Probabilités et Applications

Benjamin JOURDAIN

28 septembre 2004

Table des matières

1 Introduction : probabilité sur un espace fini **1**

1.1 Probabilité sur un espace fini, événements . . . 1

1.1.1 Définitions . . . 1

1.1.2 Probabilités uniformes . . . 4

1.2 Probabilité conditionnelle et indépendance . . . 5

1.2.1 Probabilité conditionnelle . . . 5

1.2.2 Indépendance . . . 7

1.3 Exercices . . . 7

1.4 Résumé . . . 9

2 Variables aléatoires discrètes **11**

2.1 Espace de probabilité . . . 11

2.2 Variables aléatoires discrètes . . . 12

2.2.1 Rappel sur les manipulations de séries : le théorème de Fubini . . . 12

2.2.2 Définition . . . 12

2.2.3 Indépendance . . . 13

2.2.4 Lois discrètes usuelles . . . 14

2.2.5 Loi marginale . . . 16

2.3 Espérance et variance . . . 17

2.3.1 Espérance . . . 17

2.3.2 Variance . . . 20

2.4 Loi conditionnelle . . . 22

2.5 Exercices . . . 23

2.6 Résumé . . . 26

3 Variables aléatoires à densité **27**

3.1 Manipulation d'intégrales multiples . . . 27

3.1.1 Théorème de Fubini . . . 27

3.1.2 Changement de variables . . . 28

3.2 Variables aléatoires réelles à densité . . . 30

3.2.1 Définition . . . 30

3.2.2 Densités réelles usuelles . . . 31

3.2.3 Espérance, variance . . . 31

3.2.4 Fonction de répartition . . . 32
3.3 Vecteurs aléatoires à densité . . . 33
3.3.1 Définition . . . 33
3.3.2 Densité marginale . . . 33
3.3.3 Changement de variables . . . 34
3.3.4 Indépendance . . . 35
3.3.5 Covariance . . . 35
3.4 Lois béta, gamma et du khi 2 . . . 37
3.5 Exercices . . . 39
3.6 Résumé . . . 42

4 Simulation 43
4.1 Simulation de variables aléatoires discrètes . . . 43
4.1.1 Loi de Bernoulli de paramètre $p \in [0,1]$ . . . 43
4.1.2 Loi binômiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$ . . . 44
4.1.3 Loi géométrique de paramètre $p \in]0,1]$ . . . 44
4.1.4 Simulation suivant une loi discrète quelconque . . . 44
4.2 Simulation de variables aléatoires à densité . . . 45
4.2.1 Loi uniforme sur $[a,b]$ avec $a < b \in \mathbb{R}$ . . . 45
4.2.2 Méthode d'inversion de la fonction de répartition . . . 45
4.2.3 Méthode polaire pour la loi normale centrée réduite . . . 46
4.2.4 Méthode du rejet . . . 47
4.3 Exercices . . . 49
4.4 Résumé . . . 51

5 Convergence et théorèmes limites 53
5.1 Convergence . . . 53
5.2 Lois des grands nombres . . . 55
5.2.1 Loi faible des grands nombres . . . 55
5.2.2 Loi forte des grands nombres . . . 55
5.3 Fonction caractéristique et convergence en loi . . . 57
5.3.1 Fonction caractéristique . . . 57
5.3.2 Convergence en loi . . . 59
5.4 Le théorème de la limite centrale . . . 61
5.4.1 Enoncé et preuve du résultat . . . 61
5.4.2 Intervalle de confiance dans la méthode de Monte-Carlo . . . 62
5.5 Exercices . . . 63
5.6 Résumé . . . 66

6 Vecteurs gaussiens 67
6.1 Définition, construction . . . 67
6.1.1 Définition . . . 67

6.1.2 Stabilité du caractère gaussien par transformation linéaire 68
6.1.3 Construction d'un vecteur gaussien de loi $\mathcal{N}_n(m, \Gamma)$ 68
6.2 Propriétés des vecteurs gaussiens 69
6.2.1 Vecteurs gaussiens et indépendance 69
6.2.2 Stabilité par convergence en loi 70
6.3 Exercices 70
6.4 Résumé 72

7 Quelques exercices corrigés 73
7.1 Énoncés 73
7.1.1 Variables aléatoires discrètes 73
7.1.2 Variables aléatoires à densité 75
7.1.3 Convergence des variables aléatoires 76
7.2 Corrigés 78
7.2.1 Variables aléatoires discrètes 78
7.2.2 Variables aléatoires à densité 85
7.2.3 Convergence des variables aléatoires 91

Chapitre 1

Introduction : probabilité sur un espace fini

Historiquement, le calcul des probabilités s'est développé à partir du XVIIème siècle autour des problèmes de jeux dans des situations où le nombre de cas possibles est fini. Les développements plus récents concernant des espaces non nécessairement finis nécessitent les outils techniques de la théorie de la mesure. Mais on peut introduire simplement sur les espaces finis toutes les notions importantes de probabilités sans avoir besoin de cet outillage.

1.1 Probabilité sur un espace fini, événements

1.1.1 Définitions

On s'intéresse à une expérience aléatoire qui conduit à la réalisation d'un seul résultat parmi un nombre fini de résultats possibles $\omega_1, \omega_2, \ldots, \omega_n$. On note $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ l'ensemble de ces résultats.

Exemple 1.1.1.
- Jet d'une pièce à pile où face : $\Omega = \{P, F\}$.
- Jet d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Si on mesure la fréquence d'apparition du résultat ω_k au cours d'un grand nombre de répétitions de l'expérience i.e. on calcule le rapport $F_k = \frac{N_k}{N}$ du nombre N_k d'expériences dont le résultat est ω_k sur le nombre total d'expériences N, on constate qu'elle fluctue de moins en moins. La limite $p_k \geq 0$ de F_k lorsque $N \to +\infty$ correspond à la notion intuitive de probabilité.

On appelle événement une partie A de Ω. La fréquence de A c'est-à-dire la proportion d'expériences dont le résultat est dans A est égale à $\sum_{k:\omega_k \in A} F_k$. On est donc amené à associer la probabilité $\sum_{k:\omega_k \in A} p_k$ à l'événement A.
Comme la fréquence de Ω vaut 1, en passant à la limite, on obtient $\sum_{k=1}^{n} p_k = 1$.

Définition 1.1.2. *Une probabilité $\mathbb{P}$ sur un ensemble fini $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ est une pondération $p_1, p_2, \ldots, p_n$ des éléments de cet ensemble t.q.*

$$\forall 1 \leq k \leq n,\ p_k \geq 0 \quad \textit{et} \quad \sum_{k=1}^{n} p_k = 1.$$

On attribue à tout événement $A \subset \Omega$ le nombre

$$\mathbb{P}(A) = \sum_{k:\omega_k \in A} p_k$$

qui est appelé probabilité de l'événement A.

Exemple 1.1.3. Jet de deux dés à six faces : $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$ où i désigne la valeur de la face supérieure du premier dé et j celle du second.
Pour des raisons de symétrie (si les dés ne sont pas pipés), on munit Ω de la pondération suivante :

$$\forall 1 \leq i,j \leq 6,\ p_{(i,j)} = \frac{1}{36}.$$

Soit A l'événement : les valeurs des deux dés sont identiques.

$$A = \{(1,1),(2,2),\ldots,(6,6)\} \quad \text{et} \quad \mathbb{P}(A) = \sum_{i=1}^{6} p_{(i,i)} = \frac{6}{36} = \frac{1}{6}.$$

On note S la somme des deux dés et $\{S = k\}$ l'événement $\{(i,j) : S(i,j) = k\}$. On a $S(i,j) = i + j$. Donc

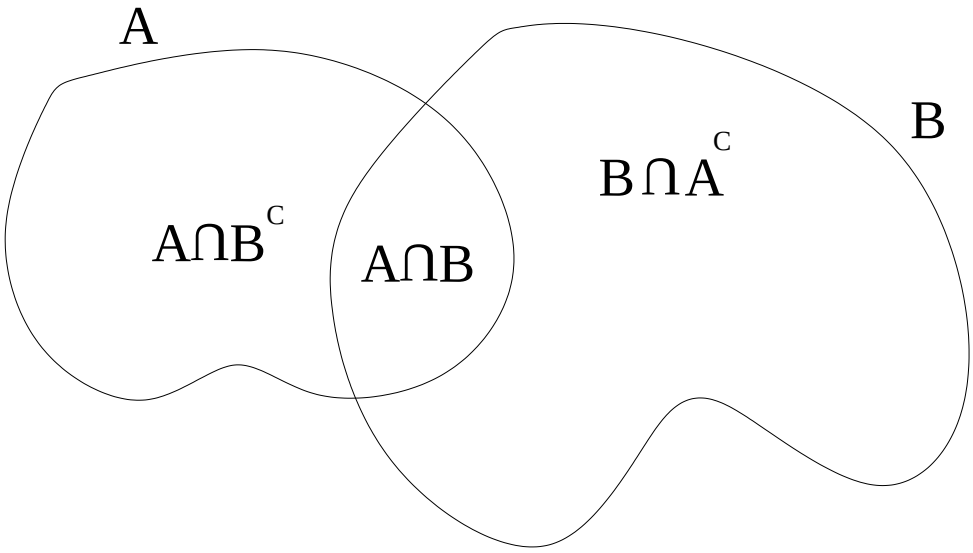
$$\begin{array}{lll}
\{S=2\} & = \{(1,1)\} & \mathbb{P}(S=2) = 1/36 \\
\{S=3\} & = \{(1,2),(2,1)\} & \mathbb{P}(S=3) = 1/18 \\
\{S=4\} & = \{(1,3),(2,2),(3,1)\} & \mathbb{P}(S=4) = 1/12 \\
\{S=5\} & = \{(1,4),(2,3),(3,2),(4,1)\} & \mathbb{P}(S=5) = 1/9 \\
\{S=6\} & = \{(1,5),(2,4),(3,3),(4,2),(5,1)\} & \mathbb{P}(S=6) = 5/36 \\
\{S=7\} & = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\} & \mathbb{P}(S=7) = 1/6 \\
\{S=8\} & = \{(2,6),(3,5),(4,4),(5,3),(6,2)\} & \mathbb{P}(S=8) = 5/36 \\
\{S=9\} & = \{(3,6),(4,5),(5,4),(6,3)\} & \mathbb{P}(S=9) = 1/9 \\
\{S=10\} & = \{(4,6),(5,5),(6,4)\} & \mathbb{P}(S=10) = 1/12 \\
\{S=11\} & = \{(5,6),(6,5)\} & \mathbb{P}(S=11) = 1/18 \\
\{S=12\} & = \{(6,6)\} & \mathbb{P}(S=12) = 1/36
\end{array}$$

Terminologie concernant les événements :

- Si $\mathbb{P}(A) = 0$, l'événement A est dit négligeable.
- Si $\mathbb{P}(A) = 1$, il est dit presque sûr.
- On appelle événement contraire de A et on note A^c l'événement $\Omega \setminus A$.
- Si $A, B \subset \Omega$, l'événement A et B (réalisé lorsque A et B le sont) est noté $A \cap B$.
- L'événement A ou B (réalisé lorsque A ou B le sont) est noté $A \cup B$.

