

Sorbonne Université
Master MEEF 1^{ère} année
2019 - 2020

Probabilités et statistiques

Raphaël Roux *

*. raphael.roux@upmc.fr

Table des matières

0	Quelques observations	7
0.1	Fréquences empiriques	7
0.1.1	Convergence de la moyenne observée	7
0.1.2	Écarts à la moyenne théorique	8
0.2	Marches aléatoires	10
0.2.1	La marche aléatoire simple	10
0.2.2	La marche aléatoire biaisée	13
0.2.3	La marche aléatoire de Cauchy	13
1	Formalisme de la théorie des probabilités	17
1.1	Introduction	17
1.2	Espaces probabilisés	18
1.3	Variables aléatoires	22
1.4	Variables aléatoires réelles	23
1.5	Exercices	26
2	Espérance d'une variable aléatoire	27
2.1	Définition de l'espérance	27
2.1.1	Le cas des variables bornées	28
2.1.2	Le cas des variables positives	30
2.1.3	Le cas des variables intégrables	31
2.2	Théorème de transfert	32
2.3	Variance, moments d'ordres supérieurs	32
2.4	Inégalité de Markov et de Bienaymé-Tchebychev	36
2.4.1	Inégalité de Markov	36
2.4.2	Inégalité de Bienaymé-Tchebychev	37
2.5	Exercices	38
3	Probabilités conditionnelles - indépendance	39
3.1	Probabilité conditionnelle	39
3.1.1	Définition	39
3.1.2	Arbres de probabilité	40
3.1.3	Formule des probabilités totales et formule de Bayes	42
3.2	Indépendance d'évènements	43

3.3	Variables aléatoires indépendantes	45
3.4	Exercices	47
4	Variables aléatoires discrètes - dénombrement	51
4.1	Espaces probabilisés dénombrables (ou finis)	51
4.1.1	Indépendance	52
4.2	Lois discrètes usuelles	53
4.2.1	Loi uniforme	53
4.2.2	Loi de Bernoulli	53
4.2.3	Loi binomiale	54
4.2.4	Loi de Poisson	55
4.2.5	Loi géométrique	55
4.3	Dénombrement	57
4.3.1	Tirages ordonnés avec remise	57
4.3.2	Tirages ordonnés sans remise	57
4.3.3	Tirages non ordonnés sans remise	57
4.3.4	Tirages non ordonnés avec remise	59
4.4	Exercices	59
5	Variables aléatoires continues	63
5.1	Variables aléatoires réelles à densité	63
5.1.1	Définitions	63
5.2	Lois à densité usuelles	64
5.2.1	Loi uniforme	65
5.2.2	Loi de Gauss ou loi normale	65
5.2.3	Loi exponentielle	66
5.2.4	Loi de Cauchy	67
5.3	Exercices	68
6	Suites de variables aléatoires - théorèmes limites	71
6.1	Suites de variables aléatoires	71
6.1.1	Convergence de suites de variables aléatoires réelles	71
6.2	Loi des grands nombres	73
6.3	Théorème limite central	74
6.4	Chaînes de Markov	75
6.5	Exercices	77
7	Estimation	79
7.1	Estimation d'une proportion	79
7.2	Intervalles de confiance	81
7.2.1	Par l'inégalité de Bienaymé-Tchebychev	81
7.2.2	Dans le cas Gaussien	82
7.2.3	Par le théorème limite central	82
7.3	Exercices	82

<i>TABLE DES MATIÈRES</i>	5
8 Tests d'hypothèses	85
8.1 Exemple introductif	85
8.2 Test pour la moyenne	86
8.3 Exercices	87
A Indications pour les exercices	89

Chapitre 0

Quelques observations

Dans ce chapitre, nous n'allons pas faire de théorie sur le calcul des probabilités, mais plutôt décrire quelques "expériences", réalisables avec un ordinateur (ou avec un dé et beaucoup de patience), qui vont illustrer les principaux résultats qui seront formalisés rigoureusement dans les prochains chapitres.

Pour être précis, nous allons étudier la sortie de programmes représentant des lancers de dés. Nous allons supposer qu'un ordinateur est capable de reproduire fidèlement les lancers d'un dé équilibré, ce qui n'est en réalité le cas qu'approximativement ^(note 1).

0.1 Fréquences empiriques

On va tout d'abord s'intéresser à la proportion des différents résultats obtenus par plusieurs lancers de dés successifs.

0.1.1 Convergence de la moyenne observée

Considérons une suite infinie de lancers de dés. On note $(X_n)_{n \in \mathbf{N}}$ la suite (aléatoire) des valeurs obtenues. Une réalisation de la suite $(X_n)_{n \in \mathbf{N}}$ ressemble donc à

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

Si, par exemple, on s'intéresse à la proportion de "6" parmi les résultats obtenus, on observe ceci :

6 · · 6 6 · · · · · 6 · · · · · 6 · · · · · 6 · 6 · · · · ·

L'intuition nous dit alors que "un dés sur six va donner le résultat 6". Pour formaliser cette affirmation, on peut s'intéresser à la proportion de "6" parmi les n premiers tirages. Définissons la suite $(\mathbf{1}_{X_n=6})_{n \in \mathbf{N}}$ par

$$\mathbf{1}_{X_n=6} = \begin{cases} 1 & \text{si } X_n = 6, \\ 0 & \text{sinon.} \end{cases}$$

La proportion de "6" parmi les n premiers tirages est alors donnée par $\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=6}$. Traçons le graphe de $n \mapsto \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=6}$ pour une suite aléatoire $(X_n)_{n \in \mathbf{N}}$ fixée. Le résultat est donné en figure 0.1.

(note 1). Un ordinateur ne sait produire que des calculs déterministes. En revanche, il existe des algorithmes produisant des suites déterministes mais *ressemblant* à des suites aléatoires.

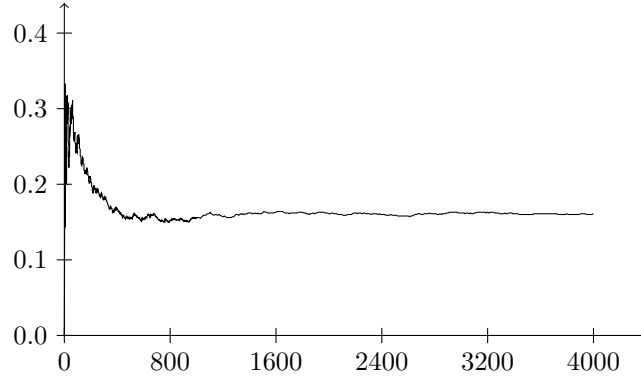


FIGURE 0.1 – Proportion de “6”, en fonction du nombre de lancers. En abscisse, le nombre de lancers ; en ordonnée, la proportion de “6” (qui est dans $[0, 1]$).

À première vue, cette suite *semble* converger. Deux remarques :

- Cette convergence, si elle peut paraître intuitive, n’est *absolument pas évidente*, et il est difficile d’explicitier un argument prouvant la convergence. Sans analyse mathématiques plus poussée, on pourrait a priori envisager que la suite $(\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=6})_{n \in \mathbf{N}}$ oscille sans jamais se stabiliser.
- On observe ici *une* réalisation de la suite $(X_n)_{n \in \mathbf{N}}$, pour laquelle on obtient une certaine valeur limite λ . A priori, il n’est pas clair que la valeur de λ soit toujours la même. On pourrait imaginer que quelle que soit la suite $(X_n)_{n \in \mathbf{N}}$ obtenue en lançant des dés, la suite $(\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=6})_{n \in \mathbf{N}}$ converge, mais que sa limite dépende de la suite de lancers de dés.

On verra plus tard que la suite est effectivement convergente, et que sa limite sera toujours la même. Ce résultat est la *loi des grands nombres*. Si l’on admet cela, on peut dans ce cas donner la valeur de la limite λ . En effet, par symétrie, les proportions de chacune des six valeurs $\{1, 2, 3, 4, 5, 6\}$ obtenues doivent toutes converger vers λ :

$$\text{pour } i = 1, \dots, 6, \quad \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=i} \xrightarrow[n \rightarrow \infty]{} \lambda.$$

De plus, la somme des six proportions est toujours égale à 1. Par conséquent, on a $6\lambda = 1$, d’où $\lambda = 1/6$. On retrouve l’intuition “on obtient le résultat 6 avec probabilité $1/6$ ”.

Si on affiche, pour plusieurs suites de lancers de dés, la convergence de la proportion de “6”, on obtient le graphe de la figure 0.2. On observe effectivement, dans tous les cas, la convergence vers la valeur fixée $1/6$ (ligne pointillée).

0.1.2 Écarts à la moyenne théorique

On a remarqué dans la partie précédente que la proportion de 6 *convergeait*, quand le nombre de lancers tend vers l’infini, vers une valeur fixée. Toutefois, sur un nombre *donné* de lancers, il n’y a aucune raison que la proportion de 6 soit exactement $1/6$. On peut donc légitimement se poser la question suivante : pour un nombre n de lancers, à quel écart peut-on s’attendre entre la valeur théorique et la

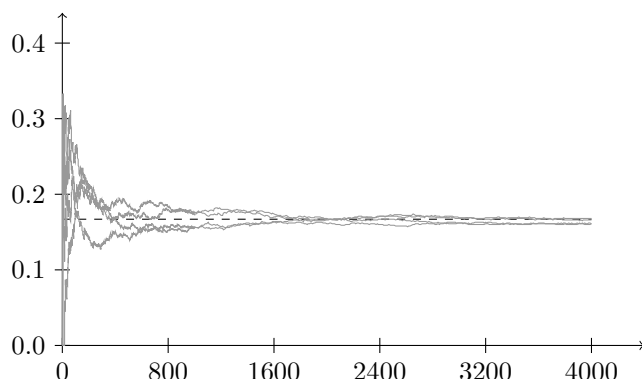


FIGURE 0.2 – Proportion de “6” sur quatre suites de lancers, avec comparaison à la valeur théorique $1/6$.

valeur observée? Plus précisément : la suite $\left| \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=6} - \frac{1}{6} \right|$ tend vers 0 quand $n \rightarrow \infty$; à quelle vitesse cette convergence a-t-elle lieu?

Pour répondre à cette question, on fait l’expérience suivante : pour un certain nombre de valeurs de n (ici, $n = 8, 16, 32, 64, \dots, 4096$ — chaque nombre est le double du précédent), on calcule un grand nombre de fois l’écart $\left| \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k=6} - \frac{1}{6} \right|$ (ici, 1000 fois), et on calcule la valeur moyenne de ces écarts, que l’on note ε_n . Autrement dit,

$$\varepsilon_n = \frac{1}{1000} \sum_{i=1}^{1000} \left| \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X_k^{(i)}=6} - \frac{1}{6} \right|,$$

où pour chaque $1 \leq i \leq 1000$, $(X_n^{(i)})_{n \in \mathbf{N}}$ désigne une suite de lancers de dés. On trace alors la courbe $n \rightarrow \varepsilon_n$, et on obtient le graphique représenté en figure 0.3, à gauche.

L’erreur semble bien tendre vers 0. Toutefois, on aimerait un résultat plus quantitatif, par exemple avoir une décroissance de la forme $\varepsilon_n \simeq Cn^{-\alpha}$ pour certaines constantes C et α . Pour voir si une telle relation existe, on utilise la transformation suivante : si ε_n est proche de $Cn^{-\alpha}$, on aura $\ln(\varepsilon_n) \simeq \ln C - \alpha \ln(n)$, de sorte que les couples $(\ln(n), \ln(\varepsilon_n))$ seront alignés sur la droite d’équation $y = \ln C - \alpha x$ (noter que α est le coefficient directeur de la droite). On a donc tracé sur le graphe de droite de la figure 0.3 les couples $(\ln(n), \ln(\varepsilon_n))$.

On obtient effectivement un tracé très proche d’une droite pour laquelle une régression linéaire donne l’équation $y = -1.204 - 0.501x$. En particulier, le coefficient directeur est très proche de $-1/2$, ce qui donnerait une décroissance de l’ordre de

$$\varepsilon_n \simeq \frac{C}{\sqrt{n}}.$$

On verra plus tard que l’erreur moyenne est en fait donnée par

$$\varepsilon_n \sim \sqrt{\frac{5}{18\pi n}}. \quad (0.1)$$

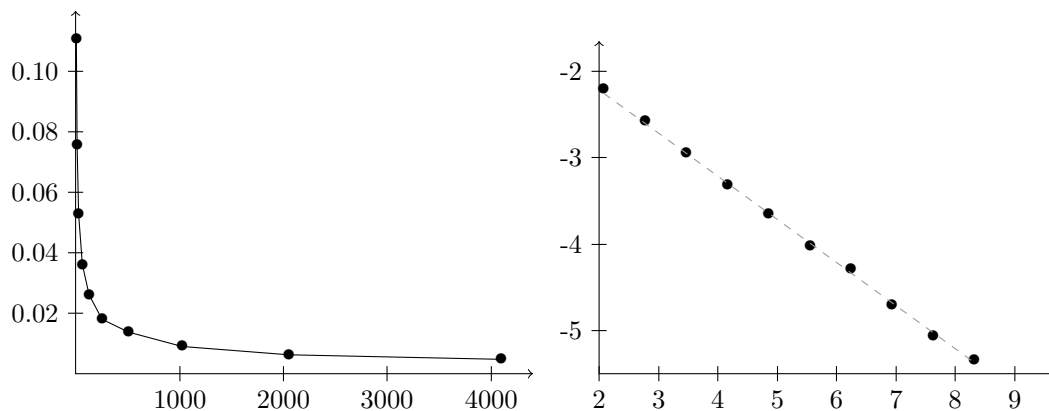


FIGURE 0.3 – Erreur moyenne ε_n entre la proportion de “6” obtenus sur n lancers et la valeur théorique $1/6$ (moyenne sur 1000 observations). À gauche : ε_n en fonction de n . À droite : tracé logarithmique ($\ln(\varepsilon_n)$ en fonction de $\ln n$), avec en pointillé la droite théorique $y = \ln\left(\sqrt{\frac{5}{18\pi}}\right) - \frac{1}{2}x$.

L’expression exacte de la constante $C = \sqrt{\frac{5}{18\pi}}$ n’a que peu d’intérêt. Le point important à noter ici est que l’erreur moyenne décroît comme $1/\sqrt{n}$.

0.2 Marches aléatoires

Dans cette partie, on va s’intéresser à l’allure de différentes trajectoires obtenues à partir de la suite $(X_n)_{n \in \mathbb{N}}$ des lancers de dés.

0.2.1 La marche aléatoire simple

On définit ce qu’on appelle une *marche aléatoire* de la façon suivante. On définit $Z_0 = 0$, puis par récurrence,

$$Z_{n+1} = \begin{cases} Z_n + 1 & \text{si } X_n \in \{1, 2, 3\}, \\ Z_n - 1 & \text{si } X_n \in \{4, 5, 6\}. \end{cases}$$

Le variable n est à comprendre comme une variable de temps. On part à l’instant initial $n = 0$ de la position 0, puis à chaque unité de temps, on se déplace de $+1$ ou de -1 , en choisissant aléatoirement et symétriquement. Notamment, au bout de N pas, la marche se situera quelque part entre $-N$ et N . Un exemple de 20 pas d’une trajectoire est présenté en figure 0.4. En abscisse, on a représenté la variable de temps n et en ordonnée, la position Z_n . La grille grise représente l’ensemble des points qui sont atteignables par la marche. La marche part initialement de la pointe, à gauche de la grille. À chaque pas de temps, la marche se déplace vers la droite soit en montant, soit en descendant.

Une question naturelle est de savoir quelle est la position la plus probable pour la marche au bout de N pas, pour N fixé. Sur la figure 0.5, où l’on a tracé 10 réalisations des 20 premiers pas de la marche, on observe que la marche semble arriver plus souvent près de 0 que de $\pm N$. La raison en est qu’il y

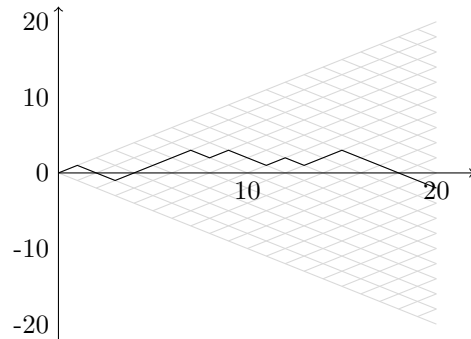


FIGURE 0.4 – Trajectoire d'une marche aléatoire.

a beaucoup plus de trajectoires possibles qui finissent près de 0 que de trajectoires qui finissent près de $\pm N$. Par exemple, il n'existe qu'une seule trajectoire qui arrive à N (il s'agit de celle qui ne fait que monter) ou à $-N$ (celle qui ne fait que descendre).

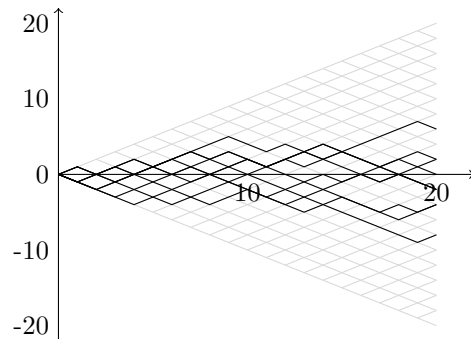


FIGURE 0.5 – Trajectoires de 10 marches aléatoires.

Analysons plus précisément ce phénomène : le tableau suivant donne la répartition des positions de 10000 marches au bout de $N = 20$ pas.

Position	-20	-18	-16	-14	-12	-10	-8	-6	-4	-2	0
Effectif	0	1	2	9	63	159	349	734	1183	1594	1752
Position	2	4	6	8	10	12	14	16	18	20	
Effectif	1606	1195	789	356	159	45	12	3	0	0	

Ce tableau est tracé sous forme d'histogramme sur la figure 0.6. On remarque d'une part que la trajectoire est beaucoup plus souvent proche de 0 que de $\pm N$, mais aussi que la probabilité de se trouver en position k semble dépendre de k de manière très régulière. Il s'agit là d'un premier aperçu du *théorème limite central*, que nous verrons au chapitre 6 (théorème 6.3.1).

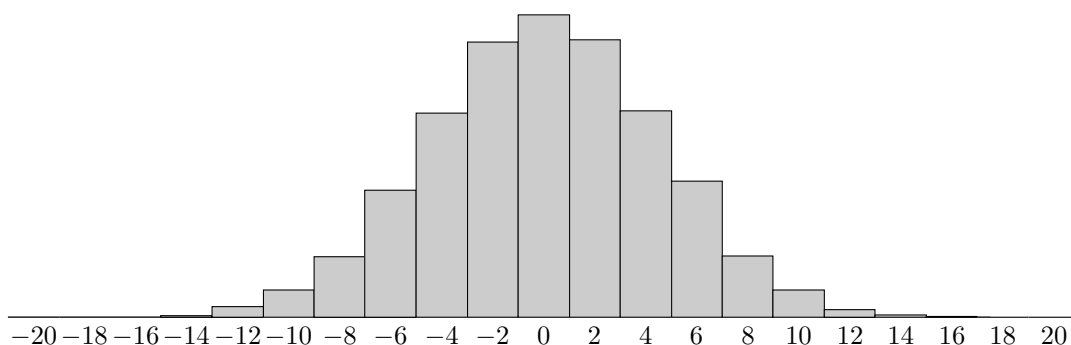


FIGURE 0.6 – Histogramme des points d'arrivée de 10000 marches aléatoires de 20 pas.

On peut alors se poser la question suivante : quand n tend vers l'infini, quel est l'ordre de grandeur de Z_n ? Par exemple, existe-t-il une suite déterministe $(u_n)_{n \in \mathbf{N}}$ telle que $(Z_n/u_n)_{n \in \mathbf{N}}$ admette une limite ? On a tracé à gauche en figure 0.7 la même chose qu'en figure 0.4, mais avec plus de pas : on est allé jusqu'à 400 pas. Par rapport aux positions atteignables extrêmes ± 400 , la trajectoire semble très proche de 0. Notamment, on peut s'attendre à ce que pour n grand, Z_n soit négligeable devant n .

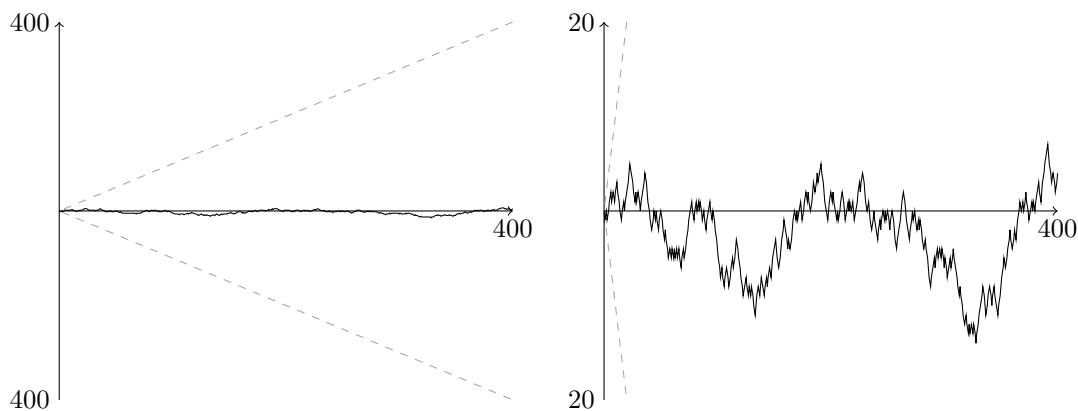


FIGURE 0.7 – Trajectoire de $(Z_n)_{n \in \mathbf{N}}$ pour $n = 400$. En pointillés, la limite de la zone atteignable par la trajectoire.

L'explication est que pour n grand, Z_n n'est pas de l'ordre de n , mais de l'ordre de \sqrt{n} . Sur le graphe de droite de la figure 0.7, on a renormalisé les ordonnées, de sorte à afficher les positions entre $-\sqrt{n}$ et \sqrt{n} (on a $\sqrt{400} = 20$). À cette échelle-là on observe une trajectoire continue aléatoire. Autrement dit, sur une échelle de temps de l'ordre de n , les variations de Z_n sont de l'ordre de \sqrt{n} . On retiendra que pour n grand, $\sum_{k=1}^n Z_k$ a un comportement aléatoire de l'ordre de \sqrt{n} .

0.2.2 La marche aléatoire biaisée

Dans cette partie, on va étudier une suite $(Y_n)_{n \in \mathbf{N}}$ dont la définition sera légèrement différente de celle de la suite $(Z_n)_{n \in \mathbf{N}}$. Plus précisément, on va poser $Y_0 = 0$ et

$$Y_{n+1} = \begin{cases} Y_n + 1 & \text{si } X_n \in \{1, 2, 3, 4\} \\ Y_n - 1 & \text{si } X_n \in \{5, 6\}. \end{cases}$$

Autrement dit, les trajectoires possibles de $(Y_n)_{n \in \mathbf{N}}$ sont les mêmes que celles de $(Z_n)_{n \in \mathbf{N}}$, mais les probabilités entre les deux directions ne sont plus symétriques : on va choisir le déplacement $+1$ deux fois plus souvent. On a tracé une trajectoire de $(Y_n)_{n \in \mathbf{N}}$ sur la figure 0.8, jusqu'à $n = 400$. On voit clairement une différence avec la figure 0.7 : la trajectoire est très proche d'une trajectoire déterministe non-nulle, qui se trouve être la suite $\frac{n}{3}$.

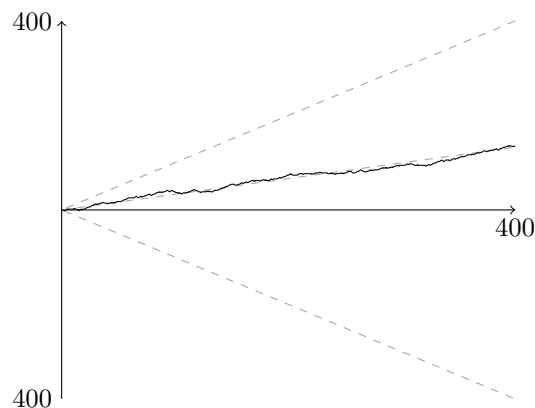


FIGURE 0.8 – Une trajectoire de la marche $(Y_n)_{n \in \mathbf{N}}$. En pointillés gris, les limites de la zone atteignable par la trajectoire, ainsi que la droite d'équation $y = x/3$.

Pour autant, la trajectoire Y_n est aléatoire, et présente des variations autour de $\frac{n}{3}$. Pour illustrer ces variations, il suffit de tracer la suite des points $Y_n - \frac{n}{3}$, ce qui est fait sur la figure 0.9.

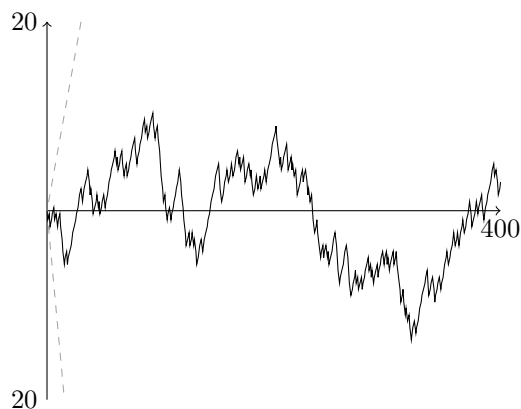
On observe ici une trajectoire visuellement très similaire à celle de la figure 0.7. En particulier, $Y_n - \frac{n}{3}$ est de l'ordre de \sqrt{n} pour n grand. On est alors dans la situation suivante, qui se trouvera en fait être assez générale :

$$Y_n = \left(\begin{array}{c} \text{partie déterministe} \\ \text{d'ordre } n \end{array} \right) + \left(\begin{array}{c} \text{partie aléatoire} \\ \text{d'ordre } \sqrt{n} \end{array} \right). \quad (0.2)$$

Ici, la partie déterministe est $n/3$. Pour $(Z_n)_{n \in \mathbf{N}}$, la partie déterministe était en fait $0 \times n$, ce qui fait que le premier terme significatif était la partie aléatoire d'ordre \sqrt{n} .

0.2.3 La marche aléatoire de Cauchy

En équation (0.2), on donne une forme assez générale du comportement asymptotique des marches aléatoires. Toutefois, cette forme admet des exceptions, dont la marche que nous allons présenter dans cette partie. Cette marche est définie de la manière suivante : on pose encore $W_0 = 0$, mais W_{n+1} est

FIGURE 0.9 – Tracé de la trajectoire de $Y_n - \frac{n}{3}$.

construit à partir de W_n de sorte que l'angle Θ_n entre le vecteur $(1, W_{n+1} - W_n)$ et l'axe des abscisses soit choisi uniformément sur l'intervalle $[-\pi/2, \pi/2]$ (note 2). Pour être plus précis, remarquons que la tangente de l'angle entre le vecteur (x, y) et l'axe des abscisses est donné par $\frac{y}{x}$, de sorte que

$$\tan(\Theta_n) = \frac{W_{n+1} - W_n}{1}.$$

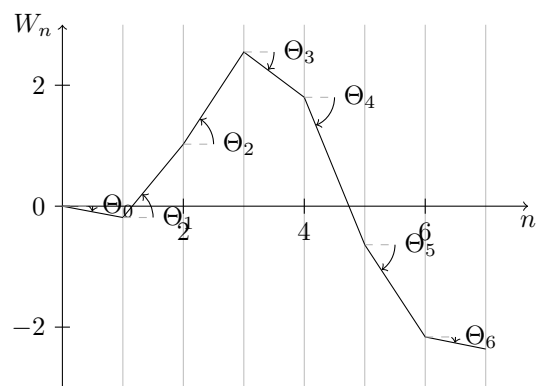
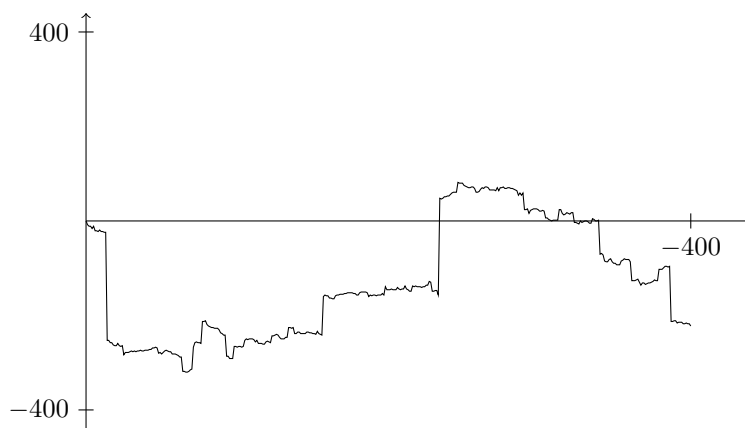
On va donc définir la suite $(W_n)_{n \in \mathbb{N}}$ par la formule

$$W_{n+1} = W_n + \tan(\Theta_n).$$

Un exemple de trajectoire obtenue est donné en figure 0.10.

Traçons maintenant une trajectoire jusqu'à $n = 400$. Le résultat est donné sur la figure 0.11. Cette fois-ci, on observe que le comportement aléatoire de W_n est de l'ordre de n . Contrairement à la décomposition (0.2), la marche W_n présente directement de l'aléa à l'échelle n . Par ailleurs (mais on ne verra pas l'explication de ce phénomène), la forme de la trajectoire sur la figure 0.11 est différente des trajectoires des figures 0.7 et 0.9 qui, elles, se ressemblent beaucoup. Par exemple, la trajectoire de la figure 0.11 n'est pas continue car elle présente des grands sauts.

(note 2). On n'a pas encore défini ce que signifiait "tirer un nombre uniformément dans un intervalle", et on ne peut pas construire Θ_n à partir de notre suite de lancers de dés, mais on va admettre qu'il y a un sens à "choisir une direction aléatoire sans privilégier de direction particulière". Un dispositif physique permettant ce type d'aléa serait une roue avec un rayon marqué, que l'on fait tourner et dont on attend l'arrêt complet. L'angle entre la position à la fin et la position initiale de la roue peut être considéré comme un angle aléatoire uniforme sur $[-\pi, \pi]$.

FIGURE 0.10 – Trajectoire de X_n , jusqu'à $n = 7$.FIGURE 0.11 – Trajectoire de $(W_n)_{n \in \mathbf{N}}$, jusqu'à $n = 400$.

Chapitre 1

Formalisme de la théorie des probabilités

1.1 Introduction

Le but de la théorie des probabilités est de donner un fondement mathématiques à la notion de hasard, dans le but de pouvoir définir et étudier précisément des modèles faisant intervenir de l'aléa. Le champ d'applications est ici très large : aujourd'hui des modèles probabilistes sont utilisés dans des domaines aussi variés que la physique, l'économie, la finance, la biologie ou la météorologie.

Le mot “hasard” désigne ici tout phénomène trop complexe pour être prévu précisément. De fait, il est souvent commode de considérer comme aléatoires des phénomènes qui sont pourtant parfaitement déterministes : si on lance un dé, la trajectoire du dé va entièrement être déterminée par les lois de la physique, du moins si l'on connaît très précisément la position initiale du dé, sa vitesse initiale, ainsi que la composition de l'air sur sa trajectoire et de la table sur laquelle il va tomber. Pour autant, prévoir ce mouvement exact est humainement impossible, et il est plus aisé de penser le résultat d'un dé comme étant le résultat du “hasard”.

Il y a plusieurs manières de définir mathématiquement la notion de hasard, qui seront plus ou moins simples et plus ou moins puissantes. La manière qui fait consensus aujourd'hui parmi les mathématiciens est l'*axiomatique de Kolmogorov*, qui est basée sur la théorie de la mesure. Comme la théorie de la mesure n'est pas au programme du CAPES, nous ne passerons que rapidement sur les notions la faisant intervenir, pour nous concentrer sur la pratique du calcul des probabilités.

L'idée de l'axiomatique de Kolmogorov est la suivante : si X est une certaine quantité “aléatoire”, elle peut être vue comme le résultat d'un processus faisant intervenir une source de “hasard”. Formellement, on se donne donc un ensemble Ω , correspondant à tous les états possibles de la source de hasard (dans le cas du lancer de dé, Ω est donc l'ensemble des manières de lancer un dé — un ensemble très complexe !) et une fonction $X : \Omega \rightarrow E$ qui a un état de la source de hasard associe un résultat. Pour le lancer de dé, E est l'ensemble $\{1, 2, 3, 4, 5, 6\}$, et la fonction X est la fonction qui a une manière de lancer le dé associe le résultat (il s'agit également d'une fonction très complexe!).

