mathbb{P}\{A | B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}}.$$

La probabilité conditionnelle de A sachant B «restreint» l'ensemble des possibilités à seulement celles qui sont incluses dans l'événement B .

EXEMPLE 3.2. Toujours en considérant un lancer simultané de deux dés :

- (a) Calculer la probabilité que la somme soit 8 sachant que le premier dé est pair.
- (b) Calculer la probabilité que le second dé soit 3 sachant que le premier dé est 5.
- (c) Calculer la probabilité que la somme soit supérieure à cinq sachant que le premier dé est strictement plus grand que le second dé.

SOLUTION. (a) Si l'événement A_1 est «la somme est 8» et l'événement A_2 est «le premier dé est pair», alors, bien sûr, $\mathbb{P}A_1 \cap A_2 = 1/12$, car il y a seulement trois façons d'obtenir ce résultat : (2, 6), (4, 4) et (6, 2).

Quant à $\mathbb{P}\{A_2\}$, la probabilité que le premier dé soit pair, on a $\mathbb{P}\{A_2\} = 1/2$, évidemment.

Finalement,

$$\mathbb{P}\{A_1 | A_2\} = \frac{1/12}{1/2} = \frac{1}{6}.$$

Remarquez que, sachant que le premier dé est pair, les sommes de 7 et 8 sont équiprobables.

2. L'hypothèse que $\mathbb{P}\{B\} > 0$ est cruciale pour cette définition ; néanmoins il pourrait arriver que l'on veuille conditionner par un événement qui ne se réalise presque jamais. Ces cas sont hautement plus délicats, et sont traités différemment selon les spécificités du problème. Souvent on se débrouille en prenant des limites, de différentes façons.

- (b) Si B_1 est «le second dé est 3» et B_2 est «le premier dé est 5», alors, $\mathbb{P}\{B_1 \cap B_2\} = 1/36$, et bien sûr $\mathbb{P}\{B_2\} = 1/6$; on a donc

$$\mathbb{P}\{B_1 \mid B_2\} = \frac{1}{6},$$

ce qui est en fait la même chose que $\mathbb{P}\{B_1\}$ sans le conditionnement. On verra plus loin que ça n'est pas anodin.

- (c) Si C_1 est «la somme est ≥ 5 » et C_2 est «le premier dé est strictement supérieur au second dé», alors, $\mathbb{P}\{C_2\} = 15/36$, et $\mathbb{P}\{C_1 \cap C_2\} = 13/36$, puisque les seules deux façons d'obtenir une somme inférieure à 5 avec deux dés dont le premier est strictement plus grand que le second sont $(2, 1)$ et $(3, 1)$. Donc, finalement,

$$\mathbb{P}\{C_1 \mid C_2\} = 13/15.$$

De façon générale, pour les espaces de probabilités avec des ensembles fondamentaux finis et une mesure de probabilités équiprobable, les probabilités conditionnelles sont (intuitivement) données par ce résultat :

PROPOSITION 3.1. *Soit Ω un ensemble fondamental fini, $\mathcal{E} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure de probabilité équiprobable sur Ω . Soient $E, F \in \mathcal{E}$ deux événements de Ω . Alors, on a que*

$$\mathbb{P}\{E \mid F\} = \frac{|E \cap F|}{|F|}.$$

DÉMONSTRATION. Exercice 3.15. □

3.2. La formule de probabilité totale

Si on multiplie l'équation (3.1.1) par $\mathbb{P}\{B\}$, on trouve que

$$(3.2.1) \quad \mathbb{P}\{A \cap B\} = \mathbb{P}\{A \mid B\} \mathbb{P}\{B\}.$$

On va utiliser ce fait à notre avantage, avec le résultat suivant :

PROPOSITION 3.2 (Formule des probabilités totales). *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. Soit $A \in \mathcal{E}$ un événement, et soit $(B_n)_{n \in I}$ une partition^{3, 4} avec $I \subseteq N$ un ensemble d'indices possiblement dénombrable) de Ω .*

Alors, on a

$$(3.2.2) \quad \mathbb{P}\{A\} = \sum_{i \in I} \mathbb{P}\{A \mid B_i\} \mathbb{P}\{B_i\}.$$

DÉMONSTRATION. On commence par remarquer que

- i. $A = \bigcup_{i \in I} (A \cap B_i)$;
- ii. $(A \cap B_i) \cap (A \cap B_j) = \emptyset$, puisque les B_i sont mutuellement disjoints.

3. *Rappel* : une partition $(B_n)_{n \in I}$ de Ω est une famille d'ensemble deux à deux disjoints dont la réunion satisfait $\bigcup_{i \in I} B_i = \Omega$. Voir la définition A.2 de l'annexe A.

4. On peut montrer (exercice 3.17) que l'équation (3.2.2) tient aussi lorsque les $(B_i)_{i \in I}$ forment une «presque-partition» – c'est à dire qu'ils sont deux-à-deux presque-disjoints et que leur réunion recouvre presque Ω (c'est à dire à l'exception d'un événement de probabilité 0)

Par l'axiome ii de la définition 2.5 (mesure de probabilités), bien sûr, puisque la mesure de probabilités est additive pour les réunions d'événements disjoints deux-à-deux,

$$\mathbb{P}\{A\} = \sum_{i \in I} \mathbb{P}\{A \cap B_i\}.$$

On complète en utilisant l'équation (3.2.1) :

$$\mathbb{P}\{A\} = \sum_{i \in I} \mathbb{P}\{A | B_i\} \mathbb{P}\{B_i\}.$$

□

COROLLAIRE 3.1. *Si $A, B \in \mathcal{E}$ sont deux événements, et que $\mathbb{P}\{B\} = 1$, alors,*

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

DÉMONSTRATION. Exercice 3.16. □

On enchaîne immédiatement avec un exemple très fameux de la formule de probabilité totale, qui est aussi un résultat surprenant et contre-intuitif :

EXEMPLE 3.3 (Le problème de Monty Hall). *Let's Make a Deal* («Faisons un marché!», Tr. libre) était un célèbre jeu télévisé, débutant dans les années '60, et présentée par l'animateur Monty Hall. Il était basé sur un concept assez simple : Monty proposait à des participant.e.s du public (qu'il choisissait au hasard), des échanges de nature plus ou moins risquées. Typiquement, Monty offrait d'abord une somme d'argent à une participante, puis il lui proposait d'échanger cette somme pour le contenu d'une vaste boîte mystérieuse, ou le prix fabuleux caché derrière un rideau... qui n'était pas toujours si fabuleux que ça : ça pouvait être une voiture neuve, aussi bien qu'une tonne de choux – un *zonk*, comme on les appelait dans l'émission.

C'est cette émission qui a démarré tout un débat, d'abord entre statisticien.ne.s dans les années '70, puis entre membres du public, en 1990, lorsqu'un magazine a publié une version du problème suivant :

- Monty donne à la personne un choix de trois portes numérotées ;
- la personne désigne l'une des trois portes.
- Monty révèle alors le prix caché derrière une autre porte. C'est toujours un *zonk* – et si il y en a deux, Monty choisit la porte à ouvrir de façon aléatoire.
- Monty offre ensuite à la personne de changer son choix ;
- le/la participant.e repart avec le prix derrière la porte qu'il/elle avait finalement choisie.

La question que plusieurs posaient à ce moment, c'était : existe-t-il une stratégie qui permet d'avoir plus de 50% de chances de remporter le prix fabuleux ?

SOLUTION. On va se doter des deux événements suivants :

- B : on choisit la bonne porte dès le départ.
- G : on gagne le prix en changeant de porte.

Bien sûr, on a $\mathbb{P}\{B\} = 1/3 = 1 - \mathbb{P}\{B^c\}$.

Par la formule des probabilités totales, la probabilité de G (on gagne le prix en changeant de porte) :

$$\mathbb{P}\{G\} = \mathbb{P}\{G | B\} \mathbb{P}\{B\} + \mathbb{P}\{G | B^c\} \mathbb{P}\{B^c\}.$$

Occupons nous d'abord du premier terme. $\mathbb{P}\{G | B\}$ est la probabilité qu'on gagne le prix en changeant de porte sachant qu'on avait choisi la bonne porte au départ. C'est donc 0,

naturellement, puisque $G \cap B = \emptyset$ – en effet, il est impossible de gagner le prix en changeant de porte ET d'avoir choisi la bonne porte au départ. Donc le premier terme est nul.

De l'autre côté, $\mathbb{P}\{G | B^c\}$ est la probabilité qu'on gagne le prix en changeant de porte sachant qu'on avait choisi la mauvaise porte au départ. Cette probabilité est égale à 1, puisque si on avait choisi la mauvaise porte au départ (c'est à dire un *zonk*), et que Monty nous a montré l'autre *zonk*, en changeant notre choix, on trouve nécessairement la porte qui cache le prix fabuleux.

Finalement, on a donc bien

$$\mathbb{P}\{G\} = \mathbb{P}\{B^c\} = \frac{2}{3}.$$

On conclue qu'il existe bel et bien une stratégie qui augmente nos chances de remporter le prix fabuleux.⁵

On obtient aussi une autre formule pratique pour calculer les probabilités d'intersections :

PROPOSITION 3.3. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et soit $(E_i \in \mathcal{E})_{i \leq n}$ une famille finie de n événements.*

Alors

$$(3.2.3) \quad \begin{aligned} \mathbb{P}\left\{\bigcap_{i=1}^n E_i\right\} &= \mathbb{P}\{E_1\} \mathbb{P}\{E_2 | E_1\} \mathbb{P}\{E_3 | E_2 \cap E_1\} \cdots \mathbb{P}\left\{E_n \middle| \bigcap_{i=1}^{n-1} E_i\right\} \\ &= \prod_{i=1}^n \mathbb{P}\left\{E_i \middle| \bigcap_{j=1}^{i-1} E_j\right\}. \end{aligned}$$

(On considère que $\mathbb{P}\{E_1 | E_0\} = \mathbb{P}\{E_1\}$ car E_0 n'est pas défini.)

DÉMONSTRATION. On fait la preuve par induction : Clairement, pour $n = 1$, c'est trivial. Pour $n = 2$, les équations (3.2.3) et (3.2.1) sont identiques.

Supposons que c'est vrai pour n . Alors,

$$\begin{aligned} \mathbb{P}\left\{\bigcap_{i=1}^{n+1} E_i\right\} &= \mathbb{P}\left\{E_{n+1} \cap \bigcap_{i=1}^n E_i\right\} \\ &= \mathbb{P}\left\{E_{n+1} \middle| \bigcap_{i=1}^n E_i\right\} \mathbb{P}\left\{\bigcap_{i=1}^n E_i\right\} \\ &= \mathbb{P}\left\{E_{n+1} \middle| \bigcap_{i=1}^n E_i\right\} \cdot \prod_{i=1}^n \mathbb{P}\left\{E_i \middle| \bigcap_{j=1}^{i-1} E_j\right\} \\ &= \prod_{i=1}^{n+1} \mathbb{P}\left\{E_i \middle| \bigcap_{j=1}^{i-1} E_j\right\} \end{aligned}$$

5. Dans *Behind Monty Hall's doors : Puzzle, Debate and Answer*, un article par John Tierney paru dans le New-York Times le 21 juillet 1991, Monty Hall explique au journaliste la raison pour laquelle le jeu fonctionne quand même : «Si le présentateur doit absolument ouvrir chaque fois une porte, et vous offrir de changer, vous devriez changer. Mais s'il a le choix de vous offrir de changer ou pas, gare à vous!» (Tr. libre)

Il faut également noter que Monty Hall n'a jamais réellement offert aux participant.e.s de changer de porte. Le problème a simplement été inspiré par l'émission.

et on a que c'est vrai pour $n + 1$. Donc, comme c'est vrai pour $n = 2$, c'est vrai pour tout $n \geq 2$. \square

EXEMPLE 3.4. On sépare un paquet de 52 cartes à jouer en 4 piles de 13 cartes (par exemple pour une partie de Bridge). Quelle est la probabilité que chaque pile reçoive un as ?

SOLUTION. On peut faire un argument combinatoire un peu compliqué, dans lequel on remarque qu'il y a $\binom{52}{13,13,13,13}$ façons de répartir les cartes en 4 piles, mais il faut ensuite diviser par $4!$ parce que l'ordre des piles n'importe pas. Puis, on dirait ensuite qu'il y a $\binom{48}{12,12,12,12}$ façons de répartir les autres cartes en 4 piles si on a mis un as dans chaque pile. La réponse serait la bonne : environ 0,105....

Mais le calcul est un peu cauchemardesque.

On va plutôt se douter des événements suivants :

- A_1 : « l'as de pique est dans l'une des quatre piles.»
- A_2 : « l'as de cœur et l'as de pique sont dans des piles différentes.»
- A_3 : « les as de pique, cœur et carreau sont dans trois piles différentes.»
- A_4 : «les quatre as sont dans des piles différentes.»

Bien entendu, on cherche $\mathbb{P}\{A_4\}$. On remarque cependant que $A_1 \supseteq A_2 \supseteq A_3 \supseteq A_4$, soit que

$$A_2 = A_1 \cap A_2, \quad A_3 = A_1 \cap A_2 \cap A_3, \quad A_4 = A_1 \cap A_2 \cap A_3 \cap A_4.$$

On utilise alors la formule des produits (proposition 3.3), et on obtient

$$\begin{aligned} \mathbb{P}\{A_4\} &= \mathbb{P}\{A_1 \cap A_2 \cap A_3 \cap A_4\} \\ &= \mathbb{P}\{A_4 | A_3\} \mathbb{P}\{A_3 | A_2\} \mathbb{P}\{A_2 | A_1\} \mathbb{P}\{A_1\}. \end{aligned}$$

Bien sûr, $\mathbb{P}\{A_1\} = 1$, puisque l'as de pique doit être quelque part.

$\mathbb{P}\{A_2 | A_1\} = \mathbb{P}\{A_2\}$ est la probabilité que l'as de pique et l'as de cœur ne soient pas dans la même pile. La probabilité que l'as de cœur ne soit pas dans la même pile que l'as de pique est simplement de $39/51$, car parmi les 51 «positions restantes» pour les 51 cartes restantes, il y en a 39 qui ne sont pas dans la pile de l'as de pique.

Donc finalement, $\mathbb{P}\{A_2 | A_1\} = 39/52$.

$\mathbb{P}\{A_3 | A_2\}$ est la probabilité que l'as de carreau ne soit pas dans les piles de l'as de cœur et l'as de pique, sachant que ces deux derniers sont dans des piles séparées. Par un raisonnement similaire, à celui pour $\mathbb{P}\{A_2 | A_1\}$, on a forcément que $\mathbb{P}\{A_3 | A_2\} = 26/50$.

De la même façon, $\mathbb{P}\{A_4 | A_3\} = 13/49$.

Au final,

$$\mathbb{P}\{A_4\} = \frac{39 \times 26 \times 13}{51 \times 50 \times 49},$$

ce qui nous donne exactement la même valeur.

REMARQUE. Ce raisonnement peut sembler plus complexe. Toutefois, les calculs sont beaucoup plus simples, et le raisonnement se transpose très facilement à plus grande échelle.

Le résultat de cet exemple nous met la puce à l'oreille pour ce corollaire :

COROLLAIRE 3.2. Si $(E_i)_{i \leq \mathbb{N}}$ sont une famille monotone décroissante d'événements, alors

$$(3.2.4) \quad \mathbb{P}\{E_n\} = \prod_{i=1}^n \mathbb{P}\{E_i | E_{i-1}\}.$$

(On considère que $\mathbb{P}\{E_1 | E_0\} = \mathbb{P}\{E_1\}$ car E_0 n'est pas défini.)

DÉMONSTRATION. Exercice 3.18. □

3.3. La formule de Bayes

L'une des principales utilités des probabilités conditionnelles vient du fait que, comme mentionné précédemment, elles sont une façon formelle de représenter comment de nouvelles informations changent les probabilités d'une expérience aléatoire.

Le problème-type est le suivant : lorsqu'on mène une expérience, on émet une hypothèse. Puis, on fait une observation. L'objectif, c'est donc d'évaluer comment l'observation a affecté la probabilité que l'hypothèse soit vraie.

Si H est l'événement « l'hypothèse est vraie », et O est l'observation, alors, le terme qu'on cherche, c'est

$$\mathbb{P}\{H \mid O\}.$$

Malheureusement, ce n'est souvent pas le plus facile à calculer ; d'autres choses sont plus simples. Par exemple :

- $\mathbb{P}\{H\}$, la probabilité que l'hypothèse soit vraie *a priori*, sans qu'on n'ait fait d'observations ;
- $\mathbb{P}\{O \mid H\}$, la probabilité qu'on fasse l'observation si l'hypothèse est vraie. Les statisticien.ne.s appellent cette valeur «la vraisemblance de l'observation O ».
- $\mathbb{P}\{O \mid H^c\}$, la probabilité qu'on fasse l'observation si l'hypothèse n'est pas vraie.

Ces valeurs sont très souvent plus faciles à calculer. L'objectif, ce sera donc d'obtenir une formule qui permettrait d'exprimer notre probabilité conditionnelle $\mathbb{P}\{H \mid O\}$.

PROPOSITION 3.4 (Formule de Bayes, version 1). *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et soient $A, B \in \mathcal{E}$ deux événements.*

Alors, on a

$$(3.3.1) \quad \mathbb{P}\{B \mid A\} = \frac{\mathbb{P}\{A \mid B\} \mathbb{P}\{B\}}{\mathbb{P}\{A \mid B\} \mathbb{P}\{B\} + \mathbb{P}\{A \mid B^c\} \mathbb{P}\{B^c\}}$$

DÉMONSTRATION. On utilise simplement la définition des probabilités conditionnelles, puis l'équation (3.2.1) pour le numérateur, et la formule des probabilités totales (équation (3.2.2)) pour le dénominateur, en remarquant que B et B^c forment une partition de Ω :

$$\begin{aligned} \mathbb{P}\{B \mid A\} &= \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{A\}} \\ &= \frac{\mathbb{P}\{A \mid B\} \mathbb{P}\{B\}}{\mathbb{P}\{A \mid B\} \mathbb{P}\{B\} + \mathbb{P}\{A \mid B^c\} \mathbb{P}\{B^c\}} \end{aligned}$$

□

Cette formule n'est rien de très nouveau en elle-même ; il s'agit simplement d'appliquer les définitions, comme on le voit dans la preuve.

Toutefois, c'est intuitivement qu'elle fait une différence. Voyons un exemple :

EXEMPLE 3.5. Un test de dépistage pour le cancer de l'œil gauche⁶ est réputé très efficace : si un.e patient.e a bel et bien le cancer de l'œil gauche, le test sera positif avec une probabilité de 99%. Et si un.e patient.e n'a pas le cancer de l'œil gauche, le test sera négatif avec une probabilité de 99% également.

6. Je ne connais rien en médecine, je ferai pas trop semblant.

Des études ont montré que la fréquence d'occurrence du cancer de l'œil gauche dans la population est de 0,1%. On sélectionne un individu au hasard et on lui fait passer le test. Si le test est positif, quelle est la probabilité que l'individu souffre bel et bien du cancer de l'œil gauche ?

SOLUTION. Ici, notre hypothèse H , c'est «l'individu a le cancer de l'œil gauche». Notre observation O , c'est «Le test est positif».

Selon l'énoncé, on a que $\mathbb{P}\{H\} = 1/1000 - a priori$, seulement un.e individu sur 1000 a le cancer de l'œil gauche.

Mais maintenant, on sait aussi que $\mathbb{P}\{O | H\}$, la probabilité que le test soit positif si l'individu a le cancer de l'œil gauche, est de 99/100. Et on sait que $\mathbb{P}\{O | H^c\}$, la probabilité que le test est positif si l'individu n'a pas le cancer de l'œil gauche, est de 1/100.

Si on applique simplement la formule de Bayes, on trouve donc

$$\begin{aligned}\mathbb{P}\{H | O\} &= \frac{\mathbb{P}\{O | H\} \mathbb{P}\{H\}}{\mathbb{P}\{O | H\} \mathbb{P}\{H\} + \mathbb{P}\{O | H^c\} (1 - \mathbb{P}\{H\})} \\ &= \frac{(99/100) \times (1/1000)}{(99/100) \times (1/1000) + (1/100) \times (999/1000)} \\ &= \frac{11}{122} \approx 0,09.\end{aligned}$$

Autrement dit, si on sait que le résultat du test est positif, la probabilité que la personne testée soit réellement atteinte du cancer de l'œil gauche est d'à peine 9% !

Ce résultat est profondément contre-intuitif pour la très grande majorité des gens. Pour arriver à s'y retrouver un peu, il faut réaliser ce qu'on vient de calculer. Le test produit de faux résultats seulement 1% du temps. Mais comme il y a énormément plus de personnes non-atteintes par la maladie que de personnes qui en souffrent vraiment, le groupe des faux positifs est proportionnellement beaucoup plus grand que le groupe des vrais positifs. Ce serait même vrai si le test identifiait de manière fiable tous les cas de la maladie ($\mathbb{P}\{O | H\} = 1$).

Dans ce cas, on avait fait une hypothèse binaire : «le patient a le cancer» ou «le patient n'a pas le cancer».

Cependant, on remarque qu'il existe une généralisation de la formule de Bayes :

PROPOSITION 3.5 (Formule de Bayes, version 2). *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. Soit A un événement, et soit $(B_i)_{i \in I}$ une partition⁷ de Ω (avec $I \subseteq \mathbb{N}$ un ensemble d'indices potentiellement dénombrable).*

Alors, pour tout $i \in I$,

$$(3.3.2) \quad \mathbb{P}\{B_i | A\} = \frac{\mathbb{P}\{A | B_i\} \mathbb{P}\{B_i\}}{\sum_{j \in I} \mathbb{P}\{A | B_j\} \mathbb{P}\{B_j\}}.$$

DÉMONSTRATION. Exactement le même raisonnement que pour la proposition 3.4. \square

Ici, notre A serait notre observation, et nos B_i sont nos hypothèses, mutuellement exclusives, et couvrant entre-elles toutes les possibilités.

EXEMPLE 3.6. À une fête foraine, un forain répartit trois pièces d'argent et trois pièces d'or dans trois sacs identiques, comme ceci :

— Un sac aura deux pièces d'or ;

7. Ça fonctionne aussi si les B_i sont une presque-partition.

- un sac aura deux pièces d'argent ;
- un sac aura une pièce d'or et une d'argent.

Puis, le forain choisit un sac de façon aléatoire, puis il retire une pièce du sac de façon aléatoire.

- (a) Quelle est la probabilité qu'il retire une pièce d'or ?
- (b) Sachant qu'il a retiré une pièce d'or, quelle est la probabilité que l'autre pièce dans le sac soit aussi une pièce d'or ?

SOLUTION. (a) On va se doter des trois événements suivants, qui forment une partition de notre ensemble fondamental :

- H_0 : «On a choisi le sac à deux pièces d'argent» ;
- H_1 : «On a choisi le sac à une pièce d'or» ;
- H_2 : «On a choisi le sac à deux pièces d'or» ;

Bien sûr, $\mathbb{P}\{H_i\} = 1/3$, puisque les sacs sont choisis équiprobablement.

Et puis on va finalement définir l'événement O : «On pique une pièce d'or.»

Alors, par la formule de probabilités totales,

$$\mathbb{P}\{O\} = \sum_{i=0}^2 \mathbb{P}\{O | H_i\} \mathbb{P}\{H_i\},$$

et on peut énumérer les termes. D'abord, $\mathbb{P}\{O | H_0\} = 0$; il est impossible de tirer une pièce d'or si on sait qu'on a tiré la pièce d'un sac qui en contenait deux d'argent.

$\mathbb{P}\{O | H_2\} = 1$ pour une raison similaire : si on sait qu'il y avait deux pièces d'or dans le sac, on a sûrement tiré une pièce d'or. Finalement, $\mathbb{P}\{O | H_1\} = 1/2$ puisqu'il y a une pièce d'or seulement dans ce sac.

Donc,

$$\mathbb{P}\{O\} = 0 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{1}{2}.$$

- (b) Maintenant on cherche $\mathbb{P}\{H_2 | O\}$, la probabilité d'avoir tiré du sac à deux pièces d'or sachant qu'on a tiré une pièce d'or.

On pourrait naïvement croire que cette probabilité est $1/2$ – après tout, il y a deux sacs. Cependant, il faut se souvenir que c'était plus probable de piger une pièce d'or dans l'un des deux sacs que dans l'autre.

Avec la formule de Bayes, on a :

$$\mathbb{P}\{H_2 | O\} = \frac{\mathbb{P}\{O | H_2\} \mathbb{P}\{H_2\}}{\mathbb{P}\{O\}} = \frac{(1/3)}{(1/2)} = \frac{2}{3}.$$

3.4. Les événements indépendants

Directement en lien avec la notion de probabilité conditionnelle, une notion centrale en théorie des probabilités est la notion d'indépendance.

L'idée, c'est qu'on remarque qu'il y a des événements qui «n'interfèrent pas entre eux», c'est à dire que l'occurrence de l'un n'influe pas sur la probabilité de l'autre, et vice-versa.

DÉFINITION 3.2. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et soient $E, F \in \mathcal{E}$ deux événements.

On dit que E et F sont **indépendants** si ils vérifient :

$$(3.4.1) \quad \mathbb{P}\{E \cap F\} = \mathbb{P}\{E\} \mathbb{P}\{F\}$$

Cette définition peut sembler arbitraire ; malgré la connexion annoncée entre la notion d'indépendance et la notion de probabilité conditionnelle, cette dernière n'intervient pas dans la définition.

Ceci dit, ça n'est pas pour autant que les deux notions ne sont pas intimement liées.

PROPOSITION 3.6. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et soient $E, F \in \mathcal{E}$ des événements avec $0 < \mathbb{P}\{F\} < 1$. Alors, les énoncés suivants sont logiquement équivalents :*

- i. E et F sont indépendants ;
- ii. $\mathbb{P}\{E | F\} = \mathbb{P}\{E\}$;
- iii. $\mathbb{P}\{E | F\} = \mathbb{P}\{E | F^c\}$;
- iv. E et F^c sont indépendants.

DÉMONSTRATION. Nous allons montrer les implications en cycle.

i \Rightarrow ii : Puisque E et F sont indépendants, alors on a que

$$\mathbb{P}\{E | F\} = \frac{\mathbb{P}\{E \cap F\}}{\mathbb{P}\{F\}} = \frac{\mathbb{P}\{E\}\mathbb{P}\{F\}}{\mathbb{P}\{F\}} = \mathbb{P}\{E\}.$$

ii \Rightarrow iii : Par la formule de probabilités totales, puisque $\mathbb{P}\{E\} = \mathbb{P}\{E | F\}$, on a :

$$\begin{aligned} \mathbb{P}\{E | F\} &= \mathbb{P}\{E | F\}\mathbb{P}\{F\} + \mathbb{P}\{E | F^c\}(1 - \mathbb{P}\{F\}) \\ &\Rightarrow \mathbb{P}\{E | F\}(1 - \mathbb{P}\{F\}) = \mathbb{P}\{E | F^c\}(1 - \mathbb{P}\{F\}), \end{aligned}$$

et on conclue en divisant par $(1 - \mathbb{P}\{F\})$ (on peut faire ça parce que $\mathbb{P}\{F\} < 1$).

iii \Rightarrow iv : Puisqu'on a $\mathbb{P}\{E | F\} = \mathbb{P}\{E | F^c\}$, on a nécessairement que

$$\begin{aligned} \mathbb{P}\{E\} &= \mathbb{P}\{E | F\}(1 - \mathbb{P}\{F^c\}) + \mathbb{P}\{E | F^c\}\mathbb{P}\{F^c\} \\ &= \mathbb{P}\{E | F^c\}(1 - \mathbb{P}\{F^c\}) + \mathbb{P}\{E | F^c\}\mathbb{P}\{F^c\} \\ &= \mathbb{P}\{E | F^c\}, \end{aligned}$$

d'où il suit que

$$\mathbb{P}\{E \cap F^c\} = \mathbb{P}\{E | F^c\}\mathbb{P}\{F^c\} = \mathbb{P}\{E\}\mathbb{P}\{F^c\},$$

et E et F^c sont indépendants.

iv \Rightarrow i : On sait que E et F^c sont indépendants. Mais on a

$$\begin{aligned} \mathbb{P}\{E \cap F\} &= \mathbb{P}\{E\} - \mathbb{P}\{E \cap F^c\} \\ &= \mathbb{P}\{E\} - \mathbb{P}\{E\}\mathbb{P}\{F^c\} \\ &= \mathbb{P}\{E\}(1 - \mathbb{P}\{F^c\}) \\ &= \mathbb{P}\{E\}\mathbb{P}\{F\}, \end{aligned}$$

et E et F sont indépendants. □

Comme on le voit, les manipulations requises dans cette preuve nécessitent que $\mathbb{P}\{F\}$ soit dans $(0, 1)$, avec les bornes exclues – c'est principalement pour cela qu'on définit l'indépendance, de façon générale, comme on le fait plus haut : ça évite les embrouilles avec les événements qui auraient probabilité 0.

Ceci dit, on a aussi le résultat suivant :

PROPOSITION 3.7. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et soit $F \in \mathcal{E}$ un événement. Alors, si $\mathbb{P}\{F\} = 0$ ou 1 , F est indépendant de tout $E \in \mathcal{E}$.*

DÉMONSTRATION. Si $\mathbb{P}\{F\} = 1$, alors F est indépendant de tous les événements. En effet, par le corollaire 3.1, pour tout $E \in \mathcal{E}$, on a

$$\mathbb{P}\{E \cap F\} = \mathbb{P}\{E\} = \mathbb{P}\{E\}\mathbb{P}\{F\}.$$

Si $\mathbb{P}\{F\} = 0$, alors on a que $0 \leq \mathbb{P}\{E \cap F\} \leq \mathbb{P}\{F\}$ par la proposition 2.1, vu que $E \cap F \subseteq F$. Mais on a donc que $\mathbb{P}\{E \cap F\} = 0 = \mathbb{P}\{E\}\mathbb{P}\{F\}$. \square

Ensemble, ces deux propositions nous montrent bien que, dans les seuls cas réellement d'intérêt, la notion d'indépendance correspond bien à l'idée intuitive qu'on en a : deux événements sont indépendants si la probabilité de l'un ne change pas lorsqu'on sait si l'autre s'est produit.

EXEMPLE 3.7. On pique une carte d'un paquet ordinaire de 52 cartes. Soient les deux événements suivants : A , «La carte est un as» et B , «la carte est un trèfle».

Les événements A et B sont indépendants, puisque $\mathbb{P}\{A \cap B\}$ soit la probabilité de piger l'as de trèfle est de $1/52$, alors que $\mathbb{P}\{A\}$, la probabilité de piger un as, est de $4/52$, et la probabilité $\mathbb{P}\{B\}$ de piger un trèfle est de $1/4$, et qu'on a bien $(1/4) \times (4/52) = (1/52)$.

EXEMPLE 3.8. Si on fait deux lancers de pile ou face, et que les quatre résultats sont équiprobables. Les événements A , «les deux lancers sont pareils», et B , «la première pièce est tombée sur 'pile'», sont indépendants, parce que la probabilité $\mathbb{P}\{A\}$ que les deux lancers soient pareils est $1/2$, la probabilité $\mathbb{P}\{B\}$ que la première pièce tombe sur «pile» est $1/2$, et la probabilité que les deux lancers soient pareils et que la première pièce tombe sur «pile» est la probabilité d'obtenir pile-pile, soit $1/4$.

EXEMPLE 3.9. Si on lance deux dés à six faces, équilibrés, de telle sorte que les 36 résultats possibles sont équiprobables, alors

- (a) Les événements A_1 , «le premier dé est un 5» et A_2 , «le second dé est un 2», sont ils indépendants ?
- (b) Soit B l'événement : «La somme des deux dés est 7». Les événements A_1 et B sont ils indépendants ? A_2 et B ?
- (c) L'événement B est-il indépendant de $A_1 \cap A_2$?

SOLUTION. (a) On a que $\mathbb{P}\{A_1\} = 1/6 = \mathbb{P}\{A_2\}$, bien sûr, et $\mathbb{P}\{A_1 \cap A_2\} = 1/36$. Donc A_1 et A_2 sont indépendants.

- (b) On a que $\mathbb{P}\{B\} = 1/6$ aussi, et bien sûr,

$$\mathbb{P}\{A_1 \cap B\} = \mathbb{P}\{A_1 \cap A_2\} = \mathbb{P}\{A_2 \cap B\} = 1/36,$$

donc on conclue que B est indépendant de A_1 et B est indépendant de A_2 .

- (c) Par contre, on a que $\mathbb{P}\{B | A_1 \cap A_2\} = 1 \neq \mathbb{P}\{B\}$, donc forcément on ne peut pas avoir que B est indépendant de $A_1 \cap A_2$, par la proposition 3.6.

Ce dernier exemple nous met un peu un bâton dans les roues. D'un côté on voudrait que la notion d'indépendance pour plusieurs événements se généralise naturellement, comme ceci :

$$\mathbb{P}\{E_1 \cap E_2 \cap \dots \cap E_n\} = \prod_{i=1}^n \mathbb{P}\{E_i\}.$$

De l'autre, ça ne semble pas suivre directement de l'indépendance «deux-à-deux», comme on vient de le voir.

DÉFINITION 3.3. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et soit $(E_i \in \mathcal{E})_{i \in I}$ une famille d'événements, avec I un ensemble d'indices, potentiellement indénombrable⁸.

On dit que la famille $(E_i)_{i \in I}$ est indépendante si pour tout $J \subseteq I$, on a que

$$\mathbb{P} \left\{ \bigcap_{j \in J} E_j \right\} = \prod_{j \in J} \mathbb{P} \{E_j\}.$$

Ainsi, par exemple, les événements E, F, G ne sont indépendants que si on a $\mathbb{P} \{E \cap F \cap G\} = \mathbb{P} \{E\} \mathbb{P} \{F\} \mathbb{P} \{G\}$, en plus des indépendances deux-à-deux.

L'intérêt de l'indépendance réside très souvent dans le fait que c'est souvent une hypothèse très naturelle à faire ; on passera donc beaucoup moins de temps à prouver que des événements sont indépendants que simplement à se servir du fait qu'ils sont indépendants.

EXEMPLE 3.10. Si on conduit une longue chaîne d'expériences, identiques et indépendantes, et pour laquelle l'événement E_i , «La i ème expérience réussit» a probabilité $\mathbb{P} \{E_i\} = p$ pour tout i , déterminer

- (a) La probabilité qu'au moins l'une des n premières expériences réussisse.
- (b) La probabilité qu'une tentative réussisse éventuellement.
- (c) La probabilité qu'exactement k des n premières expériences réussissent.
- (d) La probabilité que toutes les tentatives réussissent.

SOLUTION. (a) Ici, on va passer par le calcul du complément. On cherche $\mathbb{P} \{\bigcup_{i=1}^n E_i\} = 1 - \mathbb{P} \{\bigcap_{i=1}^n E_i^c\}$, et bien sûr, puisque les E_i sont indépendants, on a que

$$\mathbb{P} \left\{ \bigcap_{i=1}^n E_i^c \right\} = \prod_{i=1}^n \mathbb{P} \{E_i^c\} = (1-p)^n.$$

Donc, la probabilité qu'au moins 1 expérience réussisse est

$$\mathbb{P} \left\{ \bigcup_{i=1}^n E_i \right\} = 1 - (1-p)^n.$$

- (b) La probabilité qu'une tentative réussisse éventuellement est

$$\mathbb{P} \left\{ \bigcup_{i \in \mathbb{N}} E_i \right\} = \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{i=1}^n E_i \right\},$$

par la continuité de la mesure de probabilités.

Donc : si $p > 0$, la probabilité qu'une tentative réussisse éventuellement est 1. Si $p = 0$, la probabilité qu'une tentative réussisse éventuellement est 0.

- (c) La probabilité qu'exactement k des n premières tentatives réussissent, c'est un peu plus sportif.

Commençons par choisir un certain $I \subseteq \{1, \dots, n\}$, un ensemble d'indices parmi les n premiers, et avec $|I| = k$.

8. Oui oui, pour vrai !

On considère l'événement $(\bigcap_{i \in I} E_i) \cap \left(\bigcap_{n \geq i \notin I} E_i^c\right)$, que tous les événements dont l'indice est dans I sont réalisés, et que tous les autres ne sont pas réalisés (parmi les n premiers). Bien sûr, par la propriété d'indépendance, on a que

$$\mathbb{P}\left\{\left(\bigcap_{i \in I} E_i\right) \cap \left(\bigcap_{n \geq i \notin I} E_i^c\right)\right\} = p^k(1-p)^{n-k}.$$

puisque il y a $n - k$ indices non-inclus dans I et plus petits que n .

Tout ce qu'il reste à faire, c'est réaliser que pour deux I différents, ces événements sont disjoints – en effet, si $i \in I \Delta I'$, alors on pourrait avoir que

$$\left(\bigcap_{i \in I} E_i\right) \cap \left(\bigcap_{n \geq i \notin I} E_i^c\right) \subseteq E_i, \quad \left(\bigcap_{i \in I'} E_i\right) \cap \left(\bigcap_{n \geq i \notin I'} E_i^c\right) \subseteq E_i^c,$$

ce qui prouve que ces événements sont disjoints.

Finalement, si F_k est l'événement «exactement k tentatives réussissent», on a que

$$F_k = \bigsqcup_{I:|I|=k} \left(\bigcap_{i \in I} E_i\right) \cap \left(\bigcap_{n \geq i \notin I} E_i^c\right),$$

et

$$\mathbb{P}\{F_k\} = \sum_{I:|I|=k} \mathbb{P}\left\{\left(\bigcap_{i \in I} E_i\right) \cap \left(\bigcap_{n \geq i \notin I} E_i^c\right)\right\}.$$

On complète en remarquant que, comme tous ces termes sont égaux à $p^k(1-p)^{n-k}$, il suffit de compter combien il y en a – c'est à dire combien il y a de façons de choisir k indices parmi n , soit $\binom{n}{k}$.

On a donc enfin de compte

$$\mathbb{P}\{F_k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

(d) La probabilité que les n premières tentatives réussissent est

$$\mathbb{P}\left\{\bigcap_{i=1}^n E_i\right\} = \prod_{i=1}^n \mathbb{P}\{E_i\} = p^n.$$

Bien sûr, l'événement «toutes les tentatives réussissent» est donné par

$$\bigcap_{i \in \mathbb{N}} E_i = \lim_{n \rightarrow \infty} \bigcap_{i=1}^n E_i,$$

et sa probabilité est donnée par

$$\mathbb{P}\left\{\bigcap_{i \in \mathbb{N}} E_i\right\} = \lim_{n \rightarrow \infty} \mathbb{P}\left\{\bigcap_{i=1}^n E_i\right\},$$

c'est à dire 0 si $p < 1$, et 1 si $p = 1$.

3.5. La probabilité conditionnelle comme nouvelle mesure

Nous terminons ce chapitre en faisant une observation astucieuse : soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. On pourrait, pour un certain $B \in \mathcal{E}$, définir $Q_B : \mathcal{E} \rightarrow \mathbb{R}$, par

$$Q_B \{A\} = \mathbb{P} \{A | B\}.$$

Et si on faisait ça, quelles seraient les propriétés de Q_B ? La réponse : **Q_B est une autre mesure de probabilités pour Ω, \mathcal{E} !**

PROPOSITION 3.8. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et soit $B \in \mathcal{E}$ un événement fixé. On définit $Q_B : \mathcal{E} \rightarrow \mathbb{R}$ par :*

$$(3.5.1) \quad Q_B \{A\} = \mathbb{P} \{A | B\},$$

pour tout $A \in \mathcal{E}$.

Alors, Q_B est une autre mesure de probabilités pour l'espace Ω avec les événements \mathcal{E} .

DÉMONSTRATION. Pour faire cette preuve, il suffit de vérifier les axiomes de la définition 2.5.

Axiome i : positivité. Cet axiome est facile à vérifier. Clairement, pour tout $A \in \mathcal{E}$, $Q_B \{A\} = \mathbb{P} \{A | B\} \geq 0$.

Axiome iii : normalisation. Cet axiome aussi est plus ou moins trivial :

$$Q_B \{\Omega\} = \mathbb{P} \{\Omega \cap B\} / \mathbb{P} \{B\} = 1,$$

puisque $\Omega \cap B = B$.

Axiome ii : additivité. Supposons $(A_i \in \mathcal{E})_{i \in \mathbb{N}}$ une famille d'événements deux-à-deux disjoints.

Alors, en utilisant la distributivité de l'intersection sur la réunion, et l'additivité pour \mathbb{P} , on trouve :

$$\begin{aligned} Q_B \left\{ \bigsqcup_{i \in \mathbb{N}} A_i \right\} &= \frac{\mathbb{P} \{(\bigsqcup_{i \in \mathbb{N}} A_i) \cap B\}}{\mathbb{P} \{B\}} \\ &= \frac{\mathbb{P} \{\bigsqcup_{i \in \mathbb{N}} (A_i \cap B)\}}{\mathbb{P} \{B\}} \\ &= \frac{\sum_{i \in \mathbb{N}} \mathbb{P} \{A_i \cap B\}}{\mathbb{P} \{B\}} \\ &= \sum_{i \in \mathbb{N}} Q_B \{A_i\} \end{aligned}$$

□

L'intérêt de le voir comme ça, c'est que Q_B hérite de toutes les propriétés qu'on a vu pour les mesures de probabilités. En vrac, on a donc les résultats suivants :

- $\mathbb{P} \{A^c | B\} = 1 - \mathbb{P} \{A | B\}$ (proposition 2.1);
- Si $E \subseteq F$, alors $\mathbb{P} \{E | B\} \leq \mathbb{P} \{F | B\}$ (même proposition);
- etc.

Toutes ces propriétés n'ont pas besoin d'être redémontrées pour les probabilités conditionnelles – on vient tout juste de montrer que notre Q_B est une mesure de probabilités, et ces résultats suivent immédiatement.

REMARQUE. La notation Q_B n'est pas «standard» – il s'agit simplement d'un choix arbitraire, mais utile pour bien illustrer le fait que les probabilités conditionnelles pour un événement fixé sont de nouvelles mesures de probabilités.

Mais, que se passerait-il si on conditionnait... avec Q_B plutôt que \mathbb{P} ? Puisque Q_B est une mesure de probabilités, qu'est-ce qui arriverait si on considérait par exemple l'expression

$$Q_B \{A | C\},$$

pour deux événements $A, C \in \mathcal{E}$?

LEMME 3.1. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soit $Q_B : \mathcal{E} \rightarrow \mathbb{R}$ définie par $Q_B \{A\} = \mathbb{P} \{A | B\}$ pour tout $A \in \mathcal{E}$, et pour tout $B \in \mathcal{E}$.

Alors,

$$Q_B \{A | C\} = \mathbb{P} \{A | B \cap C\} = Q_{B \cap C} \{A\}.$$

DÉMONSTRATION.

$$\begin{aligned} Q_B \{A | C\} &= \frac{\mathbb{P} \{A \cap C | B\}}{\mathbb{P} \{C | B\}} \\ &= \frac{\mathbb{P} \{A \cap B \cap C\} / \mathbb{P} \{B\}}{\mathbb{P} \{B \cap C\} / \mathbb{P} \{B\}} \\ &= \frac{\mathbb{P} \{A \cap B \cap C\}}{\mathbb{P} \{B \cap C\}} \\ &= \mathbb{P} \{A | B \cap C\} \\ &= Q_{B \cap C} \{A\} \end{aligned}$$

□

Autrement dit, conditionner successivement par B , puis par C , revient exactement au même que de conditionner simultanément par $B \cap C$, soit par B «et» C en même temps.

Ce lemme est très important pour nous, puisqu'il permet de montrer une nouvelle version de la formule de probabilité totale, pour les probabilités conditionnelles :

PROPOSITION 3.9. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soient $A, B \in \mathcal{E}$ deux événements, et soient $(C_i)_{i \in I}$ une partition de Ω , avec $I \subseteq \mathbb{N}$ un ensemble d'indices potentiellement dénombrable.

Alors, on a

$$(3.5.2) \quad \mathbb{P} \{A | B\} = \sum_{i \in I} \mathbb{P} \{A | B \cap C_i\} \mathbb{P} \{C_i | B\}.$$

DÉMONSTRATION. Soit Q_B définie comme dans le lemme 3.1 ; alors, Q_B est une mesure de probabilités (par la proposition 3.8), et on doit donc avoir la formule de probabilités totales pour Q_B :

$$Q_B \{A\} = \sum_{i \in I} Q_B \{A | C_i\}.$$

On conclue en appliquant la définition de Q_B et le lemme 3.1. □

Finalement, on a également la notion d'indépendance conditionnelle :

DÉFINITION 3.4 (Indépendance conditionnelle). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et soient A, B, C trois événements avec $\mathbb{P}\{C\} > 0$. On dit que A et B sont **indépendants conditionnellement à C** si

$$\mathbb{P}\{A \cap B \mid C\} = \mathbb{P}\{A \mid C\} \mathbb{P}\{B \mid C\}.$$

Il arrive souvent que la connaissance d'informations rendent indépendants des événements qui ne l'étaient pas précédemment.

Voici un exemple :

EXEMPLE 3.11. À la kermesse du village, un forain vous présente une boîte contenant deux dés à six faces. Le premier dé a quatre faces rouges et deux faces noires. L'autre dé a deux faces rouges et quatre noires. Il choisit au hasard l'un des deux dés, le lance, puis annonce la couleur de la face du dessus. Puis, il le relance. Quelle est la probabilité que le dé montre une face rouge si c'était rouge la première fois ? Les premier et second lancers de dé sont ils indépendants ?

SOLUTION. Soit R_1 l'événement «le premier lancer est rouge», et R_2 , «le second lancer est rouge».

Alors, on cherche $\mathbb{P}\{R_2 \mid R_1\}$. Soit D_1 l'événement «le forain choisit le premier dé», et D_2 , «le forain choisit le second dé».

Par la formule de probabilité totale, on a

$$\mathbb{P}\{R_2 \mid R_1\} = \mathbb{P}\{R_2 \mid R_1 \cap D_1\} \mathbb{P}\{D_1 \mid R_1\} + \mathbb{P}\{R_2 \mid R_1 \cap D_2\} \mathbb{P}\{D_2 \mid R_1\}.$$

L'intérêt, c'est que, conditionnellement à D_1 , les événements R_1 et R_2 sont indépendants : si on sait quel dé on a choisi, les lancers ne dépendent plus les uns des autres. Donc,

$$\mathbb{P}\{R_2 \mid R_1 \cap D_1\} = \mathbb{P}\{R_2 \mid D_1\} = \frac{2}{3},$$

puisque si on a pris le premier dé, évidemment la probabilité de tomber sur une face rouge est de $2/3$. De façon similaire,

$$\mathbb{P}\{R_2 \mid R_1 \cap D_2\} = \mathbb{P}\{R_2 \mid D_2\} = \frac{1}{3}.$$

Il reste maintenant à employer la formule de Bayes pour déterminer $\mathbb{P}\{D_1 \mid R_1\}$.

On a

$$\mathbb{P}\{D_1 \mid R_1\} = \frac{\mathbb{P}\{R_1 \mid D_1\} \mathbb{P}\{D_1\}}{\mathbb{P}\{R_1 \mid D_1\} \mathbb{P}\{D_1\} + \mathbb{P}\{R_1 \mid D_2\} \mathbb{P}\{D_2\}}.$$

Bien sûr, $\mathbb{P}\{R_1 \mid D_1\} = 2/3$ et $\mathbb{P}\{R_1 \mid D_2\} = 1/3$. Bien sûr, puisque les dés sont choisis de façon équiprobable, on a aussi $\mathbb{P}\{D_1\} = \mathbb{P}\{D_2\} = 1/2$, et par conséquent,

$$\mathbb{P}\{D_1 \mid R_1\} = \frac{(2/3) \times (1/2)}{(2/3) \times (1/2) + (1/3) \times (1/2)} = 2/3.$$

Évidemment, puisque D_2 est complémentaire à D_1 , on a $\mathbb{P}\{D_2 \mid R_1\} = 1/3$. Finalement,

$$\mathbb{P}\{R_2 \mid R_1\} = (2/3) \times (2/3) + (1/3) \times (1/3) = 5/9.$$

Par contraste, on remarque que

$$\begin{aligned}\mathbb{P}\{R_2\} &= \mathbb{P}\{R_2 \mid D_1\} \mathbb{P}\{D_1\} + \mathbb{P}\{R_2 \mid D_2\} \mathbb{P}\{D_2\} \\ &= \frac{2}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \\ &= \frac{1}{2},\end{aligned}$$

ce qui est inférieur à $5/9$ – donc les événements R_1 et R_2 ne sont pas indépendants ; l'événement R_1 donne plus d'informations sur l'événement R_2 .

Intuitivement, ça se comprend : si on sait qu'on est tombés sur une face rouge, ça augmente la vraisemblance de l'hypothèse qu'on a choisi le dé qui a plus de faces rouges – et donc ça augmente la probabilité que le second lancer sera aussi rouge.

3.6. Exercices

EXERCICE 3.1. Si on roule deux dés à six faces, quelle est

- (a) la probabilité que le premier dé tombe sur 6 sachant que la somme des deux dés est de n (calculer pour toutes les valeurs de n de 2 à 12) ?
- (b) la probabilité qu'au moins l'un des deux dés tombe sur 6 sachant que la somme des deux dés est de n (calculer pour toutes les valeurs de n de 2 à 12) ?

EXERCICE 3.2. Une urne contient 12 boules dont 8 blanches. On tire 4 boules. Quelle est la probabilité que les première et troisième boules soient blanches sachant que trois des quatre boules sont blanches si :

- (a) On fait le tirage avec remise.
- (b) On fait le tirage sans remise.
- (c) Y a-t-il une différence ? Pourquoi, ou pourquoi pas ?

EXERCICE 3.3. Une étudiante récemment graduée veut prendre les trois examens d'une société d'actuaires. Elle doit réussir les trois afin d'être reçue ; un seul échec et c'est fini. La probabilité qu'elle réussisse le premier examen est de 9/10 ; sachant qu'elle le réussit, la probabilité qu'elle réussisse le second est de 4/5. La probabilité qu'elle réussisse le troisième test sachant qu'elle a réussi les deux premiers est de 7/10.

- (a) Quelle est la probabilité qu'elle réussisse tous les examens ?
- (b) Sachant qu'elle n'a pas réussi, quelle est la probabilité qu'elle ait échoué au second ?

EXERCICE 3.4. 98% des bébés survivent à l'accouchement.⁹ Cependant, 15% des naissances nécessitent une intervention par césarienne et dans ces cas-là, 96% des bébés survivent. Supposons qu'une femme (choisie aléatoirement dans la population) accouche sans subir d'intervention par césarienne.

- (a) Intuitivement (c.à.d. sans calculs), est-ce que cela augmente les chances de survie de son bébé ?
- (b) Quelle est la probabilité que le bébé survive sachant que la mère n'a pas subi d'intervention par césarienne ? (OK, maintenant faites les calculs !)

EXERCICE 3.5. Un dé jaune, un dé rouge et un dé bleu (tous à six faces) sont lancés. Si B dénote le résultat du dé bleu, J celui du dé jaune, et R celui du dé rouge, on s'intéresse à la probabilité que $B < J < R$.¹⁰

- (a) Quelle est la probabilité qu'au moins deux dés affichent le même résultat ?
- (b) Quelle est la probabilité que $B < J < R$ sachant qu'aucun des trois dés n'affichent le même résultat ?

9. Cette statistique est factice – comme toutes les autres incluses dans ce problème.

10. Les symboles B , J et R représentent des objets mathématiques précis, qu'on appelle des variables aléatoires. Plus de détails au chapitre 4.

- (c) Quelle est la probabilité que $B < J < R$?

EXERCICE 3.6. On brasse un paquet de 52 cartes, puis on retourne les cartes une par une. Sachant que le premier as est la 20e carte du paquet, quelle est la probabilité que la carte immédiatement après soit :

- (a) l'as de pique ?
- (b) le deux de trèfle ?

EXERCICE 3.7 (La règle de succession de Laplace). Il y a $k + 1$ pièces de monnaie dans une boîte (numérotées de 0 à k), et lorsqu'on la lance dans les airs, la pièce $\#i$ retombe sur pile avec probabilité i/k . On choisit une pièce aléatoirement parmi les $k + 1$ pièces dans la boîte, puis on la lance plusieurs fois de suite. Si les n premiers lancers donnent tous pile, quelle est la probabilité que le $(n + 1)$ ième lancer donnera aussi pile ?

EXERCICE 3.8. Un modèle simple décrivant le prix d'une action fonctionne comme suit : à chaque jour le prix de l'action augmente de 1 unité avec probabilité p , ou diminue d'une unité avec probabilité $(1 - p)$. Les mouvements quotidiens sont indépendants.

- (a) Quelle est la probabilité que l'action ait retrouvé sa valeur initiale après deux jours ?
- (b) Quelle est la probabilité que l'action ait pris 1 unité de valeur après 3 jours ?
- (c) Sachant que l'action a augmenté de 1 unité après 3 jours, quelle est la probabilité que l'action a augmenté le premier jour ? (On assume que $p > 0$.)

EXERCICE 3.9. Dans un modèle génétique hyper-simplifié, la couleur des yeux d'une personne est déterminée par une seule paire de chromosomes. Si une personne porte le gène L sur chacun de ses chromosomes, ses yeux seront bleus. Mais si elle porte le gène R sur au moins l'un de ces deux chromosomes, alors ses yeux seront bruns. Pour cette raison on dit que le gène R est dominant sur le gène L .

Une personne reçoit indépendamment un gène de chacun de ses parents biologiques, et le gène reçu de chaque parent respectivement est équiprobablement l'un ou l'autre des gènes portés par le parent en question.

Supposons que Quentin et ses parents (Pauline et Oswald) ont tous trois les yeux bleus, mais que sa sœur Sabrina a les yeux bleus.

- (a) Quelle est la probabilité que Quentin ait un gène L ?
- (b) Supposons que Quentin soit hétérosexuel, et que sa femme Tatiana ait les yeux bleus. Quelle est la probabilité que leur premier enfant, Ursule, ait les yeux bleus ?
- (c) Quentin et Tatiana ont un second enfant : Victor. Intuitivement (c.à.d. sans faire de calculs), si on sait qu'Ursule a les yeux bleus, quel impact cela a-t-il sur les chances de Victor d'avoir aussi les yeux bleus ?
- (d) Quelle est la probabilité que Victor ait les yeux bleus sachant qu'Ursule, a les yeux bleus ?

EXERCICE 3.10. Soient E et F deux événements mutuellement exclusifs (disjoints) d'une expérience aléatoire.

Si on répète de manière indépendante cette expérience aléatoire plusieurs fois, montrer que la probabilité que l'événement E survienne pour la première fois avant F est simplement

$$\frac{\mathbb{P}\{E\}}{\mathbb{P}\{E\} + \mathbb{P}\{F\}}.$$

EXERCICE 3.11. Axel et Bénédicte jouent une série de parties d'échecs. Chaque partie est indépendante. Axel gagne avec probabilité p , et Bénédicte, avec probabilité $1 - p$. On arrête de jouer lorsque la différence entre leurs nombres totaux de gains est de plus ou moins 2, et le plus grand nombre de victoires à ce moment remporte la partie.

- (a) Quelle est la probabilité qu'exactement 4 parties soient jouées?
- (b) Quelle est la probabilité qu'Axel gagne?

EXERCICE 3.12. Une aiguille est dans l'une de n bottes de foin; elle est dans la i ème botte de foin avec probabilité P_i (et bien sûr, $\sum_{i=1}^n P_i = 1$).

Si notre aiguille est bien dans la i ème botte de foin, on la trouvera avec probabilité α_i .

Montrer que la probabilité que l'aiguille soit dans la j ème botte de foin, sachant qu'on ne l'a pas trouvée dans la i ème botte, est donnée par

$$\frac{P_j}{1 - \alpha_i P_i} \text{ si } i \neq j, \quad \frac{(1 - \alpha_i)P_i}{1 - \alpha_i P_i} \text{ si } i = j$$

EXERCICE 3.13. Des tentatives indépendantes qui résultent en un succès avec probabilité p chaque fois sont réalisées jusqu'à ce qu'on obtienne un total de r succès.

Montrer que la probabilité qu'exactement n tentatives soient nécessaires est de

$$\binom{n-1}{r-1} p^n (1-p)^{n-r}.$$

Indice : Pour obtenir r succès après exactement n tentatives, combien y a-t-il eu de succès dans les $n - 1$ premières tentatives?

EXERCICE 3.14. Des tentatives indépendantes qui résultent en un succès avec probabilité p chaque fois sont appelées «tentatives de Bernoulli». Soit P_n la probabilité que n tentatives de Bernoulli résultent en un nombre pair de succès (0 est pair). Montrer que

$$P_n = p(1 - P_{n-1}) + (1 - p)P_{n-1}.$$

Utiliser cette formule pour montrer que

$$P_n = \frac{1 + (1 - 2p)^n}{2}.$$

EXERCICE 3.15. Montrer la proposition 3.1 :

PROPOSITION. Soit Ω un ensemble fondamental fini, $\mathcal{E} = \mathcal{P}(\Omega)$ et \mathbb{P} la mesure de probabilité équiprobable sur Ω . Soient $E, F \in \mathcal{E}$ deux événements de Ω . Alors, on a que

$$\mathbb{P}\{E \mid F\} = \frac{|E \cap F|}{|F|}.$$

EXERCICE 3.16. Montrer le corollaire 3.1 :

COROLLAIRE. Si $A, B \in \mathcal{E}$ sont deux événements, et que $\mathbb{P}\{B\} = 1$, alors,

$$\mathbb{P}\{A\} = \mathbb{P}\{A \mid B\} = \mathbb{P}\{A \cap B\}.$$

EXERCICE 3.17. Nous allons d'abord faire la définition suivante :

DÉFINITION. Soit une famille d'événements $(E_i \in \mathcal{E})_{i \in I}$, avec un certain ensemble d'indices $I \subseteq \mathbb{N}$ potentiellement dénombrable.

On dira que les E_i forment une **presque-partition** de l'ensemble fondamental Ω si et seulement si :

- i. $\mathbb{P}\{\bigcup_{i \in I} E_i\} = 1$;
- ii. $\mathbb{P}\{E_i \cap E_j\} = 0$ pour tous $i, j \in I$ tels que $i \neq j$.

Avec cette définition, montrer le corollaire suivant pour la formule des probabilités totales (proposition 3.2) :

COROLLAIRE. Soit $A \in \mathcal{E}$ un événement, et soit $(B_i)_{i \in I}$ (avec $I \subseteq \mathbb{N}$ un ensemble d'indices potentiellement dénombrables) une presque-partition de l'ensemble fondamental Ω .

Alors, on a

$$\mathbb{P}\{A\} = \sum_{i \in I} \mathbb{P}\{A \mid B_i\} \mathbb{P}\{B_i\}.$$

EXERCICE 3.18. Montrer le corollaire 3.2 :

COROLLAIRE. Si $(E_i)_{i \leq \mathbb{N}}$ sont une famille monotone décroissante d'événements, alors

$$\mathbb{P}\{E_n\} = \prod_{i=1}^n \mathbb{P}\{E_i \mid E_{i-1}\}.$$

(On considère que $\mathbb{P}\{E_1 \mid E_0\} = \mathbb{P}\{E_1\}$ car E_0 n'est pas défini.)

Deuxième partie

Les variables aléatoires

Chapitre 4

Les variables aléatoires

Nous avons plusieurs fois utilisé l'exemple d'un lancer de deux dés équilibrés à six faces, et considéré plusieurs événements, par exemple :

- le premier dé montre un résultat pair ;
- la somme des deux dés est de 5.

Jusqu'ici, nous avons utilisé simplement des événements pour calculer les probabilités pertinentes. Toutefois, ces événements ont tous une chose en commun : ils sont définis en référant à des valeurs numériques spécifiques qui dépendent du résultat de notre expérience aléatoire – ce que l'on appellera des *variables aléatoires*.

4.1. Les variables aléatoires

DÉFINITION 4.1 (Variable aléatoire). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. Une **variable aléatoire** est une fonction $X : \Omega \rightarrow \mathbb{R}$.^{1, 2}

Vous avez bien lu : les variables aléatoires sont en fait des... fonctions ! Mais pensez-y : ça fait tout plein de sens. Une *variable aléatoire*, c'est une valeur numérique qui dépend du résultat d'une expérience aléatoire. C'est donc une fonction, qui prend comme argument le résultat ω de l'expérience aléatoire, et qui retourne la valeur en question.

EXEMPLE 4.1. Si on lance deux dés, on peut définir les variables aléatoires X_1 et X_2 , qui montrent les résultats pour le premier et le second dé. On peut aussi définir S , la somme des deux dés, D , la différence, etc.

EXEMPLE 4.2. La température, la pression, la force et la direction du vent pour le lendemain sont toutes des variables aléatoires dans un modèle stochastique de météorologie.

EXEMPLE 4.3. Le nombre de clics enregistrés par un compteur Geiger en une minute est une variable aléatoire.

EXEMPLE 4.4. Si on pige sans remise k boules d'un sac qui en contient n , dont m mauves, on pourrait avoir M , le nombre de boules mauves qu'on a pigé.

EXEMPLE 4.5. Si on conduit une expérience aléatoire, et qu'on considère un événement E , alors la variable aléatoire $\mathbb{1}_E$ – la **fonction indicatrice**³ de E – vaut 1 si l'événement est réalisé, et 0 sinon.

1. La « vraie » définition requiert que les ensembles de la forme $\{\omega \in \Omega : X(\omega) < \lambda\}$ soient dans \mathcal{E} . Il s'agit d'une condition technique dont nous ne nous préoccupons pas vraiment.

2. Une variable aléatoire peut aussi être à image dans \mathbb{C} , ou dans n'importe quel autre ensemble. Toutefois, certains outils analytiques ne peuvent être utilisés que lorsque la variable aléatoire est à image dans un « espace mesurable ». Pour nous, les variables aléatoires seront toujours réelles (ou, quelque fois, complexes).

3. Consulter la section A.5 au sujet des fonctions indicatrices.

EXEMPLE 4.6. Si on joue à pile ou face de façon répétée, on pourrait considérer X , le nombre de lancers à réaliser pour obtenir «pile». On pourrait aussi considérer Y , le nombre de fois qu'on a obtenu «pile» en n lancers.

Attention de bien faire la différence entre une variable aléatoire et un événement! Les événements sont des « choses qui peuvent se produire », tandis que les variables aléatoires sont des valeurs numériques précises qui dépendent du résultat de l'expérience aléatoire.

Mathématiquement, ça n'aurait pas de sens de considérer « la probabilité du résultat du premier dé » ; on calcule la probabilité d'un événement.

4.1.1. Des raccourcis dans la notation. L'intérêt des variables aléatoires, c'est qu'elles nous permettent de continuer de rendre le langage des probabilités plus près de l'intuition. Comme mentionné en introduction de ce chapitre, les variables aléatoires sont des objets très naturels pour définir des événements.

Plusieurs des événements que nous avons considéré peuvent être écrits de la façon suivante :

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

c'est à dire comme le pré-image d'un certain ensemble de valeurs A pour une variable aléatoire X .⁴

Par exemple, les événements suivants :

- « La somme des deux dés est d'au moins 7 » ;
- « On a obtenu pile au moins trois fois en 6 lancers » ;
- « On n'a pigé aucune boule mauve » ;
- « Le mercure excède la barre des 30 °C » ;

sont tous des événements formulés comme étant « les pré-images d'ensembles de valeurs par une variable aléatoire. »

Par conséquent, on peut noter la probabilité d'un tel événement de la façon suivante :

$$\mathbb{P}\{E\} = \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\},$$

ce qui est parfaitement cohérent avec notre compréhension des événements et de leurs probabilités jusqu'à date. On lirait simplement « la probabilité de l'événement qui regroupe tous les éléments de l'ensemble fondamental pour lesquels la valeur de la variable aléatoire X est contenue dans la plage de valeurs A ».

C'est très lourd. On fera donc le raccourci de notation suivant :

$$\mathbb{P}\{E\} = \mathbb{P}\{X \in A\},$$

qu'on lira tout simplement « la probabilité que X soit dans A ».

De façon général, c'est un raccourci que l'on fera très souvent puisqu'il est extrêmement pratique. Voici des exemples pour bien comprendre :

EXEMPLE 4.7. Si on lance deux dés équilibrés à six faces, que X_1 est le résultat du premier dé, et que X_2 est le résultat du second dé. Alors, $S = X_1 + X_2$ est la somme des deux dés, et l'événement $\{\omega \in \Omega : S(\omega) = 7\}$ que la somme des deux dés donne 7 peut être abrégé par $\{S = 7\}$.

Ainsi, on notera simplement $\mathbb{P}\{S = 7\}$ la probabilité que la somme des deux dés donne 7.

4. Si la notion de pré-image (ou la notation) vous embête, il serait bon de consulter la section A.5.

EXEMPLE 4.8. Si on joue à pile ou face de façon répétée, n fois, et qu'on note N le nombre de fois qu'on a obtenu « pile », alors la probabilité

$$\mathbb{P}\left\{N \geq \frac{n}{2}\right\}$$

est la probabilité de l'événement que « on a obtenu pile à plus de la moitié des lancers ».

EXEMPLE 4.9. Si T est la température qu'il fera demain, on pourrait noter

$$\mathbb{P}\{T \geq 30^\circ\text{C}\},$$

la probabilité que la température de demain dépassera les 30°C .

Attention à ne pas se méprendre ! Les probabilités sont **toujours** – je répète, **T O U J O U R S** – des probabilités d'**événements**. Jamais de rien d'autre ! On a simplement allégé la notation du mieux qu'on peut.

D'ailleurs, tant qu'on y est, on va souvent calculer la probabilité conjointe de plusieurs événements. Jusqu'ici, on utilisait toujours la notation ensembliste, \cap , pour l'intersection. C'est très bien, mais encore une fois, ça obscurcit parfois un peu l'intuition.

On va donc dorénavant utiliser le raccourci suivant :

$$\mathbb{P}\{E, F\} := \mathbb{P}\{E \cap F\},$$

en gardant toujours à l'esprit qu'il ne s'agit que d'un raccourci dans la notation.

EXEMPLE 4.10. Si X_1 et X_2 sont les résultats respectifs des deux dés à six faces lors d'un lancer, alors,

$$\mathbb{P}\{X_1 = 1, X_2 \geq 2\},$$

est la probabilité que le premier dé donne 1 et que le second est supérieur à 2. C'est plus court et plus intelligible que d'écrire :

$$\mathbb{P}\{\{\omega \in \Omega : X_1(\omega) = 1\} \cap \{\omega \in \Omega : X_2(\omega) \geq 2\}\}.$$

Mais ces expressions signifient **exactement la même chose !**

4.1.2. La distribution d'une variable aléatoire. Évidemment, la théorie des probabilités ne permet pas de prédire avec certitude le résultat d'une expérience aléatoire. Le mieux qu'on peut faire, concernant un événement, c'est de connaître sa probabilité.

De même, le mieux qu'on puisse faire, concernant une variable aléatoire, c'est de connaître *sa distribution statistique*.

La distribution, c'est un objet mathématique qui nous dit, pour n'importe quel ensemble de valeurs⁵, la probabilité que notre variable se retrouve dedans.⁶

Heureusement, il suffit de se limiter à connaître les probabilités pour les parties de \mathbb{R} de la forme $(-\infty, x]$.

DÉFINITION 4.2 (Fonction de répartition, complémentaire). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire quelconque.

La **fonction de répartition** de X est la fonction $F : \mathbb{R} \rightarrow [0, 1]$ donnée par

$$F(x) = \mathbb{P}\{X \leq x\}.$$

On l'appelle aussi parfois **fonction de distribution**, ou **fonction de distribution cumulative**.

5. Ensemble mesurable. Mais, encore une fois, nous ne nous en soucierons pas trop...

6. Les étudiant.e.s plus curieux sont invité.e.s à contempler la possibilité de définir une mesure de probabilité $\mu_X(A) = \mathbb{P}\{X \in A\}$. Voir l'exercice 4.19.

La **fonction de répartition complémentaire** de X est la fonction $\bar{F} : \mathbb{R} \rightarrow [0, 1]$ donnée par

$$\bar{F}(x) = \mathbb{P}\{X > x\} = 1 - F(x).$$

Dans divers contextes, on l'appelle aussi parfois la **queue** de X .

On déduit immédiatement des axiomes et des propriétés de la mesure de probabilités la proposition suivante :

PROPOSITION 4.1. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire, et $F(x) = \mathbb{P}\{X \leq x\}$ sa fonction de répartition.*

- i. F est croissante.
- ii. $F(x) \leq 1$ pour tout x .
- iii. $\lim_{x \rightarrow a^-} F(x) = \mathbb{P}\{X < a\}$ pour tout $a \in \mathbb{R}$.
- iv. $\lim_{x \rightarrow a^+} F(x) = F(a)$ pour tout $a \in \mathbb{R}$.⁷

DÉMONSTRATION. i. On montre cette propriété simplement en remarquant que pour $x \leq y$, $\{X \leq x\} \subseteq \{X \leq y\}$, puisque si $X \leq x \leq y$, par transitivité, alors $X \leq y$. Mais alors, $F(x) = \mathbb{P}\{X \leq x\} \leq \mathbb{P}\{X \leq y\} = F(y)$ dès que $x \leq y$, et F est croissante.

- ii. Cela est vrai puisque $F(x)$ est une probabilité pour tout x .
- iii. Puisque F est bornée et monotone, clairement la limite existe. Si on considère une suite monotone croissante x_n convergeant vers a , alors $\{X \leq x_n\}$ est une suite monotone croissante d'événements, et

$$\lim_{n \rightarrow \infty} \{X \leq x_n\} = \bigcup_{n \in \mathbb{N}} \{X \leq x_n\} = \{X < a\}.$$

Par la propriété de continuité pour la mesure de probabilités, on trouve donc

$$\lim_{x \rightarrow a^-} F(x) = \lim_{n \rightarrow \infty} \mathbb{P}\{X \leq x_n\} = \mathbb{P}\{X < a\}.$$

- iv. Si on considère une suite monotone décroissante x_n qui converge vers a , alors

$$\lim_{n \rightarrow \infty} \{X \leq x_n\} = \bigcap_{n \in \mathbb{N}} \{X \leq x_n\} = \{X \leq a\},$$

et par la propriété de continuité pour la mesure de probabilités, on trouve que

$$\lim_{x \rightarrow a^+} F(x) = \lim_{n \rightarrow \infty} \mathbb{P}\{X \leq x_n\} = \mathbb{P}\{X \leq a\} = F(a).$$

□

On remarque que F n'est pas forcément continue. En fait, il arrive même très souvent que F soit discontinue. On notera

$$F(x^-) = \lim_{y \rightarrow x^-} F(y);$$

Alors, il suit toujours le résultat suivant :

$$(4.1.1) \quad \mathbb{P}\{X = x\} = F(x) - F(x^-).$$

7. Ces deux dernières propriétés définissent en fait les fonctions continues à droite avec limite à gauche – ou simplement, *càdlàg*. Voir la section B.12 pour plus de détails.

Contrairement à la majorité des situations en mathématiques, ici, ce sera plus facile de traiter le cas où F est discontinue. En fait, on va commencer par le cas où F est une fonction en escalier...

4.2. Les variables aléatoires discrètes

On va commencer notre étude des variables aléatoires tout en douceur par l'étude d'une catégorie particulière de variables aléatoires : les variables aléatoires dites « discrètes ».

Pour nos fins, nous les définirons comme suit :

DÉFINITION 4.3 (Variable discrète). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. On dit que X est **discrète** si l'image $X(\Omega)$ de l'ensemble fondamental par X est finie ou dénombrable.⁸

Pour nous, le plus souvent, les variables aléatoires discrètes prendront des valeurs entières, mais ce n'est pas forcément le cas. Le plus important, c'est surtout qu'on peut « énumérer » les valeurs possibles pour la variable X , comme dans une « liste » – même si celle-ci est potentiellement infinie.

Pour ce qui suit, on va se doter d'une variable aléatoire discrète X quelconque sur notre espace de probabilités $(\Omega, \mathcal{E}, \mathbb{P})$.

On va noter $V = \{v_1, v_2, v_3, \dots\}$ la liste de valeurs que peut prendre notre variable aléatoire X .

On appellera V le **support** de X , dans le contexte des variables aléatoires discrètes.⁹

On peut alors définir la *fonction de masse* de notre variable aléatoire X :

DÉFINITION 4.4 (Fonction de masse). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire discrète pouvant prendre les valeurs v_1, v_2, v_3, \dots

Alors, la **fonction de masse** $p_X : \mathbb{R} \rightarrow [0, 1]$ pour notre variable aléatoire X est donnée par

$$p_X(x) = \begin{cases} \mathbb{P}\{X = v_i\} & \text{si } x = v_i \text{ pour un certain } i \\ 0 & \text{pour toute autre valeur de } x. \end{cases}$$

La fonction de masse est appelée ainsi en raison d'une analogie avec la physique – on s'imagine qu'une masse totale de $1kg$ est répartie en un certain nombre (potentiellement infini, dénombrable) de poids fixés à la tige à certaines positions. Le poids fixé à la position v_i est de masse $m_i = p_X(v_i)$ (figure 4.1)

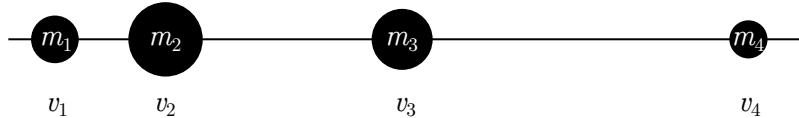


FIGURE 4.1. Les masses fixées à chaque point représentent la probabilité que notre variable prenne la valeur correspondante.

Avant de continuer, voici quelques exemples :

8. Le terme *discret* renvoie à la topologie *discrète* qui est la topologie héritée par le *support* de la variable aléatoire. Nous parlerons plus loin de variables aléatoires « continues » ; ce ne sont pas forcément des fonctions continues. La terminologie ici est assez mélangeante, et est surtout due à de vieilles habitudes datant d'avant que toutes ces notions deviennent claires.

9. Il existe une définition du *support* d'une variable aléatoire dans un contexte plus général, mais il fait intervenir des notions de topologie plus avancées. Nous en reparlerons plus loin.

EXEMPLE 4.11. En supposant qu'on lance n pièces de monnaie équilibrées, indépendamment les unes des autres, et que N dénote le nombre d'entre elles qui sont tombées sur pile.

Quelle est la fonction de masse pour la variable aléatoire N ?

SOLUTION. Les valeurs possibles pour N sont $V = \{0, 1, 2, 3, \dots, n\}$.

On va donc calculer $p_N(k) = \mathbb{P}\{N = k\}$, pour k de 0 à n .

Quelle est la probabilité d'obtenir k fois « pile » ?

On a 2^n configurations équiprobables de « pile » et de « face ». Parmi elles, il y en a $\binom{n}{k}$ où exactement k pièces tombent sur « pile ».

Donc, on doit avoir que

$$p_N(k) = \frac{\binom{n}{k}}{2^n}.$$

EXEMPLE 4.12. On lance deux dés équilibrés à 6 faces.

- (a) Quelle est la fonction de masse pour le résultat du premier dé ? Le résultat du second dé ?
- (b) Quelle est la fonction de masse pour la somme des deux dés ?

SOLUTION. On va noter X_1, X_2 les résultats des premier et second dés respectivement, et $S = X_1 + X_2$ la somme des deux.