Probabilité de l'événement $A \cup B$:

Par définition, $\mathbb{P}(A \cup B) = \sum_{k:\omega_k \in A \cup B} p_k$.



Comme $A \cup B$ est égal à l'union disjointe $(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$,

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \sum_{k:\omega_k \in A \cap B^c} p_k + \sum_{k:\omega_k \in A \cap B} p_k + \sum_{k:\omega_k \in A^c \cap B} p_k \\
&= \left(\sum_{k:\omega_k \in A \cap B^c} p_k + \sum_{k:\omega_k \in A \cap B} p_k \right) + \left(\sum_{k:\omega_k \in A^c \cap B} p_k + \sum_{k:\omega_k \in A \cap B} p_k \right) - \sum_{k:\omega_k \in A \cap B} p_k \\
&= \sum_{k:\omega_k \in A} p_k + \sum_{k:\omega_k \in B} p_k - \sum_{k:\omega_k \in A \cap B} p_k \\
&= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).
\end{aligned}$$

Ainsi

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

Fonction indicatrice :

On appelle fonction indicatrice de l'événement A la fonction $1_A : \Omega \to \{0, 1\}$ définie par

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

Exercice 1.1.4. Quel est l'événement $\{\omega : 1_A(\omega) \times 1_B(\omega) = 1\}$ que l'on note aussi de façon condensée $\{1_A \times 1_B = 1\}$?
Conclure que

$$\boxed{1_{A \cap B} = 1_A \times 1_B.}$$

Montrer également que

$$\boxed{1_{A^c} = 1 - 1_A \quad \text{et} \quad 1_{A \cup B} = 1_A + 1_B - 1_{A \cap B}.}$$

1.1.2 Probabilités uniformes

Dans le cas où les symétries font que tous les résultats possibles $\omega_1, \omega_2, \dots \omega_n$ jouent le même rôle, ces résultats doivent avoir la même pondération $1/\text{Card}\ (\Omega)$. On dit alors qu'il sont équiprobables.
On a alors pour tout événement $A \subset \Omega$,

$$\boxed{\mathbb{P}(A) = \sum_{k:\omega_k \in A} \frac{1}{\text{Card}\ (\Omega)} = \frac{\text{Card}\ (A)}{\text{Card}\ (\Omega)}.}$$

Cette probabilité $\mathbb{P}$ s'appelle *probabilité uniforme* sur Ω.

Exemple 1.1.5. Dans le cas du jet de deux dés non pipés, $\Omega = \{(i,j) :\ 1 \leq i, j \leq 6\}$ est muni de la probabilité uniforme.

Remarque 1.1.6. Si on s'intéresse à la somme des deux dés, on peut choisir $\Omega = \{2, 3, 4 \dots, 12\}$, ensemble des valeurs prises par cette somme. Mais faute de propriétés de symétrie, on ne sait pas munir cet espace d'une probabilité naturelle.
Dans l'exemple 1.1.3, en travaillant sur l'espace plus gros $\{(i,j) :\ 1 \leq i, j \leq 6\}$ des couples des valeurs des deux dés muni de la probabilité uniforme, nous avons pu construire la pondération naturelle sur les valeurs de la somme des deux dés. Cette pondération n'a rien d'uniforme.
Cet exemple permet de bien comprendre l'importance du choix de l'espace de probabilité sur lequel on travaille.

Dans le cas des probabilités uniformes, les calculs se ramènent à du dénombrement.

Rappels de dénombrement

On se donne $n, p \in \mathbb{N}^*$ avec $p \leq n$.
- Le nombre de permutations d'un ensemble à n éléments est $n!$.
- De façon plus générale, le nombre d'injections d'un ensemble à p éléments dans un ensemble à n éléments est

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

  Le facteur n (resp. $n-1$,..., resp. $n-p+1$) vient du choix de l'image du 1er (resp 2nd,..., $(n-p)^{\text{ème}}$) élément.
- Le nombre de parties à p éléments d'un ensemble à n éléments est

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

Exercice résolu 1.1.7. *Dans une classe de $n \leq 365$ élèves, quelle est la probabilité de l'événement : "2 élèves au moins sont nés le même jour" que l'on note A ?*

On choisit comme espace de probabilité $\Omega = \{f : [1,n] \to [1,365]\}$ où pour $1 \leq i \leq n$, $f(i)$ représente le jour d'anniversaire du $i$ème élève dans l'ordre alphabétique.
Même si les naissances ne sont pas vraiment équiréparties au long de l'année, on munit Ω

de la probabilité uniforme. On a Card $(\Omega) = 365^n$.
Pour calculer la probabilité de A, on peut calculer la probabilité de l'événement contraire A^c : "tous les élèves ont des dates d'anniversaire différentes". En effet comme $A \cup A^c = \Omega$ et $A \cap A^c = \emptyset$,

$$\mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) - \mathbb{P}(A \cap A^c) \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c).$$

On a $A^c = \{f : [1, n] \to [1, 365] \text{ injective}\}$. Donc Card $(A^c) = A^n_{365}$ et

$$\mathbb{P}(A^c) = \frac{\text{Card }(A^c)}{\text{Card }(\Omega)} = \frac{365!}{(365-n)!365^n} = \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365-n+1}{365},$$

et

$$\mathbb{P}(A) = 1 - \frac{365}{365} \times \frac{364}{365} \times \ldots \times \frac{365-n+1}{365}.$$

On peut vérifier que dès que $n \geq 23$, cette probabilité est supérieure à $1/2$.

1.2 Probabilité conditionnelle et indépendance

1.2.1 Probabilité conditionnelle

La notion de probabilité conditionnelle permet de prendre en compte l'information dont on dispose (à savoir qu'un événement B est réalisé) pour actualiser la probabilité que l'on donne à un événement A :

Définition 1.2.1. *Soit Ω muni d'une probabilité $\mathbb{P}$ et $A, B \subset \Omega$. La probabilité conditionnelle de l'événement A sachant l'événement B est notée $\mathbb{P}(A|B)$ et définie par*

$$\boxed{\mathbb{P}(A|B) = \begin{cases} \mathbb{P}(A \cap B)/\mathbb{P}(B) & \textit{si } \mathbb{P}(B) > 0 \\ \mathbb{P}(A) & \textit{sinon.} \end{cases}}$$

Exercice résolu 1.2.2. 1. *Dans une famille qui comporte deux enfants, l'un est une fille. On cherche la probabilité que l'autre soit un garçon.*
On choisit $\Omega = \{FF, FG, GF, GG\}$ où par exemple FG signifie que l'aîné des enfants est une fille et le second un garçon.
Cet espace est muni de la probabilité uniforme. On note

$$A = \{\text{un des enfants est un garçon}\} = \{FG, GF, GG\}$$
$$B = \{\text{un des enfants est une fille}\} = \{FF, FG, GF\}.$$

On a $\mathbb{P}(B) = \frac{\text{Card }(B)}{\text{Card }(\Omega)} = \frac{3}{4}$. Comme $A \cap B = \{FG, GF\}$, $\mathbb{P}(A \cap B) = \frac{\text{Card }(A \cap B)}{\text{Card }(\Omega)} = \frac{1}{2}$.
Donc la probabilité recherchée est

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

2. *On suppose maintenant que l'aîné des enfants est une fille. On veut alors connaître la probabilité pour que l'autre soit un garçon.*
En reprenant la démarche ci-dessus, on obtient que cette probabilité vaut $1/2$.

Dans certains problèmes, ce sont les probabilités conditionnelles que l'on connaît naturellement et on est amené à utiliser la définition sous la forme

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

qui se généralise en

$$\boxed{\mathbb{P}(A_1 \cap A_2 \cap .. \cap A_m) = \mathbb{P}(A_m|A_1 \cap .. \cap A_{m-1})\mathbb{P}(A_{m-1}|A_1 \cap .. \cap A_{m-2}) \dots \mathbb{P}(A_2|A_1)\mathbb{P}(A_1)}$$

pour m événements $A_1, \dots, A_m$.