L'axiomatique de Kolmogorov utilise aussi un autre objet, appelé *mesure de probabilité* et noté \mathbf{P} . Il s'agit d'une fonction qui à un sous-ensemble de Ω associe la probabilité que celui-ci se produise. Cet objet permet de décrire de quelle manière un élément de Ω est choisi “au hasard”.

On retiendra de l'exemple du lancer de dé que l'ensemble Ω est très complexe à décrire. Pour autant, cela n'est nullement un problème, puisqu'en pratique *il est inutile de connaître l'ensemble Ω* . La seule chose importante est *la façon dont X transporte la mesure de probabilité \mathbf{P} de Ω vers E* . Cela correspond à la notion de *loi* d'une variable aléatoire. Dans l'exemple du lancer de dés, la seule chose importante à savoir sur l'ensemble Ω est qu'il contient 6 sous-ensembles A_1, A_2, \dots, A_6 correspondant aux 6 valeurs possibles prises par le dés : A_i est l'ensemble des ω tels que $X(\omega) = i$. Pour coller à l'intuition que chaque résultat se produit avec probabilité $1/6$, il sera alors naturel d'imposer $\mathbf{P}(A_i) = 1/6$.

On notera aussi que l'on ne va pas donner de sens au concept de *probabilité*. On va associer à chaque sous-ensemble de Ω un nombre réel que l'on appellera sa "probabilité", sans expliquer ce que cela signifie. En revanche, on retrouvera la notion *empirique* de probabilité dans la loi des grands nombres (théorème 6.2.1) : si on répète une même expérience un grand nombre de fois de manière indépendante, alors la proportion de réalisations d'un certain évènement va être (proche de) la probabilité de cet évènement.

1.2 Espaces probabilisés

Donnons maintenant des définitions plus précises. Étant donné un ensemble Ω , une mesure de probabilité sur Ω devrait être une application de $\mathcal{P}(\Omega)$ ^(note 1) dans $[0, 1]$, vérifiant $\mathbf{P}(\emptyset) = 0$, $\mathbf{P}(\Omega) = 1$ et

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$$

si A et B sont deux sous-ensembles disjoints de Ω . Par ailleurs, pour pouvoir faire de l'analyse à partir de là, on va avoir besoin de faire des *passages à la limite*. Notamment, on aimerait que la fonction \mathbf{P} soit "continue", au sens où $\mathbf{P}(A_n)$ devrait converger vers $\mathbf{P}(A)$ si A_n "tend" vers A . Cependant, il n'y a pas de notion vraiment naturelle de convergence d'un ensemble vers un autre. On a toutefois le résultat suivant :

Lemme 1.2.1. *Soit Ω un ensemble et \mathbf{P} une fonction de $\mathcal{P}(\Omega)$ dans $[0, 1]$ vérifiant $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ dès que A et B sont disjoints. Alors, on a équivalence entre les cinq propriétés suivantes :*

1. Pour toute suite $(A_n)_{n \in \mathbf{N}}$ de sous-ensembles de Ω , on a

$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} A_n\right) = \lim_N \mathbf{P}\left(\bigcup_{n=0}^N A_n\right) ;$$

2. Pour toute suite $(B_n)_{n \in \mathbf{N}}$ de sous-ensembles de Ω , on a

$$\mathbf{P}\left(\bigcap_{n \in \mathbf{N}} B_n\right) = \lim_N \mathbf{P}\left(\bigcap_{n=0}^N B_n\right) ;$$

3. Pour toute suite $(C_n)_{n \in \mathbf{N}}$ croissante de sous-ensembles de Ω (c'est-à-dire vérifiant $C_n \subset C_{n+1}$), on a

$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} C_n\right) = \lim_n \mathbf{P}(C_n) ;$$

(note 1). On note $\mathcal{P}(\Omega)$ l'ensemble des sous-ensembles de Ω .

4. Pour toute suite $(D_n)_{n \in \mathbf{N}}$ décroissante de sous-ensembles de Ω (c'est-à-dire vérifiant $D_{n+1} \subset D_n$), on a

$$\mathbf{P} \left(\bigcap_{n \in \mathbf{N}} D_n \right) = \lim_n \mathbf{P}(D_n) ;$$

5. Pour toute suite $(E_n)_{n \in \mathbf{N}}$ de sous-ensembles disjoints de Ω , on a

$$\mathbf{P} \left(\bigcup_{n \in \mathbf{N}} E_n \right) = \sum_{n \in \mathbf{N}} \mathbf{P}(E_n). \quad (1.1)$$

Démonstration. Remarquons tout d'abord que la relation $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ pour des ensembles disjoints implique $\mathbf{P}(\Omega) = \mathbf{P}(A) + \mathbf{P}(A^c)$. Notamment, la probabilité $\mathbf{P}(A^c)$ est donnée par $\mathbf{P}(A^c) = \mathbf{P}(\Omega) - \mathbf{P}(A)$, ce qui permet donc de "passer au complémentaire" dans une probabilité. L'équivalence entre 1. et 2. se prouve alors en posant $A_n = B_n^c$ (et $B_n = A_n^c$), et de même, l'équivalence entre 3. et 4. se voit en posant $C_n = D_n^c$ (et $D_n = C_n^c$). Il reste donc à prouver 1. \Rightarrow 3. \Rightarrow 5. \Rightarrow 1.

Pour 1. \Rightarrow 3., on remarque que pour une suite croissante $(C_n)_{n \in \mathbf{N}}$, on a $\bigcup_{n=0}^N C_n = C_N$, de sorte que

$$\mathbf{P} \left(\bigcup_{n \in \mathbf{N}} C_n \right) = \lim_N \mathbf{P} \left(\bigcup_{n=0}^N C_n \right) = \lim_N \mathbf{P}(C_N).$$

Pour 3. \Rightarrow 5., on pose $C_n = \bigcup_{k=0}^n E_k$, où les $(E_n)_{n \in \mathbf{N}}$ sont supposés disjoints. La suite $(C_n)_{n \in \mathbf{N}}$ est alors croissante avec $\bigcup_{n \in \mathbf{N}} C_n = \bigcup_{n \in \mathbf{N}} E_n$, et on obtient alors

$$\mathbf{P} \left(\bigcup_{n=0}^{\infty} E_n \right) = \mathbf{P} \left(\bigcup_{n=0}^{\infty} C_n \right) = \lim_N \mathbf{P}(C_N) = \lim_N \mathbf{P} \left(\bigcup_{n=0}^N E_n \right) = \lim_N \sum_{n=0}^N \mathbf{P}(E_n) = \sum_{n=0}^{\infty} \mathbf{P}(E_n).$$

L'avant-dernière égalité se déduit par récurrence sur n de la propriété $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ pour des ensembles A et B disjoints.

Pour montrer 5. \Rightarrow 1. on pose pour tout n , $E_n = A_n \setminus \bigcup_{k=0}^{n-1} A_k$, de sorte que les E_n sont disjoints avec $\bigcup_{n \in \mathbf{N}} E_n = \bigcup_{n \in \mathbf{N}} A_n$ et $\bigcup_{n=1}^N E_n = \bigcup_{n=0}^N A_n$. On a alors

$$\mathbf{P} \left(\bigcup_{n \in \mathbf{N}} A_n \right) = \mathbf{P} \left(\bigcup_{n \in \mathbf{N}} E_n \right) = \lim_N \sum_{n=1}^N \mathbf{P}(E_n) = \lim_N \mathbf{P} \left(\bigcup_{n=1}^N E_n \right) = \lim_N \mathbf{P} \left(\bigcup_{n=1}^N A_n \right).$$

□

Au vu du lemme précédent, il est naturel de définir une mesure de probabilité comme étant une application de $\mathcal{P}(\Omega)$ dans $[0, 1]$ vérifiant les cinq propriétés (équivalentes) de continuité ci-dessus.

Définition 1.2.2. On appelle espace probabilisé un couple (Ω, \mathbf{P}) où Ω est un ensemble et \mathbf{P} est une application $\mathbf{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ telle que :

- $\mathbf{P}(\emptyset) = 0$, $\mathbf{P}(\Omega) = 1$;
- si les ensembles A et B sont disjoints, alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$;
- les propriétés de continuité énoncées dans le lemme 1.2.1 sont vérifiées.

On dit que \mathbf{P} est une mesure de probabilité sur Ω .

Quelques remarques :

— On a passé sous silence un point important de théorie de la mesure : dans de nombreux cas, il n'est pas possible de définir de manière cohérente une mesure \mathbf{P} sur tout $\mathcal{P}(\Omega)$. Dans ces cas-là, on devrait seulement supposer que \mathbf{P} est défini sur un sous-ensemble de $\mathcal{P}(\Omega)$ qui dispose d'une structure adéquate (il s'agit de la structure de *tribu*).

On ne se préoccupera pas de cette subtilité dans ce cours, et on fera comme si toutes les probabilités étaient définies sur $\mathcal{P}(\Omega)$ entier. Cela est d'ailleurs toujours possible tant que Ω est *fini* ou *dénombrable*.

— L'égalité (1.1) implique la relation $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ (pour le voir, prendre la suite $E_0 = A$, $E_1 = B$, $E_n = \emptyset$ pour $n \geq 2$) et le fait que $\mathbf{P}(\emptyset) = 0$ (prendre la suite $E_n = \emptyset$ pour tout n). On pourrait donc aussi définir une mesure de probabilité comme une fonction $\mathbf{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$ vérifiant $\mathbf{P}(\Omega) = 1$ et la propriété 5. du lemme (1.2.1).

Donnons un peu de vocabulaire :

- Les sous-ensembles de Ω seront appelés les *événements* de l'espace probabilisé (Ω, \mathbf{P}) .
- Les éléments de Ω , souvent notés ω s'appellent des *issues*.

Puisque les événements sont des ensembles, on peut leur appliquer les opérations ensemblistes habituelles, avec la correspondance suivante entre les terminologies ensembliste et probabiliste.

Notation	Terminologie ensembliste	Terminologie probabiliste
Ω	ensemble entier	espace des états, événement certain
\emptyset	ensemble vide	événement impossible
ω	élément de Ω	issue
A	sous-ensemble de Ω	événement
$\omega \in A$	ω appartient à A	si ω est le résultat de l'expérience, alors A est réalisé
$A \subset B$	A est inclus dans B	si A est réalisé, alors B aussi
$A \cup B$	réunion de A et B	A est réalisé ou B est réalisé
$A \cap B$	intersection de A et B	A et B sont réalisés
A^c	complémentaire de A	A n'est pas réalisé
$A \cap B = \emptyset$	A et B sont disjoints	A et B ne peuvent se produire simultanément

À partir de la seule définition 1.2.2, nous pouvons déjà donner un certain nombre de propriétés des mesures de probabilité.

Propriété 1.2.3. *Une mesure de probabilité \mathbf{P} sur Ω est une fonction croissante, au sens où si $A \subset B$, alors*

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

Démonstration. Les ensembles A et $B \setminus A$ sont disjoints, et l'inclusion $A \subset B$ donne l'égalité

$$A \cup (B \setminus A) = B.$$

On a donc $\mathbf{P}(A) + \mathbf{P}(B \setminus A) = \mathbf{P}(B)$. Comme \mathbf{P} est une mesure de probabilité, on a $\mathbf{P}(B \setminus A) \geq 0$, d'où $\mathbf{P}(A) \leq \mathbf{P}(B)$. \square

Propriété 1.2.4. *Soit (Ω, \mathbf{P}) un espace probabilisé. Si A et B sont deux événements, alors :*

- $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$;
- $\mathbf{P}(A \setminus B) + \mathbf{P}(B \setminus A) = \mathbf{P}(A \cup B) - \mathbf{P}(A \cap B)$.

$$- \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B);$$

Démonstration. Pour le premier point, le fait que A et A^c soient disjoints montre que $\mathbf{P}(A) + \mathbf{P}(A^c) = \mathbf{P}(A \cup A^c) = \mathbf{P}(\Omega) = 1$, d'où le résultat.

Pour le deuxième point, on utilise l'écriture

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

où les trois ensembles $A \setminus B$, $A \cap B$ et $B \setminus A$ sont disjoints, de sorte que

$$\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus B) + \mathbf{P}(A \cap B) + \mathbf{P}(B \setminus A).$$

Pour le troisième point, on remarque que $A = (A \setminus B) \cup (A \cap B)$, où l'union est disjointe, de sorte que $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$ (et de même $\mathbf{P}(B \setminus A) = \mathbf{P}(B) - \mathbf{P}(A \cap B)$). On conclut en utilisant le deuxième point. \square

On a aussi la propriété suivante, qui peut souvent servir.

Propriété 1.2.5. *Si $(A_n)_{n \in \mathbf{N}}$ est une suite d'évènements, alors on a la majoration*

$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} A_n\right) \leq \sum_{n \in \mathbf{N}} \mathbf{P}(A_n).$$

Démonstration. On va montrer par récurrence sur N l'inégalité

$$\mathbf{P}\left(\bigcup_{n=0}^N A_n\right) \leq \sum_{n=0}^N \mathbf{P}(A_n).$$

La propriété s'en déduira alors en passant à la limite en $N \rightarrow \infty$ et en utilisant la propriété 1. du lemme 1.2.1.

Le cas $N = 0$ est trivial. Si on suppose vraie l'inégalité pour l'entier N , alors

$$\begin{aligned} \mathbf{P}\left(\bigcup_{n=0}^{N+1} A_n\right) &= \mathbf{P}\left(\left(\bigcup_{n=0}^N A_n\right) \cup A_{N+1}\right) = \mathbf{P}\left(\bigcup_{n=0}^N A_n\right) + \mathbf{P}(A_{N+1}) - \mathbf{P}\left(\left(\bigcup_{n=0}^N A_n\right) \cap A_{N+1}\right) \\ &\leq \mathbf{P}\left(\bigcup_{n=0}^N A_n\right) + \mathbf{P}(A_{N+1}) \\ &\leq \sum_{n=0}^N \mathbf{P}(A_n) + \mathbf{P}(A_{N+1}), \end{aligned}$$

ce qui est la propriété au rang $N + 1$. \square

Une conséquence directe :

Corollaire 1.2.6. *Si les A_n sont des évènements de probabilité nulle ($\mathbf{P}(A_n) = 0$), alors l'évènement $\bigcup_{n \in \mathbf{N}} A_n$ est lui aussi de probabilité nulle.*

1.3 Variables aléatoires

Définition 1.3.1. Soit (Ω, \mathbf{P}) un espace probabilisé et E un ensemble. On appelle variable aléatoire à valeurs dans E une fonction de Ω dans E .

Comme on l'a déjà dit, la description exacte de l'ensemble Ω est en générale peu pertinente. En fait, les objets "utiles" seront les évènements pouvant se décrire à partir des variables aléatoires définies sur Ω . On va introduire une notation commode pour manipuler ces évènements.

Définition 1.3.2. Soit $X : \Omega \rightarrow E$ une variable aléatoire, I un sous ensemble de E et x un élément de E . On note alors^(note 2)

$$\begin{aligned}\{X \in I\} &= X^{-1}(I) = \{\omega \in \Omega, X(\omega) \in I\}, \\ \{X = x\} &= X^{-1}(\{x\}) = \{\omega \in \Omega, X(\omega) = x\}.\end{aligned}$$

De manière générale, si \mathcal{P} est une certaine propriété, on notera

$$\{X \text{ vérifie } \mathcal{P}\} = \{\omega \in \Omega, X(\omega) \text{ vérifie } \mathcal{P}\}.$$

Par exemple, si $E = \mathbf{R}$, on utilisera souvent la notation

$$\{X \leq x\} = X^{-1}(]-\infty, x]) = \{\omega \in \Omega, X(\omega) \leq x\}$$

ainsi que les notations similaires $\{X < x\}$, $\{x \leq X \leq y\}$, etc.

La notation introduite en définition 1.3.2 permet de manipuler des évènements qui ont une interprétation pour le problème que l'on considère. En effet, si $X : \Omega \rightarrow E$ est une variable aléatoire, on n'a pas besoin de pouvoir représenter précisément l'ensemble Ω , mais seulement les évènements qui ont une signification par rapport à X : il s'agit des éléments de la forme $\{X \in I\}$.

Propriété 1.3.3. Si (Ω, \mathbf{P}) est un espace probabilisé, E est un ensemble et X est une variable aléatoire à valeurs dans E , alors la variable X induit une mesure de probabilité μ_X sur l'ensemble E , définie par

$$\mu_X(I) = \mathbf{P}(X \in I).$$

La mesure de probabilité μ_X est appelée loi de X .

Pour ne pas alourdir l'écriture, on a noté $\mathbf{P}(X \in I)$ pour $\mathbf{P}(\{X \in I\})$.

Démonstration. On a $\{X \in \emptyset\} = \emptyset$ et $\{X \in \Omega\} = E$. Par conséquent, on a bien

$$\mu_X(\emptyset) = \mathbf{P}(\emptyset) = 0 \text{ et } \mu_X(E) = \mathbf{P}(\Omega) = 1.$$

Ensuite, si $(I_n)_{n \in \mathbf{N}}$ est une suite de sous-ensembles disjoints de E , alors les ensembles $A_n = \{X \in I_n\}$ sont eux-aussi disjoints, d'où

$$\mu_X \left(\bigcup_{n \in \mathbf{N}} I_n \right) = \mathbf{P} \left(\bigcup_{n \in \mathbf{N}} A_n \right) = \sum_{n \in \mathbf{N}} \mathbf{P}(A_n) = \sum_{n \in \mathbf{N}} \mu_X(I_n).$$

□

(note 2). On rappelle que la notation $X^{-1}(I)$ désigne l'image réciproque de la partie I par l'application X . Cela ne sous-entend pas que X est bijective!

L'idée de la notion de loi de probabilité est la suivante : la mesure \mathbf{P} permet de "choisir au hasard" un ω dans Ω , c'est-à-dire, un état possible de la source de hasard. La variable aléatoire X associe à ce ω une observation $X(\omega) \in E$. D'une certaine manière, on a donc choisi, par le biais de X , un élément aléatoire de E . La mesure μ_X représente la manière dont cet élément aléatoire est tiré.

La plupart du temps, quand on manipulera des variables aléatoires, les seules hypothèses que l'on fera concerneront la loi de X . Les énoncés que l'ont présentera commencerons donc souvent par "Soit X une variable aléatoire de loi μ ". Pour que tout cela ait un sens, on doit s'assurer que, étant donnée une mesure de probabilité μ sur un ensemble E , il existe une variable aléatoire dont la loi est μ . C'est ce que l'on énonce dans le résultat suivant.

Propriété 1.3.4. *Soit E un ensemble et μ une mesure de probabilité sur E . Alors, il existe un espace probabilisé (Ω, \mathbf{P}) sur lequel est définie une variable aléatoire X de loi μ .*

Démonstration. Il suffit de choisir $\Omega = E$, $\mathbf{P} = \mu$ et de prendre pour X la fonction identité. Dans ce cas, (Ω, \mathbf{P}) est bien un espace probabilisé et X est une variable aléatoire.

Par définition la loi de X est la mesure μ_X définie pour $A \subset X$ par $\mu_X(A) = \mathbf{P}(X^{-1}(A))$. Or, avec notre choix de \mathbf{P} et X , on a $\mathbf{P} = \mu$ et $X^{-1}(A) = A$. On a donc bien $\mu_X(A) = \mu(A)$. \square

1.4 Variables aléatoires réelles

On appellera variable aléatoire *réelle* toute variable aléatoire à valeurs dans \mathbf{R} .

Comme l'outil fondamental en théorie des probabilité est la mesure de probabilité, il est utile d'avoir des manières de caractériser une mesure. Plus précisément, si X et Y sont deux variables aléatoires, y a-t-il une façon économique de montrer qu'elles ont la même loi, c'est-à-dire de montrer $\mu_X(I) = \mu_Y(I)$ pour tout $I \subset E$? Dans le cas où les variables sont à valeurs dans \mathbf{R} , le théorème suivant répond à cette question :

Théorème 1.4.1. *Soient μ et ν deux mesures de probabilité sur \mathbf{R} . Si pour tout $t \in \mathbf{R}$ on a*

$$\mu(]-\infty, t]) = \nu(]-\infty, t]),$$

alors $\mu = \nu$.

Démonstration. On voudrait montrer que $\mu(I) = \nu(I)$ pour tout $I \subset \mathbf{R}$. On va seulement faire la preuve dans le cas où I est un intervalle. Si I est de la forme $]-\infty, t]$ on a $\mu(I) = \nu(I)$ par hypothèse. Si I est de la forme $]t, \infty[$, il suffit de passer au complémentaire :

$$\mu(I) = 1 - \mu(]-\infty, t]) = 1 - \nu(]-\infty, t]) = \nu(I).$$

Pour $I =]-\infty, t[= \bigcup_{n \geq 1}]-\infty, t - 1/n]$, on écrit

$$\mu(I) = \lim_n \mu(]-\infty, t - 1/n]) = \lim_n \nu(]-\infty, t - 1/n]) = \nu(I),$$

et en passant au complémentaire, on a le résultat pour $I = [t, \infty[$. Enfin, les intervalles bornés se traitent de la manière suivante, par exemple pour $I = [s, t]$:

$$\begin{aligned} \mu(I) &= \mu(]-\infty, t] \setminus]-\infty, s]) = \mu(]-\infty, t]) - \mu(]-\infty, s]) = \nu(]-\infty, t]) - \nu(]-\infty, s]) \\ &= \nu(]-\infty, t] \setminus]-\infty, s]) \\ &= \nu(I). \end{aligned}$$

On a donc $\mu(I) = \nu(I)$ pour tout intervalle I . Pour montrer le même résultat pour tout $I \subset \mathbf{R}$, l'idée serait de montrer que tout ^(note 3) sous-ensemble de \mathbf{R} peut se “décrire” à partir d'intervalles. \square

On est donc amenés à la définition suivante :

Définition 1.4.2. Soit X une variable aléatoire de loi μ_X à valeurs dans \mathbf{R} . On appelle fonction de répartition de la loi μ_X ou encore, par abus, fonction de répartition de X , l'application F_X définie sur \mathbf{R} par :

$$\forall t \in \mathbf{R}, F_X(t) = \mu_X(]-\infty, t]) = \mathbf{P}(X \leq t).$$

Avec la définition précédente, le théorème 1.4.1 revient à dire que si deux variables aléatoires X et Y ont la même fonction de répartition, alors elles ont la même loi.

Propriété 1.4.3. Une fonction de répartition F_X satisfait les propriétés suivantes :

- F_X est croissante ;
- F_X est continue à droite et admet une limite à gauche en tout point ;
- $\lim_{t \rightarrow -\infty} F_X(t) = 0$, $\lim_{t \rightarrow +\infty} F_X(t) = 1$.

Démonstration. Le premier point découle de la propriété de croissance des probabilités : si $s \leq t$, alors $]-\infty, s] \subset]-\infty, t]$, d'où $\mu_X(]-\infty, s]) \leq \mu_X(]-\infty, t])$.

Pour montrer le deuxième point, on considère une suite $(t_n)_{n \in \mathbf{N}}$ croissante, de limite t , avec $t_n < t$ ainsi qu'une suite décroissante $(\tilde{t}_n)_{n \in \mathbf{N}}$ de limite t avec $t_n > t$. Pour tout $n \geq 1$, on définit

$$C_n =]-\infty, t_n] \text{ et } D_n =]-\infty, \tilde{t}_n].$$

Alors $(C_n)_{n \geq 1}$ est une suite croissante vérifiant $\bigcup_{n \geq 1} C_n =]-\infty, t[$, et la suite $(D_n)_{n \geq 1}$ est décroissante avec $\bigcap_{n \geq 1} D_n =]-\infty, t]$. Donc, d'après le lemme 1.2.1, on sait que :

$$\lim_{n \rightarrow +\infty} \mu_X(C_n) = \mu_X(]-\infty, t]) \text{ et } \lim_{n \rightarrow +\infty} \mu_X(D_n) = \mu_X(]-\infty, t]),$$

d'où le résultat.

De même, pour démontrer le dernier point il suffit de montrer

$$\lim_{n \rightarrow \infty} F_X(-n) = 0 \text{ et } \lim_{n \rightarrow \infty} F_X(n) = 1.$$

La preuve est similaire à celle du troisième point, en utilisant les ensembles $C_n =]-\infty, n]$ et $D_n =]-\infty, -n]$. \square

En fait, les propriétés listées dans la propriété 1.4.3 caractérisent les fonctions de répartition :

Théorème 1.4.4. Toute application F de \mathbf{R} dans $[0, 1]$ vérifiant les trois propriétés listées dans la propriété 1.4.3 est la fonction de répartition d'une loi de probabilité sur \mathbf{R} .

En vertu du théorème 1.4.1, la mesure de probabilité μ dont F est la fonction de répartition est unique. En somme, on a construit une bijection entre un ensemble d'objets abstraits (les mesures de probabilités), et un ensemble d'objets beaucoup plus concrets (des fonctions croissantes).

Démonstration. Admis. Il s'agirait de construire une mesure de probabilité sur $(\mathbf{R}, \mathcal{P}(\mathbf{R}))$. Or, on a déjà évoqué le fait que construire explicitement une mesure de probabilité sur un ensemble indénombrable (ici \mathbf{R}) pouvait être très complexe. \square

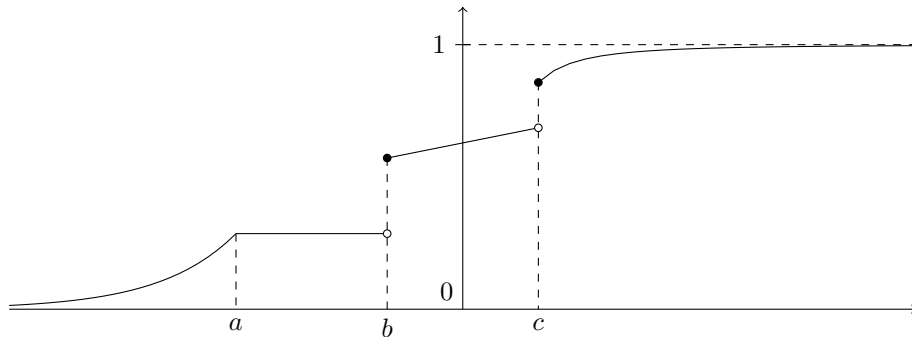


FIGURE 1.1 – Un exemple de fonction vérifiant les conditions de la propriété 1.4.3. Le théorème 1.4.1 assure qu’il existe une mesure μ dont c’est la fonction de répartition. Les valeurs de la fonction en b et c sont données par les points noirs (continuité à droite). Les points blancs représentent les limites à gauche en b et en c .

Un exemple de fonction de répartition est donné sur la figure 1.1. D’après le théorème 1.4.1, la fonction F caractérise entièrement la mesure μ dont elle est la fonction de répartition. Il est donc en théorie possible de lire n’importe quelle valeur $\mu(A)$ sur le graphique de la figure 1.1. En pratique, certaines de ces valeurs sont très simples à exprimer : il s’agit du cas où A est un intervalle, pour lequel $\mu(A)$ peut s’exprimer comme un accroissement de F .

Propriété 1.4.5. Soit (Ω, \mathbf{P}) un espace probabilisé, et X une variable aléatoire sur Ω de fonction de répartition F_X , alors :

- $\mathbf{P}(X > x) = 1 - F_X(x)$,
- $\mathbf{P}(x < X \leq y) = F_X(y) - F_X(x)$,
- $\mathbf{P}(X < x) = F_X(x-)$.
- $\mathbf{P}(X = x) = F_X(x) - F_X(x-)$.

La notation $F_X(x-)$ désigne la limite à gauche de la fonction F_X au point x , dont l’existence a été donnée dans la propriété 1.4.3.

De la propriété 1.4.5, on peut retenir que les points de discontinuité de F ^(note 4) sont les x tels que $\mathbf{P}(X = x) > 0$. Dans ce cas, la valeur $\mathbf{P}(X = x)$ vaut $F(x) - F(x-)$: il s’agit de la taille du saut dans la fonction de répartition. Inversement, un “plat” dans la fonction de répartition correspond à un intervalle I tel que $\mathbf{P}(X \in I) = 0$. Par exemple, on voit que la mesure μ associée à la fonction de répartition tracée en figure 1.1 vérifie $\mu(\{b\}) > 0$ et $\mu(\{c\}) > 0$ (la fonction n’est pas continue en b et en c , ainsi que $\mu(]a, b]) = 0$ (la fonction est constante sur $]a, b])$.

(note 3). C’est en fait seulement le cas de certains sous-ensembles (les ensembles Boréliens). Ceci est lié au fait que, en toute rigueur, une mesure de probabilité sur \mathbf{R} n’est pas définie sur $\mathcal{P}(\mathbf{R})$ tout entier.

(note 4). Qui sont les x tels que $F(x-) \neq F(x)$, puisque F est continue à droite et admet une limite à gauche en tout point.

1.5 Exercices

Exercice 1.1.

Soient A et B deux évènements d'un même espace probabilisé. Montrer les relations suivantes entre indicatrices :

$$\mathbf{1}_{A^c} = 1 - \mathbf{1}_A, \quad \mathbf{1}_{A \cap B} = \mathbf{1}_A \mathbf{1}_B, \quad \mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_{A \cap B}.$$

Exercice 1.2.

Soient A et B deux évènements. Montrer que les trois évènements $A \setminus B$, $B \setminus A$ et $A \cap B$ sont disjoints et que leur union est $A \cup B$.

Exercice 1.3.

Soient A, B, C trois évènements d'un même espace probabilisé. Exprimer les phrases suivantes en termes ensemblistes, en fonction de A, B et C .

- A est réalisé mais ni B ni C ne le sont ;
- au moins un des trois évènements est réalisé ;
- au plus deux des trois évènements sont réalisés ;
- si A est réalisé, alors un des évènements B et C l'est ;
- si A est réalisé, alors au plus un des évènements B et C l'est.

Exercice 1.4.

Soit T une variable aléatoire à valeurs dans $\mathbf{N} \cup \{\infty\}$. Montrer que $\mathbf{P}(T < \infty) = \sum_{n \in \mathbf{N}} \mathbf{P}(T = n)$.

Exercice 1.5.

Soit N une variable aléatoire à valeurs dans \mathbf{N} , et soit A un évènement. Montrer l'égalité

$$\mathbf{P}(A) = \sum_{n \in \mathbf{N}} \mathbf{P}(A \cap \{N = n\}).$$

Exercice 1.6.

Soit X une variable aléatoire réelle. Montrer que les deux suites $(\mathbf{P}(X \leq \frac{1}{n}))_{n \in \mathbf{N}}$ et $(\mathbf{P}(X \geq \frac{1}{n}))_{n \in \mathbf{N}}$ convergent et donner leurs limites.

Exercice 1.7.

Les fonctions suivantes sont-elles des fonctions de répartition ? Si non, quelle modification simple apporter pour que ce soit le cas ?

$$\arctan(x) ; 1 - e^{-x} ; x \mathbf{1}_{[0,1]} ; e^{-e^{-x}} ; \frac{x}{x+1} ; \arcsin(x) ; E(1/x)^{-1} \mathbf{1}_{]0,1]}(x)$$

(E est la fonction partie entière).

Chapitre 2

Espérance d'une variable aléatoire

2.1 Définition de l'espérance

On veut donner un sens à la notion de valeur moyenne d'une variable aléatoire X . Pour désigner cette valeur moyenne, on parlera de l'*espérance* de X , que l'on notera $\mathbf{E}X$ ^(note 1). Notre but dans ce chapitre va être de définir une valeur $\mathbf{E}X$ pour autant de variables X que possible. On voudrait que la fonction $X \mapsto \mathbf{E}X$ vérifie au moins les trois propriétés suivantes :

1. pour une variable de la forme ^(note 2) $\mathbf{1}_A$, on a $\mathbf{E}\mathbf{1}_A = \mathbf{P}(A)$;
2. \mathbf{E} est linéaire : $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$ et $\mathbf{E}(\alpha X) = \alpha \mathbf{E}X$;
3. \mathbf{E} est croissante : si deux variables X et Y vérifient $X \leq Y$, alors $\mathbf{E}X \leq \mathbf{E}Y$.