- (a) On a évidemment que $p_{X_1}(i) = 1/6$ pour $i = 1, \dots, 6$, et 0 pour toute autre valeur, même chose pour p_{X_2} .
- (b) On veut calculer $\mathbb{P}\{S = k\}$. Les valeurs possibles sont évidemment $V = \{2, 3, 4, \dots, 12\}$.

Par la formule des probabilités totales, on a nécessairement

$$\mathbb{P}\{S = k\} = \sum_{i=1}^n \mathbb{P}\{X_1 = i, S = k\},$$

et bien sûr, $S = X_1 + X_2$. Donc $\{X_1 = i, S = k\} = \{X_1 = i, X_2 = k - i\}$. Puis, puisque les lancers sont indépendants, les événements $\{X_1 = i\}$ et $\{X_2 = k - i\}$ sont indépendants, et finalement,

$$p_S(k) = \sum_{i=1}^6 p_{X_1}(i)p_{X_2}(k - i).$$

Il suffit maintenant de compter combien de termes sont non-nuls – ceux-là vaudront $1/36$ chacun, en raison des fonctions de masse calculées plus haut.

Pour que $p_{X_2}(k - i) = 1/6$, il faut que $1 \leq k - i \leq 6$, soit que $k - 6 \leq i \leq k - 1$. On en est donc à compter combien de termes sont non-nuls, soit combien de i entre 1 et 6 sont aussi entre $k - 6$ et $k - 1$. Ce nombre correspond à la différence (plus un) entre les bornes de l'intersection de ces deux intervalles, soit :

$$\begin{aligned} \min\{6, k - 1\} - \max\{1, k - 6\} + 1 &= \frac{(5 + k) - |7 - k|}{2} - \frac{k - 5 + |7 - k|}{2} + 1 \\ &= 6 - |k - 7| \end{aligned}$$

Pour faire ce calcul, on a utilisé les identités bien connues :

$$\max\{a, b\} = \frac{a + b + |a - b|}{2}, \quad \min\{a, b\} = \frac{a + b - |a - b|}{2}.$$

Finalement, on a donc

$$p_S(k) = \frac{6 - |k - 7|}{36},$$

pour k entre 2 et 12. Pour toutes les autres valeurs, c'est évidemment 0.

Le lien entre la fonction de répartition et la fonction de masse de notre variable aléatoire, c'est que pour une variable aléatoire X quelconque,¹⁰

$$(4.2.1) \quad p_X(x) = F(x) - F(x^-).$$

Autrement dit, notre fonction de masse, c'est une fonction qui nous donne les « sauts » dans la fonction de répartition.

À l'inverse, pour une variable X discrète, on aura toujours que :

$$(4.2.2) \quad F(x) = \sum_{v \in V: v \leq x} p_X(v).$$

4.2.1. La condition de normalisation. Toujours avec notre variable aléatoire discrète X , pouvant prendre ses valeurs dans V , on remarque que les événements $(\{X = v\})_{v \in V}$ forment une partition de Ω .

Dès lors, il est nécessaire d'avoir que

$$(4.2.3) \quad \sum_{v \in V} p_X(v) = 1.$$

C'est ce qu'on appelle la condition de normalisation. On peut vérifier aisément, dans les exemples précédents, que cette condition tient. Voici un autre exemple :

EXEMPLE 4.13. On considère N le nombre de clics enregistrés par un compteur Geiger en une minute à proximité d'une source radioactive de particules β . La fonction de masse de N est donnée par $p_N(k) = c \frac{\lambda^k}{k!}$, pour des entiers $k \geq 0$.

- (a) Trouver la valeur de c .
- (b) Donner la probabilité qu'il y ait eu au moins un clic.

SOLUTION. (a) Pour trouver la valeur de c , on remarque que notre fonction de masse doit respecter la condition de normalisation :

$$\sum_{k=0}^{\infty} p_N(k) = c \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = ce^\lambda = 1.$$

On conclue donc que

$$c = e^{-\lambda}.$$

- (b) On doit avoir que $\mathbb{P}\{N \geq 1\} = 1 - \mathbb{P}\{N = 0\} = 1 - p_N(0) = 1 - c$, encore une fois par la condition de normalisation. On a donc $\mathbb{P}\{N \geq 1\} = 1 - e^{-\lambda}$.

10. L'équation (4.2.1) est toujours vraie, dans un contexte général et pas seulement pour les variables aléatoires discrètes. Nous en reparlerons plus loin.

4.3. L'espérance

La notion d'espérance est un autre outil absolument fondamental pour l'étude de la théorie des probabilités. Elle est toutefois... difficile... à définir proprement à ce niveau.

Du point de vue mathématique, l'espérance est une **fonctionnelle** : c'est une fonction à valeur réelle, dont l'argument est une variable aléatoire.

Sur un espace de probabilités $(\Omega, \mathcal{E}, \mathbb{P})$, si Ω est fini ou dénombrable et qu'on a une variable aléatoire X , on voudrait l'**espérance de X** :¹¹

$$(4.3.1) \quad \mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}\{\{\omega\}\}.$$

Intuitivement, ce qu'on veut faire, c'est trouver le *centre de masse* de notre distribution de probabilités. Pour parvenir à cela, ce qu'on peut faire, c'est prendre une moyenne de toutes les valeurs possibles, « pondérées par leurs probabilités respectives ».

Si on reprend l'analogie de la tige, où des poids de masse m_i sont fixés à des positions v_i , le centre de masse serait :

$$C.M = \sum_i v_i \frac{m_i}{M},$$

avec $M = \sum_i m_i = 1$.

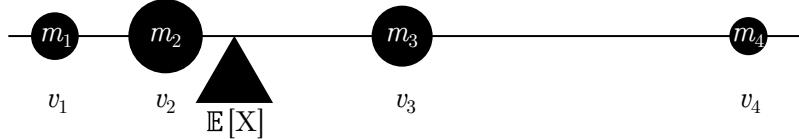


FIGURE 4.2. L'espérance $\mathbb{E}[X]$ correspond au centre de masse de la distribution.

En fait, pour les variables discrètes, l'espérance, c'est exactement ça :¹²

$$(4.3.2) \quad \mathbb{E}[X] = \sum_{v \in V} v p_X(v).$$

C'est la moyenne de toutes les valeurs possibles que peut prendre la variable aléatoire X , pondérée par les probabilités de chacune des valeurs respectivement. On peut prendre la somme puisque les valeurs en question sont dénombrables.

REMARQUE (La différence entre la moyenne et l'espérance). L'espérance d'une variable aléatoire n'est pas sa *moyenne* – la moyenne d'un ensemble de valeurs est la somme de toutes les valeurs, divisée par le nombre total de valeurs.

L'espérance d'une variable aléatoire, c'est un nombre associé à sa distribution. La confusion existe, parce qu'en statistiques, pour estimer l'espérance de la distribution de laquelle on a échantillonné plusieurs valeurs, on fait la moyenne des valeurs obtenues.

11. La réelle définition de l'espérance fait appel à la notion d'intégrale au sens de Lebesgue – en fait, l'espérance de X n'est simplement que l'intégrale de Lebesgue de X sur Ω , pour la mesure \mathbb{P} – mais, encore une fois, cela dépasse largement le contenu de ce cours. Toutefois, dans le cas particulier où Ω est un ensemble au plus dénombrable, l'équation (4.3.1) équivaut à la définition.

12. L'équation (4.3.2) n'est pas, néanmoins, une *définition* de l'espérance. C'est une formule qui permet de la calculer pour une variable aléatoire discrète.

REMARQUE. On remarque aussi que l'espérance ne dépend uniquement que de la distribution de X – pas nécessairement de X précisément. Deux variables aléatoires qui ont la même distribution statistique ont la même espérance, même si elles n'ont pas forcément les mêmes valeurs.

EXEMPLE 4.14. Supposons qu'on réalise une expérience aléatoire et que l'événement E ait une probabilité p de se réaliser. On considère la variable aléatoire $X = \mathbb{1}_E$, l'indicatrice de l'événement E .

Alors, l'espérance de X est

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

REMARQUE. L'exemple précédent n'est pas anodin. En théorie des probabilités, ce fait est très souvent utilisé : **L'espérance de l'indicatrice d'un événement est la probabilité de cet événement.**

EXEMPLE 4.15. Si X est le résultat affiché par un dé équilibré à n faces, trouver $\mathbb{E}[X]$.

SOLUTION. Les valeurs possibles sont $V = \{1, 2, 3, \dots, n\}$, et bien sûr la fonction de masse pour X est $p_X(k) = 1/n$ pour $k \in V$.

Par conséquent, l'espérance de X est donnée par

$$\begin{aligned} \mathbb{E}[X] &= \sum_{v \in V} vp_X(v) \\ &= \sum_{k=1}^n k \cdot \frac{1}{n} \\ &= \frac{1}{n} \sum_{k=1}^n k \\ &= \frac{1}{n} \cdot \frac{n(n+1)}{2} \\ &= \frac{n+1}{2}. \end{aligned}$$

EXEMPLE 4.16. Des écologistes ont recensé 117 lièvres appartenant à 3 espèces différentes : il y en a 58 de l'espèce 1, 21 de l'espèce 2, 38 de l'espèce 3. Un chasseur passe sur le territoire où a eu lieu le recensement et abat un lièvre. Si X est le nombre de lièvres appartenant à l'espèce de l'animal abattu par le chasseur, quelle est l'espérance de X ?

SOLUTION. Les valeurs possibles pour X sont $v_1 = 58$, $v_2 = 21$ et $v_3 = 38$. La fonction de masse pour X correspond à $p_X(v_i)$, la probabilité que le chasseur ait abattu un lièvre de l'espèce 1. C'est $p_X(v_1) = 58/117$. De même, $p_X(v_2) = 21/117$ et $p_X(v_3) = 38/117$.

En tout, on a donc

$$\mathbb{E}[X] = \frac{58^2 + 21^2 + 38^2}{117} = \frac{5249}{117} \approx 44,86.$$

On remarque que le nombre moyen de lièvres par espèce est de $117/3 = 39$ – significativement plus bas.

Cela est dû au fait que le chasseur n'abat pas un lièvre choisi de l'une des trois espèces équiprobablement. Le chasseur abat plus probablement un lièvre qui appartient à une espèce plus populeuse.

4.4. Les fonctions de variables aléatoires

Bien sûr, on peut manipuler les variables aléatoires à l'aide de fonctions de nombres réelles – comme on est habitué.e.s de le faire pour des nombres ordinaires.

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction quelconque. On peut considérer la variable aléatoire $Y = f \circ X$, la composition de f avec la variable aléatoire X – qui souvenez-vous, est une fonction de Ω dans \mathbb{R} .

Encore une fois, pour alléger la notation, et pour la rendre plus intuitive, on notera $f(X) := f \circ X$ si $f : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction et X est une variable aléatoire à valeurs réelles.

Intuitivement, $f(X)$ est exactement ce qu'on s'attend que ce soit : si X vaut une certaine valeur v , alors $f(X)$ vaut $f(v)$.

Si X est une variable aléatoire discrète, alors $Y = f(X)$ est aussi une variable aléatoire discrète, et si le support de X est l'ensemble de valeurs réelles V (au plus dénombrable), alors le support de Y est V' , l'image de V par $f : V' = f(V)$.

On peut dès lors calculer la fonction de masse pour Y :

$$(4.4.1) \quad p_Y(v') = \mathbb{P}\{f(X) = v'\} = \mathbb{P}\{X \in f^{-1}(v')\} = \sum_{v \in V : f(v) = v'} p_X(v).$$

EXEMPLE 4.17. Soit X une variable aléatoire avec $\mathbb{P}\{X = -1\} = 1/5 = \mathbb{P}\{X = 1\}$, et $\mathbb{P}\{X = 0\} = 3/5$. On considère maintenant $Y = X^2$.

Les valeurs possibles pour Y sont $\{0, 1\}$. On a que $p_Y(0) = p_X(0) = 3/5$, et $p_Y(1) = p_X(-1) + p_X(1) = 2/5$.

EXEMPLE 4.18. Soit X une variable aléatoire discrète avec $\mathbb{P}\{X = k\} = e^{-\lambda} \lambda^k / k!$ pour tout $k \geq 0$.

On considère $Y = \cos(\pi X)$. Calculer $\mathbb{P}\{Y = 1\}$.

SOLUTION. On a que $\mathbb{P}\{Y = 1\} = \mathbb{P}\{X \text{ est pair}\}$.

On a par ailleurs

$$\begin{aligned} \mathbb{P}\{X \text{ est pair}\} &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{2k!} \\ &= e^{-\lambda} \cosh(\lambda) \\ &= e^{-\lambda} \frac{e^\lambda + e^{-\lambda}}{2} \\ &= \frac{1 + e^{-2\lambda}}{2} \end{aligned}$$

Bien sûr, $\mathbb{P}\{Y = -1\} = 1 - \mathbb{P}\{Y = 1\}$.

EXEMPLE 4.19. Avec les mêmes variables qu'à l'exemple précédent, trouver $\mathbb{E}[Y]$.

SOLUTION. On a clairement

$$\mathbb{E}[Y] = \mathbb{P}\{Y = 1\} - \mathbb{P}\{Y = -1\} = e^{-2\lambda}.$$

Le dernier calcul est un calcul d'espérance, et on aurait pu le faire autrement. En effet, nous avons la proposition suivante :

PROPOSITION 4.2. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire discrète de support V (au plus dénombrable), et soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction.

Alors, nous avons que

$$(4.4.2) \quad \mathbb{E}[f(X)] = \sum_{v \in V} f(v)p_X(v).$$

DÉMONSTRATION. Soit $Y = f(X)$, de support $V' = f(V)$. Alors, nous avons que

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{v' \in V'} v' p_Y(v') \\ &= \sum_{v' \in V'} v' \sum_{v \in V : f(v)=v'} p_X(v) \\ &= \sum_{v' \in V'} \sum_{v \in V : f(v)=v'} f(v)p_X(v) \\ &= \sum_{v \in V} f(v)p_X(v). \end{aligned}$$

Le passage de l'avant-dernière ligne à la dernière ligne s'effectue en remarquant que, puisque $V' = f(V)$, il doit pour tout $v' \in V'$ exister au moins un $v \in V$ tel que $f(v) = v'$. \square

EXEMPLE 4.20. On peut refaire le même exemple que précédemment : si $\mathbb{P}\{X = k\} = e^{-\lambda}\lambda^k/k!$, et si $Y = \cos(\pi X)$, alors bien sûr

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[\cos(\pi X)] \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \cos(\pi k) \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(-1)^k \lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} \\ &= e^{-\lambda} e^{-\lambda} = e^{-2\lambda}. \end{aligned}$$

DÉFINITION 4.5 (Moment). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. Son **n**ième **moment** (pour $n > 0$) est donné par

$$\mathbb{E}[X^n].$$

Dans le cas d'une variable avec support V (au plus dénombrable) et fonction de masse p , on a :

$$\mathbb{E}[X^n] = \sum_{v \in V} v^n p(v).$$

EXEMPLE 4.21. Soit X une variable aléatoire avec $p_X(k) = \frac{90}{\pi^4 k^4}$ pour $k \geq 1$.

- (a) Trouver $\mathbb{E}[X^2]$.
- (b) Trouver $\mathbb{E}[X^3]$.

SOLUTION. (a) On calcule :¹³

$$\begin{aligned}\mathbb{E}[X^2] &= \frac{90}{\pi^4} \sum_{k=1}^{\infty} \frac{k^2}{k^4} \\ &= \frac{90}{\pi^4} \frac{\pi^2}{6} \\ &= \frac{15}{\pi^2}\end{aligned}$$

(b) On calcule encore :

$$\begin{aligned}\mathbb{E}[X^3] &= \frac{90}{\pi^4} \sum_{k=1}^{\infty} \frac{k^3}{k^4} \\ &= \frac{90}{\pi^4} \sum_{k=1}^{\infty} \frac{1}{k}\end{aligned}$$

(*Effet musical dramatique de films muets :) Tan tan TAAAAAAAAAAAAAN!* La série diverge ! Que fait-on ?

Eh bien on ne peut rien faire ; ça signifie simplement que l'espérance de X^3 est infinie. Ça arrive assez souvent.

REMARQUE. On dit de variables aléatoires dont l'espérance est infinie qu'elles ont une (ou des) *queue(s) lourde(s)*, en référence à la lente décroissance de leur fonction de répartition complémentaire \bar{F} (et possiblement la lente décroissance de $F(-x)$ lorsque $x \rightarrow \infty$).

À noter : lorsqu'un échantillon est vraisemblablement issu d'une distribution à queue lourde, la moyenne de l'échantillon est une mauvaise mesure de la tendance centrale ; en effet, celle-ci divergerait avec la taille de l'échantillon.

C'est la raison pour laquelle on a souvent plus tendance à s'intéresser, par exemple, au revenu médian des ménages, plutôt qu'au revenu moyen : le revenu des ménages est distribué sur plusieurs ordres de grandeur comme une variable aléatoire à queue lourde – le revenu moyen est donc une sur-estimation de la tendance centrale.

On termine cette section avec la proposition suivante :

PROPOSITION 4.3. Soient $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire discrète avec fonction de masse p_X et support V , et $a, b \in \mathbb{R}$ deux constantes réelles.

Alors,

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

DÉMONSTRATION. Pour faire la preuve, on va utiliser la proposition 4.2, ainsi que la condition de normalisation :

$$\begin{aligned}\mathbb{E}[aX + b] &= \sum_{v \in V} (av + b)p_X(v) \\ &= a \sum_{v \in V} vp_X(v) + b \sum_{v \in V} p_X(v) \\ &= a\mathbb{E}[X] + b.\end{aligned}$$

13. On utilise les valeurs connues de la fonction ζ de Riemann : $\zeta(2) = \pi^2/6$, et $\zeta(4) = \pi^4/90$.

□

4.5. La variance

On utilise souvent les moments pour décrire la forme « globale » d'une distribution statistique. Par exemple, le premier moment (l'espérance) d'une variable aléatoire donne souvent une bonne idée de la tendance centrale d'une distribution (pourvu qu'elle existe).

Supposons qu'on ait une variable aléatoire X . Alors, on pourrait noter $\mu = \mathbb{E}[X]$ le premier moment de X .

À partir de là, on peut définir les moments centrés :

DÉFINITION 4.6 (Moments centrés). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. Le *nième moment centré* de X est donné par

$$\mathbb{E}[(X - \mu)^n].$$

Évidemment, le premier moment centré est toujours nul – en effet,

$$\mathbb{E}[(X - \mathbb{E}[X])] = \mathbb{E}[X] - \mathbb{E}[X] = 0.$$

Le second moment centré, cependant, est d'importance capitale – on lui a même donné un nom :

DÉFINITION 4.7 (Variance). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. La **variance** de X est le second moment centré de X :

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

La variance est nommée ainsi parce qu'elle donne une indication de la « dispersion » des valeurs prises par la variable aléatoire. La proposition suivante donne une expression pratique pour la variance :

PROPOSITION 4.4. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire.*

Alors,

$$(4.5.1) \quad \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

DÉMONSTRATION. Cette proposition est vraie pour toute variable aléatoire, indistinctement. Nous allons ici faire une preuve seulement dans le cas où X est une variable aléatoire discrète.

Dans ce cas, on assume que p_X est la fonction de masse, et V le support (au plus dénombrable) de X . On va noter $\mu = \mathbb{E}[X]$ pour raccourcir la notation. On a alors :

$$\begin{aligned} \text{Var}[X] &= \sum_{v \in V} (v - \mu)^2 p_X(v) \\ &= \sum_{v \in V} (v^2 - 2v\mu + \mu^2) p_X(v) \\ &= \sum_{v \in V} v^2 p_X(v) - 2\mu \sum_{v \in V} v p_X(v) + \mu^2 \sum_{v \in V} p_X(v) \end{aligned}$$

Mais, par l'équation (4.3.2), $\sum_{v \in V} v p_X(v) = \mu$, et par la condition de normalisation, $\sum_{v \in V} p_X(v) = 1$. De plus, par la proposition 4.2, $\sum_{v \in V} v^2 p_X(v) = \mathbb{E}[X^2]$.

En tout, ça fait donc :

$$\text{Var}[X] = \mathbb{E}[X^2] - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

□

Avant de nous plonger dans un gros bloc de ce chapitre, nous allons voir un petit exemple :

EXEMPLE 4.22. Soit X le résultat affiché par un dé équilibré à n faces. La variance de X est

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Nous avons déjà calculé que $\mathbb{E}[X] = \frac{n+1}{2}$. Il faut maintenant calculer $\mathbb{E}[X^2]$.

On a

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=1}^n \frac{k^2}{n} \\ &= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} \\ &= \frac{(n+1)(2n+1)}{6}.\end{aligned}$$

Finalement,

$$\text{Var}[X] = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}.$$

On simplifie et on obtient

$$\text{Var}[X] = \frac{n+1}{2} \cdot \frac{n-1}{6} = \frac{n^2-1}{12}.$$

On termine en mentionnant au passage la définition de l'*écart-type* :

DÉFINITION 4.8 (Écart-type). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire. L'**écart-type** de X (noté σ) est donné par

$$\sigma = \sqrt{\text{Var}[X]}.$$

4.6. Plusieurs distributions discrètes

Dans cette section, nous allons explorer plus amplement plusieurs distributions de variables aléatoires discrètes importantes. Toutefois, avant de commencer, ça vaut la peine de discuter brièvement de la notion de *distribution* d'une variable aléatoire un peu plus en détails.

DÉFINITION. Soient $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire.

La **distribution** de X fait référence à la mesure μ sur \mathbb{R} définie par

$$\mu(A) = \mathbb{P}\{X \in A\}$$

pour tous¹⁴ $A \subseteq \mathbb{R}$.

Cette définition peut paraître obscure et intractable – et pour cause. C'est très difficile de spécifier directement une mesure.

C'est pour cette raison qu'on se sert plutôt d'autre objets mathématiques qui *caractérisent* une distribution. Nous en avons déjà rencontrés plus tôt dans le chapitre 4 :

14. Pas « pour tous », en fait. On doit se restreindre à la tribu engendrée par les intervalles – qu'on appelle la tribu Borélienne. Tout ça est hors du cadre du cours.

- La fonction de répartition d'une variable aléatoire caractérise complètement sa distribution.
- La fonction de masse caractérise complètement la distribution d'une variable aléatoire discrète.

Ce qu'on veut dire, lorsqu'on dit qu'un objet *caractérise* une distribution, c'est que deux variables aléatoires qui ont cet objet en commun ont forcément la même distribution. Ici, par exemple, deux variables aléatoires qui ont la même fonction de répartition ont forcément la même distribution. De même, deux variables aléatoires discrètes qui ont la même fonction de masse ont aussi forcément la même distribution.

Donc, pour décrire une distribution, il suffit de préciser l'un de ces objets. Pour les variables aléatoires discrètes, la fonction de masse procure une intuition beaucoup plus forte quant à la distribution à laquelle elle est attachée. C'est pour cette raison que, dans ce qui suit, chaque fois que nous étudierons une nouvelle distribution, nous commencerons toujours par donner sa fonction de masse.

Les paramètres d'une distribution. On a déjà vu qu'on peut calculer les moments d'une variable aléatoire. Ces valeurs ne dépendent pas de la variable aléatoire elle-même – en effet,

$$\mathbb{E}[X^k] = \sum_{v \in V} v^k p_X(v),$$

et comme on le voit dans cette formule, ces valeurs seront identiques pour deux variables aléatoires ayant la même distribution, puisque celles-ci auront la même fonction de masse.

De telles quantités, qui ne dépendent que de la distribution, sont appelés des **paramètres** de la distribution. Nous en connaissons déjà une poignée :

- Les moments, les moments centrés ;
- Les valeurs du support ;
- Le mode (la valeur la plus probable) ;
- L'étendue (la distance entre la valeur maximale possible et la valeur minimale possible) ;
- La médiane ; les quantiles ;¹⁵
- etc.

Notre étude des différentes distributions se penchera donc chaque fois sur des paramètres pertinents.

Les « paramètres caractéristiques ». Nous verrons en fait plusieurs « types » de distributions différents, mais chaque fois, certains paramètres de la distribution pourront varier.

J'appelle ces paramètres « les paramètres caractéristiques » de la distribution.

Par exemple :

- Pour la distribution de Bernoulli (cannonique), le paramètre caractéristique est $p = p_X(1)$;
- pour la distribution binomiale, les paramètres caractéristiques sont n et p ;
- pour la distribution géométrique, le paramètre caractéristique est p ;
- pour la distribution de Poisson, le paramètre caractéristique est λ ;
- etc.

15. Bien que ces paramètres n'ont pas été définis dans le cours, vous en avez probablement déjà entendu parler...

4.6.1. La distribution de Bernoulli. C'est en quelque sorte la distribution la plus simple : on dit que X est une variable aléatoire de **distribution Bernoulli** si son support est $V = \{0, 1\}$ – c'est à dire que la variable X peut prendre deux valeurs.¹⁶

Ainsi, la distribution d'une variable de Bernoulli est entièrement caractérisé par un unique paramètre : la probabilité $p \in [0, 1]$ que $X = 1$. Sa fonction de masse est donc

$$p_X(0) = (1 - p), \quad p_X(1) = p.$$

Les variables aléatoires de Bernoulli sont souvent utilisées pour représenter le succès ou l'échec d'une « tentative » – plus souvent qu'autrement, elles apparaissent naturellement comme les variables aléatoires *indicatrices* d'événements d'intérêt.¹⁷ En fait, les variables Bernoulli peuvent toujours être érites comme des indicatrices pour un certain événement.

On a bien sûr que, pour une variable Bernoulli X , de paramètre p ,

$$(4.6.1) \quad \mathbb{E}[X] = p$$

$$(4.6.2) \quad \text{Var}[X] = p(1 - p).$$

Au chapitre 7, nous verrons comment il sera pratique d'utiliser les variables de Bernoulli (et spécifiquement les indicatrices) pour décortiquer l'analyse de problèmes plus complexes.

4.6.2. La distribution binomiale. Supposons maintenant qu'on réalise n tentatives indépendantes d'une expérience ayant une probabilité p de résulter en un « succès », et $1 - p$ d'« échouer ».

On voudrait compter X le nombre de succès obtenus en n tentatives. La distribution de X est appelée **distribution binomiale**.

On calcule

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1 - p)^{n-k}.$$

En effet, il existe $\binom{n}{k}$ façons distinctes de choisir lesquelles de nos n tentatives seront un succès, et pour chacune de ces façons, la probabilité d'obtenir le résultat sera $p^k(1 - p)^{n-k}$, puisque les tentatives sont indépendantes les unes des autres, et qu'on doit avoir k succès et $n - k$ échecs.

On a donc que la fonction de masse d'une variable aléatoire binomiale X est donnée par

$$(4.6.3) \quad p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

pour $k = 0, \dots, n$.

EXEMPLE 4.23. Lors d'un examen, un étudiant mal préparé décide qu'il répondra aléatoirement aux questions à choix multiples. Il y a 5 questions, toutes à 4 choix de réponse.

Quelle est la probabilité que l'étudiant obtienne la note de passage (au moins 3/5) ?

SOLUTION. Ici, supposons que X est le nombre de réponses correctes de l'étudiant. Alors, X est une variable de distribution Binomiale, puisqu'elle correspond au nombre de succès en $n = 5$ tentatives. La probabilité de succès correspond à la probabilité que l'étudiant tombe sur la bonne réponse à la question – c'est donc $p = 1/4$. On a donc $p_X(k) = \binom{5}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{5-k}$, pour $k = 0, 1, 2, 3, 4, 5$.

16. On voit parfois aussi des variables supportées sur $\{-1, 1\}$ appelées des « variables de Bernoulli ». Pour nous, ce sera toujours sur $\{0, 1\}$.

17. Encore une fois, consulter la section A.5.2 sur les fonctions indicatrices.

On cherche $\mathbb{P}\{X \geq 3\} = \mathbb{P}\{X = 3\} + \mathbb{P}\{X = 4\} + \mathbb{P}\{X = 5\}$. Calculons séparément :

$$\begin{aligned}\mathbb{P}\{X = 3\} &= \binom{5}{3} \left(\frac{1}{4}\right)^3 \left(\frac{3}{4}\right)^2 \\ &= 10 \cdot \frac{9}{1024} = \frac{90}{1024} \\ \mathbb{P}\{X = 4\} &= \binom{5}{4} \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right) \\ &= 5 \cdot \frac{3}{1024} = \frac{15}{1024} \\ \mathbb{P}\{X = 5\} &= \binom{5}{5} \left(\frac{1}{4}\right)^5 \\ &= \frac{1}{1024}.\end{aligned}$$

En tout, la somme donne

$$\mathbb{P}\{X \geq 3\} = \frac{106}{1024},$$

soit légèrement plus de 10%.

Les moments de la distribution binomiale. On peut obtenir les moments de la variable aléatoire binomiale en observant une récurrence.

En effet, en se servant de l'identité suivante (démontrée dans l'exercice 1.14) :

$$i \binom{n}{i} = n \binom{n-1}{i-1},$$

on voit que si X est une binomiale de paramètres n, p , son k ième moment est donné par

$$\begin{aligned}\mathbb{E}[X^k] &= \sum_{i=0}^n i^k \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i^{k-1} \left(i \binom{n}{i} \right) p^i (1-p)^{n-i} \\ &= np \sum_{i=1}^n i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-1-(i-1)} \\ &= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \\ &= np \mathbb{E}[(Y+1)^{k-1}],\end{aligned}\tag{4.6.4}$$

où Y est une variable aléatoire binomiale de paramètres $(n-1), p$.

En particulier, si X est une binomiale n, p ,

$$\mathbb{E}[X] = np \mathbb{E}[(Y+1)^0] = np,\tag{4.6.5}$$

puisque Y étant binomiale $(n-1), p$, $Y \geq 0$ et $Y+1 > 0$ et, par conséquent, $(Y+1)^0 = 1$.

On trouve aussi que

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

puisque par l'équation (4.6.5), $\mathbb{E}[Y+1] = \mathbb{E}[Y] + 1 = (n-1)p + 1$.

On conclue que, en vertu de la proposition 4.4,

$$(4.6.6) \quad \text{Var}[X] = n(n-1)p^2 + np - n^2p^2 = np(1-p).$$

L'unimodalité de la distribution binomiale. La distribution binomiale est « unimodale » – c'est à dire qu'elle n'a qu'un maximum global, et aucun extrema secondaire.

Pour le constater, il suffit d'abord de remarquer que (exercice 4.18) :

$$(4.6.7) \quad \frac{\mathbb{P}\{X=k\}}{\mathbb{P}\{X=k-1\}} = \frac{(n-k+1)p}{k(1-p)}.$$

Et bien entendu, ce ratio est supérieur à 1 tant que

$$(4.6.8) \quad \begin{aligned} (n-k+1)p &\geq k(1-p) \\ \Rightarrow (n+1)p &\geq k \end{aligned}$$

Inversement, lorsque $k > (n+1)p$, on a que $\mathbb{P}\{X=k\}/\mathbb{P}\{X=k-1\} < 1$.

On peut donc conclure que la fonction de masse $p_X(k)$ augmente en k (pour des bonds entiers, à partir de 0), jusqu'à environ $(n+1)p$, puis diminue ensuite.

REMARQUE. L'équation (4.6.7) donne une méthode plus facile pour calculer rapidement les valeurs pour la fonction de masse (ou la fonction de répartition) d'une variable aléatoire de distribution binomiale : par récurrence.

4.6.3. La distribution géométrique. Supposons que, lors d'une expérience aléatoire, une tentative de Bernoulli soit un succès avec probabilité p . Notre question maintenant, c'est : « Combien de fois faut-il réaliser l'expérience pour obtenir un succès ? »

Il est évident que la réponse à cette question doit être donnée sous la forme d'une variable aléatoire X – notre ensemble fondamental serait l'ensemble des suites infinies de résultats pour notre expérience aléatoire, et l'événement $\{X=k\}$ correspondrait alors à l'ensemble de toutes les suites où le premier succès apparaît à la k ième expérience. Évidemment, la variable X est supportée sur les nombres naturels : il faut toujours au moins une tentative pour obtenir un succès.

On aurait donc :

$$(4.6.9) \quad \mathbb{P}\{X=k\} = (1-p)^{k-1}p,$$

puisque les $k-1$ premières tentatives doivent toutes avoir résulté en un échec, et seulement la k ième doit avoir été un succès.

Une telle variable aléatoire est dite suivre une **distribution géométrique**. L'unique paramètre qui caractérise une distribution géométrique est donc la valeur p , la probabilité d'un succès.

REMARQUE. Il faut faire bien attention de distinguer entre X le nombre de tentatives nécessaires pour obtenir un succès, et le nombre $Y = X - 1$ de tentatives nécessaires *avant* d'obtenir un succès. Tout dépendant de l'ouvrage de référence, on verra l'un ou l'autre désigné comme ayant une distribution géométrique. C'est naturel, mais il faut simplement faire attention de bien saisir la distinction, et de vérifier que les raisonnements employés sont toujours cohérents.

EXEMPLE 4.24. Supposons qu'on joue à Pile ou Face avec une pièce de monnaie équilibrée.

Quelle est la probabilité qu'il faille lancer la pièce de monnaie au moins n fois avant qu'elle tombe sur Face ?

SOLUTION. On va définir X le nombre de fois qu'il faut lancer la pièce pour obtenir Face une première fois. Évidemment, X est une variable aléatoire de loi géométrique, avec probabilité de succès $p = \frac{1}{2}$. On cherche

$$\begin{aligned}\mathbb{P}\{X \geq n\} &= \sum_{k=n}^{\infty} \left(\frac{1}{2}\right)^{k-1} \frac{1}{2} \\ &= \left(\frac{1}{2}\right)^n \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} \\ &= \frac{1}{2^n} \cdot \left(\frac{1}{1 - \frac{1}{2}}\right) \\ &= \frac{1}{2^{n-1}}.\end{aligned}$$

Pour $n = 1$, évidemment, c'est 1. Pour $n = 2$, c'est $1/2$, ce qui est encore naturel ; en effet, la probabilité qu'il faille lancer la pièce de monnaie au moins deux fois avant de voir Face, est simplement la probabilité que le première résultat ait été Pile – c'est donc $1/2$.

On vérifie que la fonction de masse respecte la condition de normalisation :

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{P}\{X = k\} &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \\ &= p \sum_{k=1}^{\infty} (1-p)^{k-1} \\ &= \frac{p}{1 - (1-p)} \\ &= \frac{p}{p} = 1.\end{aligned}$$

Les moments de la distribution géométrique. Écrive directement, l'expression pour le n ième moment de la distribution géométrique est

$$(4.6.10) \quad \mathbb{E}[X^n] = \sum_{k=1}^{\infty} k^n (1-p)^{k-1} p$$

Cette somme peut sembler difficile à calculer, mais on peut s'en tirer assez facilement en remarquant une astuce. Pour $n = 1$, on choisit $q = (1-p)$, et on a :¹⁸

18. Certain.e.s d'entre vous pourraient avoir des doutes quant à la possibilité de dériver terme par terme une série infinie. C'est très justifié ! Mais comme vous le verrez dans le cours d'Analyse II, on peut le faire ici sans soucis.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} kq^{k-1}p \\
&= p \sum_{k=1}^{\infty} \left(\frac{d}{dq} q^k \right) \\
&= p \frac{d}{dq} \left(q \sum_{k=1}^{\infty} q^{k-1} \right) \\
&= p \frac{d}{dq} \left(\frac{q}{1-q} \right) \\
&= p \left(\frac{q}{(1-q)^2} + \frac{1}{1-q} \right) \\
&= p \left(\frac{1}{(1-q)^2} \right) \\
&= p \frac{1}{p^2} \\
&= \frac{1}{p}.
\end{aligned} \tag{4.6.11}$$

On aurait pu aussi voir ce résultat sans l'entourloupe de la différentiation par q , simplement comme suit :

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= 1 + \sum_{k=1}^{\infty} (k-1)(1-p)^{k-1}p \\
&= 1 + \sum_{k=1}^{\infty} k(1-p)^k p \\
&= 1 + (1-p) \sum_{k=1}^{\infty} k(1-p)^{k-1}p \\
&= 1 + (1-p)\mathbb{E}[X]
\end{aligned} \tag{4.6.12}$$

et en réarrangeant l'équation (4.6.12), on trouve

$$p\mathbb{E}[X] = 1,$$

ce qui correspond évidemment à (4.6.11).

Pour calculer les moments supérieurs, on peut encore une fois s'en tirer en se servant des propriétés de la fonction de masse « auto-similaire » de la distribution géométrique :

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{k=1}^{\infty} k^2(1-p)^{k-1}p \\
 &= \sum_{k=1}^{\infty} (k-1)^2(1-p)^{k-1}p + 2\sum_{k=1}^{\infty} (k-1)(1-p)^{k-1}p + \sum_{k=1}^{\infty} (1-p)^{k-1}p \\
 &= (1-p)\sum_{k=1}^{\infty} k^2(1-p)^{k-1}p + 2(1-p)\sum_{k=1}^{\infty} k(1-p)^{k-1}p + 1 \\
 (4.6.13) \quad &= (1-p)\mathbb{E}[X^2] + 2(1-p)\mathbb{E}[X] + 1.
 \end{aligned}$$

En réarrangeant (4.6.13), on obtient

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

ou encore, en remplaçant pour $\mathbb{E}[X]$ et en réarrangeant encore,

$$(4.6.14) \quad \mathbb{E}[X^2] = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2},$$

et la variance de la variable aléatoire géométrique est donnée par

$$(4.6.15) \quad \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Propriété d'absence de mémoire.

DÉFINITION 4.9 (Absence de mémoire). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire discrète entière non-négative.

On dit que X est **sans mémoire** si et seulement si on a que, pour tous $m, n \in \mathbb{N}$,

$$(4.6.16) \quad \mathbb{P}\{X > m+n \mid X > m\} = \mathbb{P}\{X > n\}.$$

L'intérêt d'une variable dite *sans mémoire*, c'est qu'elle « ne se souvient pas » qu'elle est plus grande qu'un certain seuil – c'est en quelque sorte la définition plus formelle de la propriété d'« auto-similarité » de la distribution qu'on avait déjà remarqué en calculant les moments pour la géométrique.

On appelle cette propriété *l'absence de mémoire* en référence au fait que c'est une propriété qu'on exige souvent de délais aléatoires.

EXEMPLE 4.25. Soit X le temps (en années) qu'il faut attendre entre l'explosion de deux géantes rouges en super-nova dans notre galaxie.

Comme on ne sait pas comment prédire ce phénomène, le temps qui s'est écoulé depuis la dernière super-nova ne peut pas influencer le temps qu'il reste à attendre avant la prochaine. autrement dit,

$$\mathbb{P}\{X - m > n \mid X > m\} = \mathbb{P}\{X > n\}$$

, c'est à dire que, sachant que la dernière super-nova a eu lieu il y a m années, ça ne change pas la probabilité que la prochaine super-nova se déroule dans n années – c'est la même chose que si la super-nova avait eu lieu aujourd'hui.

On a la proposition suivante :

PROPOSITION 4.5. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire entière positive. Alors, X est sans mémoire si et seulement si X suit une distribution géométrique.

DÉMONSTRATION. $\Rightarrow :$ On montre que X est sans mémoire si X est géométrique.

On calcule d'abord $\mathbb{P}\{X > k\}$, pour tout k :

$$\begin{aligned} \mathbb{P}\{X > k\} &= \sum_{i=k+1}^{\infty} (1-p)^{i-1} p \\ &= (1-p)^k \sum_{i=1}^{\infty} (1-p)^{i-1} p \\ (4.6.17) \quad &= (1-p)^k. \end{aligned}$$

Évidemment, il suit que pour tous $m, n \in \mathbb{N}$, on a

$$\begin{aligned} \mathbb{P}\{X > m+n \mid X > m\} &= \frac{\mathbb{P}\{X > m+n\}}{\mathbb{P}\{X > m\}} \\ &= \frac{(1-p)^{m+n}}{(1-p)^m} \\ &= (1-p)^n \\ &= \mathbb{P}\{X > n\} \end{aligned}$$

$\Leftarrow :$ On montre que si X est sans mémoire, alors X est géométrique.

Si X est sans mémoire, alors, pour tous $m, n \in \mathbb{N}$, on a que $\mathbb{P}\{X > m+n \mid X > m\} = \mathbb{P}\{X > n\}$, et avec $n = 1$, en particulier, on trouve que

$$\mathbb{P}\{X > m+1 \mid X > m\} = \frac{\mathbb{P}\{X > m+1\}}{\mathbb{P}\{X > m\}} = \mathbb{P}\{X > 1\}.$$

Si on note $C = \mathbb{P}\{X > 1\}$, alors $C \in [0, 1]$, et il suit que pour tout m ,

$$\mathbb{P}\{X > m\} = C^m,$$

et bien sûr, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}\{X = k\} = \mathbb{P}\{X > k-1\} - \mathbb{P}\{X > k\} = C^{k-1}(1-C).$$

On voit alors que X est une variable aléatoire géométrique, avec $p = 1 - C$. \square

Autrement dit, les variables aléatoires géométriques sont les seules variables aléatoires qui possèdent la propriété d'absence de mémoire.

4.6.4. La distribution de Poisson. Soit X une variable aléatoire discrète non-négative (supportée sur $\{0, 1, 2, \dots\}$) avec

$$(4.6.18) \quad \mathbb{P}\{X = k\} = e^{-\lambda} \frac{\lambda^k}{k!},$$

pour un certain $\lambda > 0$. Alors, on dit que X suit une **distribution de Poisson**.

Les variables qui suivent cette distribution surviennent naturellement dans plusieurs contextes. Le plus souvent, il s'agit de compter le nombre de fois qu'un événement s'est produit en une période de temps prolongée :

- Le nombre d'accidents sur une route en une journée ;
- le nombre de clics comptés par un compteur Geiger en une période de 1 minute ;

- le nombre d'appels téléphoniques reçus par une centrale en quatre heures ;
- etc.

Pour l'instant nous ne nous attarderons pas sur les raisons qui poussent de telles variables à assumer une telle distribution ; nous nous contenterons de l'équation (4.6.18).

On vérifie à l'aide de l'expansion en série de Taylor pour la fonction exponentielle, que

$$\sum_{k=0}^{\infty} \mathbb{P}\{X = k\} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1.$$

EXEMPLE 4.26. Le nombre de tempêtes de neige durant un hiver donné est une variable aléatoire de Poisson, avec $\lambda = 6$. Trouver la probabilité qu'il y ait au moins quatre tempêtes cet hiver, sachant qu'il y en a déjà eu deux.

SOLUTION. Si X est le nombre de tempêtes, on cherche alors

$$\mathbb{P}\{X \geq 4 \mid X \geq 2\} = \frac{\mathbb{P}\{X \geq 4\}}{\mathbb{P}\{X \geq 2\}},$$

puisque si $X \geq 4$, alors $X \geq 2$.

On trouve que

$$\mathbb{P}\{X \geq 2\} = 1 - \mathbb{P}\{X = 0\} - \mathbb{P}\{X = 1\} = 1 - e^{-6} - 6e^{-6} = 1 - 7e^{-6},$$

et que

$$\mathbb{P}\{X \geq 4\} = 1 - \sum_{k=0}^3 \mathbb{P}\{X = k\} = 1 - e^{-6} \left(1 + 6 + \frac{36}{2} + \frac{216}{6} \right) = 1 - 61e^{-6}.$$

Finalement,

$$\mathbb{P}\{X \geq 4 \mid X \geq 2\} = \frac{1 - 61e^{-6}}{1 - 7e^{-6}} \approx 86,4\%.$$

On voit que, comme le prédit la proposition 4.5, la distribution de Poisson (qui n'est pas une distribution géométrique) n'a pas la propriété d'absence de mémoire : si on sait qu'il y a déjà eu deux tempêtes, la probabilité qu'il y en ait 4 est plus probable que si on ne savait rien sur le nombre total de tempêtes.

Les moments de la distribution de Poisson. On peut calculer les quelques premiers moments de la distribution de Poisson assez facilement. Si X est une variable de distribution Poisson, avec paramètre $\lambda > 0$,

$$\begin{aligned}
 \mathbb{E}[X] &= e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \\
 &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
 &= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
 &= \lambda.
 \end{aligned} \tag{4.6.19}$$

Pour le second moment, c'est un peu plus compliqué. On y parvient néanmoins en appliquant le même genre de procédé :

$$\begin{aligned}
 \mathbb{E}[X^2] &= e^{-\lambda} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} \\
 &= e^{-\lambda} \sum_{k=1}^{\infty} k(k-1) \frac{\lambda^k}{k!} + e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \\
 &= \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \mathbb{E}[X] \\
 (4.6.20) \quad &= \lambda(\lambda + 1).
 \end{aligned}$$

Ainsi, la variance d'une variable de Poisson est donnée par

$$(4.6.21) \quad \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda.$$

Les poissons et les binomiales. Il existe un lien pratique entre les variables aléatoires binomiales et les variables aléatoires de Poisson : on peut approximer les binomiales par des variables de Poisson : lorsque n est grand et que p est relativement petit, on peut approximer une binomiale n, p par une poisson de paramètre $\lambda = np$.

Formellement, on a le résultat suivant :

PROPOSITION 4.6. *Soient X_n des variables aléatoires binomiales, de paramètres $n, p_n = \lambda/n$. Soit Y une variable aléatoire de Poisson.¹⁹*

Alors, pour tout $k \geq 0$, on a que

$$(4.6.22) \quad \lim_{n \rightarrow \infty} \mathbb{P}\{X_n = k\} = \mathbb{P}\{Y = k\}.$$

DÉMONSTRATION. On a que

$$\begin{aligned}
 \mathbb{P}\{X_n = k\} &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
 &= \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
 &= \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \cdot \left(\frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}\right).
 \end{aligned}$$

Et on prend le produit des limites :

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \mathbb{P}\{X_n = k\} &= \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} \cdot \lim_{n \rightarrow \infty} \left(\frac{n \cdot (n-1) \cdots (n-k+1)}{n^k}\right) \\
 &= \frac{\lambda^k}{k!} \cdot e^{-\lambda} \cdot 1 \\
 &= \mathbb{P}\{Y = k\}.
 \end{aligned}$$

□

L'intérêt d'un tel résultat, c'est qu'on peut l'employer pour faire des approximations de calculs rapides ; dans un sens ou dans l'autre, ça va assez vite – mais très souvent, on va se servir de la Poisson pour approximer les binomiales.

19. Nous verrons au chapitre 8 qu'un tel résultat est dit « de convergence en loi ».

EXEMPLE 4.27. Donner une borne approximative sur la probabilité qu'au moins une personne célèbre son anniversaire le 4 ou le 7 février.

SOLUTION. On sait que le nombre de personnes dans la classe qui célèbrent leur anniversaire le 4 ou le 7 février est une variable aléatoire binomiale de paramètre $N, p = 2/365$, où N est le nombre de personnes total dans la classe (on va faire les calculs avec $N = 130$).

On peut donc faire l'approximation que le nombre de personnes qui célèbrent leur anniversaire le 4 ou le 7 mars est une variable aléatoire de Poisson de paramètre $\lambda = 2N/365$.

La probabilité qu'il y a au moins une personne qui célèbre son anniversaire à l'une de ces deux dates est donc

$$1 - e^{-2N/365} \approx_{(N=130)} 50,95\%.$$

On pourrait également calculer cette valeur exactement (exercice !) : ce serait

$$1 - \frac{363^N}{365^N} \approx_{(N=130)} 51,05\%.$$

On remarque que le résultat est assez proche.

REMARQUE. Historiquement, l'approximation était très utile, parce que les nombres 363^N et 365^N sont énormes et peu pratiques à manipuler à la main. Par contraste, le nombre $2N/365$ ne pose pas de réel problème, et l'approximation de l'exponentielle pouvait se faire en référant à des tables pré-existantes. Aujourd'hui, Mathematica exécute le gros calcul en un clin d'œil, mais en 1950, on était très content.e.s d'avoir de telles approximations sous la main.

Les trois distributions qui suivent sont moins utiles que les quatre précédentes, mais nous les couvrirons brièvement quand même.

4.6.5. La distribution hypergéométrique. Supposons une expérience aléatoire où on retire, sans remise, n boules d'un sac contenant N boules en tout, dont M sont mauves.

On peut nommer X le nombre de boules mauves retirées. Alors, X est supportée sur $\{0, 1, 2, 3, \dots, \min\{n, M\}\}$. On aurait immédiatement que

$$(4.6.23) \quad \mathbb{P}\{X = m\} = \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}.$$

L'exemple le plus amusant d'utilisation de cette variable aléatoire est le suivant :

EXEMPLE 4.28. On cherche à comprendre la taille d'une population de lièvres sur un territoire. Pour parvenir à nos fins, on en attrape un nombre M , auxquels on pose une bague sur la patte.

Puis, la semaine suivante, on revient sur le territoire et on attrape n lièvres. De ces lièvres, m ont une bague à la patte. Sachant cela, peut-on estimer la taille de la population de lièvres ?

SOLUTION. Si N est la population totale de lièvres, et X est le nombre de lièvres recapturés qui ont une bague (parmi les n recapturés), alors

$$\mathbb{P}\{X = m \mid N = k\} = \frac{\binom{M}{m} \binom{k-M}{n-m}}{\binom{k}{n}} =: P_m(k).$$

Cette valeur est la probabilité d'avoir observé $X = m$ sachant qu'il y avait $N = k$ lièvres en tout. On peut alors estimer N en choisissant pour N la valeur k qui maximise cette expression – en statistiques, c'est ce qu'on appelle une *estimation par maximum de vraisemblance*.

On remarque d'abord que, en appliquant la définition du coefficient binomial,

$$\frac{P_m(k)}{P_m(k-1)} = \frac{(k-M)(k-n)}{k(k-M-n+m)}.$$

Ce ratio excède 1 lorsque $(k-M)(k-n) \geq k(k-M-n+m)$, soit lorsque

$$k \leq \frac{Mn}{m}.$$

Ainsi, lorsque k franchit le seuil de Mn/m , la valeur $P_m(k)$ commence à décroître.

Ainsi, le meilleur estimateur pour N est Mn/m où, on le rappelle, M est le nombre de lièvres initialement bagués, n est le nombre de lièvres attrapés la seconde fois et m est le nombre de lièvres bagués attrapés la seconde fois.

Les moments. On peut montrer (exercice 4.16) que les moments de la variable hypergéométrique satisfont :

$$(4.6.24) \quad \mathbb{E}[X^k] = \frac{nM}{N} \mathbb{E}[(Y+1)^{k-1}],$$

où X est une variable binomiale négative de paramètres N, M, n et Y est une binomiale négative de paramètres $N-1, M-1, n-1$.

Ainsi, on trouve que

$$(4.6.25) \quad \mathbb{E}[X] = \frac{nM}{N}.$$

Il suffit de rappliquer cette récurrence pour trouver que

$$(4.6.26) \quad \mathbb{E}[X^2] = \frac{nM}{N} \left(\frac{(n-1)(M-1)}{(N-1)} + 1 \right).$$

On a donc que la variance d'une variable hypergéométrique de paramètres N, M, n est donnée par

$$(4.6.27) \quad \text{Var}[X] = \frac{nM}{N} \left(\frac{(n-1)(M-1)}{N-1} + 1 - \frac{nM}{N} \right).$$

Pour y voir un peu plus clair, on peut noter $p = M/N$, la proportion des boules dans le sac qui sont mauves. Alors, l'espérance d'une variable hypergéométrique est simplement

$$\mathbb{E}[X] = np.$$

On constate que c'est égal à l'espérance d'une binomiale n, p – en effet, si on remettait les boules dans le sac, X serait une variable binomiale. Mais même sans remise, l'espérance ne change pas.

Toutefois, la variance d'une hypergéométrique est

$$\text{Var}[X] = np(1-p) \left(1 - \frac{n-1}{N-1} \right).$$

Ici, ça n'est plus identiquement égal à la variance d'une binomiale. En fait, lorsque n est négligeable devant N , on tend vers la variance d'une binomiale (puisque les tirages ont presque tous les mêmes probabilités, vu que la remise ne fait pas de grande différence).

Mais lorsque n et N sont proches, la variance diminue, vu que si on pique sans remise, on finira toujours par piquer toutes les boules lorsque n s'approche de N .

4.6.6. La variable binomiale négative. Supposons qu'on réalise des tentatives de Bernoulli indépendantes, qui résultent avec probabilité p en un succès. On définit X le nombre de tentatives qu'il faut pour obtenir r succès.

X est donc supportée sur $\{r, r+1, r+2, \dots\}$ – en effet, il va falloir au moins r tentatives pour obtenir r succès, dans le meilleur des scénarios où on réussit toujours.

On a que, pour tout $k \geq r$,

$$(4.6.28) \quad \mathbb{P}\{X = k\} = \binom{k-1}{r-1} p^r (1-p)^{k-r}.$$

Une variable aléatoire dont la fonction de masse est donnée par (4.6.28) est dite une variable aléatoire **binomiale négative** de paramètres r, p .

Le raisonnement justifiant cette fonction de masse était précisément l'objet de l'exercice 3.13.

Notez la ressemblance avec la fonction de masse pour les variables aléatoires binomiales, supportées sur $\{0, n\}$.

On remarque en outre que les variables binomiales négatives sont une généralisation des variables géométriques. En effet, pour $r = 1$, on retrouve exactement la fonction de masse d'une variable aléatoire géométrique.

4.6.6.1. Les moments de la binomiale négative. On peut montrer (exercice 4.17) que les moments de variables aléatoires binomiales négatives satisfont :

$$(4.6.29) \quad \mathbb{E}[X^k] = \frac{r}{p} \mathbb{E}[(Y-1)^{k-1}],$$

où X est une binomiale négative de paramètres r, p et Y est une binomiale négative de paramètres $r+1, p$.

En prenant $k = 1$, on trouve immédiatement

$$(4.6.30) \quad \mathbb{E}[X] = \frac{r}{p}.$$

En prenant $k = 2$, on obtient

$$(4.6.31) \quad \mathbb{E}[X^2] = \frac{r}{p} \left(\frac{r+1}{p} - 1 \right),$$

et la variance de X est donc donnée par

$$(4.6.32) \quad \text{Var}[X] = \frac{r}{p} \left(\frac{r+1}{p} - 1 \right) - \frac{r^2}{p^2} = \frac{r(1-p)}{p^2}.$$

4.6.7. La distribution Zeta. Soit X une variable aléatoire avec la fonction de masse donnée par

$$(4.6.33) \quad \mathbb{P}\{X = k\} = p_X(k) = \frac{1}{\zeta(1+\alpha)k^{1+\alpha}}.$$

Une telle variable aléatoire est dite suivre une distribution de Zeta (en l'honneur de la fonction ζ de Riemann, qui intervient dans sa définition). On dit parfois aussi « distribution de Zipf ».

On sait que la fonction ζ de Riemann satisfait

$$\zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s},$$

et bien sûr la condition de normalisation est vérifiée.

La distribution de Zipf est célèbre pour son apparition dans plusieurs situations apparemment non-reliées.

- La fréquence d'occurrence de mots dans des textes ;
- La taille de la population d'une ville ;
- Le nombre de foyers touchés par une panne d'électricité ;
- Le nombre d'ordinateurs infectés par un virus informatique ;
- etc.

Cette distribution émerge souvent dans des contextes où des systèmes dynamiques complexes sont à l'équilibre près d'un *point critique*²⁰.

On a bien sûr que

$$(4.6.34) \quad \mathbb{E}[X^\gamma] = \frac{1}{\zeta(1+\alpha)} \sum_{k=1}^{\infty} \frac{k^\gamma}{k^{1+\alpha}} = \frac{\zeta(1+\alpha-\gamma)}{\zeta(1+\alpha)},$$

ce qui signifie que $\mathbb{E}[X^\gamma]$ n'existe que lorsque $\gamma < \alpha$ – autrement, la série divergerait.

Ainsi, pour $\alpha < 2$, la variance de X est « infinie ». Pour $\alpha < 1$, l'espérance de X est infinie, et X est une variable à queue lourde.

Bon. Nous nous sommes familiarisé.e.s avec les variables aléatoires discrètes – c'est bien-tôt le temps de clore ce chapitre. Avant de faire cela, par contre, nous allons nous pencher sur une dernière notion importante.

4.7. Espérance de sommes de variables aléatoires

Le sujet des variables aléatoires simultanées sera abordé en beaucoup plus de détails à partir du chapitre 6 – toutefois, nous allons anticiper une partie de son contenu rapidement ici, afin de nous rendre la vie beaucoup plus simple et agréable.

La question sous considération est celle de l'espérance de sommes de plusieurs variables aléatoires.

En effet, supposons un espace de probabilités $(\Omega, \mathcal{E}, \mathbb{P})$ sur lequel on aurait deux variables aléatoires $X, Y : \Omega \rightarrow \mathbb{R}$.

La question c'est : que peut-on dire sur $\mathbb{E}[X + Y]$?

Si on assume que notre espace Ω est dénombrable ou fini, alors on peut utiliser l'équation (4.3.1), notre « proto-définition de l'espérance » pour répondre à la question :

²⁰ Les points critiques sont, dans l'espace des paramètres d'un système dynamique, les points où apparaissent les bifurcations (transitions de phase, etc.). Vous êtes excusé.e.s si ça ne vous dit absolument rien ; je ne suis moi-même pas du tout un expert.

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) \mathbb{P}\{\{\omega\}\} \\
&= \sum_{\omega \in \Omega} X(\omega) \mathbb{P}\{\{\omega\}\} + \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}\{\{\omega\}\} \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

Ce résultat demeure vrai pour n'importe quel espace fondamental, et pour n'importe quelles deux variables aléatoires. En fait, on a le résultat suivant :

PROPOSITION 4.7. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et soient $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires. Soit a une constante réelle.*

Alors, on a toujours

$$(4.7.1) \quad \mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y].$$

La démonstration de cette proposition dans toute sa généralité dépasse le cadre de ce cours, bien que nous reviendrons sur la définition exacte de l'espérance au chapitre 7. Toutefois, nous la voyons ici car cette proposition nous sera singulièrement bienfaisante.

EXEMPLE 4.29. Si on réalise n expériences indépendantes, que E_i est l'événement que la i ème expérience réussit, et que $\mathbb{P}\{E_i\} = p$ pour tout i , alors le nombre X de succès est donné par

$$X = \sum_{k=1}^n \mathbb{1}_{E_i},$$

puisque si on additionne 1 à chaque succès, on finit forcément avec le nombre total de succès en n tentatives.

X est une variable aléatoire binomiale de paramètres n, p . On a calculé son espérance en employant une méthode de récurrence sur les moments, et c'était bien joli, mais on peut faire plus facile.

En utilisant la linéarité de l'espérance, on voit que

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[\mathbb{1}_{E_i}].$$

Mais on se souvient que l'espérance d'une fonction indicatrice est donnée par la probabilité de l'événement dans l'indicatrice :

$$\mathbb{E}[\mathbb{1}_{E_i}] = \mathbb{P}\{E_i\} = p,$$

pour tout i . Donc, on a finalement

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{P}\{E_i\} = np.$$

REMARQUE. Les E_i n'ont pas à être indépendants pour que le résultat soit le même ! En effet, si on s'imagine qu'on pige n boules d'un sac qui contient N boules dont M sont mauves, et que E_i est l'événement : « la i ème boule est mauve », alors on a toujours que

$$X = \sum_{i=1}^n \mathbb{1}_{E_i},$$

et donc $\mathbb{E}[X] = np$. C'est ce qui justifie que l'espérance d'une hypergéométrique et d'une binomiale sont identiques.

EXEMPLE 4.30. Dans la même veine, si on réalise des tentatives de Nernoulli indépendantes avec probabilité de succès p , et que X est le nombre de tentatives nécessaires pour obtenir r succès, alors bien sûr, on peut définir T_1 le nombre de tentatives nécessaires pour obtenir le premier succès, puis T_2 le nombre de tentatives nécessaires pour obtenir le second succès après qu'on ait eu le premier. Puis T_3, T_4 , etc. jusqu'à T_r .

Pour tout i , T_i est une variable aléatoire géométrique de paramètre p , puisque c'est le nombre de tentatives nécessaires pour obtenir un succès. On a donc $\mathbb{E}[T_i] = 1/p$ pour tout i .

Mais $X = \sum_{i=1}^r T_i$, puisque X est le nombre de tentatives nécessaires pour obtenir r succès ! Donc,

$$\mathbb{E}[X] = \sum_{i=1}^r \mathbb{E}[T_i] = r/p.$$

Ouf ! C'était beaucoup de choses. Au chapitre 5, nous poursuivons notre étude des variables aléatoires, en parlant cette fois de variables aléatoires dites « continues ».

4.8. Exercices

EXERCICE 4.1. Une pièce de monnaie retombe sur « pile » avec probabilité p . On lance la pièce jusqu'à ce que soit « pile », soit « face » soit apparu deux fois (pas forcément consécutives).

Calculer l'espérance du nombre de lancers nécessaires.

EXERCICE 4.2. Il y a deux pièces de monnaie dans une boîte. L'une tombe sur « pile » avec probabilité $6/10$. L'autre tombe sur « pile » avec probabilité $3/10$. On sélectionne équiprobablement l'une des pièces. Sans savoir quelle pièce sera choisie, on pariera un montant d'argent d'au plus $10\$$. On remporte le double de notre mise si la pièce tombe sur « face », mais on perd notre mise si la pièce tombe sur « pile ».

L'organisateur du jeu est cependant prêt à nous vendre l'information quant à la pièce choisie, pour un montant de $C\$$. Quelle est l'espérance du gain si on choisit d'acheter l'information ? Pour quelles valeurs de C vaut-il mieux acheter l'information ?

EXERCICE 4.3. Pour $n \geq 1$, $p \in [0, 1]$, on dira que $B(n, p)$ est une variable aléatoire binomiale de paramètres n, p .

Argumenter que

$$\mathbb{P}\{B(n, p) \leq k\} = 1 - \mathbb{P}\{B(n, 1-p) \leq n-k-1\}.$$

Indice : L'affirmation que le nombre de succès en n tentatives est inférieur à k est équivalente à quelle affirmation sur le nombre d'échecs ?

EXERCICE 4.4. (a) Un entier N est choisi uniformément, au hasard, parmi les 10^3 premiers entiers naturels. Quelle est la probabilité que N soit divisible par 3 ? 5 ? 7 ? 15 ? 105 ?

Donner la limite de ces probabilités lorsque N est choisi équiprobablement parmi les 10^k premiers entiers, pour k tendant vers l'infini.

(b) La fonction de Möbius est une fonction étudiée principalement en théorie analytique des nombres. Elle est liée à l'un des problèmes ouverts les plus connus en mathématiques ; l'hypothèse de Riemann. On la définit comme ceci :

$$\mu(n) = \begin{cases} 0 & \text{si l'un des facteurs premiers de } n \text{ est dégénéré} \\ -1 & \text{si } n \text{ a un nombre impair de facteurs premiers} \\ 1 & \text{si } n \text{ a un nombre pair de facteurs premiers.} \end{cases}$$

On dit qu'un facteur premier de n est *dégénéré* si celui-ci apparaît de multiples fois dans la factorisation de n en facteurs premiers. Par exemple, $10 = 2 \times 5$ n'a pas de facteur premiers dégénérés, mais $12 = 2 \times 2 \times 3$ en a un. On aurait en fait $\mu(10) = 1$ et $\mu(12) = 0$.

Si on choisit N_k uniformément parmi les 10^k premiers entiers, estimer la valeur de

$$\lim_{k \rightarrow \infty} \mathbb{P}\{\mu(N_k) = 0\}.$$

Indice : utiliser que

$$\zeta(2) = \frac{\pi^2}{6},$$

où $\zeta(s)$ est la fonction de Riemann, définie pour les nombres complexes par

$$\zeta(s) = \prod_{i=1}^{\infty} \frac{P_i^s}{P_i^s - 1},$$

avec P_i le i ème nombre premier.

EXERCICE 4.5. Quatre autous transportent 148 étudiant.e.s vers la cabane-à-sucre. Ils transportent respectivement 40, 25, 33 et 50 étudiant.e.s.

On choisit l'un.e des étudiant.e.s équiprobablement, et on dénote par X le nombre de personnes qu'il y a dans son autobus. Puis on choisit un.e chauffeur.e de façon équiprobable et on dénote par Y le nombre de personnes dans son autobus.

- (a) Sans faire de calcul, laquelle de $\mathbb{E}[X]$ et $\mathbb{E}[Y]$ est la plus grande ?
- (b) Calculer et comparer.

EXERCICE 4.6. À une kermesse, on vous propose un jeu simple : on place une pièce de 1\$ sur la table – c'est le montant en jeu. Puis, on tire à pile ou face. Si la pièce fait « face », le montant en jeu est doublé. Si la pièce fait « pile », vous gagnez immédiatement le montant en jeu.

- (a) Calculer l'espérance du gain net si vous devez d'abord débourser un montant de $C\$$ pour jouer.
- (b) Seriez-vous prêt.e.s à débourser $1M\$$ pour jouer une fois ?
- (c) Seriez-vous prêt.e.s à vous endetter de $1M\$$ par partie pour jouer autant que vous voulez ?

REMARQUE. Ce problème célèbre s'appelle le paradoxe de St-Pétersbourg. D'ici la fin du cours, nous aurons vu suffisamment de matière pour prouver rigoureusement que, peu importe le coût par partie, on peut toujours faire un gain net positif en un nombre de parties d'espérance finie.

EXERCICE 4.7. Si $\mathbb{E}[X] = 1$ et $\text{Var}[X] = 1/2$, trouver

- (a) $\mathbb{E}[(2 + X)^2]$;
- (b) $\text{Var}[4 + 3X]$.

EXERCICE 4.8. On suppose que pour condamner un.e accusé.e, il faut le vote de 9 des 12 membres d'un jury. La probabilité qu'un.e juré.e croie à l'innocence d'un.e coupable est de 20%²¹, tandis que la probabilité qu'un.e juré.e croie à la culpabilité d'un.e innocent.e est de 10%.

Si chaque juré.e est indépendant.e et la probabilité qu'un.e accusé.e soit effectivement coupable est de 65%, quelle est la probabilité que le jury rende la bonne décision ?

EXERCICE 4.9. Supposons que le nombre d'accidents qui surviennent sur une autoroute un jour donné est une variable aléatoire X de Poisson, avec paramètre $\lambda = 3$.

- (a) Quelle est la probabilité que plus de trois accidents se produisent aujourd'hui ?
- (b) Et si on sait qu'un accident s'est déjà produit ?

21. Les statistiques données dans ce problème sont factices.

EXERCICE 4.10. La probabilité de recevoir une main pleine (*full-house*) au poker est d'environ 14/10000. Trouver une approximation pour la probabilité de recevoir 2 fois une main-pleine en 1000 rondes de poker.

EXERCICE 4.11. Si X a une fonction de répartition $F(x) = \mathbb{P}\{X \leq x\}$, quelle est la fonction de répartition pour la variable e^X ?

EXERCICE 4.12. Si X est une variable aléatoire avec $\mathbb{P}\{X = 1\} = p = 1 - \mathbb{P}\{X = -1\}$, trouver une constante c non-triviale telle que $\mathbb{E}[c^X] = 1$.

EXERCICE 4.13. Soit X une variable aléatoire binomiale de paramètres n, p . Montrer que

$$\mathbb{E}\left[\frac{1}{X+1}\right] = \frac{1 - (1-p)^{n+1}}{(n+1)p}.$$

EXERCICE 4.14. (a) Soit X une variable aléatoire binomiale (n, p) . Montrer que

$$\mathbb{P}\{X \text{ est pair}\} = \frac{1 + (2p - 1)^n}{2}.$$

(Voir l'exercice 3.14).

(b) Utiliser le résultat précédent pour montrer que si Y est une variable aléatoire de loi Poisson avec paramètre $\lambda > 0$, alors

$$\mathbb{P}\{Y \text{ est pair}\} = \frac{1 + e^{-2\lambda}}{2}.$$

(c) Vérifier cette formule directement en utilisant le développement en série de Taylor pour le cosinus hyperbolique.

EXERCICE 4.15. Soit X une variable aléatoire de loi Poisson. Quel paramètre λ maximise $\mathbb{P}\{X = k\}$?

EXERCICE 4.16. À l'aide des identités suivantes :

$$k \binom{M}{k} = M \binom{M-1}{k-1}, \quad n \binom{N}{n} = N \binom{N-1}{n-1},$$

montrées à l'exercice 1.14, montrer l'équation (4.6.24) :

$$(4.6.24) \quad \mathbb{E}[X^k] = \frac{nM}{N} \mathbb{E}[(Y+1)^{k-1}],$$

où X est une variable hypergéométrique de paramètre N, M, n et Y est une variable hypergéométrique de paramètres $N-1, M-1, n-1$.

EXERCICE 4.17. À l'aide de l'identité

$$k \binom{k-1}{r-1} = r \binom{k}{r},$$

montrée à l'exercice 1.14, montrer l'équation (4.6.29) :

$$(4.6.29) \quad \mathbb{E}[X^k] = \frac{r}{p} \mathbb{E}[(Y-1)^{k-1}],$$

où X est une binomiale négative de paramètres r, p et Y est une binomiale négative de paramètres $r+1, p$.

EXERCICE 4.18. Montrer l'équation (4.6.7) pour la fonction de masse d'une variable binomiale :

$$(4.6.7) \quad \frac{\mathbb{P}\{X = k\}}{\mathbb{P}\{X = k - 1\}} = \frac{(n - k + 1)p}{k(1 - p)}$$

EXERCICE 4.19. Montrer la proposition suivante :

PROPOSITION. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire.

La fonction d'ensembles μ définie par $\mu_X(A) := \mathbb{P}\{X \in A\}$ est une mesure de probabilités sur \mathbb{R} .²²

REMARQUE. Cet objet μ_X est l'objet fondamental qu'on considère lorsqu'on parle de la *distribution* de X . Heureusement, pour les variables aléatoires réelles, cet objet est entièrement décrit par la fonction de répartition ; donc, pas trop de soucis.

²². Le domaine de μ_X n'est pas très clair – typiquement, on considérerait l'ensemble des Boréliens, soit la plus petite tribu mesurable qui contient tous les ouverts de \mathbb{R} . C'est de l'analyse beaucoup plus avancée – sentez-vous libres de faire vos propres recherches si ça vous intéresse !

Chapitre 5

Les variables aléatoires continues

Dans le chapitre 4, on s'est intéressé aux variables aléatoires discrètes. Une telle variable pouvait prendre des valeurs dans un ensemble V , qu'on appelait le **support** de la variable, et V était toujours un ensemble fini ou dénombrable.

Dans ce chapitre-ci, on étendra notre étude aux variables aléatoires qui peuvent prendre une infinité dénombrable de valeurs.

Pour ces variables aléatoires, le **support** (l'ensemble des valeurs qu'elles peuvent prendre) ne sera plus forcément dénombrable. En fait, la plupart du temps, ce sera un intervalle dans les nombres réels – ou carrément l'ensemble des nombres réels.

Par exemple,

- Le temps qu'il reste avant l'arrivée d'un autobus ;
- La température qu'il fera demain ;
- La distance entre l'incendie et la caserne de pompiers
- etc.

sont toutes des variables aléatoires qui peuvent prendre n'importe quelles valeurs dans un intervalle de nombres réels.

5.1. Fonctions de répartition et densité

La fonction de répartition $F(x)$, introduite au chapitre précédent (définition 4.2), peut nous aider à régler ce problème :

$$\mathbb{P}\{a < X \leq b\} = F(b) - F(a).$$

Ça, c'était déjà vrai au chapitre précédent. La différence, maintenant, c'est qu'on va considérer que, au chapitre 4, on ne considérait que les variables aléatoires qui avaient une fonction de répartition F dite « en escaliers », ou « constante par morceaux ».

Maintenant, on ne fera plus cette hypothèse. La fonction de répartition peut donc, *a priori*, être absolument n'importe quoi – tant qu'elle respecte les axiomes des probabilités (et donc qu'elle satisfait la proposition 4.1).

5.1.1. Les atomes.

DÉFINITION 5.1 (Atome). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire quelconque avec fonction de répartition F et fonction de masse de probabilité p_X . Soit $x \in \mathbb{R}$.

On dit que x est un **atome** de la distribution de X si la fonction de masse $p_X(x) = F(x) - F(x^-) > 0$.

Autrement dit, les atomes d'une distribution sont les points qui ont, à eux seuls, une probabilité positive. Ils correspondent exactement aux points où la fonction de répartition F est discontinue.

Les variables aléatoires discrètes sont un cas particulier de variable aléatoire où toute la distribution de probabilité est sur des atomes. Mais ça n'est pas forcément le cas.

Dans ce qui suit, nous nous concentrerons exclusivement sur les distributions qui n'ont pas d'atomes – c'est à dire des distributions pour lesquelles la fonction de répartition F est continue partout.

5.1.2. La densité. On remarque que, lorsque la fonction de répartition F d'une variable aléatoire X est continue en x , cela signifie que

$$\mathbb{P}\{X = x\} = p_X(x) = F(x) - F(x^-) = 0,$$

puisque à cet endroit, $F(x^-) = \lim_{y \rightarrow x^-} F(y) = F(x)$.

A priori, ça peut avoir l'air décourageant – on se dit, « mais si toutes les valeurs ont probabilité 0, toutes les valeurs sont équiprobables ! »

Et à strictement parler, c'est vrai, mais ça gâche un peu notre intuition. Par exemple, on s'attendrait à ce que la probabilité que le prochain bus passe dans 6 minutes soit « plus grande » que la probabilité qu'il passe dans 3 jours.

Ça nous prend donc un autre outil pour se faire une bonne idée.

Soit X une variable aléatoire et soit x un point dans \mathbb{R} .

Notre idée, c'est que pour de très petits intervalles $(x-h, x+h]$ autour de x , la probabilité que x se retrouve dans un tel intervalle devrait être à peu près proportionnel à la largeur de l'intervalle. On peut donc considérer

$$f(x) = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}\{x-h < X \leq x+h\}}{2h} = \lim_{h \rightarrow 0^+} \frac{F(x+h) - F(x-h)}{2h}.$$

Bien sûr, lorsque cette limite existe, $f(x) = F'(x)$.

On fait la définition suivante :

DÉFINITION 5.2 (Densité). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire avec fonction de répartition F quelconque.

Alors, pour tout $x \in \mathbb{R}$ tel que F est différentiable, on définit la **densité** de X en x par

$$(5.1.1) \quad f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{\mathbb{P}\{x < X \leq x+h\}}{h}.$$

EXEMPLE 5.1. Soit X une variable aléatoire avec fonction de répartition

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1. \end{cases}$$

Alors, la densité de X sur l'intervalle $(0, 1)$ est 1. C'est 0 partout ailleurs, sauf en 0 et en 1 où elle n'est pas définie.

REMARQUE. Dans le précédent exemple, on voit bien que la densité d'une variable aléatoire n'est pas toujours définie partout. Il existe un théorème qui dit que, puisque F est non-décroissante, dans n'importe quel intervalle il existe au plus un nombre fini de points où $f = F'$ n'est pas définie. Donc, pas de soucis.

EXEMPLE 5.2. Soit X une variable aléatoire avec densité $f(x) = \lambda e^{-\lambda x}$ pour $x > 0$, et 0 partout ailleurs.

Alors, la fonction de répartition de X est

$$\mathbb{P}\{X \leq x\} = \int_0^x f(t)dt = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0. \end{cases}$$

On voit que $\mathbb{P}\{X \leq x\}$ tend vers 1 à mesure que x tend vers l'infini.

Intuitivement, la densité d'une variable aléatoire continue, c'est l'équivalent de sa fonction de masse – ça nous donne directement une idée de quelles valeurs sont les plus probables. C'est pour cette raison que, malgré qu'elle ne soit pas toujours aussi bien définie que la fonction de répartition, c'est plus souvent la densité de probabilités qu'on utilisera pour comprendre rapidement l'allure d'une distribution pour une variable aléatoire continue.