Exercice résolu 1.2.3. *Parmi 10 pièces mécaniques, 4 sont défectueuses. On prend successivement deux pièces au hasard dans le lot (sans remise). Quelle est la probabilité pour que les deux pièces soient correctes.*

On note A_1 l'événement la première pièce est bonne et A_2 l'événement la seconde pièce est bonne.
Comme, au départ, il y a 6 pièces bonnes sur 10, $\mathbb{P}(A_1) = 6/10 = 3/5$. Lorsque l'on a retiré une pièce bonne, il reste 5 pièces bonnes sur 9. D'où $\mathbb{P}(A_2|A_1) = 5/9$. On conclut que la probabilité cherchée est

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) = \frac{5}{9} \times \frac{3}{5} = \frac{1}{3}.$$

Il est à noter et c'est rassurant que l'on peut retrouver ce résultat en posant $\Omega = \{\text{sous-ensembles comportant 2 pièces de l'ensemble des 10 pièces}\}$ que l'on munit de la probabilité uniforme. L'événement A dont on cherche la probabilité est $\{\text{sous-ensembles comportant 2 pièces de l'ensemble des 6 pièces correctes}\}$. On a alors

$$\mathbb{P}(A) = \frac{\text{Card }(A)}{\text{Card }(\Omega)} = \frac{C_6^2}{C_{10}^2} = \frac{6!\ 8!\ 2!}{10!\ 4!\ 2!} = \frac{6 \times 5}{10 \times 9} = \frac{1}{3}.$$

Enfin le résultat suivant qui porte le nom de formule de Bayes est souvent utile

Proposition 1.2.4. *Soit $B_1, \dots, B_m$ une partition de Ω (i.e. des sous-ensembles disjoints de Ω dont la réunion est Ω) et $A \subset \Omega$ t.q. $\mathbb{P}(A) > 0$. Alors pour tout $1 \leq i \leq n$,*

$$\boxed{\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^m \mathbb{P}(A|B_j)\mathbb{P}(B_j)}.}$$

Démonstration : Le numérateur du second membre est égal à $\mathbb{P}(A \cap B_i)$. Le dénominateur vaut $\sum_{j=1}^m \mathbb{P}(A \cap B_j)$ et comme les B_j forment une partition de Ω il est égal à $\mathbb{P}(A)$. Donc le second membre est bien égal à $\mathbb{P}(A \cap B_i)/\mathbb{P}(A)$. □

Exercice 1.2.5. Pour dépister une maladie, on applique un test sanguin. Si le patient est atteint, le test donne un résultat positif dans 99% des cas. Mais il se peut aussi que le résultat du test soit positif alors que le patient est en bonne santé et cela se produit dans 2% des cas. La proportion de personnes malades dans la population soumise au test est de 10^{-3}. Calculer la probabilité pour qu'un patient soit en bonne santé sachant que le résultat de son test est positif.

1.2.2 Indépendance

Définition 1.2.6. *Soit Ω muni d'une probabilité $\mathbb{P}$. Deux événements A et B sont dits indépendants si*

$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).}$$

Remarque 1.2.7. L'indépendance de A et B se caractérise aussi par les relations $\mathbb{P}(A|B) = \mathbb{P}(A)$ ou $\mathbb{P}(B|A) = \mathbb{P}(B)$, c'est-à-dire que la probabilité donnée à l'événement A (resp. B) n'est pas modifiée par l'information que l'événement B (resp. A) est réalisé.

Définition 1.2.8. *m évènements $A_1, \ldots, A_m$ sont dits indépendants si*

$$\forall I \subset [1, m],\ \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Attention

- Il ne suffit pas que $\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_m) = \prod_{i=1}^{m} \mathbb{P}(A_i)$ pour que les événements soient indépendants.
- Pour que 3 événements soient indépendants, il ne suffit pas qu'il soient 2 à 2 indépendants.

Exemple 1.2.9. Jet de deux pièces à Pile ou Face : $\Omega = \{PP, PF, FP, FF\}$ où par exemple PF signifie que la première pièce donne Pile et la seconde Face. Cet espace est muni de la probabilité uniforme.
On note A l'événement "la première pièce donne Pile", B l'événement "la seconde pièce donne Face" et C l'événement "les deux pièces donnent le même résultat".

$$\begin{array}{rlrll}
A & = \{PP, PF\} & \mathbb{P}(A) & = 1/2 & \\
B & = \{PF, FF\} & \mathbb{P}(B) & = 1/2 & \\
C & = \{PP, FF\} & \mathbb{P}(C) & = 1/2 & \\
A \cap B & = \{PF\} & \mathbb{P}(A \cap B) & = 1/4 & = \mathbb{P}(A)\mathbb{P}(B) \\
A \cap C & = \{PP\} & \mathbb{P}(A \cap C) & = 1/4 & = \mathbb{P}(A)\mathbb{P}(C) \\
B \cap C & = \{FF\} & \mathbb{P}(B \cap C) & = 1/4 & = \mathbb{P}(B)\mathbb{P}(C) \\
A \cap B \cap C & = \emptyset & \mathbb{P}(A \cap B \cap C) & = 0 & \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).
\end{array}$$

Ainsi les événements A, B et C sont 2 à 2 indépendants mais pas indépendants.

1.3 Exercices

Exercice 1.3.1. Deux événements A et B disjoints ($A \cap B = \emptyset$) et de probabilités non nulles peuvent-ils être indépendants ?

Exercice 1.3.2. On tire successivement et sans remise 4 lettres du mot "ATTACHANT". Quelle est la probabilité d'obtenir "CHAT" ?

Exercice 1.3.3. Une personne rentre chez elle après une soirée un peu trop arrosée. Elle ne sait plus laquelle des n clés qui se trouvent dans sa poche ouvre la porte de son domicile. Elle essaie donc les clés successivement. Déterminer la probabilité pour que la k-ième clé soit la bonne ($1 \leq k \leq n$).

Exercice 1.3.4. On note T la valeur obtenue lors du jet d'un dé à douze faces (numérotées de 1 à 12) et S le total obtenu lors du jet de deux dés à six faces (numérotées de 1 à 6). Calculez $\mathbb{P}(S > T)$, $\mathbb{P}(S = T)$ et $\mathbb{P}(S < T)$. Dans un jeu à deux joueurs où celui qui obtient le plus grand total gagne, on vous propose de choisir entre le dé à douze faces et les deux dés à six faces. Quel est votre choix ?

Exercice 1.3.5. 1. Soient $A_1, A_2, ..., A_n$ des événements. Montrer la formule du crible (ou formule de Poincaré)

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

On pourra commencer par établir que

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}(A_n) + \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) - \mathbb{P}\left(\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right).$$

2. Un service à thé comporte n ensembles "tasse + sous-tasse" de couleurs différentes. A la sortie du lave-vaisselle, on pose au hasard les tasses sur les sous-tasses. On numérote les couleurs de 1 à n et pour $1 \leq i \leq n$, on note A_i l'événement : "la tasse de couleur i est sur la sous-tasse de même couleur. Soit N le nombre d'ensembles "tasses+sous-tasses" assortis.
   (a) Pour $1 \leq k \leq n$, et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, calculer $\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})$.
   (b) En déduire $\mathbb{P}(N = 0)$.
   (c) A l'aide du résultat précédent, montrer que

$$\mathbb{P}(\{N = k\} \cap A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!} \left(1 - \sum_{l=1}^{n-k} \frac{(-1)^{l+1}}{l!}\right).$$

   (d) Pour $1 \leq k \leq n$, donner $\mathbb{P}(N = k)$.

Exercice 1.3.6. Trois chasseurs tirent en même temps sur un éléphant lors d'un safari. La bête meurt atteinte par deux balles. On estime la valeur d'un chasseur par sa probabilité d'atteindre la cible en un coup. Ces probabilités sont respectivement $\frac{1}{4}$, $\frac{1}{2}$ et $\frac{3}{4}$. Trouver pour chacun des chasseurs, la probabilité d'avoir raté l'éléphant.

Exercice 1.3.7. On dispose d'une carte comportant 2 faces rouges, d'une carte comportant une face rouge et une face blanche et d'une carte comportant 2 faces blanches. Une des trois cartes est tirée au sort et une des faces de cette carte (également choisie au hasard) est exposée. Cette face est rouge. On vous demande de parier sur la couleur de la face cachée. Choisissez-vous rouge ou blanc ?

Exercice 1.3.8. La couleur des yeux d'une personne est déterminée par 2 gènes dont l'un est transmis par le père et l'autre par la mère. On suppose qu'il y a deux formes possibles pour les gènes : la forme B (bleue) et la forme M (marron). La forme B est récessive c'est à dire qu'une personne qui a le génotype BM ou MB a les yeux marrons. Vos parents ont tous deux les yeux marrons mais votre sœur a les yeux bleus. Quelle est la probabilité pour que vous ayez les yeux marrons ? On suppose en plus que votre conjoint a les yeux bleus tandis que votre premier enfant a les yeux marrons. Quelle est la probabilité pour que votre second enfant ait les yeux bleus ?

1.4 Résumé

- Soit $A, B \subset \Omega$.
$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$
Pour $B = A^c$, cette égalité se récrit
$$\boxed{\mathbb{P}(A) = 1 - \mathbb{P}(A^c).}$$

- Pour $A, B \subset \Omega$, on appelle probabilité conditionnelle de l'événement A sachant l'événement B le nombre
$$\boxed{\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.}$$

- Les événements $A_1, \ldots, A_m$ sont dits indépendants si
$$\forall I \subset [1, m],\ \mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$
En particulier les événements A et B sont dits indépendants si
$$\boxed{\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).}$$

Chapitre 2

Variables aléatoires discrètes

2.1 Espace de probabilité

Dans le cas d'un espace Ω fini, nous avons défini un événement comme une partie quelconque de Ω. Mais si on souhaite modéliser le temps de première obtention de Pile dans une suite de jets d'une pièce à Pile ou Face, on choisit naturellement $\Omega = \{P, F\}^{\mathbb{N}^*}$, enesemble qui est infini. Dès que Ω est infini, pour pouvoir construire une probabilité qui satisfasse des propriétés intuitives, il est souvent utile de restreindre les événements que l'on considère à une sous-classe de $\mathcal{P}(\Omega)$ appelée *tribu*.
Il est naturel de demander que si A et B sont dans cette sous-classe alors A^c, $A \cup B$ et $A \cap B$ le sont aussi. Dans la définition d'une tribu, on demande même la stabilité par union ou intersection dénombrable (un ensemble est dénombrable s'il existe une surjection de $\mathbb{N}$ sur cet ensemble) :

Définition 2.1.1. *Une tribu $\mathcal{A}$ sur Ω est une classe de parties de Ω qui vérifie les trois propriétés suivantes :*

i) $\emptyset, \Omega \in \mathcal{A}$.

ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.

iii) *Si $(A_i)_{i \in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ alors $\bigcup_{i \in I} A_i$ et $\bigcap_{i \in I} A_i$ sont dans $\mathcal{A}$.*

On appelle événements les éléments de $\mathcal{A}$.

Exemple 2.1.2.
- $\{\emptyset, \Omega\}$ est la plus petite tribu sur Ω. On l'appelle *tribu grossière*.
- $\mathcal{P}(\Omega)$ est la plus grosse tribu sur Ω. On l'appelle *tribu discrète*.
- Si $A \subset \Omega$, $\{\emptyset, A, A^c, \Omega\}$ est une tribu sur Ω.

Définition 2.1.3. *Soit Ω muni d'une tribu $\mathcal{A}$. On appelle probabilité sur $(\Omega, \mathcal{A})$ une application $\mathbb{P} : \mathcal{A} \to [0, 1]$ qui vérifie*

i) $\mathbb{P}(\Omega) = 1$.

ii) *Si $(A_i)_{i \in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints ($\forall i \neq j \in I,\ A_i \cap A_j = \emptyset$) alors*

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

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ *s'appelle un espace de probabilité, la propriété ii) la* σ*-additivité.*

Dans toute la suite de ce cours, nous travaillerons sans le spécifier nécessairement systématiquement sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Nous pourrons également considérer sans inquiétude que toutes les parties de Ω qui nous intéressent sont des événements (i.e. sont dans $\mathcal{A}$). En effet, s'il est difficile de construire une probabilité sur Ω muni de la tribu discrète $\mathcal{P}(\Omega)$ dès que Ω n'est pas dénombrable, en revanche on peut le faire sur Ω muni d'une tribu suffisamment grosse pour qu'elle contienne tous les sous-ensembles de Ω définis sans l'aide de l'axiome du choix.
Enfin, en utilisant la propriété de σ-additivité ii), on peut facilement vérifier que l'on a toujours pour tout couple d'événements A et B :

$$\boxed{\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).}$$

Les définitions de la probabilité conditionnelle de A sachant B et celle de l'indépendance des événements données au chapitre précédent restent également valables. Comme nous pouvons maintenant être amenés à considérer des familles infinies d'événements, nous préciserons simplement qu'une telle famille est indépendante si toute sous-famille finie l'est.

2.2 Variables aléatoires discrètes

2.2.1 Rappel sur les manipulations de séries : le théorème de Fubini

Théorème 2.2.1. *Soit* F *et* G *deux ensembles dénombrables.*

- *Si* $(a(x,y))_{(x,y)\in F\times G}$ *est une famille de nombres* **positifs**, *alors*

$$\boxed{\sum_{(x,y)\in F\times G} a(x,y) = \sum_{x\in F}\left(\sum_{y\in G} a(x,y)\right) = \sum_{y\in G}\left(\sum_{x\in F} a(x,y)\right).}$$

  Cela signifie que les trois membres sont soit simultanément finis et égaux soit simultanément égaux à $+\infty$.
- *Si* $(a(x,y))_{(x,y)\in F\times G}$ *est sommable au sens où* $\sum_{(x,y)\in F\times G} |a(x,y)| < +\infty$, *alors l'égalité précédente reste vraie.*

2.2.2 Définition

Définition 2.2.2. *On appelle variable aléatoire discrète une application* $X : \Omega \to F$ *où* F *est un ensemble dénombrable. Pour* $x \in F$, *on note de façon concise* $\{X = x\}$ *l'événement* $\{\omega \in \Omega : \ X(\omega) = x\}$. *La famille des nombres* $(\mathbb{P}(X = x))_{x\in F}$ *s'appelle la loi de* X.

Remarque 2.2.3.
- En toute rigueur, pour pouvoir considérer $(\mathbb{P}(X = x))_{x\in F}$, il faut imposer dans la définition que $\forall x \in F$, $\{X = x\} \in \mathcal{A}$.
- En général l'ensemble F est égal à $\mathbb{N}$ ou $\mathbb{Z}$ ou à une partie de $\mathbb{Z}$.

– Notons que $(\{X = x\})_{x\in F}$ est une famille dénombrable d'événements deux à deux disjoints t.q. $\bigcup_{x\in F}\{X = x\} = \Omega$. Donc par la propriété de σ-additivité,

$$\sum_{x\in F} \mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{x\in F}\{X = x\}\right) = \mathbb{P}(\Omega) = 1.$$

Exemple 2.2.4. – Dans le cas du jet de deux dés, la somme S des deux dés est une variable aléatoire discrète à valeurs dans $F = [2, 12]$ dont nous avons calculé la loi dans l'exemple 1.1.3 :

$$\begin{array}{ll} \mathbb{P}(S = 2) = \mathbb{P}(S = 12) = 1/36 & \mathbb{P}(S = 3) = \mathbb{P}(S = 11) = 1/18 \\ \mathbb{P}(S = 4) = \mathbb{P}(S = 10) = 1/12 & \mathbb{P}(S = 5) = \mathbb{P}(S = 9) = 1/9 \\ \mathbb{P}(S = 6) = \mathbb{P}(S = 8) = 5/36 & \mathbb{P}(S = 7) = 1/6. \end{array}$$

Il faut noter que la loi de S est la probabilité naturelle dont on doit munir l'ensemble $[2, 12]$ lorsque l'on s'intéresse à la somme de deux dés.

– Soit $A \subset \Omega$ un événement. Sa fonction indicatrice $1_A : \Omega \to \{0, 1\}$ définie par

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

est une variable aléatoire discrète de loi :

$$\mathbb{P}(1_A = 1) = \mathbb{P}(A) \quad \text{et} \quad \mathbb{P}(1_A = 0) = 1 - \mathbb{P}(A).$$

2.2.3 Indépendance

Définition 2.2.5. – *Deux variables aléatoires discrètes X et Y à valeurs respectivement dans F et G sont dites indépendantes si*

$$\boxed{\forall x \in F,\ \forall y \in G,\ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y).}$$

– *n variables aléatoires discrètes $X_1, X_2, \ldots, X_n$ à valeurs respectivement dans $F_1, F_2, \ldots, F_n$ sont dites indépendantes si*

$$\boxed{\forall x_1 \in F_1, \ldots,\ \forall x_n \in F_n,\ \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).} \tag{2.1}$$

– *Une famille quelconque de variables aléatoires discrètes est dite indépendante si toute sous-famille finie l'est.*

Exemple 2.2.6. Jet de 2 dés : $\Omega = \{(i, j) :\ 1 \leq i, j \leq 6\}$ muni de la probabilité uniforme. Si on note X_1 la valeur du premier dé et X_2 celle du second, on a $X_1(i, j) = i$ et $X_2(i, j) = j$. On vérifie facilement que

$$\forall 1 \leq i \leq 6,\ \mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = i) = \frac{1}{6},$$

ce qui n'a rien de surprenant. Comme

$$\forall 1 \leq i, j \leq 6,\ \mathbb{P}(X_1 = i, X_2 = j) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = \mathbb{P}(X_1 = i) \times \mathbb{P}(X_2 = j),$$

les variables X_1 et X_2 sont indépendantes.

Remarque 2.2.7. – Il peut sembler étonnant que dans la définition de l'indépendance de n variables aléatoires discrètes, contrairement à celle de l'indépendance des événements (définition 1.2.8), on ne regarde pas de propriétés portant sur des sous-ensembles des n variables. La raison est la suivante : en sommant par exemple sur $x_n \in F_n$ l'égalité (2.1), et en utilisant la propriété de σ-additivité on obtient

$$\mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n \in F_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_{n-1} = x_{n-1})\mathbb{P}(X_n \in F_n)$$
$$\text{i.e. } \mathbb{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \prod_{i=1}^{n-1} \mathbb{P}(X_i = x_i),$$

c'est-à-dire une relation analogue à (2.1) mais dans laquelle la variable X_n a disparu. De façon plus générale pour $1 \leq d < n$, $\mathbb{P}(X_1 = x_1, \ldots, X_d = x_d) = \prod_{i=1}^{d} \mathbb{P}(X_i = x_i)$.

– La remarque précédente permet de montrer que si les variables aléatoires discrètes $X_1, \ldots, X_n$ sont indépendantes, pour $1 \leq d < n$, les deux variables aléatoires discrètes $(X_1, \ldots, X_d)$ et $(X_{d+1}, \ldots, X_n)$ sont indépendantes : en effet,

$$\begin{aligned}\mathbb{P}((X_1, \ldots, X_d) = (x_1, \ldots, x_d))\mathbb{P}((X_{d+1}, \ldots, X_n) = (x_{d+1}, \ldots, x_n)) \\ = \prod_{i=1}^{d} \mathbb{P}(X_i = x_i) \prod_{j=d+1}^{n} \mathbb{P}(X_j = x_j) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n).\end{aligned}$$

Ce résultat se généralise bien sûr de la façon suivante : $\forall m \in [1, n]$, $\forall 1 \leq d_1 < d_2 < \ldots < d_m \leq n$, les variables aléatoires discrètes $(X_1, \ldots, X_{d_1})$, $(X_{d_1+1}, \ldots, X_{d_2})$,..., $(X_{d_{m-1}+1}, \ldots, X_{d_m})$ et $(X_{d_m+1}, \ldots, X_n)$ sont indépendantes.

2.2.4 Lois discrètes usuelles

Ce sont des lois qui portent sur $F \subset \mathbb{N}$.

Loi de Bernoulli de paramètre $p \in [0, 1]$

On dit que X suit la loi de Bernoulli de paramètre p et on note $X \sim \mathcal{B}(p)$ si :

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

On a alors

$$\forall x \in \{0, 1\},\ \mathbb{P}(X = x) = p^x(1 - p)^{1-x} \quad \text{(convention : } 0^0 = 1). \tag{2.2}$$

Exemple 2.2.8. Si A est un événement, $1_A \sim \mathcal{B}(\mathbb{P}(A))$.