Les deux premières conditions ci-dessus imposent déjà la valeur de $\mathbf{E}X$ si X ne prend qu'un nombre fini de valeurs. Une telle variable X est dite étagée :

Définition 2.1.1. Une variable aléatoire est dite étagée si elle ne prend qu'un nombre fini de valeurs. De manière équivalente, une variable aléatoire X est étagée si et seulement si elle peut s'écrire

$$X = \sum_{k=1}^n x_k \mathbf{1}_{A_k} \tag{2.1}$$

pour des réels $(x_k)_{k=1,\dots,n}$ et des événements $(A_k)_{k=1,\dots,n}$

Pour voir qu'une variable aléatoire est étagée si et seulement si elle est de la forme (2.1), on remarque d'une part qu'une variable aléatoire de la forme (2.1) prend ses valeurs dans l'ensemble fini $\{\sum_{k \in I} x_k, I \subset \{1, \dots, n\}\}$. D'autre part, si une variable ne prend que les valeurs $(x_k)_{k \in \{1, \dots, n\}}$, alors elle vérifie

$$X = \sum_{k=1}^n x_k \mathbf{1}_{X=x_k}.$$

(note 1). Ou $\mathbf{E}(X)$ ou $\mathbf{E}[X]$.

(note 2). Par définition, $\mathbf{1}_A(\omega)$ vaut 1 si $\omega \in A$ et vaut 0 sinon. La variable aléatoire $\mathbf{1}_A$ s'appelle l'*indicatrice* de A .

À partir des conditions 1. et 2. ci-dessus, l'espérance d'une variable étagée $X = \sum_{k=1}^n x_k \mathbf{1}_{A_k}$ est forcément donnée par

$$\mathbf{E}X = \sum_{k=1}^n x_k \mathbf{P}(A_k). \quad (2.2)$$

Pour rendre cette définition complètement rigoureuse, il faudrait vérifier que l'expression (2.2) ne dépend pas de la décomposition (2.1) que l'on a choisie, car une variable aléatoire étagée peut s'écrire de plusieurs manières sous cette forme. Par exemple, pour A et B disjoints, on a $\mathbf{1}_{A \cup B} = \mathbf{1}_A + \mathbf{1}_B$. Toutefois, comme \mathbf{P} est une mesure de probabilité, on a bien $\mathbf{E}\mathbf{1}_{A \cup B} = \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) = \mathbf{E}(\mathbf{1}_A + \mathbf{1}_B)$.

L'équation (2.2) est illustrée sur la figure 2.1. On a représenté Ω comme étant l'axe des abscisses, et une fonction de Ω dans \mathbf{R} qui prend un nombre fini de valeurs peut se voir comme une fonction "en escalier". Dans ce cas, l'espérance de X est schématiquement "l'aire sous la courbe" (note 3) : on ajoute les $x_k \times \mathbf{P}(A_k)$, où x_k est sur le dessin la hauteur du k ème rectangle, et $\mathbf{P}(A_k)$ sa largeur.

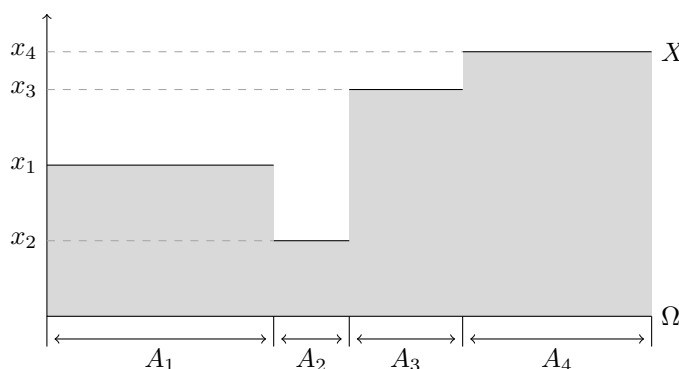


FIGURE 2.1 – Espérance d'une variable étagée.

On voudrait maintenant étendre la notion d'espérance à une classe plus large de variables aléatoires.

2.1.1 Le cas des variables bornées

On va tout d'abord pouvoir étendre la définition de l'espérance à toute variable aléatoire *bornée*. Pour cela, on remarque la chose suivante :

Propriété 2.1.2. *Soit X une variable aléatoire bornée, et soit ε un réel positif. Il existe deux variables aléatoires étagées Y et Z telles que*

$$Y \leq X \leq Z, \text{ et } 0 \leq \mathbf{E}(Z - Y) \leq \varepsilon.$$

Ce résultat signifie donc qu'on peut "approcher" n'importe quelle variable aléatoire bornée par des variables étagées, avec une erreur de petite espérance. Ce résultat est illustré sur la figure 2.2, sur lequel la bande grise, correspondant à $\mathbf{E}(Z - Y)$, peut être rendue d'aire aussi petite que voulu. La variable $Z - Y$ est étagée, donc $\mathbf{E}(Z - Y)$ a bien un sens ici.

(note 3). Cette remarque fait penser à une intégrale. En fait, l'espérance est une intégrale : dans le cadre de la théorie de l'intégration de Lebesgue, l'espérance de X est l'intégrale de X contre la mesure \mathbf{P} .

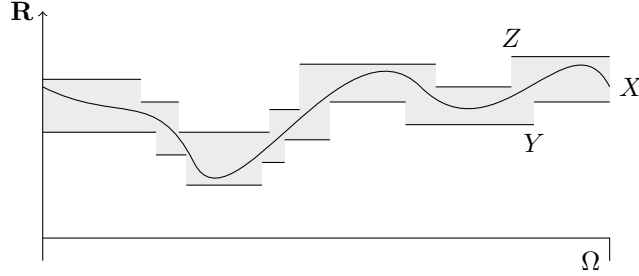


FIGURE 2.2 – Une variable aléatoire X bornée, encadrée par deux variables étagée Y et Z . Quitte à prendre des fonctions étagées prenant suffisamment de valeurs différentes, il est possible de rendre la surface grisée aussi petite que voulu.

Démonstration. Quitte à effectuer une transformation affine, on peut supposer que X prend ses valeurs dans $]0, 1[$, et on peut fixer un $n \in \mathbf{N}$ tel que $1/n \leq \varepsilon$.

On pose alors :

$$Y = \sum_{k=1}^n \frac{k-1}{n} \mathbf{1}_{X \in [(k-1)/n, k/n[} \text{ et } Z = \sum_{k=1}^n \frac{k}{n} \mathbf{1}_{X \in [(k-1)/n, k/n[}.$$

On a bien $Y \leq X \leq Z$ puisque pour un ω donné, si $X(\omega)$ est dans $[(k-1)/n, k/n[$, alors $Y(\omega)$ vaut $(k-1)/n$ et $Z(\omega)$ vaut k/n .

Par ailleurs, on a :

$$Z - Y = \sum_{k=1}^n \left(\frac{k}{n} - \frac{k-1}{n} \right) \mathbf{1}_{X \in [(k-1)/n, k/n[} = \sum_{k=1}^n \frac{1}{n} \mathbf{1}_{X \in [(k-1)/n, k/n[} = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{X \in [(k-1)/n, k/n[} = \frac{1}{n}.$$

En effet, la somme $\sum_{k=1}^n \mathbf{1}_{X \in [(k-1)/n, k/n[}$ vaut 1 puisqu'on a supposé que X prenait ses valeurs dans $]0, 1[$. On a donc bien $0 \leq \mathbf{E}(Z - Y) = 1/n \leq \varepsilon$. \square

Pour définir l'espérance d'une variable étagée, nous n'avons utilisé que les conditions 1. et 2. ci-dessus, et pas la condition de croissance 3. C'est maintenant que cette condition va être importante. Si Y et Z sont deux variables aléatoires étagées encadrant X comme dans la proposition 2.1.2, on souhaiterait donc que l'espérance de X , si elle était définie, vérifie la relation $\mathbf{E}Y \leq \mathbf{E}X \leq \mathbf{E}Z$. En particulier, on aurait :

$$\sup_{Y \leq X, Y \text{ étagé}} \mathbf{E}Y \leq \mathbf{E}X \leq \inf_{Z \geq X, Z \text{ étagé}} \mathbf{E}Z,$$

où les bornes supérieure et inférieure portent respectivement sur les variables étagées vérifiant $Y \leq X$ et $X \leq Z$. Cependant, d'après la proposition 2.1.2, pour tout $\varepsilon > 0$, on peut trouver deux telles variables vérifiant $0 \leq \mathbf{E}(Z - Y) \leq \varepsilon$, ce qui implique $\sup_Y \mathbf{E}Y = \inf_Z \mathbf{E}Z$. Par conséquent, la valeur de l'espérance d'une variable aléatoire bornée nous est imposée : cela ne peut être que la valeur $\sup_Y \mathbf{E}Y = \inf_Z \mathbf{E}Z$. On va donc définir comme cela l'espérance d'une variable bornée :

Définition 2.1.3. Si X est une variable aléatoire réelle bornée, on définit l'espérance de X comme étant la valeur commune de $\sup_Y \mathbf{E}Y = \inf_Z \mathbf{E}Z$ où les bornes supérieure et inférieure portent respectivement sur les variables étagées minorant ou majorant X .

On remarquera que les ensembles sur lesquels on prend les bornes supérieure et inférieure sont non-vides puisque la variable X est bornée. En effet, si $a \leq X \leq b$, alors les variables aléatoires constantes $Y = a$ et $Z = b$ sont bien des variables aléatoires étagées encadrant X .

Cette définition s'est imposée en cherchant à définir une espérance linéaire et croissante. On peut vérifier que ces deux propriétés sont bien vérifiées :

Propriété 2.1.4. Soient X et Y deux variables aléatoires bornées. Alors :

- On a $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$, et si α est un réel, alors $\mathbf{E}(\alpha X) = \alpha \mathbf{E}X$.
- Si $X \leq Y$, alors $\mathbf{E}X \leq \mathbf{E}Y$.

Propriété 2.1.5. Soit $\varepsilon > 0$. Par définition de $\mathbf{E}X$ et $\mathbf{E}Y$, il existe des variables étagées X_0 et Y_0 vérifiant $X_0 \leq X$, $Y_0 \leq Y$ avec $\mathbf{E}X_0 \geq \mathbf{E}X - \varepsilon$ et $\mathbf{E}Y_0 \geq \mathbf{E}Y - \varepsilon$. La variable $X_0 + Y_0$ est alors une variable aléatoire étagée inférieure à $X + Y$. On a donc $\mathbf{E}(X + Y) \geq \mathbf{E}(X_0 + Y_0) \geq \mathbf{E}X + \mathbf{E}Y - 2\varepsilon$. Ceci étant vrai pour tout ε , on a donc $\mathbf{E}(X + Y) \geq \mathbf{E}X + \mathbf{E}Y$. En raisonnant de la même manière avec des variables étagées majorant X et Y , on obtient $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$.

Pour montrer $\mathbf{E}(\alpha X) = \alpha \mathbf{E}X$, en supposant $\alpha > 0$, il suffit de remarquer que pour X_0 est une variable étagée, on a $X_0 \leq X$ si et seulement si $\alpha X_0 \leq \alpha X$ et que $\mathbf{E}(\alpha X_0) = \alpha \mathbf{E}X_0$. Le raisonnement est le même pour $\alpha < 0$, à ceci près que $X_0 \leq X$ si et seulement si $\alpha X_0 \geq \alpha X$.

Enfin, si $X \leq Y$, l'ensemble des variables aléatoires étagées vérifiant $Z \leq X$ est inclus dans l'ensemble des variables aléatoires étagées vérifiant $Z \leq Y$, de sorte que $\mathbf{E}X = \sup_{Z \leq X} \mathbf{E}Z \leq \sup_{Z \leq Y} \mathbf{E}Z = \mathbf{E}Y$.

2.1.2 Le cas des variables positives

Se limiter à des variables bornées sera rapidement trop contraignant, on va donc étendre la définition de la partie précédente à toutes les variables majorées ou minorées. On va présenter la définition et les principales propriétés de l'espérance pour des variables *positives* (et donc minorées par 0). Tous ces résultats pourraient s'étendre aux variables majorées ou minorées par ajout d'une constante et/ou changement de signe.

Si nous reprenons la définition 2.1.3, la quantité $\sup_{Y \leq X, Y \text{ étagé}} \mathbf{E}Y$ a encore un sens pour un X positif, alors que $\inf_{Z \geq X, Z \text{ étagé}} \mathbf{E}Z$ n'en a pas forcément. En effet, pour une variable aléatoire X positive et non majorée, il n'existe pas de variable aléatoire Z étagée vérifiant $X \leq Z$. On reprend donc la définition 2.1.3, en ne considérant que la borne supérieure des espérances des Y :

Définition 2.1.6. Soit X une variable aléatoire à valeurs dans $[0, \infty[$. On appelle espérance de X la valeur

$$\mathbf{E}X = \sup_{0 \leq Y \leq X} \mathbf{E}Y,$$

où la borne supérieure est prise sur l'ensemble des variables aléatoires Y étagées vérifiant $0 \leq Y \leq X$. Pour un tel Y , $\mathbf{E}Y$ est défini par l'équation (2.2).

On retiendra que l'espérance d'une variable *positive* a toujours un sens, même si on peut avoir $\mathbf{E}X = \infty$. En particulier, quelle que soit la variable aléatoire réelle X l'expression $\mathbf{E}|X|$ est bien définie.

On vérifie maintenant que cette définition étendue d'espérance vérifie toujours les propriétés de monotonie et de linéarité.

Propriété 2.1.7. Si X et \tilde{X} sont deux variables aléatoires positives avec $X \leq \tilde{X}$, alors $\mathbf{E}X \leq \mathbf{E}\tilde{X}$.

Démonstration. Si $0 \leq X \leq \tilde{X}$, alors toute variable aléatoire Y vérifiant $0 \leq Y \leq X$ vérifiera aussi $0 \leq Y \leq \tilde{X}$. Par conséquent, la borne supérieure apparaissant dans la définition de $\mathbf{E}\tilde{X}$ portera sur un ensemble plus grand que celle qui apparaît dans la définition de $\mathbf{E}X$. Cette borne supérieure sera donc plus grande. \square

Propriété 2.1.8. Soient X et Y deux variables aléatoires positives et $a > 0$. On a

1. $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$, avec la convention $x + \infty = \infty$ pour tout x de $[0, \infty]$;
2. $\mathbf{E}(aX) = a\mathbf{E}X$, avec la convention $a \times \infty = \infty$.

Démonstration. On admet la propriété d'additivité.

Pour le dixième point, on écrit

$$\mathbf{E}(aX) = \sup_{0 \leq Y \leq aX} \mathbf{E}Y = \sup_{0 \leq \frac{Y}{a} \leq X} \mathbf{E}Y = \sup_{0 \leq Z \leq X} \mathbf{E}(aZ) = a \times \sup_{0 \leq Z \leq X} \mathbf{E}Z = a\mathbf{E}X,$$

où chacune des bornes supérieures porte sur l'ensemble des Y (ou Z) étagés vérifiant les inégalités données. \square

2.1.3 Le cas des variables intégrables

On va maintenant étendre la notion d'espérance à des variables qui ne sont pas nécessairement positives. Cette fois-ci, on ne pourra par contre pas forcément donner un sens à l'espérance de n'importe quelle variable aléatoire. Une des raisons est qu'il est possible de trouver des variables aléatoires X et Y positives avec $\mathbf{E}X = \mathbf{E}Y = \infty$ pour lesquelles $\mathbf{E}(X - Y)$ n'a pas de sens : en effet, on est en présence d'une forme indéterminée $\infty - \infty$.

Définition 2.1.9. Soit X une variable aléatoire réelle. On dit que X est intégrable si $\mathbf{E}|X| < \infty$. Dans ce cas on appelle espérance de X la quantité

$$\mathbf{E}X = \mathbf{E}(X\mathbf{1}_{\{X \geq 0\}}) - \mathbf{E}(-X\mathbf{1}_{\{X < 0\}}). \quad (2.3)$$

Une variable aléatoire d'espérance nulle est dite centrée.

Cette définition a bien un sens. En effet, d'une part, la variable $|X|$ est positive, de sorte que $\mathbf{E}|X|$ est bien définie, sans hypothèse supplémentaire. D'autre part, les deux variables $X\mathbf{1}_{\{X \geq 0\}}$ et $-X\mathbf{1}_{\{X < 0\}}$ sont positives et sont majorées par $|X|$, de sorte que

$$\mathbf{E}(X\mathbf{1}_{\{X > 0\}}) \leq \mathbf{E}|X| < \infty \text{ et } \mathbf{E}(-X\mathbf{1}_{\{X < 0\}}) \leq \mathbf{E}|X| < \infty,$$

ce qui fait que l'on n'est pas en présence d'une forme indéterminée $\infty - \infty$. Enfin, on remarquera que pour X positive, les deux définitions se rejoignent. La forme de l'expression (2.3) est motivée par l'égalité

$$X = X(\mathbf{1}_{\{X \geq 0\}} + \mathbf{1}_{\{X < 0\}}) = X\mathbf{1}_{\{X \geq 0\}} - (-X)\mathbf{1}_{\{X < 0\}}.$$

Si on interprète l'espérance comme une aire sous une courbe, cette généralisation correspond à compter négativement les surfaces situées en dessous de l'axe des abscisses.

On notera qu'une variable aléatoire bornée est *toujours* intégrable (si $-a < X < a$, alors $|X| < a$, donc $\mathbf{E}|X| < a < \infty$).

Les résultats énoncés dans les propriétés 2.1.7 et 2.1.8 sont encore valables pour des variables aléatoires intégrables.

Propriété 2.1.10. Soient X et Y deux variables aléatoires intégrables et $a > 0$.

- Si $X \leq Y$, alors $\mathbf{E}X \leq \mathbf{E}Y$. En particulier, on a $|\mathbf{E}X| \leq \mathbf{E}|X|$.
- Les variables aléatoires $X + Y$ et aX sont intégrables et on a

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y \text{ et } \mathbf{E}(aX) = a\mathbf{E}X.$$

Démonstration. Pour la première propriété, on remarque que si $X \leq Y$, alors $X\mathbf{1}_{\{X \geq 0\}} \leq Y\mathbf{1}_{\{Y \geq 0\}}$ et $(-X)\mathbf{1}_{\{X > 0\}} \geq (-Y)\mathbf{1}_{\{Y > 0\}}$.

Pour la deuxième propriété, on voit que $X + Y$ est intégrable car $|X + Y| \leq |X| + |Y|$, d'où

$$\mathbf{E}|X + Y| \leq \mathbf{E}|X| + \mathbf{E}|Y| < \infty.$$

On découpe ensuite Ω en fonction des signes de X , Y et $X + Y$.

Pour aX , cela découle de $\mathbf{E}|aX| = |a|\mathbf{E}|X| < \infty$. □

2.2 Théorème de transfert

Les définitions 2.1.3, 2.1.6 et 2.1.9 *ne sont pas* les définitions qui seront utilisées *en pratique* pour calculer une espérance. Le théorème 2.2.1 ci-dessous est une première étape pour donner une expression plus explicite de l'espérance d'une variable aléatoire. Il sera énoncé dans des cas plus généraux par les théorèmes 4.1.2 et 5.1.5.

Théorème 2.2.1 (Théorème de transfert, version 1). Soit X une variable aléatoire à valeurs dans E et f une fonction de E dans \mathbf{R} . Si X prend un nombre fini de valeurs $\{x_1, \dots, x_n\}$, alors

$$\mathbf{E}f(X) = \sum_{k=1}^n f(x_k)\mathbf{P}(X = x_k) = \sum_{k=1}^n f(x_k)\mu_X(\{x_k\}).$$

Démonstration. Si la variable X ne prend qu'un nombre fini de valeurs x_1, \dots, x_n , alors c'est également le cas de $f(X)$ et on peut écrire

$$f(X) = \sum_{k=1}^n f(x_k)\mathbf{1}_{\{X=x_k\}}.$$

Le résultat se déduit alors de la définition (2.2) de l'espérance d'une variable étagée. □

On remarque en particulier que $\mathbf{E}f(X)$ ne dépend que de f et de la loi μ_X de X : si X et Y sont deux variables aléatoires de même loi, alors $\mathbf{E}f(X) = \mathbf{E}f(Y)$.

2.3 Variance, moments d'ordres supérieurs

Certaines propriétés intéressantes des variables aléatoires s'expriment à partir des espérances des puissances de X , qui s'appellent les *moments* de X :

Définition 2.3.1. Pour $n \in \mathbf{N}^*$, si X^n est intégrable, la quantité $\mathbf{E}(X^n)$ est appelée moment d'ordre n de la variable X . On dira alors que X admet un moment d'ordre n .

En particulier, pour $n = 2$, si X^2 est intégrable, on dira que X est de carré intégrable.

Le fait d'admettre un moment d'ordre n est d'autant plus restrictif que n est grand :

Propriété 2.3.2. *Soit X une variable aléatoire réelle et $n < m$ deux entiers. Si X admet un moment d'ordre m , alors elle admet un moment d'ordre n .*

En particulier une variable aléatoire de carré intégrable est intégrable.

Démonstration. Soit $x \geq 0$. Si $x \leq 1$, alors $x^n \leq 1 \leq 1 + x^m$. Si $1 < x$, alors $1 < x^{m-n}$, d'où $x^n < x^m < 1 + x^m$. On a donc dans tous les cas $x^n \leq 1 + x^m$. Par conséquent :

$$\mathbf{E}(|X|^n) \leq \mathbf{E}(1 + |X|^m) = 1 + \mathbf{E}(|X|^m) < \infty.$$

□

Propriété 2.3.3 (Inégalité de Cauchy-Schwarz). *Si X et Y sont deux variables aléatoires réelles de carré intégrable, alors la variable $|XY|$ est intégrable et on a*

$$\mathbf{E}|XY| \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$$

En particulier, pour $Y = 1$, on trouve :

$$\mathbf{E}|X| \leq \sqrt{\mathbf{E}(X^2)}.$$

Cette dernière égalité est une autre manière de voir qu'une variable aléatoire de carré intégrable est intégrable.

Démonstration. Pour tous réels x et y , on a l'inégalité $0 \leq (|x| - |y|)^2 = x^2 + y^2 - 2|xy|$. En particulier, on a $|xy| \leq \frac{1}{2}(x^2 + y^2)$. Par conséquent $\mathbf{E}|XY| \leq \frac{1}{2}(\mathbf{E}X^2 + \mathbf{E}Y^2) < \infty$. La variable aléatoire XY est donc intégrable.

Par linéarité de l'espérance, on a pour tout réel λ

$$\mathbf{E}((|X| + \lambda|Y|)^2) = \mathbf{E}(X^2) + 2\lambda\mathbf{E}(|XY|) + \lambda^2\mathbf{E}(Y^2).$$

Vue comme une fonction de λ , cette expression est un polynôme de degré 2 qui ne prend que des valeurs positives. Par conséquent, son discriminant est négatif, ce qui s'écrit

$$4\mathbf{E}(|XY|)^2 - 4\mathbf{E}(Y^2)\mathbf{E}(X^2) \leq 0.$$

Cette inégalité n'est qu'une réécriture de celle que l'on cherchait à démontrer. □

Corollaire 2.3.4. *Si deux variables aléatoires réelles X et Y sont de carré intégrable, alors $X + Y$ aussi.*

Démonstration. On a $(X + Y)^2 = X^2 + 2XY + Y^2 \leq X^2 + 2|XY| + Y^2$. Par conséquent

$$\mathbf{E}((X + Y)^2) \leq \mathbf{E}(X^2) + 2\mathbf{E}|XY| + \mathbf{E}(Y^2) \leq \mathbf{E}(X^2) + 2\sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)} + \mathbf{E}(Y^2) < \infty.$$

□

Si X et Y sont deux variables aléatoires réelles de carré intégrable, on peut définir une “distance” entre elles par la formule

$$\delta(X, Y) = \mathbf{E}((X - Y)^2).$$

Par exemple, on verra avec la propriété 2.4.2, que $\delta(X, Y) = 0$ si et seulement si $\mathbf{P}(X = Y) = 1$.

Pour une variable aléatoire réelle X de carré intégrable donnée, on peut se demander quelle est la constante dont elle est la plus proche, au sens de δ . C'est-à-dire, pour quelle(s) valeur(s) λ_0 a-t-on

$$\delta(X, \lambda_0) = \inf_{\lambda \in \mathbf{R}} \delta(X, \lambda).$$

En développant, on voit que $\delta(X, \lambda)$ est polynômiale par rapport à λ :

$$\delta(X, \lambda) = \mathbf{E}((X - \lambda)^2) = \mathbf{E}(X^2) - 2\lambda \mathbf{E}X + \lambda^2.$$

Ce polynôme atteint son minimum en $\lambda = \mathbf{E}X$. Autrement dit, *l'espérance de X peut s'interpréter comme la variable constante la plus “proche” de X* . La quantité $\delta(X, \mathbf{E}X) = \mathbf{E}((X - \mathbf{E}X)^2)$ mesure donc à quel point X est proche d'être constante. Cela motive la définition suivante :

Définition 2.3.5. *Si X est une variable aléatoire de carré intégrable, on définit sa variance, notée $\mathbf{V}(X)$, par :*

$$\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}X)^2).$$

L'écart-type, noté $\sigma(X)$, est défini par $\sqrt{\mathbf{V}(X)}$.

Une variable aléatoire de variance égale à 1 est dite réduite.

On remarquera que si la variable aléatoire X a une dimension, alors l'écart-type a la même dimension, tandis que la variance a la dimension au carré, d'où l'intérêt dans la pratique de travailler avec l'écart-type. Par exemple, si X représente la taille (en cm) d'un individu pris au hasard dans une population, alors l'espérance $\mathbf{E}X$ s'exprime en centimètres, de même que l'écart-type $\sigma(X)$, alors que la variance s'exprime en centimètres carrés. On verra dans la propriété 2.4.3 que l'écart-type correspond à un écart caractéristique à la moyenne. Par exemple, sur la figure 2.3 qui représente les courbes de croissances des enfants jusqu'à trois ans^(note 4), on a tracé les graphes de la moyenne à laquelle est ajouté ou soustrait l'écart-type ou son double. On verra plus tard que l'on peut donner des résultats précis sur la proportion d'enfants qui se situeront entre ces courbes.

Donnons quelques propriétés algébriques de la variance :

Propriété 2.3.6. *Soit X une variable aléatoire de carré intégrable, alors :*

- $\mathbf{V}(X) \geq 0$;
- La variance est invariante par translation : pour tout réel a , on a $\mathbf{V}(X + a) = \mathbf{V}(X)$;
- La variance est quadratique : pour tout réel a , on a $\mathbf{V}(aX) = a^2 \mathbf{V}(X)$;
- On a une autre expression de la variance, qui simplifie souvent certains calculs^(note 5) :

$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 ;$$

- La variance de X est nulle si et seulement si X est une variable aléatoire constante^(note 6).

(note 4). Ce sont les graphiques que l'on retrouve dans les carnets de santé, avec les mêmes courbes.

(note 5). Cette égalité est parfois appelée *formule de König-Huygens*.

(note 6). Plus précisément, si et seulement si il existe un réel λ tel que $\mathbf{P}(X = \lambda) = 1$. Il est a priori possible que X ne soit pas constante, mais avec $\mathbf{P}(X \neq \lambda) = 0$.

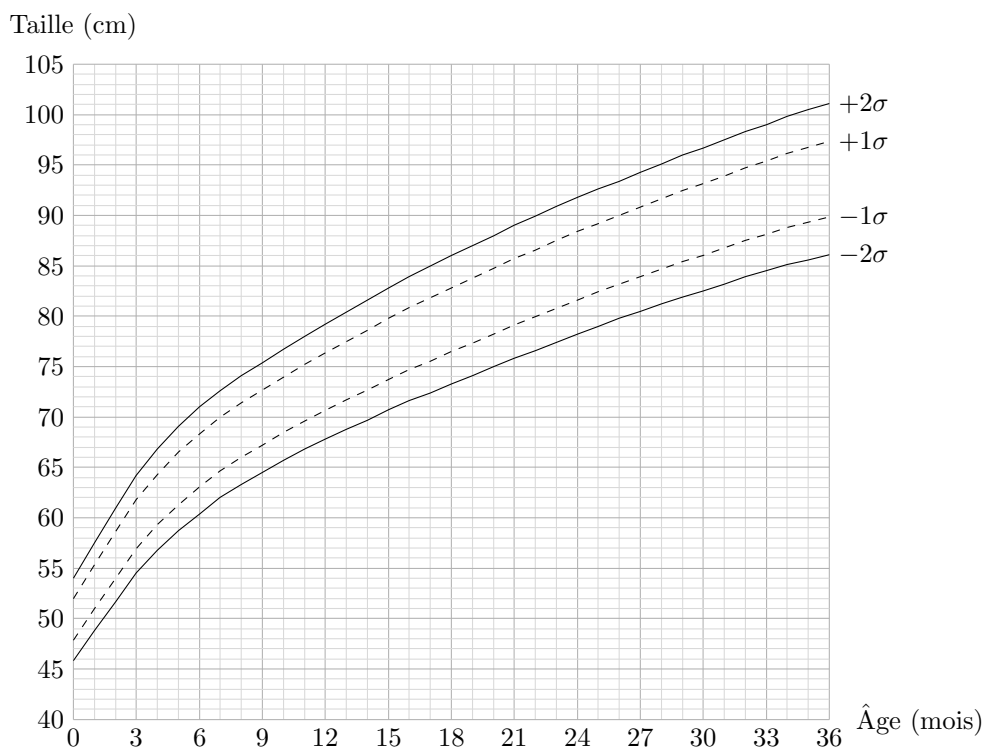


FIGURE 2.3 – Courbe de croissance des enfants de 0 à 36 mois.

Démonstration. La positivité de $\mathbf{V}(X)$ découle de la monotonie de l'espérance et du fait que $(X - \mathbf{E}(X))^2 \geq 0$.

En utilisant la définition de la variance et la linéarité de l'espérance, on a, pour l'invariance par translation :

$$\mathbf{V}(X + a) = \mathbf{E}((X + a - \mathbf{E}(X + a))^2) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{V}(X),$$

et pour le caractère quadratique :

$$\mathbf{V}(aX) = \mathbf{E}((aX - \mathbf{E}(aX))^2) = \mathbf{E}(a^2(X - \mathbf{E}(X))^2) = a^2\mathbf{V}(X).$$

La deuxième expression se prouve en développant :

$$\begin{aligned} \mathbf{V}(X) &= \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2 - 2\mathbf{E}(X)X + \mathbf{E}(X)^2) \\ &= \mathbf{E}(X^2) - 2\mathbf{E}(X)^2 + \mathbf{E}(X)^2 \\ &= \mathbf{E}(X^2) - \mathbf{E}(X)^2. \end{aligned}$$

Pour le dernier point, on considère une variable X de variance 0. La variable $(X - \mathbf{E}X)^2$ est donc une variable positive d'espérance nulle, ce qui implique par la propriété 2.4.2 que l'évènement $\{(X - \mathbf{E}X)^2 = 0\}$ a probabilité 1. Par conséquent, $\mathbf{P}(X = \mathbf{E}X) = 1$, et X est une variable constante. \square

2.4 Inégalité de Markov et de Bienaymé-Tchebychev

On présente dans cette partie deux inégalités permettant de “localiser” les valeurs prises par une variable aléatoire, en fonction de sa moyenne et de sa variance.

2.4.1 Inégalité de Markov

La première inégalité formalise l'intuition suivante : si une variable aléatoire est intégrable alors elle “n'est pas trop grande”, et dans ce cas, sa probabilité de prendre des “grandes” valeurs est “petite”.

Propriété 2.4.1 (Inégalité de Markov). *Soit X une variable aléatoire positive. Alors,*

$$\forall a > 0, \quad \mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a}.$$

Plus généralement, si X est une variable aléatoire admettant un moment d'ordre $n \in \mathbf{N}^*$ et si $a > 0$, alors

$$\mathbf{P}(|X| > a) \leq \frac{\mathbf{E}(|X|^n)}{a^n}.$$

Démonstration. Pour tout réel positif x , on a l'inégalité $a\mathbf{1}_{\{x \geq a\}} \leq x$, comme on peut le voir sur la figure 2.4. En appliquant cette inégalité à la variable X , on obtient $\mathbf{1}_{X \geq a} \leq \frac{X}{a}$.

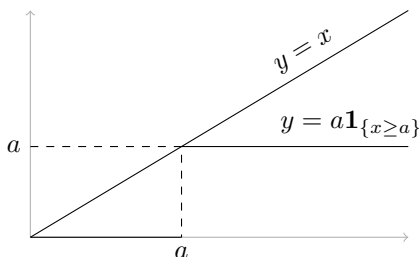


FIGURE 2.4 – L'inégalité $a\mathbf{1}_{x \geq a} \leq x$.