Évidemment, comme nous l'avons vu dans le dernier exemple, pour une variable aléatoire X continue (c'est à dire dont la fonction de répartition F est continue en tout point), on a bien sûr que¹

$$(5.1.2) \quad F(x) = \mathbb{P}\{X \leq x\} = \int_{-\infty}^x f(t)dt.$$

Évidemment, si X est une variable continue à valeurs dans \mathbb{R} , alors

$$(5.1.3) \quad 1 = \mathbb{P}\{X \in \mathbb{R}\} = \lim_{x \rightarrow \infty} F(x) = \int_{-\infty}^{\infty} f(t)dt.$$

C'est l'équivalent de la condition de normalisation pour des variables aléatoires continues.

On a bien entendu que, si X est une variable aléatoire continue, que f est sa densité et que $[a, b]$ est un intervalle,

$$(5.1.4) \quad \mathbb{P}\{X \in [a, b]\} = \mathbb{P}\{X \in (a, b)\} = \int_a^b f(t)dt.$$

REMARQUE. Évidemment, en principe, par le théorème fondamental du calcul,

$$\int_a^b f(t)dt = F(b) - F(a) = \mathbb{P}\{X \in (a, b]\}.$$

Cependant, bien entendu, puisque $\mathbb{P}\{X = a\} = \mathbb{P}\{X = b\} = 0$, vu que X est une variable continue, on a que

$$\mathbb{P}\{X \in (a, b)\} = \mathbb{P}\{X \in (a, b]\} = \mathbb{P}\{X \in [a, b]\}.$$

Par conséquent, *seulement dans le cas où* X est une variable aléatoire *continue*, on aura

$$\mathbb{P}\{X \leq a\} = \mathbb{P}\{X < a\}, \quad \mathbb{P}\{X \geq a\} = \mathbb{P}\{X > a\}.$$

Ainsi, on pourra se soucier moins rigoureusement du caractère strict ou pas des inégalités dans ce contexte.

EXEMPLE 5.3. Soit X une variable aléatoire continue avec la fonction de densité

$$f(x) = \begin{cases} cx(1-x) & \text{si } x \in (0, 1) \\ 0 & \text{sinon.} \end{cases}$$

1. L'équation (5.1.2) tient toujours, car, comme mentionné dans une remarque précédente, si F est continue en tout point et que F est non-décroissante partout, alors pour n'importe quel intervalle borné, F est différentiable partout sauf en un nombre fini de points, et on peut par conséquent intégrer quand même en négligeant les points où $f = F'$ n'est pas défini.

- (a) Trouver c .
- (b) Trouver la fonction de répartition de X .
- (c) Trouver la probabilité que $X \geq \frac{3}{4}$.

SOLUTION. (a) Ici, la forme de la fonction de densité est définie, mais on cherche la valeur de la constante. Pour la trouver, il suffit d'utiliser la condition de normalisation :

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(t)dt \\ &= c \int_0^1 t(1-t)dt \\ &= c \left[\frac{t^2}{2} - \frac{t^3}{3} \right]_0^1 \\ &= c \left(\frac{1}{2} - \frac{1}{3} \right) \\ &= \frac{c}{6}. \end{aligned}$$

Il convient donc de conclure que $c = 6$.

- (b) Pour trouver la fonction de répartition, il suffit de réaliser l'intégrale, carrément :

$$F(x) = \int_{-\infty}^x f(t)dt.$$

Pour $x < 0$, c'est clair, $F(x) = 0$. Pour $x > 1$, $F(x) = 1$; en effet, $\int_0^1 f(t)dt = 1$, vu que c'est sur cet intervalle que f est non-nulle.

Supposons maintenant que $x \in [0, 1]$.

Alors,

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t)dt \\ &= 6 \int_0^x t(1-t)dt \\ &= 6 \left(\frac{x^2}{2} - \frac{x^3}{3} \right) \\ &= 3x^2 - 2x^3. \end{aligned}$$

- (c) Pour trouver la probabilité que $X \geq \frac{3}{4}$, on calcule :

$$\mathbb{P}\left\{X \leq \frac{3}{4}\right\} = F\left(\frac{3}{4}\right) = 3 \cdot \frac{9}{16} - 2 \cdot \frac{27}{64} = \frac{27}{32}.$$

Bein sûr, $\mathbb{P}\left\{X > \frac{3}{4}\right\} = \frac{5}{32}$, mais puisque X est une variable aléatoire continue,

$$\mathbb{P}\left\{X > \frac{3}{4}\right\} = \mathbb{P}\left\{X \geq \frac{3}{4}\right\} = \frac{5}{32}.$$

EXEMPLE 5.4. Si X est une variable aléatoire continue avec fonction de répartition F_X et fonction de densité f_X ,

- (a) Trouver la fonction de répartition et la densité de $Y = X + a$, où $a \in \mathbb{R}$.

- (b) Trouver la fonction de répartition et la densité de $Y = cX$, où $c > 0$.
- (c) Trouver la fonction de répartition et la densité de $Y = -X$.
- (d) Trouver la fonction de répartition et la densité de $Y = e^X$.

SOLUTION. (a) La fonction de répartition de $Y = X + a$ est

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{X + a \leq y\} = \mathbb{P}\{X \leq y - a\} = F_X(y - a).$$

Ainsi, par la règle de dérivation en chaîne, la fonction de densité de $Y = X + a$ est

$$f_Y(y) = F'_Y(y) = F'_X(y - a) = f_X(y - a).$$

- (b) La fonction de répartition de $Y = cX$ est

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{cX \leq y\} = \mathbb{P}\left\{X \leq \frac{y}{c}\right\} = F_X\left(\frac{y}{c}\right).$$

Ainsi, par la règle de dérivation en chaîne, la fonction de densité de $Y = cX$ est

$$f_Y(y) = F'_Y(y) = \frac{1}{c}F'_X\left(\frac{y}{c}\right) = \frac{1}{c}f_X\left(\frac{y}{c}\right).$$

- (c) La fonction de répartition de $Y = -X$ est

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{-X \leq y\} = \mathbb{P}\{X \geq -y\} = 1 - F_X(-y).$$

Ainsi, par la règle de dérivation en chaîne, la fonction de densité de $Y = -X$ est

$$f_Y(y) = F'_Y(y) = F'_X(-y) = f_X(-y).$$

- (d) La fonction de répartition de $Y = e^X$ est

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{e^X \leq y\} = \mathbb{P}\{X \leq \log y\} = F_X(\log y),$$

pour $y > 0$ – clairement $e^X > 0$ donc pour $y \leq 0$, $F_Y(y) = 0$.

Ainsi, par la règle de dérivation en chaîne, pour $y > 0$, on a que

$$f_Y(y) = F'_Y(y) = \frac{1}{y}F'_X(\log y) = \frac{1}{y}f_X(\log y).$$

5.1.3. Le support d'une variable aléatoire continue. Au chapitre 4, on avait défini le support de X comme simplement « l'image de Ω par X », soit l'ensemble des valeurs que peut prendre notre variable aléatoire. À toutes fins pratiques, on aurait pu rajouter, « toutes les valeurs que notre variable peut prendre avec probabilité supérieure à 0. »

Encore une fois, dans le cas continu, la généralisation de la notion de « support » n'est pas directe – puisque pour tout $x \in \mathbb{R}$, $\mathbb{P}\{X = x\} = 0$, on ne peut pas définir le support aussi simplement.

DÉFINITION 5.3 (Support (v.a. continues)²). Pour une variable aléatoire $X : \Omega \rightarrow \mathbb{R}$, avec densité f , on appellera **support** l'ensemble V défini par :

$$V := \{x \in \mathbb{R} : f(x) > 0\}.$$

2. En général, dans un espace topologique mesuré, le **support** de la mesure est l'intersection de tous les ouverts qui ont une mesure strictement positive. Ce que l'on appelle le *support d'une variable aléatoire* X n'est en fait que le support de la mesure $\mu_X(A) := \mathbb{P}\{X \in A\}$ définie à l'exercice 4.19.

5.2. Espérance et Variance

Au chapitre 4, on a défini l'espérance pour des variables aléatoires définies sur un espace fondamental au plus dénombrable. On avait également vu que pour X une variable discrète de support $V = \{v_1, v_2, v_3, \dots\}$, on avait

$$\mathbb{E}[X] = \sum_{v \in V} vp_X(v).$$

On s'intéresse maintenant à des variables aléatoires dont le support est non-dénombrable ; en particulier, on s'intéresse à des variables aléatoires dont la fonction de répartition est continue, et on aimerait étendre la notion d'espérance à de telles variables.

Nous ne ferons pas la preuve rigoureusement, mais on va se donner une idée de comment ça fonctionne.

Supposons que $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire continue avec fonction de répartition F et densité f . Ce qu'on va faire, c'est approximer X par X_n , qui sera une variable aléatoire discrète. Voici comment on procède :

On divise \mathbb{R} en petits intervalles de longueur $\frac{1}{n}$. Et puis, pour tout $k \in \mathbb{Z}$, on va avoir

$$\mathbb{P}\left\{X_n = \frac{k}{n}\right\} := \mathbb{P}\left\{\frac{k-1}{n} < X \leq \frac{k}{n}\right\} = F\left(\frac{k}{n}\right) - F\left(\frac{k-1}{n}\right) \approx \frac{1}{n}f\left(\frac{k}{n}\right)$$

où la dernière égalité (approximative) est lorsque n est très grand.

Mais alors,

$$\mathbb{E}[X_n] = \frac{1}{n} \sum_{k \in \mathbb{Z}} \frac{k}{n} f\left(\frac{k}{n}\right).$$

L'idée, c'est qu'on va obtenir $\mathbb{E}[X]$ en prenant la limite lorsque n tend vers l'infini – et la série va devenir une intégrale.

Finalement, pour X une variable aléatoire continue avec densité f , on a donc³

$$(5.2.1) \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

EXEMPLE 5.5. Soit $n \in \mathbb{N}$. Trouver $\mathbb{E}[X]$ lorsque X est une variable aléatoire continue avec densité

$$f(x) = \begin{cases} nx^{n-1} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{ailleurs.} \end{cases}$$

SOLUTION. Par l'équation (5.2.1), on a que

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^1 x \cdot nx^{n-1} dx \\ &= n \int_0^1 x^n dx \\ &= \frac{n}{n+1}. \end{aligned}$$

3. Encore une fois, il ne s'agit pas d'une définition mais d'une *conséquence* de la définition de l'espérance, définition que nous n'avons pas vue parce qu'elle implique des notions de la théorie de l'intégration de Lebesgue qui dépassent le cadre du présent cours.

On va maintenant montrer une proposition d'une très grande utilité :

PROPOSITION 5.1. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire continue non-négative avec fonction de répartition F et fonction de répartition complémentaire $\bar{F} = 1 - F$.*

*Alors,*⁴

$$(5.2.2) \quad \mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\} dt = \int_0^\infty \bar{F}(t) dt.$$

DÉMONSTRATION. Pour prouver ce résultat, on utilise la notion que, puisque X est positive, si $f = F'$ est la densité de X , alors

$$\mathbb{P}\{X > t\} = \bar{F}(t) = \int_t^\infty f(u) du.$$

Et si on fait l'intégrale, on trouve, en inversant l'ordre d'intégration, que

$$\begin{aligned} \int_0^\infty \bar{F}(t) dt &= \int_0^\infty \left(\int_t^\infty f(u) du \right) dt \\ &= \int_0^\infty \int_0^u f(u) dt du \\ &= \int_0^\infty \left(\int_0^u dt \right) f(u) du \\ &= \int_0^\infty u f(u) du \\ &= \mathbb{E}[X]. \end{aligned}$$

□

EXEMPLE 5.6. Soit X une variable aléatoire avec densité

$$f(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \\ 0 & \text{ailleurs} \end{cases} = \mathbb{1}_{[0,1]}(x).$$

Trouver $\mathbb{E}[-\log X]$.

SOLUTION. On doit d'abord trouver la densité de $Y = -\log X$.

On commence par trouver sa fonction de répartition $F_Y(y)$. Pour $y \leq 0$, clairement $F_Y(y) = 0$, puisque X étant entre 0 et 1, son logarithme sera toujours négatif, et Y sera toujours positive. Pour des valeurs de y positives,

$$\begin{aligned} F_Y(y) &= \mathbb{P}\{Y \leq y\} \\ &= \mathbb{P}\{-\log X \leq y\} \\ &= \mathbb{P}\{\log X \geq -y\} \\ &= \mathbb{P}\{X \geq e^{-y}\} \\ &= 1 - \mathbb{P}\{X < e^{-y}\} \\ &= 1 - \mathbb{P}\{X \leq e^{-y}\} \\ &= 1 - e^{-y}. \end{aligned}$$

4. L'équation (5.2.2) tient également pour les variables discrètes non-négatives ! La preuve est laissée pour l'exercice 5.18

La dernière égalité vient du fait que $\mathbb{P}\{X \leq x\} = \int_0^x dt = x$ pour tout x entre 0 et 1
Finalement, $\overline{F_Y}(y) = 1 - (1 - e^{-y}) = e^{-y}$, et

$$\mathbb{E}[-\log X] = \mathbb{E}[Y] = \int_0^\infty \overline{F_Y}(y)dy = \int_0^\infty e^{-y}dy = [-e^{-y}]_0^\infty = 1.$$

La méthode employée ici est légèrement encombrante, et, comme pour les variables discrètes, on aimerait avoir un résultat qui permet de calculer l'espérance pour une fonction de notre variable aléatoire. Et, comme pour les variables aléatoires discrètes, on en a un !

PROPOSITION 5.2. *Soient $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire continue avec densité f et $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction intégrable.*

Alors,

$$(5.2.3) \quad \mathbb{E}[g(X)] = \int_{-\infty}^\infty g(x)f(x)dx.$$

DÉMONSTRATION. Nous allons commencer par le montrer pour une fonction g positive – c'est à dire telle que $g(x) > 0$ pour tout x .

Dans ce cas, bien sûr, par la proposition 5.1, on a que

$$\begin{aligned} \mathbb{E}[g(X)] &= \int_0^\infty \mathbb{P}\{g(X) > t\} dt \\ &= \int_0^\infty \left(\int_{\{u:g(u)>t\}} f(u)du \right) dt \\ &= \int_{-\infty}^\infty \int_0^{g(u)} f(u)dtdu \\ &= \int_{-\infty}^\infty g(u)f(u)du. \end{aligned}$$

Dans le cas où g est une fonction quelconque, on sépare g en deux fonctions :

$$(5.2.4) \quad g^+(x) := \max\{g(x), 0\}; \quad g^-(x) := \max\{-g(x), 0\}.$$

L'équation (5.2.4) est très utile en général ; les fonctions g^+ et g^- sont positives et ont les propriétés suivantes :

$$g^+ - g^- = g; \quad g^+ + g^- = |g|.$$

En particulier,

$$\mathbb{E}[g(X)] = \mathbb{E}[g^+(X) - g^-(X)] = \mathbb{E}[g^+(X)] - \mathbb{E}[g^-(X)].$$

Mais puisque g^+ et g^- sont des fonctions positives, on a déjà montré comment calculer l'espérance de $g^+(X)$ et $g^-(X)$. On a donc finalement

$$\begin{aligned} \mathbb{E}[g(X)] &= \int_{-\infty}^\infty g^+(x)f(x)dx - \int_{-\infty}^\infty g^-(x)f(x)dx \\ &= \int_{-\infty}^\infty (g^+(x) - g^-(x))f(x)dx \\ &= \int_{-\infty}^\infty g(x)f(x)dx. \end{aligned}$$

□

On peut donc calculer le k ième moment d'une variable continue X qui a une densité f :

$$(5.2.5) \quad \mathbb{E}[X^k] = \int_{-\infty}^{\infty} x^k f(x) dx.$$

En particulier, si $\mu = \mathbb{E}[X]$, alors la variance d'une variable aléatoire continue X est donnée par

$$(5.2.6) \quad \text{Var}[X] = \mathbb{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx.$$

Il sera cependant toujours pratique de calculer la variance avec l'équation (4.5.1) – en rappel :

$$(4.5.1) \quad \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

EXEMPLE 5.7. Soit X une variable aléatoire continue avec densité f donnée par

$$f(x) = \begin{cases} nx^{n-1} & \text{si } 0 \leq x \leq 1 \\ 0 & \text{ailleurs.} \end{cases}$$

Trouver $\text{Var}[X]$.

SOLUTION. On calcule simplement :

$$\begin{aligned} \mathbb{E}[X^2] &= \int_0^1 x^2 \cdot nx^{n-1} dx \\ &= n \int_0^1 x^{n+1} dx \\ &= \frac{n}{n+2}. \end{aligned}$$

On avait calculé précédemment que $\mathbb{E}[X] = \frac{n}{n+1}$ – on déduit donc que

$$\text{Var}[X] = \frac{n}{n+2} - \frac{n^2}{(n+1)^2}.$$

On constate évidemment que $\text{Var}[X]$ tend vers 0 lorsque n est grand.

Finalement, on montre une proposition très utile pour le calcul de la variance :

PROPOSITION 5.3. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire quelconque et $a, b \in \mathbb{R}$ deux constantes.

Alors,

$$(5.2.7) \quad \text{Var}[aX + b] = a^2 \text{Var}[X].$$

DÉMONSTRATION. Posons $\mu = \mathbb{E}[X]$. Soit $X' = aX + b$ et $\mu' = \mathbb{E}[X'] = a\mu + b$. Alors,

$$\begin{aligned} \text{Var}[X'] &= \mathbb{E}[(X' - \mu')^2] \\ &= \mathbb{E}[(aX + b - a\mu - b)^2] \\ &= \mathbb{E}[a^2(X - \mu)^2] \\ &= a^2 \mathbb{E}[(X - \mu)^2] \\ &= a^2 \text{Var}[X]. \end{aligned}$$

□

5.3. Plusieurs distributions courantes

Pareil comme pour les variables aléatoires discrètes, il existe autant de distributions pour les variables aléatoires continues qu'on peut en imaginer – et, comme pour les variables aléatoires discrètes, il y a des distributions qui reviennent plus souvent que d'autres en raison de certaines propriétés particulières.

5.3.1. La distribution uniforme. On dit que X a une distribution **uniforme** si sa densité f est constante sur son support.

En particulier, la variable aléatoire X est distribuée uniformément sur l'intervalle (a, b) si sa densité est donnée par

$$(5.3.1) \quad f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in (a, b) \\ 0 & \text{ailleurs.} \end{cases}$$

La fonction de répartition d'une telle variable aléatoire est donnée par $F(x) = 0$ pour $x \leq a$, $F(x) = 1$ pour $x \geq b$, et, lorsque $x \in (a, b)$,

$$(5.3.2) \quad \begin{aligned} F(x) &= \frac{1}{b-a} \int_a^x dt \\ &= \frac{x-a}{b-a} \end{aligned}$$

PROPOSITION 5.4. Soit X une variable aléatoire de distribution uniforme sur (a, b) . Alors, $Y = \frac{X-a}{b-a}$ est de distribution uniforme sur $(0, 1)$.

DÉMONSTRATION. On fait la preuve en identifiant la distribution de Y par sa fonction de répartition : pour $x \leq 0$, $\mathbb{P}\{Y \leq x\} = 0$ et pour $x \geq 1$, $\mathbb{P}\{Y \leq x\} = 1$. Pour $x \in (0, 1)$,

$$\begin{aligned} \mathbb{P}\{Y \leq x\} &= \mathbb{P}\left\{\frac{X-a}{b-a} \leq x\right\} \\ &= \mathbb{P}\{X \leq (b-a)x + a\} \\ &= \frac{(b-a)x + a - a}{b-a} \\ &= x \end{aligned}$$

et la fonction de répartition de Y est celle d'une variable uniforme sur l'intervalle $(0, 1)$. \square

COROLLAIRE 5.1. Soit X une variable aléatoire de distribution uniforme sur $(0, 1)$. Alors, avec $a < b$, $Y = (b-a)X + a$ est une variable aléatoire uniforme sur (a, b) .

Autrement dit, on peut utiliser des transformations linéaires simples pour transformer les variables uniformes d'un intervalle à un autre.

Les moments des variables aléatoires uniformes. On a que, si X est distribuée uniformément sur $(0, 1)$,

$$(5.3.3) \quad \mathbb{E}[X] = \int_0^1 x dx = \frac{1}{2}.$$

Par la proposition 5.4, on sait que si X est une variable uniforme (a, b) , il existe une variable aléatoire Y uniforme $(0, 1)$ telle que $X = (b-a)Y + a$. Par conséquent, si X est

une variable uniforme (a, b) ,

$$(5.3.4) \quad \mathbb{E}[X] = \mathbb{E}[(b-a)Y + a] = a + \frac{b-a}{2} = \frac{a+b}{2}.$$

De même, lorsque X est une variable aléatoire uniforme sur $(0, 1)$,

$$(5.3.5) \quad \mathbb{E}[X^2] = \int_0^1 x^2 dx = \frac{1}{3},$$

et sa variance est

$$(5.3.6) \quad \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Et encore une fois, par les propositions 5.4 et 5.3, si X est uniforme sur (a, b) , il existe une variable Y uniforme sur $(0, 1)$ telle que $X = (b-a)Y + a$ et

$$(5.3.7) \quad \text{Var}[X] = \text{Var}[(b-a)Y + a] = (b-a)^2 \text{Var}[Y] = \frac{(b-a)^2}{12}.$$

EXEMPLE 5.8. Entre 7h et 8h, les autobus sont aux 6 minutes sur la ligne 18, avec le premier autobus à 7h 02. en direction ouest. Thomas arrive à l'arrêt de bus à un temps T distribué uniformément entre 7h et 7h 15. Si $A(t)$ est le temps d'attente de l'autobus,

- (a) Donner une expression pour $A(t)$.
- (b) Calculer $\mathbb{E}[A(T)]$ l'espérance du temps d'attente de l'autobus. $\mathbb{E}[A]$.

SOLUTION. On va dire que T est une variable aléatoire uniforme sur l'intervalle (7h, 7h 15). Alors, sa densité est $f(x) = \frac{1}{15 \text{ min.}} \mathbb{1}_{(7h, 7h\ 15)}(x)$.

- (a) On a que

$$A(t) = \begin{cases} 7h\ 02 - t & \text{si } 7h < t \leq 7h\ 02 \\ 7h\ 08 - t & \text{si } 7h\ 02 < t \leq 7h\ 08 \\ 7h\ 14 - t & \text{si } 7h\ 08 < t \leq 7h\ 14 \\ 7h\ 20 - t & \text{si } 7h\ 14 < t \leq 7h\ 15. \end{cases}$$

- (b) Il suffit de calculer. Si on compte en minutes depuis 7h, alors on a que

$$\begin{aligned} \mathbb{E}[A(T)] &= \frac{1}{15} \int_0^{15} A(t) dt \\ &= \frac{1}{15} \left[\int_0^2 (2-t) dt + \int_2^8 (8-t) dt + \int_8^{14} (14-t) dt + \int_{14}^{15} (20-t) dt \right] \\ &= \frac{1}{30} [4 + 36 + 36 + 11] \\ &= \frac{87}{30} = \frac{174}{60} = 2 \text{ min. } 54 \text{ s..} \end{aligned}$$

5.3.2. La distribution normale. On dit que X est de distribution **normale** avec paramètres (μ, σ^2) lorsque la densité f de la variable aléatoire X est donnée pour tout $x \in \mathbb{R}$ par :

$$(5.3.8) \quad f(x) = C(\sigma^2) e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

avec $C(\sigma^2)$ une constante de normalisation.

LEMME 5.1. *La constante de normalisation pour la densité de la distribution normale avec paramètres (μ, σ^2) est*

$$C(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}.$$

DÉMONSTRATION. On calcule d'abord $C(\sigma^2)^2$: par la condition de normalisation,

$$\begin{aligned} 1 &= \left(C(\sigma^2) \int_{-\infty}^{\infty} e^{-(x-\mu)^2/(2\sigma^2)} dx \right) \left(C(\sigma^2) \int_{-\infty}^{\infty} e^{-(y-\mu)^2/(2\sigma^2)} dy \right) \\ &= C(\sigma^2)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(u^2+v^2)/(2\sigma^2)} du dv \\ &= C(\sigma^2)^2 \int_{-\pi}^{\pi} \int_0^{\infty} e^{-r^2/(2\sigma^2)} r dr d\theta \\ &= C(\sigma^2)^2 \cdot 2\pi \int_0^{\infty} r e^{-r^2/(2\sigma^2)} dr \\ &= C(\sigma^2)^2 \cdot 2\pi \int_0^{\infty} e^{-t/\sigma^2} dt \\ &= C(\sigma^2)^2 \cdot 2\pi\sigma^2. \end{aligned}$$

On conclue donc que

$$C(\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}.$$

□

Si X est une variable aléatoire de loi normale (μ, σ^2) avec $\mu = 0$, on dit qu'elle est **centrée**. Si $\sigma^2 = 1$, on dit qu'elle est **réduite**.

La fonction de répartition de la distribution normale. On note

$$(5.3.9) \quad \Phi(x) = \mathbb{P}\{Z \leq x\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt,$$

où Z est une variable aléatoire de loi normale centrée réduite ($\mu = 0, \sigma^2 = 1$).

Cette fonction n'est pas descriptible en forme fermée – c'est pourquoi nous lui attribuerons le symbole Φ pour la suite des choses. On notera aussi $\phi = \Phi'$ la densité d'une variable aléatoire normale centrée réduite :

$$(5.3.10) \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

PROPOSITION 5.5. *Soit Z une variable normale centrée et réduite. Alors, pour $a, b \in \mathbb{R}$, avec $a \neq 0$, $X = aZ + b$ est une variable aléatoire normale de paramètres $(\mu = b, \sigma^2 = a^2)$.*

DÉMONSTRATION. En supposant que $a > 0$, la fonction de répartition de X est donnée par

$$\mathbb{P}\{X \leq x\} = \mathbb{P}\left\{Z \leq \frac{x-b}{a}\right\} = \Phi\left(\frac{x-b}{a}\right).$$

En dérivant on trouve que

$$f(x) = \frac{1}{a} \Phi'\left(\frac{x-b}{a}\right) = \frac{1}{\sqrt{2\pi}a^2} e^{-(x-b)^2/(2a^2)},$$

ce qui correspond à la densité d'une variable aléatoire normale.

Si on suppose plutôt que $a < 0$, alors

$$\mathbb{P}\{X \leq x\} = \mathbb{P}\left\{Z \geq \frac{x-b}{a}\right\} = 1 - \Phi\left(\frac{x-b}{a}\right).$$

et en dérivant on trouve que

$$f(x) = \frac{1}{-a} \Phi'((x-b)/(a)) = \frac{1}{\sqrt{2\pi a^2}} e^{-(x-b)^2/(2a^2)}.$$

Dans tous les cas, on obtient que X a la densité d'une variable aléatoire normale de paramètres $\mu = b$ et $\sigma^2 = a^2$. \square

COROLLAIRE 5.2. *Si X a une distribution normale de paramètres (μ, σ^2) , alors pour $a \neq 0$ et b quelconque, $X' = aX + b$ est une variable aléatoire normale de paramètres $\mu' + a\mu + b$ et $\sigma'^2 = a^2\sigma^2$.*

En particulier, $Z = \frac{X-\mu}{\sigma}$ est une variable aléatoire normale centrée réduite.

Au final, si X est une variable aléatoire normale de paramètres (μ, σ^2) , et que Z est une normale centrée réduite, alors on voit que la fonction de répartition de X est donnée par :

$$(5.3.11) \quad \mathbb{P}\{X \leq x\} = \mathbb{P}\{\sigma Z + \mu \leq x\} = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Les moments de la distribution normale. On a que l'espérance d'une variable normale centrée réduite est de 0 – pour ce faire, on utilise le lemme suivant :

LEMME (Proposition B.26)). *Soient $f : (-a, a) \rightarrow \mathbb{R}$ une fonction paire et $g : (-a, a) \rightarrow \mathbb{R}$ une fonction impaire, toutes deux absolument intégrables, avec $0 < a \leq \infty$.*

Alors :

$$i. \int_{-a}^a f(t)dt = 2 \int_0^a f(t)dt;$$

$$ii. \int_{-a}^a g(t)dt = 0;$$

Ce lemme (un résultat d'analyse réelle) nous sera souvent utile pour simplifier les calculs d'intégrales, en employant des symétries dans les intégrandes.

En l'occurrence, $\phi(x)$ est paire et x est impaire – donc $x\phi(x)$ est une fonction impaire, et

$$(5.3.12) \quad \begin{aligned} \mathbb{E}[Z] &= \int_{-\infty}^{\infty} x\phi(x)dx \\ &= 0. \end{aligned}$$

Il suit, par le corolaire 5.2, que si X est normalement distribuée avec paramètres (μ, σ^2) , alors, avec $Z = \frac{X-\mu}{\sigma}$ est une variable normale centrée réduite, et

$$\mathbb{E}[Z] = \frac{\mathbb{E}[X] - \mu}{\sigma} = 0,$$

ce qui conduit directement à la conclusion que, pour X une variable aléatoire normale de paramètres (μ, σ^2) , on a que

$$(5.3.13) \quad \mathbb{E}[X] = \mu.$$

Si Z est une variable aléatoire normale centrée réduite, par calcul direct (en intégrant par parties),

$$\begin{aligned}
 \mathbb{E}[Z^2] &= \int_{-\infty}^{\infty} x^2 \phi(x) dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot x e^{-x^2/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \left(\left[-xe^{-x^2/2} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} dx \right) dx \\
 &= 1.
 \end{aligned} \tag{5.3.14}$$

Et, puisque $\mu = 0$ et que $\text{Var}[Z] = \mathbb{E}[(Z - \mu)^2] = \mathbb{E}[Z^2]$, on a directement que

$$\text{Var}[Z] = 1. \tag{5.3.15}$$

Évidemment, encore une fois grâce au corollaire 5.2 et à la proposition 5.3, on voit directement que, si X est une normale de paramètres (μ, σ^2) , alors $Z = \frac{X-\mu}{\sigma}$ est une normale centrée réduite et

$$\text{Var}[Z] = \text{Var}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma^2} \text{Var}[X] = 1,$$

d'où l'on tire que

$$\text{Var}[X] = \sigma^2. \tag{5.3.16}$$

EXEMPLE 5.9. On s'imagine que la hauteur T d'un humain mâle adulte est une variable aléatoire distribuée normalement, avec une espérance de $\mu = 177$ cm et une variance de $\sigma^2 = 100$ cm².

- (a) La distribution normale est supportée sur $(-\infty, \infty)$ – mais bien sûr, la véritable taille d'une personne ne peut pas être inférieure à 0. Trouver $\mathbb{P}\{T \leq 0\}$. L'erreur occasionnée est-elle significative ?
- (b) Le 27 juin 1940, Robert Wadlow mesurait 272 centimètres – soit environ 2 mètres 3/4, ou 8'11". Calculer $\mathbb{P}\{T \geq 272\}$, la probabilité d'excéder cette taille.

SOLUTION. (a) La variance est $\sigma^2 = 100$ cm², donc l'écart-type est $\sigma = \sqrt{\sigma^2} = 10$ cm. Par l'équation (5.3.11), la fonction de répartition de T est

$$\mathbb{P}\{T \leq x\} = \Phi\left(\frac{x - 177}{10}\right).$$

Donc, $\mathbb{P}\{T \leq 0\} = \Phi\left(-\frac{177}{10}\right) \approx 2,09 \times 10^{-70}$.

Non, l'erreur ainsi réalisée n'est pas significative.

- (b) De même, $\mathbb{P}\{T \geq 272\} = 1 - \mathbb{P}\{T \leq 272\} = 1 - \Phi\left(\frac{272 - 177}{10}\right) \approx 1,05 \times 10^{-21}$.

L'approximation de Binomiales par la loi normale. Si $X_{n,p}$ est une variable aléatoire de distribution binomiale avec paramètres (n, p) , alors le théorème de De Moivre–Laplace nous permet d'approximer les probabilités pour notre variable par celles d'une variable aléatoire de densité normale.⁵

PROPOSITION 5.6 (Théorème de De Moivre–Laplace). *Soit $X_{n,p}$ une variable aléatoire de loi binomiale avec paramètres (n, p) .*

Alors, on a que pour tout $x \in \mathbb{R}$,

$$(5.3.17) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{X_{n,p} - np}{\sqrt{np(1-p)}} \leq x \right\} = \Phi(x)$$

EXEMPLE 5.10. On lance une pièce de monnaie équilibrée dix mille fois. Si N est le nombre de « pile » obtenus, alors N est une variable binomiale de paramètres $n = 10000, p = 1/2$.

Trouver la probabilité d'obtenir entre 4900 et 5100 fois « pile ».

SOLUTION. On utilise l'approximation de De Moivre–Laplace (proposition 5.6) :

$$\begin{aligned} \mathbb{P}\{4900 \leq N \leq 5100\} &= \mathbb{P} \left\{ \frac{4900 - 10000 \cdot (1/2)}{\sqrt{10000 \cdot (1/2) \cdot (1/2)}} \leq \frac{N - 10000 \cdot (1/2)}{\sqrt{10000 \cdot (1/2) \cdot (1/2)}} \leq \frac{5100 - 10000 \cdot (1/2)}{\sqrt{10000 \cdot (1/2) \cdot (1/2)}} \right\} \\ &= \mathbb{P} \left\{ -2 \leq \frac{N - 5000}{50} \leq 2 \right\} \\ &\approx \Phi(2) - \Phi(-2) \\ &\approx 95,45\%. \end{aligned}$$

REMARQUE. En travaillant avec des variables aléatoires normalement distribuées, l'écart-type σ devient très utile – par exemple, il sera toujours vrai que, pour de grandes valeurs du paramètre n , une Binomiale (n, p) se retrouvera à moins de deux σ avec probabilité $\approx 95,45\%$.

5.3.3. La distribution exponentielle. Soit X une variable aléatoire continue de densité f . On dit que X est de distribution **exponentielle** si, pour un certain $\lambda > 0$, on a que

$$(5.3.18) \quad f(x) = \begin{cases} C(\lambda)e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0. \end{cases}$$

Et on détermine rapidement la constante $C(\lambda)$ appropriée :

LEMME 5.2. *On a que*

$$(5.3.19) \quad C(\lambda) = \lambda.$$

DÉMONSTRATION. Il suffit de calculer :

$$1 = C(\lambda) \int_0^\infty e^{-\lambda x} dx = \frac{C(\lambda)}{\lambda}.$$

□

5. Ce théorème n'est en fait qu'un cas particulier du théorème de la limite centrale (Théorème 8.3).

Soit $F(x) = \mathbb{P}\{X \leq x\}$ la fonction de répartition d'une telle variable aléatoire.

Alors, bien sûr, étant donné la densité, on sait immédiatement que $F(x) = 0$ pour tout $x \leq 0$.

Pour $x > 0$,

$$\begin{aligned} \mathbb{P}\{X \leq x\} &= \lambda \int_0^x e^{-\lambda t} dt \\ &= \lambda \left[-\frac{e^{-\lambda t}}{\lambda} \right]_0^x \\ &= 1 - e^{-\lambda x}. \end{aligned} \tag{5.3.20}$$

PROPOSITION 5.7. Soit X une variable aléatoire de loi exponentielle avec paramètre $\lambda > 0$, et soit $c > 0$ une constante réelle.

Alors, $Y = cX$ est une variable aléatoire de loi exponentielle avec paramètre λ/c .

DÉMONSTRATION. La preuve est laissée pour l'exercice 5.14. \square

Les moments de la distribution exponentielle. Si X est une variable aléatoire de loi exponentielle avec paramètre $\lambda > 0$, X est une variable aléatoire non-négative. Sa fonction de répartition complémentaire est

$$\bar{F}(x) = e^{-\lambda x}$$

pour $x > 0$ (et 1 pour $x \leq 0$), et par la proposition 5.1, on a que

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty \bar{F}(x) dx \\ &= \int_0^\infty e^{-\lambda x} dx \\ &= \frac{1}{\lambda}. \end{aligned} \tag{5.3.21}$$

Le second moment se calcule facilement :

$$\begin{aligned} \mathbb{E}[X^2] &= \lambda \int_0^\infty x^2 e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} \int_0^\infty u^2 e^{-u} du \\ &= \frac{1}{\lambda^2} \left([-u^2 e^{-u}]_0^\infty + 2 \int_0^1 ue^{-u} du \right) \\ &= \frac{2}{\lambda^2}. \end{aligned} \tag{5.3.22}$$

On déduit donc immédiatement que la variance est donnée par

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}. \tag{5.3.23}$$

EXEMPLE 5.11. Dans un centre d'appels, le temps T entre la réception de deux appels est une variable exponentielle de paramètre $\lambda = 1/5 \text{ min.}^{-1}$.

- (a) Déterminer la probabilité de ne recevoir aucun appel pour au moins 10 minutes.
- (b) Déterminer la probabilité que le prochain appel arrive dans entre 3 et 5 minutes.

- (c) Ça fait 5 minutes depuis le dernier appel. Quel est la probabilité que le prochain appel n'arrive pas avant 3 autres minutes ?

SOLUTION. (a) La probabilité de ne recevoir aucun appel pour au moins 10 minutes est

$$\mathbb{P}\{T \geq 10\} = e^{-2} \approx 13,5\%.$$

- (b) La probabilité de recevoir le prochain appel dans entre 3 et 5 minutes est de

$$\mathbb{P}\{3 \leq T \leq 5\} = e^{-3/5} - e^{-1} \approx 18,09\%.$$

- (c) On cherche

$$\mathbb{P}\{T - 5 \geq 3 \mid T \geq 5\} = \frac{\mathbb{P}\{T \geq 8\}}{\mathbb{P}\{T \geq 5\}}.$$

C'est

$$\mathbb{P}\{T - 5 \geq 3 \mid T \geq 5\} = \frac{e^{-8/5}}{e^{-1}} = e^{-3/5} \approx 54,88\%.$$

On remarque que c'est exactement la même probabilité que si on venait juste de recevoir l'appel et qu'on devait attendre au moins 3 minutes – autrement dit, le temps déjà attendu ne fait aucune différence.

DÉFINITION 5.4 (Absence de mémoire (v.a. continues)). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et soit $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire continue non-négative.

On dit que X est **sans mémoire** si pour tout $t, \delta > 0$, on a que

$$(5.3.24) \quad \mathbb{P}\{X > t + \delta \mid X > t\} = \mathbb{P}\{X > \delta\}.$$

La définition d'absence de mémoire pour les variables continues est en tous points analogue à celle pour les variables discrètes (définition 4.9).

De même, on va montrer que les variables aléatoires exponentielles sont les seules à posséder cette propriété pour les variables aléatoires continues.

PROPOSITION 5.8. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire continue non-négative.

Alors, X est sans mémoire si et seulement si X est une variable aléatoire exponentielle.

DÉMONSTRATION. \Leftarrow : Si X est une variable aléatoire exponentielle, alors

$$\mathbb{P}\{X > u\} = e^{-\lambda u}$$

pour tout $u > 0$.

On a alors que pour tous $t, \delta > 0$,

$$\mathbb{P}\{X > t + \delta \mid X > t\} = \frac{e^{-\lambda(t+\delta)}}{e^{-\lambda t}} = e^{-\lambda\delta} = \mathbb{P}\{X > \delta\},$$

et X est sans mémoire.

\Rightarrow : Si X est sans mémoire, alors on a que pour tous $t, \delta > 0$,

$$\mathbb{P}\{X > t + \delta \mid X > t\} = \frac{\bar{F}(t + \delta)}{\bar{F}(t)} = \bar{F}(\delta) = \mathbb{P}\{X > \delta\}.$$

On déduit que pour tout $t, \delta > 0$,

$$\bar{F}(t + \delta) = \bar{F}(t)\bar{F}(\delta).$$

Les seules fonctions qui satisfont cette équation fonctionnelle sont des fonctions de la forme $\bar{F}(x) = e^{\alpha x}$ pour un certain paramètre α – en particulier, on doit avoir que \bar{F} est décroissante et que $\bar{F}(0) = 1$. On doit donc avoir que $\alpha < 0$ et on prend $\lambda = -\alpha > 0$.

Alors, X doit avoir $\bar{F}(x) = e^{-\lambda x}$ – donc X doit être une variable aléatoire de distribution exponentielle. \square

EXEMPLE 5.12. Au supermarché, il y a deux caisses, qui opèrent au même rythme – chacune d'entre elles sert chaque client en un temps distribué selon une loi exponentielle.

Lorsqu'on arrive aux caisses, elles sont toutes deux occupées, et nous sommes les prochains en file. Quelle est la probabilité qu'on ressorte en dernier du supermarché (après les deux autres client.e.s) ?

SOLUTION. Au moment où un.e client.e aura terminé sa transaction, nous passerons à sa caisse. À ce moment là, le temps restant à notre transaction sera une variable aléatoire exponentielle de paramètre λ . Mais à partir de ce moment-là, le temps restant à la transaction de la personne à l'autre caisse sera *aussi* une exponentielle de paramètre λ – vu que la distribution exponentielle est sans mémoire, le fait que la personne était là avant ne change rien !

Donc, par symétrie, il doit être tout aussi probable que ce soit nous ou l'autre qui sortions en premier.⁶ Donc la probabilité qu'on sorte avant est de $1/2$.

5.3.4. La distribution Gamma. Une variable aléatoire X est dite suivre une distribution Gamma de paramètres (λ, α) si sa densité f est donnée par

$$(5.3.25) \quad f(x) = C(\lambda, \alpha) e^{-\lambda x} x^{\alpha-1}.$$

LEMME 5.3. *On a que*⁷

$$(5.3.26) \quad C(\lambda, \alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)}.$$

DÉMONSTRATION. On doit avoir

$$\begin{aligned} 1 &= C(\lambda, \alpha) \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx \\ &= C(\lambda, \alpha) \cdot \frac{1}{\lambda^\alpha} \int_0^\infty u^{\alpha-1} e^{-u} du \\ &= C(\lambda, \alpha) \cdot \frac{\Gamma(\alpha)}{\lambda^\alpha}. \end{aligned}$$

On conclue donc que $C(\lambda, \alpha) = \frac{\lambda^\alpha}{\Gamma(\alpha)}$. \square

REMARQUE. On a que si $\alpha = 1$, la distribution Gamma $(\lambda, \alpha = 1)$ se réduit à la distribution exponentielle.

PROPOSITION 5.9. *Soit X une variable aléatoire suivant une loi Gamma de paramètres $\lambda > 0, \alpha > 0$. Soit $c > 0$ une constante réelle positive.*

Alors, cX est une variable aléatoire de loi Gamma avec paramètres $(\lambda/c, \alpha)$.

6. Ici on fait sans le dire une hypothèse supplémentaire – que les temps d'attente sont *interchangeables* (par exemple parce qu'ils seraient indépendants). Plus aux chapitres 6 et 7.

7. Consulter la section B.8.1 au sujet de la fonction Γ .

DÉMONSTRATION. La preuve est laissée pour l'exercice 5.15. \square

On calcule l'espérance d'une variable aléatoire de loi Gamma en utilisant les propriétés de la fonction Gamma (voir la section B.8.1) : si X suit une loi Gamma (λ, α) ,

$$\begin{aligned}
 \mathbb{E}[X] &= \int_0^\infty x \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} dx \\
 &= \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda x} (\lambda x)^\alpha dx \\
 &= \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\lambda} \int_0^\infty e^{-u} u^\alpha du \\
 &= \frac{1}{\lambda} \cdot \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \\
 &= \frac{1}{\lambda} \cdot \frac{\alpha \Gamma(\alpha)}{\Gamma(\alpha)} \\
 &= \frac{\alpha}{\lambda}.
 \end{aligned} \tag{5.3.27}$$

De façon similaire,

$$\begin{aligned}
 \mathbb{E}[X^2] &= \int_0^\infty x^2 \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1} dx \\
 &= \frac{\Gamma(\alpha+2)}{\lambda^2 \Gamma(\alpha)} \\
 &= \frac{\alpha(\alpha+1)}{\lambda^2},
 \end{aligned} \tag{5.3.28}$$

et par la proposition 4.4, on obtient

$$\text{Var}[X] = \frac{\alpha}{\lambda^2}. \tag{5.3.29}$$

Nous ne nous attarderons pas sur la distribution Gamma pour l'instant, mais son interprétation pratique reviendra au chapitre 6.

5.3.5. La distribution de Cauchy. X est une variable aléatoire **de Cauchy** avec paramètres $\theta \in \mathbb{R}, \sigma > 0$ si on a que sa densité f est donnée pour tout $x \in \mathbb{R}$ par

$$f(x) = C(\sigma) \frac{1}{1 + \left(\frac{x-\theta}{\sigma}\right)^2}. \tag{5.3.30}$$

LEMME 5.4. *On a que*

$$C(\sigma) = \frac{1}{\pi \sigma}. \tag{5.3.31}$$

DÉMONSTRATION. Pour faire cette preuve il suffit de faire le changement de variables $u = (x - \theta)/\sigma$, en se souvenant que

$$\int_0^x \frac{dt}{1+t^2} = \arctan x.$$

$$\begin{aligned}
1 &= C(\sigma) \int_{-\infty}^{\infty} \frac{dx}{1 + \left(\frac{x-\theta}{\sigma}\right)^2} \\
&= C(\sigma) \cdot \sigma \int_{-\infty}^{\infty} \frac{du}{1 + u^2} \\
&= C(\sigma) \cdot \sigma [\arctan u]_{-\infty}^{\infty} \\
&= C(\sigma) \cdot \pi\sigma.
\end{aligned}$$

On conclue donc que $C(\sigma) = \frac{1}{\pi\sigma}$. □

De la même façon analogue, on détermine que

$$(5.3.32) \quad \mathbb{P}\{X \leq x\} = \frac{1}{\pi\sigma} \int_{-\infty}^x \frac{dt}{1 + \left(\frac{t-\theta}{\sigma}\right)^2} = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\theta}{\sigma}\right).$$

On dira que X est une variable de Cauchy **standard** si $\theta = 0$ et $\sigma = 1$.

PROPOSITION 5.10. *Soit X une variable aléatoire de Cauchy de paramètres θ, σ , et soit $X' = aX + b$ avec $a \neq 0$ et $b \in \mathbb{R}$.*

Alors, X' est une variable aléatoire de Cauchy de paramètres $\theta' = a\theta + b$ et $\sigma' = |a|\sigma$.

DÉMONSTRATION. La preuve est laissée pour l'exercice 5.17. □

Pour une variable X de Cauchy standard, évidemment la densité est

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

Le souci, c'est que

$$|xf(x)| = \frac{1}{\pi} \frac{|x|}{1 + x^2} = O\left(\frac{1}{|x|}\right) \quad (x \rightarrow \pm\infty).$$

Donc, bien que la fonction $xf(x)$ soit localement intégrable, elle n'est pas intégrable sur tout \mathbb{R} .

Il suit que

$$(5.3.33) \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx \text{ n'est pas définie.}$$

EXEMPLE 5.13. On s'imagine qu'un pointeur laser est situé à distance 1 d'un mur de longueur infini. L'angle d'incidence du pointeur laser sur le mur est un angle Θ de loi uniforme entre $-\pi/2$ et $\pi/2$. On mesure la distance H entre le point du laser sur le mur et le point d'origine sur le mur.

Montrer que H suit une distribution de Cauchy.

SOLUTION. On a que $H = \tan \Theta$, par la géométrie du problème.

On veut déterminer la loi de H . Pour ce faire, on va calculer sa fonction de répartition :

$$\begin{aligned}
\mathbb{P}\{H \leq h\} &= \mathbb{P}\{\tan \Theta \leq h\} \\
&= \mathbb{P}\{\Theta \leq \arctan h\} \\
&= \frac{\arctan(h) + \frac{\pi}{2}}{\pi} \\
&= \frac{1}{2} + \frac{1}{\pi} \arctan(h).
\end{aligned}$$

Et on voit que H suit une loi de Cauchy standard ($\theta = 0, \sigma = 1$).

5.3.6. La distribution Beta. On dit que X suit une loi Beta de paramètres α, β lorsque sa densité f est donnée par

$$(5.3.34) \quad f(x) = \begin{cases} C(\alpha, \beta)x^{\alpha-1}(1-x)^{\beta-1} & \text{si } 0 < x < 1 \\ 0 & \text{sinon.} \end{cases}$$

LEMME 5.5. *On a que*

$$(5.3.35) \quad C(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

DÉMONSTRATION. On doit avoir que

$$\frac{1}{C(\alpha, \beta)} = B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

On considère

$$\begin{aligned}
\Gamma(x)\Gamma(y) &= \left(\int_0^x nfty u^{x-1} e^{-u} du \right) \left(v^{y-1} e^{-v} dv \right) \\
&= \int_0^\infty \int_0^\infty u^{x-1} v^{y-1} e^{-u-v} dudv.
\end{aligned}$$

Mais maintenant, si on fait le changement de variables $w = u + v$ et $r = \frac{u}{u+v}$ (de sorte que les transformations inverses soient $u = rw$ et $v = (1-r)w$), alors $dudv = |J(w, r)| dr dw$ avec le déterminant jacobien

$$|\det J(w, r)| = \left| \det \begin{pmatrix} \frac{\partial u}{\partial w} & \frac{\partial u}{\partial r} \\ \frac{\partial v}{\partial w} & \frac{\partial v}{\partial r} \end{pmatrix} \right| = w$$

Donc, $dudv = wdrdw$.

Évidemment, w va de 0 à l'infini puisque u et v vont de 0 à l'infini – mais r va de 0 à 1. On a donc

$$\begin{aligned}
\Gamma(x)\Gamma(y) &= \int_0^\infty \int_0^\infty e^{-u-v} u^{x-1} v^{y-1} dudv \\
&= \int_0^\infty \int_0^1 e^{-w} (wr)^{x-1} (w(1-r))^{y-1} wdrdw \\
&= \int_0^\infty e^{-w} w^{x+y-1} dw \cdot \int_0^1 r^{x-1} (1-r)^{y-1} dr \\
&= \Gamma(x+y) B(x, y).
\end{aligned}$$

D'où on tire que

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

et

$$C(\alpha, \beta) = \frac{1}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}.$$

□

REMARQUE. Dans le cas particulier où $\beta = 1 - \alpha$, on a que $C(\alpha, 1 - \alpha) = \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} = \frac{\pi}{\sin \pi \alpha}$. La densité devient donc :

$$f(x) = \frac{\sin(\pi\alpha)}{\pi} x^{\alpha-1} (1-x)^{-\alpha}, \quad 0 < x < 1.$$

Cette densité est connue en tant que **loi de l'arcsius généralisée** – lorsque $\alpha = 1/2$, on trouve en effet que

$$f(x) = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$$

pour $x \in (0, 1)$, et la fonction de répartition pour ces valeurs est donnée par $F(x) = \frac{1}{2} + \frac{1}{\pi} \arcsin(2x - 1)$.

On calcule l'espérance de notre variable :

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{B(\alpha, \beta)} \int_0^1 x \cdot x^{\alpha-1} (1-x)^{\beta-1} dx \\ &= \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} \\ &= \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{\alpha}{\alpha+\beta}. \end{aligned} \tag{5.3.36}$$

De façon similaire, le k ième moment est donné par

$$\begin{aligned} \mathbb{E}[X^k] &= \frac{B(\alpha+k, \beta)}{B(\alpha, \beta)} \\ &= \frac{\Gamma(\alpha+k)\Gamma(\beta)}{\Gamma(\alpha+\beta+k)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\ &= \frac{\alpha+k-1}{\alpha+\beta+k-1} \mathbb{E}[X^{k-1}], \end{aligned} \tag{5.3.37}$$

et en particulier,

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\ &= \frac{(\alpha^2+\alpha)(\alpha+\beta) - \alpha^2(\alpha+\beta) - \alpha^2}{(\alpha+\beta)^2(\alpha+\beta+1)} \\ &= \frac{\alpha\beta}{(\alpha+\beta)^3 + (\alpha+\beta)^2}. \end{aligned} \tag{5.3.38}$$

5.3.7. Une notation pratique. Pour clore cette section, nous allons introduire encore un peu de notation. Tout au long du chapitre, nous avons utilisé de façon très répétitive des expressions comme « X est une variable aléatoire suivant une loi normale de paramètres μ, σ^2 ». C'est très long.

Pour se raccourcir la vie, nous allons d'abord introduire la notion d'équivalence en loi :

DÉFINITION 5.5 (Équivalence en loi). Soient X et Y deux variables aléatoires définies sur un même espace de probabilités $(\Omega, \mathcal{E}, \mathbb{P})$.

On dit que X et Y sont **équivalentes en loi** – noté $X \sim Y$ – si X et Y suivent la même distribution.

La proposition suivante nous permet d'identifier quand on peut dire que deux variables aléatoires sont équivalentes en loi :

PROPOSITION 5.11. *Soient X et Y deux variables aléatoires quelconques.*

Les énoncés suivants sont logiquement équivalents :

- i. $X \sim Y$: X et Y sont équivalentes en loi.
- ii. X et Y ont la même fonction de densité et la même fonction de masse de probabilité.
- iii. X et Y ont la même fonction de répartition.
- iv. X et Y ont la même fonction de répartition complémentaire.

DÉMONSTRATION. On a que i et iii sont équivalents par définition de la loi d'une variable aléatoire.

On a que iii et ii sont équivalents parce que la fonction de répartition est une primitive de la densité et que la fonction de masse de probabilité détermine uniquement les sauts aux discontinuités de la fonction de répartition.

Finalement, on a que iii et iv sont équivalents parce que la fonction de répartition et la fonction de répartition complémentaire sont définies la seconde par la première. \square

La proposition 5.11 permet d'utiliser n'importe laquelle de ces méthodes pour déterminer si deux variables aléatoires ont la même loi.

Finalement, ce qu'on va faire, c'est se doter d'une notation simple pour désigner rapidement une variable aléatoire générique pour chacune des lois qu'on a vues. Ils sont répertoriés au tableau 1.⁸

Ainsi, par exemple, l'affirmation

« la variable X suit une loi normale avec paramètres μ, σ^2 »

pourra être synthétisée par

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

8. Bien qu'il soit d'usage courant de définir de tels symboles, les symboles du tableau 1 sont entièrement arbitraire et ne font pas l'objet d'un large consensus dans la communauté mathématique. Ils seront principalement utilisés dans le reste des notes pour alléger le texte, mais ne correspondent pas nécessairement à ceux employés dans d'autres ouvrages de référence.

Distribution	Symbole	Paramètres
<i>Distributions discrètes :</i>		
Bernoulli	$Bern(p)$	$p \in [0, 1]$
Binomiale	$Bin(n, p)$	$n \in \mathbb{N}$ $p \in [0, 1].$
Géométrique	$Geom(p)$	$p \in [0, 1]$
Poisson	$Poi(\lambda)$	$\lambda > 0$
		$N \in \mathbb{N}$
Hypergéométrique	$HG(N, M, n)$	$M \in \mathbb{N}, \leq N$ $n \in \mathbb{N}, \leq N.$
Binomiale négative	$BN(r, p)$	$r \in \mathbb{N}, p \in [0, 1]$
Zeta	$Zeta(\alpha)$	$\alpha > 0$
<i>Distributions continues :</i>		
Uniforme	$\mathcal{U}(I)$ ou $\mathcal{U}(a, b)$	I un ouvert de \mathbb{R} OU (a, b) avec $a < b$
Normale	$\mathcal{N}(\mu, \sigma^2)$	$\mu \in \mathbb{R}$ $\sigma^2 > 0$
Exponentielle	$\mathfrak{E}(\lambda)$	$\lambda > 0$
Gamma	$\Gamma(\lambda, \alpha)$	$\lambda > 0$ $\alpha > 0$
Cauchy	$\mathfrak{C}(\theta, \sigma)$	$\theta \in \mathbb{R}$ $\sigma > 0$
Beta	$\mathfrak{B}(\alpha, \beta)$	$\alpha > 0$ $\beta > 0$

TABLE 1. Liste des symboles que nous utiliserons pour représenter des variables aléatoires génériques issues de diverses distributions.

5.4. Fonctions d'une variable aléatoire continue

À plusieurs reprises nous avons fait des exemples où nous cherchions à savoir la distribution, non pas d'une variable aléatoire directement, mais plutôt d'une « transformation » de cette variable aléatoire par le biais d'une fonction quelconque.

À peu près à chaque fois, la méthode était la même : déterminer la fonction de répartition ou la densité, en utilisant la transformation inverse.

Voici d'autres exemples :

EXEMPLE 5.14. Si $X \sim \mathcal{U}(0, 1)$, trouver la densité de X^n .

SOLUTION. On a que la fonction de répartition de X est $\mathbb{P}\{X \leq x\} = x$ pour $0 < x < 1$. On va donc avoir

$$\mathbb{P}\{X^n \leq x\} = \mathbb{P}\left\{X \leq x^{1/n}\right\} = x^{1/n}$$

pour $x \in (0, 1)$, et sa densité sera, pour $x \in (0, 1)$,

$$f(x) = \frac{d}{dx}x^{1/n} = \frac{1}{n}x^{1/n-1}.$$

EXEMPLE 5.15. Si X a une densité f_X , trouver la fonction de répartition pour X^2 et $|X|$.

SOLUTION. On a que

$$\begin{aligned}\mathbb{P}\{X^2 \leq x\} &= \mathbb{P}\{-\sqrt{x} \leq X \leq \sqrt{x}\} \\ &= \int_{-\sqrt{x}}^{\sqrt{x}} f_X(t) dt.\end{aligned}$$

On a également

$$\begin{aligned}\mathbb{P}\{|X| \leq x\} &= \mathbb{P}\{-x \leq X \leq x\} \\ &= \int_{-x}^x f_X(t) dt.\end{aligned}$$

PROPOSITION 5.12. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire continue, avec fonction de répartition F_X et densité f_X .

Supposons que $g : \mathbb{R} \rightarrow \mathbb{R}$ est non-décroissante et continue (donc différentiable⁹).

Alors, la variable aléatoire $Y = g(X)$ a la densité f_Y donnée par

$$(5.4.1) \quad f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) & \text{si } y \in g(\mathbb{R}) \\ 0 & \text{sinon.} \end{cases}$$

Ici, $g^{-1}(y)$ est l'**inverse continue à droite** de la fonction g :

$$(5.4.2) \quad g^{-1}(y) := \inf\{x \in \mathbb{R} : g(x) > y\}.$$

DÉMONSTRATION. On a que

$$\begin{aligned}F_Y(y) &= \mathbb{P}\{g(X) \leq y\} \\ &= \mathbb{P}\{X \leq g^{-1}(y)\} \\ &= F_X(g^{-1}(y)).\end{aligned}$$

Il suit immédiatement, par la règle de la dérivation en chaîne, que pour toutes valeurs de $y \in g(\mathbb{R})$, on a que

$$f_Y(y) = F'_Y(y) = F'_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y).$$

□

La proposition suivante est un exemple crucial de l'application de la proposition 5.12 :

PROPOSITION 5.13. Soit $X \sim \mathcal{U}(0, 1)$ une variable aléatoire uniforme sur l'intervalle $(0, 1)$. Soit Y une autre variable aléatoire, avec fonction de répartition F , et supportée sur (a, b) (avec $-\infty \leq a < b \leq \infty$).

Alors, $Y \sim F^{-1}(X)$ – c'est à dire que $F^{-1}(X)$ a la même distribution que Y .

9. Comme mentionné précédemment, le théorème existe pour justifier cette affirmation, mais il dépasse le cadre de ce cours.

DÉMONSTRATION. Soit $g = F^{-1}$ avec F^{-1} définie comme dans (5.4.2).

Puisque F est non-décroissante, $g = F^{-1}$ est non-décroissante ; on a que pour tout $x \in (0, 1)$, $g(x) = \inf\{y \in \mathbb{R} : F(y) > x\}$, et donc forcément pour $x' \geq x$, $g(x') \geq g(x)$. On a aussi que g est continue puisque F est continue et non-décroissante.

On considère maintenant

$$g^{-1}(x) = \inf\{y \in (0, 1) : \inf\{z \in \mathbb{R} : F(z) > y\} > x\}$$

Si $-\infty < a$, et qu'on a $x \leq a$. Alors $g^{-1}(x) = 0$. Si $b < \infty$ et qu'on a $x \geq b$, Alors $g^{-1}(x) = 1$.

Si $x \in (a, b)$, on a que $g^{-1}(x) = F(x)$.

Mais alors, on a que la densité de $F^{-1}(X)$ est donnée par

$$f_{F^{-1}(X)}(x) = f_X(F(x)) \cdot \frac{d}{dx}F(x) = F'(x),$$

soit exactement la densité de Y .

Donc, $Y \sim F^{-1}(X)$. □

EXEMPLE 5.16. Soit $X \sim \mathcal{U}(-\frac{\pi}{2}, \frac{\pi}{2})$. Trouver la distribution de $\tan X$.

SOLUTION. Pour $x \in (-\pi/2, \pi/2)$, on a que $PR\{X \leq x\} = \frac{1}{2} + \frac{x}{\pi}$.

Bien sûr,

$$\mathbb{P}\{\tan X \leq x\} = \mathbb{P}\{X \leq \arctan x\} = \frac{1}{2} + \frac{1}{\pi} \arctan x$$

Donc, $\tan X \sim \mathfrak{C}(0, 1)$.

EXEMPLE 5.17. Soit $X \sim \mathcal{N}(0, 1)$. On considère $Y = e^X$. Trouver la distribution de Y .

SOLUTION. Ici $g(x) = e^x$ est évidemment non-décroissante, continue, et différentiable. Sur $(0, \infty)$, la fonction $g^{-1}(y) = \log y$ est donc celle qui nous intéresse.

Par la proposition 5.12, on trouve que

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y) = \frac{1}{\sqrt{2\pi}} \frac{e^{-\log^2 y/2}}{y}, \quad y > 0.$$

Cette distribution est appelée la distribution **log-normale**, puisque le logarithme de Y est distribué de façon normale.

5.5. Bonus : les distributions mixtes.

Nous venons de passer le chapitre à discuter de variables aléatoires dont la fonction de répartition est continue. Au chapitre 4, on s'était concentré sur les variables aléatoires discrètes, dont la fonction de répartition est une fonction dite « en escaliers ».

Dans cette courte section, on explique ce qui se passe dans le cas le plus général : les variables aléatoires à distribution mixte.

Ici, F peut être discontinue, mais F n'est pas pour autant une fonction en escaliers.

EXEMPLE 5.18. Soit X une variable aléatoire dont la fonction de répartition est donnée par :

$$F(x) = \begin{cases} 0 & \text{si } x < -1/3 \\ \frac{1}{3} + x & \text{si } -1/3 \leq x \leq 0 \\ 1 - \frac{1}{3}e^{-x} & \text{si } x > 0 \end{cases}$$

On va définir $V_d = \{x \in \mathbb{R} : p_X(x) > 0\}$ l'ensemble des atomes de X – ce sont l'ensemble des valeurs de la distribution de X qui ont, à elles seules, une probabilité positive.

On va définir également $V_c = \{x \in \mathbb{R} : F'(x) > 0\}$ l'ensemble des valeurs où F est continue et différentiable, et où $F'(x) > 0$.

Pour l'exemple 5.18, $V_d = \{0\}$ – la distribution de X n'a qu'un atome, en 0. $V_c = (-1/3, 0) \cup (0, \infty)$.

5.5.1. Densité et fonction de masse. La densité et la fonction de masse d'une variable aléatoire sont toujours définies, pour toute variable aléatoire – il s'avère simplement que dans le cas des variables aléatoires discrète, la densité est nulle partout où elle est définie, et dans le cas des variables continues, la fonction de masse est nulle partout.

Pour les variables aléatoires à distribution mixte, par contre, les deux sont non-triviales. En effet, on a que la densité f d'une variable X quelconque est donnée par $f = F'$ partout où F' est différentiable.

Cependant, on n'a pas immédiatement la propriété de normalisation comme avant. À la place, on a que, pour une variable aléatoire X quelconque,

$$(5.5.1) \quad \int_{V_c} f(x)dx = \mathbb{P}\{X \in V_c\}.$$

De façon analogue, la propriété de normalisation ne tient plus pour la fonction de masse de probabilités non plus. On est réduit.e.s à

$$(5.5.2) \quad \sum_{v \in V_d} p_X(v) = \mathbb{P}\{X \in V_d\}.$$

La vraie condition de normalisation est donc

$$(5.5.3) \quad \sum_{v \in V_d} p_X(v) + \int_{V_c} f(x)dx = 1.$$

5.5.2. L'espérance de variables mixtes. Pareillement, le calcul de l'espérance devient mixte : pour n'importe quelle fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ intégrable, on aura

$$(5.5.4) \quad \mathbb{E}[X] = \sum_{v \in V_d} g(v)p_X(v) + \int_{V_c} g(x)f(x)dx.$$

La section B.12 présente brièvement la notion de l'intégrale de Stieltjes, qui permet de prouver rigoureusement ces formules.

Dans notre exemple, on aurait que

$$\begin{aligned}\mathbb{E}[X] &= \int_{-1/3}^0 x dx + \frac{1}{3} \int_0^\infty xe^{-x} dx + 0 \times \mathbb{P}\{X = 0\} \\ &= \left[\frac{x^2}{2} \right]_{-1/3}^0 + \frac{1}{3} \\ &= \frac{1}{3} - \frac{1}{18} = \frac{5}{18}.\end{aligned}$$

5.6. Exercices

EXERCICE 5.1. Un système peut fonctionner pendant une durée de temps de X mois, où X est une variable aléatoire de densité

$$f(x) = \begin{cases} Cxe^{-x/2} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

Quelle est la probabilité que ce système fonctionne au moins 5 mois ?

EXERCICE 5.2. La densité de X , la durée de vie (en heures) d'un appareil électronique, est donnée par

$$f(x) = \begin{cases} \frac{10}{x^2} & \text{si } x > 10 \\ 0 & \text{sinon.} \end{cases}$$

- (a) Trouver $\mathbb{P}\{X > 20\}$.
- (b) Trouver la fonction de répartition de X .
- (c) Quelle est la probabilité que, pour 6 tels appareils, au moins 3 fonctionnent pendant au moins 15 heures ?

EXERCICE 5.3. La densité d'une variable aléatoire X est donnée par

$$f(x) = \begin{cases} a + bx^2 & \text{pour } x \in (0, 1) \\ 0 & \text{sinon.} \end{cases}$$

Si $\mathbb{E}[X] = \frac{3}{5}$, trouver a et b .

EXERCICE 5.4. Les trains à destination de Québec arrivent à intervalle de 15 minutes à partir de 7h. Les trains à destination de Toronto arrivent aussi à la gare à intervalles de 15 minutes, mais à partir de 7h 05.

- (a) Si une certaine voyageuse arrive à la gare à un moment aléatoire distribué uniformément entre 7h et 8h et qu'elle prend le premier train arrivé, quelle est la probabilité qu'elle se rende à Québec ?
- (b) Et si elle arrive à un temps uniformément distribué entre 7h 10 et 8h 10 ?

EXERCICE 5.5. On choisit un point uniformément sur un bâton de longueur L , et on casse le bâton à ce point. Trouver la probabilité que le segment le plus long soit au moins quatre fois la longueur du segment le plus court.

EXERCICE 5.6. Les précipitations annuelles (en centimètres) dans une région sont distribuées normalement, avec $\mu = 100$ et $\sigma = 10$. Quelle est la probabilité que cela prenne exactement 10 ans avant qu'il tombe plus de 125 cm en une année ? Quelles hypothèses doit-on faire pour résoudre le problème ?

EXERCICE 5.7. Réaliser 1000 lancers d'un dé équilibré. Calculer une approximation de la probabilité que le 6 apparaisse entre 150 et 200 fois inclusivement. Sachant que le 6 est apparu 200 fois, quelle est la probabilité que le 5 apparaîtra moins de 150 fois ?

EXERCICE 5.8. (a) Une caserne de pompiers est construite sur une route de longueur A avec A finie. La position d'un incendie le long de la route est une variable aléatoire X distribuée uniformément sur $(0, A)$.

Trouver la valeur c pour la position de la caserne qui minimise l'espérance $\mathbb{E}[|X - c|]$ de la distance à parcourir pour se rendre à l'incendie.

- (b) Si la route est demi-infinie (c.à.d. qu'elle s'étend de 0 à $+\infty$ dans une direction) et que la position X de l'incendie est de loi exponentielle avec paramètre λ , trouver la position c qui minimise $\mathbb{E}[|X - c|]$.

REMARQUE. En général, la valeur c qui minimise $\mathbb{E}[|X - c|]$ est la médiane de la distribution de X . Pouvez-vous le prouver ?

EXERCICE 5.9. Si $X \sim \mathcal{U}(-1, 1)$,

- (a) Trouver $\mathbb{P}\{|X| > \frac{1}{2}\}$.
 (b) Trouver la fonction de répartition de $|X|$.

EXERCICE 5.10. Trouver la densité de $Y = R \sin \theta$ où $\theta \sim \mathcal{U}(-\pi/2, \pi/2)$ et R est une constante réelle.

EXERCICE 5.11. Soit X une variable aléatoire supportée sur l'intervalle $[0, L]$ (c.à.d. que $\mathbb{P}\{0 \leq X \leq L\} = 1$).

Montrer que peu importe la distribution de X , on a toujours

$$\text{Var}[X] \leq \frac{L^2}{4}.$$

Indice : Commencer par montrer que $\mathbb{E}[X^2] \leq L\mathbb{E}[X]$, puis montrer que

$$\text{Var}[X] \leq L^2\alpha(1 - \alpha)$$

avec $\alpha = \mathbb{E}[X]/L$.

EXERCICE 5.12. Soit $Z \sim \mathcal{N}(0, 1)$. Montrer que

- (a) $\mathbb{P}\{Z > x\} = \mathbb{P}\{Z < -x\}$;
 (b) $\mathbb{P}\{|Z| > x\} = 2\mathbb{P}\{Z > x\}$;
 (c) $\mathbb{P}\{|Z| \leq x\} = 2\mathbb{P}\{Z \leq x\} - 1$.

EXERCICE 5.13. Soit $Z \sim \mathcal{N}(0, 1)$ et soit $g : \mathbb{R} \rightarrow \mathbb{R}$ différentiable partout.

- (a) Montrer que $\mathbb{E}[g'(Z)] = \mathbb{E}[Zg(Z)]$.
- (b) Montrer que $\mathbb{E}[Z^{n+1}] = n\mathbb{E}[Z^{n-1}]$.
- (c) Montrer que $\mathbb{E}[Z^{2k+1}] = 0$ pour tout k .
- (d) Trouver $\mathbb{E}[Z^4]$.

EXERCICE 5.14. Montrer la proposition 5.7 :

PROPOSITION. Soit X une variable aléatoire de loi exponentielle avec paramètre $\lambda > 0$, et soit $c > 0$ une constante réelle.

Alors, $Y = cX$ est une variable aléatoire de loi exponentielle avec paramètre λ/c .

EXERCICE 5.15. Montrer la proposition 5.9 :

PROPOSITION. Soit X une variable aléatoire suivant une loi Gamma de paramètres $\lambda > 0$, $\alpha > 0$. Soit $c > 0$ une constante réelle positive.

Alors, cX est une variable aléatoire de loi Gamma avec paramètres $(\lambda/c, \alpha)$.

EXERCICE 5.16. On prend un bâton de longueur 1, qu'on marque d'un point à distance $d \in (0, 1)$ de l'une des extrémités. Puis, on casse le bâton.

La position X de la fracture est une variable aléatoire ayant la densité

$$f(x) = \begin{cases} 6x(1-x) & \text{si } 0 < x < 1 \\ 0 & \text{ailleurs.} \end{cases}$$

- (a) Montrer que $1 - X$ a la même distribution que X .
- (b) Soit $L_p(x)$ la longueur du morceau qui porte la marque (en p) lorsqu'on casse le bâton en x . Donner une expression pour $L_p(x)$.
- (c) Trouver $\mathbb{E}[L_p(X)]$.

EXERCICE 5.17. Montrer la proposition 5.10 :

PROPOSITION. Soit X une variable aléatoire de Cauchy de paramètres θ, σ , et soit $X' = aX + b$ avec $a \neq 0$ et $b \in \mathbb{R}$.

Alors, X' est une variable aléatoire de Cauchy de paramètres $\theta' = a\theta + b$ et $\sigma' = |a|\sigma$.

EXERCICE 5.18. Soit X une variable aléatoire discrète non-négative de support $V = \{v_1, v_2, \dots\}$ (on assume que $v_1 < v_2 < \dots$), avec fonction de masse de probabilité p_X , et soit $\bar{F}(x) = \mathbb{P}\{X > x\}$ sa fonction de répartition complémentaire.

On suppose également que $\lim_{x \rightarrow \infty} x\bar{F}(x) = 0$.

- (a) Montrer que

$$\int_0^\infty \bar{F}(t)dt = v_1 + \sum_{k=1}^{\infty} (v_{k+1} - v_k)\bar{F}(v_k).$$

(b) Montrer que

$$\bar{F}(v_k) - \bar{F}(v_{k+1}) = p_X(v_{k+1})$$

(c) Le lemme suivant énonce la technique de sommation d'Abel :

LEMME. Soient $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites de nombres, et soient $A_n = a_{n+1} - a_n$ et $B_n = b_{n+1} - b_n$ les différences de termes consécutifs.

Alors, pour $N > M$,

$$\sum_{n=M}^N a_n B_n = a_{N+1} b_{N+1} - a_M b_M - \sum_{n=M}^N A_n b_{n+1}.$$

Utiliser ce lemme pour démontrer que, pour X une variable aléatoire discrète non-négative,

$$\mathbb{E}[X] = \int_0^\infty \bar{F}(t) dt.$$

Chapitre 6

Variables aléatoires simultanées

Au cours des chapitres précédents, nous avons établi les notions de base de variables aléatoires. La plupart du temps, nous ne considérons qu'une variable à la fois.

Dans ce chapitre, nous nous penchons sur des espaces de probabilités sur lesquels on définira plusieurs variables aléatoires simultanément. *Avertissement* : il sera très important

d'être à l'aise avec les notions de calcul intégral à multiples variables. Si ce n'est pas déjà fait, il serait pertinent de réviser ces notions.

6.1. Fonctions de répartition et densités jointes

DÉFINITION 6.1 (Fonction de répartition jointe, marginale). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. et $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires quelconques.

La **fondation de répartition jointe** de X et Y est donnée par

$$(6.1.1) \quad F(x, y) = \mathbb{P}\{X \leq x, Y \leq y\}.$$

La **fondation de répartition marginale** de X est donnée par

$$(6.1.2) \quad F_X(x) = \mathbb{P}\{X \leq x\} = \lim_{y \rightarrow +\infty} F(x, y).$$

La fondation de répartition marginale de Y est donnée de façon similaire par

$$F_Y(y) = \mathbb{P}\{Y \leq y\} = \lim_{x \rightarrow +\infty} F(x, y).$$

Il s'agit simplement de la fondation de répartition de la variable X toute seule.

REMARQUE. On peut généraliser ces notion à n variables, ou même à un nombre dénombrable de variables. Voyez-vous comment ?

On voit qu'avec une telle définition et la formule du crible, on a que pour X, Y deux variables aléatoires avec fondation de répartition jointe $F(x, y)$, on trouve que

$$(6.1.3) \quad \mathbb{P}\{a < X \leq b, c < Y \leq d\} = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$

6.1.1. Le cas des variables aléatoires discrètes. On peut définir une fondation de masse de probabilité jointe :

DÉFINITION 6.2. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires. Alors leur fondation de masse de probabilité jointe $p : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ est

$$(6.1.4) \quad p(x, y) = \mathbb{P}\{X = x, Y = y\}$$

Lorsqu'une des deux variables aléatoires est continues, évidemment $p(x, y) = 0$ partout. Cette fondation n'est donc vraiment utile que lorsque X et Y ont des atomes qui peuvent se réaliser simultanément.