Loi binômiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$

C'est la loi de la somme $S = X_1 + \ldots + X_n$ de n variables de Bernoulli de paramètre p indépendantes $X_1, \ldots, X_n$. On a alors pour $k \in F = \{0, \ldots, n\}$,

$$\begin{aligned}
\mathbb{P}(S = k) = \mathbb{P}(X_1 + \ldots + X_n = k) &= \mathbb{P}\left(\bigcup_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} \{X_1 = x_1, \ldots, X_n = x_n\} \right) \\
&= \sum_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \ \text{ par } \sigma\text{-additivité}, \\
&= \sum_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n) \ \text{ par indépendance des } X_i, \\
&= \sum_{\substack{x_i \in \{0,1\} \\ x_1 + \ldots + x_n = k}} p^{x_1 + \ldots + x_n} (1-p)^{n - x_1 - \ldots - x_n} \ \text{ d'après (2.2)}, \\
&= p^k (1-p)^{n-k} \text{Card } (\{(x_1, \ldots, x_n) \in \{0,1\}^n : \ x_1 + \ldots + x_n = k\}) \\
&= C_n^k \, p^k (1-p)^{n-k}.
\end{aligned}$$

$$\boxed{\text{Si } \forall 0 \leq k \leq n, \ \mathbb{P}(S = k) = C_n^k \, p^k (1-p)^{n-k}, \text{ on note } S \sim \mathcal{B}(n,p).}$$

Loi de Poisson de paramètre $\lambda > 0$

On dit que N suit la loi de Poisson de paramètre $\lambda > 0$ et on note $N \sim \mathcal{P}(\lambda)$ si

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

Loi géométrique de paramètre $p \in]0,1]$

C'est la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes où la probabilité de succès est p.
Une telle suite se modélise à l'aide d'une suite $(X_i)_{i \geq 1}$ de variables indépendantes et identiquement distribuées (abréviation : I.I.D.) suivant la loi de Bernoulli de paramètre p. L'événement "la ième expérience est un succès" s'écrit alors $\{X_i = 1\}$ et le temps T de premier succès est donné par

$$T = \inf\{i \geq 1 : \ X_i = 1\}.$$

Pour $k \geq 1$, en utilisant l'indépendance des X_i on obtient

$$\begin{aligned}
\mathbb{P}(T = k) &= \mathbb{P}(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1) \\
&= \mathbb{P}(X_1 = 0) \ldots \mathbb{P}(X_{k-1} = 0) \mathbb{P}(X_k = 1) = (1-p)^{k-1} p.
\end{aligned}$$

$$\boxed{\text{Si } \forall k \in \mathbb{N}^*, \ \mathbb{P}(T = k) = p(1-p)^{k-1}, \text{ on note } T \sim \mathcal{G}eo(p).}$$

On vérifie que

$$\mathbb{P}(T \in \mathbb{N}^*) = \sum_{k\in\mathbb{N}^*} \mathbb{P}(T = k) = p \sum_{k\in\mathbb{N}^*} (1-p)^{k-1} = p \sum_{l\in\mathbb{N}} (1-p)^l = p\, \frac{1}{1-(1-p)} = 1.$$

Exercice 2.2.9. Soit $S \sim \mathcal{G}eo(p)$ et $T \sim \mathcal{G}eo(q)$ deux variables indépendantes. On cherche la loi de $Z = \min(S, T)$.

1. Pour $k \in \mathbb{N}^*$, calculer $\mathbb{P}(S \geq k)$.
2. En déduire $\mathbb{P}(Z \geq k)$.
3. Quelle est la loi de Z ?
4. Retrouver ce résultat en raisonnant en termes de temps de premier succès.

2.2.5 Loi marginale

Si X est une variable aléatoire discrète à valeurs dans F et Y une variable discrète à valeur dans G alors, comme le produit de deux ensembles dénombrables est dénombrable, (X, Y) est une variable aléatoire discrète à valeurs dans $F \times G$. Mais la connaissance de la loi de X et de la loi de Y ne suffit pas à connaître la loi de (X, Y). Il faut rajouter de l'information comme par exemple le caractère indépendant de X et Y qui entraîne $\mathbb{P}((X, Y) = (x, y)) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ pour obtenir la loi du couple.

Exemple 2.2.10. Soit $X \sim \mathcal{B}(1/2)$. Alors il est facile de voir que $Y = 1 - X \sim \mathcal{B}(1/2)$. Ainsi Y a même loi que X, ce que l'on note aussi $Y \overset{\mathcal{L}}{=} X$.
Donc les premières coordonnées des couples (X, X) et (X, Y) ont même loi de même que les secondes coordonnées. Mais les deux couples n'ont bien sûr pas la même loi :

$$\mathbb{P}((X, Y) = (1, 0)) = \mathbb{P}(X = 1) = \frac{1}{2} \neq 0 = \mathbb{P}((X, X) = (1, 0)).$$

En revanche, si l'on connaît la loi du couple discret (X, Y) on en déduit la loi de X (et celle de Y) par la formule dite de la *loi marginale* :

Proposition 2.2.11. *Soit (X, Y) un couple discret à valeurs dans $F \times G$. Alors*

$$\boxed{\forall x \in F,\ \mathbb{P}(X = x) = \sum_{y\in G} \mathbb{P}(X = x, Y = y).}$$

On somme sur les valeurs prises par la variable Y dont on souhaite se débarrasser.

Démonstration : Le résultat se déduit de la propriété de σ-additivité en remarquant que $\{X = x\}$ est l'union disjointe de la famille dénombrable $\{X = x, Y = y\}_{y\in G}$. □

Remarque 2.2.12. Si la loi de (X, Y) se met sous forme produit i.e.

$$\forall (x, y) \in F \times G,\ \mathbb{P}(X = x, Y = y) = c\mu(x)\nu(y),$$

alors les variables X et Y sont indépendantes de lois respectives $(\mu(x)/\sum_{x'\in F}\mu(x'))_{x\in F}$ et $(\nu(y)/\sum_{y'\in G}\nu(y'))_{y\in G}$.
En effet, comme $\sum_{(x',y')\in F\times G}\mathbb{P}(X=x',Y=y')=1$,

$$c=1\Big/\left(\left(\sum_{x'\in F}\mu(x')\right)\left(\sum_{y'\in G}\nu(y')\right)\right).$$

La formule de la loi marginale permet de vérifier que $\mathbb{P}(X=x)=\mu(x)/\sum_{x'\in F}\mu(x')$ et $\mathbb{P}(Y=y)=\nu(y)/\sum_{y'\in G}\nu(y')$ et on a bien $\mathbb{P}(X=x,Y=y)=\mathbb{P}(X=x)\times\mathbb{P}(Y=y)$.

2.3 Espérance et variance

2.3.1 Espérance

Définition 2.3.1. *Soit $X:\Omega\to F\subset\mathbb{R}$ une variable aléatoire discrète à valeurs réelles. Elle est dite intégrable si $\sum_{x\in F}|x|\mathbb{P}(X=x)<+\infty$. Dans ce cas, on définit son espérance $\mathbb{E}(X)$ par*

$$\boxed{\mathbb{E}(X)=\sum_{x\in F}x\mathbb{P}(X=x).}$$

Remarque 2.3.2.
- Le caractère intégrable et l'espérance d'une variable aléatoire ne dépendent que de sa loi : $X\overset{\mathcal{L}}{=}Y\Rightarrow\mathbb{E}(X)=\mathbb{E}(Y)$.
- Si on note $|F|=\{|x|:x\in F\}$, alors

$$\sum_{x\in F}|x|\mathbb{P}(X=x)=\sum_{y\in|F|}(|y|\mathbb{P}(X=y)+|-y|\mathbb{P}(X=-y))=\sum_{y\in|F|}y\mathbb{P}(|X|=y).$$

  Donc X est intégrable si et seulement si $|X|$ l'est et dans ce cas, $|\mathbb{E}(X)|\leq\mathbb{E}(|X|)$.
- Soit A un événement. Sa fonction indicatrice qui est à valeurs dans $\{0,1\}$ est bien sûr intégrable et

$$\boxed{\mathbb{E}(1_A)=1\times\mathbb{P}(1_A=1)+0\times\mathbb{P}(1_A=0)=\mathbb{P}(A).}$$

  Ainsi l'espérance de la fonction indicatrice d'un événement est égale à la probabilité de cet événement. On peut donc dire que la notion d'espérance prolonge celle de probabilité.

Exemple 2.3.3.
- Si $X\sim\mathcal{B}(p)$,

$$\mathbb{E}(X)=1\times\mathbb{P}(X=1)+0\times\mathbb{P}(X=0)=p.$$

- Si $Y\sim\mathcal{P}(\lambda)$,

$$\mathbb{E}(Y)=\sum_{n\in\mathbb{N}}n\exp(-\lambda)\frac{\lambda^n}{n!}=\lambda\exp(-\lambda)\sum_{n\in\mathbb{N}^*}\frac{\lambda^{n-1}}{(n-1)!}=\lambda\exp(-\lambda)\sum_{k\in\mathbb{N}}\frac{\lambda^k}{k!}=\lambda.$$

- Si $Z\sim\mathcal{G}eo(p)$,

$$\mathbb{E}(Z)=\sum_{n\in\mathbb{N}^*}np(1-p)^{n-1}=p(-f'(p))\ \text{ où }\ f(x)=\sum_{k\in\mathbb{N}}(1-x)^k=\frac{1}{x}.$$

$$\text{D'où } E(Z) = p\,\frac{1}{p^2} = \frac{1}{p}.$$

Le fait que le temps moyen d'attente du premier succès décroisse avec la probabilité de succès est parfaitement intuitif.

Propriétés 2.3.4. *1. Linéarité : si X et Y sont deux variables discrètes à valeurs réelles intégrables et $\lambda \in \mathbb{R}$, alors $X + \lambda Y$ est intégrable et*

$$\boxed{\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y).}$$

2. Condition suffisante d'intégrabilité : Si X et Y deux variables aléatoires discrètes réelles t.q. $\mathbb{P}(|X| \leq Y) = 1$ et Y est intégrable, alors X l'est aussi.