Par conséquent,

$$\mathbf{P}(X \geq a) = \mathbf{E}\mathbf{1}_{\{X \geq a\}} \leq \mathbf{E}\left(\frac{X}{a}\right) = \frac{\mathbf{E}X}{a}.$$

Pour la seconde inégalité, il suffit de remarquer l'égalité entre évènements $\{|X| \geq a\} = \{|X|^n \geq a^n\}$ et d'appliquer la première égalité à $|X|^n$ et a^n . \square

On peut aussi lire l'inégalité de Markov comme “une variable aléatoire positive a peu de chances d'être beaucoup plus grande que sa moyenne”. En effet, en prenant $a = \alpha\mathbf{E}X$, l'inégalité se récrit

$$\mathbf{P}(X \geq \alpha\mathbf{E}X) \leq \frac{1}{\alpha}.$$

L'inégalité de Markov a la conséquence suivante :

Propriété 2.4.2. Si X est une variable positive telle que $\mathbf{E}X = 0$, alors on a $\mathbf{P}(X = 0) = 1$.

Démonstration. On a la décomposition

$$\{X > 0\} = \bigcup_{n=1}^{\infty} \{X \geq 1/n\}.$$

Par l'inégalité de Markov, on obtient

$$\mathbf{P}\left(X \geq \frac{1}{n}\right) \leq n\mathbf{E}X = 0.$$

Donc pour tout n , $\mathbf{P}(X \geq 1/n) = 0$. Par le corollaire 1.2.6, on en déduit $\mathbf{P}(X > 0) = 0$. Comme X est positive, on a donc $\mathbf{P}(X = 0) = 1 - \mathbf{P}(X > 0) = 1$. \square

2.4.2 Inégalité de Bienaymé-Tchebychev

La deuxième inégalité formalise l'intuition suivante : la variance mesure l'écart entre une variable aléatoire et sa moyenne. Par conséquent si une variable a une "petite" variance, elle aura une "grande" probabilité de se trouver "près" de sa moyenne.

Propriété 2.4.3 (Inégalité de Bienaymé-Tchebychev). Soient X une variable aléatoire de carré intégrable et $a > 0$. Alors,

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq a) \leq \frac{\mathbf{V}(X)}{a^2}.$$

Démonstration. L'inégalité de Bienaymé-Tchebychev est une conséquence de l'inégalité de Markov. La variable aléatoire X étant de carré intégrable, il en est de même pour la variable aléatoire $X - \mathbf{E}X$. On applique alors l'inégalité de Markov avec la variable aléatoire $Y = |X - \mathbf{E}X|^2$: pour tout $a > 0$,

$$\mathbf{P}(|X - \mathbf{E}X| \geq a) = \mathbf{P}(|X - \mathbf{E}X|^2 \geq a^2) \leq \frac{\mathbf{E}(|X - \mathbf{E}X|^2)}{a^2} = \frac{\mathbf{V}(X)}{a^2}.$$

\square

En appliquant la propriété précédente à $a = 2\sigma(X)$ et $a = 3\sigma(X)$, on obtient, pour toute variable aléatoire réelle X ,

$$\begin{aligned} \mathbf{P}(|X - \mathbf{E}X| \leq 2\sigma(X)) &\geq \frac{3}{4} = 0.75 \\ \mathbf{P}(|X - \mathbf{E}X| \leq 3\sigma(X)) &\geq \frac{8}{9} = 0.888\dots \end{aligned}$$

Ces propriétés sont encore une illustration du fait que l'écart-type mesure la dispersion de la variable aléatoire autour de sa moyenne. Par exemple, la première nous dit qu'au moins les trois quarts des enfants ont leur taille comprise entre les deux courbes en trait plein de la figure 2.3.

2.5 Exercices

Exercice 2.1.

Soit X une variable aléatoire intégrable. Montrer que les variables aléatoires suivantes sont elles aussi intégrables :

$$X \sin(X) ; \sqrt{|X|} ; \ln(1 + |X|) ; \frac{X^3}{1 + X^2}.$$

Exercice 2.2.

Soit X une variable aléatoire à valeurs dans $[-1, 1]$.

1. Montrer que la variance de X est inférieure à 1. Notamment, l'écart-type de X est aussi inférieur à 1.
2. Si on suppose de plus $\mathbf{V}(X) = 1$, que peut-on dire sur la loi de X ?
3. Refaire les deux questions précédente, en remplaçant dans les hypothèses le segment $[-1, 1]$ par un segment $[a, b]$ arbitraire. La borne supérieure de la variance et de l'écart-type sera à adapter.

Chapitre 3

Probabilités conditionnelles - indépendance

3.1 Probabilité conditionnelle

Dans de nombreuses situations, on observe un phénomène aléatoire dont on connaît une information partielle. Par exemple, pendant une partie de cartes à jouer, la distribution des cartes est faite aléatoirement, mais chaque joueur connaît ses propres cartes et dispose donc d'une information sur le jeu des autres.

On voudrait donc, à partir d'un premier espace probabilisé (Ω, \mathbf{P}) , définir une nouvelle mesure de probabilité sur Ω portée par un évènement B : plus spécifiquement, on veut donner une probabilité nulle aux évènements disjoints de B . Cette idée est formalisée dans la notion de *probabilité conditionnelle*.

3.1.1 Définition

À partir d'un espace probabilisé (Ω, \mathbf{P}) , on veut donc construire une nouvelle probabilité qui “ressemble” à \mathbf{P} , mais telle que les sous-ensembles de B^c aient une probabilité nulle. Une manière de faire pourrait être de poser $\mathbf{P}_B(A) = \mathbf{P}(A \cap B)$, de sorte que $\mathbf{P}_B(A) = \mathbf{P}(A)$ pour tout $A \subset B$. Le problème est que l'objet obtenu ne serait pas une mesure de probabilité, puisqu'on aurait $\mathbf{P}_B(\Omega) = \mathbf{P}(B) \neq 1$ (a priori). La solution consiste à renormaliser par la quantité $\mathbf{P}(B)$.

Définition 3.1.1. Soit (Ω, \mathbf{P}) un espace probabilisé, et B un évènement tel que $\mathbf{P}(B) > 0$. Pour tout $A \subset \Omega$, on définit la probabilité conditionnelle de A sachant B , notée $\mathbf{P}_B(A)$ ou $\mathbf{P}(A|B)$, par :

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}. \quad (3.1)$$

La condition $\mathbf{P}(B) > 0$ permet de diviser par $\mathbf{P}(B)$. En pratique, on a $0 < \mathbf{P}(B) < 1$. En effet, le cas $\mathbf{P}(B) = 1$ a peu d'intérêt, puisqu'on obtient alors $\mathbf{P}_B = \mathbf{P}$.

La notation $\mathbf{P}(A|B)$ est la plus couramment utilisée ; en revanche, la notation figurant au programme du lycée est $\mathbf{P}_B(A)$. L'intérêt de la notation $\mathbf{P}_B(A)$ est qu'elle souligne bien le fait que \mathbf{P}_B est une mesure de probabilité que l'on applique à l'évènement A . En outre, la notation $\mathbf{P}(A|B)$ pourrait faire croire

que l'on calcule la probabilité d'un évènement noté $A|B$. Or, on ne peut pas donner de sens à un "évènement $A|B$ ".

Propriété 3.1.2. *L'application \mathbf{P}_B est une probabilité sur Ω .*

Démonstration. Il faut montrer que \mathbf{P}_B satisfait les trois axiomes caractérisant une probabilité.

- Par hypothèse, on a $\mathbf{P}(B) > 0$. De plus, pour tout $A \subset \Omega$, on a $A \cap B \subset B$, d'où l'inégalité $0 \leq \mathbf{P}(A \cap B) \leq \mathbf{P}(B)$, et donc :

$$0 \leq \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \leq 1.$$

- Comme $\Omega \cap B = B$, et $\Omega \cap \emptyset = \emptyset$, on a bien $\mathbf{P}_B(\Omega) = 1$ et $\mathbf{P}_B(\emptyset) = 0$
- Soit $(A_n)_{n \geq 1}$ une famille dénombrable d'évènements deux à deux disjoints. Alors on a l'égalité $(\bigcup_{n \geq 1} A_n) \cap B = \bigcup_{n \geq 1} (A_n \cap B)$. Comme les évènements $(A_n)_{n \geq 1}$ sont deux à deux disjoints, il en est de même pour les évènements $(A_n \cap B)_{n \geq 1}$. Par conséquent, on a

$$\mathbf{P} \left(\bigcup_{n \geq 1} (A_n \cap B) \right) = \sum_{n \geq 1} \mathbf{P}(A_n \cap B),$$

et on conclut :

$$\mathbf{P}_B \left(\bigcup_{n \geq 1} A_n \right) = \frac{\mathbf{P} \left((\bigcup_{n \geq 1} A_n) \cap B \right)}{\mathbf{P}(B)} = \sum_{n \geq 1} \frac{\mathbf{P}(A_n \cap B)}{\mathbf{P}(B)} = \sum_{n \geq 1} \mathbf{P}_B(A_n).$$

□

La définition de la probabilité conditionnelle peut se récrire $\mathbf{P}(A \cap B) = \mathbf{P}_B(A)\mathbf{P}(B)$. Autrement dit, on peut calculer la probabilité d'une intersection en conditionnant par un des évènements. En appliquant plusieurs fois la même égalité, on arrive à l'équation suivante :

$$\mathbf{P} \left(\bigcap_{k=1}^n A_k \right) = \mathbf{P}(A_1)\mathbf{P}_{A_1}(A_2)\mathbf{P}_{A_1 \cap A_2}(A_3) \cdots \mathbf{P}_{A_1 \cap \cdots \cap A_{n-1}}(A_n), \quad (3.2)$$

où les $(A_k)_{1 \leq k \leq n}$ sont des évènements tels que $\mathbf{P} \left(\bigcap_{k=1}^{n-1} A_k \right) > 0$. L'égalité (3.2) peut se récrire de manière intuitive à l'aide d'arbres de probabilité, qui permettent de calculer aisément les probabilités de certains évènements.

3.1.2 Arbres de probabilité

Les arbres de probabilité sont un outil permettant de calculer simplement diverses probabilités concernant une même expérience aléatoire.

Donnons d'abord une définition de la notion d'arbre.

Définition 3.1.3. *Un arbre (orienté) est un ensemble \mathcal{A} , dont un des éléments x (appelé racine de l'arbre) est spécifié, et dans lequel on associe à chaque élément de \mathcal{A} (on parlera aussi de sommet) un sous-ensemble de \mathcal{A} qui est l'ensemble de ses enfants.*

On suppose que chaque sommet est l'enfant d'un seul autre sommet, à l'exception de la racine, qui n'a pas de père.

Par ailleurs, on suppose qu'en partant d'un sommet puis en allant à son père, puis au père de son père, et ainsi de suite, on ne retombera jamais sur le sommet initial. En particulier, on finira toujours par atteindre la racine de l'arbre.

Un exemple graphique est probablement plus parlant : on représentera les sommets d'un arbre en les reliant d'une flèche à leurs enfants. Sur la figure 3.1, le graphe de gauche est un arbre dont la racine est R . Les enfants de R sont A et B . Les enfants de B sont C , D et E . Les sommets A , C , D et E n'ont pas d'enfants. Le graphe de droite n'est pas un arbre car le sommet C est l'enfant de plusieurs sommets.

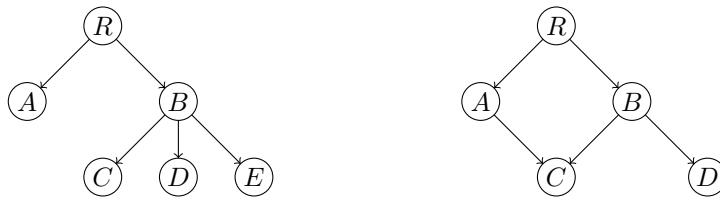


FIGURE 3.1 – Exemple et contre-exemple d'arbres.

Définition 3.1.4. *Un arbre de probabilité est un arbre orienté dont chaque sommet est un évènement de probabilité non nulle tel que :*

- la racine de l'arbre est l'évènement Ω ;
- si A est un sommet de l'arbre, alors l'ensemble des enfants de A constitue une partition^(note 1) de A .

À chaque arête reliant un sommet A à un de ses enfants B , on associe la probabilité $\mathbf{P}_A(B)$.

Dans de nombreux cas, les enfants d'un sommet A sont de la forme $A \cap B_i$, où les B_i constituent une partition de Ω . On a alors bien $\bigcup_i A \cap B_i = A \cap \bigcup_i B_i = A \cap \Omega = A$, avec une union disjointe.

La somme des probabilités des branches issues d'un même sommet est 1. En effet, si \mathcal{E} est l'ensemble des enfants d'un sommet A , on a par hypothèse $\bigcup_{B \in \mathcal{E}} B = A$, où l'union est disjointe. Comme \mathbf{P}_A est une mesure de probabilité, on a donc

$$\sum_{B \in \mathcal{E}} \mathbf{P}_A(B) = \mathbf{P}_A\left(\bigcup_{B \in \mathcal{E}} B\right) = \mathbf{P}_A(A) = 1.$$

On donne un exemple d'arbre de probabilité en figure 3.2, sur laquelle on a les décompositions $\Omega = A_1 \cup A_2$, ainsi que $A_2 = A_{21} \cup A_{22} \cup A_{23}$ et $A_{21} = A_{211} \cup A_{212}$ où les unions sont disjointes. Comme les sommes des probabilités partant d'un sommet doivent valoir 1, on a aussi $a + b = c + d + e = f + g = 1$.

L'intérêt des arbres de probabilité est que l'on peut retrouver la probabilité de chacun des sommets par simple produit, en vertu de l'équation (3.2) :

Propriété 3.1.5. *Soit (A_0, A_1, \dots, A_n) un chemin de l'arbre, au sens où A_i est toujours le père de A_{i+1} , avec $A_0 = \Omega$. Alors $\mathbf{P}(A_n)$ est donné par le produit des probabilités sur les branches formant le chemin.*

(note 1). On rappelle qu'une *partition* de A est une famille $(B_i)_{i \in I}$ d'ensembles *disjoints* tels que $\bigcup_{i \in I} B_i = A$.

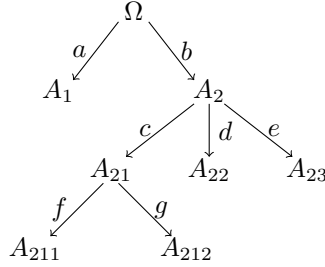


FIGURE 3.2 – Un exemple d’arbre de probabilité.

Démonstration. C’est une récurrence sur la longueur du chemin : si le chemin est de longueur 0, il est réduit à $A_0 = \Omega$, et alors $\mathbf{P}(A_0) = \mathbf{P}(\Omega) = 1$. Or, il n’y a pas de branche sur le chemin, et un produit vide vaut 1.

Ensuite, on suppose la propriété vraie pour les chemins de longueur n . Si (A_0, \dots, A_{n+1}) est un chemin de longueur $n + 1$, alors $\mathbf{P}(A_{n+1}) = \mathbf{P}_{A_n}(A_{n+1})\mathbf{P}(A_n)$. Or, par hypothèse de récurrence, $\mathbf{P}(A_n)$ est le produit des n premières probabilités présentes sur les branches du chemin, et $\mathbf{P}_{A_n}(A_{n+1})$ est la $n + 1$ ème probabilité. \square

Pour illustration, sur la figure 3.2, on trouve par exemple $\mathbf{P}(A_{211}) = bcf$, $\mathbf{P}(A_{23}) = be$ ou encore $\mathbf{P}(A_{21}) = bc$.

3.1.3 Formule des probabilités totales et formule de Bayes

Dans cette partie, on présente deux formules importantes faisant intervenir des probabilités conditionnelles.

La première formule permet de calculer la probabilité d’un évènement à partir d’une partition de Ω .

Théorème 3.1.6 (Formule des probabilités totales). *Soit $(B_n)_{n \in \mathbf{N}}$ une partition de Ω , telle que pour tout $n \in \mathbf{N}$, $\mathbf{P}(B_n) > 0$, et soit $A \subset \Omega$. Alors :*

$$\mathbf{P}(A) = \sum_{n \in \mathbf{N}} \mathbf{P}(A \cap B_n) = \sum_{n \in \mathbf{N}} \mathbf{P}_{B_n}(A)\mathbf{P}(B_n).$$

La formule reste valide si on remplace la partition infinie dénombrable $(B_n)_{n \in \mathbf{N}}$ par une partition finie $(B_n)_{n \in \{1, \dots, N\}}$.

Démonstration. Comme $(B_n)_{n \in \mathbf{N}}$ est une partition de Ω , on a :

$$A = A \cap \Omega = A \cap \left(\bigcup_{n \in \mathbf{N}} B_n \right) = \bigcup_{n \in \mathbf{N}} (A \cap B_n),$$

où les évènements $(A \cap B_n)_{n \in \mathbf{N}}$ sont deux à deux disjoints. D’où le résultat. \square

Un exemple courant d'application de la formule des probabilités totale est lorsque la famille $(B_n)_{n \in \mathbf{N}}$ est de la forme $B_n = \{N = n\}$, pour une variable aléatoire N . La formule s'écrit alors

$$\mathbf{P}(A) = \sum_{n \in \mathbf{N}} \mathbf{P}_{\{N=n\}}(A) \times \mathbf{P}(N = n).$$

Intuitivement, cela signifie que l'on "rend N déterministe", en considérant une par une toutes les valeurs qu'elle peut prendre, en pondérant à chaque fois par la probabilité que N prenne la valeur en question.

La deuxième formule présentée dans cette partie permet d'exprimer la probabilité conditionnelle de A sachant B à l'aide de la probabilité conditionnelle de B sachant A .

Théorème 3.1.7 (Formule de Bayes). *Soient A et B deux évènements, tels que $\mathbf{P}(A) > 0$ et $\mathbf{P}(B) > 0$, alors :*

$$\mathbf{P}_B(A) = \frac{\mathbf{P}_A(B)\mathbf{P}(A)}{\mathbf{P}(B)}.$$

Démonstration. Par définition de la probabilité conditionnelle, on a :

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}_A(B)\mathbf{P}(A)}{\mathbf{P}(B)}.$$

□

3.2 Indépendance d'évènements

Définition 3.2.1. *Deux évènements A et B sont dits indépendants, si*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B). \quad (3.3)$$

Si A est un évènement de probabilité non nulle, l'équation (3.3) peut se récrire $\mathbf{P}(B) = \mathbf{P}_A(B)$. Cette équation se comprend comme le fait que la connaissance de A ne modifie pas la probabilité de B .

Par exemple, les évènements \emptyset et Ω sont indépendants de n'importe quel autre évènement A . En effet,

$$\mathbf{P}(\emptyset \cap A) = \mathbf{P}(\emptyset) = 0 = 0 \times \mathbf{P}(A) = \mathbf{P}(\emptyset)\mathbf{P}(A),$$

de même

$$\mathbf{P}(\Omega \cap A) = \mathbf{P}(A) = 1 \times \mathbf{P}(A) = \mathbf{P}(\Omega)\mathbf{P}(A).$$

L'indépendance est stable par passage au complémentaire :

Propriété 3.2.2. *Si les évènements A et B sont indépendants, alors il en est de même des évènements A^c et B , A et B^c , A^c et B^c .*

Démonstration. On démontre seulement que A^c et B sont indépendants, les autres cas s'en déduisent. Comme $B = (A \cap B) \cup (A^c \cap B)$ avec une union disjointe, on a

$$\begin{aligned} \mathbf{P}(A^c \cap B) &= \mathbf{P}(B) - \mathbf{P}(A \cap B) = \mathbf{P}(B) - \mathbf{P}(A)\mathbf{P}(B) \\ &= \mathbf{P}(B)(1 - \mathbf{P}(A)) \\ &= \mathbf{P}(B)\mathbf{P}(A^c). \end{aligned}$$

□

Quand on a plus que deux évènements, il y a deux généralisations de la notion d'indépendance, qui ne sont pas équivalentes.

Définition 3.2.3. Soient $(A_n)_{n \in \mathbf{N}}$ une suite d'évènements.

— Les $(A_n)_{n \in \mathbf{N}}$ sont dits mutuellement indépendants, si pour tout entier $k \geq 1$ et tout k -uplet d'entiers $1 \leq i_1 < \dots < i_k$, on a

$$\mathbf{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \cdots \mathbf{P}(A_{i_k}).$$

— Les $(A_n)_{n \in \mathbf{N}}$ sont dits deux à deux indépendants, si pour tous entiers $1 \leq i_1 < i_2$, on a :

$$\mathbf{P}(A_{i_1} \cap A_{i_2}) = \mathbf{P}(A_{i_1}) \mathbf{P}(A_{i_2}).$$

Propriété 3.2.4. Des évènements mutuellement indépendants sont deux à deux indépendants.

Démonstration. La définition de l'indépendance deux à deux est un cas particulier de l'indépendance mutuelle, pour $k = 2$. \square

Il existe toutefois des familles d'évènements qui sont indépendants deux à deux sans être mutuellement indépendants. Par exemple, si on se donne deux variables aléatoires X et Y telles que $\mathbf{P}(X = \pm 1, Y = \pm 1) = \frac{1}{4}$ (pour tous les choix de signes possibles), alors les évènements $\{X = 1\}$, $\{Y = 1\}$ et $\{XY = 1\}$ sont indépendants deux à deux, mais pas mutuellement indépendants. En effet, dans ce cas, on a

$$\mathbf{P}(X = 1) = \mathbf{P}(X = 1, Y = 1) + \mathbf{P}(X = 1, Y = -1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

On a de même $\mathbf{P}(Y = 1) = \frac{1}{2}$. On a donc bien

$$\mathbf{P}(X = 1, Y = 1) = \frac{1}{4} = \mathbf{P}(X = 1)\mathbf{P}(Y = 1).$$

On a également

$$\mathbf{P}(XY = 1) = \mathbf{P}(X = 1, Y = 1) + \mathbf{P}(X = -1, Y = -1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

d'où

$$\mathbf{P}(X = 1, XY = 1) = \mathbf{P}(X = 1, Y = 1) = \frac{1}{4} = \mathbf{P}(X = 1)\mathbf{P}(XY = 1),$$

et de même, on a $\mathbf{P}(Y = 1, XY = 1) = \mathbf{P}(Y = 1)\mathbf{P}(XY = 1)$. L'indépendance deux à deux est donc bien vérifiée. En revanche, ce n'est pas le cas de l'indépendance mutuelle puisque :

$$\mathbf{P}(X = 1, Y = 1, XY = 1) = \mathbf{P}(X = 1, Y = 1) = \frac{1}{4} \neq \mathbf{P}(X = 1)\mathbf{P}(Y = 1)\mathbf{P}(XY = 1).$$

On peut expliquer cet exemple de la manière suivante : X et Y valent 1 ou -1 avec même probabilité, et connaître X ne donne aucune information sur la valeur de Y (et inversement). On remarque aussi que connaître X ne donne aucune information sur la valeur de XY (et de même, connaître Y ne donne aucune information sur la valeur de XY). En revanche, connaître X et Y nous donne la valeur de XY .

3.3 Variables aléatoires indépendantes

Dans l'exemple de la fin de la partie 3.2, les évènements indépendants étaient obtenus à partir de variables aléatoires. Cela nous amène à la notion de *variables aléatoires indépendantes*.

Définition 3.3.1. *Deux variables aléatoires X et Y à valeurs dans E sont dites indépendantes, si pour tous sous-ensembles I et J de E les évènements $\{X \in I\}$ et $\{Y \in J\}$ sont indépendants :*

$$\mathbf{P}(\{X \in I\} \cap \{Y \in J\}) = \mathbf{P}(X \in I)\mathbf{P}(Y \in J). \quad (3.4)$$

Pour le cas de plusieurs variables on a la même généralisation que dans le cas des évènements.

Définition 3.3.2. *Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires à valeurs dans E .*

- *On dit que les X_n sont deux à deux indépendantes si, pour tout n et m , les variables X_n et X_m sont indépendantes.*
- *On dit que les variables X_n sont mutuellement indépendantes si pour toute suite $(I_n)_{n \in \mathbf{N}}$ de sous-ensembles de E , les évènements $\{X_n \in I_n\}$ sont mutuellement indépendants.*

Reprenons l'exemple de la fin de la partie 3.2, où l'on avait deux variables X et Y vérifiant $\mathbf{P}(X = \pm 1, Y = \pm 1) = \frac{1}{4}$. Dans ce cas, les trois variables X , Y et XY sont indépendantes deux à deux, mais pas mutuellement.

On peut récrire la définition de l'indépendance en utilisant des espérances. En effet, l'équation (3.4) peut se récrire

$$\mathbf{E}(\mathbf{1}_I(X)\mathbf{1}_J(Y)) = \mathbf{E}(\mathbf{1}_I(X))\mathbf{E}(\mathbf{1}_J(Y)).$$

Par conséquent, si on applique cette relation à plusieurs sous-ensembles $(I_n)_{n=1, \dots, N}$ et $(J_m)_{m=1, \dots, M}$ de E , on obtient par linéarité la relation suivante :

$$\mathbf{E} \left(\sum_{n=1}^N \alpha_n \mathbf{1}_{I_n}(X) \sum_{m=1}^M \beta_m \mathbf{1}_{J_m}(Y) \right) = \mathbf{E} \left(\sum_{n=1}^N \alpha_n \mathbf{1}_{I_n}(X) \right) \mathbf{E} \left(\sum_{m=1}^M \beta_m \mathbf{1}_{J_m}(Y) \right)$$

Si on définit alors deux fonctions f et g de E dans \mathbf{R} par :

$$f(x) = \sum_{i=1}^N \mathbf{1}_{I_i}(x) \text{ et } g(y) = \sum_{m=1}^M \mathbf{1}_{J_m}(y), \quad (3.5)$$

on a montré la relation

$$\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y)),$$

pour toutes fonctions de la forme (3.5), c'est-à-dire, toutes fonctions ne prenant qu'un nombre fini de valeurs. Cette relation se généralise en fait à toutes fonctions f et g .

Propriété 3.3.3. *Soit X et Y deux variables aléatoires. Les assertions suivantes sont équivalentes.*

1. *Les variables aléatoires X et Y sont indépendantes ;*
2. *Pour toutes fonctions bornées f et g de E dans \mathbf{R} , on a :*

$$\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y)).$$

3. Pour toutes fonctions f et g de E dans $[0, \infty[$, on a :

$$\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y)) ;$$

4. Pour toutes fonctions f et g de E dans \mathbf{R} telles que $f(X)$, $g(Y)$ et $f(X)g(Y)$ soient intégrables, on a :

$$\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y)).$$

Les conditions sur f et g permettent de s'assurer que les espérances sont bien définies : dans le cas 3, il n'y a que des variables positives, dont l'espérance est bien définie ; dans le cas 2, il n'y a que des variables aléatoires bornées, donc intégrables.

Démonstration. 2. \Rightarrow 1. L'indépendance se déduit de l'égalité entre les espérances en choisissant $f = \mathbf{1}_I$ et $g = \mathbf{1}_J$.

1. \Rightarrow 2. On a vu que l'égalité est bien vérifiée si f et g sont de la forme $f(x) = \sum_{n=1}^N \alpha_n \mathbf{1}_{I_n}$ et $g(x) = \sum_{m=1}^M \beta_m \mathbf{1}_{J_m}$. Comme dans la propriété 2.1.2, on peut approcher des fonctions bornées f et g aussi près que l'on veut par des fonctions de cette forme, ce qui fait que l'égalité $\mathbf{E}(f(X)g(Y)) = \mathbf{E}(f(X))\mathbf{E}(g(Y))$ reste vraie pour les fonctions bornées.

Les propriétés 1. et 2. sont donc équivalentes. L'équivalence entre les propriétés 2., 3. et 4., est admise. \square

On déduit de la propriété 3.3.3 le fait suivant : si les variables aléatoires X et Y sont indépendantes, alors pour toutes fonctions f et g , les variables aléatoires $f(X)$ et $g(Y)$ sont aussi indépendantes.

Propriété 3.3.4. Soient X et Y deux variables aléatoires de carré intégrable. Alors, si X et Y sont indépendantes,

1. $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$,
2. $\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y)$.

Démonstration. La première égalité découle de la propriété 3.3.3 appliquée à $f(x) = x$ et $g(y) = y$.

Pour la deuxième, on suppose, quitte à changer X en $X - \mathbf{E}X$, que les variables sont centrées. On a alors

$$\begin{aligned} \mathbf{V}(X + Y) &= \mathbf{E}((X + Y)^2) - (\mathbf{E}(X + Y))^2 \\ &= \mathbf{E}(X^2) + \mathbf{E}(Y^2) + 2\mathbf{E}(XY) - (\mathbf{E}X)^2 - (\mathbf{E}Y)^2 - 2\mathbf{E}X\mathbf{E}Y \\ &= [\mathbf{E}(X^2) - (\mathbf{E}X)^2] + [\mathbf{E}(Y^2) - (\mathbf{E}Y)^2] + [2\mathbf{E}(XY) - 2\mathbf{E}X\mathbf{E}Y] \\ &= \mathbf{V}(X) + \mathbf{V}(Y) + 2 \times 0. \end{aligned}$$

\square

Attention, l'égalité $\mathbf{E}(XY) = \mathbf{E}X\mathbf{E}Y$ n'implique pas nécessairement que les variables aléatoires X et Y sont indépendantes. Par exemple, soient X et Y deux variables indépendantes avec $\mathbf{P}(X = \pm 1) = \frac{1}{2}$ et $\mathbf{P}(Y = 1) = \mathbf{P}(Y = 2) = \frac{1}{2}$. Par conséquent, $\mathbf{E}X = \frac{1}{2} \times 1 + \frac{1}{2} \times (-1) = 0$. On a donc

$$\mathbf{E}(Y \times (XY)) = \mathbf{E}(Y^2 \times X) = \mathbf{E}(Y^2)\mathbf{E}X = \mathbf{E}(Y^2) \times 0 = 0$$

et

$$\mathbf{E}Y\mathbf{E}(XY) = (\mathbf{E}Y)^2\mathbf{E}X = 0.$$

On a donc bien $\mathbf{E}(Y \times (XY)) = \mathbf{E}Y \times \mathbf{E}(XY)$, or les deux variables Y et XY ne sont pas indépendantes. En effet, $\mathbf{P}(Y = 1, XY = 2) = 0$, alors que $\mathbf{P}(Y = 1)\mathbf{P}(XY = 2) = \frac{1}{2} \times \frac{1}{4}$.

Nous finissons ce chapitre sur un résultat semblable dans l'esprit à la propriété 1.3.4. On va souvent donner des hypothèses sous la forme "Soient X et Y deux variables aléatoires indépendantes de loi (...)". En toute rigueur, on doit s'assurer qu'il existe un espace probabilisé sur lequel on puisse définir de telles variables aléatoires, sans quoi le résultat énoncé serait vide. C'est le but du théorème suivant :

Théorème 3.3.5. Soient μ et ν deux mesures de probabilité sur un ensemble E . Alors il existe un espace probabilisé (Ω, \mathbf{P}) et deux variables aléatoires X et Y sur (Ω, \mathbf{P}) telles que X et Y sont indépendantes et de loi respectives μ et ν .

Démonstration. On pose $\Omega = E \times E$, et on définit

$$\left\{ \begin{array}{l} X : E \times E \rightarrow E, \\ (x, y) \mapsto x, \end{array} \right. \text{ et } \left\{ \begin{array}{l} Y : E \times E \rightarrow E, \\ (x, y) \mapsto y. \end{array} \right.$$

On doit alors définir une probabilité sur $\Omega = E \times E$ telle que pour tout I et J inclus dans E , on ait

$$\mathbf{P}(X \in I, Y \in J) = \mathbf{P}(\{(x, y) \in E \times E, x \in I, y \in J\}) = \mu(I)\nu(J).$$

Dans le cas où E est indénombrable cette construction est complexe et nous ne la ferons pas. Dans le cas où E est dénombrable, on peut remarquer que $E \times E$ est également dénombrable. On verra au chapitre 4 (théorème 4.1.1) qu'il suffit alors de définir $\mathbf{P}(\{(x, y)\})$ pour tout élément (x, y) de $E \times E$. Ici le choix est nécessairement $\mathbf{P}(\{(x, y)\}) = \mu(\{x\})\nu(\{y\})$. \square

En appliquant plusieurs fois le résultat précédent, on peut construire plusieurs variables mutuellement indépendantes de lois données.

3.4 Exercices

Exercice 3.1.