En particulier, on utilisera beaucoup la fondation de masse de probabilité lorsque X et Y sont deux variables discrètes.

EXEMPLE 6.1. On tire au hasard et sans remise trois boules d'une urne qui contient 3 boules bleues, 4 boules blanches et 5 boules rouges.

Soit X le nombre de boules bleues, et Y le nombre de boules rouges, trouver $p(x, y)$ la fonction de masse jointe de X et Y .

SOLUTION. Avec la convention que $\binom{n}{k} = 0$ si $k < 0$ ou si $k > n$, on trouve :

$$p(i, j) = \frac{\binom{3}{i} \binom{5}{j} \binom{4}{3-i-j}}{\binom{12}{3}}$$

On peut organiser les résultats dans une matrice, simplement :

$$(p(i, j))_{(i,j)} = \begin{pmatrix} \frac{1}{55} & \frac{3}{22} & \frac{2}{11} & \frac{1}{22} \\ \frac{9}{55} & \frac{3}{3} & \frac{3}{22} & 0 \\ \frac{11}{3} & \frac{11}{3} & \frac{22}{22} & 0 \\ \frac{55}{3} & \frac{44}{44} & 0 & 0 \\ \frac{1}{220} & 0 & 0 & 0 \end{pmatrix}$$

Dans le cas où on a deux variables aléatoires X, Y discrètes avec fonction de masse de probabilité jointe $p(x, y)$, si V_X est le support de X et V_Y est celui de Y , alors on trouve que les fonctions de masse de probabilité marginales pour X et Y sont données par :

$$(6.1.5) \quad p_X(x) = \sum_{y \in V_Y} p(x, y); \quad p_Y(y) = \sum_{x \in V_X} p(x, y).$$

EXEMPLE 6.2. Dans notre exemple précédent, on voit que

$$p_X(0) = \frac{21}{55}, \quad p_X(1) = \frac{27}{55}, \quad p_X(2) = \frac{27}{220}, \quad p_X(3) = \frac{1}{220}.$$

En comparant avec la fonction de masse d'une variable hypergéométrique de paramètres $n = 3, N = 12, M = 3$, on voit bien que $X \sim HG(3, 12, 3)$. C'est normal. Si X est le nombre de boules bleues, on se fiche de savoir combien de boules rouges il y a eu. Le nombre de boules bleues pigées sera une hypoérgométrique, comme on avait vu à la section 4.6.5.

De même, on pourrait voir, en comparant, que $Y \sim HG(3, 12, 5)$.

6.1.2. Le cas des variables continues. Si X et Y sont des variables aléatoires continues, alors leur fonction de répartition jointe F est aussi continue sur \mathbb{R}^2 , et non-décroissante.

De façon analogue à ce qu'on avait fait à la section 5.1 pour une seule variable, on peut se pencher sur la probabilité que le couple (X, Y) se trouve « proche » du point (x, y) , en s'imaginant que la probabilité que (X, Y) se retrouve dans le petit carré $(x - h, x + h] \times (y - h, y + h]$ est proportionnelle à la surface $4h^2$ de ce petit carré, on trouve que le facteur de proportionnalité tend vers

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}\{x - h < X \leq x + h, y - h < Y \leq y + h\}}{4h^2} = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

On fait la définition suivante :

DÉFINITION 6.3. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soient $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires sur cet espace.

Partout où F est différentiable en x et en y , on définit f la **densité jointe** de X et Y par :

$$(6.1.6) \quad f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^1}{\partial y \partial x} F(x, y),$$

soit la dérivée mixte de F en x puis en y – l'ordre de différentiation n'est pas important.

Il suit de cette définition que

$$(6.1.7) \quad \mathbb{P}\{X \in (a, b), Y \in (c, d)\} = \int_a^b \int_c^d f(x, y) dy dx.$$

De façon plus générale, si $B \subseteq \mathbb{R}^2$ est un ouvert de \mathbb{R}^2 ,

$$(6.1.8) \quad \mathbb{P}\{(X, Y) \in B\} = \iint_B f(x, y) dy dx.$$

Et l'ordre d'intégration n'a pas d'importance.

Il suit la condition de normalisation :

$$(6.1.9) \quad 1 = \mathbb{P}\{(X, Y) \in \mathbb{R}^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx.$$

Finalement, on peut obtenir les densités marginales pour X et Y de façon analogue au cas discret ;

$$(6.1.10) \quad f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy; \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Encore une fois, ces densités ne sont que les densité respectives de X et Y , si on s'était intéressé seulement à l'une ou l'autre de ces variables sans se soucier de ce qui arrive à l'autre.

EXEMPLE 6.3. Soient X, Y un couple de variables aléatoires continues avec densité jointe

$$f(x, y) = \begin{cases} Ce^{-x}e^{-2y} & \text{si } x, y > 0 \\ 0 & \text{ailleurs.} \end{cases}$$

- (a) Trouver C .
- (b) Trouver $\mathbb{P}\{X > 1, Y < 1\}$.
- (c) Trouver $\mathbb{P}\{X < Y\}$.
- (d) Trouver $\mathbb{P}\{X < a\}$.

SOLUTION. (a) Pour trouver C , il suffit d'utiliser la condition de normalisation :

$$\begin{aligned} 1 &= C \int_0^{\infty} \int_0^{\infty} e^{-x}e^{-2y} dy dx \\ &= C \cdot \int_0^{\infty} e^{-x} dx \cdot \int_0^{\infty} e^{-2y} dy \\ &= C \cdot 1 \cdot \frac{1}{2}. \end{aligned}$$

On conclue que $C = 2$.

(b) On a que

$$\begin{aligned}
 \mathbb{P}\{X > 1, Y < 1\} &= \int_1^\infty \int_0^1 f(x, y) dy dx \\
 &= 2 \int_1^\infty e^{-x} dx \cdot \int_0^1 e^{-2y} dy \\
 &= 2e^{-1} \cdot \left[-\frac{1}{2}e^{-2y} \right]_0^1 \\
 &= 2e^{-1} \cdot \frac{1}{2}(1 - e^{-2}) \\
 &= e^{-1}(1 - e^{-2})
 \end{aligned}$$

(c) On veut $(X, Y) \in \{(x, y) : x < y\}$.

On a que

$$\begin{aligned}
 \mathbb{P}\{X < Y\} &= \int_0^\infty \int_x^\infty f(x, y) dy dx \\
 &= \int_0^\infty e^{-x} \cdot \left(\int_x^\infty 2e^{-2y} dy \right) dx \\
 &= \int_0^\infty e^{-x} \cdot e^{-2x} dx \\
 &= \int_0^\infty e^{-3x} dx \\
 &= \frac{1}{3}.
 \end{aligned}$$

(d) La densité marginale de X est donnée par

$$f_X(x) = \int_{-\infty}^\infty f(x, y) dy = e^{-x} \cdot \int_0^\infty 2e^{-2y} dy = e^{-x}, \quad (x > 0).$$

Donc, pour $a \leq 0$, $\mathbb{P}\{X < a\} = 0$.

Pour $a > 0$,

$$\mathbb{P}\{X < a\} = \int_0^a e^{-x} dx = 1 - e^{-a}.$$

On constate que $X \sim \mathfrak{E}(1)$ et $Y \sim \mathfrak{E}(2)$.

EXEMPLE 6.4. Soit X, Y un couple de variables aléatoires continues avec densité

$$f(x, y) = \begin{cases} C & \text{si } x^2 + y^2 < R^2 \\ 0 & \text{si } x^2 + y^2 \geq R^2. \end{cases}$$

(a) Trouver C .

(b) Calculer les densités marginales de X et Y .

(c) Calculer la probabilité que la distance du point (X, Y) à l'origine soit plus petite que a .

SOLUTION. (a) On a que

$$1 = C \iint_{\{(x,y):x^2+y^2 < R^2\}} dydx = C\pi R^2.$$

Donc, $C = \frac{1}{\pi R^2}$.

(b) Pour $x \in (-R, R)$, on a que

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \frac{1}{\pi R^2} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \\ &= \frac{2}{\pi R^2} \sqrt{R^2 - x^2} \quad (x \in (-R, R)) \end{aligned}$$

Par symétrie du problème, on voit que

$$f_Y(y) = \frac{2}{\pi R^2} \sqrt{R^2 - y^2}, \quad (y \in (-R, R)).$$

(c) On cherche $\mathbb{P}\{X^2 + Y^2 < a^2\}$.

On a bien sûr que

$$\mathbb{P}\{X^2 + Y^2 < a^2\} = \iint_{\{(x,y):x^2+y^2 < a^2\}} f(x, y) dydx.$$

Pour $a > R$, c'est évidemment 1.

Pour $a < R$, c'est bien sûr

$$\begin{aligned} \mathbb{P}\{X^2 + Y^2 < a^2\} &= \frac{1}{\pi R^2} \iint_{\{(x,y):x^2+y^2 < a^2\}} dydx \\ &= \frac{\pi a^2}{\pi R^2} \\ &= \left(\frac{a}{R}\right)^2. \end{aligned}$$

EXEMPLE 6.5. Soient X, Y deux variables aléatoires avec densité jointe

$$f(x, y) = e^{-x-y}, \quad x, y > 0.$$

Déterminer la densité de X/Y .

SOLUTION. Pour résoudre ce problème on doit d'abord trouver la fonction de répartition de X/Y . Premier constat : X/Y est toujours positive – en effet, puisque X et Y sont positives, X/Y le sera aussi.

Supposons que $u > 0$. Alors,

$$\begin{aligned}\mathbb{P}\{X/Y \leq u\} &= \mathbb{P}\{Y \geq X/u\} \\ &= \int_0^\infty \int_{x/u}^\infty e^{-x-y} dy dx \\ &= \int_0^\infty e^{-x} \left(\int_{x/u}^\infty e^{-y} dy \right) dx \\ &= \int_0^\infty e^{-x} e^{-x/u} dx \\ &= \int_0^\infty e^{-x(1+\frac{1}{u})} dx \\ &= \frac{u}{u+1}.\end{aligned}$$

Et on déduit que $f_{X/Y}(u) = \frac{d}{du} \mathbb{P}\{X/Y \leq u\}$, soit

$$f_{X/Y}(u) = \frac{1}{(1+u)^2}, \quad u > 0.$$

Évidemment le résultat probablement le plus intriguant ici est la fait que, puisque $\mathbb{P}\{X/Y > u\} = O(1/u)$, on a nécessairement que $\mathbb{E}[X/Y] = +\infty$.

6.1.3. Le cas mixte. Supposons que X soit une variable aléatoire discrète, et Y une variable aléatoire continue. Puisque X est discrète, alors $\frac{\partial^2}{\partial x \partial y} F(x, y) = 0$ partout où F est différentiable. Mais puisque Y est continue, de même, $p(x, y) = 0$ partout sur \mathbb{R}^2 .

Si on suppose que $F(x, y)$ est la fonction de répartition jointe de X et Y (toujours bien définie), et que V est le support de X , alors la solution la plus simple consiste à considérer la famille de fonctions

$$(6.1.11) \quad (f_v)_{v \in V} : f_v(y) := \frac{\partial}{\partial y} \left(F(v, y) - \lim_{x \rightarrow v^-} F(x, y) \right).$$

On aura alors que

$$(6.1.12) \quad \mathbb{P}\{X = v, Y \leq x\} = \int_{-\infty}^x f_v(t) dt,$$

ou, plus généralement, que

$$(6.1.13) \quad \mathbb{P}\{X \in A, Y \in B\} = \sum_{v \in A} \int_B f_v(t) dt.$$

La condition de normalisation dans ce cas est donc donnée par

$$(6.1.14) \quad 1 = \mathbb{P}\{(X, Y) \in \mathbb{R}^2\} = \sum_{v \in V} \int_{-\infty}^\infty f_v(t) dt.$$

La fonction de masse de probabilités marginale de X sera donnée par

$$(6.1.15) \quad p_X(v) = \int_{-\infty}^\infty f_v(x) dx,$$

et la densité marginale de Y est donnée par

$$(6.1.16) \quad f_Y(y) = \sum_{v \in V} f_v(y).$$

EXEMPLE 6.6. Soient X, Y deux variables aléatoires, avec X discrète et Y continue, dont la distribution jointe est décrite par les fonctions

$$f_k(x) = e^{-2x} x^k k!, \quad k \in \mathbb{N}_0, x > 0.$$

Trouver les distributions marginales de X et de Y .

SOLUTION. On sait que

$$\begin{aligned} p_X(k) &= \int_0^\infty f_k(x) dx \\ &= \frac{1}{k!} \int_0^\infty x^k e^{-2x} dx \\ &= \frac{1}{k! 2^{k+1}} \int_0^\infty u^k e^{-u} du \\ &= \frac{1}{k! 2^{k+1}} \Gamma(k+1) \\ &= \frac{1}{2^{k+1}}. \end{aligned}$$

Il suit que $X \sim Geom(1/2) - 1$.

D'un autre côté,

$$\begin{aligned} f_Y(x) &= \sum_{k=0}^{\infty} f_k(x) \\ &= e^{-2x} \sum_{k=0}^{\infty} \frac{x^k}{k!} \\ &= e^{-x} \end{aligned}$$

lorsque $x > 0$.

Donc, $Y \sim \mathfrak{E}(1)$.

Nous allons voir à la section 6.4 qu'il sera plus souvent facile de traiter les cas mixtes en conditionnant d'abord par la valeur de la variable discrète.

6.1.4. Généralisation à n variables aléatoires. Comme mentionné précédemment, on peut généraliser les notions introduites plus haut au cas où on a n variables aléatoires – supposons qu'on a $X_1, X_2, X_3, \dots, X_n$.

Alors, la fonction de répartition jointe sera

$$(6.1.17) \quad F(x_1, \dots, x_n) = \mathbb{P}\{X_1 \leq x_1, \dots, X_n \leq x_n\}.$$

Dans le cas où les variables aléatoires sont discrètes, on a la fonction de masse jointe

$$(6.1.18) \quad p(x_1, \dots, x_n) = \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\},$$

avec les fonctions de masse marginales

$$(6.1.19) \quad p_{X_i}(x) = \mathbb{P}\{X_i = x\} = \sum_{v=(v_j)_{j \leq n}: v_i=x} p(v).$$

Dans le cas où les variables aléatoires sont continues, on a la fonction densité jointe

$$(6.1.20) \quad f(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \dots, x_n).$$

On a les fonctions de densité marginales

$$(6.1.21) \quad f_{X_i}(x_i) = \iiint \cdots \int_{\mathbb{R}^{n-1}} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

De façon complètement analogue, on obtient la probabilité que le vecteur (X_1, \dots, X_n) soit dans la région $C \subseteq \mathbb{R}^n$ en intégrant la densité jointe sur la région C .

En fait, on peut voir $\mathbf{X} = (X_1, \dots, X_n)$ comme un vecteur aléatoire dans \mathbb{R}^n , avec sa densité de probabilité $f : \mathbb{R}^n \rightarrow \mathbb{R}$ et

$$(6.1.22) \quad \mathbb{P}\{\mathbf{X} \in C\} = \int_C f(\mathbf{x}) d^n \mathbf{x},$$

ou, si \mathbf{X} est un vecteur de variables discrètes supporté sur V et $C \subseteq V$, sa fonction de masse $p : \mathbb{R}^n \rightarrow \mathbb{R}$ telle que $p(\mathbf{x}) = \mathbb{P}\{\mathbf{X} = \mathbf{x}\}$ et

$$(6.1.23) \quad \mathbb{P}\{\mathbf{X} \in C\} = \sum_{\mathbf{v} \in C} p(\mathbf{v}).$$

Si $J \subseteq \{1, \dots, n\} =: I$ est un sous-ensemble des indices de 1 à n , on peut s'intéresser au sous-vecteur $\mathbf{X}_J := (X_j)_{j \in J}$.

Dans le cas où les X_i sont des variables aléatoires continues, on peut trouver la densité marginale de \mathbf{X}_J :

$$(6.1.24) \quad f_{\mathbf{X}_J}(\mathbf{x}_J) = \iiint \cdots \int_{\mathbb{R}^{n-|J|}} f(\mathbf{x}) d^{n-|J|} \mathbf{x}_{I \setminus J},$$

où l'intégrale est sur toutes les variables d'indices non dans J .

Dans le cas où les X_i sont des variables aléatoires discrètes, on trouve la fonction de masse marginale de \mathbf{X}_J :

$$(6.1.25) \quad p_{\mathbf{X}_J}(\mathbf{v}_J) = \sum_{\mathbf{v}_{I \setminus J}} p(\mathbf{v}),$$

où la somme est prise sur toutes les variables d'indices non dans J .

6.2. Variables aléatoires indépendantes

DÉFINITION 6.4 (Indépendance). Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités. Soient $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires.

On dit que X est **indépendante** de Y lorsque, pour tous¹ $A, B \subseteq \mathbb{R}$, on a

$$(6.2.1) \quad \mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\}.$$

1. Pas tous, mais les exceptions sont pathologiques et de peu d'intérêt pour nous présentement. En temps normal on exigerait que A et B soient parties de la tribu des Boréliens sur \mathbb{R} – vous pouvez fouiller si ça vous chante.

Autrement dit, X et Y sont indépendants si tous les événements de la forme $\{X \in A\}$ sont respectivement indépendants de tous les événements de la forme $\{Y \in B\}$.

Cette définition a tout plein de sens – on dit que deux variables aléatoires sont indépendantes si et seulement si peu importe ce qu'on sait sur la valeur de l'une, ça ne peut rien nous dire sur la valeur de l'autre.

On montre immédiatement un lemme qui sera très pratique :

LEMME 6.1. *Soient X, Y deux variables aléatoires indépendantes. Soient $g, h : \mathbb{R} \rightarrow \mathbb{R}$. Alors, $g(X)$ et $h(Y)$ sont aussi des variables aléatoires indépendantes.*

DÉMONSTRATION. Soit Puisque X et Y sont indépendantes, alors pour tout $A, B \subseteq \mathbb{R}$, on a que

$$\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\}.$$

Prenons maintenant $A, B \subseteq \mathbb{R}$ quelconques :

$$\begin{aligned} \mathbb{P}\{g(X) \in A, h(Y) \in B\} &= \mathbb{P}\{X \in g^{-1}(A), Y \in h^{-1}(B)\} \\ &= \mathbb{P}\{X \in g^{-1}(A)\} \mathbb{P}\{Y \in h^{-1}(B)\} \\ &= \mathbb{P}\{g(X) \in A\} \mathbb{P}\{h(Y) \in B\}. \end{aligned}$$

Donc, $g(X)$ et $h(Y)$ sont indépendantes. □

Évidemment, en général il ne sera pas pratique de travailler toujours avec *tous* les ensembles imaginables de \mathbb{R} en même temps. Pour nous aider, nous avons cette proposition :

PROPOSITION 6.1. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires avec fonction de répartition jointe $F(x, y) = \mathbb{P}\{X \leq x, Y \leq y\}$.*

Alors, X et Y sont indépendantes si et seulement si

$$(6.2.2) \quad F(x, y) = F_X(x)F_Y(y),$$

où F_X et F_Y sont les fonction de répartition marginales de X et Y respectivement.

DÉMONSTRATION. (\Rightarrow) : Dans cette direction, c'est facile – puisque X et Y sont indépendantes, on a que pour tout $A, B \subseteq \mathbb{R}$, $\mathbb{P}\{X \in A, Y \in B\} = \mathbb{P}\{X \in A\} \mathbb{P}\{Y \in B\}$. On ne fait que choisir $A = (-\infty, x]$ et $B = (-\infty, y]$ pour finir la preuve.

(*Leftarrow*) : Dans ce sens, c'est plus compliqué, mais on peut montrer que toutes les parties de \mathbb{R} peuvent être grossièrement découpées en intervalles, et que comme on peut calculer les probabilités d'intervalles à l'aide des fonctions de répartitions, celles-ci déterminent toutes les probabilités possibles, etc. □

PROPOSITION 6.2. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires discrètes avec fonction de masse jointe p et fonctions de masse marginales p_X et p_Y respectivement.*

Alors, X et Y sont indépendantes si et seulement si pour tout $x, y \in \mathbb{R}$,

$$(6.2.3) \quad p(x, y) = p_X(x)p_Y(y).$$

DÉMONSTRATION. D'abord, si X et Y sont indépendantes, alors en particulier,

$$p(x, y) = \mathbb{P}\{X = x, Y = y\} = \mathbb{P}\{X = x\} \mathbb{P}\{Y = y\} = p_X(x)p_Y(y).$$

Dans l'autre sens, supposons que $p(x, y) = p_X(x)p_Y(y)$. Alors, si F est la fonction de répartition jointe de X, Y , que V est le support de X et V' celui de Y ,

$$\begin{aligned} F(x, y) &= \sum_{v \in V: v \leq x} \sum_{v' \in V': v' \leq y} p(v, v') \\ &= \sum_{v \in V: v \leq x} \sum_{v' \in V': v' \leq y} p_X(v)p_Y(v') \\ &= \left(\sum_{v \in V: v \leq x} p_X(v) \right) \left(\sum_{v' \in V': v' \leq y} p_Y(v') \right) \\ &= F_X(x)F_Y(y). \end{aligned}$$

Par la proposition 6.1, les variables X et Y sont donc indépendantes. \square

PROPOSITION 6.3. *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et $X, Y : \Omega \rightarrow \mathbb{R}$ deux variables aléatoires continues avec densité jointe f et densités marginales.*

Alors, X et Y sont indépendantes si et seulement si

$$(6.2.4) \quad f(x, y) = f_X(x)f_Y(y).$$

DÉMONSTRATION. Si on suppose que X et Y sont indépendantes, alors par la proposition 6.1, on doit avoir que si F, F_X et F_Y sont respectivement les fonctions de répartition jointe, marginale de X et marginale de Y , alors

$$F(x, y) = F_X(x)F_Y(y).$$

Il suit que la densité jointe f est donnée par

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} (F_X(x)F_Y(y)) = \frac{d}{dx} F_X(x) \cdot \frac{d}{dy} F_Y(y) = f_X(x)f_Y(y),$$

où f_X et f_Y sont les densités marginales respectives de X et Y .

Dans l'autre sens, si $f(x, y) = f_X(x)f_Y(y)$, alors,

$$\begin{aligned} F(x, y) &= \int_{-\infty}^x \int_{-\infty}^y f(x, y) dy dx \\ &= \int_{-\infty}^x f_X(x) dx \cdot \int_{-\infty}^y f_Y(y) dy \\ &= F_X(x)F_Y(y). \end{aligned}$$

et par la proposition 6.1, on a que X et Y sont indépendantes. \square

EXEMPLE 6.7. On admet que le nombre de personnes qui entre dans un immeuble en une journée est une variable aléatoire de loi Poisson avec paramètre $\lambda > 0$. Si la probabilité qu'une personne porte un chapeau est p , et qu'on note X le nombre de personnes portant un chapeau, et Y le nombre de personnes ne portant pas de chapeau, déterminer les lois de X et Y , et montrer qu'elles sont indépendantes.

SOLUTION. On a que $\mathbb{P}\{X = k \mid X + Y = n\} = \binom{n}{k} p^k (1-p)^{n-k}$. – si on sait qu'il est entré n personnes, le nombre de personnes portant un chapeau X est une variable binomiale de paramètres n, k .

Par la formule de probabilités totale,

$$\mathbb{P}\{X = k\} = \sum_{n=k}^{\infty} \mathbb{P}\{X = k \mid X + Y = n\} \mathbb{P}\{X + Y = n\}.$$

Mais $X + Y$ est le nombre total de personnes entré.e.s dans l'édifice. Il s'agit d'une variable aléatoire de loi Poisson.

Finalement,

$$\begin{aligned} \mathbb{P}\{X = k\} &= \sum_{n=k}^{\infty} \mathbb{P}\{X = k \mid X + Y = n\} \mathbb{P}\{X + Y = n\} \\ &= \sum_{n=k}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{n=k}^{\infty} e^{-\lambda} \frac{(\lambda p)^k}{k!} \cdot \frac{(\lambda(1-p))^{n-k}}{(n-k)!} \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{(\lambda(1-p))^{n-k}}{(n-k)!} \\ &= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{n=0}^{\infty} \frac{(\lambda(1-p))^n}{n!} \\ &= e^{-\lambda+\lambda(1-p)} \frac{(\lambda p)^k}{k!} \\ &= e^{-\lambda p} \frac{(\lambda p)^k}{k!} \end{aligned}$$

et $X \sim \text{Pois}(\lambda p)$.

De façon complètement analogue, $Y \sim \text{Pois}(\lambda(1-p))$.

Or,

$$\begin{aligned} \mathbb{P}\{X = i, Y = j\} &= \mathbb{P}\{X = i \mid X + Y = i + j\} \mathbb{P}\{X + Y = i + j\} \\ &= \binom{i+j}{i} p^i (1-p)^j \frac{e^{-\lambda} \lambda^{i+j}}{(i+j)!} \\ &= e^{-\lambda p - \lambda(1-p)} \frac{(\lambda p)^i}{i!} \frac{(\lambda(1-p))^j}{j!} \\ &= \mathbb{P}\{X = i\} \mathbb{P}\{Y = j\}. \end{aligned}$$

et X et Y sont indépendantes.

6.2.1. Généralisation à n variables aléatoires. Dans le cas le plus général, on va vouloir considérer des familles $\mathbf{X} = (X_i)_{i \in I}$ de variables aléatoires indexées par un ensemble d'indices I quelconque – typiquement, ça sera \mathbb{N} mais ça pourrait aussi être \mathbb{Z} ou \mathbb{R} !

On veut une notion plus générale de ce que constituerait l'indépendance pour une famille de variables aléatoires quelconque.

DÉFINITION 6.5 (Famille indépendante). Soit $\mathbf{X} = (X_i)_{i \in I}$ une famille de variables aléatoires indexée par un ensemble d'indices I quelconque. Pour tout $J \subseteq I$, on note la sous-famille $\mathbf{X}_J = (X_j)_{j \in J}$.

Alors, on a que \mathbf{X} est une **famille indépendante** de variables aléatoires si et seulement si pour toute famille $(A_i)_{i \in I}$ d'ouverts² et pour tout $J \subseteq I$ fini, on a que

$$(6.2.5) \quad \mathbb{P} \left\{ \mathbf{X}_J \in \prod_{j \in J} A_j \right\} = \prod_{j \in J} \mathbb{P} \{X_j \in A_j\},$$

soit que, si $J = \{j_1, \dots, j_n\}$, on a

$$(6.2.6) \quad \mathbb{P} \{X_{j_1} \in A_{j_1}, \dots, X_{j_n} \in A_{j_n}\} = \mathbb{P} \{X_{j_1} \in A_{j_1}\} \cdots \mathbb{P} \{X_{j_n} \in A_{j_n}\}.$$

On généralise évidemment les propositions d'équivalence comme suit :

PROPOSITION 6.4. Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités et soit $\mathbf{X} = (X_i)_{i \in I}$ une famille de variables aléatoires indexée par un ensemble d'indices I quelconque. Soit F_i la fonction de répartition marginale de X_i pour tout i .

Pour $J \subseteq I$ fini, on note

$$F_J(\mathbf{x}_J) = \mathbb{P} \{X_j \leq x_j \ \forall j \in J\}.$$

i. La famille \mathbf{X} est indépendante si et seulement si pour toute famille $\mathbf{x} = (x_i)_{i \in I}$ et pour tout $J \subseteq I$ fini, avec $\mathbf{x}_J = (x_j)_{j \in J}$ on a que

$$(6.2.7) \quad F_J(\mathbf{x}_J) = \prod_{j \in J} F_j(x_j).$$

ii. Si les X_i sont des variables aléatoires continues avec densités marginales f_i et qu'on note f_J la densité marginale du sous-vecteur $\mathbf{X}_J = (X_j)_{j \in J}$. alors, la famille \mathbf{X} est indépendante si et seulement si pour tout $\mathbf{x} = (x_i)_{i \in I}$ et pour tout $J \subseteq I$ fini, on a que

$$(6.2.8) \quad f_J(\mathbf{x}_J) = \prod_{j \in J} f_j(x_j).$$

iii. Si les X_i sont des variables aléatoires discrètes avec fonctions de masse marginales p_i et qu'on note p_J la fonction de masse marginale du sous-vecteur $\mathbf{X}_J = (X_j)_{j \in J}$, alors, la famille \mathbf{X} est indépendante si et seulement si pour tout $\mathbf{x} = (x_i)_{i \in I}$ et pour tout $J \subseteq I$ fini, on a que

$$(6.2.9) \quad p_J(\mathbf{x}_J) = \prod_{j \in J} p_j(x_j).$$

Ce résultat suit directement de la définition de la généralisation de l'indépendance à des familles quelconques de variables et n'est pas particulièrement difficile. Nous nous passerons de la preuve.

REMARQUE. *Attention !* Il ne suffit pas que toutes les variables aléatoires soient indépendantes deux à deux pour déclarer immédiatement qu'elles forment une famille indépendante.

2. Comprendre : d'intervalles.

EXEMPLE 6.8. Soient X, Y, Z trois variables aléatoires de Bernoulli avec fonction de masse jointe

$$\begin{aligned} p(0, 0, 0) &= \frac{1}{4} & p(1, 0, 0) &= 0; \\ p(0, 0, 1) &= 0; & p(1, 0, 1) &= \frac{1}{4}; \\ p(0, 1, 0) &= 0; & p(1, 1, 0) &= \frac{1}{4}; \\ p(0, 1, 1) &= \frac{1}{4}; & p(1, 1, 1) &= 0; \end{aligned}$$

- (a) Montrer que X, Y et Z ont la même distribution marginale.
- (b) Montrer que (X, Y) sont indépendantes.
- (c) Montrer que (X, Y, Z) n'est pas indépendant.

SOLUTION. (a) Les trois variables sont des variables de Bernoulli. Toutes les permutations de (X, Y, Z) ont la même distribution jointe, puisque p est invariante sous permutations de ses trois arguments.

En particulier, $\mathbb{P}\{X = 0\} = \mathbb{P}\{X = 1\} = \frac{1}{2}$, et c'est pareil pour Y et Z .

- (b) On a que

$$\mathbb{P}\{X = Y = 0\} = \mathbb{P}\{X = 0, Y = 1\} = \mathbb{P}\{X = 1, Y = 0\} = \mathbb{P}\{X = Y = 1\} = \frac{1}{4},$$

et on a donc que X et Y sont indépendantes.

- (c) On a que

$$\mathbb{P}\{X = Y = Z = 0\} = \frac{1}{4} \neq \frac{1}{8} = \mathbb{P}\{X = 0\} \mathbb{P}\{Y = 0\} \mathbb{P}\{Z = 0\}.$$

EXEMPLE 6.9. Soient X, Y, Z trois variables aléatoires de loi uniforme sur $(0, 1)$ indépendantes.

Trouver $\mathbb{P}\{Z > XY\}$.

SOLUTION. Pour effectuer ce calcul, on commence par constater que si X, Y, Z sont indépendantes, alors leur densité jointe est donnée par $f(x, y, z) = f_X(x)f_Y(y)f_Z(z)$ – c'est

$$f(x, y, z) = \mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(y)\mathbb{1}_{(0,1)}(z) = \mathbb{1}_{(0,1)^3}(x, y, z).$$

Par conséquent,

$$\begin{aligned} \mathbb{P}\{Z > XY\} &= \int_0^1 \int_0^1 \int_{xy}^1 dz dy dx \\ &= \int_0^1 \int_0^1 (1 - xy) dy dx \\ &= \int_0^1 \left(1 - \frac{x}{2}\right) dx \\ &= 1 - \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

6.3. Sommes de variables aléatoires indépendantes

On se penche maintenant sur les sommes de variables aléatoires indépendantes. En effet, celles-ci sont très souvent utiles dans la théories des probabilités et des processus stochastiques, ou encore en statistiques.

Soient X et Y deux variables aléatoires continues indépendantes de fonctions de répartitions respectives F_X et F_Y , et de densités respectives f_X et f_Y . Alors, on sait que la densité jointe $f(x, y) = f_X(x)f_Y(y)$.

Par conséquent, avec $Z = X + Y$, la fonction de répartition de Z est donnée par :

$$\begin{aligned} F_Z(z) &= \mathbb{P}\{X + Y \leq z\} = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_X(x)f_Y(y)dydx \\ (6.3.1) \quad &= \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)dx. \end{aligned}$$

En dérivant par z de part et d'autre³, on trouve finalement que la densité de Z est donnée par

$$\begin{aligned} f_Z(z) &= F'_Z(z) = \frac{d}{dz} \int_{-\infty}^{\infty} f_X(x)F_Y(z-x)dx \\ &= \int_{-\infty}^{\infty} f_X(x)F'_Y(z-x)dx \\ (6.3.2) \quad &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx. \end{aligned}$$

REMARQUE. En analyse, on note

$$f * g(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt;$$

c'est la **convolution** des fonctions f et g (plus de détails à la section B.9).

En ces termes, la densité de la somme Z de deux variables aléatoires indépendantes n'est que la convolution des densités de X et Y .

On voit qu'un raisonnement tout à fait similaire donne

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy;$$

C'est bien pratique puisqu'on s'attend à ce que $X + Y$ et $Y + X$ aient la même distribution, bien entendu.

EXEMPLE 6.10. Trouver la distribution de la somme de deux variables aléatoires uniformes $(0, 1)$.

3. Les jolies propriétés de f_X et F_Y nous garantissent que c'est permis de passer la dérivée dans l'intégrale. Juré craché.

SOLUTION. On considère deux variables aléatoires X, Y indépendantes avec densités $f_X(t) = f_Y(t) = \mathbb{1}_{(0,1)}(t)$. On considère $Z = X + Y$. Alors, pour $z \in (0, 2)$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(t)f_Y(z-t)dt \\ &= \int_{-\infty}^{\infty} \mathbb{1}_{(0,1)}(t)\mathbb{1}_{(0,1)}(z-t)dt \\ &= \int_{-\infty}^{\infty} \mathbb{1}_{(0,1)}(t)\mathbb{1}_{(z-1,z)}(t)dt \\ &= \int_{\max\{0,z-1\}}^{\min\{z,1\}} dt \\ &= \min\{z,1\} - \max\{0,z-1\} \\ &= 1 + \min\{z-1,0\} - \max\{z-1,0\} \\ &= 1 - (\max\{0,z-1\} - \min\{0,z-1\}) \\ &= 1 - |z-1|. \end{aligned}$$

6.3.1. Sommes de variables indépendantes de loi Gamma. On se penche maintenant sur le cas particulier de la somme de variables aléatoires de loi Gamma. On se souvient (voir la section 5.3.4) que la densité d'une variable aléatoire de loi Gamma avec paramètres α, λ est donnée par

$$(5.3.25) \quad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.$$

On montre donc la proposition suivante :

PROPOSITION 6.5. *Soient X, Y deux variables aléatoires indépendantes, avec $X \sim \Gamma(s, \lambda)$ et $Y \sim \Gamma(t, \lambda)$.*

Alors, $X + Y \sim \Gamma(s+t, \lambda)$.

DÉMONSTRATION. Pour faire la preuve il suffit de calculer la convolution : par (6.3.2), si $Z = X + Y$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(u)f_Y(z-u)du \\ &= \frac{\lambda^{s+t}}{\Gamma(s)\Gamma(t)} \int_{-\infty}^{\infty} e^{-\lambda u} e^{-\lambda(z-u)} u^{s-1} (z-u)^{t-1} \mathbb{1}_{(0,\infty)}(u) \mathbb{1}_{(0,\infty)}(z-u) du \\ &= \frac{\lambda^{s+t}}{\Gamma(s)\Gamma(t)} e^{-\lambda z} \int_0^z u^{s-1} (z-u)^{t-1} du. \end{aligned}$$

Avec le changement de variables $u = zv$, on a $du = zdv$

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{s+t}}{\Gamma(s)\Gamma(t)} e^{-\lambda z} z^{s+t-1} \int_0^1 v^{s-1} (1-v)^{t-1} dv \\ &= \frac{\lambda^{s+t} B(s, t)}{\Gamma(s)\Gamma(t)} e^{-\lambda z} z^{s+t-1} \\ &= \frac{\lambda^{s+t}}{\Gamma(s+t)} z^{s+t-1} e^{-\lambda z} \end{aligned}$$

et Z suit bel et bien une loi Gamma de paramètres $(s+t, \lambda)$. \square

On déduit immédiatement le corollaire suivant :

COROLLAIRE 6.1. *Soient X_1, \dots, X_n des variables aléatoires indépendantes, telles que X_i suit une loi Gamma de paramètres (α_i, λ) pour tout i .*

Alors, $S_n = \sum_{k=1}^n X_k$ suit une loi Gamma de paramètres $(\sum_{k=1}^n \alpha_k, \lambda)$.

DÉMONSTRATION. La preuve est laissée à l'exercice 6.25. \square

EXEMPLE 6.11. Soient $X_1, X_2, X_3, \dots, X_n$ toutes des variables aléatoires de loi exponentielles de paramètre λ formant une famille indépendante.

Alors, $S_n = \sum_{k=1}^n X_k$ est une variable aléatoire de loi $\text{Gamma}(n, \lambda)$.

En effet, la loi exponentielle n'est en fait que la loi Gamma avec paramètres $(1, \lambda)$.

EXEMPLE 6.12 (Le processus de Points de Poisson). Dans une centrale téléphonique, on suppose que les temps T_i entre la réception des $(i-1)$ -es et i -èmes appels sont distribués de façon indépendante pour chaque i selon une loi exponentielle de paramètre $\lambda = 2 \text{ min}^{-1}$.

Montrer que le nombre d'appels reçus en 1 minute est une variable aléatoire de Poisson de paramètre $\lambda = 2$.

SOLUTION. Supposons qu'on dénote par N le nombre d'appels reçus en 1 minute.

On va définir $S_n = \sum_{i=1}^n T_i$ – c'est l'heure à laquelle on reçoit l' n ième appel, puisque c'est la somme de tous les délais entre les appels consécutifs.

Alors, l'événement $\{N \leq n\}$ est exactement l'événement que $\{S_{n+1} > 1\}$ – en effet, si on sait que le $(n+1)$ ième appel est arrivé *après* une minute, ça veut dire qu'en une minute on a pu recevoir au plus n appels.

Or, S_{n+1} est une somme de $n+1$ variables aléatoires exponentielles indépendantes de paramètre λ – donc, S_{n+1} est une variable aléatoire de loi Gamma, de paramètres $(n+1, \lambda)$.

Alors on calcule :

$$\begin{aligned}
\mathbb{P}\{N \leq n\} &= \mathbb{P}\{S_{n+1} > 1\} \\
&= \frac{\lambda^{n+1}}{\Gamma(n+1)} \int_1^\infty x^n e^{-\lambda x} dx \\
&= \frac{\lambda^{n+1}}{\Gamma(n+1)} \int_0^\infty (y+1)^n e^{-\lambda(y+1)} dy \\
&= \frac{\lambda^{n+1}}{\Gamma(n+1)} e^{-\lambda} \int_0^\infty \sum_{k=0}^n \binom{n}{k} y^k e^{-\lambda y} dy \\
&= \frac{\lambda^{n+1}}{n!} e^{-\lambda} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \int_0^\infty y^k e^{-\lambda y} dy \\
&= e^{-\lambda} \sum_{k=0}^n \frac{\lambda^{n+1}}{k!(n-k)!} \frac{\Gamma(k+1)}{\lambda^{k+1}} \\
&= e^{-\lambda} \sum_{k=0}^n \frac{\lambda^{n-k}}{(n-k)!} \\
&= e^{-\lambda} \sum_{k'=0}^n \frac{\lambda^{k'}}{k'!}.
\end{aligned}$$

Et

$$\mathbb{P}\{N = n\} = \mathbb{P}\{N \leq n\} - \mathbb{P}\{N \leq n-1\} = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Et N est une variable aléatoire de loi Poisson avec paramètre λ .

EXEMPLE 6.13 (La loi du chi-carré). Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes de loi normale centrée réduite.

On note $Y = \sum_{k=1}^n X_k^2$. Trouver la loi de Y .

SOLUTION. Ici, on commence par considérer la loi de X_1^2 – Si on écrit $R_1 = X_1^2$, alors bien sûr, la densité de R_1 est donnée, par la proposition 5.12, par

$$f_R(r) = \phi(\sqrt{r}) \cdot \frac{1}{2\sqrt{r}} = \frac{1}{2\sqrt{2\pi}} r^{1/2-1} e^{-r/2}, \quad r > 0.$$

R_1 suit ce que l'on appelle la distribution du χ_1^2 – c'est simplement une loi Gamma de paramètres $(1/2, 1/2)$, comme on le voit à la forme de sa densité.

Il en va de même pour tous les X_i – ils suivent des distributions Gamma de paramètres $(1/2, 1/2)$. On a donc que leur somme, $Y = \sum_{k=1}^n X_k$ suit une loi Gamma de paramètres $(n/2, 1/2)$ – c'est ce qu'on appelle la loi du χ^2 à n degrés de liberté, ou la loi du χ_n^2 .

6.3.2. Les sommes de variables aléatoires normales indépendantes. Une autre conséquence importante de l'équation (6.3.2) est la proposition suivante :

PROPOSITION 6.6. Soient X une variable aléatoire de loi normale de paramètres μ_X, σ_X^2 et Y une variable aléatoire de loi normale de paramètres μ_Y, σ_Y^2 , indépendantes.

Alors, $Z = X + Y$ suit une loi normale de paramètres $\mu = \mu_X + \mu_Y$, $\sigma^2 = \sigma_X^2 + \sigma_Y^2$.

Pour prouver cette proposition, nous allons d'abord montrer le lemme suivant :

LEMME 6.2. *Soient X, Y deux variables aléatoires indépendantes, où X suit une loi normale centrée réduite et Y suit une loi normale d'espérance 0 et de variance σ^2 . Alors $X + Y$ est une variable aléatoire normale centrée de variance $1 + \sigma^2$.*

DÉMONSTRATION. On prouve ceci en faisant simplement la convolution. Par (6.3.2), la densité f_Z de la somme $Z = X + Y$ est donnée par

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(t) f_Y(z-t) dt \\ &= \frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} e^{-\frac{(z-t)^2}{2\sigma^2}} dt \\ &= \frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(t^2 + \frac{1}{\sigma^2}z^2 - 2\frac{zt}{\sigma^2} + \frac{1}{\sigma^2}t^2)} dt \\ &= \frac{1}{2\pi\sigma} e^{-\frac{z^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{\sigma^2+1}{\sigma^2}(t^2 - 2\frac{zt}{\sigma^2+1})} dt \\ &= \frac{1}{2\pi\sigma} e^{-\frac{z^2}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{\sigma^2+1}{\sigma^2}\left(\left[t - \frac{z}{1+\sigma^2}\right]^2 - \frac{z^2}{(1+\sigma^2)^2}\right)} dt \\ &= \frac{1}{2\pi\sigma} e^{-\frac{z^2}{2\sigma^2}} e^{\frac{z^2}{2\sigma^2(1+\sigma^2)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{\sigma^2+1}{\sigma^2}\left[t - \frac{z}{\sigma^2+1}\right]^2} dt \\ &= \frac{1}{2\pi\sigma} e^{-\frac{z^2}{2(1+\sigma^2)}} \cdot \sqrt{2\pi \frac{\sigma^2}{\sigma^2+1}} \\ &= \frac{1}{\sqrt{2\pi(\sigma^2+1)}} e^{-\frac{z^2}{2(1+\sigma^2)}}. \end{aligned}$$

Et on conclue que Z suit une loi normale centrée de variance $1 + \sigma^2$. \square

DÉMONSTRATION DE LA PROPOSITION 6.6. Si X est d'espérance μ_X et de variance σ_X^2 , alors $X' = \frac{X - \mu_X}{\sigma_X}$ est une variable aléatoire normale centrée réduite.

Si Y est d'espérance μ_Y et de variance σ_Y^2 , alors $Y' = \frac{Y - \mu_Y}{\sigma_Y}$ est une variable aléatoire normale centrée de variance $\sigma_0^2 = \frac{\sigma_Y^2}{\sigma_X^2}$.

Il suit que $Z' = X' + Y'$ est une variable aléatoire normale centrée de variance $1 + \frac{\sigma_Y^2}{\sigma_X^2}$.

Mais maintenant, si $Z = X + Y$, on a que

$$Z = X + Y = \sigma_X Z' + \mu_X + \mu_Y,$$

et par la proposition 5.5, Z est une variable aléatoire de loi normale avec espérance $\mu_X + \mu_Y$ et variance $\sigma_X^2 + \sigma_Y^2$. \square

On a évidemment le corollaire suivant :

COROLLAIRE 6.2. *Si X_1, X_2, \dots, X_n sont une famille de variables aléatoires indépendante de lois marginales normales, avec $\mathbb{E}[X_i] = \mu_i$ et $\text{Var}[X_i] = \sigma_i^2$ pour tout i de 1 à n ,*

alors $S_n = \sum_{i=1}^n X_i$ suit une loi normale avec paramètres $\mu = \sum_{i=1}^n \mu_i$ et $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.

DÉMONSTRATION. La preuve est faite par induction et est laissée à l'exercice 6.26. \square

EXEMPLE 6.14. En statistiques, si on a un échantillon X_1, X_2, \dots, X_n de n valeurs, dont on assume qu'elles sont tirées indépendamment de la même distribution, alors on utilise $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne arithmétique des valeurs de l'échantillon pour estimer l'espérance $\mathbb{E}[X_1]$ de la distribution d'origine.

Prenons par exemple un échantillon de n humains mâles adultes. Nous avons déjà mentionné que l'on peut assumer que leurs tailles respectives seront des variables aléatoires distribuées selon une loi normale de paramètres $\mu = 177$ cm et $\sigma^2 = 100$ cm².

Quelle sera la variance de \bar{X} ?

SOLUTION. La variance de $\sum_{i=1}^n X_i$ sera de $n\sigma^2$, par le corollaire 6.2. Mais par la proposition 5.3, la variance de $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ sera donc $\frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$.

De façon surprenante, la variance de \bar{X} diminue graduellement à mesure qu'on prend des échantillons de plus en plus grands.

Cette observation est un cas très particulier qui motivera ce qu'on appelle la « loi des grands nombres »⁴.

EXEMPLE 6.15. On assume que, d'année en année, le taux de rendement annuel d'un placement sera distribué indépendamment selon une loi normale⁵ de paramètres $\mu = 0,01$, $\sigma^2 = 0,0009$.

Si R_i est le taux de rendement du placement pour l'année i , et que X_i est la valeur du placement au début de l'année i , alors

$$X_{i+1} = X_i(1 + R_i).$$

Nous allons faire l'approximation⁶

$$X_{i+1} = X_i e^{R_i}.$$

Ainsi, les ratios X_{i+1}/X_i suivent des distributions log-normales indépendantes.

Trouver la probabilité que le placement se soit apprécié de plus de 5% en 5 ans.

SOLUTION. On cherche $\mathbb{P}\{X_6/X_1 \geq 1,05\}$. Avec notre approximation, on a bien sûr que

$$\frac{X_6}{X_1} = e^{R_1+R_2+R_3+R_4+R_5}.$$

Les variables aléatoires R_i suivent des normales de paramètres $\mu = 0,01$ et de variance $\sigma^2 = 0,0009$. Par conséquent, $S := R_1 + R_2 + R_3 + R_4 + R_5 \sim \mathcal{N}(\mu = 0,05; \sigma^2 = 0,0045)$.

On cherche maintenant la probabilité que

$$\mathbb{P}\left\{\frac{X_6}{X_1} = e^S \geq 1,05\right\} = \mathbb{P}\{S \geq \log(1,05)\} = 1 - \Phi\left(\frac{\log(1,05) - 0,05}{\sqrt{0,0045}}\right) \approx 50\%.$$

4. Voir les théorèmes 8.1 et 8.2.

5. Je ne connais rien à la finance et je n'ai donc aucune idée de la validité de ce modèle. C'est juste pour l'exemple.

6. Ici, R_i serait le rendement annuel nominal moyen pour l'année i – l'approximation revient à calculer le rendement composé continûment sur toute l'année à un taux nominal moyen de R_i .

6.3.3. Les sommes de variables aléatoires binomiales indépendantes. Nous venons de voir comment on peut déterminer la densité de la somme de deux variables aléatoires continues indépendantes en faisant tout simplement la convolution de leurs densités marginales respectives.

Idéalement, on aimerait avoir une façon de faire le même genre de chose pour les variables aléatoires discrètes.

De façon générale, évidemment, si X et Y sont deux variables aléatoires discrètes indépendantes, avec fonctions de masse p_X et p_Y et supports V_X et V_Y respectivement, et que $Z = X + Y$ a la fonction de masse p et le support $V = V_X + V_Y = \{x + y : x \in V_X, y \in V_Y\}$, alors pour tout $v \in V$,

$$(6.3.3) \quad p(v) = \sum_{x \in V_X} p_X(x)p_Y(v - x).$$

Lorsque $V_X, V_Y \subseteq \mathbb{Z}$, on peut écrire de façon plus systématique. D'abord, on remarque qu'en effet, $V = V_X + V_Y \subseteq \mathbb{Z}$ également (puisque les entiers négatifs sont fermés sous l'addition). On peut donc faire, de façon beaucoup plus systématique le décompte des combinaisons possibles d'entiers x, y qui ont la somme désirée : pour tout $v \in \mathbb{Z}$:

$$(6.3.4) \quad p(v) = \sum_{k \in \mathbb{Z}} p_X(k)p_Y(v - k).$$

L'exemple désigné est celui des variables aléatoires binomiales.

PROPOSITION 6.7. *Si X est une variable aléatoire binomiale de paramètres (n, p) et que Y est une variable aléatoire de paramètres (m, p) , alors $X + Y$ est une variable aléatoire binomiale de paramètres $(m + n, p)$.*

DÉMONSTRATION (PAR CALCUL). On utilise l'équation (6.3.4), tout en remarquant que pour tout $i > k$ ou $i < 0$, $\mathbb{P}\{X = i, Y = k - i\} = 0$. On obtient donc :

$$\begin{aligned} \mathbb{P}\{X + Y = k\} &= \sum_{i=0}^k \mathbb{P}\{X = i\} \mathbb{P}\{Y = k - i\} \\ &= \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-k+i} \\ &= p^k (1-p)^{n+m-k} \sum_{i=0}^k \binom{n}{i} \binom{m}{k-i} \end{aligned}$$

Pour terminer le raisonnement, il suffit d'appliquer l'identité de Vandermonde, qui fait l'objet de l'exercice 1.17 :

$$(1.3.1) \quad \binom{n+m}{k} = \sum_{i=1}^k \binom{n}{i} \binom{m}{k-i}.$$

On obtient ainsi :

$$\mathbb{P}\{X + Y = k\} = p^k (1-p)^{m+n-k} \binom{m+n}{k}$$

pour autant que $0 \leq k \leq m + n$.

On a donc que $X + Y$ est une variable aléatoire discrète dont la fonction de masse est exactement celle d'une variable aléatoire binomiale de paramètres $(m + n, p)$. \square

Une autre façon de voir ce résultat est bien sûr de considérer les variables aléatoires binomiales comme des « nombres de succès en un nombre de tentatives ». Dans ce cadre, il est assez facile de constater que si X est le nombre de succès en n tentatives indépendantes, et que Y est le nombre de succès (de même probabilité) en m autres tentatives indépendantes (toutes différentes), alors $X+Y$ est simplement le nombre total de succès en $m+n$ tentatives indépendantes, et c'est bien une variable aléatoire binomiale de paramètres $(m+n, p)$.

On a évidemment le corollaire suivant :

COROLLAIRE 6.3. *Soient X_1, X_2, \dots, X_n des variables binomiales indépendantes, respectivement de paramètres (n_i, p) pour tout i .*

Alors, $S_k = \sum_{i=1}^k X_i$ est une variable aléatoire binomiale de paramètres $(n = \sum_{i=1}^k n_i, p)$.

DÉMONSTRATION. La preuve est laissée pour l'exercice 6.27. \square

EXEMPLE 6.16. Si chaque étudiant.e d'une classe de N étudiant.e.s doit répondre à 6 « Vrai ou Faux » à l'examen intra d'un cours de probabilité, et qu'ils et elles répondent tou.te.s de façon complètement aléatoire en choisissant chaque fois « Vrai » ou « Faux » de façon équiprobable,

- (a) quelle est la distribution pour le résultat d'une personne à cet exercice ?
- (b) Quelle est la distribution pour le nombre total de questions réussies par toute la classe ?

SOLUTION. (a) Ce sera une variable aléatoire binomiale de paramètres $n = 6, p = \frac{1}{2}$.

- (b) Ce sera une variable aléatoire binomiale de paramètres $n = 6N, p = \frac{1}{2}$.

6.3.4. Les sommes de variables de Poisson indépendantes. On peut faire la même passe aux variables aléatoires de Poisson.

PROPOSITION 6.8. *Si X, Y sont deux variables aléatoires indépendantes de distribution de Poisson avec paramètres λ et μ respectivement, alors $Z = X + Y$ suit une distribution de Poisson de paramètres $\lambda + \mu$.*

DÉMONSTRATION. On a bien sûr par l'équation (6.3.4) que

$$\begin{aligned}\mathbb{P}\{X + Y = n\} &= \sum_{k=0}^n \mathbb{P}\{X = k\} \mathbb{P}\{Y = n - k\} \\ &= e^{-\lambda} e^{-\mu} \sum_{k=0}^n \frac{\lambda^k}{k!} \frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda^k \mu^{n-k} \\ &= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}.\end{aligned}$$

Cela coïncide parfaitement avec la fonction de masse pour une variable aléatoire de Poisson avec paramètre $\lambda + \mu$. \square

REMARQUE. La proposition 6.8 et les résultats montrés dans les exemples 6.7 et 6.12 sont à la base d'une notion cruciale en théorie des processus stochastiques : le processus de points de Poisson. L'exercice 6.24 explore ce processus plus en profondeur.

6.4. Distributions conditionnelles

Si on considère le vecteur aléatoire (X, Y) , on peut se demander ce qu'on connaît sur la distribution de X « sachant Y » – par analogie directe avec les probabilités conditionnelles au chapitre 3.

On introduit donc des outils pour discuter des « distributions conditionnelles » de variables aléatoires.

Plus concrètement, on veut pouvoir dire rigoureusement des choses comme : « Si un groupe de N personnes entre dans un ascenseur, où N est une variable aléatoire, et que chacune d'entre ces personnes porte un chapeau indépendamment avec probabilité p , alors, sachant N , le nombre X de personnes portant des chapeaux suit une loi binomiale N, p . »

Ici, il est important d'insister : la loi *marginal* de X ne sera pas forcément une loi binomiale. Mais *conditionnellement à N* , c'est une binomiale de paramètre N, p .

6.4.1. Le cas des variables discrètes. Dans le cas des variables aléatoires X, Y discrètes avec fonction de masse jointe p et fonctions de masse marginales respectives p_X, p_Y , on peut définir la **fonction de masse conditionnelle** de X sachant Y :

$$(6.4.1) \quad p_{X|Y}(x | y) = \mathbb{P}\{X = x | Y = y\} = \frac{\mathbb{P}\{X = x, Y = y\}}{\mathbb{P}\{Y = y\}} = \frac{p(x, y)}{p_Y(y)}.$$

Ceci est évidemment défini seulement à condition que $\mathbb{P}\{Y = y\} > 0$, mais il s'agit bien sûr des seuls cas intéressants.

On a également la notion analogue de la fonction de répartition conditionnelle :

$$(6.4.2) \quad F_{X|Y}(x | y) = \mathbb{P}\{X \leq x | Y = y\} = \sum_{z \in V_X: z \leq x} p_{X|Y}(z | y).$$

REMARQUE. Par la proposition 6.2, on voit que X et Y sont indépendantes si et seulement si $p_{X|Y}(x | y) = p_X(x)$ pour tous x, y ; c'est à dire que X et Y sont indépendantes si et seulement si la distribution conditionnelle de X sachant Y ne change pas en fonction de la valeur de Y – ce qui a tout plein de sens.

EXEMPLE 6.17. Soient X, Y deux variables aléatoires de Poisson indépendantes, respectivement de paramètres λ et μ . On note $Z = X + Y$.

Trouver la distribution conditionnelle de X sachant Z .

SOLUTION. On a que

$$\begin{aligned} p_{X|Z}(k | n) &= \mathbb{P}\{X = k | X + Y = n\} \\ &= \frac{\mathbb{P}\{X = k, X + Y = n\}}{\mathbb{P}\{X + Y = n\}} \\ &= \frac{\mathbb{P}\{X = k, Y = n - k\}}{\mathbb{P}\{X + Y = n\}}. \end{aligned}$$

On sait que, puisque X et Y sont des variables de Poisson indépendantes de paramètres respectifs λ et μ , alors $Z = X + Y$ est une variable de Poisson de paramètre $\lambda + \mu$. On a

donc

$$\begin{aligned} p_{X|Z}(k \mid n) &= \frac{e^{-\lambda}\lambda^k}{k!} \frac{e^{-\mu}\mu^{n-k}}{(n-k)!} \frac{n!}{e^{-\lambda-\mu}(\lambda+\mu)^n} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{\lambda+\mu}\right)^k \left(\frac{\mu}{\lambda+\mu}\right)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}, \end{aligned}$$

avec $p = \frac{\lambda}{\lambda+\mu}$.

On a donc que, sachant $X + Y$, la variable aléatoire X suit une distribution binomiale de paramètres $(X + Y, p)$.

À noter qu'on peut tout à fait déterminer la distribution jointe d'un vecteur aléatoire en fournissant, plutôt que sa fonction de masse jointe, les fonctions de masse conditionnelles et marginales nécessaires.

EXEMPLE 6.18. On définit Z une variable aléatoire de loi binomiale négative avec paramètres $r = 2$ et $p \in (0, 1)$.

Puis, sachant Z on définit X une variable aléatoire équiprobablement choisie parmi les entiers entre 1 et $Z - 1$.

- (a) Montrer que X suit une loi géométrique de paramètre p .
- (b) Montrer que $Y = Z - X$ suit une loi géométrique de paramètre p .
- (c) Montrer que X et Y sont indépendantes.

SOLUTION. (a) On a évidemment que, si $n \geq k + 1$,

$$\mathbb{P}\{X = k \mid Z = n\} = \frac{1}{n-1}.$$

De plus, on sait que Z est une binomiale négative de paramètre $r = 2$ et p . Donc,

$$\mathbb{P}\{Z = n\} = (n-1)p^2(1-p)^{n-2}.$$

Par la formule de probabilités totale, on a que

$$\begin{aligned} \mathbb{P}\{X = k\} &= \sum_{n \geq k+1} \mathbb{P}\{X = k \mid Z = n\} \mathbb{P}\{Z = n\} \\ &= \sum_{n \geq k+1} \frac{1}{n-1} (n-1)p^2(1-p)^{n-2} \\ &= p^2 \sum_{n \geq k+1} (1-p)^{n-2} \\ &= p^2 \sum_{n' \geq 0} (1-p)^{k+1+n'-2} \\ &= p^2 (1-p)^{k-1} \sum_{n' \geq 0} (1-p)^{n'} \\ &= (1-p)^{k-1} p. \end{aligned}$$

- (b) Clairement, si $Y = Z - X$, on a que Y est aussi distribuée équiprobablement sur les entiers entre 1 et $Z - 1$. En effet, pour tout k, n tels que $n \geq k + 1$,

$$\mathbb{P}\{Z - X = k \mid Z = n\} = \mathbb{P}\{X = n - k \mid Z = n\} = \frac{1}{n - 1}.$$

Donc, Y a la même distribution conditionnelle que X et donc la même distribution marginale : Y est aussi une variable aléatoire géométrique de paramètre p .

- (c) On calcule la distribution jointe : pour $k, k' \geq 1$,

$$\begin{aligned} \mathbb{P}\{X = k, Y = k'\} &= \mathbb{P}\{X = k, Z = k + k'\} \\ &= \mathbb{P}\{X = k \mid Z = k + k'\} \mathbb{P}\{Z = k + k'\} \\ &= \frac{1}{k + k' - 1} (k + k' - 1)p^2(1-p)^{k+k'-2} \\ &= ((1-p)^{k-1}p) \cdot ((1-p)^{k'-1}p) \\ &= \mathbb{P}\{X = k\} \mathbb{P}\{Y = k'\}. \end{aligned}$$

et on a montré que X et Y sont indépendantes (proposition 6.2).

6.4.2. Le cas des variables aléatoires continues. Dans le cas de vecteurs aléatoires (X, Y) où les variables aléatoires sont continues, évidemment l'expression

$$\frac{\mathbb{P}\{X = x, Y = y\}}{\mathbb{P}\{Y = y\}}$$

n'est pas bien définie, vu que les probabilités au numérateur et au dénominateur son 0.

Cependant, de façon tout à fait analogue, on peut obtenir la **densité conditionnelle**

$$(6.4.3) \quad f_{X|Y}(x \mid y) := \frac{f(x, y)}{f_Y(y)},$$

où f est la densité jointe de X, Y et f_Y est la densité marginale de Y . Pour motiver cette expression, on commence en constatant que :

$$\frac{\mathbb{P}\{x \leq X < x + \epsilon \mid y \leq Y < y + \eta\}}{\epsilon} = \frac{\frac{\mathbb{P}\{X \in [x, x+\epsilon], Y \in [y, y+\eta]\}}{\epsilon\eta}}{\frac{\mathbb{P}\{Y \in [y, y+\eta]\}}{\eta}},$$

puis en prenant la limite lorsque ϵ, η tend vers 0.

REMARQUE. Encore une fois, il suit directement de la proposition 6.3 que X et Y sont indépendantes si et seulement si $f_{X|Y}(x \mid y) = f_X(x)$ pour tous x, y où la densité conditionnelle est définie.

On peut alors commettre l'abus de notation suivant :

$$\text{“} \mathbb{P}\{X \in A \mid Y = y\} = \int_A f_{X|Y}(x \mid y) dx \text{”}.$$

REMARQUE. *Attention !* Il s'agit d'un abus de notation – en effet, la probabilité conditionnelle n'est bel et bien pas définie, puisque $\mathbb{P}\{Y = y\} = 0$.

Toutefois, dans le cadre de ce cours et pour nos fins, nous nous permettrons, dans un tel contexte, de faire un tel « conditionnement ».

Ainsi, la **fonction de répartition conditionnelle** sera définie, prévisiblement, par

$$(6.4.4) \quad F_{X|Y}(x | y) = \int_{-\infty}^x f_{X|Y}(t | y) dt, \quad " = \mathbb{P}\{X \leq x | Y = y\}$$

EXEMPLE 6.19. Soient X, Y deux variables aléatoires de densité jointe

$$f(x, y) = \begin{cases} xe^{-x(y+1)} & \text{si } x, y > 0 \\ 0 & \text{sinon.} \end{cases}$$

- (a) Trouver la densité marginale de X .
- (b) Trouver la densité conditionnelle de Y sachant X .

SOLUTION. (a) Pour trouver la densité marginale de X il suffit d'intégrer par rapport à Y . Pour $x > 0$, on a que :

$$\begin{aligned} f_X(x) &= \int_0^\infty xe^{-xy-x} dy \\ &= e^{-x} \int_0^\infty xe^{-xy} dy \\ &= e^{-x}. \end{aligned}$$

et X est une variable aléatoire exponentielle de paramètre 1 (une exponentielle standard).

- (b) On trouve la densité conditionnelle de Y sachant X en appliquant simplement la définition de l'équation (6.4.3) :

$$f_{Y|X}(y | x) = \frac{f(x, y)}{f_X(x)} = \frac{xe^{-x(y+1)}}{e^{-x}} = xe^{-xy},$$

pour $x, y > 0$.

Ainsi, sachant X , on a que Y est une variable aléatoire exponentielle de paramètre X .

6.4.3. Le cas mixte. Nous avons brièvement mentionné à la section 6.1.3 une façon de décrire la distribution jointe d'un vecteur aléatoire X, Y où l'une des variables est discrète et l'autre est continue. La plupart du temps, cependant, ce sera beaucoup plus facile de s'intéresser à la densité conditionnelle.

Soit (X, Y) un vecteur aléatoire avec X une variable aléatoire continue, et Y une variable aléatoire discrète de support V (au plus dénombrable).

Alors, on définit la fonction de répartition conditionnelle de X sachant Y par

$$(6.4.5) \quad F_{X|Y}(x | y) = \mathbb{P}\{X \leq x | Y = y\}.$$

Le membre de droite de l'équation (6.4.5) est bien défini puisque si $y \in V$, alors $\mathbb{P}\{Y = y\} > 0$.

Il suffit dès lors de dériver pour obtenir la densité conditionnelle de X sachant Y :

$$(6.4.6) \quad f_{X|Y}(x | y) = \frac{\partial}{\partial x} F_{X|Y}(x | y).$$

Tout ça va très bien. Le souci, c'est dans l'autre sens – quand on veut connaître des choses sur la distribution conditionnelle de Y sachant X .

Idéalement, on voudrait une fonction de masse $p_{Y|X}$ de Y sachant X .

Pour y parvenir, on considère :

$$p_{Y|X}(y | x) = \lim_{\epsilon \rightarrow 0^+} \frac{\frac{\mathbb{P}\{Y=y, x \leq X < x + \epsilon\}}{\epsilon}}{\frac{\mathbb{P}\{x \leq X < x + \epsilon\}}{\epsilon}}$$

Bien sûr,

$$\mathbb{P}\{Y = y, x \leq X < x + \epsilon\} = \mathbb{P}\{x \leq X < x + \epsilon | Y = y\} \mathbb{P}\{Y = y\} \approx \epsilon f_{X|Y}(x | y) \mathbb{P}\{Y = y\}.$$

On devrait donc déduire

$$(6.4.7) \quad p_{Y|X}(y | x) = \frac{f_{X|Y}(x | y) \mathbb{P}\{Y = y\}}{f_X(x)}, \text{ " } = \mathbb{P}\{Y = y | X = x\} \text{ ,}$$

où $f_X(x)$ est la densité marginale de X .

En rappel de la section 6.1.3 : Nous avions que la distribution jointe des variables X et Y était décrite par une famille de fonctions

$$(6.1.11) \quad (f_y)_{y \in V} : f_y(x) = \frac{\partial}{\partial x} \left(F(x, y) - \lim_{z \rightarrow y^-} F(x, z) \right),$$

où F est la fonction de répartition jointe $F(x, y) = \mathbb{P}\{X \leq x, Y \leq y\}$, toujours définie.

Ici, la famille des $(f_y)_{y \in V}$ joue un peu le rôle d'un espèce d'hybride entre la fonction de masse jointe et la fonction de densité jointe.

En particulier, on avait

$$(6.1.15) \quad \mathbb{P}\{Y = y\} = \int_{-\infty}^{\infty} f_y(x) dx,,$$

et

$$(6.1.16) \quad f_X(x) = \sum_{y \in V} f_y(x).$$

On peut maintenant ajouter la relation suivante :

$$(6.4.8) \quad f_y(x) = f_{X|Y}(x | y) \mathbb{P}\{Y = y\} = p_{Y|X}(y | x) f_X(x).$$

EXEMPLE 6.20. Soit X une variable aléatoire continue de densité marginale $f_X(x) = e^{-x}$ pour $x > 0$ et soit Y de fonction de masse conditionnelle donnée, pour $k \geq 0$ entier, par

$$p_{Y|X}(k | x) = e^{-x} \frac{x^k}{k!}.$$

Autrement dit, X est une exponentielle standard et, sachant X , Y est une variable aléatoire de Poisson de paramètre X .

Trouver la loi marginale de Y .

SOLUTION. Par les équations (6.1.15) et (6.4.8), on trouve que pour $k \geq 0$,

$$\begin{aligned}\mathbb{P}\{Y = k\} &= \int_{-\infty}^{\infty} f_k(x) dx \\ &= \int_0^{\infty} p_{Y|X}(k | x) f_X(x) dx \\ &= \int_0^{\infty} e^{-x} \frac{x^k}{k!} e^{-x} dx \\ &= \frac{1}{k!} \int_0^{\infty} x^k e^{-2x} dx \\ &= \frac{1}{k!} \frac{k!}{2^k} = \frac{1}{2^k}.\end{aligned}$$

Soit que $Y + 1 \sim Geom(\frac{1}{2})$.

EXEMPLE 6.21. On réalise n tentatives indépendantes de probabilité de succès P , où P est une variable aléatoire uniforme comprise entre 0 et 1.

Soit X le nombre de succès. Déterminer la densité conditionnelle de X sachant que $X = k$.

SOLUTION. Sachant P , on a que X est de loi binomiale n, P .

Donc,

$$p_{X|P}(k | p) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Maintenant, la densité marginale de P est $\mathbb{1}_{(0,1)}(p)$. Par conséquent, avec l'équation (6.4.7),

$$f_{P|X}(p | k) = \frac{p_{X|P}(k | p) f_P(p)}{\mathbb{P}\{X = k\}} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\mathbb{P}\{X = k\}}.$$

On a donc que, finalement, $f_{P|X}(p | k) = C p^k (1-p)^{n-k}$ et C ne dépend pas de p . Donc, sachant que $X = k$, on a que P suit une loi Beta de paramètres $k+1, n-k+1$.

Mais bien sûr, $C = \frac{1}{B(k+1, n-k+1)}$, et donc, pour $0 \leq k \leq n$,

$$\begin{aligned}\mathbb{P}\{X = k\} &= B(k+1, n-k+1) \binom{n}{k} \\ &= \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \cdot \frac{n!}{k!(n-k)!} \\ &= \frac{1}{n+1}\end{aligned}$$

Donc, la distribution marginale de X est uniforme sur les entiers de 0 à n .

6.5. Statistiques d'ordre

Considérons un vecteur de n variables aléatoires $(X_1, X_2, X_3, \dots, X_n)$. Notre objectif sera de comprendre ce qui arrive lorsqu'on les place « dans l'ordre ».

On notera $(X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)})$ le **vecteur des statistiques d'ordre**. Ici, pour tout i tel que $1 \leq i \leq n$, $X_{(i)}$ correspond à la i ème plus petite valeur parmi les X_i .

La question maintenant, c'est de comprendre quelle distribution ont les $X_{(i)}$. Dans le cas général, c'est très difficile de dire quoi que ce soit – ça dépend beaucoup de la structure de dépendance entre les variables. Pour la suite des choses, on va supposer que les X_i sont des variables aléatoires continues indépendantes et identiquement distribuées, et que la densité marginale de X_1 (et donc de X_i , pour tout i), est f_1 .

Puisque les variables aléatoires sont indépendantes, la densité jointe est donc

$$f(x_1, x_2, x_3, \dots, x_n) = f_1(x_1)f_1(x_2)f_1(x_3) \cdots f_1(x_n).$$

PROPOSITION 6.9. *Soit $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ un vecteur aléatoire de n variables aléatoires continues indépendantes et identiquement distribuées, et soit f_1 la densité marginale de chacune de ces variables, et f la densité jointe de \mathbf{X} . Soit $\mathbf{X}_() = (X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)})$ le vecteur de statistiques d'ordre associé à \mathbf{X} .*

Alors, la densité jointe $f()$ de $\mathbf{X}_()$ est donnée par

$$(6.5.1) \quad f_()(\mathbf{x}) = n!f_1(x_1)f_1(x_2)f_1(x_3) \cdots f_1(x_n)\mathbb{1}_{\{x_1 < \dots < x_n\}}(\mathbf{x})$$

où $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n) \in \mathbb{R}^n$ et où $\mathbb{1}_{\{x_1 < \dots < x_n\}}$ est une fonction indicatrice qui vaut 1 seulement si les coordonnées de \mathbf{x} sont en ordre strictement croissant, et 0 sinon.

DÉMONSTRATION. *Rappel sur les permutations :* On dit que $\sigma : A \rightarrow A$ est une permutation des éléments de A si σ est une fonction bijective. On peut noter $\mathbb{S}[A]$ l'ensemble des permutations des éléments de A . Dans ce qui suit, on choisira $A = \{1, \dots, n\}$ l'ensemble des n premiers entiers, et on notera $\mathbb{S}_n = \mathbb{S}[A]$, l'ensemble des permutations des n premiers entiers. On note I la permutation identité $i \mapsto i$.

On va noter $\mathbf{X}_\sigma = (X_{\sigma(1)}, X_{\sigma(2)}, X_{\sigma(3)}, \dots, X_{\sigma(n)})$ le vecteur \mathbf{X} , mais où on a ré-ordonné les variables selon la permutation σ .

Étape 1 : \mathbf{X} et \mathbf{X}_σ ont la même loi. Cette preuve sera revisitée en détails à la section 6.7 ; pour l'instant on se contente de le remarquer de façon « intuitive » : les variables aléatoires X_i sont indépendantes et identiquement distribuées. Si on les « ré-étiquette », en permutant les indices, ça ne changera pas le fait que le nouveau vecteur \mathbf{X}_σ est un vecteur de variables aléatoires indépendantes et identiquement distribuées, avec la même distribution marginale pour chacune des entrées. Donc la distribution jointe de X_σ est la même que celle de \mathbf{X} .

Étape 2 : Les ordres possibles. On a que $|\mathbb{S}_n| = n!$ – sans surprise, le nombre de permutations possibles de n entiers est $n!$.

Puisque les variables aléatoires sont continues, pour $i, j \leq n$, avec $i \neq j$, on a que $\mathbb{P}\{X_i = X_j\} = 0$. On peut donc assumer qu'il existe une certaine permutation $\sigma \in \mathbb{S}_n$ telle que

$$X_{\sigma(1)} < X_{\sigma(2)} < X_{\sigma(3)} < \cdots < X_{\sigma(n)}.$$

Cette permutation est aléatoire, bien sûr, mais elle est chaque fois unique – en effet, il ne peut pas en même temps y avoir une permutation $\sigma' \neq \sigma$ qui a aussi la propriété que $X_{\sigma'(1)} < \cdots < X_{\sigma'(n)}$, puisqu'il n'existe qu'une façon de placer les valeurs $X_1, X_2, X_3, \dots, X_n$ en ordre croissant.

En particulier, ce que cela signifie, c'est que, pour un $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$ quelconque avec $x_1 < x_2 < x_3 < \cdots < x_n$, et des $\epsilon_i > 0$ suffisamment petits, on a que

i. D'une part, les événements

$$\{x_i \leq X_{\sigma(i)} < x_i + \epsilon_i, \forall i \leq n\}$$

sont disjoints.

En effet, lorsque les ϵ_i sont plus petits que la plus petite distance $x_{i+1} - x_i$, les intervalles $[x_i, x_i + \epsilon_i)$ sont disjoints, et il ne peut pas y avoir deux ordres différents des valeurs X_i tels qu'elles se retrouvent dans les intervalles $[x_i, x_i + \epsilon_i)$.

ii. D'autre part, la réunion de ces événements est

$$\bigcup_{\sigma \in \mathbb{S}_n} \{x_i \leq X_{\sigma(i)} < x_i + \epsilon_i, \forall i \leq n\} = \{x_i \leq X_{(i)} < x_i + \epsilon_i \forall i \leq n\}.$$

En effet, la seule façon d'avoir que les statistiques d'ordres se retrouvent dans les intervalles $[x_i, x_i + \epsilon_i)$ est qu'il existe au moins une permutation σ pour laquelle les entrées permutées se retrouvent dans ces intervalles. À l'inverse, si on sait qu'il existe une permutation qui place les valeurs des entrées permutées dans ces intervalles, puisqu'ils sont en ordre croissant, on sait alors que les statistiques d'ordre seront dans ces intervalles également.