3. Positivité : si X est une variable aléatoire discrète intégrable et presque sûrement positive au sens où $\mathbb{P}(X \geq 0) = 1$ alors

- $\mathbb{E}(X) \geq 0$.
- $\mathbb{E}(X = 0) \Longrightarrow \mathbb{P}(X = 0) = 1$ *(On dit alors que X est presque sûrement nulle).*

4. Croissance : si X et Y sont deux variables intégrables t.q. $\mathbb{P}(X \geq Y) = 1$ alors $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

Exercice 2.3.5. Soit $Y \sim \mathcal{B}(n, p)$. Montrer sans calcul que $\mathbb{E}(Y) = np$.

Démonstration :
Linéarité : Si F et G désignent les ensembles dénombrables dans lesquels X et Y prennent leurs valeurs, alors $Z = X + \lambda Y$ est une variable aléatoire discrète à valeurs dans $H = \{x + \lambda y : (x, y) \in F \times G\}$. En utilisant le fait que pour $z \in H$, l'événement $\{Z = z\}$ est l'union disjointe de la famille dénombrable $\{X = x, Y = y\}_{(x,y) \in F \times G : x + \lambda y = z}$, on obtient par σ-additivité que

$$\mathbb{P}(Z = z) = \sum_{\substack{(x,y) \in F \times G \\ x + \lambda y = z}} \mathbb{P}(X = x, Y = y) = \sum_{(x,y) \in F \times G} 1_{\{x + \lambda y = z\}} \mathbb{P}(X = x, Y = y). \tag{2.3}$$

Vérifions d'abord l'intégrabilité de Z : en utilisant (2.3) et le Théorème de Fubini 2.2.1, on a

$$\begin{aligned}
\sum_{z \in H} |z| \mathbb{P}(Z = z) &= \sum_{z \in H} |z| \sum_{(x,y) \in F \times G} 1_{\{x + \lambda y = z\}} \mathbb{P}(X = x, Y = y) \\
&= \sum_{z \in H} \sum_{(x,y) \in F \times G} 1_{\{x + \lambda y = z\}} |x + \lambda y| \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y) \in F \times G} \left(\sum_{z \in H} 1_{\{x + \lambda y = z\}} \right) |x + \lambda y| \mathbb{P}(X = x, Y = y) \\
&= \sum_{(x,y) \in F \times G} |x + \lambda y| \mathbb{P}(X = x, Y = y) \\
&\leq \sum_{(x,y) \in F \times G} (|x| + |\lambda||y|) \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in F} |x| \sum_{y \in G} \mathbb{P}(X = x, Y = y) + |\lambda| \sum_{y \in G} |y| \sum_{x \in F} \mathbb{P}(X = x, Y = y) \\
&= \sum_{x \in F} |x| \mathbb{P}(X = x) + |\lambda| \sum_{y \in G} |y| \mathbb{P}(Y = y) \text{ d'après la proposition 2.2.11.}
\end{aligned}$$

Le dernier membre est fini par intégrabilité de X et Y.
Pour montrer que $\mathbb{E}(Z) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$, on reprend le calcul que l'on vient d'effectuer en enlevant les valeurs absolues.
C.S. d'intégrabilité :
Soit X et Y deux variables aléatoires discrètes réelles à valeurs respectives dans F et G t.q. $\mathbb{P}(|X| \leq Y) = 1$. Alors $\forall (x,y) \in F \times G$ t.q. $|x| > |y|$, $\mathbb{P}(X = x, Y = y) = 0$. En utilisant la formule de la loi marginale, on en déduit que

$$\begin{aligned}\sum_{x\in F} |x|\mathbb{P}(X = x) &= \sum_{(x,y)\in F\times G} |x|\mathbb{P}(X = x, Y = y)\\ &\leq \sum_{(x,y)\in F\times G} |y|\mathbb{P}(X = x, Y = y) = \sum_{y\in G} |y|\mathbb{P}(Y = y).\end{aligned}$$

Donc si Y est intégrable, alors X l'est aussi.
Positivité et croissance :
Si X est intégrable, par définition $\mathbb{E}(X) = \sum_{x\in F} x\mathbb{P}(X = x)$. Si en outre $\mathbb{P}(X \geq 0) = 1$ alors

$$\mathbb{P}(X < 0) = \sum_{\substack{x\in F\\ x<0}} \mathbb{P}(X = x) = 0 \quad \text{d'où } \mathbb{E}(X) = \sum_{\substack{x\in F\\ x>0}} x\mathbb{P}(X = x).$$

Ainsi $\mathbb{E}(X) \geq 0$.
On en déduit en outre que si $\mathbb{E}(X) = 0$, alors $\forall x \in F \cap]0, +\infty[$, $\mathbb{P}(X = x) = 0$, ce qui entraîne que $\mathbb{P}(X > 0) = \sum_{\substack{x\in F\\ x>0}} \mathbb{P}(X = x) = 0$ puis $\mathbb{P}(X = 0) = \mathbb{P}(X \geq 0) - \mathbb{P}(X > 0) = 1$.
Lorsque X et Y intégrables vérifient $\mathbb{P}(X \geq Y) = 1$, alors $\mathbb{P}(X - Y \geq 0) = 1$ ce qui entraîne $\mathbb{E}(X - Y) \geq 0$. On en déduit par linéarité de l'espérance que ,

$$\mathbb{E}(X) = \mathbb{E}(X - Y) + \mathbb{E}(Y) \geq \mathbb{E}(Y).$$

□

Le résultat suivant qui exprime l'espérance de $f(X)$ en fonction de la loi de X est très utile en pratique.

Théorème 2.3.6. *Soit $X : \Omega \to F$ une variable aléatoire discrète et $f : F \to \mathbb{R}$. Alors la variable $f(X)$ est intégrable si et seulement si $\sum_{x\in F} |f(x)|\mathbb{P}(X = x) < +\infty$ et alors,*

$$\boxed{\mathbb{E}(f(X)) = \sum_{x\in F} f(x)\mathbb{P}(X = x).}$$

Remarque 2.3.7.
- Si f est bornée sur F alors $\sum_{x\in F} |f(x)|\mathbb{P}(X = x) \leq \sup_{x\in F} |f(x)|$ et $f(X)$ est intégrable.
- $\mathbb{E}(f(X))$ ne dépend que de la loi de X : $X \overset{\mathcal{L}}{=} Y \Rightarrow \mathbb{E}(f(X)) = \mathbb{E}(f(Y))$.

Démonstration : La variable aléatoire $Y = f(X)$ est discrète à valeurs dans $f(F)$.

$$\sum_{y\in f(F)} |y|\mathbb{P}(Y = y) = \sum_{y\in f(F)} |y| \sum_{\substack{x\in F\\ f(x)=y}} \mathbb{P}(X = x) = \sum_{x\in F} |f(x)|\mathbb{P}(X = x).$$

D'où la caractérisation de l'intégrabilité de $f(X)$. Dans le cas où cette variable aléatoire est effectivement intégrable, on peut reprendre le calcul précédent en enlevant les valeurs absolues pour obtenir l'expression désirée de son espérance $\square$

La caractérisation suivante de l'indépendance au travers des espérances fait intervenir un produit. D'une façon très générale l'indépendance permet d'effectuer des factorisation et il est bon d'avoir en tête le lien "indépendance = produit".

Proposition 2.3.8. *Deux variables aléatoires discrètes X et Y à valeurs respectivement dans F et G sont indépendantes si et seulement si pour toutes fonctions $f : F \to \mathbb{R}$ et $g : G \to \mathbb{R}$ t.q. $f(X)$ et $g(Y)$ sont intégrables, alors $f(X)g(Y)$ est intégrable et*

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

Démonstration :
<u>Condition suffisante :</u> Soit $x \in F$ et $y \in G$. On pose $f(w) = 1_{\{w=x\}}$ et $g(z) = 1_{\{z=y\}}$. En utilisant la remarque 2.3.2, l'égalité de l'énoncé se récrit

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

<u>Condition nécessaire :</u> Supposons que $f(X)$ et $g(Y)$ sont intégrables avec X et Y indépendantes. L'indépendance entraîne que

$$\begin{aligned}\sum_{(x,y)\in F\times G} |f(x)g(y)|\mathbb{P}(X = x, Y = y) &= \sum_{(x,y)\in F\times G} |f(x)||g(y)|\mathbb{P}(X = x)\mathbb{P}(Y = y)\\ &= \left(\sum_{x\in F} |f(x)|\mathbb{P}(X = x)\right)\left(\sum_{y\in G} |g(y)|\mathbb{P}(Y = y)\right).\end{aligned}$$

Le dernier membre est fini par intégrabilité de $f(X)$ et $g(Y)$ et d'après le théorème 2.3.6. Toujours d'après ce résultat, on en déduit que $f(X)g(Y)$ est intégrable. En reprenant le calcul qui précède sans valeurs absolues, on conclut que l'espérance du produit est égale au produit des espérances. $\square$

Exercice 2.3.9. Montrer que si X et Y sont deux variables aléatoires discrètes indépendantes à valeurs respectives dans F et G, et $\varphi : F \to F'$, $\psi : G \to G'$ alors les variables $\varphi(X)$ et $\psi(Y)$ sont indépendantes.

2.3.2 Variance

Définition 2.3.10. – *Soit $X : \Omega \to F \subset \mathbb{R}$ une variable aléatoire discrète à valeurs réelles. Alors X est dite de carré intégrable si X^2 est intégrable i.e. si*

$$\sum_{x\in F} x^2\mathbb{P}(X = x) < +\infty.$$

Dans ce cas, on définit la variance de X par

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

– *La racine carrée de la variance est appelée écart-type.*

Remarque 2.3.11. – Si X est de carré intégrable alors comme $|X| \leq (1+X^2)/2$, d'après la condition suffisante d'intégrabilité donnée dans les propriétés 2.3.4, cette variable est intégrable ce qui donne un sens à $\mathbb{E}(X)$. En outre, par linéarité de l'espérance , $(X-\mathbb{E}(X))^2 = X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2$ est intégrable. Ainsi la variance est bien définie.
– La variance et l'écart-type mesurent l'étalement de la variable aléatoire X autour de son espérance : plus ils sont grands et plus X est étalée.
–

Exercice 2.3.12. – Montrer que

$$\boxed{\text{Var }(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.} \tag{2.4}$$

– Montrer que $\forall a, b \in \mathbb{R}$, Var $(aX+b) = a^2$Var (X).