On considère le lancer de deux dés dont les résultats sont notés X et Y . Pour $2 \leq k \leq 12$, que vaut $\mathbf{P}(X = n | X + Y = k)$, pour $n = 1, \dots, 6$?

Exercice 3.2.

Soient X et Y deux variables aléatoires telles que $\mathbf{P}(X = n, Y = m) = \frac{1}{N^2}$ pour tous entiers n et m de $\{1, \dots, N\}$. On pose $U = \min(X, Y)$ et $V = \max(X, Y)$.

1. Les variables U et V sont elles indépendantes ?
2. Calculer $\mathbf{P}(U = n, V = m)$ pour tous $1 \leq n, m \leq N$.
3. Calculer $\mathbf{P}(V = m)$ pour tout $1 \leq m \leq N$.
4. Que vaut $\mathbf{P}(U = n | V = m)$, pour n et m dans $\{1, \dots, N\}$?
5. Que vaut $\mathbf{P}(X = n | V = m)$, pour n et m dans $\{1, \dots, N\}$?

Exercice 3.3.

Le tableau suivant récapitule les taux de succès de deux traitements permettant de soigner les calculs rénaux ^(note 2). Les traitements seront appelés *traitement A* et *traitement B*. Par ailleurs, on regroupe les patients en deux groupes, suivant la taille des calculs rénaux.

	Traitement A	Traitement B
Petits calculs	81/87	234/270
Gros calculs	192/263	55/80

- Calculer la probabilité de succès conditionnellement à l'utilisation de chacun des traitements ;
- Calculer la probabilité de succès conditionnellement à l'utilisation de chacun des traitements *et* de la taille des calculs ;
- Explication ?

Exercice 3.4.

Une urne contient trois boules blanches et deux boules noires. On tire successivement et “au hasard” trois boules dans cette urne, en respectant le protocole suivant : on remet la boule dans l’urne si elle est noire, on ne la remet pas si elle est blanche.

1. Quelle est la probabilité de n’obtenir aucune boule blanche ?
2. Quelle est la probabilité d’obtenir trois boules blanches ?
3. Quelle est la probabilité d’obtenir exactement une boule blanche ?

Exercice 3.5.

On choisit une famille “au hasard” parmi toutes les familles ayant deux enfants.

1. Sachant que la famille choisie a au moins un garçon, quelle est la probabilité qu’elle ait deux garçons ?
2. Sachant que l’aîné de la famille choisie est un garçon, quelle est la probabilité que le plus jeune soit aussi un garçon ?

Exercice 3.6.

On considère trois cartes à jouer de même forme. Les deux faces de la première carte ont été colorées en noir, les deux faces de la deuxième en rouge tandis que la troisième porte une face noire et une face rouge. On mélange les trois cartes au fond d’un chapeau puis une carte est tirée au hasard et placée sur la table. Si la face apparente est rouge, quelle est la probabilité que l’autre soit noire ?

Exercice 3.7.

Une urne contient 9 boules indiscernables, numérotée de 1 à 9. On tire une boule “au hasard”. Les événements suivants sont-ils indépendants ?

1. A : “la boule tirée porte un numéro pair”,
2. B : “le numéro tiré est multiple de 3”.

Répondre à la même question lorsque l’urne contient 12 boules.

Exercice 3.8.

On considère une urne contenant deux boules noires et une boule blanche. On effectue trois fois de suite les opérations suivantes :

(note 2). Il s’agit de données réelles, voir Charig CR, Webb DR, Payne SR, Wickham JE, *Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy*, Br Med J (Clin Res Ed). 1986 Mar 29 ;292(6524) :879-82.

— On tire une boule de l'urne ;
— On replace dans l'urne la boule tirée et on y ajoute une boule supplémentaire de la même couleur.
Ainsi, à chaque étape, l'urne contient une boule de plus.

1. Construire l'arbre de probabilité associé à l'expérience.
2. Quelle est la probabilité de tirer plus de boules noires que de boules blanches ?
3. Pour chaque étape, quelle est le nombre moyen de boules noires dans l'urne ?

Exercice 3.9.

On dispose d'une urne dans laquelle sont placées n boules noires ou blanches, n étant ici un entier fixé. On ne connaît pas la composition exacte de l'urne, au sens où le nombre de boules noires est une variable aléatoire K vérifiant $\mathbf{P}(K = k) = \frac{1}{n+1}$, pour tout entier $0 \leq k \leq n$. Le nombre de boules blanches est donc $n - K$.

1. On tire une boule dans l'urne. Quelle est la probabilité que cette boule soit noire ?
2. On remet la première boule dans l'urne, puis on tire de nouveau une boule. Quelle est la probabilité que les deux boules soient de la même couleur ?

Exercice 3.10.

Une certaine maladie touche une personne sur 10000 dans une population. On dispose pour cette maladie d'un test de dépistage qui n'est pas totalement fiable :

- Si une personne est malade, elle sera testée négative à la maladie avec probabilité 2% ;
- Si une personne est saine, elle sera testée positive à la maladie avec probabilité 1%.

Un individu choisi au hasard est testé positif. Quelle est la probabilité qu'il soit effectivement malade ?

Chapitre 4

Variables aléatoires discrètes - dénombrément

4.1 Espaces probabilisés dénombrables (ou finis)

Dans ce chapitre, nous allons considérer des variables aléatoires à valeurs dans un ensemble E fini ou dénombrable. Les exemples les plus courants seront $E = \{1, \dots, N\}$ ou $E = \mathbf{N}$. Il se trouve que dans le cas d'un ensemble E dénombrable ^(note 1), il est simple de décrire l'ensemble des mesures de probabilité sur E .

Théorème 4.1.1. *Soit μ une mesure de probabilité sur un ensemble E dénombrable. Alors pour tout sous-ensemble A de E , on a*

$$\mu(A) = \sum_{x \in A} \mu(\{x\}).$$

Inversement, si $(p_x)_{x \in E}$ est une famille d'éléments de $[0, 1]$ avec $\sum_{x \in E} p_x = 1$, alors il existe une unique mesure de probabilité μ sur E telle que

$$\mu(\{x\}) = p_x.$$

Les sommes qui apparaissent dans ce théorème ont bien un sens car il s'agit de sommes dénombrables de termes positifs.

Démonstration. L'expression de $\mu(A)$ découle du fait que $A = \bigcup_{x \in A} \{x\}$, où l'union est disjointe et dénombrable.

Montrons maintenant qu'une famille $(p_x)_{x \in E}$ vérifiant les conditions de l'énoncé définit bien une mesure de probabilité. On pose pour tout $A \subset E$

$$\mu(A) = \sum_{x \in A} p_x.$$

(note 1). Dans ce chapitre, "dénombrable" sera à comprendre comme "fini ou dénombrable".

On va montrer que A est bien une probabilité. Comme $0 \leq p_x$ et $\sum_{x \in E} p_x = 1$, on a bien $\mu(A) \in [0, 1]$. Ensuite, si $(A_n)_{n \in \mathbf{N}}$ est une suite d'ensembles disjoints on a

$$\mu\left(\bigcup_{n \in \mathbf{N}} A_n\right) = \sum_{x \in \bigcup_{n \in \mathbf{N}} A_n} p_x = \sum_{n \in \mathbf{N}} \sum_{x \in A_n} p_x = \sum_{n \in \mathbf{N}} \mu(A_n)$$

(le fait que les A_n sont disjoints est utilisé dans la deuxième égalité). \square

Pour une variable aléatoire à valeurs dans un ensemble dénombrable, l'espérance peut se mettre sous la forme d'une somme, d'après le résultat ci-dessous, qui précise le théorème 2.2.1.

Théorème 4.1.2 (Théorème de transfert, version discrète). *Soit X est une variable à valeurs dans un ensemble dénombrable $E = \{x_n, n \in \mathbf{N}\}$ et soit f une fonction de E dans \mathbf{R} .*

1. *Si f est bornée ou à valeurs positives, alors*

$$\mathbf{E}f(X) = \sum_{n \in \mathbf{N}} f(x_n) \mathbf{P}(X = x_n).$$

2. *Dans le cas général, $f(X)$ est intégrable si et seulement si*

$$\sum_{n \in \mathbf{N}} |f(x_n)| \mathbf{P}(X = x_n) < \infty.$$

Dans ce cas, on a alors

$$\mathbf{E}f(X) = \sum_{n \in \mathbf{N}} f(x_n) \mathbf{P}(X = x_n).$$

Démonstration. Commençons par le cas f bornée, en supposant pour simplifier $0 \leq f \leq 1$. Définissons $Y_N = \sum_{n=1}^N f(x_n) \mathbf{1}_{\{X=x_n\}}$ et $Z_N = Y_N + \sum_{n=N+1}^{\infty} \mathbf{1}_{\{X=x_n\}}$. On a $Y_N \leq f(X) \leq Z_N$, d'où $\mathbf{E}Y_N \leq \mathbf{E}f(X) \leq \mathbf{E}Z_N$. Or $\mathbf{E}(Z_N - Y_N) = \mathbf{P}(\exists n > N, X = x_n) \rightarrow 0$. Par conséquent,

$$\mathbf{E}f(X) = \lim_N \mathbf{E}Y_N = \lim_N \sum_{n=1}^N f(x_n) \mathbf{P}(X = x_n) = \sum_{n=1}^{\infty} f(x_n) \mathbf{P}(X = x_n).$$

On admet le cas général. \square

4.1.1 Indépendance

Si X et Y sont deux variables aléatoires à valeurs dans un ensemble dénombrable, alors le couple (X, Y) prend aussi ses valeurs dans un ensemble dénombrable. Dans ce cas, on peut alors caractériser simplement l'indépendance des variables X et Y .

Propriété 4.1.3. *Soit E un ensemble dénombrable et soient X et Y deux variables aléatoires à valeurs dans E . Alors, X et Y sont indépendantes si et seulement si pour tous éléments x et y de E on a*

$$\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x) \mathbf{P}(Y = y). \quad (4.1)$$

Démonstration. L'implication directe découle de la définition de l'indépendance.

Pour la réciproque, on suppose que X et Y vérifient (4.1). Soient A et B deux sous-ensembles de E . On a

$$\begin{aligned} \mathbf{P}(X \in A, Y \in B) &= \sum_{x \in A} \sum_{y \in B} \mathbf{P}(X = x, Y = y) = \sum_{x \in A} \sum_{y \in B} \mathbf{P}(X = x) \mathbf{P}(Y = y) \\ &= \left(\sum_{x \in A} \mathbf{P}(X = x) \right) \left(\sum_{y \in B} \mathbf{P}(Y = y) \right) \\ &= \mathbf{P}(X \in A) \mathbf{P}(Y \in B). \end{aligned}$$

Les variables X et Y sont donc bien indépendantes. \square

4.2 Lois discrètes usuelles

Dans cette partie on donne quelques propriétés des variables aléatoires discrètes les plus courantes.

4.2.1 Loi uniforme

Définition 4.2.1. Soit E un ensemble fini. On appelle loi uniforme sur E la mesure de probabilité définie par

$$\mu(A) = \frac{\text{card}A}{\text{card}E}.$$

Autrement dit, on a $\mu(\{x\}) = \frac{1}{\text{card}E}$, pour tout $x \in E$.

Dans le cas de la loi uniforme, un calcul de probabilité se ramène donc à un calcul de cardinal, ce qui sera le sujet de la partie 4.3.

Lorsqu'une variable aléatoire suit la loi uniforme sur un ensemble E , on dit aussi que l'on est en situation d'équiprobabilité. On notera aussi que dire qu'un élément de E est tiré "au hasard" pour signifier qu'il suit une loi uniforme sur E peut être ambigu.

Dans le cas de la loi uniforme sur $\{0, \dots, n\}$, on peut donner les résultats suivants :

Propriété 4.2.2. Soit X une variable aléatoire de loi uniforme sur $\{0, \dots, n\}$. Alors X est de carré intégrable et vérifie

$$\mathbf{E}X = \frac{n}{2} \text{ et } \mathbf{V}X = \frac{n(n+1)}{12}.$$

4.2.2 Loi de Bernoulli

Soit $0 \leq p \leq 1$ et une variable aléatoire $X : \Omega \rightarrow \{0, 1\}$ de loi de probabilité :

$$\mu_X(\{1\}) = p, \quad \mu_X(\{0\}) = 1 - p.$$

Alors la loi de X est appelée loi de Bernoulli de paramètre p .

Si on néglige les évènements de probabilité 0, les variables aléatoires de Bernoulli sont exactement les variables aléatoires de la forme $\mathbf{1}_A$. En effet, pour X une variable aléatoire, on peut partitionner Ω en

$$\Omega = \{X = 0\} \cup \{X = 1\}^c \cup \{X \notin \{0, 1\}\},$$

et si X suit une loi de Bernoulli, les variables X et $\mathbf{1}_{\{X=1\}}$ coïncident sur $\{X = 1\}$ et $\{X = 0\}$, et ne diffèrent que sur $\{X \notin \{0, 1\}\}$ qui est de probabilité nulle. En particulier, une variable de la forme $\mathbf{1}_A$ est une variable de Bernoulli de paramètre $\mathbf{P}(A)$.

Propriété 4.2.3. Une variable aléatoire X de loi de Bernoulli de paramètre p vérifie

$$\mathbf{E}X = p \text{ et } \mathbf{V}(X) = p(1 - p).$$

4.2.3 Loi binomiale

On veut étudier la répétition de plusieurs expériences identiques et indépendantes. Considérons donc n variables aléatoires X_1, \dots, X_n indépendantes de loi de Bernoulli de même paramètre p . On note S_n la somme de ces n variables : $S_n = \sum_{k=1}^n X_k$. Autrement dit, on considère une expérience qui réussit avec probabilité p , et on réalise cette expérience n fois, de manière indépendante. La variable S_n compte le nombre de succès parmi les n . Notamment, S_n prend ses valeurs dans $\{0, \dots, n\}$.

Propriété 4.2.4. On considère la variable S_n définie ci-dessus. Pour k dans $\{0, \dots, n\}$ on a

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

(la notation $\binom{n}{k}$ est définie en partie 4.3.3). La loi de S_n est appelée loi binomiale de paramètres n et p .

Démonstration. On a

$$\{S_n = k\} = \left\{ \text{card}\{i, X_i = 1\} = k \right\} = \bigcup_{\substack{I \subset \{1, \dots, n\} \\ \text{card} I = k}} \left\{ \{i, X_i = 1\} = I \right\},$$

où la dernière union est disjointe. Par conséquent, par indépendance des X_i

$$\begin{aligned} \mathbf{P}(S_n = k) &= \sum_{\substack{I \subset \{1, \dots, n\} \\ \text{card} I = k}} \mathbf{P}(\{i, X_i = 1\} = I) = \sum_{\substack{I \subset \{1, \dots, n\} \\ \text{card} I = k}} \prod_{i \in I} p \prod_{i \notin I} (1 - p) = \sum_{\substack{I \subset \{1, \dots, n\} \\ \text{card} I = k}} p^k (1 - p)^{n-k} \\ &= \binom{n}{k} p^k (1 - p)^{n-k}. \end{aligned}$$

□

Propriété 4.2.5. Une variable aléatoire X de loi binomiale de paramètres n et p vérifie :

$$\mathbf{E}X = np \text{ et } \mathbf{V}X = np(1 - p).$$

Démonstration. Par définition, X à la même loi que $\sum_{k=1}^n Y_k$, où les Y_k sont indépendantes et de loi de Bernoulli de paramètre p . On a donc $\mathbf{E}X = \sum_{k=1}^n \mathbf{E}Y_k = n\mathbf{E}Y_1 = np$. Par ailleurs, l'indépendance des Y_k donne $\mathbf{V}X = \sum_{k=1}^n \mathbf{V}Y_k = n\mathbf{V}Y_1 = np(1 - p)$. □

Propriété 4.2.6. Si X et Y sont deux variables aléatoires indépendantes de loi binomiale de paramètres (n, p) et (m, p) , alors $X + Y$ est suit la loi binomiale de paramètre $(n + m, p)$.

4.2.4 Loi de Poisson

Soit θ un réel strictement positif. Si X est une variable aléatoire de loi binomiale de moyenne θ , alors ses paramètres sont nécessairement n et $\frac{\theta}{n}$. Quand n tend vers l'infini, on est donc en train de regarder un grand nombre de fois un phénomène qui a une chance très faible de se produire. On peut par exemple penser au nombre de cas d'une maladie rare dans un pays : on a une grande population, dont chaque individu est touché par la maladie avec une probabilité faible. En conséquence, on observe seulement quelques cas sur tout le territoire (avec les notations ci-dessus, en moyenne θ cas).

La probabilité d'observer k cas est :

$$\binom{n}{k} \left(\frac{\theta}{n}\right)^k \left(1 - \frac{\theta}{n}\right)^{n-k} = \frac{n(n-1)\dots(n-k+1)\theta^k}{k!n^k} \exp\left((n-k)\ln\left(1 - \frac{\theta}{n}\right)\right). \quad (4.2)$$

Le quotient $\frac{n(n-1)\dots(n-k+1)}{n^k}$ tend vers 1 quand n tend vers ∞ . Par ailleurs, on a le développement asymptotique en $n \rightarrow \infty$:

$$\exp\left((n-k)\ln\left(1 - \frac{\theta}{n}\right)\right) = \exp\left(-\left(n-k\right)\left(\frac{\theta}{n} + o\left(\frac{1}{n}\right)\right)\right) = \exp(-\theta + o(1)).$$

Par conséquent, l'expression en (4.2) tend vers $\frac{\theta^k}{k!}e^{-\theta}$ quand $n \rightarrow \infty$. Par ailleurs, on remarque que $\sum_{k \in \mathbf{N}} \frac{\theta^k}{k!}e^{-\theta} = 1$, de sorte que la famille $(\frac{\theta^k}{k!}e^{-\theta})_{k \in \mathbf{N}}$ définit une loi de probabilité. On est donc amenés à la définition suivante.

Définition 4.2.7. Soit $\theta > 0$. On dit qu'une variable aléatoire X à valeurs dans \mathbf{N} suit la loi de Poisson ^(note 2) de paramètre θ si pour tout entier k , on a

$$\mathbf{P}(X = k) = e^{-\theta} \frac{\theta^k}{k!}.$$

Propriété 4.2.8. Si X et Y sont deux variables indépendantes de loi de Poisson de paramètres θ et η , alors $X + Y$ suit une loi de Poisson de paramètre $\theta + \eta$.

Propriété 4.2.9. Si X suit la loi de Poisson de paramètre θ , alors $\mathbf{E}X = \theta$ et $\mathbf{V}X = \theta$.

4.2.5 Loi géométrique

On répète indéfiniment une expérience ayant une probabilité p de succès. Au bout de combien de temps va-t-on observer le premier succès? Pour modéliser cela, on considère $(X_n)_{n \in \mathbf{N}}$ une suite infinie ^(note 3) de variables aléatoires indépendantes de loi de Bernoulli de paramètre p . On pose $T = \inf\{n \in \mathbf{N}, X_n = 1\}$.

Propriété 4.2.10. On a $\mathbf{P}(T = n) = p(1-p)^n$, pour tout $n \in \mathbf{N}$. La loi de T est appelée loi géométrique ^(note 4) de paramètre p .

(note 2). "Poisson" prend une majuscule, puisqu'il s'agit de Siméon Denis Poisson.

(note 3). On explique brièvement qu'une telle suite existe au début du chapitre 6.

(note 4). Plusieurs convention coexistent sur la définition de la loi géométrique. Une autre convention courante est d'appeler loi géométrique la loi de $T + 1$, portée par \mathbf{N}^* . De même, c'est parfois $1 - p$ qui est appelé le paramètre de la loi.

Démonstration. L'évènement $\{T = n\}$ peut se récrire

$$\{T = n\} = \{X_n = 1\} \cap \bigcap_{i=0}^{n-1} \{X_i = 0\}.$$

Sa probabilité est donc, par indépendance,

$$\mathbf{P}(T = n) = \mathbf{P}\left(\{X_n = 1\} \cap \bigcap_{i=0}^{n-1} \{X_i = 0\}\right) = \mathbf{P}(X_n = 1) \times \prod_{i=0}^{n-1} \mathbf{P}(X_i = 0) = p(1-p)^n.$$

□

Propriété 4.2.11. Une variable aléatoire T de loi géométrique de paramètre p est de carré intégrable et

$$\mathbf{E}T = \frac{1}{p} - 1, \quad \mathbf{V}(T) = \frac{1-p}{p^2}.$$

La loi géométrique vérifie la propriété suivante, dit d'absence de mémoire.

Propriété 4.2.12. Soient T une variable aléatoire de loi géométrique, et n et m deux entiers naturels. Alors

$$\mathbf{P}_{\{T \geq n\}}(T \geq n+m) = \mathbf{P}(T \geq m). \quad (4.3)$$

Inversement, si une variable aléatoire T à valeur dans \mathbf{N} vérifie la relation (4.3) pour tous entiers n et m , alors T suit une loi géométrique.

Démonstration. On peut commencer par remarquer que

$$\mathbf{P}(T \geq n) = \sum_{k=n}^{\infty} p(1-p)^k = p(1-p)^n \frac{1}{1-(1-p)} = (1-p)^n.$$

On a donc

$$\mathbf{P}_{\{T \geq n\}}(T \geq n+m) = \frac{\mathbf{P}(\{T \geq n\} \cap \{T \geq n+m\})}{\mathbf{P}(T \geq n)} = \frac{\mathbf{P}(\{T \geq n+m\})}{\mathbf{P}(T \geq n)} = \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m.$$

Pour la réciproque, on remarque que la relation (4.3) pour $m = 1$ se récrit $\mathbf{P}(T \geq n)\mathbf{P}(T \geq 1) = \mathbf{P}(T \geq n+1)$. Par conséquent on a $\mathbf{P}(T \geq n) = \mathbf{P}(T \geq 1)^n$. En notant $p = 1 - \mathbf{P}(T \geq 1)$, on trouve bien

$$\mathbf{P}(T = n) = \mathbf{P}(T \geq n) - \mathbf{P}(T \geq n+1) = (1-p)^n - (1-p)^{n+1} = p(1-p)^n.$$

□

Propriété 4.2.13. Si X et Y sont deux variables indépendantes de loi géométrique de paramètres respectifs p et q , alors la variable $\min(X, Y)$ suit la loi géométrique de paramètre $1 - (1-p)(1-q) = p + q - pq$.

Démonstration. On écrit

$$\mathbf{P}(\min(X, Y) \geq n) = \mathbf{P}(X \geq n, Y \geq n) = (1-p)^n(1-q)^n = (1-p-q+pq)^n.$$

Pour conclure, on remarque que

$$\mathbf{P}(\min(X, Y) = n) = \mathbf{P}(\min(X, Y) \geq n) - \mathbf{P}(\min(X, Y) \geq n+1).$$

□

4.3 Dénombrement

Le dénombrement consiste à calculer le cardinal d'un ensemble. La plupart du temps, cela revient à trouver une bijection entre l'ensemble dont on cherche le cardinal et un ensemble dont le cardinal est connu. En effet, deux ensembles en bijection ont même cardinal.

Nous présentons donc dans cette partie les cardinaux de divers ensembles "de référence".

4.3.1 Tirages ordonnés avec remise

Propriété 4.3.1. *L'ensemble des fonctions de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$ a n^k éléments.*

Démonstration. Par récurrence sur k . Pour $k = 1$, choisir une fonction de $\{1\}$ dans $\{1, \dots, n\}$ revient à choisir un élément de $\{1, \dots, n\}$; il y a donc n possibilités.

Supposons le résultat vrai pour un k fixé. Choisir une fonction de $\{1, \dots, k+1\}$ dans $\{1, \dots, n\}$ revient à choisir une fonction de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$ (soit n^k possibilités par hypothèse de récurrence), et à choisir l'image de $k+1$ (n possibilités). Il y a donc bien $n^k \times n = n^{k+1}$ telles fonctions. \square

On remarquera que les fonctions de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$ sont en bijection avec les k -uplets d'éléments de $\{1, \dots, n\}$.

4.3.2 Tirages ordonnés sans remise

Propriété 4.3.2. *L'ensemble des injections de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$ a pour cardinal $\frac{n!}{(n-k)!}$.*

Démonstration. Par récurrence sur k . Pour $k = 1$, une injection de $\{1\}$ dans $\{1, \dots, n\}$ peut s'assimiler à un élément de $\{1, \dots, n\}$, on a donc bien $n = \frac{n!}{(n-1)!}$ éléments.

On suppose le résultat vrai pour un k fixé avec $1 \leq k < n$. Se donner une injection de $\{1, \dots, k+1\}$ dans $\{1, \dots, n\}$ revient à se donner une injection de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$ et se donner l'image de $k+1$, qui doit être différente des k éléments déjà choisis. Il reste donc $n-k$ choix pour l'image de $k+1$, et on a donc $\frac{n!}{(n-k)!} \times (n-k) = \frac{n!}{(n-k-1)!}$ éléments. \square

Le cas particulier $k = n$ de la propriété 4.3.2 se réécrit comme suit.

Corollaire 4.3.3. *L'ensemble des bijections de $\{1, \dots, n\}$ dans lui-même a $n!$ éléments.*

On remarquera que l'ensemble des injections de $\{1, \dots, k\}$ dans $\{1, \dots, n\}$ est en bijection avec l'ensemble des k -uplets d'éléments tous distincts de $\{1, \dots, n\}$.

4.3.3 Tirages non ordonnés sans remise

Propriété 4.3.4. *Soient k et n deux entiers avec $0 \leq k \leq n$. L'ensemble des sous-ensembles à k éléments de $\{1, \dots, n\}$ a pour cardinal $\frac{n!}{k!(n-k)!}$. Cette valeur est appelée coefficient binomial et est noté $\binom{n}{k}$ ou C_n^k .*

Démonstration. On va montrer par récurrence sur n la propriété "Pour tout entier k avec $0 \leq k \leq n$, le nombre d'éléments est $\frac{n!}{k!(n-k)!}$ ".

Pour $n = 0$, on a un sous-ensemble à 0 éléments. Pour $n = 1$, on a un sous-ensemble à 0 éléments, et un sous-ensemble à un élément.

Supposons le résultat vrai pour un n fixé. Un sous-ensemble à k éléments de $\{1, \dots, n+1\}$ est :

- soit un sous-ensemble de $\{1, \dots, n\}$ à k éléments ;
- soit l'union de $\{n+1\}$ et d'un sous-ensemble de $\{1, \dots, n\}$ à $k-1$ éléments.

Le nombre d'éléments est donc

$$\begin{aligned} \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} &= \frac{n!}{(k-1)!(n-k)!} \left(\frac{1}{k} + \frac{1}{n-k+1} \right) \\ &= \frac{n!}{k!(n-k+1)!} \times (n+1) \\ &= \frac{(n+1)!}{k!(n-k+1)!}, \end{aligned}$$

ce qui achève la preuve. □

Dans la preuve de la proposition 4.3.4, on a en particulier montré la relation suivante :

Propriété 4.3.5. *Les coefficients binomiaux vérifient la relation*

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

Cette relation permet de calculer simplement les coefficients binomiaux, en remplissant ligne par ligne le tableau ci-dessous, contenant les valeurs des $\binom{n}{k}$:

$n \backslash k$	0	1	2	3	4	5	...
0	1	0	0	0	0	0	
1	1	1	0	0	0	0	
2	1	2	1	0	0	0	
3	1	3	3	1	0	0	
4	1	4	6	4	1	0	
5	1	5	10	10	5	1	
⋮							

Chaque coefficient est la somme du coefficient situé au-dessus de lui et de celui à la gauche de ce dernier (parmi les cases encadrées, 10 s'obtient comme $6+4$).

Partitionnement

Propriété 4.3.6. *Soit n un entier positif, et r_1, \dots, r_k des entiers tels que $r_1 + \dots + r_k = n$. L'ensemble des partitions de $\{1, \dots, n\}$ en k ensembles de cardinaux respectifs r_1, \dots, r_k , a pour cardinal*

$$\frac{n!}{r_1! \cdots r_k!}.$$

On remarque que cas particulier $k=2$ correspond à la propriété 4.3.4. En effet, choisir un sous-ensemble de $\{1, \dots, n\}$ à k éléments revient à partitionner $\{1, \dots, n\}$ en un sous-ensemble à k élément et un autre sous-ensemble à $n-k$ éléments (le complémentaire de l'ensemble initial).

Démonstration. On raisonne par récurrence sur k . Pour $k = 1$, on a alors $r_1 = n$ et il n'y a qu'un seul choix possible, or $1 = \frac{n!}{r_1!}$.

Supposons le résultat vrai pour un certain $k \geq 1$. On se donne $k + 1$ entiers vérifiant

$$r_1 + \dots + r_k + r_{k+1} = 1.$$

Partitionner $\{1, \dots, n\}$ en $k + 1$ ensembles de cardinaux r_1, \dots, r_{k+1} revient à partitionner $\{1, \dots, n\}$ en k ensembles de cardinaux $r_1, \dots, r_{k-1}, (r_k + r_{k+1})$, puis à partitionner le dernier ensemble (à $r_k + r_{k+1}$ éléments) en un ensemble à r_k éléments et un à r_{k+1} éléments. Le nombre d'éléments est donc

$$\frac{n!}{r_1! \cdots r_{k-1}! (r_k + r_{k+1})!} \times \frac{(r_k + r_{k+1})!}{r_k! r_{k+1}!} = \frac{n!}{r_1! \cdots r_{k-1}! r_k! r_{k+1}!}.$$

□

4.3.4 Tirages non ordonnés avec remise

Propriété 4.3.7. *L'ensemble des n -uplets d'entiers (positifs ou nuls) dont la somme vaut k a $\binom{k+n-1}{k}$ éléments.*

Démonstration. Fixons n et k . On va représenter l'entier k par une suite de k étoiles. Par exemple, pour $k = 11$

Pour représenter les entiers dont la somme fera k , on va séparer cet ensemble d'étoiles à l'aide de barres verticales :

*|***|**|****|*|

Ici, on a représenté la décomposition $1 + 3 + 0 + 2 + 4 + 1 + 0 = 11$. Dans cet exemple, on a tracé 6 barres verticales, pour obtenir une décomposition en somme de $n = 7$ entiers. Chaque disposition de k étoiles et $n - 1$ barres fournit une décomposition de k en somme de n entiers. On a donc $k + n - 1$ objets à placer, k étoiles et $n - 1$ barres. La propriété 4.3.4 nous dit qu'il y a $\frac{(k+n-1)!}{k!(n-1)!} = \binom{k+n-1}{k}$ tracés possibles. □

La situation de la propriété 4.3.7 correspond à un tirage non ordonné avec remise de k éléments parmi $\{1, \dots, n\}$. Autrement dit, on peut tirer un même élément plusieurs fois, et on ne s'intéresse qu'au nombre de fois où chaque élément a été tiré, et pas à l'ordre dans lequel ils ont été tirés.

En effet, le k -uplet (r_1, \dots, r_k) correspond au tirage non ordonné où l'on a tiré r_i fois le nombre i .

4.4 Exercices

Exercice 4.1.

Quelle est la probabilité que dans un groupe de 23 personnes, au moins deux d'entre elles aient leur anniversaire le même jour ?

Exercice 4.2.

Dans une course, n chevaux sont au départ. On suppose que tous les ordres d'arrivée à la course sont équiprobables. Calculer la probabilité de gagner le tiercé avec un ticket :

1. dans l'ordre ;
2. dans l'ordre ou dans un ordre différent ;
3. dans un ordre différent.

Exercice 4.3.

Un joueur de poker reçoit une main de 5 cartes d'un jeu de 32 cartes. Quelle est la probabilité qu'il reçoive :

1. une seule paire (deux cartes de même hauteur) ;
2. deux paires ;
3. un brelan (trois cartes de même hauteur et pas de paire ni de carré) ;
4. un carré (quatre cartes de même hauteur) ;
5. un full (une paire et un brelan) ?

Exercice 4.4.

Au loto, le joueur doit choisir 6 nombres dans $\{1, \dots, 49\}$. Un tirage consiste à extraire, sans remise, 6 boules numérotées d'une urne, dont les numéros sont dits gagnants, et une septième boule fournissant le numéro dit complémentaire.

1. Quelle est la probabilité de tirer les 6 numéros gagnants ?
2. Quelle est la probabilité de tirer 5 numéros gagnants et le numéro complémentaire ?

Exercice 4.5.

On place r boules dans n urnes, chaque boule ayant la même probabilité d'être placée dans chaque urne, et les boules étant placées indépendamment les unes des autres. Quelle est la probabilité qu'une urne donnée contienne exactement k boules ?