Étape 3 : Calcul de la densité. Le résultat net, c'est que

$$\mathbb{P}\{x_i \leq X_{(i)} < x_i + \epsilon_i \forall i \leq n\} = \sum_{\sigma \in \mathbb{S}_n} \mathbb{P}\{x_i \leq X_{\sigma(i)} < x_i + \epsilon_i \forall i \leq n\}.$$

Mais les vecteurs \mathbf{X} et \mathbf{X}_σ suivent la même loi ! Donc, pour tout $\sigma \in \mathbb{S}_n$, on a que

$$\mathbb{P}\{x_i \leq X_{\sigma(i)} < x_i + \epsilon_i \forall i \leq n\} = \mathbb{P}\{x_i \leq X_i < x_i + \epsilon_i \forall i \leq n\}.$$

Finalement, on a

$$\mathbb{P}\{x_i \leq X_{(i)} < x_i + \epsilon_i \forall i \leq n\} = n! \mathbb{P}\{x_i \leq X_i < x_i + \epsilon_i \forall i \leq n\}.$$

En divisant chaque côté par $\epsilon_1 \epsilon_2 \epsilon_3 \cdots \epsilon_n$, puis en prenant la limite lorsque $\epsilon = (\epsilon_1, \epsilon_2, \epsilon_3, \dots, \epsilon_n)$ tend vers 0, on trouve que

$$f_{()}(\mathbf{x}) = n! f(\mathbf{x}) \mathbb{1}_{\{x_1 < \dots < x_n\}}(\mathbf{x}),$$

ce qu'il fallait démontrer. □

EXEMPLE 6.22. Le long d'une route de 1 km, trois accidents se produisent à des points X_1, X_2, X_3 indépendants et uniformément distribués.

Trouver la probabilité que les trois accidents se retrouvent tous dans le même segment de longueur d .

SOLUTION. On a que la densité jointe de $\mathbf{X} = (X_1, X_2, X_3)$ est simplement $\mathbb{1}_{(0,1)^3}(x_1, x_2, x_3)$.

Par conséquent, la densité jointe de $\mathbf{X}_{()} = (X_{(1)}, X_{(2)}, X_{(3)})$, le vecteur qui nous indique en ordre de parcours les points où les accidents se sont produits, est $3! \mathbb{1}_{\{0 < x_1 < x_2 < x_3 < 1\}}(x_1, x_2, x_3)$.

On cherche $\mathbb{P}\{X_{(3)} - X_{(1)} \leq d\}$. On calcule donc

$$\begin{aligned} \mathbb{P}\{X_{(3)} - X_{(1)} \leq d\} &= 6 \int_{x_1=0}^1 \int_{x_3=x_1}^{\min\{x_1+d,1\}} \int_{x_2=x_1}^{x_3} dx_2 dx_3 dx_1 \\ &= 6 \left(\int_0^{1-d} \int_{x_1}^{x_1+d} (x_3 - x_1) dx_3 dx_1 + \int_{1-d}^1 \int_{x_1}^1 (x_3 - x_1) dx_3 dx_1 \right) \\ &= 6 \left(\int_0^{1-d} \left(\frac{1}{2}d^2 \right) dx_1 + \int_{1-d}^1 \left(\frac{1}{2}(x_1 - 1)^2 \right) dx_1 \right) \\ &= 6 \left(\frac{1}{2}d^2(1-d) + \frac{1}{2} \int_{-d}^0 u^2 du \right) \\ &= 6 \left(\frac{1}{2}d^2(1-d) + \frac{d^3}{6} \right) \\ &= 3d^2(1-d) + d^3. \end{aligned}$$

6.5.1. La densité marginale de la i ème statistique d'ordre. On s'intéresse maintenant spécifiquement à la loi de $X_{(i)}$, la i ème statistique d'ordre du vecteur $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$, soit la i ème plus petite coordonnée de \mathbf{X} .

Notons $f_{(i)}$ la densité marginale de $X_{(i)}$. On a alors le résultat suivant :

PROPOSITION 6.10. Soit $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ un vecteur de variables aléatoires continues indépendantes et identiquement distribuées, où F_1 est la fonction de répartition marginale des coordonnées individuelles de \mathbf{X} , f_1 est la densité marginale pour une coordonnée. Soit $\mathbf{X}_{()} = (X_{(1)}, X_{(2)}, X_{(3)}, \dots, X_{(n)})$ le vecteur des statistiques d'ordre de \mathbf{X} , avec les densités marginales $f_{(i)}$ et fonctions de répartition marginales $F_{(i)}$ respectivement pour les $X_{(i)}$, et la densité jointe $f_{()}$.

Alors, on a que pour tout i , la densité marginale de $X_{(i)}$ est donnée par

$$(6.5.2) \quad f_{(i)}(x) = \binom{n}{i-1, n-i, 1} (F_1(x))^{i-1} (1 - F_1(x))^{n-i} f_1(x).$$

De plus, la fonction de répartition marginale de $X_{(i)}$ est donnée par

$$(6.5.3) \quad F_{(i)}(x) = \sum_{k=i}^n \binom{n}{k} (F_1(x))^k (1 - F_1(x))^{n-k}$$

DÉMONSTRATION. La preuve peut être effectuée de deux façons. On peut intégrer la densité jointe pour $\mathbf{X}_{()}$. Ça fonctionnerait, mais ça ne serait pas très joli.

À la place, on va simplement raisonner. On commence par prouver l'équation (6.5.2).

Étape 1 : Un événement dont on peut calculer la proba. On va obtenir la densité comme dérive de la fonction de répartition. Donc, on commence en regardant l'événement

$$\{x \leq X_{(i)} < x + \epsilon\}.$$

Pour que la i ème valeur soit entre x et $x + \epsilon$, il faut

- que $i - 1$ valeurs soient plus petites que x ;
- que $n - i$ valeurs soient plus grandes que $x + \epsilon$;
- que 1 valeur soit entre x et $x + \epsilon$.

Encore une fois, notre objectif sera de partitionner notre espace en événements disjoints. On va choisir un ensemble I de $i - 1$ indices parmi $\{1, \dots, n\}$, un ensemble J de $n - i$ indices parmi ceux qui restent, puis finalement un indice k . On définit l'événement

$$E_{I,J,k} = \{X_l < x \ \forall l \in I, X_k \in [x, x + \epsilon), X_l > x + \epsilon \ \forall l \in J\}.$$

Les événements $E_{I,J,k}$ sont disjoints pour tous choix possibles de I, J, k distincts. On a aussi que, pour tous I, J, k respectant les conditions énoncées plus haut,

$$\mathbb{P}\{E_{I,J,k}\} = (F_1(x))^{i-1} (F_1(x + \epsilon) - F_1(x)) (1 - F_1(x + \epsilon))^{n-i}.$$

On remarque également que

$$\bigsqcup_{I,J,k} E_{I,J,k} = \{x \leq X_{(i)} < x + \epsilon\}.$$

Par conséquent,

$$\mathbb{P}\{x \leq X_{(i)} < x + \epsilon\} = \sum_{I,J,k} \mathbb{P}\{E_{I,J,k}\}.$$

Mais la probabilité de $E_{I,J,k}$ est la même pour tout trio I, J, k qui satisfait nos conditions ! Donc il faut simplement compter combien il y a de façons de répartir n indices en un groupe de $i - 1$ (I), un groupe de $n - i$ (J) et un tout seul (k). Bien entendu, il s'agit du coefficient multinomial $\binom{n}{i-1, n-i, 1}$.

Finalement, on a que

$$\mathbb{P}\{x \leq X_{(i)} < x + \epsilon\} = \binom{n}{i-1, n-i, 1} (F_1(x))^{i-1} (F_1(x + \epsilon) - F_1(x)) (1 - F_1(x + \epsilon))^{n-i}.$$

Étape 2 : on prend la limite. Bien sûr, la densité $f_{(j)}(x)$ doit être la limite

$$f_{(j)}(x) = \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{P}\{x \leq X_{(j)} < x + \epsilon\}}{\epsilon}.$$

On calcule donc :

$$\begin{aligned} f_{(j)}(x) &= \lim_{\epsilon \rightarrow 0^+} \binom{n}{i-1, n-i, 1} (F_1(x))^{i-1} (1 - F_1(x + \epsilon))^{n-i} \frac{F_1(x + \epsilon) - F_1(x)}{\epsilon} \\ &= \binom{n}{i-1, n-i, 1} (F_1(x))^{i-1} (1 - F_1(x))^{n-i} f_1(x). \end{aligned}$$

Étape 3 : Pour la fonction de répartition, on fait pareil. On raisonne de façon similaire pour la fonction de répartition. On veut calculer

$$\mathbb{P}\{X_{(i)} \leq x\}.$$

Pour que $X_{(i)} \leq x$, il faut que

- k des coordonnées de \mathbf{X} soient inférieures à x , pour un certain $k \geq i$.
- toutes les $n - k$ coordonnées restantes de \mathbf{X} soient supérieures à x .

Pour un k fixé, on va choisir $I \subseteq \{1, \dots, n\}$ les indices des coordonnées qui seront inférieures ou égales à x , avec $|I| = k$, et $J = \{1, \dots, n\} \setminus I$ les autres coordonnées.

Alors, si on note

$$G_{I,J} = \{X_l \leq x \ \forall l \in I, X_l > x \ \forall l \in J\},$$

on voit que d'une part, les $G_{I,J}$ sont disjoints pour des choix différents de I et J , et

$$\mathbb{P}\{G_{I,J}\} = (F_1(x))^k (1 - F_1(x))^{n-k}.$$

On voit aussi que si $H_k = \bigsqcup_{I,J} G_{I,J}$, alors

$$\mathbb{P}\{H_k\} = \sum_{I,J} \mathbb{P}\{G_{I,J}\} = \binom{n}{k} (F_1(x))^k (1 - F_1(x))^{n-k}.$$

On remarque que H_k correspond à l'événement « il y a exactement k des n coordonnées de \mathbf{X} qui sont inférieures ou égales à x ». Les H_k sont disjoints entre eux pour toutes les valeurs de k fixées. De plus, on remarque que

$$\bigsqcup_{k=i}^n H_k = \{X_{(i)} \leq x\},$$

puisque c'est l'événement « au moins k des n coordonnées de \mathbf{X} sont inférieures ou égales à x ».

Donc,

$$F_{(i)}(x) = \mathbb{P}\{X_{(i)} \leq x\} = \sum_{k=i}^n \binom{n}{k} (F_1(x))^k (1 - F_1(x))^{n-k}.$$

□

EXEMPLE 6.23. Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires indépendantes distribuées uniformément sur l'intervalle $(0, 1)$.

Montrer que la i ème statistique d'ordre $X_{(i)}$ suit une loi Beta de paramètres $(\alpha = i, \beta = n + 1 - i)$.

SOLUTION. Pour voir cela, il suffit de réaliser que la fonction de répartition marginale F pour nos variables X_j est simplement $F(x) = x$ pour $x \in (0, 1)$. De plus, la densité marginale est $\mathbb{1}_{(0,1)}(x)$.

Donc, la densité de la i ème statistique d'ordre est :

$$f_{(i)}(x) = \frac{n!}{(i-1)!(n-i)!} x^{i-1} (1-x)^{(n-i+1)-1} \mathbb{1}_{(0,1)}(x),$$

et en se souvenant que $\Gamma(k+1) = k!$ pour tout $k \geq 0$, on a bien sûr que

$$\frac{n!}{(i-1)!(n-i)!} = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} = \frac{1}{B(i, n-i+1)}.$$

Finalement, la densité marginale de $X_{(i)}$ correspond exactement à celle d'une variable aléatoire de loi Beta avec paramètres $i, n+1-i$.

Comme on s'y attend (peut-être pas?), l'espérance de $X_{(i)}$ est bien sûr

$$\mathbb{E}[X_{(i)}] = \frac{i}{n+1}.$$

6.6. Fonctions de vecteurs aléatoires et changements de variables

Dans cette section nous étudions ce qui se passe lorsqu'on étudie une fonction d'un vecteur de variables aléatoires continues.

Supposons que $\mathbf{X} = (X, Y)$ est un vecteur de variables aléatoires continues de densité jointe $f_{\mathbf{X}}$. On va considérer $\mathbf{U} = (U, V) = g(X, Y)$, avec $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ inversible et différentiable.

On peut écrire

$$U = u(X, Y); \quad V = v(X, Y),$$

avec $u, v : \mathbb{R}^2 \rightarrow \mathbb{R}$.

On va appeler g la **transformation directe** – elle prend pour argument \mathbf{X} , le vecteur dont on connaît la densité, et donne comme valeur \mathbf{U} , un nouveau vecteur aléatoire.

On appelle également $g^{-1} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ la **transformation inverse** – il s'agit de la transformation qui prend pour argument le vecteur \mathbf{U} , puis retourne le vecteur \mathbf{X} .

On va noter $g^{-1}(u, v) = (x(u, v), y(u, v))$, avec $x(u, v)$ et $y(u, v)$ les transformations inverses pour les coordonnées x, y à partir des coordonnées u, v .

Rappel du cours de calcul : Le Jacobien de la transformation g est :

$$\begin{aligned} J_g(x, y) &= \det \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \\ (6.6.1) \qquad \qquad \qquad &= \frac{\partial u}{\partial x} \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} \frac{\partial v}{\partial x} \end{aligned}$$

Le Jacobien de la transformation inverse g^{-1} est :

$$\begin{aligned} J_{g^{-1}}(u, v) &= \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \\ (6.6.2) \qquad \qquad \qquad &= \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \end{aligned}$$

Avec $\mathbf{u} = (u, v)$ et $\mathbf{x} = (x, y)$, on a que $\mathbf{u} = g(\mathbf{x})$, et on a la relation suivantes entre les Jacobiens de g et g^{-1} .

$$(6.6.3) \qquad [J_g(g^{-1}(\mathbf{u}))]^{-1} = \frac{1}{J_g(x(u, v), y(u, v))} = J_{g^{-1}}(u, v) = J_{g^{-1}}(\mathbf{u}).$$

La question maintenant, c'est de savoir comment trouver la densité jointe du vecteur \mathbf{U} . On a la proposition suivante :

PROPOSITION 6.11. *Soit $\mathbf{X} = (X, Y)$ un vecteur de variables aléatoires continues de densité jointe $f_{\mathbf{X}}$ et soit $g : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ un changement de coordonnées inversible différentiable, avec inverse g^{-1} également inversible et différentiable.*

On définit $\mathbf{U} = (U, V)$ un nouveau vecteur de variables aléatoires continues, avec

$$\mathbf{U} = g(\mathbf{X}),$$

où $g(\mathbf{X}) = g(X, Y) = (u(X, Y), v(X, Y))$, avec $u(x, y)$ et $v(x, y)$ les coordonnées de la transformation g . On notera conversement $x(u, v)$ et $y(u, v)$ les coordonnées de la transformation inverse g^{-1} .

Alors, la densité jointe $f_{\mathbf{U}}$ du vecteur \mathbf{U} est donnée par

$$(6.6.4) \qquad f_{\mathbf{U}}(u, v) = f_{\mathbf{X}}(g^{-1}(u, v)) \cdot |J_{g^{-1}}(u, v)|$$

DÉMONSTRATION. Commençons par noter l'analogie quasi-directe avec la proposition 5.12.

La démonstration rigoureuse de la méthode des changements de variables à plusieurs coordonnées est faite dans le cours Analyse III, et est couverte plus sommairement dans le cours de Calcul 1. Nous ne la referons pas ici.

On va toutefois noter l'intuition suivante :

On doit avoir que la probabilité de trouver (U, V) dans la région $[u, u + \epsilon] \times [v, v + \eta]$ est approximativement

$$\mathbb{P}\{U \in [u, u + \epsilon], V \in [v, v + \eta]\} \approx f_{\mathbf{U}}(u, v)\epsilon\eta,$$

par définition de la densité.

Mais d'un autre côté,

$$\mathbb{P}\{(U, V) \in [u, u + \epsilon] \times [v, v + \eta]\} = \mathbb{P}\{(X, Y) \in g^{-1}([u, u + \epsilon] \times [v, v + \eta])\},$$

et lorsque ϵ et η sont petits, le membre de droite est donné par

$$f_{\mathbf{X}}(x(u, v), y(u, v)) \cdot \text{Aire}(g^{-1}([u, u + \epsilon] \times [v, v + \eta])),$$

où $\text{Aire}(g^{-1}([u, u + \epsilon] \times [v, v + \eta]))$ est l'aire de la préimage de $[u, u + \epsilon] \times [v, v + \eta]$.

Or, encore une fois, lorsque ϵ et η sont petits, on a que

$$\text{Aire}(g^{-1}([u, u + \epsilon] \times [v, v + \eta])) = |J_{g^{-1}}(u, v)|\epsilon\eta.$$

On a donc finalement que

$$\mathbb{P}\{(X, Y) \in g^{-1}([u, u + \epsilon] \times [v, v + \eta])\} \approx f_{\mathbf{X}}(x(u, v), y(u, v)) \cdot |J_{g^{-1}}(u, v)|\epsilon\eta,$$

lorsque ϵ et η sont petits. On concluerait en divisant de part et d'autre par $\epsilon\eta$ et en prenant la limite lorsque ϵ et η tendent vers 0.

Alors, on obtiendrait

$$f_{\mathbf{U}}(u, v) = f_{\mathbf{X}}(g^{-1}(u, v)) |J_{g^{-1}}(u, v)|.$$

□

EXEMPLE 6.24. Soit $\mathbf{X} = (X, Y)$ un vecteur de variables aléatoires continues indépendantes et identiquement distribuées avec densité marginale f . Pour $\phi \in [0, 2\pi)$, on définit

$$U = \cos(\phi)X - \sin(\phi)Y; \quad V = \sin(\phi)X + \cos(\phi)Y.$$

Trouver la densité jointe de (U, V) . Calculer explicitement dans le cas où X et Y suivent respectivement une loi normale centrée réduite.

SOLUTION. On remarque que si $u = \cos(\phi)x - \sin(\phi)y$ et $v = \sin(\phi)x + \cos(\phi)y$, alors les transformations inverses sont données par

$$\begin{aligned} \cos(-\phi)u - \sin(-\phi)v &= \cos^2(\phi)x - \sin(\phi)\cos(\phi)y + \sin^2(\phi)x + \sin(\phi)\cos(\phi)y \\ &= x \\ \sin(-\phi)u + \cos(-\phi)v &= -\sin(\phi)\cos(\phi)x + \sin^2(\phi)y + \cos(\phi)\sin(\phi)x + \cos^2(\phi)y \\ &= y \end{aligned}$$

et $|J_{g^{-1}}| = |\cos^2(-\phi) + \sin^2(-\phi)| = 1$.

Puisque

$$f_{\mathbf{X}}(x, y) = f(x)f(y),$$

la densité de \mathbf{U} est donc donnée par

$$f_{\mathbf{U}}(u, v) = f(\cos(-\phi)u - \sin(-\phi)v) \cdot f(\sin(-\phi)u + \cos(-\phi)v).$$

Si X, Y suivent une loi normale centrée réduite, alors

$$f_{\mathbf{X}}(x, y) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(x^2 + y^2) \right],$$

et on voit que notre changement de variables garantit $x^2 + y^2 = u^2 + v^2$. On a alors

$$f_{\mathbf{U}}(u, v) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(u^2 + v^2) \right].$$

On conclue que U et V sont aussi deux variables aléatoires indépendantes, chacune de loi marginale normale centrée réduite.

EXEMPLE 6.25. Soient $\mathbf{X} = (X, Y)$ un vecteur de variables aléatoires distribué.e.s uniformément sur une région $B \subseteq \mathbb{R}^2$ – autrement dit, la densité jointe de \mathbf{X} est

$$f_{\mathbf{X}}(x, y) = \frac{\mathbb{1}_B(x, y)}{\text{Aire}(B)}$$

. Soit $A \in \mathbb{R}^{2 \times 2}$ une matrice inversible. Soit $\mathbf{U} = A\mathbf{X}$.

Montrer que \mathbf{U} est aussi distribué uniformément sur une certaine région B' .

SOLUTION. Supposons que

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Alors, puisque $(U, V) = \mathbf{U} = g(\mathbf{X}) = A\mathbf{X}$, on a que $U = aX + bY$ et $V = cX + dY$. Par conséquent,

$$J_g = ad - bc = C,$$

c'est-à-dire que le Jacobien est une constante C ; donc le jacobien de la transformation inverse est $1/C$ et c'est aussi une constante.

Il suit que la densité du vecteur $\mathbf{U} = A\mathbf{X}$ est une fonction constante sur une certaine région $B' = \{Ax : x \in B\}$.

EXEMPLE 6.26. Soient X, Y deux variables aléatoires indépendantes chacune de loi normale centrée réduite.

On s'intéresse au point $\mathbf{X} = (X, Y)$ aléatoire ainsi choisi, mais on veut décrire la position de ce point en coordonnées polaires – c'est à dire avec $R = \sqrt{X^2 + Y^2}$ la distance du point à l'origine, et Θ l'angle positif entre « l'axe des x » et le rayon reliant \mathbf{X} à l'origine, dans le sens anti-horaire.

Bien sûr, les transformations inverses sont données par

$$x = r \cos \theta; \quad y = r \sin \theta.$$

On a donc que le Jacobien des transformations inverses est

$$J_{g^{-1}} = \frac{\partial x}{\partial r} \frac{\partial y}{\partial \theta} - \frac{\partial x}{\partial \theta} \frac{\partial y}{\partial r} = r \cos^2 \theta + r \sin^2 \theta = r.$$

On a que la densité de \mathbf{X} est, pour $x \in \mathbb{R}, y \in \mathbb{R}$,

$$f_{\mathbf{X}}(x, y) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(x^2 + y^2) \right].$$

En utilisant nos transformations, on trouve que

$$f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} r \exp\left[-\frac{1}{2}r^2\right],$$

pour $r > 0$ et $\theta \in [0, 2\pi]$.

On remarque que cette densité est factorisable :

$$f_{R,\Theta}(r, \theta) = \left(\frac{1}{2\pi} \mathbb{1}_{[0,2\pi)}(\theta)\right) \cdot \left(re^{-\frac{1}{2}r^2} \mathbb{1}_{[0,\infty)}(r)\right),$$

Et il s'ensuit que R et Θ sont des variables aléatoires indépendantes, et que Θ est de distribution uniforme sur $[0, 2\pi]$.

La distribution de R est un peu plus étrange – sa densité marginale est donnée par

$$f_R(r) = re^{-\frac{1}{2}r^2} \mathbb{1}_{(0,\infty)}(r).$$

Par contre, on remarque que, puisque R est positive, la densité marginale de R^2 est

$$\begin{aligned} f_{R^2}(x) &= \frac{d}{dx} \mathbb{P}\{R^2 \leq x\} \\ &= \frac{d}{dx} \mathbb{P}\{R^2 \leq \sqrt{x}\} \\ &= f_R(\sqrt{x}) \cdot \frac{1}{2\sqrt{x}} \\ &= \frac{1}{2} e^{-\frac{1}{2}x}. \end{aligned}$$

Autrement dit, R^2 suit une loi exponentielle de paramètre $\lambda = \frac{1}{2}$. Donc, la variable aléatoire $\frac{1}{2}R^2$ suit une loi exponentielle standard.

Autre fait intéressant : $\tan \Theta = \frac{X}{Y}$. Mais puisqu'on sait que Θ est de loi uniforme sur $(0, 2\pi)$, clairement $(\Theta - \pi) \pmod{\pi/2}$ est uniforme sur $(-\pi/2, \pi/2)$ et $\tan \Theta = \tan(\Theta - \pi \pmod{\pi/2})$ suit une distribution de Cauchy (voir l'exemple 5.13).

EXEMPLE 6.27. On considère deux variables aléatoires X, Y indépendantes qui suivent respectivement des lois Gamma de paramètres (α, λ) et (β, λ) .

On définit $U = X + Y$ et $V = \frac{X}{X+Y}$.

Trouver la densité jointe de U, V , et les densités marginales de U et V . Est-ce que U et V sont indépendantes ?

SOLUTION. On a bien sûr que la densité jointe de $\mathbf{X} = (X, Y)$ est

$$f_{\mathbf{X}}(x, y) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)}. \quad x, y > 0$$

Les transformations directes de coordonnées sont $u = x + y$ et $v = x/(x + y)$. Les transformations inverses sont donc

$$x = uv; \quad y = u(1 - v).$$

On va avoir que (U, V) sera supporté sur $(0, \infty) \times (0, 1)$. Donc, la valeur absolue du Jacobien de la transformation inverse

$$|J_{g^{-1}}(u, v)| = |u(-v) - u(1 - v)| = u,$$

Finalement, la densité jointe de (U, V) est donnée par

$$f_{\mathbf{U}}(u, v) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} u(uv)^{\alpha-1} (u(1-v))^{\beta-1} e^{-\lambda u}.$$

En réarrangeant et en factorisant, on trouve

$$f_{\mathbf{U}}(u, v) = \left(\frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} u^{\alpha+\beta-1} e^{-\lambda u} \right) \cdot \left(\frac{1}{B(\alpha, \beta)} v^{\alpha-1} (1-v)^{\beta-1} \right).$$

On a alors évidemment que $f_U(u) = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)} u^{\alpha+\beta-1} e^{-\lambda u}$ et $f_V(v) = \frac{1}{B(\alpha, \beta)} v^{\alpha-1} (1-v)^{\beta-1}$; U et V sont indépendantes et U suit une loi $\Gamma(\alpha + \beta)$ (on le savait déjà à cause de la proposition 6.5), et V suit une loi Beta de paramètres α, β .

6.6.1. Généralisation à n variables aléatoires. Jusqu'ici on a travaillé seulement avec des vecteurs aléatoires à deux coordonnées aléatoires X et Y .

On peut généraliser sans difficulté la proposition 6.11 à des vecteurs à n coordonnées aléatoires.

PROPOSITION 6.12. Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vecteur à n coordonnées aléatoires, de densité jointe $f_{\mathbf{X}}$. Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une transformation de coordonnées différentiable et inversible, et g^{-1} sa transformation inverse.

On définit $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$ un vecteur aléatoire, par $Y_i = g_i(\mathbf{X})$ pour $1 \leq i \leq n$, où g_i est la transformation de la i ème coordonnée. On notera g_i^{-1} la transformation inverse de la i ème coordonnée.

Le Jacobien de la transformation inverse est donnée par le déterminant

$$J_{g^{-1}}(\mathbf{y}) = \det \left(\frac{\partial g_i^{-1}}{\partial y_j} \right)_{1 \leq i, j \leq n} = \det \begin{pmatrix} \frac{\partial g_1^{-1}}{\partial y_1} & \cdots & \frac{\partial g_1^{-1}}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n^{-1}}{\partial y_1} & \cdots & \frac{\partial g_n^{-1}}{\partial y_n} \end{pmatrix}.$$

La densité jointe du vecteur $\mathbf{Y} = g(\mathbf{X})$ est alors donnée par

$$(6.6.5) \quad f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(g^{-1}(\mathbf{y})) \cdot |J_{g^{-1}}(\mathbf{y})|.$$

DÉMONSTRATION. La preuve suit les mêmes idées que celle pour deux variables aléatoires. \square

6.7. Variables aléatoires interchangeables

Pour terminer ce chapitre, nous allons nous pencher sur un cas très particulier du changement de coordonnées pour n variables aléatoires : les permutations des coordonnées.

6.7.1. Les permutations. Pour un ensemble S quelconque, on dit que $\sigma : S \rightarrow S$ est une **permutation** des éléments de S lorsque σ est une fonction bijective. En effet, si tel est le cas, alors l'effet de σ est bel et bien de « permutez » les éléments de S entre eux.

Le plus souvent, on va considérer \mathbb{S}_n , l'ensemble des permutations des entiers $\{1, \dots, n\}$.⁷

7. \mathbb{S}_n est en fait ce qu'on appelle un *groupe*, lorsqu'on le considère avec l'opération de composition \circ entre les permutations. Plus de détails à venir dans un cours d'Algèbre I!

EXEMPLE 6.28. Une permutation des entiers $\{1, \dots, n\}$ peut être représentée par un ensemble de cycles qui indiquent quel nombre est envoyé vers quel nombre. On noterait par exemple :

$$\sigma = (1, 4, 5)(2, 3)(6)$$

la permutation des entiers de 1 à 6 telle que

$$\begin{array}{lll} \sigma(1) = 4 & \sigma(4) = 5 & \sigma(5) = 1 \\ \sigma(2) = 3 & \sigma(3) = 2 & \sigma(6) = 6 \end{array}$$

Le cours d'Algèbre I passe beaucoup de temps à décrire en grand détail l'ensemble \mathbb{S}_n et sa structure particulière. Pour nous, il s'agira simplement d'un ensemble intéressant.

6.7.2. Permutation des coordonnées. À la section 6.5, on avait introduit l'idée du vecteur aléatoire $\mathbf{X}_\sigma = (X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)})$ qui, pour un certain vecteur aléatoire $\mathbf{X} = (X_1, X_2, \dots, X_n)$ et une permutation σ , consistait simplement à considérer les mêmes variables aléatoires, mais en « changeant les variables de place dans le vecteur ».

L'idée, maintenant, c'est d'analyser cette situation comme un changement de coordonnées. Pour une permutation $\sigma \in \mathbb{S}_n$, et les notations $\mathbf{x} = (x_1, x_2, \dots, x_n)$ et $\mathbf{x}_\sigma = (x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$, on définit le changement de coordonnées

$$(6.7.1) \quad \pi_\sigma(\mathbf{x}) := \mathbf{x}_\sigma.$$

Les changements de coordonnées individuels (directs et inverses) sont :

$$(6.7.2) \quad \pi_{\sigma,i}(\mathbf{x}) = x_{\sigma(i)}; \quad \pi_{\sigma,i}^{-1}(\mathbf{y}) = y_{\sigma^{-1}(i)}.$$

Ce que ces changements de coordonnées ont de particulier, c'est que

$$\left| J_{\pi_\sigma^{-1}}(\mathbf{y}) \right| = 1,$$

pour tout $\sigma \in \mathbb{S}_n$.

Dès lors, par la proposition 6.12, si on fait la transformation $\mathbf{Y} = \pi_\sigma(\mathbf{X})$, tout ce qui change dans la densité, c'est qu'on doit permute les coordonnées.

EXEMPLE 6.29. Soient X_1, X_2, X_3 trois variables aléatoires avec densité jointe

$$f_{\mathbf{X}}(x_1, x_2, x_3) = Cx_1x_2 + x_3, \quad x_1, x_2, x_3 > 2, x_1 + x_2 + x_3 \leq 6,$$

pour une certaine constante C .

Alors, si on définit $Y_1 = X_2, Y_2 = X_3, Y_3 = X_1$, la densité jointe de (Y_1, Y_2, Y_3) est simplement

$$f_{\mathbf{Y}}(y_1, y_2, y_3) = Cy_2y_3 + y_1, \quad y_2, y_3, y_1 > 2, y_2 + y_3 + y_1 \leq 6.$$

6.7.3. L'interchangeabilité. Jusqu'ici, tout est relativement simple. Mais les résultats seront frappants. On va maintenant utiliser les symétries de la densité $f_{\mathbf{X}}$ à notre avantage. Définissons d'abord l'invariance de f sous permutation de ses coordonnées :

DÉFINITION 6.6 (Invariance sous permutation σ). Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction.

On dit que f est **invariante sous permutation** σ de ses coordonnées si et seulement si

$$(6.7.3) \quad f(\mathbf{x}) = f(\pi_\sigma(\mathbf{x})).$$

Si $G \subseteq \mathbb{S}_n$ est un ensemble de permutations, on dit que f est invariante sous les permutations de G si et seulement si f est invariante sous les permutations σ de ses coordonnées pour tout $\sigma \in G$.⁸

On va maintenant voir ce que ça signifie pour nos vecteurs aléatoires.

DÉFINITION 6.7 (Interchangeabilité). Soit $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ un vecteur aléatoire avec fonction de répartition jointe $F_{\mathbf{X}}$.

On dit que les variables aléatoires X_i et X_j sont **interchangeables** dans \mathbf{X} si il existe une permutation $\sigma \in \mathbb{S}_n$ telle que

- $F_{\mathbf{X}}$ est invariante sous permutation de ses coordonnées par σ ;
- $\sigma(i) = j$.

REMARQUE. Il est facile de montrer que,

- si \mathbf{X} est un vecteur de variables aléatoires continues, on peut remplacer $F_{\mathbf{X}}$ par $f_{\mathbf{X}}$ la densité jointe dans la définition 6.7 ;
- si \mathbf{X} est un vecteur de variables aléatoires discrètes, on peut remplacer $F_{\mathbf{X}}$ par $p_{\mathbf{X}}$ la fonction de masse de probabilité jointe.

EXEMPLE 6.30. Supposons qu'on considère le vecteur $\mathbf{X} = (X_1, X_2, X_3)$, avec la densité jointe

$$f_{\mathbf{X}} = C(x_1x_2 + x_2x_3 + x_3x_1), (x_1, x_2, x_3) \in (0, 1)^3.$$

Alors, toutes les paires de variables sont interchangeables entre elles. En effet, il suffit soit de choisir $\sigma = (1, 2, 3)$ ou $\sigma = (1, 3, 2)$, et on a immédiatement que la densité $f_{\mathbf{X}}$ est invariante sous ces permutations de ses coordonnées, mais aussi que $\sigma(i) = j$ pour n'importe quelle paire i, j .

EXEMPLE 6.31. Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vecteur de n variables aléatoires indépendantes et identiquement distribuées. Alors, toutes les variables de \mathbf{X} sont interchangeables deux à deux.

L'intérêt principal de la notion d'interchangeabilité, c'est la proposition suivante :

PROPOSITION 6.13. Soit $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ un vecteur aléatoire avec fonction de répartition jointe $F_{\mathbf{X}}$.

Si les variables aléatoires X_i et X_j sont interchangeables, alors elles ont la même distribution marginale.

8. On peut montrer que l'ensemble maximal $\hat{G} \subseteq \mathbb{S}_n$ de toutes les permutations de coordonnées qui laissent f invariantes forme un *sous-groupe* de \mathbb{S}_n . Encore une fois, Algèbre I ! Si ça ne vous dit rien, pas de souci !

DÉMONSTRATION. Si σ est notre permutation qui laisse $F_{\mathbf{X}}$ invariant et avec $\sigma(i) = j$, alors :

$$\begin{aligned}\mathbb{P}\{X_j \leq x\} &= \lim_{\substack{x_k \rightarrow \infty \\ \forall k \neq j}} F_{\mathbf{X}}(\mathbf{x}) \\ &= \lim_{\substack{x_k \rightarrow \infty \\ \forall k \neq j}} F_{\mathbf{X}}(\pi_{\sigma}(\mathbf{x})) \\ &= \lim_{\substack{x_k \rightarrow \infty \\ \forall k \neq \sigma^{-1}(j)}} F_{\mathbf{X}}(\mathbf{x}) \\ &= \lim_{\substack{x_k \rightarrow \infty \\ \forall k \neq i}} F_{\mathbf{X}}(\mathbf{x}) \\ &= \mathbb{P}\{X_i \leq x\}.\end{aligned}$$

□

6.7.4. Pour faire une histoire courte ... Ce qui est important de retenir de cette section est très simple, même si le jargon technique pour faire les preuves peut être un peu aride.

- i. On dit que des variables sont **interchangeables** si on peut échanger les variables dans la fonction de répartition (resp. densité, fonction de masse de probabilité), sans que la fonction ne change.
- ii. Lorsque deux variables sont interchangeables, alors elles ont forcément la même distribution marginale.

EXEMPLE 6.32. Dans notre exemple précédent, où on avait que $\mathbf{X} = (X_1, X_2, X_3)$ avait la densité

$$f_{\mathbf{X}}(x_1, x_2, x_3) = C(x_1 x_2 + x_2 x_3 + x_3 x_1), \quad (x_1, x_2, x_3) \in (0, 1)^3,$$

toutes les variables étaient interchangeables entre elles. Au final, donc, toutes les variables ont la même distribution marginale.

REMARQUE. *Attention !* Les variables dans cet exemple ne sont pas indépendantes ! Mais elles sont interchangeables grâce à la symétrie de la densité.

EXEMPLE 6.33. Une carte de BINGO est formée de cinq colonnes de cinq cases, chacune arborant un nombre différent compris entre 1 et 75. Chaque colonne est étiquetée, dans l'ordre, de l'une des lettres du mot BINGO.

- La colonne du B contient des nombres de 1 à 15.
- La colonne du I contient des nombres de 16 à 30.
- La colonne du N contient des nombres de 31 à 45.
- La colonne du G contient des nombres de 46 à 60.
- La colonne du O contient des nombres de 61 à 75.

Le boulier contient 75 boules numérotées, et pour retrouver plus facilement les nombres sur la carte, l'annonceur annonce d'abord la lettre de la colonne, inscrite sur la boule.

Si on tire 8 boules de Bingo de suite sans remise, et qu'on considère $(X_B, X_I, X_N, X_G, X_O)$ le nombre de boules sorties respectivement pour les colonnes B, I, N, G et O, alors

- Clairement, les variables X_B, X_I, X_N, X_G et X_O ne sont pas indépendantes – si on pige plus de boules en B on en pige forcément moins d'une autre colonne.
- Par contre, ces variables sont interchangeables. En effet, leur fonction de masse est

$$p(b, i, n, g, o) = \frac{\binom{15}{b} \binom{15}{i} \binom{15}{n} \binom{15}{g} \binom{15}{o}}{\binom{75}{8}},$$

et on voit que n'importe quelle permutation des arguments b, i, n, g, o dans la fonction de masse jointe nous redonne exactement la même expression.

EXEMPLE 6.34. On considère un vecteur aléatoire $\mathbf{X} = (X_1, X_2, X_3)$ avec fonction de densité marginale

$$f(x_1, x_2, x_3) = C(x_1^{x_2} + x_2^{x_1} + x_3), \quad 0 < x_1, x_2, x_3 < 1.$$

Trouver $\mathbb{P}\{X_1 > X_2\}$ sans calcul.

SOLUTION. Cette densité serait vraiment pénible à intégrer. Il y aurait des logarithmes au dénominateur, tout un tas de trucs déplaisants...

Heureusement, on sait que les variables aléatoires X_1 et X_2 sont interchangeables ! Cela signifie que le vecteur $(Y_1 = X_2, Y_2 = X_1, Y_3 = X_3)$ a la même densité jointe que le vecteur (X_1, X_2, X_3) .

On a donc

$$\mathbb{P}\{X_2 > X_1\} = \mathbb{P}\{Y_1 > Y_2\} = \mathbb{P}\{X_1 > X_2\},$$

mais on doit avoir que

$$1 = \mathbb{P}\{X_1 > X_2\} + \mathbb{P}\{X_2 > X_1\} + \mathbb{P}\{X_1 = X_2\},$$

et puisque $\mathbb{P}\{X_1 = X_2\} = 0$ (vu que X_1 et X_2 sont continues, on a donc que

$$\mathbb{P}\{X_1 > X_2\} = \mathbb{P}\{X_2 > X_1\} = \frac{1}{2}.$$

6.8. Exercices

EXERCICE 6.1. Armé d'un algorithme choisi, un ordinateur peut facilement générer une approximation d'une suite de variables aléatoires uniformes sur $(0, 1)$ indépendantes. La question, c'est de savoir comment un ordinateur pourrait faire pour choisir aléatoirement k objets parmi n , de sorte que toutes les combinaisons sont équiprobables.

Par exemple, au jeu *Démineur*, l'ordinateur doit sélectionner k cases parmi $n = L \times H$ cases (dans une grille de $L \times H$ cases).

Comment faire pour choisir exactement k cases distinctes parmi n , de telle sorte que toutes les combinaisons soient équiprobables ?

Note : Ceci est un exemple tiré du manuel de Ross. Vous pouvez donc aller lire la solution par vous-mêmes si vous le désirez.

EXERCICE 6.2. Trois boules sont choisies au hasard sans remise dans une urne de 5 boules jaunes et 8 boules pourpres. Soit X_i une variable aléatoire de Bernoulli qui vaut 1 si la i ème boule pigée est pourpre.

- (a) Trouver les fonctions de masse jointes de (X_1, X_2) .
- (b) Est-ce que X_1 et X_2 sont indépendantes ?
- (c) Est-ce que X_1 et X_2 sont interchangeables ?

EXERCICE 6.3. Soient X, Y deux variables aléatoires avec densité jointe

$$f(x, y) = e^{-x-y}, \quad x, y > 0.$$

- (a) Trouver $\mathbb{P}\{X > Y\}$.
- (b) Trouver $\mathbb{P}\{X \leq a\}$.

EXERCICE 6.4. Le nombre de personnes qui entrent dans l'autobus à l'arrêt au coin de Beaubien et Papineau est une variable aléatoire de loi Poisson avec paramètre $\lambda = 10$. Supposons que chacune de ces personnes porte un chapeau indépendamment avec probabilité $p = 1/10$.

Sachant que 12 personnes sont entrées sans chapeau, donner la probabilité qu'au moins trois personnes sont entrées en portant un chapeau.

EXERCICE 6.5. Le vecteur $\mathbf{X} = (X, Y)$ est dit être uniformément distribué sur une région $R \subseteq \mathbb{R}^2$ du plan si et seulement si sa densité est constante dans cette région et nulle partout ailleurs. Autrement dit, si f est la densité jointe de \mathbf{X} , on a que

$$f(\mathbf{x}) = \begin{cases} c & \text{si } \mathbf{x} \in R \\ 0 & \text{sinon.} \end{cases}$$

- (a) Montrer que $1/c = \text{Aire}(R)$.
- (b) Supposons que $R = (-1, 1)^2$ est le carré où x et y sont comprises entre -1 et 1 .
- (c) Montrer que X et Y sont indépendantes et uniformément distribuées sur $(-1, 1)$.

- (c) Quelle est la probabilité que la distance de (X, Y) à l'origine soit inférieure à 1 ?

REMARQUE. En utilisant ce résultat en conjonction avec la loi des grands nombres (théorèmes 8.1 et 8.2), on peut calculer π en lançant au hasard des cailloux sur une grande cible carrée, puis en comptant combien d'entre eux tombent dans le cercle inscrit à l'intérieur !

EXERCICE 6.6. Supposons que l'on choisisse n points indépendamment et uniformément sur la circonférence d'un cercle.

On cherche la probabilité qu'il existe un diamètre du cercle qui sépare le cercle en deux demi-cercles, dont l'un contient tous les points qu'on a choisis. On va dénoter les points par P_1, \dots, P_n .

Soit A l'événement qu'il existe un diamètre qui sépare le cercle en deux demi-cercles dont l'un contient tous les points.

On va noter A_i l'événement que pour tout $j \neq i$, l'angle (dans le sens anti-horaire) entre les points P_i et P_j est inférieur à π .

- (a) Donner A en termes des A_i .
- (b) Les A_i sont ils disjoints ? Presque-disjoints ?
- (c) Trouver $\mathbb{P}\{A\}$.

EXERCICE 6.7. On sélectionne trois points X_1, X_2, X_3 de façon indépendante et uniforme sur un segment de droite de longueur L .

Trouver la probabilité que X_2 soit entre X_1 et X_3 .

EXERCICE 6.8. Soit (X, Y) un vecteur aléatoire avec la densité jointe

$$f(x, y) = \frac{C}{x}, \quad 0 < y < x < 1.$$

- (a) Trouver C .
- (b) Trouver les densités marginales de X et Y .
- (c) Trouver $\mathbb{E}[X]$ et $\mathbb{E}[Y]$.
- (d) Trouver la densité conditionnelle de Y sachant X .

EXERCICE 6.9. Soient X, Y deux variables aléatoires indépendantes de lois exponentielles de paramètres λ et μ respectivement. On définit $Z = \frac{X}{Y}$.

- (a) Trouver $\mathbb{P}\{X < Y\}$.
- (b) Trouver la loi de Z .

EXERCICE 6.10. Soient X, Y deux variables aléatoires indépendantes continues et positives de densités marginales respectives f_X et f_Y .

- (a) Exprimer la densité de $Z = X/Y$ en fonction des densités de X et Y .

- (b) Exprimer la densité de $W = XY$ en fonction des densités de X et Y .

EXERCICE 6.11. Soient X, Y deux variables aléatoires continues avec densité jointe f . Montrer que la densité de $Z = X + Y$ est donnée par

$$f_Z(z) = \int_{-\infty}^{\infty} f(t, z-t) dt.$$

EXERCICE 6.12. Soient X, Y deux variables aléatoires normales centrées réduites indépendantes.

On définit $U = X$ et $V = \frac{X}{Y}$.

- (a) Trouver la densité jointe du vecteur (U, V) .
 (b) Utiliser ce résultat pour montrer que X/Y suit une distribution de Cauchy.

EXERCICE 6.13. Une ambulance fait des allers-retours sur une portion de route de longueur L . À un moment donné, un accident se produit à quelque part sur la route.

On note A la position de l'ambulance, et X la position de l'accident. On présume que A et X sont indépendantes et distribuées uniformément.

Déterminer la distribution de la distance $|X - A|$ entre l'ambulance et l'accident.

EXERCICE 6.14. On tire sans remise 3 boules d'une urne qui en contient 5 jaunes et 8 pourpres. Soit X_i une variable aléatoire de Bernoulli qui vaut 1 lorsque la i ème boule est pourpre, et 0 si elle est jaune.

Donner la fonction de masse conditionnelle de X_1 sachant

- (a) que $X_2 = 1$;
 (b) que $X_2 = 0$;

EXERCICE 6.15. La fonction de masse jointe de (X, Y) est donnée par

$$\begin{array}{ll} p(1, 1) = \frac{1}{8} & p(1, 2) = \frac{1}{4} \\ p(2, 1) = \frac{1}{8} & p(2, 2) = \frac{1}{2}. \end{array}$$

- (a) Trouver la fonction de masse conditionnelle de X sachant $Y = i, i = 1, 2$.
 (b) X et Y sont-elles indépendantes ?
 (c) Trouver $\mathbb{P}\{XY \leq 3\}$.
 (d) Trouver $\mathbb{P}\{X + Y > 2\}$.
 (e) Trouver $\mathbb{P}\{X/Y > 1\}$.

EXERCICE 6.16. Soient X, Y deux variables aléatoires avec densité jointe

$$f(x, y) = C(x^2 - y^2)e^{-x}, \quad 0 < |y| < x < +\infty.$$

Trouver la densité conditionnelle de Y sachant X .

EXERCICE 6.17. Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes avec fonction de répartition marginale F .

(a) Trouver la fonction de répartition de $\max\{X_i : i \leq n\}$.

(b) Trouver la fonction de répartition de $\min\{X_i : i \leq n\}$.

Supposons maintenant que $F(x) = 1 - e^{-\lambda x}$ pour $x > 0$, et 0 pour $x \leq 0$.

(c) Donner les fonctions de répartition pour $\max\{X_i : i \leq n\}$ et $\min\{X_i : i \leq n\}$.

EXERCICE 6.18. Soient U, V deux variables aléatoires uniformes sur $(0, 1)$ indépendantes. On construit

$$\begin{aligned} X &= \sqrt{-2 \log U} \cos(2\pi V) \\ Y &= \sqrt{-2 \log U} \sin(2\pi V). \end{aligned}$$

Montrer que X et Y sont deux variables aléatoires normales centrées réduites indépendantes.

REMARQUE. Ce résultat est pratique puisqu'il permet de générer facilement des variables aléatoires de distribution normale avec un ordinateur – pourvu que l'ordinateur sache calculer des racines, des logarithmes et des fonctions trigonométriques.

EXERCICE 6.19. Soient X, Y deux variables aléatoires exponentielles indépendantes de paramètres respectifs λ et μ .

On définit $Z = X + Y$, et $W = e^X$.

Trouver la densité jointe de (Z, W) .

EXERCICE 6.20. Soient $X_{(1)} < X_{(2)} < X_{(3)} < \dots < X_{(n)}$ les statistiques d'ordre du vecteur aléatoire $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$, où les X_i sont des variables aléatoires uniformes sur $(0, 1)$ indépendantes.

Montrer que pour $k = 1, \dots, n+1$, on a que

$$\mathbb{P}\{X_{(k)} - X_{(k-1)} > t\} = (1-t)^n,$$

avec $X_{(0)} = 0$ et $X_{(n+1)} = 1$.

EXERCICE 6.21. Soient X_1, \dots, X_n des variables aléatoires indépendantes identiquement distribuées avec fonction de répartition marginale F , et $X_{(1)}, \dots, X_{(n)}$ les statistiques d'ordre.

Si Y est une variable aléatoire telle que $(X_1, X_2, \dots, X_n, Y)$ forme également une famille indépendante, et que la fonction de répartition marginale de Y est aussi F ,

- (a) déterminer $\mathbb{P}\{Y > X_{(n)}\}$;
- (b) déterminer $\mathbb{P}\{Y > X_{(1)}\}$;
- (c) déterminer $\mathbb{P}\{X_{(i)} < Y < X_{(j)}\}$, pour $i < j$.

EXERCICE 6.22. Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées avec fonction de répartition marginale F et densité marginale f .

On appelle **milieu d'étendue** (en anglais *midrange*) la variable aléatoire

$$M = \frac{X_{(1)} + X_{(n)}}{2},$$

qui donne la moyenne du minimum et du maximum d'un échantillon.

Montrer que la fonction de répartition de M est donnée par

$$F_M(m) = n \int_{-\infty}^m [F(2m - x) - F(x)]^{n-1} f(x) dx.$$

EXERCICE 6.23. Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées avec fonction de répartition marginale F et densité marginale f .

- (a) Trouver $\mathbb{P}\{X_n > X_1 \mid X_1 = \max\{X_i, i \leq n-1\}\}$.
- (b) Trouver $\mathbb{P}\{X_n > X_2 \mid X_1 = \max\{X_i, i \leq n-1\}\}$.

EXERCICE 6.24. À l'exemple 6.12, on définit une suite $(T_i)_{i \in \mathbb{N}}$ de variables aléatoires indépendantes et identiquement distribuées, de loi exponentielle avec paramètre λ .

Intuitivement, on voit T_i comme le temps d'attente entre la $(i-1)$ ième et l' i ème occurrence d'un événement imprévisible. T_1 est le temps d'attente avant la première occurrence.

Définissons maintenant $S_n = \sum_{i=1}^n T_i$. On définit également $N(t) = \max\{n \in \mathbb{N} : S_n \leq t\}$.

- (a) Quelle interprétation peut-on faire de S_n ?
- (b) Quelle interprétation peut-on faire de $N(t)$?
- (c) Expliquer pourquoi $\{N(t) \leq n\} = \{S_{n+1} > t\}$.
- (d) Calculer $\mathbb{P}\{N(t) \leq n\}$ et $\mathbb{P}\{N(t) = n\}$ et déduire que $N(t)$ suit une loi de Poisson de paramètre λt .

REMARQUE. Le processus $N(t)$ est appelé « processus de points de Poisson ». C'est un processus central à l'étude des processus stochastiques à temps continu.

EXERCICE 6.25. Montrer le corollaire 6.1 :

COROLLAIRE. Soient X_1, \dots, X_n des variables aléatoires indépendantes, telles que X_i suit une loi Gamma de paramètres (α_i, λ) pour tout i .

Alors, $S_n = \sum_{k=1}^n X_k$ suit une loi Gamma de paramètres $(\sum_{k=1}^n \alpha_k, \lambda)$.

EXERCICE 6.26. Montrer le corollaire 6.2 :

COROLLAIRE. Si X_1, X_2, \dots, X_n sont une famille de variables aléatoires indépendante de lois marginales normales, avec $\mathbb{E}[X_i] = \mu_i$ et $\text{Var}[X_i] = \sigma_i^2$ pour tout i de 1 à n , alors $S_n = \sum_{i=1}^n X_i$ suit une loi normale avec paramètres $\mu = \sum_{i=1}^n \mu_i$ et $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.

EXERCICE 6.27. Montrer le corollaire 6.3 :

COROLLAIRE. Soient X_1, X_2, \dots, X_n des variables binomiales indépendantes, respectivement de paramètres (n_i, p) pour tout i .

Alors, $S_k = \sum_{i=1}^k X_i$ est une variable aléatoire binomiale de paramètres $(n = \sum_{i=1}^k n_i, p)$.

Troisième partie

Les outils analytiques

Chapitre 7

L'espérance

Notablement absente du chapitre 6, l'espérance fait maintenant son grand retour. Dans les sections qui suivent, nous allons voir comment déterminer pour des fonctions de plusieurs variables aléatoires. Nous allons découvrir des méthodes pour calculer les espérances plus facilement, mais nous allons aussi voir comment on peut utiliser l'espérance pour faire d'autres calculs.

7.1. L'espérance de fonctions de plusieurs variables aléatoires

Bien sûr, on sait déjà comment calculer l'espérance d'une fonction d'une seule variable aléatoire – si X est une variable continue (resp. discrète) de densité f_X (resp. fonction de masse p_X), alors

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx \quad \left(\text{resp. } \sum_{v \in V} g(v)p_X(v) \right).$$

Par conséquent, si on a deux variables aléatoires (disons, X et Y), il suffit de trouver leur distribution marginale puis de calculer leurs espérances à partir de là.

Le souci avec cette méthode, c'est qu'elle ne fonctionne pas très bien pour les variables aléatoires « construites » à partir de plusieurs variables aléatoires.

Par exemple, comment ferait-on pour calculer $\mathbb{E}[XY]$?

On pourrait décider de trouver la densité de XY d'abord, puis de calculer. Mais il y a plus simple :

PROPOSITION 7.1. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire de variables continues avec densité jointe f . Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles.*

On note $Y = g(\mathbf{X})$. On a que

$$(7.1.1) \quad \mathbb{E}[Y] = \mathbb{E}[g(\mathbf{X})] = \iiint \cdots \int_{\mathbb{R}^n} g(\mathbf{x})f(\mathbf{x})d^n\mathbf{x}$$

DÉMONSTRATION. On commence par montrer que c'est vrai pour g positive, comme pour la proposition 5.2.

Si g est positive,

$$\mathbb{E}[g(\mathbf{X})] = \int_0^{\infty} \mathbb{P}\{g(\mathbf{X}) > u\} du.$$

On écrit

$$\mathbb{P}\{g(\mathbf{X}) > u\} = \iiint \cdots \int_{\mathbf{x}: g(\mathbf{x}) > u} f(\mathbf{x})d^n\mathbf{x}.$$

On se retrouve alors avec

$$\begin{aligned}\mathbb{E}[g(\mathbf{X})] &= \int_0^\infty \left(\iiint \cdots \int_{\mathbf{x}: g(\mathbf{x}) > u} f(\mathbf{x}) d^n \mathbf{x} \right) du \\ &= \iiint \cdots \int_{\mathbb{R}^n} \left(\int_0^{g(\mathbf{x})} du \right) f(\mathbf{x}) d^n \mathbf{x} \\ &= \iiint \cdots \int_{\mathbb{R}^n} g(\mathbf{x}) f(\mathbf{x}) d^n \mathbf{x}.\end{aligned}$$

La preuve générale pour g quelconque procède de façon exactement identique à la preuve de la proposition 5.2. \square

EXEMPLE 7.1. Soient X et Y deux variables aléatoires indépendantes uniformément distribuées sur l'intervalle $(0, 1)$.

On cherche $\mathbb{E}[|X - Y|]$.

SOLUTION. On a bien sûr que

$$\begin{aligned}\mathbb{E}[|X - Y|] &= \int_{-\infty}^\infty \int_{-\infty}^\infty |x - y| \mathbb{1}_{(0,1)^2}(x, y) dy dx \\ &= \int_0^1 \int_0^1 |x - y| dy dx \\ &= \int_0^1 \left(\int_0^x (x - y) dy + \int_x^1 (y - x) dy \right) dx \\ &= \int_0^1 \left(\frac{x^2}{2} + \frac{1 - x^2}{2} - x(1 - x) \right) dx \\ &= \int_0^1 \left(\frac{1}{2} - x(1 - x) \right) dx \\ &= \frac{1}{2} - \left[\frac{1}{2} - \frac{1}{3} \right] \\ &= \frac{1}{3}.\end{aligned}$$

Remarque : On aurait également pu faire l'argument que si $X_{(1)}$ et $X_{(2)}$ est le vecteur de statistiques d'ordre de (X, Y) , alors $|X - Y| = X_{(2)} - X_{(1)}$. Or, comme vu à l'exemple 6.23, les $X_{(i)}$ suivent des lois Beta $(i, n + 1 - i)$. Par conséquent,

$$\mathbb{E}[|X - Y|] = \mathbb{E}[X_{(2)} - X_{(1)}] = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

Évidemment, on peut calculer les espérances de variables directement en utilisant cette méthode ; par exemple, si $\mathbf{X} = (X, Y)$ est de densité jointe f ,

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x, y) dy dx \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} xf_X(x) dx.\end{aligned}$$

On a une proposition directement analogue pour les variables discrètes :

PROPOSITION 7.2. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur de variables aléatoires discrètes avec fonction de masse jointe p et support $V \subseteq \mathbb{R}^n$ au plus dénombrable. Soit $g : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction à valeurs réelles. On note $Y = g(\mathbf{X})$. On a que*

$$(7.1.2) \quad \mathbb{E}[Y] = \mathbb{E}[g(\mathbf{X})] = \sum_{\mathbf{v} \in V} g(\mathbf{v})p(\mathbf{v}).$$

DÉMONSTRATION. La preuve est directement analogue à celle pour les variables continues, et nous ne nous y attarderons pas. \square

7.1.1. Propriétés de l'espérance. Nous allons maintenant démontrer quelques propriétés de l'espérance qui seront très utiles pour la suite des choses. À noter que nous ferons les preuves seulement dans le cas des variables aléatoires continues, malgré que ces propriétés soient des propriétés générales pour toutes les variables aléatoires.¹

La linéarité. La première propriété importante a déjà été mentionnée, et nous l'avons utilisé quelques fois déjà. Il s'agit de la **linéarité** de l'espérance (proposition 4.7) : si X et Y sont deux variables aléatoires, et a est une constante réelle,

$$(4.7.1) \quad \mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y].$$

PREUVE DE LA PROPOSITION 4.7. Si X, Y ont densité jointe f , alors on a bien sûr que

$$\begin{aligned}\mathbb{E}[aX + Y] &= \iint_{\mathbb{R}^2} (ax + y)f(x, y) dy dx \\ &= a \iint_{\mathbb{R}^2} xf(x, y) dy dx + \iint_{\mathbb{R}^2} yf(x, y) dy dx \\ &= a\mathbb{E}[X] + \mathbb{E}[Y].\end{aligned}$$

\square

On peut immédiatement déduire le corollaire suivant :

COROLLAIRE 7.1. *Soient X_1, \dots, X_n des variables aléatoires, et soient a_1, \dots, a_n des constantes réelles.*

Alors,

$$(7.1.3) \quad \mathbb{E} \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i \mathbb{E}[X_i].$$

1. Pour faire les preuves des propositions plus générales, il faut connaître la **vraie définition de l'espérance**.

DÉMONSTRATION. La preuve est faite par induction. Elle est l'objet de l'exercice 7.18. \square

En principe, on ne peut pas généraliser directement le corollaire 7.1 à des séries infinies. Toutefois, dans le cas qui nous préoccupe, on a la proposition suivante :

PROPOSITION 7.3. *Soient $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires vérifiant au moins l'une des conditions suivantes :*

- $X_i > 0$ presque sûrement ;
- $\sum_{i=1}^{\infty} \mathbb{E}[|X_i|] < +\infty$.

Alors, on a que

$$(7.1.4) \quad \mathbb{E} \left[\sum_{i=1}^{\infty} X_i \right] = \sum_{i=1}^{\infty} \mathbb{E}[X_i],$$

où on considère les possibilités que la série des espérance « converge » vers $\pm\infty$.

DÉMONSTRATION. La preuve de cette proposition dépasse le cadre du cours. Voici au moins une idée de l'endroit où l'on trouverait les difficultés. Par exemple, dans le cas de variables aléatoires continues, on aurait que :

$$\mathbb{E} \left[\lim_{n \rightarrow \infty} \sum_{i=1}^n X_i \right] = \iiint \cdots \int_{\mathbb{R}^n} \left(\lim_{n \rightarrow \infty} \sum_{i=1}^n x_i \right) f(\mathbf{x}) d^n \mathbf{x}$$

Le problème, c'est la convergence de l'intégrale. On voudrait pouvoir intervertir la limite et l'intégrale. Or, il s'avère que, lorsque nos variables aléatoires X_i respectent au moins l'une des conditions dans l'hypothèse, on a bel et bien le droit d'intervertir la limite et l'intégrale. On trouverait

$$\iiint \cdots \int_{\mathbb{R}^n} \left(\lim_{n \rightarrow \infty} \sum_{i=1}^n x_i \right) f(\mathbf{x}) d^n \mathbf{x} = \lim_{n \rightarrow \infty} \iiint \cdots \int_{\mathbb{R}^n} \left(\sum_{i=1}^n x_i \right) f(\mathbf{x}) d^n \mathbf{x}.$$

On a donc que

$$\mathbb{E} \left[\lim_{n \rightarrow \infty} \sum_{i=1}^n X_i \right] = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}[X_i],$$

et on peut conclure. \square

La monotonie. On va maintenant voir une propriété particulièrement importante de l'espérance : **la monotonie**.

Si X, Y sont deux variables aléatoires, on définit l'abus de notation $X \geq Y$ si et seulement si $X(\omega) \geq Y(\omega)$ pour tout $\omega \in \Omega$. Autrement dit, on détermine que la variable aléatoire X est plus grande que Y si il est *toujours* vrai que $X \geq Y$.

Par exemple, si X, Y sont les résultats de deux lancers de dés, $X + Y \geq X$.

On a la proposition suivante :

PROPOSITION 7.4. *Soient X, Y deux variables aléatoires, avec $X \geq Y$.² Alors, on a que*

$$(7.1.5) \quad \mathbb{E}[X] \geq \mathbb{E}[Y].$$

2. Ça fonctionne aussi avec l'hypothèse légèrement plus faible que $\mathbb{P}\{X > Y\} = 1$.

DÉMONSTRATION (DANS LE CAS DE VARIABLES CONTINUES). Si X, Y ont densité jointe f , alors puisque $X \geq Y$, on doit avoir que le support de (X, Y) est contenu dans $\{(x, y) \in \mathbb{R}^2 : x \geq y\}$.

On a donc que, par monotonicité de l'intégrale,

$$\begin{aligned}\mathbb{E}[X] &= \iint_{\{(x,y) \in \mathbb{R}^2 : x \geq y\}} xf(x, y) dy dx \\ &\geq \iint_{\{(x,y) \in \mathbb{R}^2 : x \geq y\}} yf(x, y) dy dx \\ &= \mathbb{E}[Y].\end{aligned}$$

□

On peut considérer les constantes réelles $a, b \in \mathbb{R}$ comme des variables aléatoires qui sont trivialement toujours égales à la même valeur – des fonctions constantes de Ω dans \mathbb{R} . Ça nous donne donc le corollaire suivant :

COROLLAIRE 7.2. *Soit X une variable aléatoire supportée sur l'intervalle $[a, b]$ – c'est à dire que X ne prend que des valeurs contenues entre a et b .³ Alors,*

$$(7.1.6) \quad \mathbb{E}[X] \in [a, b].$$

DÉMONSTRATION. Puisque $a \leq X \leq b$, alors $\mathbb{E}[a] \leq \mathbb{E}[X] \leq \mathbb{E}[b]$, et le résultat suit. □

L'espérance et l'indépendance. La proposition 7.1 peut également être employée pour démontrer le résultat suivant :

PROPOSITION 7.5. *Soit $\mathbf{X} = (X, Y)$ un vecteur aléatoire et soient $g, h : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions. Alors, si X et Y sont indépendantes, on a que*

$$(7.1.7) \quad \mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)].$$

DÉMONSTRATION. On a que, puisque X et Y sont indépendantes,

$$\begin{aligned}\mathbb{E}[g(X)h(Y)] &= \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dy dx \\ &= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx \right) \left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy \right) \\ &= \mathbb{E}[g(X)] \mathbb{E}[h(Y)].\end{aligned}$$

□

REMARQUE. *Attention !* : l'inverse n'est pas vrai en général ; on ne peut pas conclure que X, Y sont indépendantes si on a deux fonctions g, h telles que $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)]$. Pour pouvoir conclure, il faudrait que ce soit vrai pour toutes fonctions f et h .

3. Encore une fois, ça fonctionne aussi avec l'hypothèse légèrement plus faible que $\mathbb{P}\{X \in [a, b]\} = 1$.

L'inégalité de Jensen. Finalement, on va parfois avoir recours à une inégalité très pratique, en lien avec les fonctions convexes.⁴

PROPOSITION 7.6 (Inégalité de Jensen). *Soit X une variable aléatoire et soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction convexe. Alors on a que*

$$(7.1.8) \quad g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)].$$

REMARQUE. Si, comme moi, vous avez du mal à vous souvenir du sens de l'inégalité, remarquez que, en particulier, on doit avoir $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$, puisque la variance d'une variable aléatoire doit toujours être positive. Et la fonction $g(x) = x^2$ est convexe.

DÉMONSTRATION. Cette fois-ci, on va faire la preuve pour une variable aléatoire discrète plutôt que continue. Dans le cas d'une variable aléatoire discrète, on a que

$$\begin{aligned} g(\mathbb{E}[X]) &= g\left(\sum_{v \in V} vp_X(v)\right) \\ &\leq \sum_{v \in V} px(v)g(v) \\ &= \mathbb{E}[g(X)]. \end{aligned}$$

L'inégalité est en vertu de la définition des fonctions convexes – vous pouvez consulter la proposition B.32.

La preuve pour les variables aléatoires continues découle d'une limite un peu comme celle qu'on faisait (de façon très peu rigoureuse) au début du chapitre 5 (section 5.2) pour justifier la formule de l'espérance. \square

En particulier, on a une sorte « d'inégalité du triangle » pour les espérances :

$$(7.1.9) \quad |\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

7.1.2. Covariances et corrélations. Lorsqu'on travaillait avec une seule variable aléatoire, on avait introduit la notion de **moment** (définition 4.5). Pour rappel, on avait que le k ième moment était simplement $\mathbb{E}[X^k]$.

De façon analogue, on fait la définition suivante :

DÉFINITION 7.1 (Moments mixtes). Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur de variable aléatoires, et soient $k_1, \dots, k_n \geq 0$. Alors le (k_1, \dots, k_n) -moment mixte est donné par

$$(7.1.10) \quad \mathbb{E}\left[X_1^{k_1} X_2^{k_2} X_3^{k_3} \cdots X_n^{k_n}\right].$$

On dit que c'est un moment **d'ordre** $\sum_{i=1}^n k_i$.

Par exemple, les moments d'ordre 2 du vecteur $\mathbf{X} = (X, Y)$ sont $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2]$ et $\mathbb{E}[XY]$.

Les moments mixtes centrés sont définis de façon tout à fait analogue : si $\mathbf{X} = (X_1, \dots, X_n)$, et $\mu_i = \mathbb{E}[X_i]$ pour tout i , on a que le (k_1, \dots, k_n) -moment mixte centré est

$$(7.1.11) \quad \mathbb{E}\left[(X_1 - \mu_1)^{k_1} (X_2 - \mu_2)^{k_2} \cdots (X_n - \mu_n)^{k_n}\right].$$

4. Consulter ?? au besoin

On remarque donc que, pour un vecteur aléatoire $\mathbf{X} = (X, Y)$ composé de deux variables aléatoires, en plus des variances de X et Y , il existe un troisième moment centré d'ordre 2. Il a un nom spécial :

DÉFINITION 7.2 (Covariance). Soit $\mathbf{X} = (X, Y)$ un vecteur aléatoire composé de deux variables aléatoires. Alors, on appelle **covariance** entre X et Y , et on note $\text{Cov}[X, Y]$ le moment mixte centré d'ordre 2 de X et Y , et on note

$$(7.1.12) \quad \text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

De manière très similaire à ce qu'on avait fait au chapitre 4 pour la variance d'une variable aléatoire, on montre maintenant qu'il existe en fait un calcul plus simple pour déterminer la covariance entre deux variables aléatoires :

PROPOSITION 7.7. *Soient $\mathbf{X} = (X, Y)$ un vecteur composé de deux variables aléatoires. Alors, la covariance entre X et Y est donnée par*

$$(7.1.13) \quad \text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

DÉMONSTRATION. On a que

$$\begin{aligned} \text{Cov}[X, Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

□

La proposition suivante donne les propriétés de la covariance :

PROPOSITION 7.8. *La covariance a les propriétés suivantes :*

- i. *La covariance est symétrique : $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ pour toutes variables aléatoires, X, Y .*
- ii. *La covariance est bilinéaire : $\text{Cov}[aX + Y, Z] = a\text{Cov}[X, Z] + \text{Cov}[Y, Z]$ (et donc $\text{Cov}[X, aY + Z] = a\text{Cov}[X, Y] + \text{Cov}[X, Z]$), pour toutes variables aléatoires X, Y, Z et constante réelle a .*
- iii. *On a que $\text{Var}[X] = \text{Cov}[X, X] \geq 0$ pour toute variable aléatoire X .*
- iv. *Si X et Y sont des variables aléatoires indépendantes, alors $\text{Cov}[X, Y] = 0$.*

DÉMONSTRATION. i. Cette propriété est triviale par la commutativité du produit.

ii. On a que

$$\begin{aligned} \text{Cov}[aX + Y, Z] &= \mathbb{E}[(aX + Y)Z] - \mathbb{E}[aX + Y]\mathbb{E}[Z] \\ &= a\mathbb{E}[XZ] + \mathbb{E}[YZ] - a\mathbb{E}[X]\mathbb{E}[Z] - \mathbb{E}[Y]\mathbb{E}[Z] \\ &= a(\mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z]) + (\mathbb{E}[YZ] - \mathbb{E}[Y]\mathbb{E}[Z]) \\ &= a\text{Cov}[X, Z] + \text{Cov}[Y, Z]. \end{aligned}$$

On prouve la linéarité sur la seconde entrée en utilisant la symétrie de la covariance.

- iii. Clairement, par la définition 7.2, cette propriété est triviale.

iv. Par la proposition 7.5, on a clairement que, si X et Y sont indépendantes,

$$\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0.$$

□

REMARQUE. *Attention!* La propriété iv n'est vraie que dans ce sens ! En général, il est faux de conclure que X, Y sont indépendantes si on a seulement que $\text{Cov}[X, Y] = 0$. Par exemple, soient X une variable aléatoire de loi normale centrée réduite, et $Y = X^2$. Alors,

$$\text{Cov}[X, Y] = \mathbb{E}X^3 - \mathbb{E}[X]\mathbb{E}[X^2] = 0.$$

Pourtant, X et Y sont loin d'être indépendantes, puisque

$$\mathbb{P}\{-1 \leq X \leq 1, Y > 1\} = 0 < \mathbb{P}\{-1 \leq X \leq 1\}\mathbb{P}\{Y > 1\}.$$

On a le corollaire suivant qui découle de la bi-linéarité :

COROLLAIRE 7.3. Soient $X_1, \dots, X_n, Y_1, \dots, Y_m$ des variables aléatoires, et soient $a_1, \dots, a_n, b_1, \dots, b_m$ des constantes réelles.

Alors,

$$(7.1.14) \quad \text{Cov} \left[\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}[X_i, Y_j].$$

DÉMONSTRATION. La preuve de ce corollaire se fait par induction sur $m + n$, et est l'objet de l'exercice 7.19. □

Dans l'exercice 7.20, on utilise ce corollaire pour montrer que si X_1, \dots, X_n sont des variables aléatoires quelconques, alors

$$(7.1.15) \quad \text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i] + 2 \sum_{j=2}^n \sum_{i=1}^{j-1} \text{Cov}[X_i, X_j].$$

En particulier, par la propriété iv de la proposition 7.8, on voit que si X_1, \dots, X_n sont indépendantes, alors

$$(7.1.16) \quad \text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i].$$

REMARQUE. Certain.e.s d'entre vous auront peut-être remarqué des parallèles forts entre la notion de covariance entre variables aléatoires, et celle d'un produit scalaire entre des vecteurs. Ce n'est pas par hasard.

En fait, si on se limite seulement aux variables aléatoires d'espérance nulle, la covariance $\mathbb{E}[XY]$ est un produit scalaire sur les variables aléatoires. L'équation (7.1.15) est simplement la « loi des cosinus » ; l'équation (7.1.16) est l'identité de Parseval – ou de Pythagore, si vous voulez !. On a la norme $\|X\| = \sqrt{\text{Var}[X]}$.

Celles et ceux qui le veulent peuvent s'amuser à repérer les autres parallèles à mesure qu'on développe la théorie...

EXEMPLE 7.2. On considère un échantillon X_1, \dots, X_n de données indépendantes identiquement distribuées, avec $\mathbb{E}[X_i] = \mu$ et $\text{Var}[X_i] = \sigma^2$. On va définir des **estimateurs**; ce sont des fonctions de l'échantillon de données, et qui nous donne des estimations pour différents paramètres de la distribution marginale « réelle » des données. En principe, on voudrait que ces estimateurs convergent vers les paramètres corrects à mesure que l'échantillon est de plus en plus grand.

Par exemple, on définit

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

la moyenne arithmétique des X_i . \bar{X}_n sert d'estimateur pour l'espérance des X_i lorsque n est grand. En effet, on a que

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mu,$$

et on a que

$$\text{Var}[\bar{X}_n] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{\sigma^2}{n}.$$

On définit maintenant $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

On a que

$$\mathbb{E}[S_n^2] = \sum_{i=1}^n \mathbb{E}[(X_i - \bar{X}_n)^2].$$

On voit aisément que, puisque les X_i sont indépendants et identiquement distribués, alors les variables $(X_i - \bar{X}_n)$ sont interchangeables, et on a que, pour tout i ,

$$\mathbb{E}[(X_i - \bar{X}_n)^2] = \mathbb{E}[(X_n - \bar{X}_n)^2].$$

On trouve donc que

$$\mathbb{E}[S_n^2] = n\mathbb{E}[(X_n - \bar{X}_n)^2].$$

On se penche maintenant sur $\mathbb{E}[(X_n - \bar{X}_n)^2]$.

On a

$$\begin{aligned} \mathbb{E}[(X_n - \bar{X}_n)^2] &= \mathbb{E}\left[\left(\frac{nX_n - \sum_{i=1}^n X_i}{n}\right)^2\right] \\ &= \mathbb{E}\left[\left(\frac{n-1}{n}(X_n - \bar{X}_{n-1})\right)^2\right] \\ &= \left(\frac{n-1}{n}\right)^2 \mathbb{E}[(X_n - \bar{X}_{n-1})^2]. \end{aligned}$$

Bien sûr, puisque $\mu = \mathbb{E}[X_n] = \mathbb{E}[\bar{X}_{n-1}]$, et $\mathbb{E}[X_n - \bar{X}_{n-1}] = 0$, alors :

$$\mathbb{E}[(X_n - \bar{X}_{n-1})^2] = \text{Var}[X_n - \bar{X}_{n-1}].$$

Mais puisque X_n est indépendant de tous les X_i , alors X_n est indépendant de \bar{X}_{n-1} , et

$$\text{Var} [X_n - \bar{X}_{n-1}] = \text{Var} [X_n] + \text{Var} [\bar{X}_{n-1}] = \sigma^2 + \frac{\sigma^2}{n-1} = \sigma^2 \cdot \frac{n}{n-1}.$$

Finalement,

$$\mathbb{E} [(X_n - \bar{X}_n)^2] = \left(\frac{n-1}{n} \right)^2 \cdot \sigma^2 \cdot \frac{n}{n-1} = \frac{n-1}{n} \cdot \sigma^2,$$

et

$$\mathbb{E} [S_n^2] = (n-1)\sigma^2.$$

Il suit que l'estimateur adéquat pour la variance est, peut être surprenamment $S_n^2/(n-1)$. On détermine la variance de $S_n^2/(n-1)$:

$$\begin{aligned} \text{Var} \left[\frac{S_n^2}{n-1} \right] &= \frac{1}{(n-1)^2} \text{Var} \left[\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]. \\ &= \frac{1}{(n-1)^2} \left(\sum_{i=1}^n \text{Var} [(X_i - \bar{X}_n)^2] - 2 \sum_{j=2}^n \sum_{i=1}^{j-1} \text{Cov} [(X_i - \bar{X}_n)^2, (X_j - \bar{X}_n)^2] \right). \end{aligned}$$

L'exercice 7.22 (difficile) consiste à montrer que la variance de $S_n^2/(n-1)$ converge vers 0 lorsque n tend vers l'infini. En faisant cela, on montre que $S_n^2/(n-1)$ est un estimateur de la variance des X_i .