Remarque 2.3.13. – L'expression (2.4) et la remarque 2.3.7 impliquent que la variance d'une variable aléatoire ne dépend que de sa loi : $X \overset{\mathcal{L}}{=} Y \Longrightarrow$ Var $(X) =$ Var (Y).
– Comme par positivité de l'espérance Var $(X) \geq 0$, l'égalité (2.4) entraîne que $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$.

Dans l'énoncé suivant, l'hypothèse d'indépendance est essentielle. Pour une fois, elle ne permet pas une factorisation mais une sommation :

Proposition 2.3.14. *Soit $X_1, \ldots, X_n$ des variables aléatoires de carré intégrable intégrable. Alors $X_1 + \ldots + X_n$ est de carré intégrable et*

$$\boxed{\textit{si les } X_i \textit{ sont } \textbf{indépendantes}, \textit{ alors } \text{Var }(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var }(X_i).}$$

Démonstration : Le fait que la somme $X_1 + \ldots + X_n$ est de carré intégrable découle de l'inégalité $(X_1 + \ldots + X_n)^2 \leq n(X_1^2 + \cdots + X_n^2)$. Par linéarité de l'espérance,

$$\begin{aligned}\text{Var }(X_1 + \ldots + X_n) &= \mathbb{E}\left(\left(\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\right)^2\right) \\ &= \mathbb{E}\left(\sum_{i,j=1}^{n}(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right).\end{aligned}$$

Si Y et Z sont deux variables de carré intégrable, comme $|YZ| \leq (Y^2+Z^2)/2$, leur produit YZ est intégrable. Donc chaque terme $(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))$ est intégrable et par linéarité de l'espérance,

$$\text{Var }(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var }(X_i) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right).$$

On conclut en remarquant que par **indépendance** des variables $X_1, \ldots, X_n$, pour $i \neq j$,

$$\mathbb{E}\left((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) = \mathbb{E}(X_i - \mathbb{E}(X_i))\mathbb{E}(X_j - \mathbb{E}(X_j)) = 0.$$

□

Exercice 2.3.15. Calculer la variance d'une variable de Bernoulli de paramètre p, d'une variable binômiale de paramètres n et p, d'une variable de Poisson de paramètre λ et d'une variable géométrique de paramètre p.

2.4 Loi conditionnelle

Définition 2.4.1. – *Soient X et Y deux variables aléatoires discrètes à valeurs dans F et G respectivement. Pour $y \in G$ t.q. $\mathbb{P}(Y = y) > 0$ on appelle loi conditionnelle de X sachant $Y = y$ la famille des nombres* $\left(\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}\right)_{x\in F}$.

– *Pour $f : F \to \mathbb{R}$ bornée (ou t.q. $f(X)$ est intégrable) on appelle espérance de $f(X)$ sachant $Y = y$ le nombre $\mathbb{E}(f(X)|Y = y)$ défini par*

$$\mathbb{E}(f(X)|Y = y) = \frac{\mathbb{E}\left(f(X)1_{\{Y=y\}}\right)}{\mathbb{P}(Y = y)}.$$

Remarque 2.4.2. Comme d'après la proposition 2.2.11, $\sum_{x\in F} \mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y)$,

$$\sum_{x\in F} \mathbb{P}(X = x|Y = y) = \sum_{x\in F} \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{1}{\mathbb{P}(Y = y)} \sum_{x\in F} \mathbb{P}(X = x, Y = y) = 1.$$

Propriétés 2.4.3. *1. Pour toute fonction $f : F \to \mathbb{R}$ bornée, on a une formule analogue à la définition de l'espérance pour calculer l'espérance conditionnelle :*

$$\mathbb{E}(f(X)|Y = y) = \sum_{x\in F} f(x)\mathbb{P}(X = x|Y = y).$$

2. Les variables X et Y sont indépendantes si et seulement si la loi conditionnelle de X sachant $Y = y$ ne dépend pas de $y \in G$ t.q. $\mathbb{P}(Y = y) > 0$.

Démonstration : 1. Comme $\sum_{x\in F} 1_{\{X=x\}} = 1$, par linéarité de l'espérance,

$$\begin{aligned}
\mathbb{E}(f(X)1_{\{Y=y\}}) &= \mathbb{E}\left(\sum_{x\in F} 1_{\{X=x\}} f(X)1_{\{Y=y\}}\right) \\
&= \sum_{x\in F} \mathbb{E}\left(f(x)1_{\{X=x,Y=y\}}\right) \\
&= \sum_{x\in F} f(x)\mathbb{P}(X = x, Y = y).
\end{aligned}$$

Il suffit de diviser par $\mathbb{P}(Y = y)$ le premier et le dernier membre de cette égalité pour obtenir le résultat souhaité.
2. Il est immédiat de vérifier la condition nécessaire. Nous allons donc nous concentrer sur la condition suffisante.
Quitte à restreindre G, nous supposerons que pour tout $y \in G$, $\mathbb{P}(Y = y) > 0$. Par hypothèse, pour tout $x \in F$, il existe $\mu(x)$ t.q. $\forall y \in G$,

$$\mathbb{P}(X = x|Y = y) = \mu(x). \tag{2.5}$$

A x fixé, en multipliant cette égalité par $\mathbb{P}(Y = y)$ et en sommant sur $y \in G$, on obtient $\mathbb{P}(X = x) = \mu(x)$. Donc (2.5) se récrit

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

□

Exercice résolu 2.4.4. *Soient $X_1, \ldots, X_n$ des variables aléatoires I.I.D. (indépendantes et identiquement distribuées) suivant la loi de Bernoulli de paramètre $p \in]0, 1[$ et $S = X_1 + \ldots + X_n$ leur somme. Pour $s \in [0, n]$, quelle est la loi conditionnelle de X_1 sachant $S = s$?*

Soit $x \in \{0, 1\}$.

$$\begin{aligned}\mathbb{P}(X_1 = x|S = s) &= \frac{\mathbb{P}(X_1 = x, S = s)}{\mathbb{P}(S = s)} = \frac{\mathbb{P}(X_1 = x, X_2 + \ldots + X_n = s - x)}{\mathbb{P}(S = s)} \\ &= \frac{\mathbb{P}(X_1 = x)\mathbb{P}(X_2 + \ldots + X_n = s - x)}{\mathbb{P}(S = s)} \text{ par indépendance des } X_i.\end{aligned}$$

On a $\mathbb{P}(X_1 = x) = p^x(1 - p)^{1-x}$.
La variable S suit la loi binômiale $\mathcal{B}(n, p)$ tandis que $X_2 + \ldots + X_n \sim \mathcal{B}(n - 1, p)$. Donc

$$\mathbb{P}(S = s) = C_n^s p^s(1-p)^{n-s} \text{ et } \mathbb{P}(X_2+\ldots+X_n = s-x) = \begin{cases} 0 \text{ si } s - x = n \text{ ou } s - x = -1 \\ C_{n-1}^{s-x} p^{s-x}(1 - p)^{n-1-s+x} \text{ sinon} \end{cases}.$$

Donc

$$\mathbb{P}(X_1 = x|S = s) = \begin{cases} 0 \text{ si } s - x = n \text{ ou } s - x = -1 \\ \frac{C_{n-1}^{s-x}}{C_n^s} = \frac{(n-1)!}{n!}\frac{s!}{(s-x)!}\frac{(n-s)!}{(n-1-s+x)!} = \frac{1}{n}s^x(n - s)^{1-x} \text{ sinon} \end{cases}.$$

Ainsi pour $x \in \{0, 1\}$, $\mathbb{P}(X_1 = x|S = s) = \left(\frac{s}{n}\right)^x \left(\frac{n-s}{n}\right)^{1-x}$.
On conclut que la loi conditionnelle de X_1 sachant $S = s$ est la loi de Bernoulli de paramètre s/n, ce qui est assez intuitif. Il faut remarquer qu'elle ne dépend pas du paramètre p de départ.

2.5 Exercices

Exercice 2.5.1. Soit N une variable aléatoire qui suit la loi de Poisson de paramètre $\lambda > 0$. Calculer $\mathbb{E}\left(\frac{1}{N+1}\right)$.

Exercice 2.5.2. Soient X et Y des variables indépendantes qui suivent des lois binomiales de paramètres respectifs (n, p) et (m, p). Quelle est la loi de la somme $S = X + Y$?

Exercice 2.5.3. Dans la population, une proportion $p << 1$ d'individus est touchée par une maladie. Lors du test sanguin qui permet de détecter cette maladie, on groupe les échantillons de sang par lots de n : si le test est négatif, les n patients du lot sont sains tandis que si le test est positif, on pratique n tests individuels pour déterminer quels sont les patients atteints. Le coût d'un test sera pris égal à 1.

1. Montrer que le coût moyen par individu est égal à $C(n) = 1 + \frac{1}{n} - (1 - p)^n$. En faisant un développement au premier ordre en utilisant $p << 1$, trouver la valeur de n qui minimise ce coût.

2. Dans le cas où $p = 0,01$, calculer cette valeur optimale ainsi que l'économie réalisée par rapport au cas $n = 1$ où on teste le sang de chaque patient séparément.