Exercice 4.6.

Soit $f : \mathbf{R}^n \rightarrow \mathbf{R}$ une fonction de n variables de classe \mathcal{C}^∞ . Quel est le nombre de dérivées partielles distinctes d'ordre r ?

Exercice 4.7.

Combien l'équation $x_1 + x_2 + x_3 = 15$ a-t-elle de solutions entières positives ou nulles ? de solutions entières strictement positives ?

Exercice 4.8.

On considère les chemins tracés sur une grille de largeur 7 et de hauteur 8 (voir figure) tels que le chemin aille toujours vers la droite ou vers le haut.

1. Combien de trajets différents peut-on emprunter entre le point A et le point B ?
2. Combien y a-t-il d'octuplets (n_1, \dots, n_8) d'entiers naturels avec $n_1 + \dots + n_8 = 8$?
3. Quel est le rapport entre les deux premières questions ?

Exercice 4.9.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et de même loi de Bernoulli de paramètre $p \in]0, 1[$. On note $T = \sup\{n \in \mathbf{N}, X_1 = X_2 = \dots = X_n\}$ et $S = \sup\{n \in \mathbf{N}, X_{T+1} = X_{T+2} = \dots = X_{T+n}\}$. Quelles sont les lois de T et S ? S et T sont-elles indépendantes ?

1. Montrer que S et T sont deux variables aléatoires indépendantes de loi géométrique de paramètre p .
2. Calculer $P(T = k, S = q | T < n < T + S)$. Les variables T et $T + S$ sont-elles indépendantes sachant $\{T < n < T + S\}$?
3. Quelle est la loi de T sachant $\{T + S = n\}$?

Exercice 4.14.

Soient X et Y deux variables aléatoires indépendantes, à valeurs dans \mathbf{N} .

1. Calculer la loi de $Z = X + Y$.
2. Calculer la loi de $T = \min(X, Y)$.

Exercice 4.15.

Soit N une variable de loi binomiale de paramètres n et q , et soit $(X_k)_{k \in \mathbf{N}}$ une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre p , et indépendantes de N .

1. (a) Quelle est la loi de $Z = \sum_{k=1}^N X_k$? Et celle de $N - Z$?
 (b) Les variables Z et $N - Z$ sont-elles indépendantes ?
2. Mêmes questions si N suit la loi de Poisson de paramètre θ .

Exercice 4.16.

Soient $(X_k)_{k=1, \dots, n}$ des variables aléatoires indépendantes de loi uniforme sur $\{1, \dots, N\}$. Quelle est la loi de $\max_{k=1}^n X_k$?

Exercice 4.17.

On lance plusieurs fois de suite un dé jusqu'à être satisfait du résultat obtenu. Le nombre total de lancers est limité à trois. Quelle est la meilleure stratégie à adopter si l'on veut maximiser l'espérance de la valeur obtenue ? Que vaut alors cette espérance ?

On pourra d'abord traiter le cas où l'on a droit à un seul lancer, que l'on peut soit conserver soit échanger contre une valeur y fixée.

Chapitre 5

Variables aléatoires continues

5.1 Variables aléatoires réelles à densité

5.1.1 Définitions

On rappelle que l'ensemble des mesures de probabilité sur \mathbf{R} est en bijection avec l'ensemble des fonctions de répartition. Comme l'ensemble des fonctions de répartition est essentiellement l'ensemble des fonctions croissantes, les fonctions de répartition qui sont dérivables ont alors une dérivée positive. Plus précisément on peut énoncer la propriété suivante :

Propriété 5.1.1. *Soit F une fonction de répartition. Si il existe une fonction $p : \mathbf{R} \rightarrow \mathbf{R}$ telle que pour tout réel x on ait*

$$F(x) = \int_{-\infty}^x p(t)dt, \quad (5.1)$$

alors la fonction p est positive et vérifie $\int_{\mathbf{R}} p(t)dt = 1$.

Inversement, si $p : \mathbf{R} \rightarrow \mathbf{R}$ est une fonction positive d'intégrale 1, alors, la fonction F définie par la relation (5.1) est une fonction de répartition.

On a donc caractérisé un sous-ensemble de l'ensemble des mesures sur \mathbf{R} , que l'on appellera l'ensemble des mesures à densité :

Définition 5.1.2. *Une variable aléatoire X est dite à densité, s'il existe une fonction p_X telle que la loi de X vérifie :*

$$\mathbf{P}(X \leq x) = \int_{-\infty}^x p_X(t)dt.$$

La fonction p_X est alors appelée densité de X .

Propriété 5.1.3. *Pour une variable aléatoire X à densité, on a les résultats suivants :*

1. *La fonction de répartition F_X est continue sur \mathbf{R} . En particulier $\mathbf{P}(X = x) = 0$, pour tout réel x ;*
2. *Pour tout intervalle I de \mathbf{R} , on a*

$$\mathbf{P}(X \in I) = \int_I p_X(x)dx ;$$

3. La fonction de répartition F_X est dérivable partout où la densité p_X est continue et, en ces points, $F'_X = p_X$.

En pratique, la relation à retenir est $F'_X = p_X$. On a notamment le résultat suivant :

Propriété 5.1.4. *Soit X une variable aléatoire de fonction de répartition F_X . Si l'application F_X est continue sur \mathbf{R} et \mathcal{C}^1 par morceaux, alors X est une variable aléatoire à densité, de densité égale à F'_X partout où elle est dérivable.*

La densité d'une variable aléatoire à densité peut également se caractériser à partir de calculs d'espérance. En effet, une fonction f en escaliers peut s'écrire

$$f(x) = \sum_{k=1}^n \alpha_k \mathbf{1}_{]a_{k-1}, a_k[},$$

avec $-\infty = a_0 < a_1 < a_2 \dots < a_{n-1} < a_n = \infty$. On a alors

$$\mathbf{E}f(X) = \sum_{k=1}^n f(\alpha_k) \mathbf{P}(a_{k-1} < X < a_k) = \sum_{k=1}^n f(\alpha_k) \int_{a_{k-1}}^{a_k} p(x) dx = \int_{\mathbf{R}} f(x) p(x) dx.$$

On admettra que ce résultat s'étend à un f quelconque. On obtient alors le résultat suivant, qui est analogue aux théorèmes 2.2.1 et 4.1.2.

Théorème 5.1.5 (Théorème de transfert, version continue). *Si f est une fonction positive ou bornée, ou si $f(X)$ est intégrable, alors*

$$\mathbf{E}f(X) = \int_{\mathbf{R}} f(x) p(x) dx.$$

Démonstration. Admis. □

On a même une réciproque au théorème 5.1.5.

Propriété 5.1.6. *Soit X une variable aléatoire telle qu'il existe une fonction $p : \mathbf{R} \rightarrow \mathbf{R}$ telle que pour toute fonction f positive, on ait*

$$\mathbf{E}f(X) = \int_{\mathbf{R}} f(x) p(x) dx.$$

Alors X admet p pour densité.

Démonstration. En prenant $f = \mathbf{1}_{]-\infty, t]}$, on trouve

$$\mathbf{P}(X \leq t) = \mathbf{E}f(X) = \int_{-\infty}^t p(x) dx,$$

ce qui correspond à la définition d'une variable à densité. □

5.2 Lois à densité usuelles

Dans cette partie on présente les lois à densité les plus usuelles, avec certaines de leurs propriétés.

5.2.1 Loi uniforme

Définition 5.2.1. Soit $a < b$ deux réels. On dit qu'une variable aléatoire suit la loi uniforme sur l'intervalle $[a, b]$ si elle admet pour densité

$$\frac{1}{b-a} \mathbf{1}_{[a,b]}.$$

Le nom de loi *uniforme* vient du fait que la probabilité que X appartienne à un sous-intervalle I de $[a, b]$ est proportionnelle à la longueur de I , mais ne dépend pas de la position de I dans $[a, b]$.

Propriété 5.2.2. Une variable de loi uniforme sur $[a, b]$ est bornée et vérifie

$$\mathbf{E}X = \frac{a+b}{2} \text{ et } \mathbf{V}X = \frac{(b-a)^2}{12}.$$

5.2.2 Loi de Gauss ou loi normale

Définition 5.2.3. Une variable aléatoire réelle X suit une loi Gaussienne ou loi normale centrée réduite si elle admet pour densité l'application p_X définie sur \mathbf{R} par :

$$\forall x \in \mathbf{R}, \quad p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Plus généralement, on définit la loi normale de moyenne m et de variance $\sigma^2 > 0$, notée $\mathcal{N}(m, \sigma^2)$, comme la loi dont la densité est donnée par

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

On verra au chapitre 6 le théorème limite central 6.3.1, qui fait apparaître très naturellement la loi normale.

Propriété 5.2.4. Une variable aléatoire de loi normale de paramètres μ et σ^2 est de carré intégrable et on a

$$\mathbf{E}X = \mu \text{ et } \mathbf{V}X = \sigma^2.$$

Propriété 5.2.5. Soient X et Y deux variables aléatoires indépendantes de loi normales de paramètres respectifs (m_X, σ_X^2) et (m_Y, σ_Y^2) , et soient a et b deux réels. Alors

- la variable aléatoire $aX + b$ est de loi normale de moyenne $am_X + b$ et de variance $a^2\sigma_X^2$.
- la variable aléatoire $X + Y$ est de loi normale de moyenne $m_X + m_Y$ et de variance $\sigma_X^2 + \sigma_Y^2$.

Propriété 5.2.6. Soit X une variable aléatoire de loi normale centrée réduite et n un entier positif. Alors :

$$\mathbf{E}X^n = \begin{cases} 0 & \text{si } n \text{ est impair,} \\ (n-1) \times (n-3) \dots \times 1 = \frac{(2k)!}{2^k k!} & \text{si } n = 2k \text{ est pair.} \end{cases}$$

Démonstration. Le cas n impair vient du fait que $\mathbf{E}X^n$ est donné par l'intégrale sur \mathbf{R} d'une fonction impaire. Pour le cas n pair, on procède par récurrence, en intégrant par parties (on dérive x^{n+1} et on primitive $x e^{-x^2/2}$) :

$$\sqrt{2\pi} \mathbf{E}(X^{n+2}) = \int_{\mathbf{R}} x^{n+2} e^{-x^2/2} dx = \left[-x^{n+1} e^{-x^2/2} \right]_{-\infty}^{\infty} + \int_{\mathbf{R}} (n+1)x^n e^{-x^2/2} dx = (n+1) \mathbf{E}(X^n).$$

□

5.2.3 Loi exponentielle

Définition 5.2.7. La variable aléatoire X suit une loi exponentielle de paramètre $\lambda > 0$, si elle admet pour densité l'application p_X définie par :

$$\forall x \in \mathbf{R}, p_X(x) = \mathbf{1}_{[0, +\infty[}(x) \lambda e^{-\lambda x}.$$

En vertu de la propriété suivante, on dit que la loi exponentielle est une loi “sans mémoire”.

Propriété 5.2.8. Soit X une variable aléatoire de loi exponentielle (de paramètre quelconque). Alors, pour tout $t_0 > 0$ et $t > 0$, on a

$$\mathbf{P}_{\{X > t_0\}}(X > t_0 + t) = \mathbf{P}(X > t). \quad (5.2)$$

Inversement, toute variable aléatoire X à valeurs dans \mathbf{R}^+ vérifiant la relation (5.2) suit une loi exponentielle.

Démonstration. On remarque tout d'abord que

$$\mathbf{P}(X > t) = \int_t^\infty \lambda e^{-\lambda s} ds = [-e^{-\lambda s}]_t^\infty = e^{-\lambda t}.$$

On a donc

$$\mathbf{P}(X > t_0 + t | X > t_0) = \frac{\mathbf{P}(X > t_0 + t)}{\mathbf{P}(X > t_0)} = \frac{e^{-\lambda(t_0+t)}}{e^{-\lambda t_0}} = e^{-\lambda t},$$

d'où le résultat.

Pour la réciproque, on remarque que l'équation (5.2) se réécrit $\mathbf{P}(X > t)\mathbf{P}(X > s) = \mathbf{P}(X > t + s)$ pour tous $s, t > 0$. La fonction $t \mapsto \mathbf{P}(X > t)$ est donc une fonction positive décroissante vérifiant $f(t)f(s) = f(t + s)$. Les seules fonctions vérifiant ces conditions sont les fonctions de la forme $f(t) = e^{-\lambda t}$. \square

On remarquera que la propriété 5.2.8 ainsi que sa preuve sont presque des recopies de la propriété 4.2.12 qui concernait la loi géométrique. En fait, la loi exponentielle est un équivalent continu de la loi géométrique, et ces deux lois partagent de nombreuses propriétés.

Propriété 5.2.9. Si X et Y sont deux variables aléatoires de loi exponentielle de paramètre respectifs λ et μ , alors $\min(X, Y)$ suit la loi exponentielle de paramètre $\lambda + \mu$.

Démonstration. On calcule, pour $t \geq 0$:

$$\mathbf{P}(\min(X, Y) > t) = \mathbf{P}(X > t, Y > t) = \mathbf{P}(X > t)\mathbf{P}(Y > t) = e^{-\lambda t}e^{-\mu t} = e^{-(\lambda+\mu)t}.$$

La fonction de répartition est alors $F(t) = 1 - e^{-(\lambda+\mu)t}$ (pour $t \geq 0$, et nulle pour $t < 0$). La propriété 5.1.4 montre alors le résultat. \square

En raisonnant par récurrence on montrerait que les X_i , $1 \leq i \leq n$ sont des variables aléatoires indépendantes de loi exponentielle de paramètres $\lambda_1, \dots, \lambda_n$ respectivement, alors $\min(X_1, \dots, X_n)$ suit la loi exponentielle de paramètre $\lambda_1 + \dots + \lambda_n$.

Propriété 5.2.10. Une variable aléatoire de loi exponentielle de paramètre λ est de carré intégrable et on a

$$\mathbf{E}X = \frac{1}{\lambda} \text{ et } \mathbf{V}X = \frac{1}{\lambda^2}.$$

Propriété 5.2.11. Si U est une variable aléatoire de loi uniforme sur $[0, 1]$, alors la variable $-\frac{1}{\lambda} \ln U$ suit la loi exponentielle de paramètre λ .

Propriété 5.2.12. Soit X une variable aléatoire de loi exponentielle de paramètre λ et n un entier naturel. Alors

$$\mathbf{E}(X^n) = \frac{n!}{\lambda^n}.$$

5.2.4 Loi de Cauchy

Définition 5.2.13. La variable aléatoire X suit une loi de Cauchy standard si elle admet pour densité l'application p_X définie sur \mathbf{R} par

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

On parle également de loi de Cauchy de paramètre c si la densité est donnée par

$$p_X(x) = \frac{c}{\pi(c^2+x^2)}.$$

Le paramètre de la loi de Cauchy correspond à un changement d'échelle :

Propriété 5.2.14. Si X est une variable aléatoire de loi de Cauchy de paramètre c , alors λX suit la loi de Cauchy de paramètre λc .

Si X et Y sont deux variables aléatoires de loi de Cauchy de paramètre respectifs c et c' . Alors $X+Y$ suit une loi de Cauchy de paramètre $c+c'$.

La loi de Cauchy n'est pas particulièrement utile, mais elle a la particularité suivante :

Propriété 5.2.15. Une variable aléatoire de loi de Cauchy n'est pas intégrable.

Démonstration. Si X suit la loi de Cauchy (de paramètre 1), alors $\mathbf{E}|x| = \int_{\mathbf{R}} \frac{|x|}{1+x^2} dx$. Or, $\frac{|x|}{1+x^2}$ est équivalent à ∞ quand x tend vers $\pm\infty$. Par conséquent cette fonction n'est pas intégrable et $\mathbf{E}|X| = \infty$. \square

Une manière "naturelle" de faire apparaître la loi de Cauchy est la suivante :

Propriété 5.2.16. Soit Θ une variable aléatoire de loi uniforme sur $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Alors $\tan(\Theta)$ suit la loi de Cauchy.

La propriété précédente s'interprète géométriquement de la manière suivante : sur la figure 5.1, le point C est situé à une distance 1 de la droite \mathcal{D} . Le point O est le projeté orthogonal de C sur \mathcal{D} . On tire un angle Θ uniformément sur $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Il existe un unique point X de \mathcal{D} tel que l'angle $(\overrightarrow{CO}, \overrightarrow{CX})$ soit égal à Θ . La distance algébrique $O\overline{X}$ suit alors une loi de Cauchy.

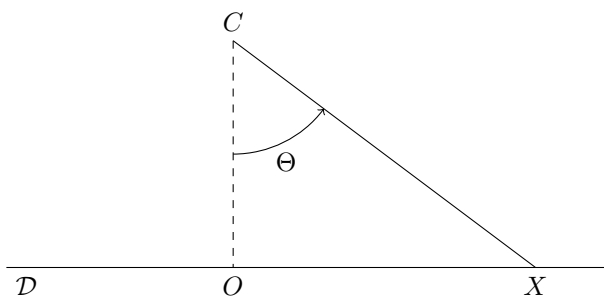


FIGURE 5.1 – Interprétation géométrique de la loi de Cauchy.

5.3 Exercices

Exercice 5.1.

Soit X une variable aléatoire réelle de densité p . Déterminer en fonction de p la densité de la variable aléatoire Y dans les cas suivants :

1. $Y = aX + b$, où a et b sont des nombres réels, $a \neq 0$;
2. $Y = X^2$;
3. $Y = \exp(X)$.

Cet exercice peut se résoudre avec le théorème de transfert. Dans le cas où la densité p est supposée continue, on peut également utiliser la fonction de répartition.

Exercice 5.2.

Soit X une variable aléatoire de loi de Cauchy standard. Déterminer la fonction de répartition des variables X et $1/X$. Quelle est la loi de $1/X$?

Exercice 5.3.

Soit U une variable aléatoire uniforme sur $[0, 1]$, et soit $\alpha \neq 0$ un réel. Montrer que la loi de U^α admet une densité que l'on explicitera. Quand elles existent, donner la valeur de l'espérance et de la variance de U^α .

Exercice 5.4.

Soit Θ une variable uniforme sur $[-\pi/2, \pi/2]$.

- Quelle est la loi de $\sin(\Theta)$?
- Quelle est la moyenne de $\sin(\Theta)$? celle de $|\sin(\Theta)|$?
- Calculer la variance de $\sin(\Theta)$.

Exercice 5.5.

Soient X_1 et X_2 deux variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Montrer que les variables aléatoires $\min(X_1, X_2)$ et $\max(X_1, X_2)$ admettent des densités que l'on explicitera.

Plus généralement, soient X_1, \dots, X_n des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Il existe une permutation (aléatoire) σ de $\{1, \dots, n\}$ qui réordonne (X_1, \dots, X_n) , c'est-à-dire que si on pose $Y_i = X_{\sigma(i)}$, alors

$$Y_1 \leq Y_2 \leq \dots \leq Y_n.$$

Quelle est la loi de Y_k , pour $k \in \{1, \dots, n\}$? son espérance? On pourra remarquer la relation $\mathbf{P}(Y_k \leq x) = \sum_{q=k}^n \mathbf{P}(Y_q \leq x, Y_{q+1} > x)$.

Exercice 5.6.

Soient n variables aléatoires indépendantes (X_1, \dots, X_n) dont la loi commune admet une densité continue p . Exprimer en fonction de p la densité de $\max_{i=1}^n X_i$.

Exercice 5.7.

On considère un parquet dont les lattes ont une largeur d . On dispose d'une aiguille de longueur l , avec $l < d$. Montrer que si on lance l'aiguille sur le parquet, elle tombera en travers d'une rainure du parquet avec probabilité $2l/\pi d$.

Cette expérience a été mise en pratique par Mario Lazzarini avec 3408 aiguilles, parmi lesquelles 1808 ont effectivement croisé une rainure. Dans son expérience, les longueurs vérifiaient la relation $l/d = 5/6$. Qu'en pensez-vous?

Exercice 5.8.

On choisit au hasard une corde sur un cercle. Quelle est la probabilité que l'angle au centre interceptant cette corde ait une mesure supérieure à $2\pi/3$?

Exercice 5.9.

Soient (X, Y, Z) un point aléatoire choisi uniformément sur la sphère unité $\{(x, y, z), x^2 + y^2 + z^2 = 1\}$. Quelle est la loi de X ?

Exercice 5.10.

Soient X et Y deux variables Gaussiennes centrées réduites indépendantes. Déterminer la fonction de répartition de la variable aléatoire $Z = \arctan\left(\frac{Y}{|X|}\right)$. Quelle est la loi de Z ?

Exercice 5.11.

Soient $(T_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes de loi exponentielle de paramètre λ . On pose $N_t = \sup\{k \geq 0, T_1 + T_2 + \dots + T_k < t\}$.

1. Montrer que la variable $T_1 + T_2 + \dots + T_k$ admet pour densité $\frac{\lambda^k}{(k-1)!} t^{k-1} e^{-\lambda t}$.
2. Montrer que N_t suit une loi de Poisson de paramètre λt .
3. Soient $s \leq t$ deux réels. Calculer $\mathbf{P}(T_1 \leq s | N_t = 1)$. Quelle est la loi de T_1 conditionnellement à $\{N_t = 1\}$?

Chapitre 6

Suites de variables aléatoires - théorèmes limites

6.1 Suites de variables aléatoires

On voudrait étudier la situation où un même phénomène aléatoire se répète indéfiniment, et de manière indépendante à chaque fois. Pour cela, il faut construire une suite (infinie) de variables de même loi indépendantes les unes des autres. Il n'est pas évident a priori qu'une telle suite de variables puisse exister. D'après, le théorème suivant, c'est effectivement le cas.

Théorème 6.1.1. *Soit $(\mu_n)_{n \in \mathbf{N}}$ une suite de mesures de probabilité sur un ensemble E . Alors il existe un espace probabilisé (Ω, \mathbf{P}) et une suite de variables aléatoires $(X_n)_{n \in \mathbf{N}}$ à valeurs dans E telles que les X_n soient indépendantes et que pour tout n , X_n ait pour loi μ_n .*

Démonstration. Admis. Il s'agirait de construire une mesure sur l'ensemble indénombrable $E^{\mathbf{N}}$ (ensemble des suites à valeurs dans E). Comme il a déjà été dit, construire une mesure sur un ensemble indénombrable peut être très complexe, et on ne le fera pas ici. \square

6.1.1 Convergence de suites de variables aléatoires réelles

Dans cette partie, on considère une suite de variables aléatoires réelles $(X_n)_{n \in \mathbf{N}}$, et on veut donner un sens à la phrase " $(X_n)_{n \in \mathbf{N}}$ converge". Il existe en fait plusieurs sens possibles à donner à cette affirmation.

La définition la plus simple a priori revient à dire que toute réalisation de cette suite est une suite convergente. C'est par exemple ce que l'on observait sur la figure 0.2.

Définition 6.1.2. *On dit que la suite $(X_n)_{n \in \mathbf{N}}$ converge presque sûrement vers une variable aléatoire X si*

$$\mathbf{P}(X_n \text{ converge vers } X) = 1.$$

Ici, $\{X_n \text{ converge vers } X\}$ désigne un évènement à savoir $\{\omega \in \Omega, X_n(\omega) \text{ converge vers } X(\omega)\}$.

Toutefois, cette notion est souvent difficile à prouver, et on va plutôt utiliser des notions plus faibles. Nous présentons donc une seconde notion, qui revient à dire que, pour n grand, X_n a de grandes chances d'être proche de X .

Définition 6.1.3. On dit que $(X_n)_{n \geq 1}$ converge en probabilité vers X si pour tout $\varepsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \mathbf{P}(|X - X_n| < \varepsilon) = 1.$$

Le résultat suivant relie les deux notions précédentes.

Propriété 6.1.4. Si $(X_n)_{n \in \mathbf{N}}$ converge presque sûrement vers X , alors elle converge aussi en probabilité vers X .

Démonstration. L'évènement $\{X_n \text{ converge vers } X\}$ peut se récrire

$$\{X_n \text{ converge vers } X\} = \{\forall \varepsilon > 0, \exists N \in \mathbf{N}, \forall n \geq N, |X_n - X| < \varepsilon\} = \bigcap_{\varepsilon > 0} \bigcup_{N \in \mathbf{N}} \bigcap_{n \geq N} \{|X_n - X| < \varepsilon\}.$$

Comme cet évènement est de probabilité 1, pour tout $\varepsilon > 0$, l'évènement plus grand

$$\bigcup_{N \in \mathbf{N}} \bigcap_{n \geq N} \{|X_n - X| < \varepsilon\}$$

est aussi de probabilité 1. Or, comme il s'agit de l'union d'une suite croissante d'évènements, on a

$$1 = \mathbf{P} \left(\bigcup_{N \in \mathbf{N}} \bigcap_{n \geq N} \{|X_n - X| < \varepsilon\} \right) = \lim_N \mathbf{P} \left(\bigcap_{n \geq N} \{|X_n - X| < \varepsilon\} \right).$$

Or $\mathbf{P} \left(\bigcap_{n \geq N} \{|X_n - X| < \varepsilon\} \right) \leq \mathbf{P}(|X_N - X| < \varepsilon)$, ce qui montre que $\mathbf{P}(|X_n - X| < \varepsilon)$ tend vers 1 pour tout ε . \square

La réciproque de la propriété 6.1.4 n'est pas vraie, comme le montre le contre-exemple que nous allons construire maintenant. On considère une suite de variables aléatoires indépendantes $(X_n)_{n \in \mathbf{N}}$ telle que X_n suive la loi uniforme sur $\{2^n, \dots, 2^{n+1} - 1\}$. On définit alors, pour tout entier k ,

$$Y_k = \begin{cases} 1 & \text{si } X_n = k \text{ pour un certain } n, \\ 0 & \text{sinon.} \end{cases}$$

Expliquons un peu cette construction : le choix des $(X_n)_{n \in \mathbf{N}}$ revient à choisir un et un seul élément uniformément dans chacun des ensembles $\{1\}$, $\{2, 3\}$, $\{4, 5, 6, 7\}$, $\{8, \dots, 15\}$, ..., $\{2^n, \dots, 2^{n+1} - 1\}$, etc, qui constituent une partition de \mathbf{N} .

k	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
X_n	×		×		×					×						...
Y_k	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	

Dans le tableau ci-dessus, les croix symbolisent un tirage possible pour les X_n . Ici, $X_0 = 1$, $X_1 = 3$, $X_2 = 5$, $X_3 = 10$, etc. La suite $(Y_k)_{k \in \mathbf{N}}$ est la suite de 0 et de 1 en dernière ligne.

On remarque alors que :

- La suite $(Y_k)_{k \in \mathbf{N}}$ ne converge pas presque sûrement vers 0. En effet, pour tout n , il existe toujours une valeur k de $\{2^n, \dots, 2^{n+1}\}$ telle que $Y_k = 1$.
- La suite $(Y_k)_{k \in \mathbf{N}}$ converge en probabilité vers 0. En effet, pour tout $\varepsilon > 0$, on a $\mathbf{P}(|Y_k - 0| < \varepsilon) \geq \mathbf{P}(Y_k = 0)$. Or, si n_k est l'entier tel que $2^{n_k} \leq k < 2^{n_k+1}$, alors

$$\mathbf{P}(Y_k = 0) = 1 - 2^{-n_k} \xrightarrow{k \rightarrow \infty} 1.$$

Par conséquent, $\mathbf{P}(|Y_k - 0| < \varepsilon)$ tend vers 1.

6.2 Loi des grands nombres

Le théorème suivant est la justification théorique des observations faites au chapitre 0 sur la figure 0.2. C'est aussi une justification de l'utilité des notions d'espérance et de probabilité : jusqu'ici, ces deux notions étaient posées comme axiomes, sans justification a priori. La loi des grands nombres les fait apparaître comme des quantité calculables.

Théorème 6.2.1 (Loi (forte) des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de même loi et intégrables. Alors, la suite $\frac{1}{n} \sum_{k=1}^n X_k$ converge presque sûrement vers $\mathbf{E}X_1$.*

Remarquons que l'hypothèse “ X_n intégrable” est tout ce dont on a besoin pour donner un sens à $\mathbf{E}X_1$.

Ce théorème constitue la “vraie” loi des grands nombres. Toutefois, sa preuve est trop complexe pour pouvoir être exposée ici en toute généralité. Par conséquent, nous ne démontrons que la version plus faible suivante.

Théorème 6.2.2 (Loi (faible) des grands nombres). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, de même loi et de carrés intégrables. La suite $\frac{1}{n} \sum_{k=1}^n X_k$ converge en probabilité vers $\mathbf{E}X_1$.*

Remarquons que le théorème 6.2.2 est plus faible que le théorème 6.2.1 à deux titres :

- ses hypothèses sont plus fortes : toute variable de carré intégrable est intégrable ;
- sa conclusion est plus faible : la convergence presque sûre entraîne la convergence en probabilité.

Par conséquent, le théorème 6.2.2 n'a en fait que peu d'intérêt par rapport au théorème 6.2.1, si ce n'est d'avoir une démonstration beaucoup plus simple.

Démonstration. On remarque que la moyenne de la variable $M_n = \frac{1}{n} \sum_{k=1}^n X_k$ est $\mathbf{E}X_1$. En effet,

$$\mathbf{E} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n \mathbf{E}X_k = \frac{1}{n} \sum_{k=1}^n \mathbf{E}X_1 = \mathbf{E}X_1.$$

De plus, la variance de M_n est

$$\mathbf{V} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n^2} \sum_{k=1}^n \mathbf{V}X_k = \frac{1}{n^2} \sum_{k=1}^n \mathbf{V}X_1 = \frac{\mathbf{V}(X_1)}{n}$$

(la variance de la somme est ici la somme des variance grâce à l'indépendance des X_i).

Soit $\varepsilon > 0$; l'inégalité de Bienaymé-Tchebychev nous donne

$$\mathbf{P}(|M_n - \mathbf{E}X_1| > \varepsilon) \leq \frac{\mathbf{V}(M_n)}{\varepsilon^2} = \frac{\mathbf{V}(X_1)}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0,$$

ce qui achève la démonstration. \square

6.3 Théorème limite central

Au vu de l'énoncé de la loi des grands nombres, une question légitime est maintenant de savoir à quelle vitesse la convergence de $\frac{1}{n} \sum_{k=1}^n X_k$ vers $\mathbf{E}X_1$ a lieu. La réponse à cette question est donnée par le théorème suivant :

Théorème 6.3.1 (Théorème limite central ^(note 1)). *Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires de carrés intégrables, indépendantes, de même loi. On suppose également que la variance commune des X_n est non nulle. On pose $M_n = \frac{1}{n} \sum_{k=1}^n X_k$. Alors pour tout intervalle I , on a*

$$\mathbf{P}\left(\sqrt{\frac{n}{\mathbf{V}(X)}}(M_n - \mathbf{E}X_1) \in I\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_I e^{-x^2/2} dx = \mathbf{P}(G \in I),$$

où G est une variable de loi normale centrée réduite.

Ce théorème a d'abord été démontré lorsque la loi commune des variables est une loi de Bernoulli de paramètre p (le théorème est dans ce cas connu sous le nom de théorème de Moivre-Laplace ^(note 2)). Dans ce cas, la variable $\sum_{k=1}^n X_k$ suit une loi binomiale de paramètres n et p . Cela explique que la distribution vue en figure 0.6 présente un profil proche de celui d'une Gaussienne.

Démonstration. Admis. On peut toutefois remarquer que la normalisation

$$\sqrt{\frac{n}{\mathbf{V}(X)}}(M_n - \mathbf{E}X_1)$$

est la bonne pour avoir une variable de moyenne 0 et de variance 1, comme la variable G qui apparaît sur le membre de droite. \square

Ce théorème dit que la variable $\sqrt{\frac{n}{\mathbf{V}(X)}}(M_n - \mathbf{E}X_1)$ a "presque la même loi" qu'une loi normale centrée réduite. On parle en fait de convergence *en loi*. On peut également le récrire sous la forme suivante, non rigoureuse mais plus intuitive :

$$\frac{1}{n} \sum_{k=1}^n X_k \simeq \mathbf{E}X_1 + \frac{\sigma(X)}{\sqrt{n}} \times G. \quad (6.1)$$

(note 1). Le nom de "théorème limite central" est une traduction de l'allemand "Zentraler Grenzwertsatz". Le terme "Grenzwertsatz" désigne un "théorème limite", c'est-à-dire un résultat de convergence d'une suite de variables aléatoires. Le nom de théorème limite *central* signifie ici que l'on a affaire à un théorème limite particulièrement important. On trouve parfois d'autres traductions, comme "théorème central limite" ou "théorème de la limite centrée", mais elles semblent moins bien correspondre au terme allemand.