On peut déterminer que la covariance entre $X_i - \bar{X}_n$ et \bar{X}_n est

$$\begin{aligned} \text{Cov} [X_i - \bar{X}_n, \bar{X}_n] &= \mathbb{E} [(X_i - \bar{X}_n)\bar{X}_n] - \mathbb{E} [X_i - \bar{X}_n] \mathbb{E} [\bar{X}_n] \\ &= \mathbb{E} [X_i \bar{X}_n] - \mathbb{E} [\bar{X}_n^2] \\ &= \mathbb{E} [X_i \bar{X}_n] - \mu^2 + \mu^2 - \mathbb{E} [\bar{X}_n^2] \\ &= \mathbb{E} [X_i \bar{X}_n] - \mu^2 - \text{Var} [\bar{X}_n] \\ &= \frac{1}{n} \sum_{j=1}^n \mathbb{E} [X_i X_j] - \mu^2 - \frac{\sigma^2}{n} \\ &= \frac{1}{n} (\text{Var} [X_i] + \mu^2) + \frac{n-1}{n} \mu^2 - \mu^2 - \frac{\sigma^2}{n} \\ &= \frac{1}{n} \text{Var} [X_i] - \frac{\sigma^2}{n} \\ &= 0. \end{aligned}$$

EXEMPLE 7.3. Supposons qu'on réalise une étude statistique et que l'on étudie le lien entre deux variables X et Y . On suspecte qu'il devrait y avoir une relation de type linéaire, du genre $Y = aX + b + N$, où a, b sont des constantes et N serait une variable aléatoire indépendante de X ; N suivrait une distribution normale centrée de variance σ^2 .

Alors, la covariance de X, Y est

$$\begin{aligned} \text{Cov} [X, Y] &= \text{Cov} [X, aX + b + N] \\ &= a\text{Var} [X]. \end{aligned}$$

Donc, la pente de notre modèle linéaire devrait être $a = \frac{\text{Cov}[X,Y]}{\text{Var}[X]}$.

7.1.2.1. *Les corrélations.* On introduit la corrélation entre deux variables aléatoires :

DÉFINITION 7.3. Soient X, Y deux variables aléatoires quelconques de variances positives. On définit la **corrélation** entre les variables X et Y par

$$(7.1.17) \quad \rho[X, Y] := \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \text{Var}[Y]}}.$$

EXEMPLE 7.4. En reprenant l'exemple précédent, on a que

$$\begin{aligned} \text{Var}[Y] &= \text{Cov}[aX + b + N, aX + b + N] \\ &= a^2\text{Var}[X] + \sigma^2. \end{aligned}$$

En particulier,

$$\begin{aligned} |\rho[X, Y]|^2 &= \left| \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X] \text{Var}[Y]}} \right|^2 \\ &= \frac{a^2\text{Var}[X]^2}{\text{Var}[X] \cdot (a^2\text{Var}[X] + \sigma^2)} \\ &= \frac{a^2\text{Var}[X]}{a^2\text{Var}[X] + \sigma^2}. \end{aligned}$$

En d'autres mots, c'est la fraction de la variance de Y qui peut être « attribuée » aux variations de X – plus σ^2 est grand, et plus cette fraction est petite. Remarquez que N n'a pas à suivre une loi normale nécessairement pour que ces calculs fonctionnent ; tant que $\text{Var}[N] = \sigma^2$, on obtiendra le même résultat.

REMARQUE. Les statisticien.ne.s parmi vous verront peut-être poindre des ressemblances avec le coefficient de détermination r^2 en statistiques...

On montre la proposition suivante :

PROPOSITION 7.9. *Soient X, Y deux variables aléatoires.*

Alors on a toujours que

$$(7.1.18) \quad |\rho[X, Y]| \leq 1.$$

DÉMONSTRATION. Pour tout λ , on a toujours que

$$\begin{aligned} 0 &\leq \text{Var}[X - \lambda Y] \\ &= \text{Cov}[X - \lambda Y, X - \lambda Y] \\ &= \text{Var}[X] - 2\lambda\text{Cov}[X, Y] + \lambda^2\text{Var}[Y]. \end{aligned}$$

Et si on choisit $\lambda = \text{Cov}[X, Y]/\text{Var}[Y]$, on trouve que

$$\begin{aligned} 0 &\leq \text{Var}[X] - 2\frac{\text{Cov}[X, Y]^2}{\text{Var}[Y]} + \frac{\text{Cov}[X, Y]^2}{\text{Var}[Y]} \\ &= \text{Var}[X] - \frac{\text{Cov}[X, Y]^2}{\text{Var}[Y]}. \end{aligned}$$

En réarrangeant, on trouve

$$\text{Var}[X] \text{Var}[Y] \geq \text{Cov}[X, Y]^2,$$

ou

$$\rho[X, Y]^2 \leq 1.$$

On conclue en prenant la racine carrée de part et d'autre. \square

7.2. Le nombre d'événements et les indicatrices

On a beaucoup développé les propriétés de linéarité de l'espérance, et ça nous facilite beaucoup la vie dans les calculs.

On va maintenant explorer une méthode de calcul des espérances qui sera très pratique en général.

Supposons que nous avons $(E_i)_{i \in I}$ une famille d'événements indexée par un ensemble d'indices I au plus dénombrable ; I peut être fini, ou ça peut être \mathbb{N} , etc.

Supposons maintenant que X soit « le nombre d'événements de notre famille $(E_i)_{i \in I}$ qui se sont réalisés. »

Par exemple, on pourrait avoir les situations suivantes :

- On lance n dés, l'événement E_i est « le i ème dé tombe sur 6 ». X est « le nombre de dés qui sont tombés sur 6. »
- On tire n boules d'un sac, sans remise. E_i est « la i ème boule pigée est mauve ». X est « le nombre de boules mauves pigées. »
- etc.

L'astuce, c'est qu'on va *représenter X comme une somme de variables indicatrices*. Cette technique est très utile, dans toutes sortes de contextes, et elle permet de faciliter le calcul des espérances. En effet, on a

$$X = \sum_{i \in I} \mathbb{1}_{E_i}.$$

Il suit immédiatement que

$$(7.2.1) \quad \mathbb{E}[X] = \sum_{i \in I} \mathbb{E}[\mathbb{1}_{E_i}] = \sum_{i \in I} \mathbb{P}\{E_i\}.$$

EXEMPLE 7.5. Si on jette n dés équilibrés à 6 faces, et que X est le nombre de dés qui donnent un 6, trouver $\mathbb{E}[X]$.

SOLUTION. On pourrait répondre en disant que X suit une distribution binomiale de paramètres $n, p = 1/6$ et donc que son espérance est $\mathbb{E}[X] = n/6$.

Avec notre méthode, les événements E_i sont « le i ème dé tombe sur 6 », ils ont tous probabilité $1/6$ et donc

$$\mathbb{E}[X] = \sum_{i=1}^6 \mathbb{P}\{E_i\} = \frac{n}{6}.$$

EXEMPLE 7.6. Si on tire sans remise 8 boules d'une urne qui contient 11 boules mauves et 17 boules jaunes, quelle est l'espérance de X , le nombre de boules mauves tirées ?

SOLUTION. Encore une fois, on pourrait dire que X est une hyper-géométrique de paramètres bla bla bla... Mais on peut aussi tout simplement écrire E_i l'événement « la i ème boule tirée est mauve ». Évidemment, les événements ne sont pas indépendants entre eux. Pas grave ; ils ont tous la même probabilité de se produire : 11/28.

Finalement, on a donc

$$\mathbb{E}[X] = \sum_{i=1}^8 \mathbb{P}\{E_i\} = \frac{88}{28}.$$

EXEMPLE 7.7. N personnes entrent à une fête et jettent leur manteau sur le lit. Puis, tout le monde repart en prenant aléatoirement un manteau dans la pile.

Soit X le nombre de personnes qui repart avec le bon manteau. Trouver l'espérance de X .

SOLUTION. Ici, on n'a pas connaissance de la distribution de X . On n'a jamais calculé son espérance avant. Mais notre nouvelle méthode va nous permettre de procéder comme si de rien n'était.

On définit d'abord l'événement E_i , « la personne numéro i repart avec son manteau ». D'abord, $\mathbb{P}\{E_i\} = \mathbb{P}\{E_1\}$ pour tout i , vu que le problème ne dépend pas de l'identité précise de chaque personne. Ça serait absurde autrement.

Mais $\mathbb{P}\{E_1\} = \frac{1}{N}$, puisque la personne numéro 1 a 1 chance sur N de choisir son manteau.

Or, on a $X = \sum_{i=1}^N \mathbb{1}_{E_i}$. Il suit que

Donc,

$$\mathbb{E}[X] = \sum_{i=1}^N \mathbb{P}\{E_i\} = 1.$$

On peut calculer la variance de X : on calcule d'abord le second moment.

$$\begin{aligned} \mathbb{E}[X^2] &= \mathbb{E}\left[\left(\sum_{i=1}^N \mathbb{1}_{E_i}\right)^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^N \sum_{j=1}^N \mathbb{1}_{E_i} \mathbb{1}_{E_j}\right] \\ &= \sum_{i=1}^N \mathbb{P}\{E_i\} + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{P}\{E_i \cap E_j\}. \end{aligned}$$

Maintenant on sait que pour tout $i \neq j$, on a que $\mathbb{P}\{E_i \cap E_j\} = \frac{1}{N(N-1)}$ – parmi les $N(N-1)$ choix de manteaux possibles ordonnées pour la personne i et la personne j , il n'y a qu'un choix où i et j ont choisi le bon manteau respectivement.

Évidemment, la somme $\sum_{i=2}^N \sum_{j=1}^{i-1} \mathbb{P}\{E_i \cap E_j\}$ compte exactement $\frac{N(N-1)}{2}$ termes. Finalement, on a donc que

$$\mathbb{E}[X^2] = 1 + 2 \frac{N(N-1)}{2} \frac{1}{N(N-1)} = 2.$$

Et $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1$.

EXEMPLE 7.8. Un groupe de N chasseurs sont cachés dans un buis, prêts à tirer sur la prochaine volée de canards.

Il passe un vol de K canards. Chaque chasseur choisit sa cible indépendamment l'un de l'autre, et atteint sa cible avec probabilité p .

Soit X le nombre de canards abattus. Trouver $\mathbb{E}[X]$.

SOLUTION. On va définir les événements E_i , « le i ème canard a été abattu ».

Le nombre de canards abattus est donc $X = \sum_{i=1}^K \mathbb{1}_{E_i}$.

Chaque canard doit avoir la même probabilité d'être abattu, puisque le problème est entièrement symétrique sous permutation des canards. Donc,

$$\mathbb{P}\{E_i\} = \mathbb{P}\{E_1\},$$

et

$$\mathbb{E}[X] = K\mathbb{P}\{E_1\}.$$

Il suffit maintenant de trouver la probabilité que le premier canard ait été abattu. On va noter Y le nombre de chasseurs qui ont choisi de viser le premier canard. Alors, par la formule des probabilités totales,

$$PR\{E_1\} = \sum_{i=0}^N \mathbb{P}\{E_1 \mid Y = i\} \mathbb{P}\{Y = i\}.$$

Y est une variable aléatoire binomiale de paramètre $N, \frac{1}{K}$.

Sachant que $Y = i$, c'est à dire que i chasseurs visaient le premier canard, on a que $\mathbb{P}\{E_1 \mid Y = i\} = 1 - (1-p)^i$, soit 1 moins la probabilité qu'ils ratent tous, $(1-p)^i$.

Finalement, on trouve que

$$\mathbb{P}\{E_1\} = \sum_{i=0}^N (1 - (1-p)^i) \left(\binom{N}{i} \left(\frac{1}{K}\right)^i \left(\frac{K-1}{K}\right)^{N-i} \right).$$

En séparant la somme et en simplifiant, on trouve que

$$\mathbb{P}\{E_1\} = 1 - \sum_{i=0}^N \binom{N}{i} \left(\frac{1-p}{K}\right)^i \left(\frac{K-1}{K}\right)^{N-i} = 1 - \left(\frac{K-p}{K}\right)^N.$$

On prend un moment pour vérifier que ça a du sens. Si il y a un seul chasseur, la probabilité que le premier canard soit abattu est la probabilité qu'il choisisse le premier canard, $(1/K)$ fois la probabilité qu'il touche le premier canard (p). Donc c'est $p/K = 1 - (K-p)/K$. Check.

Si il y a un seul canard, et que la probabilité de toucher le canard est p , alors la probabilité que le canard soit touché est 1 moins la probabilité que tout le monde rate : $(1-p)^N$. Ça marche encore.

Si on met la probabilité de toucher le canard à 1, la proba que le premier canard soit abattu est la proba que au moins un chasseur a choisi le premier canard. C'est donc 1 moins la proba que tout le monde a choisi un autre canard, soit $1 - (K-1/K)^N$. Check.

Dans la limite où on fait tendre le nombre de canards à l'infini, la proba que le premier canard est touché tend vers 0. Check. Dans la limite où on fait plutôt tendre le nombre de chasseurs vers l'infini, la probabilité que le premier canard soit touché tend vers 1. Check.

Finalement, le nombre de canards abattus X a l'espérance

$$\mathbb{E}[X] = K \left(1 - \left(\frac{K-p}{K} \right)^N \right).$$

Par exemple, avec $N = 10$ chasseurs, $K = 10$ canards et une probabilité de succès de $p = 1/2$, l'espérance du nombre de canards abattus est $\approx 4.01263\dots$

De façon générale, avec $K = N$, on aurait pu s'attendre à ce que $\mathbb{E}[X] = Np$ – toutefois, il faut tenir compte du fait que plusieurs chasseurs vont choisir le même canard. Donc, on trouve que si $N = K$, on a que

$$\mathbb{E}[X] = N \left(1 - \left(1 - \frac{p}{N} \right)^N \right) \sim N(1 - e^{-p})$$

lorsque N est grand.

7.3. Espérances conditionnelles

Nous arrivons maintenant à une notion d'une importance capitale en théorie des probabilités : la notion d'espérance conditionnelle.

À la section 6.4, nous avons abordé la notion de distribution conditionnelle d'une variable aléatoire. L'espérance conditionnelle est un outil qui s'inscrit dans la même lignée.

REMARQUE. Comme pour la définition de l'espérance, la « vraie » définition de l'espérance conditionnelle requiert des notions d'analyse qui dépassent très largement le cadre de ce cours.

Pour cette raison, nous allons nous limiter aux cas spécifiques des variables aléatoires discrètes et continues, que nous traiterons séparément.

7.3.1. L'espérance conditionnelle comme une fonction d'une valeur réelle. Supposons qu'on ait deux variables aléatoires X et Y . Une question naturelle à poser serait : « Quelle est l'espérance de X , si on sait que Y prend la valeur y ? »

Avec les notations auxquelles nous sommes habitué.e.s, on pourrait écrire simplement :

$$(7.3.1) \quad \mathbb{E}[X \mid Y = y].$$

Bien sûr, l'espérance $\mathbb{E}[X]$ tout-court est une valeur réelle constante. Il devrait suivre qu'on peut noter

$$h_{X|Y}(y) = \mathbb{E}[X \mid Y = y];$$

en effet, l'espérance conditionnelle de X sachant que $Y = y$ peut dépendre de la valeur de y .

Dans le cas où X est une variable aléatoire discrète, on aura $p_{X|Y}(x \mid y)$ la fonction de masse conditionnelle de X sachant Y . Celle-ci est définie pour Y continue ou discrète, donc pas de soucis.

Alors, il est naturel d'avoir.

$$(7.3.2) \quad h_{X|Y}(y) = \mathbb{E}[X \mid Y = y] = \sum_{v \in V} v p_{X|Y}(v \mid y).$$

EXEMPLE 7.9. Supposons que X est une variable aléatoire de loi Poisson avec paramètre Θ , où Θ suit une loi uniforme sur l'intervalle $[1, 2]$.

Trouver l'espérance conditionnelle de X sachant que $\Theta = \theta$.

SOLUTION. On a bien sûr que $p_{X|\Theta}(k | \theta) = e^{-\theta}\theta^k/k!$ pour tout $k > 0$.
Donc, $\mathbb{E}[X | \Theta = \theta] = \theta$.

EXEMPLE 7.10. Supposons que X et Y soient deux variables aléatoires binomiales indépendantes de paramètres respectifs (n, p) et (m, p) .

Trouver l'espérance conditionnelle de X sachant que $Z = X + Y = k$.

SOLUTION. On commence par chercher la loi conditionnelle de X sachant Z . Z suit une loi binomiale de paramètres $(m+n, p)$, et on a donc :

$$\begin{aligned} p_{X|Z}(i | k) &= \mathbb{P}\{X = i | X + Y = k\} \\ &= \frac{\mathbb{P}\{X = i, Y = k - i\}}{\mathbb{P}\{X + Y = k\}} \\ &= \frac{\binom{n}{i} p^i (1-p)^{n-i} \binom{m}{k-i} p^{k-i} (1-p)^{m-k+i}}{\binom{m+n}{k} p^k (1-p)^{m+n-k}} \\ &= \frac{\binom{n}{i} \binom{m}{k-i}}{\binom{m+n}{k}}. \end{aligned}$$

En d'autres termes, sachant que $X + Y = k$, alors la loi conditionnelle de X est Hypergéométrique de paramètres $k, n+m, n$.

On a donc que

$$h_{X|Z}(k) = \mathbb{E}[X | X + Y = k] = \frac{kn}{n+m}$$

Dans le cas où X est une variable aléatoire continue, on aura plutôt

$$(7.3.3) \quad h_{X|Y}(y) = \mathbb{E}[X | Y = y] = \int_{\mathbb{R}} x f_{X|Y}(x | y) dx.$$

Cela est bien défini pour Y continue ou discrète, peu importe.

EXEMPLE 7.11. Soient X, Y deux variables aléatoires avec densité jointe

$$f(x, y) = \frac{1}{y} e^{-x/y} e^{-y}, \quad x, y > 0.$$

Déterminer l'espérance de X sachant $Y = y$.

SOLUTION. On trouve que la densité marginale de Y est

$$f_Y(y) = e^{-y} \cdot \frac{1}{y} \int_0^\infty e^{-x/y} dx = e^{-x},$$

pour $x > 0$.

On a donc que la densité conditionnelle de X sachant Y est

$$f_{X|Y}(x | y) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{y} e^{-x/y}, \quad x > 0.$$

En d'autres mots, sachant Y , X suit une loi exponentielle de paramètre $1/Y$.

On trouve donc que

$$h_{X|Y}(y) = \mathbb{E}[X | Y = y] = \int_0^\infty x \cdot \frac{1}{y} e^{-x/y} dx = y.$$

Les espérances conditionnelles sont des espérances, et elles ont donc les mêmes propriétés. Nous admettrons donc sans la montrer la proposition suivante :

PROPOSITION 7.10. *Les propriétés suivantes sont vraies :*

- i. Pour $X_1, X_2, X_3, \dots, X_n, Y$ des variables aléatoires quelconques et a_1, \dots, a_n des constantes, on a que

$$(7.3.4) \quad \mathbb{E} \left[\sum_{i=1}^n a_i X_i \mid Y = y \right] = \sum_{i=1}^n a_i \mathbb{E}[X_i \mid Y = y];$$

ii. Si X, Y, Z sont des variables aléatoires quelconques avec $X \geq Y$, alors

$$(7.3.5) \quad \mathbb{E}[X \mid Z = z] \geq \mathbb{E}[Y \mid Z = z];$$

iii. Si X, Y sont deux variables aléatoires et $g : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction, alors

— Si X est discrète avec support V et fonction de masse conditionnelle $p_{X|Y}$,

$$(7.3.6) \quad \mathbb{E}[g(X) \mid Y = y] = \sum_{v \in V} g(v) p_{X|Y}(v \mid y);$$

— Si X est continue avec fonction de densité conditionnelle $f_{X|Y}$,

$$(7.3.7) \quad \mathbb{E}[g(X) \mid Y = y] = \int_{\mathbb{R}} g(x) f_{X|Y}(x \mid y) dx;$$

iv. Si X, Y sont des variables aléatoires et g est une fonction convexe,

$$(7.3.8) \quad g(\mathbb{E}[X \mid Y = y]) \leq \mathbb{E}[g(X) \mid Y = y].$$

La prochaine proposition donne des indications concernant le comportement de l'espérance conditionnelle avec l'indépendance.

PROPOSITION 7.11. *Soient X, Y des variables aléatoires, et $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction quelconque. On a les propriétés suivantes :*

- i. Si X et Y sont indépendantes, alors

$$(7.3.9) \quad \mathbb{E}[g(X) \mid Y = y] = \mathbb{E}[g(X)];$$

ii. On a que

$$(7.3.10) \quad \mathbb{E}[Xg(Y) \mid Y = y] = g(y) \mathbb{E}[X \mid Y = y].$$

DÉMONSTRATION. On fait la preuve pour des variables aléatoires continues, mais le protocole est en tout point similaire pour les variables discrètes.

- i. Si X et Y sont indépendantes, alors $g(X)$ et Y sont indépendantes, et $f_{X|Y}(x \mid y) = f_X(x)$.

Il suit que

$$\mathbb{E}[g(X) \mid Y = y] = \int_{\mathbb{R}} g(x) f_{X|Y}(x \mid y) dx = \int_{\mathbb{R}} g(x) f_X(x) dx = \mathbb{E}[g(X)].$$

ii. Si X et Y sont des variables aléatoires quelconques, alors⁵

$$\mathbb{E}[Xg(Y) \mid Y = y] = \iint_{\mathbb{R}} xg(y) f_{X|Y}(x \mid y) dx = g(y) \mathbb{E}[X \mid Y = y].$$

5. Ce raisonnement peut paraître insatisfaisant. Nous n'avons en effet pas très bien couvert les distributions conditionnelles pour des vecteurs de variables aléatoires. Toutefois l'intuition devrait être claire : sachant $Y = y$, on peut traiter la variable Y comme une constante dans l'expression.

□

En particulier,

$$\mathbb{E}[g(X) \mid X = x] = g(x),$$

tout simplement, et si X et Y sont indépendantes,

$$\mathbb{E}[g(X)h(Y) \mid Y = y] = \mathbb{E}[g(X)]h(y).$$

7.3.2. Les espérances conditionnelles vues comme des variables aléatoires.

La vraie puissance des espérances conditionnelles provient de ce nouvel outil : plutôt que de définir des espérances conditionnelles par rapport à l'événement $Y = y$, on va définir simplement des espérances conditionnelles par rapport à la variable aléatoire Y . Ça va ressembler à :

$$\mathbb{E}[X \mid Y].$$

Mais quel sens donner à une telle expression ? Eh bien voilà : si on avait une fonction

$$h_{X|Y}(y) = \mathbb{E}[X \mid Y = y],$$

bien définie, alors on va maintenant définir

$$(7.3.11) \quad \mathbb{E}[X \mid Y] = h_{X|Y}(Y).$$

Donc, l'espérance conditionnelle $\mathbb{E}[X \mid Y]$ est *une variable aléatoire* !

Qui plus est, c'est une variable aléatoire qui ne dépend uniquement que de Y , et qui est entièrement déterminée par Y .

EXEMPLE 7.12. Soit X une variable aléatoire suivant une loi de Poisson de paramètre Θ , où Θ est une variable aléatoire uniforme sur l'intervalle $[1, 2]$.

Déterminer $\mathbb{E}[X \mid \Theta]$.

SOLUTION. Nous avions vu dans un exemple précédent que $h_{X|\Theta}(\theta) = \mathbb{E}[X \mid \Theta = \theta] = \theta$.

Donc, l'espérance conditionnelle $\mathbb{E}[X \mid \Theta]$ est donnée par

$$\mathbb{E}[X \mid \Theta] = h_{X|\Theta}(\Theta) = \Theta.$$

EXEMPLE 7.13. Soient X, Y deux variables aléatoires dont la densité jointe est donnée par

$$f(x, y) = \frac{1}{y} e^{-y}, \quad 0 < x < y.$$

Trouver l'espérance de X sachant Y .

SOLUTION. On commence par trouver la densité marginale de Y . C'est

$$f_Y(y) = \int_0^y \frac{1}{x} e^{-x} dx = e^{-y}, \quad y > 0.$$

Donc, Y suit une loi exponentielle standard. La densité conditionnelle de X est

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{y}, \quad 0 < x < y,$$

et sachant $Y = y$, X suit une loi uniforme sur l'intervalle $(0, y)$.

On a donc que

$$h_{X|Y}(y) = \mathbb{E}[X | Y = y] = \frac{1}{y} \int_0^y x dx = \frac{y}{2}.$$

Finalement,

$$\mathbb{E}[X | Y] = h_{X|Y}(Y) = \frac{Y}{2}.$$

Encore une fois : $\mathbb{E}[X | Y]$ est une *variable aléatoire*, qui est entièrement déterminée par la valeur de Y .

Probablement la plus utile des propositions sur les espérances conditionnelles est celle-ci :

PROPOSITION 7.12. *Soient X, Y, Z trois variables aléatoires quelconques.*

Alors, on a les propriétés suivantes :

- i. $\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$.
- ii. $\mathbb{E}[\mathbb{E}[X | Y, Z] | Z] = \mathbb{E}[X | Z]$.

DÉMONSTRATION. Encore une fois, nous allons faire la preuve pour des variables aléatoires continues.

D'abord, puisque $\mathbb{E}[X | Y] = h_{X|Y}(Y)$, alors bien sûr

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \int_{\mathbb{R}} h_{X|Y}(y) f_Y(y) dy \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x f_{X|Y}(x | y) dx \right) f_Y(y) dy \\ &= \iint_{\mathbb{R}^2} x f_{X|Y}(x | y) f_Y(y) dy dx \\ &= \iint_{\mathbb{R}^2} x f(x, y) dy dx \\ &= \mathbb{E}[X]. \end{aligned}$$

La seconde partie suit puisque l'espérance conditionnelle est une espérance.⁶

□

EXEMPLE 7.14. Dans l'exemple 7.12, trouver l'espérance de X .

SOLUTION. *A priori*, ça peut sembler ardu, vu qu'on ne connaît pas la loi marginale de X , et que la loi jointe de X et Θ est difficile à travailler. Par contre, la proposition 7.12 vient à notre rescousse :

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X | \Theta]] \\ &= \mathbb{E}[\Theta] \\ &= \frac{3}{2}. \end{aligned}$$

6. Encore une fois, ce raisonnement, quoique globalement vrai, peut sembler insatisfaisant. Et encore une fois, la raison en est que pour tout faire comme il faut, il faudrait avoir des notions d'analyse plus poussées dont on ne dispose pas dans ce cours.

EXEMPLE 7.15. Dans l'exemple 7.13, trouver l'espérance de X .

SOLUTION. Encore une fois, on ne connaît pas la loi marginale de X , mais on connaît celle de Y , et on connaît la loi conditionnelle de X sachant Y . En particulier, on sait que, sachant Y , X suit une loi uniforme sur $(0, Y)$ et donc que son espérance conditionnelle est

$$\mathbb{E}[X \mid Y] = \frac{1}{2}Y.$$

Mais alors

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]] = \frac{1}{2}\mathbb{E}[Y].$$

et puisque Y est une variable aléatoire exponentielle standard, $\mathbb{E}[Y] = 1$ et

$$\mathbb{E}[X] = \frac{1}{2}.$$

EXEMPLE 7.16. On lance un dé équilibré à 6 faces. Si on tombe sur 1 ou 2, on gagne 5\$ et on relance. Si on tombe sur 3 ou 4, on gagne 10\$ et on relance. Si on tombe sur 5 ou 6, on ne gagne rien et on s'arrête.

Quelle est l'espérance du gain ?

SOLUTION. On définit X_i le gain réalisé au lancer i . *A priori*, $\mathbb{P}\{X_i = 0\} = \mathbb{P}\{X_i = 5\} = \mathbb{P}\{X_i = 10\} = \frac{1}{3}$.

On définit maintenant $N = \min\{i \in \mathbb{N} : X_i = 0\}$, le premier lancer où on n'a rien gagné (et donc où on a arrêté).

Le gain total est $G = \sum_{i=1}^{N-1} X_i$.

Sachant $N = k$, la fonction de masse conditionnelle de X_i est $p_{X_i|N}(j \mid k) = \mathbb{P}\{X_i = j \mid N = k\}$.

Pour tout $i < k$, on a donc $p_{X_i|N}(5 \mid k) = p_{X_i|N}(10 \mid k) = \frac{1}{2}$ – puisque sachant que $N = k$, on sait que $X_i \neq 0$.

Pour $i = k$, on a que $p_{X_k|N}(0 \mid k) = 1$, puisque forcément, $X_k = 0$ si $N = k$, par définition de N .

Donc, $G = \sum_{i=1}^{N-1} X_i$, et

$$\mathbb{E}[G] = \mathbb{E}[\mathbb{E}[G \mid N]].$$

Mais par linéarité, on a donc

$$\mathbb{E}[G] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{N-1} X_i \mid N\right]\right] = \mathbb{E}\left[\sum_{i=1}^{N-1} \mathbb{E}[X_i \mid N]\right].$$

Bien sûr, $\mathbb{E}[X_i \mid N] = \frac{15}{2}$, et il suit par conséquent que

$$\mathbb{E}[G] = \mathbb{E}\left[(N-1)\frac{15}{2}\right].$$

Mais N suit une loi géométrique de paramètre $1/3$. Par conséquent, $\mathbb{E}[N-1] = 2$ et

$$\mathbb{E}[G] = 15.$$

EXEMPLE 7.17. Un nombre aléatoire N de client.e.s entrent dans un magasin. Chaque client.e fait une dépense d'un montant X_i , indépendant et identiquement distribué.

Quelle est l'espérance du total des transactions au magasin ?

SOLUTION. Le total dépensé est $T = \sum_{i=1}^N X_i$, où les X_i et N sont des variables aléatoires indépendantes, et les X_i sont identiquement distribuées.

Supposons que $\mathbb{E}[X_1] = \mu$.

Alors, on a que

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}[\mathbb{E}[T | N]] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^N X_i \mid N\right]\right] \\ &= \mathbb{E}\left[\sum_{i=1}^N \mathbb{E}[X_i | N]\right].\end{aligned}$$

Mais puisque les X_i sont indépendantes de N , alors $\mathbb{E}[X_i | N] = \mathbb{E}[X_i]$ et

$$\begin{aligned}\mathbb{E}[T] &= \mathbb{E}\left[\sum_{i=1}^N \mu\right] \\ &= \mu\mathbb{E}[N].\end{aligned}$$

Et comme on pourrait s'y attendre, l'espérance du total des transactions dans le magasin sera donné par $\mu\mathbb{E}[N]$.

EXEMPLE 7.18. Trouver l'espérance et la variance d'une variable aléatoire géométrique sans gros calcul.

SOLUTION. Soit X une variable aléatoire géométrique de paramètre p . X est le nombre de tentatives indépendantes nécessaires pour obtenir un succès (probabilité p de succès).

Si Y est la première tentative, alors

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]].$$

On trouve alors que

$$\mathbb{E}[X] = \mathbb{E}[X | Y = 0]\mathbb{P}\{Y = 0\} + \mathbb{E}[X | Y = 1]\mathbb{P}\{Y = 1\}.$$

Bien sûr, $\mathbb{P}\{Y = 0\} = 1 - p$, $\mathbb{P}\{Y = 1\} = p$, et $\mathbb{E}[X | Y = 1] = 1$ – si on sait que la première tentative est un succès, alors le nombre de tentatives pour obtenir un succès est 1.

On a aussi que $\mathbb{E}[X | Y = 0] = 1 + \mathbb{E}[X]$ – en effet, si on sait que la première tentative a échoué, alors ça nous a pris une tentative ; puis il faut recommencer comme si on ne l'avait jamais faite.

Donc,

$$\mathbb{E}[X] = (1 + \mathbb{E}[X])(1 - p) + p,$$

et en réarrangeant on trouve

$$p\mathbb{E}[X] = (1 - p) + p = 1,$$

d'où $\mathbb{E}[X] = \frac{1}{p}$.

On peut trouver la variance également.

On commence par trouver $\mathbb{E}[X^2] = \mathbb{E}[\mathbb{E}[X^2 | Y]]$. Le raisonnement change peu : la loi de X sachant que $Y = 0$ est la même que la loi de $X + 1$. Donc, $\mathbb{E}[X^2 | Y = 0] = \mathbb{E}[(X + 1)^2]$.

Au final, si on écrit $\mu = \mathbb{E}[X]$, on a que

$$\mathbb{E}[X^2 \mid Y = 0] = \mathbb{E}[X^2] + 2\mu + 1.$$

On trouve donc

$$\mathbb{E}[X^2] = \mathbb{E}[X^2 \mid Y = 0](1-p) + \mathbb{E}[X^2 \mid Y = 1]p = (\mathbb{E}[X^2] + 2\mu + 1)(1-p) + p.$$

Finalement, on réarrange pour trouver

$$p\mathbb{E}[X^2] = \frac{2(1-p)}{p} + \frac{p(1-p)}{p} + \frac{p^2}{p},$$

et on simplifie pour obtenir

$$\mathbb{E}[X^2] = \frac{2-p}{p^2},$$

et finalement

$$\text{Var}[X] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

EXEMPLE 7.19. Soit U_1, U_2, U_3, \dots une suite de variables aléatoires indépendantes uniformément distribuées sur $(0, 1)$.

on note, pour $x \in (0, 1)$ la variable aléatoire

$$N(x) = \min \left\{ n \in \mathbb{N} : \sum_{i=1}^n U_i > x \right\}$$

Trouver $\mathbb{E}[N(x)]$.

SOLUTION. On va conditionner par U_1 :

$$\mathbb{E}[N(x)] = \mathbb{E}[\mathbb{E}[N(x) \mid U_1]].$$

On calcule :

$$\mathbb{E}[N(x)] = \int_0^1 \mathbb{E}[N(x) \mid U_1 = y] dy.$$

Bien sûr,

$$\mathbb{E}[N(x) \mid U_1 = y] = \begin{cases} 1 & \text{si } y > x \\ \mathbb{E}[1 + N(x-y)] & \text{si } y \leq x. \end{cases}$$

En effet, si U_1 n'excède pas x , alors ça prend au moins U_1 , plus le nombre d'uniformes nécessaires pour excéder ce qui reste, c'est à dire $N(x-y)$.

Donc, si on écrit $g(x) = \mathbb{E}[N(x)]$,

$$\begin{aligned} g(x) &= \mathbb{E}[N(x)] \\ &= \int_0^x (1 + \mathbb{E}[N(x-y)]) dy + \int_x^1 dy \\ &= \int_0^x (1 + g(x-y)) dy + (1-x) \\ &= x + \int_0^x g(x-y) dy + (1-x) \\ &= 1 + \int_0^x g(x-y) dy \\ &= 1 + \int_0^x g(u) du. \end{aligned}$$

Et en dérivant de part et d'autre par x , on trouve que

$$g'(x) = g(x),$$

avec la condition que $g(0) = 1$.

La solution à cette équation différentielle est $g(x) = e^x$.

Donc, pour $x \in (0, 1)$, on a que

$$\mathbb{E}[N(x)] = e^x.$$

7.3.3. Probabilités conditionnelles. Évidemment, on peut utiliser notre astuce d'espérance conditionnelle pour calculer non seulement des espérances, mais aussi des probabilités. En effet, pour tout événement A , et une variable aléatoire X quelconque,

$$(7.3.12) \quad \mathbb{P}\{A\} = \mathbb{E}[\mathbb{E}[\mathbb{1}_A | X]].$$

Donc, lorsque X est une variable aléatoire, on se dote du raccourci de notation

$$\mathbb{P}\{A | X\} := \mathbb{E}[\mathbb{1}_A | X].$$

REMARQUE. *Attention!* Il ne faut pas confondre avec la notation similaire pour les probabilités conditionnelles à des *événements*! Ici, X est une *variable aléatoire*!.

À noter : si on considère une partition E_1, E_2, \dots, E_n de Ω , et qu'on considère $X = \sum_{i=1}^n i \mathbb{1}_{E_i}$, alors l'équation 7.3.12 se résume à :

$$\begin{aligned} \mathbb{P}\{A\} &= \mathbb{E}[\mathbb{E}[\mathbb{1}_A | X]] \\ &= \sum_{i=1}^n \mathbb{E}[\mathbb{1}_A | X = i] \mathbb{P}\{X = i\} \\ &= \sum_{i=1}^n \mathbb{P}\{A | E_i\} \mathbb{P}\{E_i\}. \end{aligned}$$

Autrement dit, il s'agit simplement de la formule de probabilité totale du chapitre 3.

EXEMPLE 7.20. Soient X, Y deux variables aléatoires indépendantes continues avec densités marginales respectives f_X et f_Y et densité jointe f .

Trouver $\mathbb{P}\{X < Y\}$ et $\mathbb{P}\{X + Y \leq z\}$.

SOLUTION. Bien sûr, $\mathbb{P}\{X < Y \mid Y = y\} = \mathbb{P}\{X < y\}$. Donc,

$$\mathbb{P}\{X < Y\} = \int_{-\infty}^{\infty} \mathbb{P}\{X < y\} f_Y(y) dy,$$

ce qui conduit bien sûr à l'expression familière :

$$\mathbb{P}\{X < Y\} = \int_{-\infty}^{\infty} \int_{-\infty}^y f_X(x) f_Y(y) dx dy.$$

De même,

$$\begin{aligned} \mathbb{P}\{X + Y \leq z\} &= \mathbb{E}[\mathbb{P}\{X + Y \leq Z \mid Y\}] \\ &= \int_{\mathbb{R}} \mathbb{P}\{X + Y \leq z \mid Y = y\} f_Y(y) dy \\ &= \int_{\mathbb{R}} \mathbb{P}\{X \leq z - y \mid Y = y\} f_Y(y) dy \\ &= \int_{\mathbb{R}} \mathbb{P}\{X \leq z - y\} f_Y(y) dy \\ &= \int_{\mathbb{R}} \int_{-\infty}^{z-y} f_X(x) f_Y(y) dx dy. \end{aligned}$$

EXEMPLE 7.21. On roule en voiture le long d'une route et il faudra mettre de l'essence d'ici les n prochaines sorties d'autoroute. À chaque sortie, un panneau annonce le prix de l'essence assez d'avance pour qu'on ait le temps de choisir de s'arrêter là.

Notre objectif, c'est de payer l'essence la moins chère possible. Pour ce faire, on fait l'hypothèse que le prix de l'essence à chaque sortie est indépendant des autres, et identiquement distribué avec les autres.

On ne fait pas demi-tour. Notre stratégie pour obtenir le meilleur prix possible sera la suivante :

- On laisse passer k sorties en regardant les prix, pour se faire une idée du marché.
- Puis, on s'arrête à la première station-services qui offre un meilleur prix que toutes les précédentes.

Quelle valeur devrait prendre k pour optimiser les chances d'obtenir le meilleur prix ?

SOLUTION. On résout ce problème en conditionnant. Soit E l'événement « on trouve le meilleur prix possible », et soit X le numéro de la sortie où se trouve le meilleur prix possible.

Si on note \mathbb{P}_k la mesure de probabilité qui décrit l'expérience où on laisse passer les k premières stations-service, alors, puisque les prix sont identiquement distribués et indépendants, on a que

$$\mathbb{P}_k\{X = i\} = \frac{1}{n}.$$

De plus,

$$\begin{aligned} PR_k\{E\} &= \sum_{i=1}^n \mathbb{P}_k\{E \mid X = i\} \mathbb{P}_k\{X = i\} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}_k\{E \mid X = i\}. \end{aligned}$$

La question est donc de trouver $\mathbb{P}_k \{E \mid X = i\}$.

La seule façon qu'on a de réussir à prendre le meilleur prix si il est à la position i (pour $i > k$), c'est que de tous les $i - 1$ premières stations services croisées, le meilleur prix se trouvait dans les k qu'on a sauté. En effet, si c'est le cas, on est garantis de ne pas s'être arrêté avant i , et puisque le meilleur prix est en i , c'est sûr que c'est là qu'on s'arrête, vu qu'on n'aura pas vu de meilleur prix que ça nulle part.

Donc,

$$\mathbb{P}_k \{E \mid X = i\} = \frac{k}{i-1},$$

lorsque $i > k$ – ça correspond à la probabilité que le meilleur des $i - 1$ premiers prix se retrouve parmi les k premières positions.

Bien sûr, pour $i \leq k$, c'est mort : $\mathbb{P}_k \{E \mid X = i\} = 0$, vu qu'on va avoir sauté la meilleure station-service.

Finalement, on a que

$$\begin{aligned}\mathbb{P}_k \{E\} &= \frac{1}{n} \sum_{i=k+1}^n \frac{k}{i-1} \\ &= \frac{k}{n} \sum_{i=k+1}^n \frac{1}{i-1}.\end{aligned}$$

Supposons qu'on choisisse $k = \alpha n$. Alors,

$$\mathbb{P}_{\alpha n} \{E\} = \alpha \sum_{i=\alpha n}^{n-1} \frac{1}{i}$$

et avec n gran,

$$\sum_{i=\alpha n}^{n-1} \sim \log(n-1) - \log(\alpha) - \log(n) \sim \log(\alpha).$$

Donc on trouve que

$$\mathbb{P}_{\alpha n} \{E\} \sim -\alpha \log \alpha$$

lorsque n est grand, et en mettant la dérivée égale à 0, on trouve que $\mathbb{P}_{\alpha n} \{E\}$ est maximal lorsque $\alpha = e^{-1}$.

On conclue donc que $k \approx ne^{-1}$ est la meilleure valeur pour maximiser la probabilité d'obtenir le meilleur prix possible.

Autrement dit, la stratégie otimale consisterait à laisser passer environ 37% des premières sorties, pour se faire une idée du marché, puis d'arrêter à la première qui offre un meilleur prix.

7.3.4. Variance conditionnelle. On va définir la variance conditionnelle de X sachant Y par

$$(7.3.13) \quad \text{Var}[X \mid Y] = \mathbb{E} \left[(X - \mathbb{E}[X \mid Y])^2 \mid Y \right].$$

Il s'agit de la même définition de la variance à laquelle nous sommes habitué.e.s, mais cette fois-ci, c'est une espérance conditionnelle qui remplace l'espérance « ordinaire ».

En appliquant les mêmes astuces de linéarité que précédemment, on trouve que

$$\text{Var}[X \mid Y] = \mathbb{E}[X^2 \mid Y] - \mathbb{E}[X \mid Y]^2.$$

Si on prend l'espérance, on trouve

$$\begin{aligned}\mathbb{E}[\text{Var}[X | Y]] &= \mathbb{E}[\mathbb{E}[X^2 | Y]] - \mathbb{E}[\mathbb{E}[X | Y]^2] \\ (7.3.14) \quad &= \mathbb{E}[X^2] - \mathbb{E}[\mathbb{E}[X | Y]^2].\end{aligned}$$

D'un autre côté,

$$\begin{aligned}\text{Var}[\mathbb{E}[X | Y]] &= \mathbb{E}[\mathbb{E}[X | Y]^2] - \mathbb{E}[\mathbb{E}[X | Y]]^2 \\ (7.3.15) \quad &= \mathbb{E}[\mathbb{E}[X | Y]^2] - \mathbb{E}[X]^2.\end{aligned}$$

Le résultat net, c'est donc que

$$(7.3.16) \quad \text{Var}[X] = \mathbb{E}[\text{Var}[X | Y]] + \text{Var}[\mathbb{E}[X | Y]].$$

EXEMPLE 7.22. Au temps t , le nombre de passagers arrivés sur le quai d'une gare est une variable aléatoire de Poisson de paramètre λt . Le train part à une heure uniformément distribuée dans l'intervalle $(0, T)$, indépendamment du nombre de passagers arrivés, trouver l'espérance et la variance du nombre de passagers qui seront à bord.

SOLUTION. L'heure de départ du train est U , donc le nombre $N(U)$ de passagers à bord est une variable aléatoire de Poisson de paramètres λU si U .

Donc, $\mathbb{E}[N(U)] = \mathbb{E}[\mathbb{E}[N(U) | U]] = \mathbb{E}[\lambda U] = \frac{\lambda T}{2}$.

Sachant U , la variance $\text{Var}[N(U) | U] = \lambda U$ aussi, puisque $N(U)$ est une poisson de paramètre λU .

Donc, $\mathbb{E}[\text{Var}[N(U) | U]] = \mathbb{E}[\lambda U] = \frac{\lambda T}{2}$. D'un autre côté,

$$\text{Var}[\mathbb{E}[X | U]] = \text{Var}[\lambda U] = \lambda^2 \frac{T^2}{12}.$$

On trouve finalement que la variance de $N(U)$ est

$$\text{Var}[N(U)] = \frac{\lambda T}{2} + \frac{\lambda^2 T^2}{12}.$$

EXEMPLE 7.23. Supposons que $(X_i)_{i \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, avec $\mathbb{E}[X_1] = \mu$. et variance σ^2 . Soit N indépendante des X_i une variable aléatoire également.

On définit $S = \sum_{i=1}^N X_i$. Trouver $\mathbb{E}[S]$ et $\text{Var}[S]$.

SOLUTION. On a bien sûr que

$$\mathbb{E}[S] = \mathbb{E}[\mathbb{E}[S | N]] = \mu \mathbb{E}[N],$$

comme nous avons déjà vu dans un autre exemple.

Pour la variance, on a que

$$\text{Var}[S | N] = N \sigma^2,$$

et

$$\mathbb{E}[\text{Var}[S | N]] = \sigma^2 \mathbb{E}[N].$$

D'un autre côté, on a que

$$\text{Var}[\mathbb{E}[S | N]] = \text{Var}[N \mu] = \mu^2 \text{Var}[N].$$

Finalement,

$$\text{Var}[S] = \mu^2 \text{Var}[N] + \sigma^2 \mathbb{E}[N].$$

7.3.5. L'espérance conditionnelle comme outil de prédition. Jusqu'ici nous nous sommes beaucoup servi de l'espérance conditionnelle comme d'un outil analytique qui nous permettait de calculer la valeur d'espérances ou de probabilités autres. Cependant, les espérances conditionnelles sont également très utiles en elles-mêmes. En effet, nous allons montrer qu'elles sont dans un certain sens « la meilleure prédition qu'on puisse faire » d'une variable X sachant Y .

PROPOSITION 7.13. *Soient X, Y deux variables aléatoires. Alors, pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$, on a toujours que*

$$(7.3.17) \quad \mathbb{E}[(X - \mathbb{E}[X | Y])^2] \leq \mathbb{E}[(X - g(Y))^2].$$

L'égalité est vraie si et seulement si $g(Y) = \mathbb{E}[X | Y]$.

REMARQUE. Supposons qu'on choisissait g pour que $\mathbb{E}[g(Y)] = \mathbb{E}[X]$. Alors, ce que l'énoncé dit, c'est que $\text{Var}[X - \mathbb{E}[X | Y]] \leq \text{Var}[X - g(Y)]$ pour toute fonction g .

Autrement dit, supposons qu'on avait un modèle prédictif de X sachant Y , qui nous disait qu'une approximation de X pourrait être faite par une valeur $g(Y)$ entièrement déterminée par la valeur de Y .

Ce que la proposition nous dit, c'est que l'erreur qu'on commet en faisant cette approximation aura la plus petite variance lorsque notre modèle g correspondra à l'espérance conditionnelle.

DÉMONSTRATION. On a d'abord que

$$\begin{aligned} \mathbb{E}[(X - g(Y))^2 | Y] &= \mathbb{E}[(X - \mathbb{E}[X | Y] + \mathbb{E}[X | Y] - g(Y))^2 | Y] \\ &= \mathbb{E}[(X - \mathbb{E}[X | Y])^2 | Y] \\ &\quad + \mathbb{E}[(\mathbb{E}[X | Y] - g(Y))^2 | Y] \\ &\quad + 2\mathbb{E}[(X - \mathbb{E}[X | Y])(\mathbb{E}[X | Y] - g(Y)) | Y] \end{aligned}$$

Si on décortique le dernier terme, on voit que l'on a

$$\begin{aligned} &\mathbb{E}[(X - \mathbb{E}[X | Y])(\mathbb{E}[X | Y] - g(Y)) | Y] \\ &= \mathbb{E}[(X - \mathbb{E}[X | Y]) | Y] \cdot (\mathbb{E}[X | Y] - g(Y)) \\ &= (\mathbb{E}[X | Y] - g(Y)) \cdot (\mathbb{E}[X | Y] - \mathbb{E}[X | Y]) \\ &= 0. \end{aligned}$$

Donc, on a

$$\begin{aligned} &\mathbb{E}[(X - g(Y))^2 | Y] - \mathbb{E}[(X - \mathbb{E}[X | Y])^2 | Y] \\ &= \mathbb{E}[(\mathbb{E}[X | Y] - g(Y))^2 | Y] \\ &\geq 0 \end{aligned}$$

et en prenant l'espérance de part et d'autre on trouve le résultat attendu. \square

EXEMPLE 7.24. Un signal analogique est transmis du point A au point B par un circuit ; cependant le circuit induit un bruit aléatoire de distribution normale centrée avec variance 1. Donc si le signal s est envoyé en A , le signal reçu en B est $s + \mathcal{N}(0, 1) \sim \mathcal{N}(s, 1)$.

Supposons que le signal envoyé en A est un signal aléatoire S qui suit une distribution normale de paramètres μ, σ^2 , et que le signal reçu en B soit le signal R .

Quelle est la meilleure approximation du signal envoyé en A , si on ne sait que la valeur du signal reçue ?

SOLUTION. Il faut d'abord identifier la fonction de densité du signal envoyé sachant le signal reçu.

Soit $Z \sim \mathcal{N}(0, 1)$ le bruit induit par le circuit, indépendant de $S \sim \mathcal{N}(\mu, \sigma^2)$. Alors, le signal reçu est $R = S + Z$.

On sait que la densité conditionnelle du signal envoyé sachant le signal reçu est

$$f_{S|R}(s | r) = \frac{f_{R|S}(r | s) f_S(s)}{f_R(r)}.$$

On a que

$$f_R(r) = C_R \exp\left(-\frac{1}{2} \frac{(r - \mu)^2}{\sigma^2 + 1}\right),$$

$$f_S(s) = C_S \exp\left(-\frac{1}{2} \frac{(s - \mu)^2}{\sigma^2}\right),$$

et

$$f_{R|S}(r | s) = C_{RS} \exp\left(-\frac{(r - s)^2}{2}\right).$$

Heureusement, $f_R(r)$ ne dépend pas de s , donc on ne s'en préoccupe pas.

On trouve que

$$f_{S|R}(s | r) = C \exp\left(-\frac{1}{2} \frac{\left(s - \frac{\mu + r\sigma^2}{1 + \sigma^2}\right)^2}{\frac{\sigma^2}{1 + \sigma^2}}\right).$$

C ne dépend pas de s .

Donc, sachant R , S est une variable aléatoire normale de paramètre $\mu' = \frac{\mu + r\sigma^2}{\sigma^2 + 1}$ et $\sigma'^2 = \frac{\sigma^2}{\sigma^2 + 1}$.

La On devrait donc assumer que le signal S était de valeur

$$\mathbb{E}[S | R] = \frac{1}{\sigma^2 + 1} \mu + \frac{\sigma^2}{\sigma^2 + 1} R,$$

une moyenne pondérée entre R la valeur reçue, et μ la valeur la plus probablement envoyée.

Les poids de la pondération sont en proportion du rapport entre l'intensité moyenne du bruit et l'intensité moyenne du signal – ce qu'on appelle le rapport signal sur bruit. Plus l'intensité moyenne du signal est en haut rapport σ^2 avec l'intensité moyenne du bruit, et plus on doit se fier à la valeur reçue plutôt qu'à la valeur la plus probablement envoyée. À l'inverse, plus le signal est de faible intensité par rapport au bruit, et plus notre valeur reçue est ignorée en faveur de ce qu'on connaît *a priori* de la source.

7.4. Fonctions génératrices

L'espérance n'a pas fini de nous étonner. Dans cette section, nous allons voir comment en prenant des espérances de façon astucieuse, on va drôlement se simplifier la vie – en particulier, on va voir comment on peut, à l'aide de l'espérance, définir des fonctions qui nous permettront d'encapsuler toute l'information d'une distribution.

On fait souvent l'erreur de s'imaginer que, si deux variables aléatoires X et Y ont la même espérance, elles ont la même distribution. Bien entendu, c'est faux, et vous pouvez trouver rapidement une myriade de contre-exemples.

Mais si elles ont aussi la même variance ? Non, c'est encore faux.

Mais si elles ont aussi le même troisième moment ? Non. Et si ...

7.4.1. La fonction génératrice des probabilités.

DÉFINITION 7.4 (Fonction génératrice des probabilités). Soit X une variable aléatoire discrète à valeurs entières non-négatives. On définit la **fonction génératrice des probabilités** de X :

$$(7.4.1) \quad \psi_X(s) = \mathbb{E}[s^X].$$

Si on développe cette expression en une série, par notre formule pour le calcul des espérances de variables discrètes, on trouve

$$(7.4.2) \quad \begin{aligned} \psi_X(s) &= \sum_{i=0}^{\infty} \mathbb{P}\{X=i\} s^i \\ &= \mathbb{P}\{X=0\} + \sum_{i=1}^{\infty} s^i \mathbb{P}\{X_i\}. \end{aligned}$$

REMARQUE. Par un théorème du cours Analyse II, on sait que le rayon de convergence de cette série est d'au moins 1, puisque les coefficients sont tous bornés par 1 (vu que ce sont des probabilités). On sait aussi que la série doit converger en 1 (on va le voir tout de suite), mais on sait aussi que la convergence est uniforme. Donc, la fonction est infiniment différentiable sur $[-1, 1]$ au moins, etc. Donc pas de souci, tout est bien défini.

On déduit immédiatement les propriétés suivantes :

PROPOSITION 7.14. Soit X une variable aléatoire entière non-négative, et soit $\psi_X(s)$ sa fonction génératrice des probabilités.

Alors, on a les propriétés suivantes :

- i. $\psi_X(0) = \mathbb{P}\{X=0\}$.
- ii. $\psi_X(1) = 1$.
- iii. $\psi_X^{(n)}(1) = \mathbb{E}[X(X-1)(X-2)\cdots(X-n+1)]$. En particulier, $\psi'_X(0) = \mathbb{E}[X]$; On a aussi que $\text{Var}[X] = \psi''_X(1) + \psi'_X(1)(1 - \psi'_X(1))$.
- iv. $\psi_X^{(n)}(0) = n! \mathbb{P}\{X=n\}$.
- v. ψ_X et toutes ses dérivées sont positives, non-décroissantes et convexe sur $(0, 1)$.

DÉMONSTRATION. i. Par l'équation (7.4.2), en remplaçant s par 0, tous les termes disparaissent sauf le premier, ce qui nous donne le résultat voulu.

- ii. Pareil, en remplaçant s par 1, on additionne tous les termes – et par la condition de normalisation, la somme doit donner 1.

- iii. D'abord, la remarque précédente nous rassure : la fonction est infiniment différentiable (lisse) sur $(-1, 1)$. En 1, on va prendre les dérivées par continuité ; c'est ce qui est le plus naturel. On s'intéresse spécialement à sa *nième* dérivée.

$$\begin{aligned}\psi'_X(s) &= \mathbb{P}\{X = 1\} + \sum_{i=2}^{\infty} is^{i-1} \mathbb{P}\{X = i\} \\ \psi''_X(s) &= 2\mathbb{P}\{X = 2\} + \sum_{i=3}^{\infty} i(i-1)s^{i-2} \mathbb{P}\{X = i\} \\ &\dots = \dots \\ \psi_X^{(n)}(s) &= n! \mathbb{P}\{X = n\} + \sum_{i=n+1}^{\infty} \frac{i!}{(i-n)!} s^{i-n} \mathbb{P}\{X = i\}.\end{aligned}$$

Évalué.e.s en $s = 1$, on recouvre les expressions pour les *moments factoriels* :

$$\psi_X^{(n)}(s) = \mathbb{E} \left[\frac{X!}{(X-n)!} \right] = \mathbb{E}[X(X-1)(X-2)\cdots(X-n+1)].$$

En particulier,

$$\psi'_X(1) = \mathbb{P}\{X = 1\} + \sum_{i=2}^{\infty} i \mathbb{P}\{X = i\} = \mathbb{E}[X],$$

et

$$\psi''_X(1) + \psi'_X(1)(1 - \psi'_X(1)) = \mathbb{E}[X(X-1)] + \mathbb{E}[X](1 - \mathbb{E}[X]) = \text{Var}[X].$$

- iv. Par le calcul précédent, on voit bien qu'en remplaçant s par 0 dans l'expression pour la *nième* dérivée, on obtient $\psi_X^{(n)}(0) = n! \mathbb{P}\{X = n\}$.
- v. Finalement, on voit que sur $(0, 1)$, $\psi_X^{(n)}(s) \geq 0$ – il suit que $\psi_X(s), \psi'_X, \psi''_X(s)$ sont toutes positives, soit que $\psi_X(s)$ est positive, non-décroissante et convexe. Mais pour tout $k \geq 1$, $\psi_X^{(k)}(s), \psi_X^{(k+1)}(s), \psi_X^{(k+2)} \geq 0$, donc $\psi_X^{(k)}$ est positive, non-décroissante et convexe.

□

REMARQUE. La propriété iv est celle qui donne leur nom aux fonctions génératrices des probabilités – en effet, en comparant $\psi_X(s)$ à sa série de MacLaurin, on aurait pu anticiper cette propriété.

C'est un peu comme si la fonction génératrice des probabilités était une grosse boîte noire avec une manivelle. À chaque fois qu'on dérive, on donne un tour de manivelle, et la boîte nous recrache le prochain coefficient : $\mathbb{P}\{X = n\}$.

EXEMPLE 7.25. On calcule des fonctions génératrices des probabilités pour les distributions discrètes les plus importantes.

- i. Soit X une variable aléatoire de Bernoulli avec probabilité p de succès. Alors, sa fonction génératrice des probabilités est simplement

$$(7.4.3) \quad \psi_X(s) = (1-p) + ps.$$

- ii. Soit X une variable aléatoire binomiale de paramètres n, p . Alors, sa fonction génératrice des probabilités est :

$$\begin{aligned}
 \psi_X(s) &= \mathbb{E}[s^X] \\
 &= \sum_{i=0}^n \binom{n}{i} s^i p^i (1-p)^{n-i} \\
 (7.4.4) \quad &= ((1-p) + ps)^n.
 \end{aligned}$$

- iii. Soit X une variable aléatoire géométrique de paramètre p . Alors, sa fonction génératrice des probabilités est :

$$\begin{aligned}
 \psi_X(s) &= \mathbb{E}[s^X] \\
 &= \sum_{i=1}^{\infty} s^i (1-p)^{i-1} p \\
 (7.4.5) \quad &= \frac{ps}{1 - (1-p)s},
 \end{aligned}$$

pour $|s(1-p)| < 1$ – le rayon de convergence est donc $1/(1-p) \geq 1$.

- iv. Soit X une variable aléatoire suivant une loi de Poisson de paramètre $\lambda > 0$. Alors, sa fonction génératrice des probabilités est :

$$\begin{aligned}
 \psi_X(s) &= \mathbb{E}[s^X] \\
 &= \sum_{i=0}^{\infty} s^i e^{-\lambda} \frac{\lambda^i}{i!} \\
 (7.4.6) \quad &= e^{\lambda(s-1)}.
 \end{aligned}$$

EXEMPLE 7.26. Soit X une variable aléatoire dont la fonction génératrice des probabilités est donnée par $\psi_X(s)$.

- (a) Trouver la fonction génératrice des probabilités de $X + k$, pour $k \in \mathbb{N}$.
- (b) Trouver la fonction génératrice des probabilités de kX , pour $k \in \mathbb{N}$.

SOLUTION. (a) On a que

$$\psi_{X+k}(s) = \mathbb{E}[s^{X+k}] = s^k \psi_X(s).$$

(b) On a que

$$\psi_{kX}(s) = \mathbb{E}[s^{kX}] = \mathbb{E}[(s^k)^X] = \psi_X(s^k).$$

EXEMPLE 7.27. Soit X une variable aléatoire avec fonction génératrice des probabilités $\psi_X(s)$. On définit maintenant Y une nouvelle variable aléatoire, avec fonction génératrice des probabilités

$$\psi_Y(s) = \frac{s\psi'_X(s)}{\psi'_X(1)}.$$

Trouver la fonction de masse, l'espérance et la variance de Y en fonction de celles de X .

SOLUTION. On dérive :

$$\begin{aligned}\psi'_Y(s) &= \frac{1}{\psi'_X(1)} (s\psi''_X(s) + \psi'_X(s)) \\ \psi''_Y(s) &= \frac{1}{\psi'_X(1)} (s\psi^{(3)}_X(s) + 2\psi''_X(s)) \\ \psi^{(3)}_Y(s) &= \frac{1}{\psi'_X(1)} (s\psi^{(4)}_X(s) + 3\psi^{(3)}_X(s)) \\ &\dots = \dots \\ \psi^{(n)}_Y(s) &= \frac{1}{\psi'_X(1)} (s\psi_X^{(n+1)}(s) + n\psi_X^{(n)}(s)).\end{aligned}$$

Puis, on calcule : en $s = 0$, on a

$$\psi_Y^{(n)}(0) = \frac{n\psi_X^{(n)}(0)}{\psi'_X(1)},$$

et par la proposition 7.14, ça devient

$$n! \mathbb{P}\{Y = n\} = \frac{n \cdot n! \mathbb{P}\{X = n\}}{\mathbb{E}[X]},$$

ou plus simplement

$$\mathbb{P}\{Y = n\} = \frac{n \mathbb{P}\{X = n\}}{\mathbb{E}[X]}.$$

On a donc que l'espérance de Y est

$$\begin{aligned}\mathbb{E}[Y] &= \frac{1}{\mathbb{E}[X]} \sum_{n=0}^{\infty} n^2 \mathbb{P}\{X = n\} \\ &= \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}.\end{aligned}$$

Le second moment de Y est

$$\begin{aligned}\mathbb{E}[Y^2] &= \frac{1}{\mathbb{E}[X]} \sum_{n=0}^{\infty} n^3 \mathbb{P}\{X = n\} \\ &= \frac{\mathbb{E}[X^3]}{\mathbb{E}[X]}.\end{aligned}$$

La variance de Y est donc

$$\begin{aligned}\text{Var}[Y] &= \frac{\mathbb{E}[X^3]}{\mathbb{E}[X]} - \frac{\mathbb{E}[X^2]^2}{\mathbb{E}[X]^2} \\ &= \frac{\mathbb{E}[X]\mathbb{E}[X^3] - \mathbb{E}[X^2]^2}{\mathbb{E}[X]^2}.\end{aligned}$$

On aurait pu obtenir les mêmes données avec les fonctions génératrices :

$$\begin{aligned}\psi'_Y(1) &= \frac{1}{\psi'_X(1)} (\psi''_X(1) + \psi'_X(1)) \\ &= 1 + \frac{\mathbb{E}[X(X-1)]}{\mathbb{E}[X]} \\ &= \frac{\mathbb{E}[X^2]}{\mathbb{E}[X]}.\end{aligned}$$

De plus,

$$\begin{aligned}\psi''_Y(1) &= \frac{1}{\psi'_X(1)} (\psi^{(3)}_X(1) + 2\psi''_X(1)) \\ &= \frac{1}{\mathbb{E}[X]} (\mathbb{E}[X(X-1)(X-2)] + 2\mathbb{E}[X(X-1)]) \\ &= \frac{1}{\mathbb{E}[X]} (\mathbb{E}[X^3] - \mathbb{E}[X^2])\end{aligned}$$

Donc,

$$\begin{aligned}\text{Var}[Y] &= \psi''_Y(1) + \psi'_Y(1)(1 - \psi'_Y(1)) \\ &= \frac{1}{\psi'_X(1)} (\psi^{(3)}_X(1) + 2\psi''_X(1)) \\ &\quad + \frac{1}{\psi'_X(1)^2} (\psi''_X(1) + \psi'_X(1)) (\psi'_X(1) - (\psi''_X(1) + \psi'_X(1))) \\ &= \frac{1}{\mathbb{E}[X]} (\mathbb{E}[X^3] - \mathbb{E}[X^2]) \\ &\quad - \frac{1}{\mathbb{E}[X]^2} \mathbb{E}[X^2] (\mathbb{E}[X^2] - \mathbb{E}[X]) \\ &= \frac{1}{\mathbb{E}[X]^2} (\mathbb{E}[X]\mathbb{E}[X^3] - \mathbb{E}[X^2]^2).\end{aligned}$$

REMARQUE. On dit que la distribution de Y est « la distribution de X , biaisée par la taille. » Par exemple, dans le problème des chauffeurs d'autobus (exercice 4.5), X est la « version biaisée par la taille » de Y .

Bien sûr, la possibilité de déduire toute la fonction de masse de probabilités d'une variable aléatoire à partir de sa fonction génératrice des probabilités n'est pas anodine. On se souviendra qu'à la fin du chapitre 5, la proposition 5.11 nous donnait déjà plusieurs façons d'identifier que deux variables étaient équivalentes en loi.

On peut maintenant en rajouter une nouvelle :

PROPOSITION 7.15. *Soient X, Y deux variables aléatoires entières non-négatives. Alors, X et Y sont équivalentes en loi si et seulement si elles ont la même fonction génératrice des probabilités.*

DÉMONSTRATION. D'une part, si X et Y sont équivalentes en loi, alors clairement elles doivent avoir la même fonction génératrice des probabilités.

D'autre part, si X et Y ont la même fonction génératrice des probabilités, alors elles ont la même fonction de masse de probabilités, puisque celle-ci est entièrement déterminée par la fonction génératrice des probabilités. \square

Cette dernière proposition est la raison pour laquelle on s'intéresse aux fonctions génératrices des probabilités : dans certaines situations, elles sont plus rapides et plus faciles à manipuler, et surtout, elles permettent plus rapidement de comprendre la distribution de certaines variables aléatoires.

PROPOSITION 7.16. *Soient X et Y deux variables aléatoires entières non-négatives indépendantes, et soient ψ_X et ψ_Y leurs fonctions génératrices des probabilités respectives. Soit $Z = X + Y$, et ψ_Z sa fonction génératrice des probabilités.*

Alors,

$$(7.4.7) \quad \psi_Z = \psi_X \psi_Y.$$

DÉMONSTRATION. Par la proposition 7.5, on a que

$$\begin{aligned} \psi_Z(s) &= \mathbb{E}[s^{X+Y}] \\ &= \mathbb{E}[s^X s^Y] \\ &= \mathbb{E}[s^X] \mathbb{E}[s^Y] \\ &= \psi_X(s) \psi_Y(s). \end{aligned}$$

\square

COROLLAIRE 7.4. *Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires indépendantes avec fonctions génératrices des probabilités respectives $\psi_{X_1}, \psi_{X_2}, \psi_{X_3}, \dots, \psi_{X_n}$, alors avec $Z = \sum_{i=1}^n X_i$ et sa fonction génératrice des probabilités ψ_Z , on a*

$$(7.4.8) \quad \psi_Z = \prod_{i=1}^n \psi_{X_i}.$$

DÉMONSTRATION. La preuve peut être faite par induction et est l'objet de l'exercice 7.21. \square

L'exemple suivant utilise les propositions 7.15 et 7.16 et le corollaire 7.4 pour tirer des conclusions sur les sommes de variables aléatoires indépendantes aux distributions connues.

EXEMPLE 7.28. *i.* On sait que si les X_i sont des Bernoulli indépendantes avec probabilité p de succès et fonction génératrice des probabilités $\psi_{X_1}(s) = (1-p) + ps$, et $n \geq 0$, alors avec $S = \sum_{i=1}^n X_i$, on a que

$$\psi_S(s) = \prod_{i=1}^n \psi_{X_i}(s) = ((1-p) + ps)^n.$$

On déduit alors directement que S suit une loi binomiale.

- ii.* Si X et Y sont des variables aléatoires indépendantes qui suivent des lois binomiales de paramètres respectifs m, p et n, p , alors la somme des deux a la fonction génératrice des probabilités :

$$\psi_{X+Y}(s) = ((1-p) + ps)^m ((1-p) + ps)^n = ((1-p) + ps)^{m+n},$$

ce qui signifie que $X + Y$ suit une loi binomiale de paramètres $m + n, p$.

- iii. Si X et Y sont des variables aléatoires indépendantes qui suivent des lois de Poisson de paramètres respectifs λ et μ , alors la somme des deux a la fonction génératrice des probabilités :

$$\psi_{X+Y}(s) = e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)},$$

et $X + Y$ suit une loi de Poisson de paramètre $\lambda + \mu$.

- iv. Si $(X_i)_{i \in \mathbb{N}}$ est une famille de variables aléatoires indépendantes de loi géométriques de paramètre p , et $r \geq 1$, alors on sait que $S = \sum_{i=1}^r X_i$ est une variable aléatoire de loi binomiale négative, de paramètre r, p . On peut donc déduire la fonction génératrice des moments d'une variable aléatoire binomiale négative :

$$\psi_S(s) = \prod_{i=1}^r \psi_{X_i}(s) = \left(\frac{ps}{1 - (1-p)s} \right)^r.$$

7.4.2. La fonction génératrice des moments. Vous aurez remarqué que, malgré que j'aie défini la fonction génératrice des probabilités dans le contexte des variables aléatoires entières non-négatives, il n'y a rien qui nous empêche, *a priori*, de considérer une fonction définie exactement de la même façon pour des variables aléatoires quelconques.

EXEMPLE 7.29. Soit X une variable aléatoire qui suit une distribution exponentielle standard. Trouver $\mathbb{E}[s^X]$.

SOLUTION. On fait simplement le calcul :

$$\begin{aligned} \mathbb{E}[s^X] &= \int_0^\infty s^x e^{-x} dx \\ &= \int_0^\infty \left(\frac{s}{e}\right)^x dx. \end{aligned}$$

Pour peu que $0 < s < e$, on sait que :

$$\begin{aligned} \mathbb{E}[s^X] &= \left[\frac{1}{\log(\frac{s}{e})} \left(\frac{s}{e}\right)^x \right]_0^\infty \\ &= \frac{1}{1 - \log s}. \end{aligned}$$

Évidemment, pour $s = 0$, l'intégrale est nulle. Pour $s = 1$, on a bien que $\mathbb{E}[s^X] = 1$.

L'ennui avec ça, c'est que les interprétations qu'on peut en faire sont beaucoup moins intéressantes – si X est une variable aléatoire continue, par exemple, la fonction $\mathbb{E}[s^X]$ sera lisse, certes, mais sa série de MacLaurin ne convergera pas vers la fonction !

Pour cette raison, on introduit plutôt la fonction génératrice des moments :

DÉFINITION 7.5. Soit X une variable aléatoire quelconque. On définit sa **fonction génératrice des moments** :

$$(7.4.9) \quad M_X(t) := \mathbb{E}[e^{tX}].$$

REMARQUE. On voit que si X est une variable aléatoire entière non-négative avec fonction génératrice des probabilités ψ_X , alors sa fonction génératrice des moments est donnée par

$$M_X(t) = \psi_X(e^t).$$

La fonction génératrice des moments est donc utile pour une classe plus large de variables aléatoires.

Si on décortique cette définition, on a que pour une variable aléatoire quelconque,

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right]. \end{aligned}$$

Par la proposition 7.3, si $X > 0$ ou si $\mathbb{E}[|X|^k] < +\infty$ pour tout k , on peut intervertir la somme et l'espérance. Dans ces cas (assez répandus), on a que

$$(7.4.10) \quad M_X(t) = 1 + \sum_{k=1}^{\infty} \frac{\mathbb{E}[X^k]}{k!} t^k.$$

REMARQUE. *Sur le rayon de convergence de la série :* Cette série converge uniformément sur tous les compacts non-vides contenus dans un certain voisinage de 0 dès que

$$\left| \frac{\mathbb{E}[X^k]}{k!} \right| = O(c^k),$$

pour une constante $0 < c < +\infty$. Si on a que $\mathbb{E}[X^k] = O(c^k)$ pour une constante c quelconque, alors la série converge sur tout \mathbb{R} , et uniformément sur tout compact.

Autrement dit, aussitôt que tous les moments d'une variable aléatoire sont finis, et pour peu qu'ils croissent moins vite que $c^k k!$ pour une certaine constante c , il existe un intervalle $(-\epsilon, \epsilon)$ tel que pour tout compact contenu dans cet intervalle, $M_X(t)$ est égale au membre de droite de l'équation (7.4.10).

Autant dire que pour à peu près toutes les distributions d'intérêt, nous n'aurons pas de problème de convergence.

En particulier⁷, $M_X(t)$ est dpmc égale à sa série de MacLaurin, au moins sur un petit intervalle autour de 0.

On a les propriétés suivantes :

PROPOSITION 7.17. *Soit X une variable aléatoire avec $\mathbb{E}[|X^k|] < \infty$ et telle que $\sum_{i=1}^{\infty} \mathbb{E}[X^i] t^i / i!$ converge sur un voisinage de 0. Soit $M_X(t)$ la fonction génératrice des moments de X sur cet intervalle.*

Alors,

- i. $M_X(0) = 1$.
- ii. $M_X^{(n)}(0) = \mathbb{E}[X^n]$. En particulier, $M'_X(0) = \mathbb{E}[X]$.

DÉMONSTRATION. i. Cela est trivial par définition.

7. Je ne le répéterai pas toujours, mais les raisonnements sont toujours contingents à la bonne convergence de la série/des intégrales appropriées.

- ii. On a vu que si $\sum_{i=0}^{\infty} \mathbb{E}[X^i] t^i / i!$ converge sur un voisinage de 0, alors $M_X(t)$ est égale à cette série, sa série de MacLaurin sur ce voisinage. En particulier, $M_X(t)$ y est infiniment différentiable, et en comparant les séries, on voit que

$$\sum_{i=0}^{\infty} \frac{M_X^{(i)}(0)}{i!} t^i = M_X(t) = \sum_{i=0}^{\infty} \frac{\mathbb{E}[X^i]}{i!} t^i.$$

Et puisque les polynômes sont linéairement indépendants, il faut nécessairement que pour tout n , $\mathbb{E}[X^n] = M_X^{(n)}(0)$.

Ou, si vous préférez :⁸ pour peu que la série converge uniformément, on peut intervertir la différentiation et l'espérance. On se retrouve alors avec

$$\begin{aligned} M'_X(t) &= \frac{d}{dt} \mathbb{E}[e^{tX}] \\ &= \mathbb{E}\left[\frac{d}{dt} e^{tX}\right] \\ &= \mathbb{E}[X e^{tX}]; \end{aligned}$$

de même,

$$M_X^{(n)}(t) = \mathbb{E}[X^n e^{tX}],$$

et en remplaçant t par 0, on retrouve ce résultat. □

Autrement dit, comme pour la fonction génératrice des probabilités, la fonction génératrice des moments nous ressort un à un les moments de notre variable aléatoire X à mesure qu'on dérive.

EXEMPLE 7.30. Nous allons calculer les fonctions génératrices des moments pour toutes les distributions qu'on aime :

- i. La fonction génératrice des moments d'une variable X de Bernoulli avec probabilité de succès p est

$$(7.4.11) \quad M_X(t) = (1-p) + pe^t = \psi_X(e^t).$$

- ii. La fonction génératrice des moments d'une variable X binomiale avec paramètres n, p est

$$(7.4.12) \quad M_X(t) = ((1-p) + pe^t)^n = \psi_X(e^t).$$

- iii. La fonction génératrice des moments d'une variable X géométrique avec paramètre p est

$$(7.4.13) \quad M_X(t) = \frac{pe^t}{1 - (1-p)e^t} = \psi_X(e^t), \quad t < -\log(1-p).$$

- iv. La fonction génératrice des moments d'une variable X binomiale négative avec paramètres r, p est

$$(7.4.14) \quad M_X(t) = \left(\frac{pe^t}{1 - (1-p)e^t}\right)^r = \psi_X(e^t), \quad t < -\log(1-p).$$

8. Ce raisonnement peut sembler plus satisfaisant, mais il nécessite aussi beaucoup de justifications pour l'interversion de la différentiation et de l'espérance.

v. La fonction génératrice des moments d'une variable X de Poisson avec paramètre λ est

$$(7.4.15) \quad M_X(t) = e^{\lambda(e^t - 1)} = \psi_X(e^t).$$

vi. La fonction génératrice des moments d'une variable X uniforme sur $(0, 1)$ est

$$(7.4.16) \quad \begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_0^1 e^{tx} dx \\ &= \begin{cases} 1 & \text{si } t = 0 \\ \frac{1}{t}(e^t - 1) & \text{sinon.} \end{cases} \end{aligned}$$

vii. La fonction génératrice des moments d'une variable X normale centrée réduite est

$$(7.4.17) \quad \begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2tx + t^2 - t^2)} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} e^{\frac{1}{2}t^2} dx \\ &= e^{\frac{1}{2}t^2}. \end{aligned}$$

viii. La fonction génératrice des moments d'une variable X exponentielle standard est

$$(7.4.18) \quad \begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_0^{\infty} e^{tx} e^{-x} dx \\ &= \frac{1}{1-t}, \quad t < 1 \end{aligned}$$

ix. La fonction génératrice des moments d'une variable X de loi Gamma avec paramètres $\alpha, 1$ est

$$(7.4.19) \quad \begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-x(1-t)} dx \\ &= \frac{1}{(1-t)^{\alpha}}, \quad t < 1. \end{aligned}$$

Il existe une façon pratique de calculer rapidement les fonctions génératrices des moments pour des transformations affines de variables aléatoires dont on connaît déjà la fonction génératrice des moments :

PROPOSITION 7.18. *Soit X une variable aléatoire quelconque avec M_X sa fonction génératrice des moments. Soient a et b des constantes réelles quelconques, et soit $Y = aX + b$. Alors, la fonction génératrice des moments de Y est*

$$(7.4.20) \quad M_Y(t) = e^{bt} M_X(at).$$

DÉMONSTRATION. On a

$$M_Y(t) = \mathbb{E} [e^{t(aX+b)}] = e^{bt} \mathbb{E} [e^{(at)X}] = e^{bt} M_X(at).$$

□

EXEMPLE 7.31. Pour toutes les distributions continues plus haut, nous n'avons pas pris la peine de calculer la fonction génératrice des moments pour des paramètres généraux. Faisons-le maintenant :

- i. Si X est de distribution uniforme sur (a, b) , alors $Y = (X-a)/(b-a)$ est de distribution uniforme sur $(0, 1)$. Donc,

$$M_Y(t) = \frac{1}{t}(e^t - 1) = e^{-\frac{at}{b-a}} M_X\left(\frac{t}{b-a}\right),$$

d'où on déduit que, lorsque $s \neq 0$,

$$\begin{aligned} M_X(s) &= \frac{e^{as}}{(b-a)s} (e^{(b-a)s} - 1) \\ (7.4.21) \quad &= \frac{1}{(b-a)s} (e^{bs} - e^{as}). \end{aligned}$$

Évidemment, lorsque $s = 0$, on a obligatoirement que $M_X(0) = 1$.

- ii. Si X est de la distribution normale avec paramètres μ, σ^2 , alors $Y = (X - \mu)/\sigma$ est de distribution normale centrée réduite, et

$$M_Y(t) = e^{\frac{1}{2}t^2} = e^{-\mu\frac{t}{\sigma}} M_X\left(\frac{t}{\sigma}\right).$$

On déduit donc que

$$(7.4.22) \quad M_X(s) = e^{\mu s + \frac{1}{2}\sigma^2 s^2}.$$

- iii. Si X suit une distribution exponentielle de paramètre λ , alors $Y = \lambda X$ suit une distribution exponentielle standard. On a donc que

$$M_Y(t) = \frac{1}{1-t} = M_X(\lambda t),$$

d'où

$$(7.4.23) \quad M_X(s) = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

- iv. Si X suit une loi Gamma de paramètres α, λ , alors $Y = \lambda X$ suit une distribution Gamma de paramètres $\alpha, 1$. On a donc que

$$M_Y(t) = \frac{1}{(1-t)^\alpha} = M_X(\lambda t),$$

et

$$(7.4.24) \quad M_X(s) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \quad t < \lambda.$$

Comme pour la fonction génératrice des probabilités, la fonction génératrice des moments a pour propriété de caractériser complètement la distribution d'une variable aléatoire⁹ :

9. Pour autant que la fonction génératrice des moments converge sur un voisinage de 0.

PROPOSITION 7.19. *Soient X et Y deux variables aléatoires dont les fonctions génératrices des moments respectives M_X et M_Y convergent et sont égales sur un intervalle ouvert contenant 0.*

Alors, X et Y sont équivalentes en loi.

DÉMONSTRATION. La démonstration de cette proposition dépasse largement le cadre du cours. \square

REMARQUE. *Attention !* On pourrait être tenté.e.s de penser que, grâce à cette proposition, il suffit de vérifier que tous les moments de X et Y correspondent pour conclure que X et Y ont la même distribution. C'est faux ! Il faut aussi que les fonctions génératrices des moments existent.

Par exemple, la fonction génératrice des moments d'une variable de distribution log-normale ne converge pas (sauf en 0). Donc, si on trouve une autre variable aléatoire qui a les mêmes moments, on ne peut pas conclure que cette autre variable aléatoire suit aussi une distribution log-normale des mêmes paramètres.

On peut se servir de la proposition 7.19 pour identifier la distribution de variables aléatoires. En particulier, pour les sommes de variables aléatoires indépendantes, cela peut être très pratique.

PROPOSITION 7.20. *Soient X, Y deux variables aléatoires indépendantes avec fonctions génératrices des moments respectives M_X et M_Y . Soit $Z = X + Y$, et M_Z sa fonction génératrice des moments. Alors,*

$$(7.4.25) \quad M_Z(t) = M_X(t)M_Y(t).$$

DÉMONSTRATION. La preuve est exactement analogue à celle de la proposition 7.16 : pour peu que les fonctions génératrices existent et convergent sur un ouvert contenant 0, pour t dans cet ouvert on trouve, à l'aide de la proposition 7.5 :

$$M_Z(t) = \mathbb{E} [e^{t(X+Y)}] = \mathbb{E} [e^{tX} e^{tY}] = \mathbb{E} [e^{tX}] \mathbb{E} [e^{tY}] = M_X(t)M_Y(t).$$

\square

EXEMPLE 7.32. *i.* On a déjà montré que les sommes de binomiales sont des binomiales, et que les sommes de Poisson sont des Poisson, à l'aide de la fonction génératrice des probabilités. Bien sûr, on pourrait refaire ces arguments à l'aide des fonctions génératrices des moments, mais on n'en apprendrait pas nécessairement tant que ça.

ii. Si X, Y sont deux variables aléatoires indépendantes de lois normales, respectivement de paramètres μ_X, σ_X^2 et μ_Y, σ_Y^2 , alors il est presque trivial de montrer que $Z = X + Y$ suit une loi normale de paramètres $\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2$:

$$\begin{aligned} M_Z(t) &= M_X(t)M_Y(t) \\ &= e^{\mu_X t + \mu_Y t + \frac{1}{2}(\sigma_X^2 t^2 + \sigma_Y^2 t^2)} \\ &= e^{(\mu_X + \mu_Y)t + \frac{1}{2}(\sigma_X^2 + \sigma_Y^2)t^2}. \end{aligned}$$

iii. Si X, Y sont deux variables aléatoires indépendantes de lois Gamma, respectivement de paramètres α, λ et β, λ , alors avec $Z = X + Y$, on a que

$$\begin{aligned} M_Z(t) &= M_X(t)M_Y(t) \\ &= \left(\frac{\lambda}{\lambda-t}\right)^\alpha \left(\frac{\lambda}{\lambda-t}\right)^\beta, \quad t < \lambda \\ &= \left(\frac{\lambda}{\lambda-t}\right)^{\alpha+\beta}, \quad t < \lambda. \end{aligned}$$

et Z suit une loi Gamma de paramètres $\alpha + \beta$.

7.4.3. Les fonctions génératrices jointes. On a travaillé avec des fonctions génératrices (des probabilités ou des moments) pour une seule variable. Il ne vous surprendra pas d'apprendre qu'on peut faire une description de la distribution jointe de plusieurs variables aléatoires avec des méthodes similaires.

DÉFINITION 7.6. Soit $\mathbf{X} = (X, Y)$ un vecteur aléatoire quelconque. Si $\mathbf{t} = (t, s)$, on définit la fonction génératrice jointe des moments de \mathbf{X} par

$$(7.4.26) \quad M_{\mathbf{X}}(\mathbf{t}) = M_{X,Y}(t, s) = \mathbb{E}[e^{tX} e^{sY}] = \mathbb{E}[e^{\mathbf{t} \cdot \mathbf{X}}].$$

De façon plus générale, si $\mathbf{X} = (X_1, \dots, X_n)$ est un vecteur aléatoire et $\mathbf{t} \in \mathbb{R}^n$, la fonction génératrice jointe des moments de \mathbf{X} est

$$(7.4.27) \quad M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{X}}(t_1, t_2, \dots, t_n) = \mathbb{E}[e^{t_1 X_1} e^{t_2 X_2} \cdots e^{t_n X_n}] = \mathbb{E}[e^{\mathbf{t} \cdot \mathbf{X}}].$$

Il suit directement que, si $\mathbf{X} = (X_1, X_2, \dots, X_n)$ est un vecteur aléatoire quelconque, on peut recouvrir la fonction génératrice marginale des moments de X_i en évaluant simplement $M_{\mathbf{X}}$ aux points où toutes les coordonnées sauf t_i sont nulles :

$$(7.4.28) \quad M_{X_i}(t) = M_{\mathbf{X}}(\mathbf{t})|_{t_i=t, t_j=0 \forall j \neq i}.$$

Si on se restreint au cas où $\mathbf{X} = (X, Y)$ est un vecteur à deux coordonnées seulement, par exemple, on aurait que

$$M_X(t) = M_{X,Y}(t, 0); \quad M_Y(s) = M_{X,Y}(0, s).$$

On a toujours, bien sûr, la propriété importante que la fonction génératrice des moments caractérise complètement une distribution :

PROPOSITION 7.21. Soient \mathbf{X} et \mathbf{Y} des vecteurs aléatoires avec fonctions génératrices jointes des moments respectives $M_{\mathbf{X}}$ et $M_{\mathbf{Y}}$ qui existent et sont égales sur une petite boule ouverte centrée en $0 \in \mathbb{R}^n$.

Alors, \mathbf{X} et \mathbf{Y} sont équivalents en loi – c'est à dire qu'ils ont la même distribution jointe.

DÉMONSTRATION. Encore une fois, la preuve de cette proposition excède largement le cadre de ce cours. \square

Au chapitre 6, nous avons vu plusieurs conditions équivalentes pour montrer l'indépendance d'une famille de variables aléatoires (propositions 6.1 à 6.4). On peut maintenant y ajouter un nouveau critère, équivalent à tous les autres :

PROPOSITION 7.22. *Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire dont les fonctions génératrices jointes et marginales des moments convergent respectivement sur des ouverts autour de 0. \mathbf{X} est une famille de variables indépendantes si et seulement si*

$$(7.4.29) \quad M_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^n M_{X_i}(t_i) = M_{X_1}(t_1)M_{X_2}(t_2) \cdots M_{X_n}(t_n).$$

DÉMONSTRATION. Si \mathbf{X} est une famille de variables aléatoires indépendantes, alors clairement (7.4.29) est vraie par la proposition 7.5.

Dans l'autre sens, si \mathbf{X} a une fonction génératrice des moments jointes comme à l'équation (7.4.29), alors \mathbf{X} a la même fonction génératrice des moments qu'un vecteur indépendant \mathbf{Y} dont chaque coordonnée Y_i a respectivement M_{X_i} pour fonction génératrice marginale des moments.

Mais comme \mathbf{X} et \mathbf{Y} ont la même fonction génératrice jointe des moments, alors \mathbf{X} et \mathbf{Y} ont la même distribution jointe. Alors, les coordonnées de \mathbf{X} forment une famille de variables aléatoires indépendantes. \square

EXEMPLE 7.33. Soient X, Y deux variables aléatoires indépendantes, chacune respectivement de loi normale centrée réduite.

Montrer que $X + Y$ et $X - Y$ sont indépendantes.

SOLUTION. Mon sait que la fonction génératrice jointe des moments de X, Y est

$$M_{X,Y}(t, s) = e^{\frac{1}{2}(t^2+s^2)}.$$

On sait que la fonction génératrice des moments de $X + Y$ et $X - Y$ sont données par

$$\begin{aligned} M_{X+Y}(t) &= \mathbb{E}[e^{t(X+Y)}] = M_{X,Y}(t, t) = e^{t^2}; \\ M_{X-Y}(t) &= \mathbb{E}[e^{t(X-Y)}] = M_{X,Y}(t, -t) = e^{t^2}. \end{aligned}$$

Si $U = X + Y$ et $V = X - Y$, alors

$$\begin{aligned} M_{U,V}(t, s) &= \mathbb{E}[e^{tU+sV}] \\ &= \mathbb{E}[e^{tX+tY+sX-sY}] \\ &= \mathbb{E}[e^{(t+s)X+(t-s)Y}] \\ &= M_X(t+s)M_Y(t-s) \\ &= e^{\frac{1}{2}((t+s)^2+(t-s)^2)} \\ &= e^{t^2+s^2} \\ &= M_U(t)M_V(s). \end{aligned}$$

7.5. Vecteurs gaussiens

Nous allons clore ce chapitre en discutant un exemple spécifique très important de l'usage des fonctions génératrices des moments : les vecteurs gaussiens.

DÉFINITION 7.7. Soit $\mathbf{X} = (X_1, X_2, X_3, \dots, X_n)$ un vecteur aléatoire. \mathbf{X} est un **vecteur gaussien** si et seulement si il existe un vecteur $\mathbf{Z} = (Z_1, Z_2, Z_3, \dots, Z_m)$ de variables aléatoires indépendantes, chacune de loi normale centrée réduite, une matrice $A \in \mathbb{R}^{n \times m}$ et un vecteur $\mathbf{m} = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ tels que

$$(7.5.1) \quad \mathbf{X} = A \mathbf{Z} + \boldsymbol{\mu}$$

Autrement dit, on a

$$\begin{aligned} X_1 &= a_{11}Z_1 + a_{12}Z_2 + \cdots + a_{1m}Z_m + \mu_1 \\ X_2 &= a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{2m}Z_m + \mu_2 \\ &\vdots \\ X_n &= a_{n1}Z_1 + a_{n2}Z_2 + \cdots + a_{nm}Z_m + \mu_n. \end{aligned}$$

Évidemment, par définition, si \mathbf{X} est un vecteur gaussien, X_i est une variable aléatoire de loi normale d'espérance μ_i .

La covariance de X_i et $X_{i'}$ est

$$\begin{aligned} \text{Cov}[X_i, X_{i'}] &= \sum_{j=1}^m \sum_{j'=1}^m a_{ij}a_{i'j'} \text{Cov}[Z_j, Z_{j'}] \\ &= \sum_{j=1}^n a_{ij}a_{i'j} \\ (7.5.2) \quad &= (A A^T)_{ii'}. \end{aligned}$$

On introduit la **matrice des covariances** :

$$(7.5.3) \quad \Sigma = (\text{Cov}[X_i, X_{i'}])_{1 \leq i, i' \leq n} = A A^T \in \mathbb{R}^{n \times n}.$$

On note aussi

$$\Sigma_{ij} = \text{Cov}[X_i, X_j].$$

Clairement, Σ est une matrice carrée, symétrique, définie positive.

On cherche maintenant la fonction génératrice des moments pour le vecteur gaussien \mathbf{X} :

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= \mathbb{E}[e^{\mathbf{t} \cdot \mathbf{X}}] \\ &= e^{\mathbf{t} \cdot \mathbf{m}} \mathbb{E}[e^{\mathbf{t} \cdot A \mathbf{Z}}] \end{aligned}$$

Quelle est la loi de $\mathbf{t} \cdot A \mathbf{Z}$? On a bien sûr que

$$\begin{aligned} \mathbf{t} \cdot A \mathbf{Z} &= \sum_{i=1}^n \left(t_i \sum_{j=1}^m a_{ij} Z_j \right) \\ &= \sum_{i=1}^n \sum_{j=1}^m t_i a_{ij} Z_j. \end{aligned}$$

Donc, $\mathbf{t} \cdot A \mathbf{Z}$ est une variable aléatoire de loi normale, puisque c'est la somme de variables aléatoires de loi normale indépendantes. Son espérance est nulle, puisque les Z_j sont de loi normale centrée.

La variance de $\mathbf{t} \cdot A \mathbf{Z}$ est donnée par

$$\begin{aligned}
\text{Var} [\mathbf{t} \cdot A \mathbf{Z}] &= \text{Cov} [\mathbf{t} \cdot A \mathbf{Z}, \mathbf{t} \cdot A \mathbf{Z}] \\
&= \text{Cov} \left[\sum_{i=1}^n \sum_{j=1}^m t_i a_{ij} Z_j, \sum_{i'=1}^n \sum_{j'=1}^m t_{i'} a_{i'j'} Z_{j'} \right] \\
&= \sum_{i=1}^n \sum_{i'=1}^n \sum_{j=1}^m \sum_{j'=1}^m t_i t_{i'} a_{ij} a_{i'j'} \text{Cov} [Z_j, Z_{j'}] \\
&= \sum_{i=1}^n \sum_{i'=1}^n \sum_{j=1}^m t_i t_{i'} a_{ij} a_{i'j} \\
&= \sum_{i=1}^n \sum_{i'=1}^n \left(\sum_{j=1}^m a_{ij} a_{i'j} \right) \\
&= \sum_{i=1}^n \sum_{i'=1}^n t_i t_{i'} \Sigma_{ii'} \\
&= \mathbf{t} \cdot \Sigma \mathbf{t}
\end{aligned}$$

Finalement, on a donc que

$$\mathbb{E} [e^{\mathbf{t} \cdot A \mathbf{Z}}] = e^{\frac{1}{2} \mathbf{t} \cdot \Sigma \mathbf{t}},$$

et

$$(7.5.4) \quad M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t} \cdot \mathbf{m} + \frac{1}{2} \mathbf{t} \cdot \Sigma \mathbf{t}},$$

ou, avec une notation plus explicite :

$$(7.5.5) \quad M_{\mathbf{X}}(\mathbf{t}) = \exp \left(\sum_{i=1}^n t_i \mu_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n t_i \Sigma_{ij} t_j \right).$$

(On se souvient que $\Sigma_{ij} = \text{Cov}[X_i, X_j]$).

La conséquence de cette trouvaille, c'est que si \mathbf{X} est un vecteur gaussien, alors sa distribution est *entièrement déterminée* par \mathbf{m} et Σ – soit le vecteur des espérances des X_i , et la matrice de covariance.

Dans ce contexte, *et dans ce contexte uniquement*, il sera suffisant de montrer que $\text{Cov}[X_i, X_j] = 0$ pour prouver que X_i et X_j sont indépendantes !

REMARQUE. *Attention !* Il ne suffit pas que X et Y soient deux variables aléatoires de loi normale pour conclure qu'elles sont indépendantes si $\text{Cov}[X, Y] = 0$. Il faut aussi démontrer qu'elles sont deux coordonnées d'un vecteur gaussien, ce qui n'est pas forcément le cas. Voir l'exercice ??.

EXEMPLE 7.34. On va reprendre notre exemple de statistiques. Supposons que $X_1, X_2, X_3, \dots, X_n$ sont des variables aléatoires indépendantes et identiquement distribuées. Cette fois-ci, on va supposer que les X_i sont de loi normale avec espérance μ et variance σ^2 .

On introduit

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i; \quad S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Les variables $(\bar{X}_n, X_1, X_2, \dots, X_n)$ forment un vecteur gaussien. Pour le voir, on constate que si on prend $Z_i = (X_i - \mu)/\sigma$, alors Z_i suit une loi normale centrée réduite, les Z_i forment une famille indépendante (puisque les X_i le sont), et on choisit alors $\mathbf{m} = (\mu, \mu, \dots, \mu) \in \mathbb{R}^{n+1}$.

On prend

$$A = \sigma \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(n+1) \times n}.$$

On obtient alors

$$\Sigma = A A^T = \frac{\sigma^2}{n} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & n & 0 & \cdots & 0 \\ 1 & 0 & n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & n \end{pmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}.$$

Puisque X_i et \bar{X}_n sont deux composantes d'un vecteur gaussien, alors toute combinaison linéaires de ces deux variables sont aussi des composantes d'un autre vecteur gaussien. Donc, vu que $\text{Cov}[X_i - \bar{X}_n, \bar{X}_n] = 0$, alors on a que \bar{X}_n et $X_i - \bar{X}_n$ sont indépendantes. En particulier, \bar{X}_n est indépendant de S_n^2 .

On regarde maintenant S_n^2 de plus près :

$$\begin{aligned} S_n^2 &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= \sum_{i=1}^n [(X_i - \mu)^2 + (\bar{X}_n - \mu)^2 - 2(X_i - \mu)(\bar{X}_n - \mu)] \\ &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X}_n - \mu)^2 - 2(\bar{X}_n - \mu) \sum_{i=1}^n (X_i - \mu) \\ &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X}_n - \mu)^2 - 2(\bar{X}_n - \mu) \cdot n(\bar{X}_n - \mu) \\ &= \sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\bar{X}_n - \mu)^2 - 2n(\bar{X}_n - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu). \end{aligned}$$

Donc,

$$\frac{S_n^2 + n(\bar{X}_n - \mu)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2.$$

On sait que $n(\bar{X} - \mu)^2/\sigma^2$ suit une loi χ_1^2 . On sait aussi que $\sum_{i=1}^n ((X_i - \mu)/\sigma)^2$ suit une loi χ_n^2 .

On a donc que, puisque S_n^2 et \bar{X}_n sont indépendantes, S_n^2/σ^2 doit suivre une loi χ_{n-1}^2 .

En particulier, $\mathbb{E}[S_n^2/(n-1)] = \sigma^2$, et $\text{Var}[S_n^2/(n-1)] = O(1/n)$. On a bien que $S^2/(n-1)$ est donc un bon estimateur de la variance, qui converge vers σ^2 lorsque n tend vers l'infini.

7.6. Exercices

EXERCICE 7.1. Soient X et Y respectivement le nombre de 1 et de 2 qui se produisent en n lancers d'un dé équilibré à six faces.

Trouver $\text{Cov}[X, Y]$ et $\rho[X, Y]$.

EXERCICE 7.2. Soit Z une variable aléatoire normale centrée réduite, et soit

$$X = \begin{cases} Z & \text{si } Z > 0 \\ 0 & \text{si } Z \leq 0. \end{cases}$$

Montrer que

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

EXERCICE 7.3. La densité jointe de X et Y est donnée par

$$f(x, y) = \frac{1}{y} e^{-y-(x/y)}, \quad x > 0, y > 0.$$

(a) Trouver $\mathbb{E}[X]$, $\mathbb{E}[Y]$ et $\text{Cov}[X, Y]$ par calcul direct.

On va maintenant voir que c'est beaucoup plus facile avec les espérances conditionnelles :

- (b) Trouver la densité marginale de Y . Quelle est la loi marginale de Y ? Donner $\mathbb{E}[Y]$.
- (c) Trouver la densité conditionnelle de X sachant Y . Décrire la loi de X sachant Y .
- (d) Donner $\mathbb{E}[X | Y]$. Déduire $\mathbb{E}[X]$.
- (e) Donner $\mathbb{E}[X^2 | Y]$. Trouver $\text{Var}[X]$.

EXERCICE 7.4. Soient X, Y, Z, W quatre variables aléatoires deux-à-deux non-corréllées, toutes avec espérance 0 et variance 1. Calculer les corrélations des couples de variables aléatoires suivantes :

- (a) $X + Y, Y + Z$.
- (b) $X + Y, Z + W$.

EXERCICE 7.5. Le nombre de personnes qui embarquent dans un ascenseur au rez-de-chaussée est une variable aléatoire de loi de Poisson avec paramètre $\lambda > 0$. Si il y a N étages au-dessus du rez-de-chaussée et si chaque personne descend à un étage choisi équiprobablement et indépendamment des autres, calculer l'espérance du nombre d'arrêts que l'ascenseur va faire pour décharger tout le monde.

EXERCICE 7.6. Une pièce de monnaie ayant probabilité p de tomber sur « pile » est lancée de façon répétée jusqu'à ce que les deux côtés soient apparus au moins une fois.

- (a) Trouver l'espérance du nombre de lancers nécessaires.

- (b) Trouver la probabilité que le dernier lancer soit tombé sur « face ».

EXERCICE 7.7. Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires continues indépendantes et identiquement distribuées. On dit que X_j est une valeur record si $X_j \geq X_i$ pour tout $i \leq j$.

Montrer que

- (a) l'espérance du nombre de valeurs-record est $\sum_{j=1}^n \frac{1}{j}$;
- (b) la variance du nombre de valeurs-record est $\sum_{j=1}^n \frac{j-1}{j^2}$.

EXERCICE 7.8. Soit X une variable aléatoire de loi normale centrée réduite. On introduit S une variable aléatoire indépendante de X , qui vaut équiprobablement 1 ou -1 .

On définit alors $Y = SX$.

- (a) Montrer que Y suit une loi normale centrée réduite.
- (b) Est-ce que Y est indépendante de S ? Justifier.
- (c) Est-ce que Y est indépendante de X ? Justifier.
- (d) Calculer $\text{Cov}[X, Y]$.

EXERCICE 7.9. Soit X une variable aléatoire avec $\mathbb{E}[X] = 0$ et $\text{Var}[X] = 0$.

Montrer que $\mathbb{P}\{X = 0\} = 1$.

EXERCICE 7.10. Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées avec fonction de répartition marginale F . Trouver

- (a) $\mathbb{E}[\max\{X_1, \dots, X_n\}]$;
- (b) $\mathbb{E}[\min\{X_1, \dots, X_n\}]$.

Puis, faites le calcul explicitement si les variables suivent des lois uniformes sur $(0, 1)$.

EXERCICE 7.11. Soient X, Y deux variables aléatoires avec densité jointe

$$f(x, y) = \frac{2e^{-2x}}{x}, \quad 0 < y < x < +\infty.$$

En vous servant de l'espérance conditionnelle,

- (a) trouver $\mathbb{E}[X]$;
- (b) trouver $\text{Cov}[X, Y]$.

EXERCICE 7.12. Soient N, X_1, X_2, X_3, \dots des variables aléatoires indépendantes, avec N suivant une loi géométrique de paramètre p , et les X_i identiquement distribuées selon une fonction de répartition marginale F .

Soit $M = \max\{X_1, X_2, \dots, X_N\}$ le maximum des N premières valeurs.

- (a) Trouver $\mathbb{P}\{M \leq x\}$ en conditionnant sur N .

On va maintenant tenter d'user d'astuce pour trouver la même quantité, mais avec moins de calculs :

- (b) Trouver $\mathbb{P}\{M \leq x \mid N = 1\}$.
 (c) Trouver $\mathbb{P}\{M \leq x \mid N > 1\}$.
 (d) Trouver $\mathbb{P}\{M \leq x\}$ avec la formule de probabilités totale.

EXERCICE 7.13. Le nombre de tempêtes hivernales dans une bonne année est une variable aléatoire de Poisson de paramètre $\lambda = 3$, alors que dans une mauvaise année, c'est une variable de Poisson de paramètre $\lambda = 5$.

Si on sait que la probabilité qu'une année soit bonne est de $4/10$, trouver l'espérance et la variance du nombre de tempêtes hivernales dans une année quelconque.

EXERCICE 7.14. Soient X, Y deux variables aléatoires dont la densité jointe est donnée par

$$f(x, y) = \frac{1}{\sqrt{2\pi}} e^{-y} e^{-(x-y)^2/2}, \quad x \in \mathbb{R}, y > 0.$$

- (a) Instinctivement, comment décririez-vous la loi du vecteur X, Y ?¹⁰
 (b) Trouver la fonction génératrice jointe des moments de X et Y .
 (c) Trouver les fonctions génératrices marginales des moments de X et Y .

EXERCICE 7.15. Soit Z une variable aléatoire de loi normale centrée réduite. On note $\mu_n = \mathbb{E}[Z^n]$.

Montrer que

$$\mu_n = \begin{cases} 0 & \text{si } n \text{ est impair} \\ \frac{(2j)!}{2^j j!} & \text{si } n = 2j. \end{cases}$$

EXERCICE 7.16. Soit X une variable aléatoire continue non-négative avec fonction de répartition F et fonction de répartition complémentaire $\bar{F} = 1 - F$. On assume de plus que $\lim_{x \rightarrow \infty} x\bar{F}(x) = 0$.

Montrer que

$$\mathbb{E}[X^n] = n \int_0^\infty x^{n-1} \bar{F}(x) dx.$$

Indice : Commencer avec l'identité

$$x^n = n \int_0^x t^{n-1} dt.$$

10. Cette question est là pour que vous réalisiez que ça vaut la peine d'essayer de comprendre un peu à quel genre de problème on a affaire avant de commencer à calculer.

EXERCICE 7.17. Soient N, X_1, X_2, X_3, \dots une famille de variables aléatoires indépendantes. Supposons que N est une variable entière non-négative avec fonction génératrice des probabilités ψ_N et supposons de plus que les X_i sont identiquement distribuées avec fonction génératrice des moments M_X .

On définit

$$S = \sum_{i=1}^N X_i.$$

(a) Montrer que

$$\mathbb{E}[e^{tS} \mid N] = M_X(t)^N.$$

(b) Déduire que la fonction génératrice des moments de S est

$$M_S(t) = \psi_N(M_X(t)) = M_N(\log(M_X(t))).$$

Supposons maintenant que les X_i sont entières et non-négatives, et que leur fonction génératrice des probabilités marginale soit ψ_X .

(c) Montrer que dans ce cas, la fonction génératrice des probabilités de S est

$$\psi_S = \psi_N \circ \psi_X.$$

(d) (*Processus branchant*) On veut modéliser la croissance d'une colonie de bactéries. À chaque nouvelle génération, chaque bactérie se sépare en un nombre aléatoire de bactéries indépendant et identiquement distribué. Supposons que la fonction génératrice des probabilités pour ce nombre de bactéries est ψ .

On va noter Z_i le nombre de bactéries de la génération i . Par exemple, on commence avec $Z_0 = 1$. La fonction génératrice des probabilités de Z_1 est ψ , puisqu'il n'y a qu'une bactérie qui a pu se multiplier.

Montrer que la fonction génératrice de Z_n est

$$\psi_{Z_n}(s) = \underbrace{\psi \circ \psi \circ \psi \circ \cdots \circ \psi}_{n \text{ fois}}(s) = \underbrace{\psi(\psi(\psi(\cdots \psi(s) \cdots)))}_{n \text{ fois}}.$$

REMARQUE. Cet exercice ouvre la porte à la théorie des processus branchants, qui est abordée dans le cours MAT2717 – Processus stochastiques.

EXERCICE 7.18. Montrer le corollaire 7.1 :

COROLLAIRE. Soient X_1, \dots, X_n des variables aléatoires, et soient a_1, \dots, a_n des constantes réelles.

Alors,

$$(7.1.3) \quad \mathbb{E} \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i \mathbb{E}[X_i].$$

EXERCICE 7.19. Montrer le corollaire 7.3 :

COROLLAIRE. Soient $X_1, \dots, X_n, Y_1, \dots, Y_m$ des variables aléatoires, et soient $a_1, \dots, a_n, b_1, \dots, b_m$ des constantes réelles.

Alors,

$$(7.1.14) \quad \text{Cov} \left[\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j \right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}[X_i, Y_j].$$

EXERCICE 7.20. Soient X_1, X_2, \dots, X_n des variables aléatoires. Montrer l'équation (7.1.15) :

$$(7.1.15) \quad \text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i] + 2 \sum_{j=2}^n \sum_{i=1}^{j-1} \text{Cov}[X_i, X_j].$$

EXERCICE 7.21. Montrer le corollaire 7.4 :

COROLLAIRE. Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires indépendantes avec fonctions génératrices des probabilités respectives $\psi_{X_1}, \psi_{X_2}, \psi_{X_3}, \dots, \psi_{X_n}$, alors avec $Z = \sum_{i=1}^n X_i$ et sa fonction génératrice des probabilités ψ_Z , on a

$$(7.4.8) \quad \psi_Z = \prod_{i=1}^n \psi_{X_i}.$$

EXERCICE 7.22. Soient $X_1, X_2, X_3, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées avec espérance μ et variance σ^2 .

On définit

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i; \quad S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

À l'exemple 7.2, on a montré que

$$\mathbb{E}[\bar{X}_n] = \mu; \quad \lim_{n \rightarrow \infty} \text{Var}[\bar{X}_n] = 0.$$

On a aussi montré que

$$\mathbb{E} \left[\frac{S_n^2}{n-1} \right] = \sigma^2.$$

(a) Montrer qu'il existe une constante C_1 telle que

$$\text{Var}[(X_n - \bar{X}_n)^2] \leq C_1$$

pour tout n .

(b) Montrer qu'il existe une constante C_2 telle que pour tout $i \neq j$, on a que

$$\text{Cov}[(X_i - \bar{X}_n)^2, (X_j - \bar{X}_n)^2] \leq \frac{C_2}{n}$$

pour tout n .

(c) Déduire qu'il existe une constante C telle que

$$\sum_{i=1}^n \text{Var} [(X_i - \bar{X}_n)^2] + 2 \sum_{j=2}^n \sum_{i=1}^{j-1} \text{Cov} [(X_i - \bar{X}_n)^2, X_j - \bar{X}_n]^2 \leq Cn.$$

(d) Déduire que

$$\lim_{n \rightarrow \infty} \text{Var} \left[\frac{S_n^2}{n-1} \right] = 0.$$

EXERCICE 7.23 (Les parallèles). *Cet exercice s'adresse aux plus motivé.e.s d'entre vous. Il requiert des notions de mathématiques avancées que nous n'avons pas forcément couvert en classe ; ce sont des notions que vous retrouverez dans les cours Algèbre Linéaire et/ou Algèbre I, mais beaucoup plus rigoureusement dans le cours gradué Analyse Fonctionnelle I. Bon plaisir !*

On note $L_0^2(\Omega, \mathcal{E}, \mathbb{P})$ l'ensemble des variables aléatoires $X : \Omega \rightarrow \mathbb{R}$ telles que

- i. $\mathbb{E}[X] = 0$.
- ii. $\mathbb{E}[|X|^2] < +\infty$.

- (a) Montrer que $L_0^2(\Omega, \mathcal{E}, \mathbb{P})$ est un espace vectoriel.
- (b) Montrer que $\text{Cov}[\cdot, \cdot]$ est un produit scalaire.
- (c) Montrer que $\sqrt{\text{Var}[\cdot]}$ est une norme.
- (d) Montrer que $\sqrt{\text{Var}[X - Y]}$ est une distance entre X et Y .
- (e) Soit Y une variable aléatoire dans $L_0^2(\Omega, \mathcal{E}, \mathbb{P})$. On considère l'ensemble des variables aléatoires

$$\sigma(Y) = \{X \in L_0^2(\Omega, \mathcal{E}, \mathbb{P}) : \exists g : \mathbb{R} \rightarrow \mathbb{R}, g(Y) = X\}.$$

Il s'agit de l'ensemble des variables aléatoires qui sont déterminées entièrement par Y .

Montrer que $\sigma(Y)$ est un sous-espace vectoriel de $L_0^2(\Omega, \mathcal{E}, \mathbb{P})$.

- (f) Soit $X \in L_0^2(\Omega, \mathcal{E}, \mathbb{P})$. Montrer que $\mathbb{E}[X | Y] \in \sigma(Y)$.
- (g) Montrer que $\mathbb{E}[X | Y]$ est la projection orthogonale de X sur l'espace $\sigma(Y)$.
- (h) On considère maintenant $L^2(\Omega, \mathcal{E}, \mathbb{P})$ l'ensemble des variables aléatoires $X : \Omega \rightarrow \mathbb{R}$ telles que $\mathbb{E}[|X|^2] < +\infty$.
- Montrer que $L^2(\Omega, \mathcal{E}, \mathbb{P})$ est un espace vectoriel. Montrer que $L_0^2(\Omega, \mathcal{E}, \mathbb{P})$ est un sous-espace de $L^2(\Omega, \mathcal{E}, \mathbb{P})$.
- (i) On considère maintenant l'espérance $\mathbb{E} : L^2(\Omega, \mathcal{E}, \mathbb{P}) \rightarrow \mathbb{R}$. Montrer que \mathbb{E} est un homomorphisme de groupes entre $(L^2(\Omega, \mathcal{E}, \mathbb{P}), +)$ et $(\mathbb{R}, +)$.
- (j) Montrer que $\text{Ker } \mathbb{E} = L_0^2(\Omega, \mathcal{E}, \mathbb{P})$.
- (k) Montrer que $\text{Im } \mathbb{E} = \mathbb{R}$.
- (l) Déduire que $L^2(\Omega, \mathcal{E}, \mathbb{P}) / L_0^2(\Omega, \mathcal{E}, \mathbb{P}) \cong \mathbb{R}$ au sens des groupes (ou au sens des espaces vectoriels).
- (m) Soit $\mathbb{F} : L^2(\Omega, \mathcal{E}, \mathbb{P}) \rightarrow \mathbb{R}$ une fonctionnelle linéaire continue de $L^2(\Omega, \mathcal{E}, \mathbb{P})$. Montrer qu'il existe $X_{\mathbb{F}} \in L^2(\Omega, \mathcal{E}, \mathbb{P})$ tel que pour tout Y ,

$$\mathbb{F}[Y] = \mathbb{E}[X_{\mathbb{F}}Y].$$

REMARQUE. L'objectif de cet exercice n'est pas tellement que vous soyez capables de le résoudre (pas tout de suite). L'objectif, c'est surtout d'illustrer que les notions théoriques qu'on définit maintenant sont la point d'icebergs très vastes. Les *sous-branches* des mathématiques sont beaucoup plus inter-connectées qu'on ne peut le voir au premier coup d'oeil, mais c'est bien d'entrevoir des connexions dès maintenant.

Chapitre 8

Théorèmes-limites

Nous y voici enfin ! Nous sommes aux portes de la grandeur, sur le seuil du sublime. Fondcombe est en vue, et le gué du Bruinen n'est plus très loin !

Dans les quelques premières sections, nous explorerons d'abord des inégalités pratiques qui nous permettront de trouver des bornes utiles sur les queues de diverses variables aléatoires. Nous découvrirons ensuite comment on fait pour parler de convergence de suites de variables aléatoires.

Nous terminerons le chapitre (et la session) en discutant de trois théorèmes absolument *fon - da - men - taux* en théorie des probabilités ; chacun de ces trois théorèmes se verra dédier sa section.

8.1. Bornes et inégalités

On a vu au chapitre 5 qu'il existe un lien important entre $\mathbb{E}[X]$ et $\bar{F}(x) = \mathbb{P}\{X > x\}$ lorsque X est une variable aléatoire positive. Pour rappel, si X est une variable aléatoire non-négative :

$$(5.2.2) \quad \mathbb{E}[X] = \int_0^\infty \bar{F}(t)dt.$$

Ce qu'on va montrer maintenant, c'est en quelque sorte l'envers de ce résultat.

PROPOSITION 8.1 (Inégalité de Tchebychev-Markov). *Soit X une variable aléatoire non-négative.*

Alors, pour tout $x > 0$,

$$(8.1.1) \quad \mathbb{P}\{X \geq x\} \leq \frac{\mathbb{E}[X]}{x}.$$

DÉMONSTRATION. On remarque que

$$\mathbb{1}_{X \geq x} \leq \frac{X}{x};$$

en effet,

- soit $0 \leq X < x$ auquel cas, $\mathbb{1}_{X \geq x} = 0$ et $0 \leq \frac{X}{x}$;
- soit $X \geq x$, auquel cas $\mathbb{1}_{X \geq x} = 1$, et $1 \leq \frac{X}{x}$.

On complète la preuve en prenant l'espérance de part et d'autre :

$$\frac{\mathbb{E}[X]}{x} \geq \mathbb{E}[\mathbb{1}_{X \geq x}] = \mathbb{P}\{X \geq x\}.$$

□

Ce résultat est connu sous le nom de Tchebychev-Markov, et il a plusieurs conséquences, certaines plus surprenantes que d'autres.

COROLLAIRE 8.1. Soit X une variable aléatoire avec $\mathbb{E}[|X|^2] < +\infty$, $\mathbb{E}[X] = \mu$ et $\text{Var}[X] = \sigma^2$.

Alors, pour $x > 0$,

$$(8.1.2) \quad \mathbb{P}\{|X - \mu| > x\} \leq \frac{\sigma^2}{x^2}.$$

DÉMONSTRATION. Par la proposition 8.1, on a que

$$\mathbb{P}\{|X - \mu| > x\} = \mathbb{P}\{|X - \mu|^2 > x^2\} \leq \frac{\mathbb{E}[|X - \mu|^2]}{x^2} = \frac{\text{Var}[X]}{x^2}.$$

□

Ce corollaire nous donne une borne sur la queue de la distance entre X et $\mathbb{E}[X]$.

EXEMPLE 8.1. Supposons que X soit une variable aléatoire positive avec $\mathbb{E}[X] = 50$.

- (a) Donner une borne pour $\mathbb{P}\{X > 75\}$.
- (b) Supposons que $\text{Var}[X] = 25$. Donner une borne pour $\mathbb{P}\{40 < X < 60\}$.

SOLUTION. (a) Par la proposition 8.1,

$$\mathbb{P}\{X > 75\} \leq \frac{\mathbb{E}[X]}{75} = \frac{2}{3}.$$

- (b) Par le corollaire 8.1

$$\mathbb{P}\{40 < X < 60\} = \mathbb{P}\{|X - 50| < 10\} \leq \frac{\text{Var}[X]}{10^2} = \frac{1}{4}.$$

Comme autre exemple de l'utilité de ce corollaire, on peut citer cette proposition :

PROPOSITION 8.2. Soit X une variable aléatoire avec $\mathbb{E}[X] = \mu$ et $\text{Var}[X] = 0$. Montrer que $\mathbb{P}\{X = \mu\} = 1$.

DÉMONSTRATION. Cette preuve faisait déjà l'objet de l'exercice 7.9 ; toutefois, nous présentons ici une preuve qui fait usage de l'inégalité de Tchebychev-Markov.

On va considérer la suite d'événements $A_n = \{|X - \mu| \geq \frac{1}{n}\}$. Évidemment, $A_n \subseteq A_{n+1}$ – si $|X - \mu| \geq \frac{1}{n}$, alors puisque $\frac{1}{n} \geq \frac{1}{n+1}$, par transitivité, $|X - \mu| \geq \frac{1}{n+1}$.

Donc, la suite d'événements A_n est monotone croissante, et par continuité de la mesure de probabilité, avec $A = \bigcup_{n \in \mathbb{N}} A_n$

$$\mathbb{P}\{A\} = \lim_{n \rightarrow \infty} \mathbb{P}\{A_n\}.$$

Mais bien sûr, A est l'événement « il existe un n tel que $|X - \mu| \geq \frac{1}{n}$ ». On a donc que $A = \{|X - \mu| \neq 0\} = \{X \neq \mu\}$.

Or, par l'inégalité de Tchebychev-Markov,

$$\mathbb{P}\{A_n\} = \mathbb{P}\left\{|X - \mu| \geq \frac{1}{n}\right\} \leq n^2 \text{Var}[X] = 0.$$

Donc,

$$\mathbb{P}\{A_n\} = 0$$

pour tout n . En particulier,

$$\mathbb{P}\{A\} = \lim_{n \rightarrow \infty} \mathbb{P}\{A_n\} = 0,$$

et $\mathbb{P}\{X \neq \mu\} = 0$.

On a donc que

$$\mathbb{P}\{X = \mu\} = \mathbb{P}\{A^c\} = 1 - \mathbb{P}\{A\} = 1.$$

□

On a aussi un autre corollaire, plus général :

COROLLAIRE 8.2. *Soit X une variable aléatoire et soit $g : \mathbb{R} \rightarrow [0, +\infty)$ une fonction strictement croissante à valeurs positives. Alors,*

$$(8.1.3) \quad \mathbb{P}\{X \geq x\} \leq \frac{\mathbb{E}[g(X)]}{g(x)}.$$

En particulier, si X est une variable aléatoire positive, pour tout $x > 0$ et $\alpha > 0$, on a

$$(8.1.4) \quad \mathbb{P}\{X \geq x\} \leq \frac{\mathbb{E}[X^\alpha]}{x^\alpha}.$$

De plus, si X est une variable aléatoire quelconque avec fonction génératrice des moments $M_X(t)$ convergente pour au moins un $t > 0$, alors pour tout t où $M_X(t)$ converge,

$$(8.1.5) \quad \mathbb{P}\{X \geq x\} \leq e^{-tx} M_X(t).$$

L'équation (8.1.5) est souvent appelée la borne de Chernoff.

DÉMONSTRATION. On a bien sûr que, puisque g est monotone croissante, et par la proposition 8.1,

$$\mathbb{P}\{X \geq x\} = \mathbb{P}\{g(X) \geq g(x)\} \leq \frac{\mathbb{E}[g(X)]}{g(x)}.$$

Pour des variables positives, $g(x) = x^\alpha > 0$ est croissante lorsque $x > 0$, et alors en appliquant ce résultat, on trouve tout de suite que

$$\mathbb{P}\{X \geq x\} \leq \frac{\mathbb{E}[X^\alpha]}{x^\alpha}.$$

Par ailleurs, si on fixe $t > 0$, la fonction $g(x) = e^{tx}$ est positive et croissante. Donc,

$$\mathbb{P}\{X \geq x\} \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}} = e^{-tx} M_X(t),$$

si $M_X(t)$ converge. □

REMARQUE. On remarque qu'il suffit que $M_X(t)$ converge pour un certain $t > 0$ pour que l'on ait une décroissance exponentielle de $\mathbb{P}\{X \geq x\}$. Ça, c'est peut-être un résultat surprenant.

On remarque aussi que plus le rayon de convergence de $M_X(t)$ est grand, et plus la borne exponentielle est serrée.

EXEMPLE 8.2. Soit Z une variable aléatoire de loi normale centrée réduite. Alors, sa fonction génératrice des moments converge pour tout t , et on a que

$$M_Z(t) = e^{\frac{1}{2}t^2}.$$

L'intérêt, bien sûr, c'est que pour $x > 0$,

$$\mathbb{P}\{Z \geq x\} \leq e^{-tx} e^{\frac{1}{2}t^2},$$

et si on prend $t = x$, on se retrouve avec

$$\mathbb{P}\{Z \geq x\} \leq e^{-\frac{1}{2}x^2}.$$

EXEMPLE 8.3. Soit X une variable aléatoire de loi Poisson de paramètre $\lambda > 0$. Alors, sa fonction génératrice des moments converge pour tout t , et on a que

$$M_X(t) = e^{\lambda(e^t - 1)},$$

donc

$$\mathbb{P}\{X \geq x\} \leq e^{-tx} e^{\lambda(e^t - 1)}.$$

Si on fixe x , on peut essayer de trouver la meilleure borne possible en tentant de trouver un minimum de $e^{-tx} M_X(t)$ en dérivant par t . On a que

$$\frac{d}{dt} \left(e^{-tx} e^{\lambda(e^t - 1)} \right) = -xe^{-tx} e^{\lambda(e^t - 1)} + e^{-tx} \left(\lambda e^t e^{\lambda(e^t - 1)} \right).$$

Et en mettant la dérivée égale à 0, on trouve

$$\begin{aligned} \lambda e^{-t(x-1)} &= xe^{-tx} \\ \Rightarrow \log(\lambda) - t(x-1) &= \log(x) - tx \\ \Rightarrow \log(\lambda/x) &= -t. \end{aligned}$$

Essayons donc $e^t = \frac{x}{\lambda}$ dans la borne de Chernoff :

$$\begin{aligned} \mathbb{P}\{X \geq x\} &\leq \left(\frac{x}{\lambda}\right)^{-x} e^{\lambda(\frac{x}{\lambda} - 1)} \\ &= e^{-\lambda} \left(\frac{e}{x}\right)^x \lambda^x. \end{aligned}$$

Bien sûr, on a que $x^x \geq \Gamma(x+1)$ lorsque $x > 1$, et

$$\mathbb{P}\{X \geq x\} \leq e^{-\lambda} \frac{(e\lambda)^x}{\Gamma(x+1)}.$$

En particulier lorsque $x = k \in \mathbb{N}$,

$$\mathbb{P}\{X \geq k\} \leq e^{-\lambda} \frac{(e\lambda)^k}{k!}.$$

Bien sûr, dans bien des cas, on connaît déjà $\bar{F}(x)$ et ces bornes ne sont pas forcément utiles. Le grand intérêt de ces bornes, c'est qu'elles sont *toujours vraies*, peu importe la distribution.

8.2. Les types de convergence

Au chapitre 4, on introduit la notion de variable aléatoire. Au chapitre 6, on commence à travailler avec plusieurs variables aléatoires à la fois. Le prochain pas logique, c'est bien sûr de travailler avec *une infinité* de variables aléatoires à la fois.

Supposons qu'on définit une suite X_1, X_2, X_3, \dots de variables aléatoires. Dans quel sens pourrait-on dire que cette suite « converge » ?

Pour les suites de nombres, c'est facile ; on a une notion de limite bien définie en analyse. Les variables aléatoires X_i , par contre, ne sont pas des nombres réels. Ce sont des fonctions de Ω dans \mathbb{R} – vous vous souvenez ?

8.2.1. La convergence ponctuelle. La première façon de définir la convergence, c'est par la notion de convergence ponctuelle :

DÉFINITION 8.1 (Convergence ponctuelle). Soit X_1, X_2, X_3, \dots une suite de variables aléatoires. On dit que la suite X_i converge **ponctuellement** vers une variable aléatoire X si et seulement si pour tout $\omega \in \Omega$,

$$(8.2.1) \quad \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

En ce qui nous concerne, la convergence ponctuelle est la convergence la plus forte que nous rencontrons. On notera simplement $\lim_{n \rightarrow \infty} X_n = X$, et, lorsque ça ne sera pas ambigu, $X_n \rightarrow X$.

EXEMPLE 8.4. Supposons que les X_i sont des variables aléatoires indépendantes choisies uniformément dans l'intervalle $(0, \frac{1}{2})$. On définit $Y_i = X_i^i$. Alors, $\lim_{n \rightarrow \infty} Y_n = 0$. On le voit car $0 \leq Y_n \leq (\frac{1}{2})^n$; on obtient le résultat par le théorème de la Sandwich.

Il n'y a pas grand chose à dire à propos de la convergence ponctuelle – on ne la rencontrera pas souvent.

8.2.2. La convergence presque sûre. Le second type de convergence qu'on va rencontrer (beaucoup plus souvent), c'est la convergence presque sûre. Il s'agit d'un type de convergence pratiquement aussi fort que la convergence ponctuelle, mais beaucoup plus applicable en probabilités.

DÉFINITION 8.2 (Convergence presque sûre). Soit X_1, X_2, X_3, \dots une suite de variables aléatoires. On dit que la suite X_i converge **presque sûrement** vers une variable aléatoire X si il existe un événement $A \in \mathcal{E}$ tel que

- $\mathbb{P}\{A\} = 1$.
- Pour tout $\omega \in A$, on a

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

On notera

$$\lim_{n \rightarrow \infty} X_n = X \text{ p.s.}$$

EXEMPLE 8.5. On lance un dé équilibré à six faces de façon répétée. On note Y_i le résultat du lancer i , puis $X_n = \max\{Y_i : i \leq n\}$ le maximum observé lors des n premiers lancers.

Alors,

$$\lim_{n \rightarrow \infty} X_n = 6 \text{ p.s.}$$

En effet, si on considère l'événement A_n : « il n'y a pas eu de 6 après n lancers », alors $A_n \supseteq A_{n+1}$, et la suite A_n est décroissante. On a en outre $\mathbb{P}\{A_n\} = (\frac{5}{6})^n$.

On note $A = \bigcap_{n \in \mathbb{N}} A_n$. C'est l'événement « Il n'y a jamais eu de 6 ». Bien sûr, par continuité de la mesure de probabilités, puisque les A_n forment une suite monotone décroissante,

$$\mathbb{P}\{A\} = \lim_{n \rightarrow \infty} \mathbb{P}\{A_n\} = 0.$$

Ça signifie que l'événement A^c : « on obtient éventuellement un 6 » a probabilité 1.

Or, pour tout $\omega \in A^c$, forcément, $\lim_{n \rightarrow \infty} X_n(\omega) = 6$, puisque lorsqu'on obtient éventuellement un 6 (disons après $k(\omega)$ lancers), alors pour tout $n > k(\omega)$, on aura $X_n(\omega) = 6$.

Donc, il existe un événement de probabilité 1 sur lequel $X_n \rightarrow X$. Donc $X_n \rightarrow X$ p.s.

8.2.3. La convergence en probabilités. La convergence en probabilités est la prochaine sur notre liste. Elle est plus faible que la convergence presque sûre : si une suite de variables converge presque sûrement, alors elle converge en probabilités, mais l'inverse n'est pas forcément vrai.

DÉFINITION 8.3 (Convergence en probabilités). Soit X_1, X_2, X_3, \dots une suite de variables aléatoires. On dit que X_n converge vers une variable aléatoire X en probabilités si et seulement si pour tout $\epsilon > 0$, on a que

$$(8.2.2) \quad \lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - X| > \epsilon\} = 0.$$

La différence est subtile : ici, la suite X_n converge vers X si les probabilités que X_n et X soient « différentes » (d'au moins $\epsilon > 0$) convergent vers 0.

EXEMPLE 8.6. Soit U une variable aléatoire uniforme sur l'intervalle $(0, 1)$. on choisit r un nombre irrationnel positif quelconque. Dans ce qui suit, on note

$$\langle t \rangle := t - \lfloor t \rfloor$$

la partie fractionnelle de t .

On considère la suite de points $x_n = \langle nr \rangle$. Cette suite de points est dense dans $(0, 1)$.¹

On définit maintenant $X_i = \mathbb{1}_{(x_i - \frac{1}{i}, x_i + \frac{1}{i})}(U)$.

Alors, pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\{|X_n - 0| > \epsilon\} = \lim_{n \rightarrow \infty} \mathbb{P}\left\{U \in \left(x_n - \frac{1}{n}, x_n + \frac{1}{n}\right)\right\} \leq \lim_{n \rightarrow \infty} \frac{2}{n} = 0.$$

On conclue donc que X_n converge vers 0 en probabilité. Par contre, $\limsup_{n \rightarrow \infty} X_n(\omega) = 1$ et $\liminf_{n \rightarrow \infty} X_n(\omega) = 0$ pour tout $\omega \in \Omega$. Donc, cette suite ne converge certainement pas presque sûrement !

L'exemple le plus fameux d'un résultat de convergence en probabilités est la loi faible des grands nombres, que nous verrons dans la section 8.3.

8.2.4. La convergence en loi. Parfois appelé **convergence faible**, la convergence en loi est sans doute le plus faible des types de convergence que l'on va rencontrer. Il a toutefois de précieux usages, notamment en sciences et en statistiques.

DÉFINITION 8.4. Soit X_1, X_2, X_3, \dots une suite de variables aléatoires. On dit que la suite des X_n converge en loi vers une variable aléatoire X si on a que pour tout ouvert² $A \subseteq \mathbb{R}$,

$$(8.2.3) \quad \lim_{n \rightarrow \infty} \mathbb{P}\{X_n \in A\} = \mathbb{P}\{X \in A\}.$$

Ici, la variable aléatoire X ne sert qu'à spécifier la « loi-limite » ; la suite des variables aléatoires ne converge pas forcément vers les valeurs que prendraient la variable X , mais sa distribution oui.

EXEMPLE 8.7. On a déjà vu, très tôt, au chapitre 4, un exemple de convergence en loi : la proposition 4.6 sur l'approximation de la loi de Poisson par des variables binomiales.

On a également vu, au chapitre 5, le théorème de DeMoivre-Laplace (proposition 5.6) : un autre exemple de convergence en loi.

1. La preuve de cette affirmation est hors-cadre, mais amusante.

2. Pour tout A mesurable, en fait, mais on ne s'en fait pas...

Évidemment il existe plusieurs façons de vérifier la convergence en loi ; la proposition suivante (que nous accepterons sans preuve) nous donne quelques méthodes pour la vérifier.

PROPOSITION 8.3. *Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires. Les énoncés suivants sont équivalents :*

- i. X_n converge en loi vers X .
- ii. Pour tout x , $\lim_{n \rightarrow \infty} \mathbb{P}\{X_n \leq x\} = \mathbb{P}\{X \leq x\}$.
- iii. Si les fonctions génératrices des moments des X_n et de X sont respectivement M_n et M , et que $\lim_{n \rightarrow \infty} M_n(t) = M(t)$ pour tout t où $M_n(t)$ et $M(t)$ convergent.

En particulier, dans le cas des variables aléatoires discrètes, la condition ii est équivalente à la convergence des fonctions de masse ; dans le cas des variables aléatoires continues, la condition ii est équivalente à la convergence des densités.

8.2.5. La convergence dans L^p . *Cette section n'est pas nécessaire et comporte une discussion plus avancée. Pas de stress, donc...*

Le dernier type de convergence est surtout un outil hérité de l'analyse fonctionnelle. Supposons qu'on considère l'espace

$$L^p(\Omega, \mathcal{E}, \mathbb{P}) = \{X : \Omega \rightarrow \mathbb{R} \mid \mathbb{E}[|X|^p] < +\infty\}.$$

Cet ensemble est un espace vectoriel sur le corps \mathbb{R} (ou même \mathbb{C}) ; ses vecteurs sont les variables aléatoires et les scalaires sont les nombres réels.

On peut définir la **norme L^p** par

$$(8.2.4) \quad \|X\|_p = \mathbb{E}[|X|^p]^{1/p}.$$

Typiquement, le paramètre p est entre $[1, +\infty)$.

Ordre des espaces L^p . L'inégalité de Jensen garantit que pour tout $p' > p$, on a que

$$(8.2.5) \quad \begin{aligned} \|X\|_{p'} &= \mathbb{E}[|X|^{p'}]^{1/p'} \\ &= \mathbb{E}[(|X|^p)^{p'/p}]^{1/p'} \\ &\geq \left(\mathbb{E}[|X|^p]^{p'/p}\right)^{1/p'} \\ &= \mathbb{E}[|X|^p]^{1/p} \\ &= \|X\|_p. \end{aligned}$$

On conclue donc que

$$L^p(\Omega, \mathcal{E}, \mathbb{P}) \subseteq L^{p'}(\Omega, \mathcal{E}, \mathbb{P}),$$

pour tout $p' > p$.

Convergence dans L^p .

DÉFINITION 8.5. Soient X_1, X_2, \dots des variables aléatoires toutes dans $L^p(\Omega, \mathcal{E}, \mathbb{P})$. On dit que les X_n convergent dans L^p vers une variable aléatoire $X \in L^p(\Omega, \mathcal{E}, \mathbb{P})$ si et seulement si

$$(8.2.6) \quad \lim_{n \rightarrow \infty} \|X_n - X\|_p = 0.$$

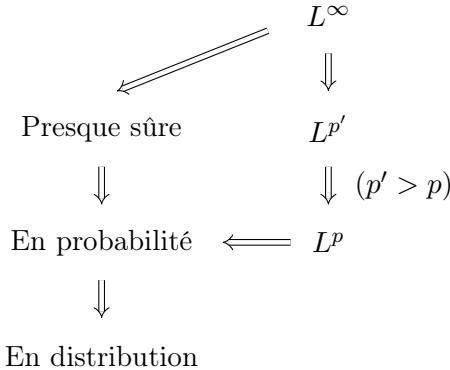


FIGURE 8.1. La hiérarchie des types de convergence.

Cette définition est équivalente à

$$\lim_{n \rightarrow \infty} \mathbb{E}[|X_n - X|^p] = 0.$$

On voit par l'équation (8.2.5) que si X_n converge vers X dans $L^{p'}$, alors pour $p < p'$ on a aussi que X_n converge vers X dans L^p .

EXEMPLE 8.8. Dans l'exemple 8.6, les X_n convergent dans L^p vers 0 pour tout p . Pourtant, on n'a pas la convergence presque sûre.

L'espace L^∞ . On définit l'espace

$$L^\infty = \{X : \Omega \rightarrow \mathbb{R} \mid \exists M > 0 : \mathbb{P}\{|X| \leq M\} = 1\}.$$

La norme sur $L^\infty(\Omega, \mathcal{E}, \mathbb{P})$ est

$$(8.2.7) \quad \|X\|_\infty = \inf\{M > 0 : \mathbb{P}\{X \leq M\} = 1\}.$$

Moralement, $L^\infty(\Omega, \mathcal{E}, \mathbb{P})$ est l'ensemble des variables aléatoires « bornées sur Ω »³. La norme L^∞ est simplement le supréumum essentiel de $|X|$.

On a que

$$L^\infty(\Omega, \mathcal{E}, \mathbb{P}) \subseteq L^p(\Omega, \mathcal{E}, \mathbb{P})$$

pour tout $p \geq 1$.

En particulier, si $X \in L^\infty(\Omega, \mathcal{E}, \mathbb{P})$, alors on a que

$$(8.2.8) \quad \lim_{p \rightarrow \infty} \|X\|_p = \|X\|_\infty.$$

La convergence dans $L^\infty(\Omega, \mathcal{E}, \mathbb{P})$ est simplement la convergence uniforme de fonctions.

REMARQUE. *Attention!* On ne peut pas conclure que $X \in L^\infty(\Omega, \mathcal{E}, \mathbb{P})$ si $X \in L^p(\Omega, \mathcal{E}, \mathbb{P})$ pour tout p – par exemple, une variable aléatoire de loi normale n'est pas dans L^∞ car elle n'est pas essentiellement bornée, mais elle est dans $L^p(\Omega, \mathcal{E}, \mathbb{P})$ pour tout p .

8.2.6. Hiérarchie des convergences. Bien que dans plusieurs circonstances, d'autres liens existent entre les différents types de convergence. La figure 8.1 présente toutes les implications qui sont vraies sans hypothèse supplémentaire.

3. Essentiellement bornées ; c'est à dire qu'il existe un événement avec probabilité 1 sur lequel la fonction est bornée.

Évidemment, il peut s'ajouter, dans certains cas, d'autres implications. Par exemple, la convergence en loi vers une constante implique la convergence en probabilité vers une constante...

8.3. La loi faible des grands nombres

Nous y sommes enfin ! Après toutes les péripéties, nous voici maintenant à montrer notre premier théorème !

THÉORÈME 8.1 (Loi faible des grands nombres). *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soit $(X_n : \Omega \rightarrow \mathbb{R})_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, avec $\mathbb{E}[|X_1|] < +\infty$, et $\mu = \mathbb{E}[X_1]$. Soit $S_n = \sum_{k=1}^n X_k$ la somme des n premières variables aléatoires.*

Alors, $\frac{S_n}{n}$ converge en probabilités vers μ – c'est à dire que pour tout $\epsilon > 0$, on a que

$$(8.3.1) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right\} = 0.$$

DÉMONSTRATION. Si on suppose que les variables ont aussi variance finie, c'est très simple. En assumant que $\text{Var}[X_i] = \sigma^2 < +\infty$, on trouve, pour $\epsilon > 0$:

$$\begin{aligned} \mathbb{P} \left\{ \left| \frac{S_n - n\mu}{n} \right| > \epsilon \right\} &\leq \frac{\text{Var} \left[\frac{S_n}{n} \right]}{\epsilon^2} \\ &= \frac{n\sigma^2}{\epsilon^2 n^2}. \end{aligned}$$

Lorsque n tend vers l'infini, bien sûr, peu importe $\epsilon > 0$, on trouve donc que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{S_n}{n} - \mu \right| > \epsilon \right\} = 0,$$

soit le résultat attendu.

Il est possible de prouver que le théorème reste vrai même si les variances sont infinies. Toutefois, cela requiert l'usage d'outils analytiques tels que la fonction caractéristique, qui ne font pas parti de ce cours. \square

Ce que la loi faible des grands nombres nous dit, c'est que, dans un échantillon de données indépendantes et identiquement distribuées, « plus on a de données, et plus la probabilité que la moyenne des données s'écarte de l'espérance tend vers 0. »

On dit que la loi des « faible » parce qu'elle donne la convergence en probabilités. Toutefois, on peut montrer un résultat beaucoup plus fort, avec les mêmes hypothèses.

8.4. La loi forte des grands nombres

THÉORÈME 8.2 (Loi forte des grands nombres). *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soit $(X_n : \Omega \rightarrow \mathbb{R})_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, avec $\mathbb{E}[|X_1|] < +\infty$, et $\mu = \mathbb{E}[X_1]$. Soit $S_n = \sum_{k=1}^n X_k$ la somme des n premières variables aléatoires.*

Alors, $\frac{S_n}{n}$ converge presque sûrement vers μ – c'est à dire que, pour tout $\epsilon > 0$, on a que

$$(8.4.1) \quad \mathbb{P} \left\{ \lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu \right\} = 1$$

DÉMONSTRATION. Nous allons assumer pour cette preuve que le quatrième moment des X_i existe.

On va d'abord faire la preuve pour $\mu = 0$ et $\sigma^2 = 1$ – on s'arrangera ensuite pour prouver le résultat plus général.

On considère

$$\mathbb{E}[S_n^4] = \sum_{i,j,k,l=1}^n \mathbb{E}[X_i X_j X_k X_l].$$

Puisque les X_i sont indépendants et identiquement distribués avec $\mathbb{E}[X_i] = 0$, alors les termes dans la sommations peuvent prendre l'une des formes suivantes :

$$X_i^4, \quad X_i^3 X_j, \quad X_i^2 X_j^2, \quad X_i^2 X_j X_k, \quad X_i X_j X_k X_l,$$

selon que certains indices sont égaux ou pas.

Or, par l'indépendance, et parce que $\mathbb{E}[X_i] = 0$ pour tout i ,

$$\mathbb{E}[X_i^3 X_j] = \mathbb{E}[X_i^3] \mathbb{E}[X_j] = 0; \quad \mathbb{E}[X_i X_j X_k X_l] = \mathbb{E}[X_i] \mathbb{E}[X_j] \mathbb{E}[X_k] \mathbb{E}[X_l] = 0.$$

Donc, les seuls termes qui survivent sont les termes de la forme

$$\mathbb{E}[X_i^2 X_j^2] = \mathbb{E}[X_i^2] \mathbb{E}[X_j^2] = 1,$$

et ceux de la forme $\mathbb{E}[X_i^4]$. Il y a $\binom{n}{2}$ façons de sélectionner des coefficients i, j différents, et $\binom{4}{2} = 6$ termes de la somme pour égaux à $\mathbb{E}[X_i^2 X_j^2]$. Il y a également n termes de notre somme de la forme $\mathbb{E}[X_i^4]$.

Disons qu'on note $\kappa = \mathbb{E}[X_1^4]$. Alors,

$$\mathbb{E}[S_n^4] = n\kappa + 3n(n-1).$$

L'intérêt d'un tel résultat, c'est que, bien sûr,

$$\mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right] = \frac{n\kappa + 3n(n-1)}{n^4},$$

ce qui garantit que la série

$$\sum_{n=1}^{\infty} \mathbb{E}\left[\left(\frac{S_n}{n}\right)^4\right]$$

converge.

Mais puisque cette série est une série de termes positifs (car $S_n^4 > 0$), par la proposition 7.3, on a que

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4\right] < +\infty.$$

Mais supposons qu'il existe un événement A tel que pour tout $\omega \in A$, on avait que

$$\limsup_{n \rightarrow \infty} \frac{S_n(\omega)^4}{n^4} > 0.$$

Alors on aurait que

$$\mathbb{E}\left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4\right] = \mathbb{E}\left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4 \mid A\right] \mathbb{P}\{A\} + \mathbb{E}\left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4 \mid A^c\right] \mathbb{P}\{A^c\}.$$

Et bien sûr, pour tout $\omega \in A$, la série $\sum_{n=1}^{\infty} \frac{S_n(\omega)^4}{n^4}$ diverge, et

$$\mathbb{E} \left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n} \right)^4 \mid A \right] = +\infty.$$

Dès lors, pour que $\mathbb{E} \left[\sum_{n=1}^{\infty} \left(\frac{S_n}{n} \right)^4 \right] < +\infty$, on doit absolument avoir $\mathbb{P}\{A\} = 0$ – ou $\mathbb{P}\{A^c\} = 1$. Mais pour tout $\omega \in A^c$, on a que

$$\limsup_{n \rightarrow \infty} \frac{S_n(\omega)^4}{n^4} = 0,$$

et, puisque $S_n^4 > 0$, cela signifie que

$$\lim_{n \rightarrow \infty} \frac{S_n(\omega)^4}{n^4} = 0.$$

Mais en prenant la racine quatrième, on a que pour tout $\omega \in A^c$,

$$\lim_{n \rightarrow \infty} \frac{S_n(\omega)}{n} = 0.$$

Donc, finalement, il existe un événement $B (= A^c)$, avec probabilité 1 tel que pour tout $\omega \in B$, $\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0$, et on doit avoir que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0, \quad \text{p.s.}$$

Maintenant si on suppose μ et σ^2 des valeurs quelconques (avec $\sigma^2 > 0$), alors on définit $\tilde{X}_i = \frac{X_i - \mu}{\sigma}$. On a que $\mathbb{E}[\tilde{X}_i] = 0$, $\text{Var}[\tilde{X}_i] = 1$, et avec $\tilde{S}_n = \sum_{k=1}^n \tilde{X}_i$, on voit que

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \frac{\tilde{S}_n}{n}, \quad \text{p.s.} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sigma n} \sum_{k=1}^n (X_k - \mu), \quad \text{p.s.} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sigma} \left(\frac{1}{n} S_n - \mu \right) \end{aligned}$$

Et il suffit de multiplier des deux côtés par σ pour obtenir que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu, \quad \text{p.s.}$$

□

La différence entre la loi forte et la loi faible est souvent confondante. Ce que la loi faible nous dit, c'est que lorsque n tend vers l'infini, la probabilité que « S_n/n est proche de μ » tend vers 1. Mais on pourrait avoir que pour une infinité de $n \in \mathbb{N}$, (même si ça se produit de plus en plus infrequentement), $|S_n/n - \mu| > \epsilon$.

La loi forte, par contre, nous dit que pour un certain $\epsilon > 0$ quelconque, la « déviation » $|S_n/n - \mu|$ sera plus grande que ϵ seulement pour un nombre fini de valeurs de n .

La loi forte des grands nombres est tellement importante qu'elle a longtemps été prise pour acquise et considérée comme un axiome fondamental de la théorie des probabilités. L'interprétation fréquentiste de la théorie des probabilités repose entièrement sur elle.

Supposons qu'on réalise N fois une expérience aléatoire, et que chaque tentative est indépendante. On définit $N(A)$ le nombre de fois où l'événement A se sera réalisé.

On va noter A_i l'événement que A se réalise à la i ème expérience. Alors,

$$N(A) = \sum_{i=1}^N \mathbb{1}_{A_i}.$$

$N(A)$ est une somme de variables aléatoires indépendantes et identiquement distribuées, et par la loi forte des grands nombres, on voit que

$$\lim_{N \rightarrow \infty} \frac{N(A)}{N} = \mathbb{E}[\mathbb{1}_{A_1}], \quad \text{p.s..}$$

Mais $\mathbb{E}[\mathbb{1}_{A_1}] = \mathbb{P}\{A\}$! Donc,

$$\lim_{N \rightarrow \infty} \frac{N(A)}{N} = \mathbb{P}\{A\}, \quad \text{p.s.}$$

Ce qu'on vient de démontrer, c'est que les axiomes des probabilités que nous avons établis au chapitre 2 sont suffisants pour conclure que l'interprétation fréquentiste est cohérente.

REMARQUE. La raison pour laquelle on a seulement la convergence presque sûre est que, par exemple dans une partie de pile ou face, il est en principe *possible* d'obtenir une séquence ininterrompue de « pile » sans jamais avoir de « face » – mais cet événement a probabilité 0. Si ça se produisait, notre estimation pour la probabilité d'obtenir « pile » serait erronnée. Mais cet événement aurait probabilité 0, donc on peut l'ignorer.

Dans la prochaine (et la dernière) section, on explore plus finement le comportement de S_n – spécifiquement, on va chercher à comprendre les fluctuations de $S_n - n\mu$.

8.5. Le théorème de la limite centrale

C'est probablement le point culminant de ce cours : on va élucider le mystère de l'importance de la distribution normale. Il existe plein de distributions en cloche. Pourquoi la distribution normale est-elle celle qu'on retient ? Pourquoi pas une autre ?

Dans les sections précédentes, on a montré que S_n/n converge presque sûrement vers μ .

Une autre façon de comprendre ce résultat, c'est de dire que $n\mu$ est une bonne approximation de S_n lorsque n est grand. Cette façon de voir les choses a plein d'avantages : S_n est aléatoire, mais n et μ ne le sont pas. Donc, *a priori*, la loi des grands nombres nous permet de faire des prédictions.

La question, c'est : à quel point ces prédictions sont-elles précises ? Et c'est ici qu'entre en jeu le théorème de la limite centrale.

Rapidement, le théorème de la limite centrale nous garantit que, pour peu que la variance des termes qui composent S_n soit finie, alors les fluctuations de S_n par rapport à $n\mu$ sont – bien sûr aléatoire, mais – d'ordre \sqrt{n} .

THÉORÈME 8.3 (Théorème de la limite centrale). *Soit $(\Omega, \mathcal{E}, \mathbb{P})$ un espace de probabilités, soit $(X_n : \Omega \rightarrow \mathbb{R})_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes et identiquement distribuées, avec $\mathbb{E}[|X_1|^2] < +\infty$, $\mu = \mathbb{E}[X_1]$ et $\sigma^2 = \text{Var}[X_1]$.*

Soit $S_n = \sum_{k=1}^n X_k$ la somme des n premières variables aléatoires.