(note 2). D'après *Abraham de Moivre*, il s'agit donc bien du théorème de Moivre-Laplace et non du théorème de Moivre-Laplace.

Autrement dit, la moyenne empirique est égale à l'espérance, avec une erreur à peu près Gaussienne de l'ordre de $\frac{1}{\sqrt{n}}$.

Si on multiplie (6.1) par n , on a l'expression

$$\sum_{k=1}^n X_k \simeq n \times \mathbf{E}X_1 + \sqrt{n} \times \sigma(X)G,$$

qui signifie qu'une somme de variables aléatoires de carré intégrable et de même loi se décompose en une partie déterministe d'ordre n , et d'une partie aléatoire d'ordre \sqrt{n} . Ceci est conforme à ce qu'on avait observé au chapitre 0, voir par exemple l'équation (0.2).

Dans le chapitre 0, on avait toutefois observé le cas de la marche aléatoire de Cauchy qui ne vérifiait pas cette décomposition. La raison en est que pour la marche aléatoire de Cauchy, les variables indépendantes et de même loi qui apparaissent sont de loi de Cauchy, et ne sont donc pas intégrables. Le comportement de l'ordre de n que l'on avait observé est dû au fait que la somme de n variables de loi de Cauchy de paramètre 1 indépendantes suit la loi de Cauchy de paramètre n (voir propriété 5.2.14).

6.4 Chaînes de Markov

Dans cette partie, nous allons effleurer la très riche théorie des phénomènes aléatoires dit *Markoviens*. Le principe est le suivant : une chaîne de Markov est une suite $(X_n)_{n \in \mathbf{N}}$ de variables aléatoires, qui n'est pas composée de variables indépendantes, mais telle que "la loi de X_{n+1} ne dépend que de la valeur de X_n ". Donnons une définition un peu plus précise :

Définition 6.4.1. *Soit E un ensemble fini. Une chaîne de Markov sur E est une suite de variables aléatoires $(X_n)_{n \in \mathbf{N}}$ telle que pour tous x_0, \dots, x_{n+1} de E , on a*

$$\mathbf{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = \mathbf{P}(X_{n+1} = x_{n+1} | X_n = x_n). \quad (6.2)$$

On résume parfois cette définition par la phrase (pas forcément très parlante) "conditionnellement au présent, le passé et le futur sont indépendants". En effet, définissons les trois évènements

$$\begin{aligned} T_- &= \{X_{n-1} = x_{n-1}, \dots, X_0 = x_0\}, \\ T_0 &= \{X_n = x_n\}, \\ T_+ &= \{X_{n+1} = x_{n+1}\}, \end{aligned}$$

correspondant respectivement au passé, au présent et au futur de la chaîne lorsque l'on est au temps n . Dans ce cas, l'équation (6.2) se réécrit

$$\mathbf{P}_{T_0}(T_+ | T_-) = \mathbf{P}_{T_0}(T_+),$$

ce qui signifie bien que si l'on se place sous la probabilité \mathbf{P}_{T_0} (autrement dit, si la position actuelle de la chaîne est connue), alors les évènements T_- et T_+ sont indépendants. On parle aussi de phénomène *sans mémoire*.

Une conséquence de la définition est que la loi de la suite $(X_n)_{n \in \mathbf{N}}$ est entièrement caractérisée par la donnée des $\mathbf{P}(X_n = y | X_{n-1} = x)$ et par la loi de X_0 . En effet :

Propriété 6.4.2. Soit x_0, \dots, x_n des éléments de E . Alors

$$\mathbf{P}(X_n = x_n, \dots, X_0 = x_0) = \mathbf{P}(X_0 = x_0) \prod_{k=1}^n \mathbf{P}(X_k = x_k | X_{k-1} = x_{k-1}).$$

C'est de ce résultat que vient le nom de *chaîne* de Markov : les relations entre X_n et X_{n+1} caractérisent la loi de toute la suite, comme la solidité d'une chaîne vient des jonctions entre deux maillons consécutifs.

Démonstration. C'est une récurrence sur n . Pour $n = 0$, il n'y a rien à démontrer.

Si le résultat est vrai pour un n donné, alors

$$\begin{aligned} \mathbf{P}(X_{n+1} = x_{n+1}, \dots, X_0 = x_0) &= \mathbf{P}(X_n = x_n, \dots, X_0 = x_0) \times \mathbf{P}(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) \\ &= \mathbf{P}(X_0 = x_0) \prod_{k=1}^n \mathbf{P}(X_k = x_k | X_{k-1} = x_{k-1}) \times \mathbf{P}(X_{n+1} = x_{n+1} | X_n = x_n), \end{aligned}$$

d'où le résultat. \square

Dans la suite, on se limitera au cas où $\mathbf{P}(X_n = y | X_{n-1} = x)$ ne dépend pas de n . On notera alors

$$P_{x,y} = \mathbf{P}(X_1 = y | X_0 = x).$$

P peut être vu comme une matrice de taille $N \times N$, où N est le nombre d'éléments de E . On l'appelle la *matrice de transition* de la chaîne de Markov. Cette matrice permet de calculer la loi de chacun des X_n .

Propriété 6.4.3. Si on note $\mu_x = \mathbf{P}(X_0 = x)$ la loi de X_0 , et si on voit μ comme un vecteur ligne, alors la loi de X_n est donnée par

$$\mathbf{P}(X_n = x) = (\mu P^n)_x.$$

On a noté μP le produit d'un vecteur ligne par une matrice :

$$(\mu P)_x = \sum_{y \in E} \mu_y P_{y,x}.$$

Démonstration. Comme on a supposé que $\mathbf{P}(X_{n+1} = y | X_n = x)$ ne dépendait pas de n , il suffit de montrer que la loi de X_1 est donnée par μP . Cela se prouve en utilisant la propriété 6.4.2 :

$$\mathbf{P}(X_1 = x) = \sum_{y \in E} \mathbf{P}(X_0 = y, X_1 = x) = \sum_{y \in E} \mathbf{P}(X_0 = y) \mathbf{P}(X_1 = x | X_0 = y) = \sum_{y \in E} \mu_y P_{y,x} = \mu P.$$

\square

Le théorème suivant donne de nombreuses informations sur le comportement asymptotique d'une chaîne de Markov.

Théorème 6.4.4. Soit P une matrice de transition sur un ensemble E fini, μ une probabilité sur E et $(X_n)_{n \in \mathbf{N}}$ une chaîne de Markov de matrice de transition P telle que X_0 soit de loi μ .

1. Si μ vérifie $\mu P = \mu$, alors X_n est de loi μ pour tout n . On dit d'un tel μ qu'il s'agit d'une mesure invariante.

2. Une chaîne de Markov admet au moins une mesure invariante.
3. Si il existe une unique mesure invariante $\bar{\mu}$, alors $\frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=x\}}$ converge presque sûrement vers $\bar{\mu}_x$.
4. Si 1 est la seule valeur propre de module 1 de P , alors il existe une unique mesure invariante $\bar{\mu}$ et la loi de $\mathbf{P}(X_n = x)$ converge vers $\bar{\mu}_x$ pour tout x .

Démonstration. Admis. □

6.5 Exercices

Exercice 6.1.

Justifier la valeur de la constante $\sqrt{5/18\pi}$ donnée en équation (0.1), page 9.

Exercice 6.2.

[Jeu de Yam's]

Cet exercice a pour but de calculer la probabilité de gagner au jeu de Yam's. On rappelle les règles de ce jeu :

- on lance 5 dés standards ;
- parmi ces 5 dés, on met de côté les dés de son choix, et on relance les autres ;
- on réitère une fois cette dernière étape avec les 5 valeurs obtenues.

On gagne si on obtient 5 dés de même valeur. On adopte la stratégie suivante qui se trouve être la stratégie maximisant la probabilité de gagner : à chaque lancer, on conserve le plus grand groupe de dés de même valeurs, et on relance tous les autres.

Une fois cette stratégie fixée, la taille X_i du plus grand groupe de dés de même valeur au bout du i ème lancer est une chaîne de Markov à valeurs dans l'ensemble $\{1, 2, 3, 4, 5\}$.

1. Écrire la matrice de transition de la chaîne $(X_i)_{i \in \mathbf{N}}$.
2. Exprimer la loi de X_1 .
3. En déduire la probabilité de gagner au jeu de Yam's, qui correspond à $\mathbf{P}(X_3 = 5)$.

Exercice 6.3.

[Marche aléatoire sur un graphe]

On considère un graphe fini. Plus précisément, on considère un ensemble $\{1, \dots, N\}$, dont certains des éléments sont reliés deux par deux. On définit alors $A_{ij} = 1$ quand deux éléments sont reliés, et $A_{ij} = 0$, sinon. La matrice A obtenue est une matrice symétrique de diagonale nulle (on considère que qu'un point ne peut pas être relié à lui-même).

On définit une marche aléatoire sur le graphe de la façon suivante : X_0 est une variable aléatoire quelconque à valeurs dans $\{1, \dots, N\}$. Conditionnellement à $X_n = i$, la variable X_{n+1} suit une loi uniforme parmi les j tels que $A_{ij} = 1$.

1. Montrer que $(X_n)_{n \in \mathbf{N}}$ est une chaîne de Markov dont on précisera la matrice de transition.
2. Montrer que la mesure $(\mu_i)_{i \in \{1, \dots, N\}}$ définie par $\mu_i = \sum_j A_{ij} / \sum_{k,j} A_{k,j}$ est une mesure de probabilité invariante pour la chaîne.

Exercice 6.4.

[Le modèle d'Ehrenfest]

On considère deux compartiments A et B , dans lesquels sont placés N particules. À chaque instant entier, on choisit uniformément une des N particules, et on la fait changer de compartiment.

1. Montrer que si X_n est le nombre de particules dans le compartiment A à l'instant n , alors $(X_n)_{n \in \mathbf{N}}$ est une chaîne de Markov dont on précisera la matrice de transition.
2. Montrer que la loi binomiale de paramètres N et $1/2$ est une mesure invariante pour cette chaîne de Markov.
3. Retrouver ce résultat en utilisant la conclusion de l'exercice précédent. On pourra représenter la position des particules par un N -uplet (x_1, \dots, x_N) où x_i vaut 1 si la i ème particule est dans le compartiment A , et vaut 0 sinon.

Exercice 6.5.

[Marche aléatoire absorbée]

On considère une marche aléatoire sur l'ensemble $\{0, \dots, N\}$. Quand la marche se trouve au point i , elle a une probabilité $1/2$ d'aller en $i - 1$ et $1/2$ d'aller en $i + 1$. Seule exception : partant des points 0 et N , la marche ne bouge plus, avec probabilité 1. On parle donc d'une marche aléatoire *absorbée* aux points 0 et N . Par conséquent, il y a *a priori* trois comportements possibles en temps long :

- ou bien la chaîne finit absorbée en 0 ;
- ou bien elle finit absorbée en N ;
- ou bien la suite n'atteint jamais les points 0 et N .

On notera $(X_n)_{n \in \mathbf{N}}$ la suite des valeurs prises par la marche.

1. Écrire la matrice de transition de $(X_n)_{n \in \mathbf{N}}$.
2. Dans cette question, on va calculer la probabilité de finir absorbé à l'une ou l'autre des extrémités. On désigne par G et D les deux événements

$$G = \{X_n = 0 \text{ à partir d'un certain rang}\} \text{ et } D = \{X_n = N \text{ à partir d'un certain rang}\}.$$

On note $\gamma_i = \mathbf{P}(G|X_0 = i)$ et $\delta_i = \mathbf{P}(D|X_0 = i)$.

- (a) Montrer les relations, pour $i \in \{1, \dots, N - 1\}$,

$$\gamma_i = \frac{1}{2}(\gamma_{i+1} + \gamma_{i-1}) \text{ et } \delta_i = \frac{1}{2}(\delta_{i+1} + \delta_{i-1})$$

ainsi que

$$\gamma_0 = 1, \quad \gamma_N = 0, \quad \delta_0 = 0, \quad \delta_N = 1.$$

- (b) En déduire l'expression de γ_i et δ_i pour tout i .
- (c) En déduire la probabilité que la marche ne soit jamais absorbée, en partant d'un point i quelconque (il s'agit de la probabilité $\mathbf{P}(D^c \cap G^c|X_0 = i)$).
3. Dans cette question, on va donner la loi explicite du temps d'absorption, pour une condition initiale particulière.

On suppose que X_0 suit la loi $(C \sin(k\pi/N))_{k=0, \dots, N}$.

- (a) Donner la valeur de C .
- (b) Quelle est la loi de X_n ?
- (c) Montrer que le temps avant que met la suite à atteindre les points 0 ou N suit une loi géométrique dont on précisera le paramètre.

Chapitre 7

Estimation

7.1 Estimation d'une proportion

On cherche à prévoir le résultat d'une élection : on a une population de N individus à laquelle sont proposés plusieurs candidats. Monsieur Michu est candidat à cette élection et veut avoir une estimation de son score $p \in [0, 1]$ (la proportion des N individus qui vont voter pour lui). Une méthode pour obtenir une telle estimation est d'interroger n personnes choisies aléatoirement dans la population, et de calculer la proportion \hat{p}_n de personnes qui vont voter pour monsieur Michu parmi ces n personnes.

Formalisons un peu le problème : pour $1 \leq k \leq n$, on va définir le réel X_k par $X_k = 1$ si la k ème personne choisie aléatoirement va voter pour monsieur Michu, et $X_k = 0$ dans le cas contraire. Si les personnes sont choisies indépendamment et uniformément dans la population ^(note 1), les variables X_k sont alors des variables de Bernoulli indépendantes de paramètres p .

L'estimation de la proportion p est donnée par

$$\hat{p}_n = \frac{1}{n} \sum_{k=1}^n X_k.$$

On dit que \hat{p}_n est un *estimateur* de p . On reconnaît la moyenne empirique d'une suite variables aléatoires. D'après la loi des grands nombres, on a convergence de \hat{p}_n vers p .

Théorème 7.1.1. *L'estimateur \hat{p}_n est convergent, au sens où \hat{p}_n converge presque sûrement et en probabilité vers p :*

$$\mathbf{P}(\hat{p}_n \rightarrow p) = 1 \text{ et } \forall \varepsilon > 0, \mathbf{P}(|\hat{p}_n - p| > \varepsilon) \rightarrow_n 0.$$

Toutefois, le résultat donné par \hat{p}_n n'a aucune raison d'être *égal* à p , et une question naturelle que monsieur Michu va se poser est "Quelle fiabilité puis-je accorder à ce résultat?". La réponse est une conséquence du théorème limite central :

Théorème 7.1.2. *Soit $c > 0$. Alors $\mathbf{P}(|\hat{p}_n - p| \leq \frac{c}{2\sqrt{n}})$ converge et on a*

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(|\hat{p}_n - p| \leq \frac{c}{2\sqrt{n}} \right) \geq \mathbf{P}(|G| \leq c) = \sqrt{\frac{2}{\pi}} \int_0^c e^{-x^2/2} dx,$$

(note 1). Cela implique en particulier qu'il serait possible d'interroger deux fois la même personne. Toutefois, cela n'arrive que rarement si le nombre de personnes interrogées est petit devant la taille de la population ($n \ll N$).

(G désigne une variable aléatoire de loi normale centrée réduite).

Démonstration. Les X_i sont ici des variables de loi de Bernoulli de paramètre p , dont la moyenne et la variance sont donc respectivement p et $p(1-p)$. D'après le théorème limite central, on a donc

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{P} \left(|\hat{p}_n - p| \leq \frac{c}{2\sqrt{n}} \right) &= \lim_{n \rightarrow \infty} \mathbf{P} \left(\sqrt{\frac{n}{p(1-p)}} |\hat{p}_n - p| \leq \frac{c}{2\sqrt{p(1-p)}} \right) = \mathbf{P} \left(|G| \leq \frac{c}{2\sqrt{p(1-p)}} \right) \\ &\geq \mathbf{P} (|G| \leq c) \\ &= \sqrt{\frac{2}{\pi}} \int_0^c e^{-x^2/2} dx. \end{aligned}$$

L'inégalité découle ici du fait que $2\sqrt{p(1-p)} \leq 1$ pour tout $p \in [0, 1]$. □

Pour un c fixé, on sait donc que, asymptotiquement, la probabilité que p soit dans l'intervalle

$$\left[\hat{p}_n - \frac{c}{2\sqrt{n}}, \hat{p}_n + \frac{c}{2\sqrt{n}} \right]$$

est d'au moins $\sqrt{\frac{2}{\pi}} \int_0^c e^{-x^2/2} dx$.

En pratique, on se fixe une valeur $\alpha \in]0, 1[$, appelé le niveau de l'intervalle de confiance, et on cherche un $c > 0$ tel que

$$\sqrt{\frac{2}{\pi}} \int_0^c e^{-x^2/2} dx = \alpha.$$

Comme la fonction $c \mapsto \sqrt{\frac{2}{\pi}} \int_0^c e^{-x^2/2} dx$ est continue, strictement croissante, envoie 0 sur 0 et converge vers 1 quand c tend vers $+\infty$, un tel c existe et est unique. Un choix courant est $\alpha = 0.95$, pour lequel $c = 1.96$ est une valeur presque exacte (note 2). Pour cette valeur de c , on peut notamment légèrement majorer $c/2$ par 1, et on obtient le résultat suivant :

Propriété 7.1.3. *La quantité*

$$\mathbf{P} \left(|p - \hat{p}_n| \leq \frac{1}{\sqrt{n}} \right)$$

converge quand n tend vers ∞ , et sa limite est supérieure à 0.95.

Le paragraphe suivant est extrait du programme de classe de seconde (ce que nous notons \hat{p}_n y est noté f) :

Le professeur peut indiquer aux élèves le résultat suivant, utilisable dans la pratique pour des échantillons de taille $n \geq 25$ et des proportions p du caractère comprises entre 0.2 et 0.8 : si f désigne la fréquence du caractère dans l'échantillon, f appartient à l'intervalle $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$ avec une probabilité d'au moins 0.95. Le professeur peut faire percevoir expérimentalement la validité de cette propriété mais elle n'est pas exigible.

À proprement parler, le résultat donné est *faux*. Par exemple, pour $n = 27$ et $p = 0.5$, on a l'équivalence (en posant $Y_n = \sum_{k=1}^n X_k$)

$$\left| \frac{Y_n}{n} - p \right| \leq \frac{1}{\sqrt{n}} \Leftrightarrow Y_n \in \{9, 10, \dots, 28\}.$$

(note 2). Avec plus de décimales, on a $c = 1.95996398\dots$

Or, le calcul donne

$$\mathbf{P}(Y_n \in \{9, \dots, 28\}) = 0.9478\dots$$

De même, pour $n = 55$ et $p = 0.5$, on a l'équivalence

$$\left| \frac{Y_n}{n} - p \right| \leq \frac{1}{\sqrt{n}} \Leftrightarrow Y_n \in \{21, 22, \dots, 34\},$$

mais on peut montrer

$$\mathbf{P}(Y_n \in \{21, \dots, 34\}) = 0.9419\dots$$

Il faut toutefois bien voir que le résultat cité par le programme est *presque vrai*, la version rigoureuse en étant la propriété 7.1.3.

7.2 Intervalles de confiance

Dans cette partie, nous allons généraliser la situation de la partie précédente à des variables aléatoires qui ne suivent pas forcément des lois de Bernoulli. On considère une suite de variables aléatoires de carré intégrable $(X_n)_{n \in \mathbf{N}}$ de même loi, dont la moyenne $m = \mathbf{E}X_1$ est inconnue, mais la variance σ^2 est connue.

On appellera *intervalle de confiance* de niveau α sur m un intervalle *aléatoire* \hat{I}_n tel que

$$\mathbf{P}(m \in \hat{I}_n) \geq \alpha.$$

Par l'expression "intervalle aléatoire", on désigne simplement un intervalle dont les bornes sont des variables aléatoires. Dans l'expression $\mathbf{P}(m \in \hat{I}_n)$, on notera bien que la valeur m est parfaitement déterministe, c'est l'intervalle \hat{I}_n qui est aléatoire.

Par exemple, la propriété 7.1.3 énonce en fait que l'intervalle $\hat{I}_n = \left[\hat{p}_n - \frac{1}{\sqrt{n}}, \hat{p}_n + \frac{1}{\sqrt{n}} \right]$ est asymptotiquement ^(note 3) un intervalle de confiance pour p de niveau 0.95.

Si on revient à l'exemple de l'élection, les échantillons habituellement utilisés lors des sondages en France sont de l'ordre de 1000 personnes. La quantité $\frac{1}{\sqrt{n}}$ correspond alors environ à 3 points de pourcentage : un sondage sur 1000 personnes donnera un score en pourcentage correct à $\pm 3\%$.

7.2.1 Par l'inégalité de Bienaymé-Tchebychev

En appliquant l'inégalité de Bienaymé-Tchebychev à la variable aléatoire $\hat{m}_n = \frac{1}{n} \sum_{k=1}^n X_k$, dont la moyenne est m et la variance est $\frac{\sigma^2}{n}$, on obtient :

$$\mathbf{P}(|\hat{m}_n - m| \leq a) \geq 1 - \frac{\sigma^2}{a^2 n}.$$

En posant $a = c \frac{\sigma}{\sqrt{n}}$, on a

$$\mathbf{P}\left(|\hat{m}_n - m| \leq \frac{c\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{c^2}.$$

Cette équation peut se récrire comme le fait que l'intervalle

$$\hat{I}_n = \left[\hat{m}_n - \frac{c\sigma}{\sqrt{n}}, \hat{m}_n + \frac{c\sigma}{\sqrt{n}} \right]$$

(note 3). Car la propriété 7.1.3 énonce une *convergence* quand n tend vers ∞ et non pas une *égalité*.

est un intervalle de confiance de niveau $\alpha = 1 - \frac{1}{c^2}$ pour m . Par exemple, pour $\alpha = 0.95$, on trouve $c \simeq 4.47$.

7.2.2 Dans le cas Gaussien

Dans le cas où les X_i sont des variables aléatoires indépendantes de loi Gaussienne, alors \hat{m}_n suit la loi Gaussienne de moyenne m et de variance $\frac{\sigma^2}{n}$. Par conséquent, on a

$$\mathbf{P}\left(|\hat{m}_n - m| \leq \frac{c\sigma}{\sqrt{n}}\right) = \mathbf{P}(|G| \leq c),$$

où G suit la loi normale centrée réduite. Autrement dit, l'intervalle

$$\hat{I}_n = \left[\hat{m}_n - \frac{c\sigma}{\sqrt{n}}, \hat{m}_n + \frac{c\sigma}{\sqrt{n}} \right]$$

est un intervalle de confiance de niveau $\mathbf{P}(|G| \leq c)$. Par exemple, pour un intervalle de niveau $\alpha = 0.95$, on trouve $c \simeq 1.96$.

On remarque que pour un même niveau, l'intervalle obtenu ici est beaucoup plus étroit que celui obtenu par l'inégalité de Bienaymé-Tchebychev (pour $\alpha = 0.95$, on a $c = 1.96$ au lieu de $c = 4.47$, soit un intervalle plus que deux fois plus fin).

7.2.3 Par le théorème limite central

Dans le cas général où les X_i ne sont pas Gaussiens, on a quand même, par le théorème limite central :

$$\mathbf{P}\left(|\hat{m}_n - m| < \frac{c\sigma}{\sqrt{n}}\right) \xrightarrow[n \rightarrow \infty]{} \mathbf{P}(|G| < c).$$

Autrement dit, l'intervalle

$$\hat{I}_n = \left[\hat{m}_n - \frac{c\sigma}{\sqrt{n}}, \hat{m}_n + \frac{c\sigma}{\sqrt{n}} \right]$$

est un intervalle de confiance asymptotique de niveau $\mathbf{P}(|G| \leq c)$. Noter que l'intervalle de confiance est le même que dans la partie 7.2.2, avec le même niveau. La différence avec le cas précédent est qu'ici l'égalité n'est vraie que pour $n \rightarrow \infty$, d'où le terme "intervalle de confiance *asymptotique*".

7.3 Exercices

Exercice 7.1.

Soit (X_1, \dots, X_n) des variables aléatoires de même loi de carré intégrable.

1. On pose $\hat{m}_n = \frac{1}{n} \sum_{k=1}^n X_k$. Montrer que \hat{m}_n est un estimateur convergent de la moyenne commune des X_k et que $\mathbf{E}\hat{m}_n = \mathbf{E}X_1$.
2. Montrer que $\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - m)^2$ est un estimateur convergent de la variance des X_k et que $\mathbf{E}\tilde{\sigma}_n^2 = \mathbf{V}(X_1)$.

3. Montrer que $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \hat{m}_n)^2$ converge la variance de l'échantillon. A-t-on $\mathbf{E}\hat{\sigma}_n^2 = \mathbf{V}(X_1)$?

Exercice 7.2.

Pour tout nombre réel θ , on définit l'application p_θ sur \mathbf{R} par :

$$p_\theta(x) = \begin{cases} 0 & \text{si } x < \theta \\ e^{-(x-\theta)} & \text{si } x \geq \theta. \end{cases}$$

1. Montrer que, pour tout nombre réel θ , l'application p_θ est une densité de probabilité sur \mathbf{R} .
2. Soit X_0 une variable aléatoire de densité p_θ . Montrer que X_0 admet un moment d'ordre deux, et expliciter son espérance et sa variance.
3. Déterminer la fonction de répartition F_0 de X_0 .
4. Soit $n \geq 1$ et (X_1, \dots, X_n) un échantillon de taille n de la loi de X_0 .
On pose $U_n = \min(X_1, \dots, X_n)$ et $M_n = \frac{1}{n} \sum_{i=1}^n X_i$.
 - (a) Pour tout nombre réel t , calculer $\mathbf{P}[U_n > t]$. En déduire la loi de U_n .
 - (b) Montrer que $M_n - 1$ et $U_n - \frac{1}{n}$ sont des estimateurs sans biais de θ .
5. Construire deux intervalles de confiance pour θ au niveau 0.95 à partir des variables M_n et U_n respectivement. Comparer les deux intervalles obtenus.

Exercice 7.3.

Soit $(X_n)_{n \in \mathbf{N}}$ une suite de variables aléatoires indépendantes de loi exponentielle de même paramètre λ . On rappelle que $\mathbf{E}X = \frac{1}{\lambda}$ et $\mathbf{V}X = \frac{1}{\lambda^2}$. On souhaiterait estimer la valeur de λ .

1. Montrer que $\hat{\lambda}_n = \frac{n}{\sum_{k=1}^n X_k}$ est un estimateur de λ convergeant presque sûrement.
2. À partir de l'estimateur $\hat{\lambda}_n$, construire un intervalle de confiance asymptotique pour λ de niveau 0.95.

Chapitre 8

Tests d'hypothèses

Dans ce chapitre, on présente à partir d'un exemple la notion de test statistique. Cette notion est un cadre général qui inclut en particulier les intervalles de fluctuation qui sont au programme du lycée.

8.1 Exemple introductif

La police doit contrôler la vitesse des véhicules sur la route et verbaliser les véhicules dépassant une certaine vitesse limite. Toutefois, les instruments mesurant la vitesse peuvent souffrir d'erreurs de mesures, et un véhicule contrôlé légèrement au dessus de la vitesse limite pouvait en fait être légèrement en dessous de cette limite.

Les policiers sont alors dans une situation où le bon comportement à adopter se déduit d'un *test statistique*. Plus précisément, on doit choisir à partir d'observations aléatoires (l'aléa venant ici de l'erreur de mesure du radar) entre deux possibilités (verbaliser ou non le conducteur), qui ne jouent pas des rôles symétriques (verbaliser demandera de mettre en branle une procédure, qui pourra avoir été lancée pour rien si le conducteur prouve qu'il n'était pas en excès de vitesse).

Dans le cadre d'un test statistique, les deux possibilités entre lesquelles on doit choisir sont notées H_0 (ou hypothèse *nulle*) et H_1 . L'hypothèse H_0 sera l'hypothèse à privilégier en cas de doute, alors que l'hypothèse H_1 sera choisie uniquement lorsque l'on aura une très forte présomption. Dans le cas du contrôle de vitesse, l'hypothèse H_0 est "pas d'excès de vitesse" et H_1 est "excès de vitesse". Autrement dit, si la vitesse mesurée est proche de la vitesse limite, on préférera considérer que le conducteur n'est pas en excès de vitesse, même si la vitesse mesurée est légèrement trop élevée.

Il s'agit donc d'éviter de choisir H_1 lorsque l'on est sous H_0 (dans notre exemple, verbaliser quelqu'un à tort). Remarquons qu'il existe une méthode très simple pour cela : il suffit de *toujours* choisir H_0 , ce qui revient, dans l'exemple du contrôle de la vitesse, à ne distribuer aucune amende. Bien entendu, ce choix n'est pas satisfaisant. Le bon choix est en fait est de contrôler la *probabilité* de choisir l'hypothèse H_1 si l'on est sous l'hypothèse H_0 .

Formalisons notre exemple du contrôle de la vitesse : on suppose que la vitesse mesurée X suit une loi normale de moyenne v (la vraie vitesse du véhicule) et de variance σ^2 (quantifiant l'incertitude de la mesure). La vitesse maximale autorisée est notée v_0 . On a donc les deux hypothèses :

$$H_0 = "v \leq v_0" \quad \text{et} \quad H_1 = "v > v_0".$$

On va considérer que le véhicule est en excès de vitesse lorsque la valeur de X observée est supérieure à un certain seuil v_s . La probabilité de déclarer l'automobiliste en excès de vitesse, en fonction de la vraie vitesse du véhicule est alors

$$p(v) = \mathbf{P}(X > v_s) = \mathbf{P}(v + \sigma G > v_s) = \mathbf{P}(G > (v_s - v)/\sigma),$$

où G suit la loi normale centrée réduite. On souhaite alors que sous H_0 (c'est-à-dire si $v \leq v_0$) cette probabilité soit inférieure à une certaine valeur α . Or on remarque que :

$$\sup_{v \leq v_0} p(v) = \mathbf{P}(G > (v_s - v_0)/\sigma).$$

On va donc choisir la valeur de seuil v_s de sorte que cette probabilité soit (inférieure ou) égale à α . Par exemple, pour $\alpha = 0.025$, on veut $(v_s - v_0)/\sigma \geq 1.96$, de sorte que l'on doit choisir $v_s \geq v_0 + 1.96\sigma$.

8.2 Test pour la moyenne

On va maintenant présenter un autre test dans un cadre un peu plus général. On observe (X_1, \dots, X_n) une suite de n variables aléatoires indépendantes et de même loi Gaussienne de moyenne m et de variance σ^2 supposée connue. On veut alors comparer la moyenne m avec une valeur de référence m_0 . Suivant les cas, on peut faire un des deux choix suivants d'hypothèse $H_1 - H_0$.

1. On se demande si m est exactement égal à m_0 , ce qui revient à tester H_0 : " $m = m_0$ " contre H_1 : " $m \neq m_0$ "
2. On sait que la valeur de m est au moins supérieure ou égale à m_0 , et on cherche à tester l'hypothèse H_0 : " $m = m_0$ " contre l'hypothèse H_1 : " $m > m_0$ "

Dans ce cadre, la variable $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ suit une loi normale de moyenne m et de variance σ^2/n . Autrement dit, si G suit une loi normale centrée réduite, \bar{X}_n a la même loi que $m + \frac{\sigma}{\sqrt{n}}G$. Dans le cas du choix 1., on acceptera l'hypothèse H_0 si \bar{X}_n est dans un intervalle de la forme

$$I_n = \left[m_0 - c \frac{\sigma}{\sqrt{n}}, m_0 + c \frac{\sigma}{\sqrt{n}} \right],$$

et on acceptera H_1 dans le cas contraire. La valeur de c sera à choisir de sorte que

$$\mathbf{P} \left(m_0 - c \frac{\sigma}{\sqrt{n}} \leq \bar{X}_n \leq m_0 + c \frac{\sigma}{\sqrt{n}} \right) = \mathbf{P}(-c \leq G \leq c) \geq 1 - \alpha,$$

où α est la probabilité d'erreur choisie a priori.