Alors, $Z_n = (S_n - n\mu)/\sqrt{n\sigma^2}$ converge en distribution vers une normale centrée réduite – c'est à dire que, pour tout $x \in \mathbb{R}$,

$$(8.5.1) \quad \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x \right\} = \Phi(x).$$

DÉMONSTRATION. Pour réaliser cette preuve, nous allons assumer que la fonction génératrice des moments existe sur un intervalle autour de 0. C'est une hypothèse très forte; pour travailler sans, il faudrait utiliser les fonctions caractéristiques, que nous n'avons pas vues et qui sont hors du cadre de ce cours.

À la proposition 8.3, nous avons vu qu'une suite de variables aléatoires converge en distribution vers une variable X si et seulement si leurs fonctions génératrices des moments convergent vers celle de X (si elles existent).

C'est cette méthode que nous allons employer.

Comme pour la preuve de la loi forte des grands nombres, nous allons assumer pour commencer que $\mathbb{E}[X_i] = 0$ pour tout i , et $\text{Var}[X_i] = \mathbb{E}[X_i^2] = 1$.

On va avoir $M_X(t) = \mathbb{E}[\exp(tX_1)]$ la fonction génératrice des moments pour les termes X_i de notre suite de variables aléatoires. Les fonctions génératrices sont toutes les mêmes, puisque les X_i sont identiquement distribuées.

Maintenant on définit

$$M_n(t) = \mathbb{E} \left[\exp \left(t \frac{S_n}{\sqrt{n}} \right) \right] = M_X \left(\frac{t}{\sqrt{n}} \right)^n.$$

Pour se faciliter la vie, on va prendre $L_X(t) = \log M_X(t)$ et $L_n(t) = \log M_n(t) = nL_X(t/\sqrt{n})$.

Ce qu'on veut montrer, c'est que $M_n(t) \rightarrow \exp(\frac{1}{2}t^2)$ lorsque $n \rightarrow \infty$; ou, alternative-
ment, que $L_n(t) = nL_X(t/\sqrt{n}) \rightarrow \frac{1}{2}t^2$.

On va obtenir la limite en raisonnant avec la règle de L'Hospital :

$$\begin{aligned} \lim_{x \rightarrow \infty} xL_X(t/\sqrt{x}) &= \lim_{x \rightarrow \infty} \frac{L_X \left(\frac{t}{\sqrt{x}} \right)}{x^{-1}} \\ &= \lim_{x \rightarrow \infty} \frac{L'_X \left(\frac{t}{\sqrt{x}} \right) \left(-\frac{1}{2}tx^{-\frac{3}{2}} \right)}{(-x^{-2})} \quad (\text{L'Hospital}) \\ &= \lim_{x \rightarrow \infty} \frac{L'_X \left(\frac{t}{\sqrt{x}} \right) t}{2x^{-\frac{1}{2}}} \\ &= \lim_{x \rightarrow \infty} \frac{L''_X \left(\frac{t}{\sqrt{x}} \right) \left(-\frac{1}{2}tx^{-\frac{3}{2}} \right) t}{2 \left(-\frac{1}{2}x^{-\frac{3}{2}} \right)} \quad (\text{L'Hospital}) \\ &= \frac{1}{2}t^2 \cdot \lim_{x \rightarrow \infty} L''_X \left(\frac{t}{\sqrt{x}} \right), \quad \text{si } \lim_{x \rightarrow \infty} L''_X(t/\sqrt{x}) \text{ existe.} \end{aligned}$$

Or, on sait que

$$L'_X(t) = \frac{M'_X(t)}{M_X(t)}; \quad L''_X(t) = \frac{M_X(t)M''_X(t) - M'_X(t)^2}{M_X(t)^2}.$$

Et ces fonctions sont continues en 0. On a que

$$M_X(0) = 1; \quad M'_X(0) = \mathbb{E}[X_1] = 0; \quad M''_X(0) = \mathbb{E}[X_1^2].$$

Il suit que

$$\lim_{x \rightarrow \infty} L''_X\left(\frac{t}{\sqrt{x}}\right) = L''_X(0) = \mathbb{E}[X_1^2] = 1.$$

On a donc finalement que

$$\lim_{n \rightarrow \infty} \log(M_n(t)) = \frac{1}{2}t^2,$$

ou carrément

$$\lim_{n \rightarrow \infty} M_n(t) = \exp\left(\frac{1}{2}t^2\right).$$

Et on a montré que si les X_i ont $\mathbb{E}[X_i] = 0$ et $\text{Var}[X_i] = 1$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\frac{S_n}{\sqrt{n}} \leq x\right\} = \Phi(x).$$

Maintenant si les X_i ont $\mathbb{E}[X_i] = \mu$ et $\text{Var}[X_i] = \sigma^2$, on considère

$$\tilde{X}_i = \frac{X_i - \mu}{\sigma}; \quad \tilde{S}_n = \sum_{i=1}^n \tilde{X}_i.$$

On vient tout juste de montrer que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\frac{\tilde{S}_n}{\sqrt{n}} \leq x\right\} = \Phi(x).$$

Mais si on développe \tilde{S}_n , on trouve que c'est directement équivalent à

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right\} = \Phi(x).$$

Et le théorème est démontré. □

Nous avons déjà vu un cas particulier du théorème de la limite centrale ; en effet, le théorème de DeMoivre-Laplace (proposition 5.6) émerge naturellement si on considère notre variable $S_n = \sum_{i=1}^n X_i$, avec les X_i des variables aléatoires de Bernoulli de paramètre p indépendantes. En effet, dans ce cas, S_n est bien une variable aléatoire de loi Binomiale, mais c'est aussi une somme de variables aléatoires indépendantes et identiquement distribuées.

EXAMPLE 8.9. Dans une classe de $N = 2k = 120$ étudiant.e.s, l'enseignant.e veut tester la solution surprenante au problème de Monty-Hall. Les étudiant.e.s sont regroupé.e.s en $k = 60$ paires. Dans chaque paire, un.e étudiant.e cache une pièce de monnaie sous l'un de trois morceaux de papiers. L'autre étudiant.e fait un premier choix, le/la premier/ère retourne l'un des papiers qui ne cache rien, puis l'étudiant.e qui avait fait un premier choix décide de changer son choix.

- (a) En moyenne combien de paires d'étudiant.e.s auront trouvé la pièce de monnaie ?
- (b) Dans quel intervalle se retrouve le nombre de paires ayant réussi (19 fois sur 20) ?

SOLUTION. (a) La probabilité de succès avec la tactique de changer son choix est de $p = 2/3$. Donc, l'espérance du nombre de succès est $\mu = 2k/3 = N/3$.

- (b) Si S_k est le nombre de succès, alors pour d'assez grandes valeurs de k , on peut approximer $(S_k - kp)/(\sigma\sqrt{k})$ comme une variable de distribution normale centrée, si σ^2 est la variance pour une tentative.

En outre, on sait qu'une variable normalement distribuée a 19 chances sur 20 de se retrouver entre -2 et 2 .

Bien sûr, on a que $\sigma^2 = p(1-p) = \frac{2}{9}$. Donc $(S_k - 2k/3)/\sqrt{2k/9}$ se retrouve entre -2 et 2 avec une probabilité de $95/100$ environ. Donc, $S_k - 2k/3$ se retrouve entre $\pm 2\sqrt{2k/9}$ et S_k se retrouve entre $2k/3 \pm 2\sqrt{2k/9}$.

Avec $k = 60$, $2\sqrt{2k/9} = 4\sqrt{10/3} \approx 7,303\dots$

On aurait donc que S_k se trouve 19 fois sur 20 entre 32 et 48.

Nous y voici enfin arrivé.e.s ! La fin du cours ! Enfin presque. Il reste une série d'exercices !

8.6. Exercices

EXERCICE 8.1. Soient X_1, X_2, \dots, X_{20} des variables aléatoires indépendantes et identiquement distribuées suivant respectivement des lois de Poisson de paramètre 1.

(a) Utiliser l'inégalité de Tchebychev-Markov pour obtenir une borne sur

$$\mathbb{P} \left\{ \sum_{i=1}^{20} X_i \geq 40 \right\}.$$

(b) Utiliser le théorème de la limite centrale pour obtenir une approximation de

$$\mathbb{P} \left\{ \sum_{i=1}^{20} X_i \geq 40 \right\}.$$

Discuter.

EXERCICE 8.2. Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires et $c \in \mathbb{R}$ une constante telle que pour tout $\epsilon > 0$, on a que

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ |X_n - c| > \epsilon \} = 0.$$

(La suite X_n converge en probabilités vers c .)

Soit $g : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et bornée.

(a) Montrer que pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \{ |g(X_n) - g(c)| > \epsilon \} = 0.$$

(b) Déduire que

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(X_n)] = g(c).$$

EXERCICE 8.3. Nous allons réaliser une partie d'une preuve célèbre de la *densité des polynômes* dans l'ensemble des fonctions continues sur un intervalle compact ; ce résultat est fondamental en mathématiques. Il établit que pour toute fonction $f : [a, b] \rightarrow \mathbb{R}$ continue, il existe pour tout $\epsilon > 0$ un polynôme p tel que $\sup_{x \in [a, b]} |p(x) - f(x)| < \epsilon$.

Soit $f : [0, 1] \rightarrow \mathbb{R}$ une fonction continue. Considérons le polynôme

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}.$$

Ces fonctions sont appelées *polynômes de Bernstein*. Montrer que

$$\lim_{n \rightarrow \infty} B_n(x) = f(x),$$

pour tout $x \in [0, 1]$.

Indice : utilisez l'exercice précédent et un choix judicieux de variables aléatoires et de fonction. *Note* : Pour prouver la densité des polynômes dans l'espace des fonctions continues (au sens mentionné plus haut), il ne suffit pas de montrer la convergence ponctuelle comme on vient de le faire ; on devrait plutôt montrer la convergence uniforme – toutefois, la convergence ponctuelle est une condition nécessaire.

EXERCICE 8.4. La borne de Chernoff pour la queue d'une variable aléatoire normale centrée réduite est

$$\mathbb{P}\{Z \geq a\} \leq e^{-a^2/2}.$$

Montrer, en considérant la densité de Z , que le côté droit de cette inégalité peut être réduit d'un facteur 2 – c'est à dire que

$$\mathbb{P}\{Z \geq a\} \leq \frac{1}{2}e^{-a^2/2}.$$

EXERCICE 8.5. Le temps que met un certain composant avant de se briser est une variable aléatoire de densité $f(x) = 2x$ pour $x \in (0, 1)$. Dès que ce composant brise, on le remplace immédiatement, de sorte que le moment du n ième bris soit donné par $S_n = \sum_{i=1}^n X_i$. À long terme, déterminer le rythme moyen auquel il faut remplacer ce composant ; c'est-à-dire, déterminer le nombre moyen de remplacement par unité de temps.

EXERCICE 8.6. On suppose qu'on a un échantillon de n données indépendantes et identiquement distribuées $X_1, X_2, X_3, \dots, X_n$. On va noter f la fonction de densité marginale des X_i , mais f est inconnue (mais on assume que les X_i sont des variables aléatoires continues).

Pour tenter de se faire une idée de la distribution des X_i , on va tracer un histogramme. L'histogramme est la fonction suivante :

$$h_{n,\epsilon}(y) = \sum_{k \in \mathbb{Z}} N_\epsilon(k, k+1) \mathbb{1}_{[k,k+1]}(y),$$

où $N_\epsilon(k, k+1) = \sum_{i=1}^n \mathbb{1}_{[\epsilon k, \epsilon(k+1)]}(X_i)$ est le nombre de données dans notre échantillon qui se situent dans l'intervalle $[\epsilon k, \epsilon(k+1)]$. Les $N_\epsilon(k, k+1)$ sont des variables aléatoires, bien entendu.

(a) Montrer que pour tout $y \in [\epsilon k, \epsilon(k+1)]$,

$$\lim_{n \rightarrow \infty} \frac{1}{n\epsilon} h_{n,\epsilon}(y) = \frac{1}{\epsilon} \int_{\epsilon k}^{\epsilon(k+1)} f(t) dt.$$

(b) Montrer que pour tout $y \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \lim_{\epsilon \rightarrow 0^+} \frac{1}{n\epsilon} h_{n,\epsilon}(y) = f(y).$$

Note : Ce qu'on vient de montrer, c'est que lorsqu'on accumule de plus en plus de données et qu'on trace un histogramme, si on prend l'histogramme de plus en plus fin, en renormalisant par $n\epsilon$ où n est la taille de l'échantillon et ϵ est la largeur des bandes de l'histogramme, l'histogramme converge vers la fonction de densité pour la distribution marginale des X_i .

EXERCICE 8.7. Soit E_1, E_2, E_3, \dots une suite d'événements. On considère l'événement

$$\limsup_{n \rightarrow \infty} E_n = \bigcap_{i \in \mathbb{N}} \left(\bigcup_{j \geq i} E_j \right).$$

Intuitivement, il s'agit de l'événement « une infinité des E_i sont réalisés. »

On a le résultat suivant :

PROPOSITION (Lemme de Borel-Cantelli). *Soit $(E_i)_{i \in \mathbb{N}}$ une suite d'événements. Si $\sum_{i \in \mathbb{N}} \mathbb{P}\{E_i\} < +\infty$, alors*

$$\mathbb{P}\left\{\limsup_{n \rightarrow \infty} E_n\right\} = 0.$$

- (a) Soit N le nombre d'événements de E_i qui sont réalisés. Montrer que

$$\mathbb{E}[N] = \sum_{i \in \mathbb{N}} \mathbb{P}\{E_i\}.$$

- (b) Utiliser ce fait pour démontrer le lemme de Borel-Cantelli.

Annexes

Annexe A

Théorie des ensembles

Un **ensemble** est un objet mathématique qui représente... un... regroupement d'objets ? La définition d'ensemble est difficile à donner précisément sans entrer dans des détails fastidieux.

On note souvent les ensembles en utilisant des lettres majuscules comme A, B, C, E, F, X, Y , etc.

On note $\emptyset = \{\}$ l'ensemble vide. Cet ensemble ne contient aucun élément.

On dit qu'un objet x est **élément** d'un ensemble X , et on note

$$x \in X.$$

On dit qu'un ensemble X est **inclus** dans un ensemble Y , et on note $X \subseteq Y$ si tous les éléments de X sont aussi des éléments de Y . On dit aussi que X est un **sous-ensemble**, ou une **partie**, de Y .

On note $\mathcal{P}(X)$ l'ensemble des parties de X .

A.1. Description d'ensembles

Il existe plusieurs façons de décrire les ensembles.

1. **En extension** : Ici, il s'agit simplement d'énumérer un à un tous les éléments de l'ensemble en question, encadrés par des accolades $\{\}$.

EXEMPLE A.1.

$$X = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31\}.$$

2. **En compréhension** : Pour décrire un sous-ensemble d'un ensemble plus général, on peut noter entre accolades
 - un « gabarit » pour les éléments, avec une ou plusieurs variables « muettes », en précisant préférablement de quel ensemble nos éléments font partie (notre ensemble plus général) ;
 - une ou des conditions que nos variables « muettes » doivent satisfaire.

EXEMPLE A.2. Le cercle-unité dans le plan complexe peut être décrit comme suit :

$$S = \{z \in \mathbb{C} : |z| = 1\}.$$

On lit : « S est l'ensemble des z dans \mathbb{C} tels que ($:)$ $|z| = 1$ ».

EXEMPLE A.3. La sphère unité dans \mathbb{R}^n peut être décrite comme suit :

$$S_n = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n : \sum_{k=1}^n x_k^2 = 1 \right\}.$$

EXEMPLE A.4. L'ensemble des nombres pairs peut-être décrit comme suit :

$$P = \{2n \in \mathbb{Z} : n \in \mathbb{Z}\}.$$

Il existe aussi certains raccourcis :

- \mathbb{N} , \mathbb{Z} , \mathbb{Q} , \mathbb{R} , \mathbb{C} , pour les nombres naturels, entiers, rationnels, réels et complexes respectivement. On utilisera aussi parfois $\mathbb{N}^* = \mathbb{N} \cup \{0\}$ pour représenter les entiers non-négatifs.
- De façon générale, pour noter un intervalle dans les nombres réels, on emploiera

$$\begin{aligned}(a, b) &:= \{x \in \mathbb{R} : a < x < b\} \\ [a, b] &:= \{x \in \mathbb{R} : a \leq x \leq b\} \\ [a, b) &:= \{x \in \mathbb{R} : a \leq x < b\} \\ (a, b] &:= \{x \in \mathbb{R} : a < x \leq b\}\end{aligned}$$

- On peut parfois commettre des abus de notation lorsque le sens est très clair. Par exemple, on pourrait noter

$$P = \{k \in \mathbb{N} : k \text{ est pair}\},$$

pour désigner les nombres pairs.

Quoiqu'il arrive, on place toujours la description d'un ensemble entre accolades.

A.2. Opérations ensemblistes

On peut construire des ensembles à l'aide d'autres ensembles en utilisant les opérations suivantes :

1. La réunion : \cup

On dénote par $X \cup Y$ la réunion des ensemble X et Y , c'est-à-dire l'ensemble qui contient tous les éléments de X et tous les éléments de Y . Donc, si $x \in X \cup Y$, c'est que $x \in X$ OU $x \in Y$:

$$X \cup Y := \{x : x \in X \text{ ou } x \in Y\}.$$

2. L'intersection : \cap

On dénote par $X \cap Y$ l'intersection entre les ensembles X et Y , c'est-à-dire l'ensemble qui contient tous les éléments qui sont à la fois dans X et dans Y :

$$X \cap Y := \{x \in X \cup Y : x \in X, x \in Y\}.$$

On dit que deux ensembles sont **disjoints** si leur intersection est l'ensemble vide – c'est à dire que les deux ensembles n'ont aucun élément en commun.

REMARQUE. On utilisera parfois la notation $X \sqcup Y$ pour signifier que l'on fait la réunion de X et Y , et que X et Y sont des ensembles disjoints. Cela n'est qu'une convention qui sert surtout de rappel visuel – ce n'est pas une opération différente de la réunion ordinaire.

3. La différence : \setminus

On dénote par $X \setminus Y$ (lire « X sans Y ») la différence d'ensembles. C'est l'ensemble des éléments qui sont dans X , mais pas dans Y .

$$X \setminus Y := \{x \in X : x \notin Y\}.$$

4. La différence symétrique Δ

On dénote par $X \Delta Y$ (lire « X xou Y ») la différence symétrique d'ensembles. C'est l'ensemble des éléments qui sont soit dans X , soit dans Y , mais pas dans les deux à la fois.

$$X \Delta Y = (X \setminus Y) \cup (Y \setminus X) = (X \cup Y) \setminus (X \cap Y).$$

La différence symétrique est commutative.

Il nous arrivera souvent de travailler surtout avec des sous-ensembles d'un ensemble «global» – appelons-le Ω . On peut alors définir le **complément** par rapport à Ω d'un sous-ensemble Y . On le note Y^c et il est défini par l'ensemble des éléments de Ω qui ne sont pas dans Y :

$$Y^c := \{x \in \Omega : x \notin Y\} = \Omega \setminus Y.$$

La notion de complémentaire fait toujours référence à un ensemble global ; elle n'a pas de sens hors-contexte.

Dès lors, on voit que pour tous sous-ensembles $X, Y \subseteq \Omega$, on a la relation suivante :

$$X \setminus Y = X \cap Y^c.$$

DÉFINITION A.1 (Ensembles disjoints). On dit que deux ensembles X et Y sont **disjoints** si et seulement si

$$X \cap Y = \emptyset.$$

Évidemment, dans le contexte d'un ensemble global Ω , on a toujours que $X \subseteq \Omega$ et son complément X^c sont disjoints. D'ailleurs, une des méthodes les plus efficaces pour montrer que deux ensembles sont disjoints dans un tel contexte est de montrer que l'un est une partie d'un ensemble, et l'autre une partie du complément...

DÉFINITION A.2 (Partition). Soient X un ensemble et $(Y_i \subseteq X)_{i \in I}$ une famille de parties de X indexée par un ensemble d'indices I (soit fini, soit dénombrable, ou même potentiellement indénombrable).

On dit que $(Y_i)_{i \in I}$ forme une **partition** de X si et seulement si

- i. Pour toute paire d'indices $i, j \in I$ avec $i \neq j$, on a que Y_i et Y_j sont disjoints (on dit que les Y_i sont disjoints deux-à-deux) ;
- ii. On a que

$$\bigcup_{i \in I} Y_i = X.$$

Intuitivement, construire une partition d'un ensemble c'est un peu comme le découper en tranches séparées.

PROPOSITION A.1 (Partitions induites). Soit Ω un ensemble, $X \subseteq \Omega$ et soit $(Y_i)_{i \in I}$ une partition de Ω . Alors, $(Z_i := Y_i \cap X)_{i \in I}$ est une partition de X .

DÉMONSTRATION. D'abord, tous les Z_i sont des parties de X puisque ce sont des intersections de X avec d'autres ensembles.

Ensuite, pour toute paire i, j dans I avec $i \neq j$, on a que $Y_i \subseteq Y_j^c$. Puisque $Z_i \subseteq Y_i$ et $Z_j \subseteq Y_j$, alors on a que $Z_i \subseteq Y_j^c$ et $Z_j \subseteq Y_i^c$, et Z_i et Z_j sont disjoints.

Finalement, par distributivité de l'intersection, on a que

$$\bigcup_{i \in I} Z_i = \bigcup_{i \in I} (X \cap Y_i) = X \cap \bigcup_{i \in I} Y_i = X \cap \Omega = X.$$

On a montré que $(Z_i)_{i \in I}$ satisfait la définition d'une partition pour X . \square

Ici on dit que $(Z_i)_{i \in I}$ est la partition induite par $(Y_i)_{i \in I}$ sur X , parce qu'elle est construite directement à partir de cette dernière.

A.3. Le produit cartésien

On peut également construire le **produit cartésien** des ensembles X, Y , noté $X \times Y$, comme ceci :

$$X \times Y := \{(x, y) : x \in X, y \in Y\}.$$

Il s'agit de l'ensemble des couples ordonnés, dont la première entrée est dans X , et la seconde dans Y . Évidemment, pour des ensembles finis, on a que $|X \times Y| = |X| |Y|$ – tel que discuté au premier chapitre.

On emploie les exposants comme raccourci de notation pour indiquer des produits cartésiens répétés d'un ensemble avec lui-même :

$$X^n = \underbrace{X \times X \times \cdots \times X}_{n \text{ fois}}.$$

A.4. Propriétés des opérations

Les propriétés suivantes sont vraies pour des sous-ensembles de Ω (les compléments sont par rapport à Ω) :

i. La commutativité de la réunion :

$$X \cup Y = Y \cup X;$$

ii. L'associativité de la réunion :

$$X \cup (Y \cup Z) = (X \cup Y) \cup Z;$$

iii. La commutativité de l'intersection :

$$X \cap Y = Y \cap X;$$

iv. L'associativité de l'intersection :

$$X \cap (Y \cap Z) = (X \cap Y) \cap Z;$$

Ajoutons avant d'aller plus loin que, puisque les réunions et les intersections sont associatives et commutatives, on peut sans ambiguïté employer les notations suivantes :

$$\bigcup_{i=1}^n E_i, \quad \bigcap_{i=1}^n E_i,$$

pour désigner les réunions de plusieurs ensembles.

v. La distributivité de l'intersection sur la réunion :

$$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z);$$

vi. La distributivité de la réunion sur l'intersection :

$$(X \cap Y) \cup Z = (X \cup Z) \cap (Y \cup Z);$$

vii. La neutralité de l'ensemble vide pour la réunion :

$$X \cup \emptyset = X;$$

viii. La neutralité de Ω pour l'intersection :

$$X \cap \Omega = X;$$

ix. Les identités de De Morgan : si X_1, X_2, \dots, X_n sont des sous-ensembles de Ω , alors

$$\left(\bigcup_{i=1}^n X_i \right)^c = \bigcap_{i=1}^n X_i^c, \quad \left(\bigcap_{i=1}^n X_i \right)^c = \bigcup_{i=1}^n X_i^c.$$

A.5. Ensembles et fonctions

On fait des liens entre ensembles au moyen de *fonctions*.

DÉFINITION A.3 (Fonctions, domaine, image). Soient deux ensembles, X et Y . On note $f : X \rightarrow Y$ une **fonction** f , de X dans Y , qui associe pour chaque élément x du **domaine** X une **image** y , qui est un élément de Y .

On note alors $y = f(x)$. Concernant l'expression $f(x)$:

- f est la fonction ;
- $f(x)$ est l'image de x par la fonction f . C'est un élément de Y .
- x est appelé **l'argument** de f .

Si $f : X \rightarrow Y$ est une fonction, pour toute partie $A \subseteq X$, on définit la notation suivante :

$$f(A) := \{f(x) \in Y : x \in A\},$$

que l'on appelle **l'image de l'ensemble A par la fonction f** . Bien entendu, on a toujours que $f(A) \subseteq Y$.

De façon générale, les fonctions préservent la relation d'inclusion sur les ensembles : si $A \subseteq B \subseteq X$ et $f : X \rightarrow Y$ est une fonction de X vers Y , alors $f(A) \subseteq f(B)$.

Pour $B \subseteq Y$, on définit également

$$f^{-1}(B) := \{x \in X : f(x) \in B\},$$

le **pré-image de l'ensemble B par la fonction f** .

A.5.1. Injections, surjections, bijections.

DÉFINITION A.4 (Fonction injective). Une fonction $f : X \rightarrow Y$ est dite **injective** (on dit aussi que c'est une injection) si et seulement si

$$f(x) = f(y) \Rightarrow x = y.$$

En d'autres mots, une fonction est injective si et seulement si deux éléments différents de l'ensemble de départ X sont toujours envoyés vers deux éléments distincts dans l'ensemble d'arrivée. De façon plus générale, si $A, B \subseteq X$ sont disjoints, alors $f(A)$ et $f(B)$ sont disjoints également.

DÉFINITION A.5 (Fonction surjective). Une fonction $f : X \rightarrow Y$ est dite **surjective** (on dit aussi que c'est une surjection) si et seulement si pour tout $y \in Y$, il existe $x \in X$ tel que $f(x) = y$.

Autrement dit, f est surjective si et seulement si $Y \subseteq f(X)$, c'est-à-dire que tous les éléments de Y ont un pré-image dans X .

DÉFINITION A.6 (Fonction bijective). Une fonction $f : X \rightarrow Y$ est dite **bijective** (on dit aussi que c'est une bijection) si et seulement si c'est une fonction à la fois injective et surjective.

On a finalement le théorème suivant :

THÉORÈME A.1 (Schröder-Bernstein). *Soient X, Y deux ensembles. Supposons qu'il existe $f : X \rightarrow Y$ et $g : Y \rightarrow X$ des fonctions injectives. Alors, il existe une bijection $h : X \rightarrow Y$.*

A.5.2. Fonctions indicatrices. On a souvent recours, en mathématiques, à ce que nous appelons les *fonctions indicatrices*. Elles sont défini dans le contexte d'un ensemble global Ω :

DÉFINITION A.7 (Fonction indicatrice). Soit $E \subseteq \Omega$ un ensemble. On définit la **fonction indicatrice** de l'ensemble E , notée $\mathbb{1}_E : \Omega \rightarrow \{0, 1\}$ la fonction qui vaut 1 lorsque son argument est dans E , et 0 autrement :

$$\mathbb{1}_E(\omega) = \begin{cases} 1 & \text{si } \omega \in E \\ 0 & \text{sinon.} \end{cases}$$

Les fonctions indicatrices ont les quelques propriétés suivantes :

- i. Pour tout $E \subseteq \Omega$, $\mathbb{1}_E + \mathbb{1}_{E^c} = 1$, où on comprend que 1 ici est la fonction constante.
- ii. Pour tous $E, F \in \Omega$, $\mathbb{1}_E \mathbb{1}_F = \mathbb{1}_{E \cap F}$.
- iii. Pour une famille $(E_i)_{i \in I}$ d'ensembles disjoints deux-à-deux, indexés par un ensemble d'indices $I \subseteq \mathbb{N}$ au plus dénombrable, alors

$$\mathbb{1}_{\bigsqcup_{i \in I} E_i} = \sum_{i \in I} \mathbb{1}_{E_i}.$$

A.6. Bonus : La cardinalité des ensembles

La cardinalité des ensembles est définie comme une classe d'équivalence. On dit que X et Y ont la même cardinalité si il existe une bijection entre les deux (c'est la relation d'équivalence). On note $|X|$ la cardinalité de X , qui est un nombre entier non-négatif lorsque X est un ensemble fini.

On dit que X est dénombrable si $|X| = |\mathbb{N}|$. Par exemple, \mathbb{N} , \mathbb{Z} et \mathbb{Q} sont dénombrables.

La relation d'ordre sur les cardinalités peut être définie en disant que $|A| \leq |B|$ si et seulement si il existe une fonction injective de A vers B . Pour les ensembles finis, cette relation d'ordre coïncide parfaitement avec la relation d'ordre pour les entiers non-négatifs.

On dit que X est non-dénombrable si $|X| > |\mathbb{N}|$.

A.7. Bonus : Les relations d'ordre

DÉFINITION A.8 (Ordre partiel). La relation \leq est un **ordre partiel** sur l'ensemble X si et seulement si il satisfait les axiomes suivants :

- i. $x \leq x$ pour tous x
- ii. $x \leq y$ et $y \leq x$ implique $x = y$.
- iii. $x \leq y$ et $y \leq z$ implique $x \leq z$.

Voici des exemples de relations d'ordre partiel :

EXEMPLE A.5. Soit Ω un ensemble non-vide et soit $\mathcal{P}(\Omega)$ l'ensemble de ses parties. La relation \subseteq définie entre les éléments de $\mathcal{P}(\Omega)$ est un ordre partiel.

EXEMPLE A.6. La relation d'ordre usuelle \leq sur les nombres réels est une relation d'ordre partiel.

EXEMPLE A.7. La relation « $a|b$ » ou « a divise b », sur \mathbb{N} , est une relation d'ordre partiel.

EXEMPLE A.8. Soit P un ensemble de propositions logiques, qui contient une contradiction F et une tautologie V . La relation « $p \rightarrow q$ », ou « p implique q », est une relation d'ordre partiel.

On dit que X est partiellement ordonné lorsqu'il est muni d'un ordre partiel.

A.7.1. Le lemme de Zorn. Soit X un ensemble partiellement ordonné, et soit \leq sa relation d'ordre.

DÉFINITION A.9 (Élément maximal (resp. minimal)). L'élément $x \in X$ est **maximal** si pour tout $y \in X$ tel que $x \leq y$ (resp. $y \leq x$), $y = x$.

DÉFINITION A.10 (Majorant (resp. minorant)). L'élément $m \in X$ est un **majorant** (resp. minorant) pour une partie $Y \subseteq X$ si pour tout $y \in Y$, on a que $y \leq m$ (resp. $m \leq y$).

DÉFINITION A.11 (Ensemble totalement ordonné). L'ensemble $Y \subseteq X$ est **totalement ordonné** si pour toute paire $x, y \in Y$, soit $x \leq y$, soit $y \leq x$.

DÉFINITION A.12 (Ensemble inductif). L'ensemble X est **inductif** si toute partie totalement ordonnée de X admet un majorant.

LEMME A.1 (Zorn). *Tout ensemble partiellement ordonné inductif non-vide admet un élément maximal.*

A.7.2. Les treillis. Les symboles \wedge et \vee font surface dans plusieurs contextes mathématiques différents, qui n'ont *a priori* rien à voir les uns avec les autres. Par exemple :

- On utilise souvent $a \vee b$ pour le maximum de nombres réels a et b , et $a \wedge b$ pour le minimum entre ces deux nombres ;
- On utilise aussi parfois $a \vee b$ pour le PPCM d'entiers naturels a et b , et $a \wedge b$ pour leur PGCD ;
- En logique, on utilise \vee pour représenter «ou», et \wedge pour représenter «et», entre les propositions. Par exemple, $p \wedge q$ signifierait « p et q ».

On pourrait penser qu'il s'agit de choix plus ou moins arbitraires, et qu'il n'existe pas vraiment de lien entre ces choix. Or, c'est faux. En effet, si on utilise le même symbole, c'est parce qu'il représente la même opération. Pour le comprendre, il faut s'intéresser à la théorie des treillis.

Soit X un ensemble partiellement ordonné et \leq sa relation d'ordre.

DÉFINITION A.13 (supremum (resp. infimum)). Pour une partie $Y \subseteq X$, on dit que Y admet un supremum (resp. infimum) si

- Y admet au moins 1 majorant (resp. minorant) ;
- L'ensemble des majorants (resp. minorants) admet un élément minimal (resp. maximal)

Dans ce cas, on note $\sup Y$ (resp. $\inf Y$) l'élément minimal (resp. maximal) des majorants (resp. minorants) de Y .

DÉFINITION A.14 (Treillis – par la relation d'ordre.). X est un treillis si toute paire $x, y \in X$ admet un supremum et un infimum.

On note alors

$$x \vee y := \sup\{x, y\}, \quad x \wedge y := \inf\{x, y\}.$$

DÉFINITION A.15 (Treillis – par l'algèbre.). L'ensemble X muni des opérations binaire \vee et \wedge est un treillis si les opérations satisfont les axiomes suivants :

i. La commutativité pour \vee et \wedge :

$$x \wedge y = y \wedge x, \quad x \vee y = y \vee x;$$

ii. L'associativité pour \vee et \wedge :

$$x \wedge (y \wedge z) = (x \wedge y) \wedge z, \quad x \vee (y \vee z) = (x \vee y) \vee z;$$

iii. L'absorptivité pour \vee et \wedge :

$$x \wedge (x \vee y) = x = x \vee (x \wedge y);$$

On peut vérifier facilement que ces deux définitions sont équivalentes. Voici des exemples de treillis bien connus :

EXEMPLE A.9. Si Ω est un ensemble non-vide, alors $\mathcal{P}(\Omega)$ muni de la relation d'ordre \subseteq est un treillis. Ici, la réunion \cup joue le rôle de l'opérateur \vee , et l'intersection \cap joue le rôle de l'opération \wedge .

EXEMPLE A.10. Si on considère l'ensemble \mathbb{R} et sa relation d'ordre \leq usuelle, alors c'est un treillis. On a alors bien que $a \wedge b = \min\{a, b\}$, et $a \vee b = \max\{a, b\}$.

EXEMPLE A.11. Si on considère l'ensemble \mathbb{N} et sa relation d'ordre $|$ (divise), alors, c'est un treillis. On a bien que $a \wedge b = \text{PGCD}\{a, b\}$ et que $a \vee b = \text{PPCM}\{a, b\}$.

EXEMPLE A.12. Si on considère un ensemble de propositions logiques complet avec une tautologie et une contradiction, alors $p \wedge q$ est bien la proposition « p et q », qui implique les deux et qui est donc un infimum, parce que toutes les propositions qui l'impliquent impliquent aussi p et q . De même, $p \vee q$ est bien la proposition « p ou q ».

A.8. Bonus : Limites d'ensembles

Si on considère une suite $(X_n)_{n \in \mathbb{N}}$ d'ensembles, on peut se demander s'il est possible de donner un sens à l'expression $\lim_{n \rightarrow \infty} X_n$.

La réponse : oui !

En fait, comme pour les suites de nombres réels auxquelles nous sommes habitué.e.s, les suites d'ensemble ne convergent pas toutes.

A.8.1. Suites monotones d'ensembles. On définit les suites monotones d'ensemble de la façon naturelle :

DÉFINITION A.16 (Suite croissante (resp. décroissante) d'ensembles). Si $(X_n)_{n \in \mathbb{N}}$, on dit que X_n est **croissante** (resp. **décroissante**) si pour tout $n \in \mathbb{N}$, on a $X_n \subseteq X_{n+1}$ (resp. \supseteq).

La particularité, c'est que les suites monotones convergent.

DÉFINITION A.17 (Limites de suites monotones d'ensembles). Si $(X_n)_{n \in \mathbb{N}}$ est une suite monotone croissante, on définit :

$$\lim_{n \rightarrow \infty} X_n = \bigcup_{k \in \mathbb{N}} X_k;$$

si X_n est plutôt décroissante, on définit :

$$\lim_{n \rightarrow \infty} X_n = \bigcap_{k \in \mathbb{N}} X_k.$$

Intuitivement :

- L'ensemble-limite d'une suite croissante est l'ensemble qui comprend tous les éléments qui se retrouvent dans au moins un des ensembles de notre suite ;
- l'ensemble-limite d'une suite décroissante est l'ensemble qui comprend tous les éléments qui se retrouvent dans TOUS les ensembles de notre suite.

A.8.2. Définition plus générale. On peut s'imaginer qu'une suite d'ensemble qui n'est pas strictement monotone, mais qui «devient» strictement monotone éventuellement, devrait avoir une limite quand même. C'est pourquoi on généralise le concept de limites.

DÉFINITION A.18 (Limite supérieure d'une suite d'ensembles). Soit $(X_n)_{n \in \mathbb{N}}$ une suite d'ensembles. On définit

$$\limsup_{n \rightarrow \infty} X_n = \bigcap_{i \in \mathbb{N}} \left(\bigcup_{j \geq i} X_j \right).$$

Si vous avez lu la section A.7.2 sur les treillis, vous pouvez repérer le lien entre la définition de la limite supérieure pour les suites de nombres réels, et celle pour les ensembles. Pour les nombres réels, on a

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \sup\{x_k : k \geq n\}.$$

Pour les ensembles, on a

$$\begin{aligned} \limsup_{n \rightarrow \infty} X_n &= \bigcap_{i \in \mathbb{N}} \left(\bigcup_{j \geq i} X_j \right) \\ &= \lim_{n \rightarrow \infty} \sup\{X_k : k \geq n\}. \end{aligned}$$

par nos définition de limite pour les suites croissantes d'ensembles, et de supremum pour les ensembles – la suite $\sup\{X_k : k \geq n\} = \bigcup_{k \geq n} X_k$ est monotone décroissante, et sa limite est $\bigcap_{n \in \mathbb{N}} \left(\bigcup_{k \geq n} X_k \right)$.

De façon analogue, on a la définition suivante :

DÉFINITION A.19 (Limite inférieure d'une suite d'ensembles). Soit $(X_n)_{n \in \mathbb{N}}$ une suite d'ensembles. On définit

$$\liminf_{n \rightarrow \infty} X_n = \bigcup_{i \in \mathbb{N}} \left(\bigcap_{j \geq i} X_j \right).$$

Ces définitions peuvent donner un peu mal à la tête. Toutefois, on peut comprendre, à force d'efforts de concentration, que

- La limite supérieure d'une suite d'ensembles est l'ensemble de tous les éléments qui se retrouvent dans une infinité des termes de la suite ;
- la limite inférieure d'une suite d'ensembles est l'ensemble de tous les éléments qui se retrouvent dans tous les termes de la suite sauf un nombre fini.

On voit alors facilement la propriété suivante :

$$\liminf_{n \rightarrow \infty} X_n \subseteq \limsup_{n \rightarrow \infty} X_n.$$

On peut maintenant définir, de façon plus générale :

DÉFINITION A.20 (Limite d'une suite d'ensembles – définition générale). Soit $(X_n)_{n \in \mathbb{N}}$ une suite d'ensembles telle que $\limsup_{n \rightarrow \infty} X_n \subseteq \liminf_{n \rightarrow \infty} X_n$.

Alors, on définit

$$\lim_{n \rightarrow \infty} X_n = \limsup_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} X_n.$$

Annexe B

Rappels d'analyse

Cet annexe regroupe des éléments du cours d'analyse qui pourraient se révéler utiles pour la compréhension du cours. Les preuves ne sont pas toujours complètes, puisque l'objectif du présent annexe est davantage de servir comme aide-mémoire.

B.1. Structure des nombres réels

On considère \mathbb{Q} l'ensemble des nombres rationnels.

On dit que $A \subseteq \mathbb{Q}$ forme une **coupure de Dedekind** de l'ensemble \mathbb{Q} (totalement ordonné), si

- i. Pour tous $x \in A, y \in B := \mathbb{Q} \setminus A$, $x < y$;
- ii. Si $x \in \mathbb{Q}$ est une borne inférieure (un minorant, voir A.7) de B , alors $x \in A$.

L'ensemble \mathbb{R} des nombres réels peut alors être construit comme l'ensemble de toutes les coupures ainsi construites. On y retrouve l'ensemble \mathbb{Q} , identifié aux coupures A qui admettent un élément maximal, mais aussi toutes les coupure qui n'admettent pas d'élément maximal.

B.1.1. Propriétés des nombres réels. L'ensemble \mathbb{R} est totalement ordonné : si $x, y \in \mathbb{R}$ sont deux coupures de Dedekind, alors $x \leq y$ si et seulement si $x \subseteq y$.

On peut définir les opérations arithmétiques suivantes :

- i. L'addition : pour $x, y \in \mathbb{R}$, $x + y = \{(p + q) \in \mathbb{Q} : p \in x, q \in y\}$;
- ii. Le produit : pour $x, y \in \mathbb{R}^+$, $xy = \mathbb{Q}_{<0} \cup \{pq \in \mathbb{Q}^+ : p \in x, q \in y, p, q > 0\}$. On étend aux nombres négatifs algébriquement.

On dénote l'inverse additif de x par $-x$, l'inverse multiplicatif de x par x^{-1} , qui existe pour tout $x \in \mathbb{R}$ excepté 0 (neutre additif, élément de \mathbb{Q}). \mathbb{R} forme donc un anneau.

PROPOSITION B.1. *Les nombres réels ont les propriétés suivantes :*

- i. *L'addition est commutative et associative.*
- ii. *Le produit est commutatif, associatif, et distributif sur l'addition.*
- iii. *Pour tout $x, y \in \mathbb{R}$ avec $x < y$, il existe $r \in \mathbb{Q}$ tel que $x < r < y$.*
- iv. *Pour tout $p, q \in \mathbb{Q}$ avec $p < q$, il existe $x \in \mathbb{R}$ tel que $p < x < q$.*
- v. *\mathbb{R} n'admet aucun élément extrémal.*

Finalement, on a la propriété suivante (axiome de complétude d'Archimède) :

PROPOSITION B.2. *Pour tout $\epsilon > 0$, il existe $n \in \mathbb{N}$ tel que $1/n < \epsilon$.*

On fait également les définitions suivantes :

DÉFINITION B.1 (Supremum, infimum, maximum, minimum). Soit E une partie de \mathbb{R} . On définit le **supréum** de E (noté $\sup E$) comme étant la plus petite borne supérieure de E .

On définit l'infimum de E (noté $\inf E$) comme étant la plus grande borne inférieure de E .

On dit que E a un **maximum** (et on le note $\max E = \sup E$) si $\sup E \in E$.

On dit que E a un **minimum** (et on le note $\min E = \inf E$) si $\inf E \in E$.

On a les propriétés suivantes :

i. Si $E \subseteq F$, alors $\inf F \leq \inf E \leq \sup E \leq \sup F$.

ii. Pour tout $\epsilon > 0$, il existe $x \in E$ tel que

$$\sup E - \epsilon < x \leq \sup E.$$

iii. Pour tout $\epsilon > 0$, il existe $x \in E$ tel que

$$\inf E \leq x < \inf E + \epsilon.$$

DÉFINITION B.2 (Valeur absolue). Soit $x \in \mathbb{R}$. On définit la **valeur absolue** de x , notée $|x|$, par

$$|x| = \max\{x, -x\}.$$

On a immédiatement les identités suivantes :

$$(B.1.1) \quad \max\{a, b\} = \frac{a + b + |a - b|}{2}, \quad \min\{a, b\} = \frac{a + b - |a - b|}{2}.$$

On a aussi la proposition suivante :

PROPOSITION B.3 (Inégalité du triangle). *Pour tous $x, y \in \mathbb{R}$, on a que*

$$(B.1.2) \quad |x + y| \leq |x| + |y|$$

B.2. Un peu de topologie

DÉFINITION B.3 (Ouvert). Dans \mathbb{R} , un **ouvert** est n'importe quelle partie $E \subseteq \mathbb{R}$ qui peut être obtenue comme :

- un intervalle (a, b) , excluant les bornes ;
- une réunion quelconque d'ouverts ;
- une intersection finie d'ouverts.

On a également que les ensembles \mathbb{R} et \emptyset sont des ouverts.

DÉFINITION B.4 (Fermé). Dans \mathbb{R} , un **fermé** est n'importe quelle partie $E \subseteq \mathbb{R}$ dont le complément relatif à \mathbb{R} ($\mathbb{R} \setminus E$) est ouvert.

REMARQUE. *Attention !* Un ensemble peut être ni ouvert, ni fermé, ou il peut être à la fois ouvert et fermé.

DÉFINITION B.5 (Intérieur, fermeture, frontière). On dit que l'**intérieur** d'une partie $E \subseteq \mathbb{R}$, noté $\text{Int } E$ est la réunion de tous les ouverts $H \subseteq E$. C'est un ouvert.

La **fermeture** d'une partie $E \subseteq \mathbb{R}$, noté \overline{E} , est l'intersection de tous les fermés $F \subseteq \mathbb{R}$ qui contiennent E . C'est un fermé.

La **frontière** d'une partie $E \subseteq \mathbb{R}$, notée ∂E , est définie par

$$\partial E := \overline{E} \setminus \text{Int } E.$$

C'est un fermé, puisque c'est le complément de la réunion entre l'intérieur de E (un ouvert) et le complément de la fermeture de E (un autre ouvert).

DÉFINITION B.6 (Voisinage). On appelle « **voisinage de x** » tout ouvert $E \subseteq \mathbb{R}$ qui contient x .

DÉFINITION B.7 (Point adhérent). On dit que x est un **point adhérent** d'une partie quelconque E de \mathbb{R} si et seulement si tout voisinage A de x a une intersection non-vide avec E .

L'ensemble des points adhérents de E est appelé l'adhérence de E , et est exactement égal à la fermeture de E .

DÉFINITION B.8 (Point d'accumulation, point isolé). On dit que $x \in \mathbb{R}$ est un point d'accumulation d'une partie quelconque $E \subseteq \mathbb{R}$ si et seulement si tout voisinage A de x contient également un point $y \in E$ distinct de x .

Si x est un point d'accumulation, alors on a que tout voisinage A de x contient une infinité d'éléments de E .

Un point adhérent de E qui n'est pas un point d'accumulation de E est appelé un **point isolé** de E .

DÉFINITION B.9 (Partie discrète). On dit que $E \subseteq \mathbb{R}$ est une **partie discrète** de \mathbb{R} si tous ses éléments sont des points isolés.

PROPOSITION B.4 (Partie bornée). *On dit qu'une partie $E \subseteq \mathbb{R}$ est **bornée** si il existe un $n \in \mathbb{N}$ tel que $E \subseteq [-n, n]$.*

DÉFINITION B.10 (Compact). Une partie $E \subseteq \mathbb{R}$ est dite **compacte** si tout recouvrement de E par des ouverts contient un sous-recouvrement fini.

PROPOSITION B.5. *Une partie $E \subseteq \mathbb{R}$ est compacte si et seulement si elle est fermée et bornée.*

B.3. Suites et séries

DÉFINITION B.11 (Suite). Une **suite** de nombres réels est une famille dénombrable de nombres, souvent notée $(x_n)_{n \in \mathbb{N}}$.

DÉFINITION B.12 (Convergence, limite). On dit q'une suite $(x_n)_{n \in \mathbb{N}}$ est **convergente** si il existe un $x \in \mathbb{R}$ tel que pour tout voisinage A de x , il existe un $N \in \mathbb{N}$ tel que $x_n \in A$ pour tout $n \geq N$. On dit que $x \in \mathbb{R}$ est la **limite** de la suite, et on note

$$\lim_{n \rightarrow \infty} x_n = x.$$

PROPOSITION B.6 (Convergence, limite, avec $\epsilon > 0$). *Soit $(x_n)_{n \in \mathbb{N}}$ une suite de nombres réels. Alors, $(x_n)_{n \in \mathbb{N}}$ converge si et seulement si il existe un $x \in \mathbb{R}$ tel que pour tout $\epsilon > 0$, il existe $N \in \mathbb{N}$ tel que $|x - x_n| < \epsilon$ pour tout $n > N$.*

DÉFINITION B.13 (Suite de Cauchy). Soit $(x_n)_{n \in \mathbb{N}}$ une suite. On dit que c'est une suite de Cauchy si pour tout $\epsilon > 0$, il existe N tel que pour tous $m, n > N$, on a que $|x_m - x_n| < \epsilon$.

PROPOSITION B.7. *Soit $(x_n)_{n \in \mathbb{N}}$ une suite de Cauchy. Alors, $(x_n)_{n \in \mathbb{N}}$ converge.*

REMARQUE. Cette dernière proposition nous donne ce qu'on appelle la **complétude** des nombres réels.¹

1. En analyse, on dit qu'un espace métrique est complet lorsque toutes les suites de Cauchy convergent.

DÉFINITION B.14 (Croissance, décroissance, monotonicité). Soit $(x_n)_{n \in \mathbb{N}}$ une suite. On dit qu'elle est croissante (resp. décroissante) si on a que

$$x_n \leq x_{n+1} \quad (\text{resp. } \geq).$$

On dit qu'une suite est monotone si elle est croissante ou décroissante.

DÉFINITION B.15 (Suite bornée). On dit qu'une suite est **bornée** si elle est contenue dans une partie bornée de \mathbb{R} .

PROPOSITION B.8. Soit $(x_n)_{n \in \mathbb{N}}$ une suite monotone et bornée. Alors, $(x_n)_{n \in \mathbb{N}}$ converge.

PROPOSITION B.9 (De la sandwich). Soient $(x_n)_{n \in \mathbb{N}}, (y_n)_{n \in \mathbb{N}}, (z_n)_{n \in \mathbb{N}}$ trois suites de nombres réels, et supposons que

- i. $x_n \leq y_n \leq z_n$ pour tout n ;
- ii. x_n et z_n convergent vers x .

Alors, y_n converge aussi vers x .

DÉFINITION B.16 (Limites supérieures, inférieures). Soit $(x_n)_{n \in \mathbb{N}}$ une suite de nombres réels. On définit sa **limite supérieure** par

$$\limsup_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \sup\{x_k : k \geq n\}.$$

On définit sa **limite inférieure** par

$$\liminf_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} \inf\{x_k : k \geq n\}.$$

On admet que ces limites peuvent être $\pm\infty$. Elles sont donc toujours définies.

PROPOSITION B.10. On a que la suite $(x_n)_{n \rightarrow \infty}$ converge si et seulement si $\limsup_{n \rightarrow \infty} x_n \leq \liminf_{n \rightarrow \infty} x_n$, auquel cas on a

$$\lim_{n \rightarrow \infty} x_n = \liminf_{n \rightarrow \infty} x_n = \limsup_{n \rightarrow \infty} x_n.$$

DÉFINITION B.17 (Sous-suite). Soit $(x_n)_{n \in \mathbb{N}}$ une suite de réels, et $(n_k)_{k \in \mathbb{N}}$ une suite croissante d'entiers naturels. Alors on dit que $(x_{n_k})_{k \in \mathbb{N}}$ est une sous-suite de la suite $(x_n)_{n \in \mathbb{N}}$.

THÉORÈME B.1 (Bolzano-Weierstrass). Une partie $E \subseteq \mathbb{R}$ est compacte si et seulement si toute suite dans E admet un point d'accumulation (et donc une sous-suite convergente).

B.3.1. Séries de nombres réels.

DÉFINITION B.18 (Série). Soit $(x_n)_{n \in \mathbb{N}}$ une suite de nombres réels. La série correspondante est la somme infinie :

$$\sum_{n=1}^{\infty} x_n.$$

DÉFINITION B.19 (Convergence d'une série, convergence absolue, conditionnelle). Soit $(x_n)_{n \in \mathbb{N}}$ une suite de nombre réels. On dit que la série $\sum_{n \in \mathbb{N}} x_n$ converge si et seulement si la suite $S_n = \sum_{k \leq n} x_k$ converge.

On dit que la série $\sum_{n \in \mathbb{N}} x_n$ converge absolument si la série $\sum_{n \in \mathbb{N}} |x_n|$ converge.

On dit qu'elle converge conditionnellement si elle converge mais pas absolument.

PROPOSITION B.11. Une série converge si elle converge absolument.

PROPOSITION B.12. Soit $(x_n)_{n \in \mathbb{N}}$ une suite. La série $\sum_{n \in \mathbb{N}} x_n$ converge seulement si $x_n \rightarrow 0$. Attention ! Ce critère est nécessaire, mais il n'est pas suffisant.

EXEMPLE B.1 (La série harmonique). La série $\sum_{n \in \mathbb{N}} \frac{1}{n^\alpha}$ diverge pour $\alpha = 1$, mais converge pour toutes valeurs de $\alpha > 1$.

EXEMPLE B.2 (La série géométrique). La série $\sum_{n \in \mathbb{N}} r^{n-1}$ converge lorsque $|r| < 1$. Sa limite est alors $1/(1-r)$.

PROPOSITION B.13 (Critère de comparaison pour la convergence). Soient $(x_n)_{n \in \mathbb{N}}$ et $(y_n)_{n \in \mathbb{N}}$ deux suites de termes positifs, avec $x_n \leq y_n$.

Si la série $\sum_{n \in \mathbb{N}} y_n$ converge, alors la série $\sum_{n \in \mathbb{N}} x_n$ converge aussi.

Si la série $\sum_{n \in \mathbb{N}} x_n$ diverge, alors la série $\sum_{n \in \mathbb{N}} y_n$ diverge aussi.

PROPOSITION B.14 (Critère de D'Alembert). Soit $(x_n)_{n \in \mathbb{N}}$ une suite de termes positifs, avec

$$\lim_{n \rightarrow \infty} \frac{x_{n+1}}{x_n} = r.$$

- Si $r < 1$, alors la série $\sum_{n \in \mathbb{N}} x_n$ converge.
- Si $r > 1$, alors la série $\sum_{n \in \mathbb{N}} x_n$ diverge.
- Si $r = 1$, on ne peut rien conclure.

PROPOSITION B.15 (Critère de Cauchy). Soit $(x_n)_{n \in \mathbb{N}}$ une suite de termes positifs, avec

$$\lim_{n \rightarrow \infty} x_n^{1/n} = r.$$

- Si $r < 1$, alors la série $\sum_{n \in \mathbb{N}} x_n$ converge.
- Si $r > 1$, alors la série $\sum_{n \in \mathbb{N}} x_n$ diverge.
- Si $r = 1$, alors on ne peut rien conclure.

B.4. Fonctions

On considère maintenant des fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$.

DÉFINITION B.20 (Continuité en un point). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. On dit que f est continue en $x \in \mathbb{R}$ si pour tout voisinage (ouvert) $E \subseteq \mathbb{R}$ de $f(x)$, le préimage $E' = f^{-1}(E)$ de E par f est un voisinage (ouvert) de x .

PROPOSITION B.16. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. f est continue en x si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que dès que $y \in \mathbb{R}$ satisfait $|y - x| < \delta$, alors $|f(y) - f(x)| < \epsilon$.

PROPOSITION B.17 (Continuité séquentielle). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. f est continue en x si et seulement si pour toute suite $(x_n)_{n \in \mathbb{N}}$ qui converge vers x , on a que la suite $f(x_n)$ converge vers $f(x)$.

DÉFINITION B.21 (Continuité). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. f est continue si elle est continue en tout point.

Intuitivement, une fonction est continue si on peut tracer son graphe « sans lever la pointe de son crayon ».

DÉFINITION B.22 (Continuité uniforme). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. On dit que f est uniformément continue si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que pour tout $x \in \mathbb{R}$, on a que pour tout $y \in \mathbb{R}$ qui satisfait $|y - x| < \delta$, alors $|f(y) - f(x)| < \epsilon$.

REMARQUE. *Attention !* La différence entre la continuité uniforme et la continuité tout court, c'est que pour qu'une fonction soit uniformément continue, il faut que ce soit *le même* δ , pour tout x – tandis que pour qu'elle soit continue partout, il suffit que pour tout x , il existe un δ (potentiellement différent pour chaque x) tel que bla bla bla.

PROPOSITION B.18. *Soit $f : A \rightarrow \mathbb{R}$ avec A compact. Alors f est uniformément continue sur A , et elle y atteint un maximum et un minimum.*

PROPOSITION B.19 (Théorème des valeurs intermédiaires). *Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction continue et y compris entre $\min\{f(a), f(b)\}$ et $\max\{f(a), f(b)\}$. Alors, il existe un $c \in [a, b]$ tel que $f(c) = y$.*

B.5. Déivation

DÉFINITION B.23 (Limites de fonctions). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction, $a \in \mathbb{R}$.

Si il existe un L tel que pour toute suite $(x_n)_{n \in \mathbb{N}}$ monotone croissante convergeant vers a on a que $\lim_{n \rightarrow \infty} f(x_n) = L$, alors on dira que

$$\lim_{x \rightarrow a^-} f(x) = L.$$

Si il existe un L tel que pour toute suite $(x_n)_{n \in \mathbb{N}}$ monotone décroissante convergeant vers a , on a que $\lim_{n \rightarrow \infty} f(x_n) = L$, alors on dira que

$$\lim_{x \rightarrow a^+} f(x) = L.$$

Supposons que pour toute suite $(x_n)_{n \in \mathbb{N}}$ quelconque convergeant vers a , on a que $\lim_{n \rightarrow \infty} f(x_n)$ converge vers la même limite (disons L), alors on définit

$$\lim_{x \rightarrow a} f(x) = L.$$

DÉFINITION B.24 (Dérivée). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue. La dérivée de f au point x , notée par $f'(x)$, est définie par la limite

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h},$$

si cette dernière existe.

On note les dérivées successives $f^{(k)}(x)$, la k ième dérivée de f au point x .

PROPOSITION B.20 (Théorème des accroissements finis). *Soit $f : [a, b] \rightarrow \mathbb{R}$ une fonction dérivable (dont la dérivée existe en tout point de son domaine). Alors, il existe $c \in [a, b]$ tel que $f'(c) = \frac{f(b) - f(a)}{b - a}$.*

PROPOSITION B.21 (La règle de L'Hospital). *Soient $f, g : (a, b] \rightarrow \mathbb{R}$ différentiables, $c \in (a, b]$ (b, c potentiellement infinis), et supposons qu'on a*

$$\lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} g(x) = 0 \text{ ou } \infty,$$

et que

$$\lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$$

existe, alors on a que

$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}.$$

B.6. Intégration au sens de Riemann

DÉFINITION B.25 (Somme de Riemann). Soit $\mathcal{P} = \{[a_0, a_1], (a_1, a_2], (a_2, a_3], \dots, (a_{n-1}, a_n]\}$ une partition de l'intervalle $[a, b]$ en intervalles consécutifs (c.à.d. avec $a = a_0 < a_1 < \dots < a_{n-1} < a_n = b$). On dit que $\mathbf{x} = (x_1, x_2, x_3, \dots, x_n)$ est un **vecteur adapté** à \mathcal{P} si \mathbf{x} est tel que $x_i \in (a_{i-1}, a_i)$ pour tout $1 \leq i \leq n$.

Le **rayon** de \mathcal{P} , noté $|\mathcal{P}|$, est

$$|\mathcal{P}| := \max\{a_k - a_{k-1} : 1 \leq k \leq n\}.$$

La **somme de Riemann** de la fonction $f : [a, b] \rightarrow \mathbb{R}$ sur l'intervalle $[a, b]$, pour la partition \mathcal{P} et le vecteur \mathbf{x} est

$$\mathcal{R}(f, \mathcal{P}, \mathbf{x}) = \sum_{i=1}^n f(x_i)(a_i - a_{i-1}).$$

DÉFINITION B.26 (Intégrale de Riemann). Soit une fonction $f : [a, b] \rightarrow \mathbb{R}$ une fonction. On dit que f est **intégrable** au sens de Riemann sur $[a, b]$ si il existe un L tel que pour tout $\epsilon > 0$ il existe un $\delta > 0$ tel que pour toute partition \mathcal{P} avec $|\mathcal{P}| < \delta$, et pour tout vecteur \mathbf{x} adapté à \mathcal{P} , on a

$$|\mathcal{R}(f, \mathcal{P}, \mathbf{x}) - L| < \epsilon.$$

On dit alors que L est l'**intégrale** de f et on note²

$$L = \int_a^b f(t)dt.$$

PROPOSITION B.22 (Propriétés de l'intégrale). Soit $f, g : [a, b] \rightarrow \mathbb{R}$ deux fonctions. On a les propriétés suivantes :

- i. Si f est intégrable, $|f|$ l'est aussi.
- ii. Si f et g sont intégrables, fg est intégrable.
- iii. f est intégrable si et seulement si elle a un nombre fini de discontinuités.
- Si f et g sont intégrables, et que $c \in (a, b)$, et que $\lambda \in \mathbb{R}$, alors
- i. $\int_a^c f(t)dt + \int_c^b f(t)dt = \int_a^b f(t)dt$.
- ii. $\int_a^b f(t)dt = - \int_b^a f(t)dt$.
- iii. $\int_a^b (\lambda f(t) + g(t))dt = \lambda \int_a^b f(t)dt + \int_a^b g(t)dt$ (Linéarité)
- iv. Si $g(x) \leq f(x)$ alors $\int_a^b g(t)dt \leq \int_a^b f(t)dt$

THÉORÈME B.2 (Théorème fondamental du calcul). Soit $F : [a, b] \rightarrow \mathbb{R}$ une fonction différentiable, et $f = F'$ continue par morceaux sur (a, b) . Alors, avec $x \in (a, b]$,

$$F(x) = F(a) + \int_a^x f(t)dt.$$

DÉFINITION B.27 (Primitive). Soit $f : (a, b) \rightarrow \mathbb{R}$ une fonction. On dit que $F : [a, b] \rightarrow \mathbb{R}$ est une primitive de f si $F' = f$ sur (a, b) .

PROPOSITION B.23 (Méthodes d'intégration). Soient $f, g : [a, b] \rightarrow \mathbb{R}$ intégrables. Alors,

2. Le t dans cette expression est une variable muette et n'importe quel symbole peut lui être substitué sans problème.

- i. On peut faire un changement de variables : soit $\phi : [c, d] \rightarrow [a, b]$ une fonction différentiable et ϕ' intégrable sur $[c, d]$.

$$\int_c^d f(\phi(t))\phi'(t)dt = \int_{\phi(c)}^{\phi(d)} f(u)du.$$

- ii. On peut faire une intégrale par parties : Si $f, g : [a, b] \rightarrow \mathbb{R}$ sont continues et différentes sur (a, b) et que f' et g' sont intégrables, alors

$$\int_a^b f(t)g'(t)dt = [f(t)g(t)]_a^b - \int_a^b f'(t)g(t)dt.$$

B.7. Symétries de fonctions.

B.7.1. Fonctions paires et impaires.

DÉFINITION B.28 (Fonction paire, impaire). Soit $f : I \rightarrow \mathbb{R}$ une fonction.

On dit que f est **paire** si

- $I = -I := \{-x : x \in I\}$ — c'est à dire si I est un domaine symétrique,
ET

- Pour tout $x \in I$, $f(x) = f(-x)$.

On dit que f est **impaire** si

- $I = -I$

ET

- Pour tout $x \in I$, $f(-x) = -f(x)$.

PROPOSITION B.24 (Décomposition en parties paires et impaires). Soit $f : I \rightarrow \mathbb{R}$ avec $I = -I$.

Alors,

- i. $f_0(x) = \frac{1}{2}(f(x) + f(-x))$ est une fonction paire ;
- ii. $f_1(x) = \frac{1}{2}(f(x) - f(-x))$ est une fonction impaire ;
- iii. $f = f_0 + f_1$.

PROPOSITION B.25 (Héritages des parités.). Soient $f, g : I \rightarrow \mathbb{R}$ des fonctions impaires, et $h, k : I \rightarrow \mathbb{R}$ des fonctions paires.

- i. $f + g$ est paire ;
- ii. $h + k$ est paire ;
- iii. fg et hk sont paires ;
- iv. fh est paire
- v. $f \circ h$ et $h \circ f$ sont paires
- vi. $f \circ g$ est paire.
- vii. $h \circ k$ est paire.

PROPOSITION B.26 (Intégration de fonctions paires, impaires). Soient $f, g : I \rightarrow \mathbb{R}$ deux fonctions, avec f paire et g impaire, toutes deux intégrables sur I , et $I = (-a, a)$ avec $0 < a \leq \infty$.

- i. $\int_{-a}^a f(t)dt = 2 \int_0^a f(t)dt$;
- ii. $\int_{-a}^a g(t)dt = 0$.

B.7.2. Fonctions périodiques.

DÉFINITION B.29 (Fonction périodique). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. On dit que f est **périodique** si il existe un $p > 0$ tel que pour tout $x \in \mathbb{R}$,

$$f(x + p) = f(x).$$

On dit que p est une **période** de f .

PROPOSITION B.27 (Héritage des périodicités). Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions périodiques, soit p une période de f et q une période de g . Soit $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction.

Alors, si $p/q \in \mathbb{Q}$, on note $r = \text{PPCM}(p, q)$, et la fonction $x \mapsto h(f(x), g(x))$ est périodique et r en est une période.

PROPOSITION B.28 (Intégrales des fonctions périodiques). Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction périodique intégrable sur n'importe quel intervalle fini (localement intégrable). Soit p une période de f .

Alors, pour tout $a \in \mathbb{R}$, $n \in \mathbb{Z}$,

$$\int_a^{a+np} f(t)dt = n \int_0^p f(t)dt.$$

B.8. Les fonctions spéciales

B.8.1. La fonction Gamma.

DÉFINITION B.30 (Fonction Gamma). On définit :

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt,$$

avec $\Re(z) > 0$ (la partie réelle de z est positive).

PROPOSITION B.29 (Propriétés de la fonction Gamma). La fonction Γ a les propriétés suivantes :

i. Pour tout z , on a que

$$\Gamma(z + 1) = z\Gamma(z).$$

ii. En particulier, pour $n \in \mathbb{N}$,

$$\Gamma(n + 1) = n!.$$

iii. On a que pour $z \notin \mathbb{Z}$,

$$\Gamma(1 - z)\Gamma(z) = \frac{\pi}{\sin(\pi z)}.$$

iv. En particulier, pour $z = 1/2$, on a que

$$\Gamma(1/2) = \sqrt{\pi}.$$

B.9. Les convolutions

DÉFINITION B.31. Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions intégrables sur \mathbb{R} . On dénote par

$$f * g(x) := \int_{-\infty}^{\infty} f(t)g(x-t)dt$$

la **convolution** des fonctions f et g .

PROPOSITION B.30. Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions intégrables sur \mathbb{R} . On a les propriétés suivantes :

- i. $f * g = g * f$.
- ii. $\int_{\mathbb{R}} f(t)dt \cdot \int_{\mathbb{R}} g(u)du = \int_{\mathbb{R}} f * g(v)dv$.

B.10. Les fonctions convexes

DÉFINITION B.32. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$. On dit que f est **convexe** si pour tout $t \in (0, 1)$ et pour toute paire $x, y \in \mathbb{R}$, on a

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

PROPOSITION B.31. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$. Les énoncés suivants sont équivalents :

- i. f est convexe.
- ii. Si f est deux fois différentiable, f'' est positive.
- iii. Pour tout vecteur (t_1, t_2, \dots, t_n) où $t_i \in (0, 1)$ pour tout i $t_1 + t_2 + \dots + t_n = 1$, et pour tout vecteur $(x_1, x_2, x_3, \dots, x_n) \in \mathbb{R}^n$,

$$f\left(\sum_{i=1}^n t_i x_i\right) \leq \sum_{i=1}^n t_i f(x_i).$$

- iv. Tout segment de droite reliant deux points du graphe de la fonction f se situe au-dessus du graphe de la fonction f .

PROPOSITION B.32. Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions convexes, et g croissante. Alors, on a les propriétés suivantes :

- i. f est continue et différentiable sauf pour un nombre dénombrable de points.
- ii. $f \circ g$ est convexe.
- iii. $f \circ (-g)$ est concave.

B.11. Les notations asymptotiques

Soient $f, g : \mathbb{R} \rightarrow \mathbb{R}$ deux fonctions. On introduit les notations suivantes :

(B.11.1)

$$f(x) = o(g(x)), \quad (x \rightarrow a) \quad \Leftrightarrow \quad \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 0.$$

(B.11.2)

$$f(x) = O(g(x)), \quad (x \rightarrow a) \quad \Leftrightarrow \quad \exists \epsilon > 0, C > 0 : \forall x \in (a - \epsilon, a + \epsilon) \quad |f(x)| \leq C |g(x)|.$$

(B.11.3)

$$f(x) \sim g(x), \quad (x \rightarrow a) \quad \Leftrightarrow \quad \lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1.$$

Toutes ces notations fonctionnent si on remplace a par $+\infty$ – dans le cas de du « grand O », on dit que $f(x) = O(g(x))$ ($x \rightarrow \infty$) si il existe N et C tel que $|f(x)| \leq C|g(x)|$ pour tout $x > N$. La définition est exactement analogue pour $-\infty$.

Par exemple, une fonction bornée es $O(1)$. Une fonction qui tend vers 0 est $o(1)$.

Si $f(x) \sim g(x)$, alors $f(x) - g(x) = o(g(x))$ et $f(x) - g(x)$ est aussi $o(f(x))$. On n'a pas toujours que $f(x) - g(x) = o(1)$, par contre. Par exemple, $x \sim x + 1$, mais $x + 1 - x = 1$, et $1 \neq o(1)$.

PROPOSITION B.33 (Arithmétique des asymptotiques). *i. Si $f(x) = o(g(x))$ et $g(x) = o(h(x))$ alors $f(x) = o(h(x))$.*

ii. Si $f(x) = O(g(x))$ et $g(x) = O(h(x))$, alors $f(x) = O(h(x))$.

iii. Si $f(x) = O(g(x))$ et $g(x) = o(h(x))$, alors $f(x) = o(h(x))$.

iv. La relation \leq_O définie par $f(x) \leq_O g(x) \Leftrightarrow f(x) = O(g(x))$ est un ordre partiel sur les expressions dépendant de x .

v. Si $f(x) = o(h(x))$ et $g(x) = o(h(x))$, alors $f(x) + g(x) = o(h(x))$.

vi. Si $f(x) = o(g(x))$, alors $f(x)^\alpha = o(g(x)^\alpha)$ pour tout α .

vii. Si $f(x) = o(h(x))$ et $g(x) = o(k(x))$, alors $f(x)g(x) = o(h(x)k(x))$

viii. Si $f(x) = O(h(x))$ et $g(x) = O(h(x))$, alors $f(x) + g(x) = O(h(x))$.

ix. Si $f(x) = O(h(x))$ et $g(x) = O(k(x))$, alors $f(x)g(x) = O(g(x)k(x))$.

x. Si il existe $C > 0$ tel que $|f(x)| \sim C|g(x)|$, alors $f(x) = O(g(x))$ et $g(x) = O(f(x))$.

xi. $f(x) \sim g(x)$ si et seulement si $|f(x) - g(x)| = o(f(x))$.

PROPOSITION B.34 (Asymptotique et convergence.). *i. Si $a_n = O(b_n)$ et $\sum_n b_n$ converge, $\sum_n a_n$ converge.*

ii. Si $f(x) = O(g(x))$, $x \rightarrow a$ et $\int_a^b g(t)dt$ converge, $\int_a^b f(t)dt$ converge.

iii. Si $a_n = O(b_n)$ et $\sum_n a_n$ diverge, $\sum_n b_n$ diverge.

iv. Si $f(x) = O(g(x))$, $x \rightarrow a$ et $\int_a^b f(t)dt$ diverge, $\int_a^b g(t)dt$ diverge.

B.12. Bonus : L'intégrale de Stieltjes

Cette section bonus présente rapidement la notion d'Intégrale de Stieltjes, de façon relativement générale. L'objectif est de donner une définition plus unifiée de l'espérance

Commençons par introduire les fonctions càdlàg :

DÉFINITION B.33 (Fonction càdlàg). Soient I un intervalle et $F : I \rightarrow \mathbb{R}$ une fonction. On dit que F est une fonction **càdlàg** (continue à droite avec limite à gauche) sur l'intervalle I si, pour tout $x \in I$

i. $\lim_{y \rightarrow x^+} F(y) = F(x)$;

ii. $\lim_{y \rightarrow x^-} F(y)$ existe;

Si F est une fonction càdlàg, on notera $F(x^-) = \lim_{y \rightarrow x^-} F(y)$.

PROPOSITION B.35. Soient I un intervalle et $F : I \rightarrow \mathbb{R}$ une fonction càdlàg, et soit D l'ensemble des points de discontinuités de F dans I .

Alors,

- i. D est au plus dénombrable ;
- ii. Soit $x_0 \in I$ un point d'accumulation de D . Alors, pour tout $\epsilon > 0$ il existe $\delta > 0$ tel que pour tout point de discontinuité $x \in (x_0 - \delta, x_0 + \delta) \cap D$, $|F(x) - F(x^-)| < \epsilon$;
- iii. En particulier, si $\inf_{x \in D} |F(x) - F(x^-)| > 0$, il suit que D n'admet aucun point d'accumulation.

DÉFINITION B.34 (Intégrale de Stieltjes). Soit $g : I \rightarrow \mathbb{R}$ une fonction quelconque sur l'intervalle fermé $I = [a, b]$, et $F : I \rightarrow \mathbb{R}$ une fonction càdlàg.

De façon analogue aux sommes de Riemann (définition B.25), pour \mathcal{P} une partition de I en n sous-intervalles et un vecteur \mathbf{x} adapté à la partition \mathcal{P} , on considère les sommes

$$\mathcal{S}(g, \mathcal{P}, \mathbf{x}, F) = \sum_{i=1}^n g(x_i)(F(a_i) - F(a_{i-1})).$$

On dit que g est **intégrable contre F sur l'intervalle I au sens de Stieltjes** si il existe un L tel que pour tout $\epsilon > 0$ il existe un $\delta > 0$ tel que pour toute partition \mathcal{P} avec un rayon $|P| < \delta$ et pour tout vecteur \mathbf{x} adapté à \mathcal{P} , on a

$$|\mathcal{S}(g, \mathcal{P}, \mathbf{x}, F) - L| < \epsilon.$$

Dans ce cas, on note

$$L = \int_a^b g(x)dF(x).$$

PROPOSITION B.36. Soit $F : I = [a, b] \rightarrow \mathbb{R}$ une fonction càdlàg continument dérivable sur I et soit $g : I \rightarrow \mathbb{R}$ une fonction intégrable contre F sur I au sens de Stieltjes.

Alors,

$$(B.12.1) \quad \int_a^b g(t)dF(t) = \int_a^b g(t)F'(t)dt.$$

PROPOSITION B.37. Soit $F : I \rightarrow \mathbb{R}$ une fonction càdlàg en escalier, et soit D l'ensemble des points de discontinuité de F . Soit $g : I \rightarrow \mathbb{R}$ intégrable contre F au sens de Stieltjes sur l'intervalle I .

Alors,

$$(B.12.2) \quad \int_a^b g(t)dF(t) = \sum_{x \in D} g(x)(F(x) - F(x^-)).$$

PROPOSITION B.38. Soit $F : I \rightarrow \mathbb{R}$ une fonction càdlàg continûment dérivable sauf sur l'ensemble D de ses points de discontinuité. On notera

$$f(x) = \begin{cases} F'(x) & x \notin D \\ 0 & x \in D \end{cases}.$$

Soit $g : I \rightarrow \mathbb{R}$ une fonction intégrable contre F au sens de Stieltjes sur l'intervalle I .

Alors,

$$(B.12.3) \quad \int_a^b g(t)dF(t) = \sum_{x \in D} g(x)(F(x) - F(x^-)) + \int_a^b g(t)f(t)dt.$$

REMARQUE. Comme nous l'avons montré à la proposition 4.1, la fonction de répartition F d'une variable aléatoire X est càdlàg. De façon générale, l'espérance est simplement

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) dF(x),$$

peu importe que la variable soit discrète ou continue.