Dans le cas du choix 2., H_0 sera acceptée si \bar{X}_n se situe dans un ensemble de la forme

$$I'_n = \left] -\infty, m_0 + c \frac{\sigma}{\sqrt{n}} \right],$$

cette fois-ci, c sera choisi de sorte que

$$\mathbf{P} \left(\bar{X}_n \leq m_0 + c \frac{\sigma}{\sqrt{n}} \right) = \mathbf{P}(G \leq c) \geq 1 - \alpha.$$

Dans le cas où la variance est inconnue, on pourra remplacer dans les intervalles ci-dessus la valeur exacte de σ^2 par son approximation $\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$. Toutefois, le résultat théorique justifiant ce remplacement (le théorème de Slutsky) ne nous est pas accessible dans le cadre de ce cours.

8.3 Exercices

Exercice 8.1.

En France, les excès de vitesse sont sanctionnés à partir d'un dépassement de 5km.h^{-1} , par exemple à partir de 95km.h^{-1} quand la vitesse maximale autorisée est de 90km.h^{-1} . En supposant que la vitesse est mesurée avec une erreur de loi normale de variance σ^2 , quelle valeur doit prendre σ pour qu'au moins 99% des personnes verbalisées aient vraiment été en excès de vitesse ?

Annexe A

Indications pour les exercices

Chapitre 1

Exercice 1.3.

Les trois premières phrases décrivent respectivement les évènements suivants :

- $A \cap B^c \cap C^c$;
- $A \cup B \cup C$;
- $A^c \cup B^c \cup C^c = (A \cap B \cap C)^c$.

Les deux dernières décrivent les relations suivantes entre évènements :

- $A \subset B \cup C$;
- $A \subset (B \cap C)^c$.

Exercice 1.6.

Les deux suites convergent respectivement vers $\mathbf{P}(X \leq 0)$ et $\mathbf{P}(X > 0)$.

Exercice 1.7.

Entre parenthèse, une version modifiée qui est une fonction de répartition

$$\text{non } \left(\frac{1}{\pi} \arctan(x) + 0.5 \right) ; \text{ non } ((1 - e^{-x})\mathbf{1}_{x>0}) ; \text{ non } (x\mathbf{1}_{[0,1]}(x) + \mathbf{1}_{]1,\infty[}(x)) ;$$

$$\text{oui } ; \text{ non } \left(\frac{x}{x+1} \mathbf{1}_{x>0} \right) ; \text{ non } \left(\left(\frac{1}{\pi} \arcsin(x) + 0.5 \right) \mathbf{1}_{[-1,1]}(x) + \mathbf{1}_{]1,\infty[}(x) \right) ;$$

$$\text{non } (\tilde{E}(1/x)^{-1} \mathbf{1}_{[0,1]}(x), \tilde{E} : \text{partie entière supérieure}).$$

Chapitre 2

Chapitre 3

Exercice 3.2.

Les réponses sont les suivantes :

1. Les variables U et V ne sont pas indépendantes,
2. On a

$$P(U = n, V = m) = \begin{cases} 2/N^2 & \text{si } n < m, \\ 1/N^2 & \text{si } n = m, \\ 0 & \text{si } n > m. \end{cases}$$

3. La probabilité $\mathbf{P}(V = m)$ vaut $(2m - 1)/N^2$.
4. $\mathbf{P}(U = n|V = m)$ vaut $2/(2m - 1)$ si $n < m$, $1/(2m - 1)$ si $n = m$, et 0 si $n > m$.
5. $\mathbf{P}(X = n|V = m)$ vaut $1/(2m - 1)$ si $n < m$, $m/(2m - 1)$ si $n = m$, et 0 si $n > m$.

Exercice 3.3.

Le taux de succès paraît meilleur pour le traitement B si on ne différencie pas selon la taille des calculs, mais le pour une taille de calcul donnée, le traitement B paraît moins efficace.

L'explication tient dans le fait que les petits calculs sont plus simples à soigner que les gros (meilleur taux de réussite pour les deux traitement), ce qui favorise sur le papier le traitement B , qui a majoritairement été testé sur des patients présentant des petits calculs.

Exercice 3.4.

Les trois probabilités demandées valent respectivement $(2/5)^3 = 0.064$, $1/10 = 0.1$ et $183/500 = 0.366$.

Exercice 3.5.

On choisit comme univers, $\Omega = \{(G, G), (G, F), (F, G), (F, F)\}$, où la première coordonnée (resp. la deuxième) donne le sexe de l'aîné (du cadet) des enfants. Comme la famille est choisie au "hasard", on munit Ω de la probabilité uniforme, notée \mathbf{P} , de sorte que pour tout événement A , $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{4}$.

Soit A l'évènement "la famille a au moins un garçon", B "l'aîné de la famille est un garçon", C "les deux enfants sont des garçons". Alors, $A \cap C = C$, $B \cap C = C$, et

$$\mathbf{P}(A) = \mathbf{P}\{(G, G), (G, F), (F, G)\} = \frac{3}{4} > 0,$$

$$\mathbf{P}(B) = \mathbf{P}\{(G, G), (G, F)\} = \frac{1}{2} > 0,$$

$$\mathbf{P}(C) = \mathbf{P}\{(G, G)\} = \frac{1}{4}.$$

En particulier les probabilités conditionnées par rapport à A ou B sont bien définies. La probabilité recherchée au point 1. est :

$$\mathbf{P}_A(C) = \frac{\mathbf{P}(A \cap C)}{\mathbf{P}(A)} = \frac{\mathbf{P}(C)}{\mathbf{P}(A)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

La probabilité recherchée au point 2. est :

$$\mathbf{P}_B(C) = \frac{\mathbf{P}(B \cap C)}{\mathbf{P}(B)} = \frac{\mathbf{P}(C)}{\mathbf{P}(B)} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Exercice 3.6.

Chaque carte possède deux faces, que l'on va distinguer. Choisissons l'univers

$$\Omega = \{R_1, R_2, N_1, N_2, R_3, N_3\},$$

où par exemple l'évènement élémentaire R_1 est réalisé si c'est la première face de la carte unicolore rouge qui est apparente. Étant donné que la carte est tirée "au hasard", on munit Ω de la probabilité uniforme, notée \mathbf{P} , de sorte que pour tout évènement A , $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{card}(A)}{6}$.

On souhaite calculer la probabilité conditionnelle de l'évènement R_3 sachant que l'évènement R_1 ou R_2 ou R_3 est réalisé. Soit $A = R_1 \cup R_2 \cup R_3$, alors $\mathbf{P}(A) = \frac{1}{2}$, et la probabilité conditionnelle $\mathbf{P}_A(R_3)$ est bien définie. D'après la définition de la probabilité conditionnelle, nous avons :

$$\mathbf{P}_A(R_3) = \frac{\mathbf{P}(A \cap R_3)}{\mathbf{P}(A)} = \frac{\mathbf{P}(R_3)}{\mathbf{P}(A)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Sans faire attention, on pourrait penser que cette probabilité est de $1/2$, pensant qu'à partir du moment où le côté rouge apparaît, il reste deux situations équiprobables.

Exercice 3.7.

Soit k le nombre de boules contenues dans l'urne, avec $k = 9$ ou $k = 12$. On choisit comme espace des états $\Omega_k = \{1, 2, \dots, k\}$. Comme la boule est tirée "au hasard", on munit Ω_k de la probabilité uniforme, notée \mathbf{P}_k , de sorte que pour tout évènement A ,

$$\mathbf{P}_k(A) = \frac{\text{card}(A)}{\text{card}(\Omega_k)} = \frac{\text{card}(A)}{k}.$$

Dans le cas où $k = 9$, les évènements $A, B, A \cap B$ s'écrivent :

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

Ainsi, $\mathbf{P}_9(A) = \frac{4}{9}$, $\mathbf{P}_9(B) = \frac{1}{3}$, $\mathbf{P}_9(A \cap B) = \frac{1}{9}$. Comme $\mathbf{P}_9(A \cap B) \neq \mathbf{P}_9(A) \mathbf{P}_9(B)$, on en déduit que les évènements A et B ne sont pas indépendants.

Dans le cas où $k = 12$, les évènements $A, B, A \cap B$ s'écrivent :

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

Ainsi, $\mathbf{P}_{12}(A) = \frac{1}{2}$, $\mathbf{P}_{12}(B) = \frac{1}{3}$, $\mathbf{P}_{12}(A \cap B) = \frac{1}{6}$. Comme $\mathbf{P}_{12}(A \cap B) = \mathbf{P}_{12}(A) \mathbf{P}_{12}(B)$, on en déduit que les évènements A et B sont indépendants.

Exercice 3.8.

La probabilité de tirer plus de boules noires que de boules blanches est de $7/10$.

Le nombre moyen de boules noires dans l'urne est 2 avant le premier lancer, puis $8/3$ avant le deuxième, $10/3$ avant le troisième, et 4 après le troisième lancer. En fait, on peut remarquer que la proportion moyenne de boules noires reste constante égale à $2/3$. Ce fait est général et ne dépend pas de la composition initiale de l'urne.

Exercice 3.9.

La probabilité que la première boule tirée soit noire est de $1/2$. La probabilité de tirer deux boules de même couleur est $(2n + 1)/3n$.

Chapitre 4

Exercice 4.1.

L'univers Ω est formé de tous les 23-uplets de jours d'anniversaire. On a donc $\Omega = \{1, \dots, 365\}^{23}$ et

$\text{card}(\Omega) = 365^{23}$. On suppose que les dates d'anniversaire sont distribuées "au hasard" sur l'année, de sorte que l'on munit Ω de la probabilité uniforme, notée \mathbf{P} . Ainsi, si $A \in \mathcal{P}(\Omega)$ est un évènement, $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.

On considère l'évènement A défini par "les personnes ont toutes leur anniversaire un jour différent". L'évènement A est formé de tous les échantillons de taille 23 sans répétition, et son cardinal est donc donné par un nombre d'arrangements : $\text{card}(A) = A_{365}^{23}$, et

$$\mathbf{P}(A) = \frac{A_{365}^{23}}{(365)^{23}}.$$

La probabilité qu'au moins deux personnes aient leur anniversaire le même jour est la probabilité de l'évènement complémentaire, et est donc égale à $1 - \frac{A_{365}^{23}}{(365)^{23}} = 0,507\dots$ Cette probabilité est d'environ 0,97 s'il y a 50 personnes, et d'environ 0,999 s'il y en a 100.

Exercice 4.2.

L'univers est l'ensemble des tiercés possibles, soit

$$\Omega = \{(i, j, k) : i, j, k \in \{1, \dots, n\}, i \neq j, j \neq k, k \neq i\}.$$

Alors, $\text{card}(\Omega) = A_n^3 = n(n-1)(n-2)$. On suppose que tous les tiercés ont la même chance de se réaliser, de sorte que l'on munit Ω de la probabilité uniforme, notée \mathbf{P} . Ainsi, pour tout $A \in \mathcal{P}(\Omega)$, $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.

1. Soit A l'évènement "obtenir le tiercé gagnant dans l'ordre". Comme il existe un unique tiercé gagnant, $\text{card}(A) = 1$, d'où la probabilité cherchée est :

$$\mathbf{P}(A) = \frac{1}{n(n-1)(n-2)}.$$

2. Soit B l'évènement "obtenir le tiercé gagnant dans l'ordre ou dans un ordre différent". Comme il y a $3!$ manières d'ordonner le tiercé gagnant, $\text{card}(B) = 6$, d'où la probabilité cherchée est :

$$\mathbf{P}(B) = \frac{6}{n(n-1)(n-2)}.$$

3. Soit C l'évènement "obtenir le tiercé gagnant dans un ordre différent". Comme il y a $3! - 1 = 5$ manières d'ordonner le tiercé gagnant dans un ordre différent, $\text{card}(C) = 5$, d'où la probabilité cherchée est :

$$\mathbf{P}(C) = \frac{5}{n(n-1)(n-2)}.$$

Exercice 4.3.

L'univers Ω est l'ensemble des mains de 5 cartes possibles, c'est-à-dire l'ensemble des combinaisons de 5 éléments pris parmi 32, de sorte que, $\text{card}(\Omega) = \binom{32}{5}$. On suppose que toutes les mains sont équiprobables et on munit Ω de la probabilité uniforme, notée \mathbf{P} . Ainsi, si $A \in \mathcal{P}(\Omega)$ est un évènement, $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.

1. Soit A l'évènement "le joueur possède une seule paire", calculons $\text{card}(A)$.
 - Choix de la hauteur de la paire : $\binom{8}{1}$.
 - Choix de la couleur (trèfle, carreau, cœur, pique) de chacune des cartes de la paire : $\binom{4}{2}$.

- Choix des hauteurs des trois autres cartes : $\binom{7}{3}$.
- Choix de la couleur de chacune des trois autres cartes : 4^3 .

Ainsi, $\text{card}(A) = \binom{8}{1} \binom{4}{2} \binom{7}{3} 4^3$, et

$$\mathbf{P}(A) = \frac{\binom{8}{1} \binom{4}{2} \binom{7}{3} 4^3}{\binom{32}{5}}.$$

2. Soit B l'évènement "le joueur possède deux paires", calculons $\text{card}(B)$.

- Choix des hauteurs des deux paires : $\binom{8}{2}$.
- Choix de la couleur de chacune des cartes de chacune des paires : $\binom{4}{2}^2$.
- Choix de la hauteur de la dernière carte : $\binom{6}{1}$.
- Choix de la couleur de la dernière carte : 4.

$$\text{Ainsi, } \mathbf{P}(B) = \frac{\binom{8}{2} \binom{4}{2}^2 \binom{6}{1} 4}{\binom{32}{5}}.$$

3. Soit C l'évènement "le joueur possède un brelan".

- Choix de la hauteur du brelan : $\binom{8}{1}$.
- Choix de la couleur de chacune des cartes du brelan : $\binom{4}{3}$.
- Choix de la hauteur des deux cartes restantes : $\binom{7}{2}$.
- Choix de la couleur de chacune des deux cartes restantes : 4^2 .

$$\text{Ainsi, } \mathbf{P}(C) = \frac{\binom{8}{1} \binom{4}{3} \binom{7}{2} 4^2}{\binom{32}{5}}.$$

4. Soit D l'évènement "le joueur possède un carré".

- Choix de la hauteur du carré : $\binom{8}{1}$.
- Choix de la couleur de chacune des cartes du carré : $\binom{4}{4} = 1$.
- Choix de la hauteur de la carte restante : 7.
- Choix de la couleur de la carte restante : 4.

$$\text{Ainsi, } \mathbf{P}(D) = \frac{8 \cdot 7 \cdot 4}{\binom{32}{5}}.$$

5. Soit E l'évènement "le joueur possède un full".

- Choix de la hauteur de la paire : 8.
- Choix de la couleur de la paire : $\binom{4}{2}$.
- Choix de la hauteur du brelan : 7.
- Choix de la couleur du brelan : $\binom{4}{3}$.

$$\text{Ainsi, } \mathbf{P}(E) = \frac{8 \cdot 6 \cdot 7 \cdot 4}{\binom{32}{5}}.$$

Exercice 4.4.

L'univers Ω est l'ensemble des combinaisons de 6 numéros choisis parmi 49, de sorte que, $\text{card}(\Omega) = \binom{49}{6}$. On suppose que toutes ces combinaisons sont équiprobables et on munit Ω de la probabilité uniforme, notée \mathbf{P} . Ainsi, si $A \in \mathcal{P}(\Omega)$ est un évènement, $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$.

Pour $k \in \{1, 2, 3\}$, soit A_k l'évènement "la grille du joueur est une grille gagnante du k -ième rang".

1. A_1 est formé des grilles qui correspondent exactement au tirage, donc $\text{card}(A_1) = 1$ et : $p_1 = \frac{1}{\binom{49}{6}} \sim 7,15 \cdot 10^{-8}$.
2. A_2 est formé des grilles qui contiennent 5 des 6 numéros du tirage et exactement le numéro complémentaire, donc : $p_2 = \frac{\binom{6}{5}}{\binom{49}{6}} \sim 4,29 \cdot 10^{-7}$.
3. A_3 est formé des grilles qui contiennent 5 des 6 numéros du tirage et un numéro autre (que les 7 tirés), donc :

$$p_3 = \frac{\binom{6}{5} \binom{49-7}{1}}{\binom{49}{6}} = \frac{\binom{6}{5} 42}{\binom{49}{6}} \sim 1,8 \cdot 10^{-5}.$$

Exercice 4.5.

L'espace des états Ω est l'ensemble de toutes les manières de distribuer r boules dans n urnes. Pour la première boule, il y a n choix d'urnes, pour la deuxième également, etc., donc :

$$\text{card}(\Omega) = n^r.$$

Comme la distribution des boules est aléatoire, on munit Ω de la probabilité uniforme, de sorte que si A est un évènement de Ω , on a $\mathbf{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$. Pour tout $k \in \{0, \dots, r\}$, on introduit A_k l'évènement "une urne donnée contient k boules". Alors il y a $\binom{r}{k}$ choix possibles pour ces k boules, et comme aucune autre boule ne peut être dans l'urne donnée, chacune a le choix entre $(n-1)$ urnes. Ainsi,

$$\text{card}(A_k) = \binom{r}{k} (n-1)^{r-k},$$

et

$$\mathbf{P}(A_k) = \binom{r}{k} \frac{(n-1)^{r-k}}{n^r} = \binom{r}{k} \left(1 - \frac{1}{n}\right)^{r-k} \left(\frac{1}{n}\right)^k.$$

On retrouve un exemple de *loi binomiale*.

Remarque : En général dans des exercices où l'on tire ou on place des boules, il est naturel, afin de coller à l'expérience, de les supposer discernables. Ceci est à mettre en perspective avec la démonstration du "tirage non ordonné avec remise", qui se transpose en un problème où l'on place des boules indiscernables dans des urnes.

Exercice 4.6.

Comme l'application f est de classe \mathcal{C}^∞ , les dérivées partielles ne dépendent pas de l'ordre dans lequel elles sont prises, mais seulement du nombre de répétitions de chacune des variables. Ainsi ce problème revient à chercher les suites d'entiers naturels r_1, \dots, r_n , tels que $r_1 + \dots + r_n = r$, où r_i représente la puissance de $\frac{\partial f}{\partial x_i}$. On cherche donc le nombre de tirages non-ordonnés avec remise de r individus parmi n , soit :

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

Exercice 4.7.

On cherche l'ensemble des suites de 3 entiers naturels de somme 15. Il y a donc $\binom{17}{2}$ solutions.

Exercice 4.8.

1. À tout chemin possible on peut associer la succession de choix que l'on fait à chaque intersection

rencontrée. Ces choix peuvent être codés par “E” (est) et “N” (nord), ces deux directions étant les seules répondant aux conditions proposées. Le chemin dessiné devient par exemple,

$$EENNEENNEENNEN.$$

Il est évident que, vice-versa, la donnée de cette suite permet de connaître sans ambiguïté le chemin correspondant puisqu’à chaque intersection, on saura où se diriger.

2. Les chemins ayant le point de départ et d’arrivée donnés ont tous en commun qu’il faut aller 8 fois vers le nord et 7 fois vers l’est. Dans ce codage, on trouvera nécessairement 7 fois le “E” et 8 fois le “N”. Donc tout codage est une suite de 15 symboles composée de 7 “E” et 8 “N” et réciproquement. Il apparaît donc une bijection entre l’ensemble des chemins et l’ensemble des suites de 15 caractères décrites ci-dessus. Le nombre de solutions est alors :

$$\binom{15}{7} = \binom{15}{8}.$$

3. À chaque chemin on peut associer la longueur des parcours que l’on fait vers le nord dans chacune des avenues que l’on rencontre. Il y a 8 avenues possibles, ce qui donnera, quel que soit le chemin, une suite de 8 entiers naturels dont la somme est 8. Par exemple, pour le chemin dessiné, on a :

$$0, 0, 2, 0, 2, 0, 2, 2.$$

À partir de ce nouveau codage, on retrouve l’ancien en remplaçant k par :

$$\underbrace{N \cdots N}_k E,$$

Le dernier nombre est uniquement remplacé par la suite de k fois “N”. Dans l’exemple ci-dessus, on remplace 0 par “E” et 2 par “NNE” (sauf, pour le dernier 2 qui est remplacé par “NN”) et on retrouve le chemin codé en “N” et “E”. Nous avons donc explicité une bijection entre l’ensemble des suites de 8 entiers naturels de somme 8 et celui des chemins.

Remarque : Les questions 1 et 3 sont deux interprétations de l’expérience qui consiste à placer 8 boules indistinguables dans 8 urnes. Dans le premier codage les parois des urnes (sauf la première et la dernière) sont remplacées par des “E” et les boules par des “N”. Dans le deuxième codage, pour $i \in \{1, \dots, 8\}$, r_i représente le nombre de boules dans la i -ième urne. La distribution de boules qui correspond à la séquence “EENNEENNEENNEN” ou “0, 0, 2, 0, 2, 0, 2, 2” est :

$$||| **|| **|| **| **|$$

Exercice 4.9.

On partitionnera l’espace Ω en $\Omega = \{X_1 = 0\} \cup \{X_1 = 1\}$.

On trouve $\mathbf{ET} = \frac{1-p}{p} + \frac{p}{1-p}$ et $\mathbf{ES} = 2$.

Exercice 4.10.

On trouve respectivement 60, 90 et 2520 anagrammes pour ces trois mots.

Exercice 4.11.

Après calculs, on trouve :

1. Pour k entier, on a la relation $\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

2. On a l'expression $\mathbf{P}(Y = m|X = k) = 0$ si $m < k$ et $\binom{n-k}{m-k}q^{m-k}(1-q)^{n-m}$ sinon.
3. Y suit la loi binomiale de paramètres n et $p + q - pq$.

Exercice 4.12.

La probabilité de $\{X = k\}$ est donnée par

$$\mathbf{P}_N(X = k) = \frac{\binom{n}{k} \binom{N-n}{n-k}}{\binom{N}{k}}.$$

L'espérance de X vaut $\frac{n^2}{N}$.

La valeur de $\mathbf{P}_N(X = k)$ est maximale pour $N = n^2/k$ ou $N = n^2/k - 1$ si n^2/k est entier. Dans le cas contraire, la valeur maximale est atteinte si N est la partie entière de n^2/k . Cela peut se voir en comparant $\mathbf{P}_N(X = k)/\mathbf{P}_{N+1}(X = k)$ avec 1, de sorte à obtenir le sens de variation de la suite $(\mathbf{P}_N(X = k))_{N \in \mathbf{N}}$.

Exercice 4.14.

On trouve $\mathbf{P}(X + Y = n) = \sum_{k=0}^n \mathbf{P}(X = k)\mathbf{P}(Y = n - k)$ et $\mathbf{P}(T = n) = P(X = n)\mathbf{P}(Y = n) + \sum_{k=n+1}^{\infty} \mathbf{P}(X = n)\mathbf{P}(Y = k) + \mathbf{P}(X = k)\mathbf{P}(Y = n)$.

Exercice 4.15.

Si N suit la loi binomiale de paramètres n et q , les variables Z et $N - Z$ ne sont pas indépendantes, mais suivent des loi binomiales dont les paramètres sont respectivement (n, qp) et $(n, (1 - q)p)$.

Si N suit la loi de Poisson de paramètre θ , les variables Z et $N - Z$ sont deux variables indépendantes de loi de Poisson, de paramètres respectifs θp et $\theta(1 - p)$.

Exercice 4.16.

On trouve $\mathbf{P}(\max_{k=1}^n X_k = q) = \left(\frac{q}{N}\right)^n - \left(\frac{q-1}{N}\right)^n$, pour $1 \leq q \leq N$.

Exercice 4.17.

Considérons tout d'abord le jeu où l'on n'a qu'un lancer, noté X , qui peut être échangé contre une valeur y . La stratégie optimale est de ne garder la valeur du dé que si elle est supérieure à y . En effet, si on note A l'ensemble des résultats pour lesquels on choisit de garder la valeur du dé, la valeur obtenue est donnée par $R = X\mathbf{1}_{X \in A} + y\mathbf{1}_{X \notin A}$. On a alors

$$\mathbf{E}[R] = \mathbf{E}[X\mathbf{1}_{X \in A}] + \mathbf{E}[y\mathbf{1}_{X \notin A}] = \sum_{i \in A} \frac{i}{6} + \sum_{i \notin A} \frac{y}{6} = y + \frac{1}{6} \sum_{i=1}^6 (i - y)\mathbf{1}_{i \in A},$$

ce qui confirme que l'espérance de R est maximale si A contient exactement les valeurs supérieures à y .

Si on remplace y par une variable aléatoire Y indépendante de X , dont on note l'espérance par y , le choix optimal de A est encore l'ensemble des valeurs supérieures à y , puisque

$$\mathbf{E}[X\mathbf{1}_{X \in A} + Y\mathbf{1}_{X \notin A}] = \mathbf{E}[X\mathbf{1}_{X \in A}] + \mathbf{E}[Y]\mathbf{E}[\mathbf{1}_{X \notin A}] = \mathbf{E}[X\mathbf{1}_{X \in A} + y\mathbf{1}_{X \notin A}],$$

et on cherche à maximiser l'espérance d'une variable ayant la même forme que précédemment.

Calculons maintenant la stratégie optimale si on a le droit à deux lancers de dés : on lance un premier dé, que l'on relance si l'on obtient moins de 3.5 (la valeur moyenne d'un lancer de dé). Autrement dit, l'ensemble A vaut ici $\{4, 5, 6\}$, et le gain moyen est

$$\frac{1}{6}(3.5 + 3.5 + 3.5 + 4 + 5 + 6) = \frac{17}{4} = 4.25.$$

Pour le jeu avec trois lancers, on lance un premier dé, que l'on garde si le résultat est supérieur au résultat moyen du jeu à deux lancers, c'est-à-dire à 4.25. On a donc $A = \{5, 6\}$, et le gain moyen est

$$\frac{1}{6}(4.25 + 4.25 + 4.25 + 4.25 + 5 + 6) = 4.666\dots$$

Au final, la stratégie gagnante est :

- garder le premier dé si sa valeur est 5 ou 6 ;
- garder le deuxième dé si sa valeur est 4, 5 ou 6 ;

avec un gain moyen de 4.666....

Chapitre 5

Exercice 5.1.

Les densités sont données respectivement par

$$p_1(x) = p\left(\frac{x-b}{a}\right), \quad p_2(x) = \frac{p(\sqrt{x}) + p(-\sqrt{x})}{2\sqrt{x}} \mathbf{1}_{x>0}, \quad p_3(x) = \frac{p(\ln(x))}{x} \mathbf{1}_{x>0}.$$

Exercice 5.2.

X et $1/X$ ont pour fonction de répartition commune $F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}$. La variable $1/X$ suit donc elle aussi la loi de Cauchy.

Exercice 5.3.

Si $\alpha > 0$, la densité de U^α est

$$\frac{1}{\alpha} x^{1/\alpha-1} \mathbf{1}_{x \in]0,1[},$$

si $\alpha < 0$ la densité est

$$\frac{1}{\alpha} x^{1/\alpha-1} \mathbf{1}_{x \in]1,\infty[}.$$

U^α est intégrable si et seulement si $\alpha > -1$, et on a alors $\mathbf{E}[U^\alpha] = \frac{1}{\alpha+1}$. U^α est donc de carré intégrable si et seulement si $2\alpha > -1$, soit encore $\alpha > -1/2$, auquel cas sa variance est

$$\mathbf{V}(U^\alpha) = \mathbf{E}[U^{2\alpha}] - \mathbf{E}[U^\alpha]^2 = \frac{1}{2\alpha+1} - \frac{1}{(\alpha+1)^2}.$$

Exercice 5.4.

La variable $\sin(\Theta)$ admet pour densité la fonction $\frac{1}{\pi}(1-x^2)^{-1/2}$. Les variables $\sin(\Theta)$ et $|\sin(\Theta)|$ ont pour espérance respectives 0 et $2/\pi$. Enfin, $\sin(\Theta)$ a pour variance $1/2$.

Exercice 5.5.

Les variables $\min(X_1, X_2)$ et $\max(X_1, X_2)$ ont pour densité respectives $2(1-x)\mathbf{1}_{[0,1]}$ et $2x\mathbf{1}_{[0,1]}$. Plus généralement, Y_k a pour densité $k\binom{n}{k}x^{k-1}(1-x)^{n-k}$ et pour moyenne $\frac{k}{n+1}$.

Exercice 5.6.

La densité de $\max_{i=1}^n X_i$ est $np(x) \left(\int_{-\infty}^x p(y) dy \right)^{n-1}$.

Exercice 5.9.

X suit une loi uniforme sur $[0, 1]$.

Exercice 5.10.

La variable Z suit une loi de Cauchy.

Exercice 5.11.

La loi de T_1 conditionnellement à $\{N_t = 1\}$ est la loi uniforme sur $[0, t]$.

Chapitre 6

Exercice 6.1.

La constante $\sqrt{5/18\pi}$ apparaît en vertu du théorème limite central en tant qu'espérance de la variable $\sqrt{p(1-p)}|G|/\sqrt{n}$, G suit la loi normale centrée réduite, avec $p = 1/6$.

Exercice 6.2.

En considérant tous les cas, on obtient la matrice de transition suivante :

$$P = \begin{pmatrix} 5/54 & 25/36 & 125/648 & 25/1296 & 1/1296 \\ 0 & 5/9 & 10/27 & 5/72 & 1/216 \\ 0 & 0 & 25/36 & 5/18 & 1/36 \\ 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

La première ligne de la matrice de transition correspond au cas où l'on relance les 5 dés. Au premier tirage, la loi du nombre de dés identiques est donc donnée par cette première ligne, soit :

$$\mu_1 = (5/54 \quad 25/36 \quad 125/648 \quad 25/1296 \quad 1/1296).$$

Les lois respectives de X_2 et X_3 s'obtiennent en multipliant à droite par la matrice de transition, soit pour X_2

$$\mu_2 = \mu_1 P = (25/2916 \quad 875/1944 \quad 28625/69984 \quad 8375/69984 \quad 221/17496)$$

et pour X_3 :

$$\begin{aligned} \mu_3 = \mu_2 P &= (125/157464 \quad 26875/104976 \quad 3419375/7558272 \quad 115625/472392 \quad 347897/7558272) \\ &\simeq (0.00079 \quad 0.25601 \quad 0.45240 \quad 0.24476 \quad 0.04603). \end{aligned}$$

On trouve que la probabilité de faire un Yam's est de $347897/7558272$, soit approximativement 0.04603.

Exercice 6.3.

La matrice de transition est donnée par $P_{ij} = A_{ij} \times (\sum_{k=1}^n A_{ik})^{-1}$.

Exercice 6.4.

La matrice de transition est donnée par

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & \dots & \dots & 0 \\ \frac{1}{n} & 0 & \frac{n-1}{n} & 0 & & & \\ 0 & \frac{2}{n} & 0 & \frac{n-2}{n} & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \frac{n-2}{n} & 0 & \frac{2}{n} & 0 \\ 0 & \dots & \dots & \dots & 0 & 1 & 0 \end{pmatrix}.$$

Pour retrouver la mesure invariante par l'exercice précédent, on peut considérer le graphe $\{0, 1\}^N$ pour lequel deux n -uplets (x_1, \dots, x_n) et (y_1, \dots, y_n) sont reliés si tous leurs coefficients sont égaux sauf un.

Chapitre 7

Exercice 7.1.

3. On remarquera que $\hat{\sigma}_n^2$ peut se récrire $\hat{\sigma}_n^2 = \frac{1}{n} (\sum_{k=1}^n x_k^2) - \hat{m}_n^2$. En revanche, $\mathbf{E}(\hat{m}_n^2) = \frac{1}{n} \mathbf{E}(X_1^2) + \frac{n-1}{n} (\mathbf{E}X_1)^2$, ce qui fait que $\mathbf{E}\hat{\sigma}_n^2 = \frac{n-1}{n} \mathbf{V}(X_1)$.

Chapitre 8

Exercice 8.1.

L'erreur de mesure du radar suit la même loi que σG , avec G de loi normale centrée réduite. Par conséquent, la probabilité de verbaliser un automobiliste roulant à vitesse v est $\mathbf{P}(v + \sigma G > 95)$. Ceci est une fonction croissante de v , par conséquent, la probabilité de verbaliser un automobiliste roulant à moins de 90 km.h^{-1} est inférieure à $\mathbf{P}(90 + \sigma G > 95) = \mathbf{P}(G > (95 - 90)/\sigma)$.

Comme le quantile d'ordre 99% de la loi normale est de 2.326 (à lire dans une table ou bien à demander à l'ordinateur), c'est-à-dire que $\mathbf{P}(G > 2.33) \simeq 0.99$, on doit donc choisir $\sigma \leq 2.150 \text{ km.h}^{-1}$.