Exercice 2.5.4. 1. Pour $n \in \mathbb{N}^*$, $0 \leq p \leq 1$ et $k \in \mathbb{N}$, on note $b_{n,p}(k)$ la probabilité pour qu'une variable qui suit la loi binomiale de paramètre (n, p) vaille k. De manière analogue, pour $\lambda > 0$ et $k \in \mathbb{N}$, on note $p_\lambda(k)$ la probabilité pour qu'une variable qui suit la loi de Poisson de paramètre λ vaille k.
Montrer que si $n \to +\infty$ et $p \to 0$ avec $np \to \lambda$ la loi de binomiale de paramètre (n, p) converge vers la loi de Poisson de paramètre λ au sens suivant :

$$\forall k \in \mathbb{N},\ \lim b_{n,p}(k) = p_\lambda(k).$$

(<u>Indication :</u> commencer par montrer que pour $k \in \mathbb{N}$ fixé $\lim_{n\to+\infty} \frac{n!}{(n-k)!n^k} = 1$)

2. Si X est une variable discrète distribuée suivant la loi de Poisson de paramètre λ, calculer $P(X = 1|X \geq 1)$.
Dans une ville de 200000 personnes, un voleur a laissé une empreinte digitale de type t. La police arrête un suspect dont les empreintes digitales sont de type t. Sachant que la probabilité pour qu'une personne ait des empreintes de ce type est 5.10^{-6}, est-il raisonnable de condamner le suspect sur la foi de ce seul indice ?
3. Aux jeux olympiques de Sydney, 300 médailles d'or ont été mises en jeu. En faisant l'hypothèse que "le nombre de médailles d'or remportées par un pays est proportionnel à sa poulation" que l'on traduit mathématiquement en supposant que ce nombre est une variable binomiale de paramètre ($n = 300, p =$ population du pays/population mondiale), calculer la probabilité pour que le nombre de médailles remporté par les athlètes français soit supérieur ou égal à 10 (la population française sera prise égale à 50 millions et la population mondiale à 5 milliards). Que concluez-vous du fait que la France a remporté 13 médailles d'or ?

Exercice 2.5.5. Une récente étude portant sur les risques encourus par les patients opérés pour une prothèse de hanche a révélé que sur 100 hôpitaux qui ont pratiqué 10 opérations de ce type, le taux moyen de décès est de 2,5% et que le plus mauvais taux de décès dans un hôpital est de 3 patients sur 10.
On peut se demander si le fait que l'un des hôpitaux a connu un taux de décès aussi élevé est dû au hasard. Pour tenter d'apporter un éclairage probabiliste sur cette question, on suppose que la probabilité qu'un patient donné meure pendant son opération est de $0,025$ et que les destins des patients sont indépendants.

- Quelle est la loi du nombre de décès dans un hopital qui pratique 10 interventions ? Quelle est la probabilité pour que 3 au moins des 10 patients d'un hôpital donné soient morts ?
- Quelle est la probabilité pour que dans l'un au moins des 100 hôpitaux, 3 patients au moins soient morts.

Commentaires ?

Exercice 2.5.6. On suppose que le nombre N de clients pendant une journée dans un grand magasin suit la loi de Poisson de paramètre $\lambda > 0$.
Chaque client a une probabilité p de se faire voler son portefeuille et ce indépendamment des autres clients, ce que l'on modèlise à l'aide d'une suite $(X_i)_{i\geq 1}$ de variables I.I.D. suivant la loi de Bernoulli de paramètre p indépendante de N : $X_i = 1$ si le ième client se fait dépouiller.

1. Exprimer le nombre V de clients volés en fonction de N et des X_i.

2. Déterminer la loi de $(V, N-V)$. En déduire la loi de V, celle de $N-V$. Ces deux variables aléatoires sont-elles indépendantes ?

Exercice 2.5.7. On considère une suite $(X_i)_{i\geq 1}$ de variables de Bernoulli de paramètre p $(0 < p < 1)$ indépendantes (schéma de Bernoulli) et on s'intéresse aux temps d'apparition

$$T_{10} = \inf\{i \geq 2,\ (X_{i-1}, X_i) = (1,0)\} \quad \text{et} \quad T_{11} = \inf\{i \geq 2,\ (X_{i-1}, X_i) = (1,1)\}$$

des séquences "10" et "11".

1. Soit $T_1 = \inf\{i \geq 1,\ X_i = 1\}$. Calculer $\mathbb{P}(T_1 = k, T_{10} - T_1 = l)$ et en déduire l'espérance de T_{10}.
2. Calculer $\mathbb{P}(T_{11} = 2)$, puis montrer en distinguant les cas $X_1 = 0$, $(X_1, X_2) = (1,0)$ et $(X_1, X_2) = (1,1)$ que $\forall k \geq 3$,

$$\mathbb{P}(T_{11} = k) = (1-p)\mathbb{P}(T_{11} = k-1) + p(1-p)\mathbb{P}(T_{11} = k-2).$$

   En déduire l'espérance de T_{11}.
3. On suppose $p = 1/2$. Comparer les espérances de T_{10} et T_{11}. En remarquant que $\{T_{11} < T_{10}\} = \{X_{T_1+1} = 1\}$ et en découpant cet événement suivant les valeurs prises par T_1, calculer $\mathbb{P}(T_{11} < T_{10})$. Commentaires ?

Exercice 2.5.8.
1. Soit X une variable aléatoire discrète qui suit la loi géométrique de paramètre p $(0 < p \leq 1)$.
   Montrer que X "n'a pas de mémoire", c'est-à-dire que la loi conditionnelle de $X - n_0$ sachant $X > n_0$ ne dépend pas de $n_0 \geq 0$.
2. Inversement, caractériser la loi d'une variable aléatoire à valeurs dans $\mathbb{N}^*$ qui "n'a pas de mémoire". On pourra pour cela montrer par récurrence que

$$\forall k \geq 1,\ P(X \geq k) = P(X \geq 2)^{k-1}.$$

Exercice 2.5.9. Soit X et Y deux variables aléatoires indépendantes distribuées respectivement suivant les lois de Poisson de paramètre $\lambda > 0$ et de paramètre $\mu > 0$. On note $S = X + Y$.

1. Déterminer la loi de S.
2. Pour tout $s \in \mathbb{N}$ déterminer la loi conditionnelle de X sachant $S = s$.

Exercice 2.5.10. Soit $A_1, \ldots A_n$ des événements.

1. Montrer que $1_{A_1 \cup \ldots \cup A_n} = 1 - \prod_{i=1}^n (1 - 1_{A_i})$.
2. En déduire la formule du crible (ou de Poincaré) énoncée dans l'exercice 1.3.5 :

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

2.6 Résumé

- **Variable aléatoire discrète** $X : \Omega \to F$ où F est dénombrable.
  La famille $(\mathbb{P}(X = x))_{x \in F}$ s'appelle la **loi** de X.

- Soit X et Y deux variables discrètes à valeurs respectives dans F et G.
  - **Loi marginale :** $\forall x \in F,\ \mathbb{P}(X = x) = \sum_{y \in G} \mathbb{P}(X = x, Y = y)$.
  - **Loi conditionnelle** de X sachant $Y = y$:
  $$\left(P(X = x | Y = y) = \frac{P(X = x, Y = y)}{\mathbb{P}(Y = y)} \right)_{x \in F}$$
  - **Indépendance** si $\forall x \in F,\ \forall y \in G,\ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$
    $\iff$ la loi conditionnelle de X sachant $Y = y$ ne dépend pas de $y \in G$.

- **Espérance et variance** $X : \Omega \to F \subset \mathbb{R}$ est dite
  - intégrable si $\sum_{x \in F} |x| \mathbb{P}(X = x) < +\infty$ et alors $\boxed{\mathbb{E}(X) = \sum_{x \in F} x \mathbb{P}(X = x).}$
  - de carré intégrable si $\sum_{x \in F} x^2 \mathbb{P}(X = x) < +\infty$ et alors
  $$\text{Var}\ (X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

  Propriétés :
  - Pour tout événement A, $\boxed{\mathbb{P}(A) = \mathbb{E}(1_A).}$
  - Linéarité : $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$.
  - Croissance : $\mathbb{P}(X \geq Y) = 1 \Longrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$.
  - X et Y indépendantes $\Longrightarrow \mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$.
  - $X_1, \ldots, X_n$ indépendantes $\Longrightarrow \text{Var}\ (X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var}\ (X_i)$.

- **Espérance de** $f(X)$ **:** $f(X)$ est intégrable ssi $\sum_{x \in F} |f(x)| \mathbb{P}(X = x) < +\infty$ et alors
  $$\boxed{\mathbb{E}(f(X)) = \sum_{x \in F} f(x) \mathbb{P}(X = x).}$$

- **Lois discrètes usuelles :**

Nom	**Loi**	$\mathbb{E}(X)$	Var (X)
Bernoulli $\mathcal{B}(p)$	$\mathbb{P}(X = 1) = p = 1 - \mathbb{P}(X = 0)$	p	$p(1-p)$
binômiale $\mathcal{B}(n, p)$	$\forall 0 \leq k \leq n,\ \mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson $\mathcal{P}(\lambda)$	$\forall n \in \mathbb{N},\ \mathbb{P}(X = n) = \exp(-\lambda)\frac{\lambda^n}{n!}$	λ	λ
géométrique $\mathcal{G}eo(p)$	$\forall n \in \mathbb{N}^*,\ \mathbb{P}(X = n) = p(1-p)^{n-1}$	$1/p$	$(1-p)/p^2$

Chapitre 3

Variables aléatoires à densité

Si on s'intéresse à un pneu de vélo crevé, en prenant la valve comme origine des angles, intuitivement, la probabilité pour que l'abcisse angulaire du point de crevaison se trouve dans l'intervalle $[\theta_1, \theta_2]$ où $0 \leq \theta_1 \leq \theta_2 \leq 2\pi$ est égale à $(\theta_2 - \theta_1)/2\pi$. On ne peut pas modéliser cette abcisse angulaire à l'aide d'une variable aléatoire discrète Θ à valeurs dans $F \subset [0, 2\pi]$[1]. En revanche, le cadre des variables aléatoires à densité est tout à fait adapté à la situation que nous venons de décrire.

Dans les calculs relatifs à ces variables aléatoires, les sommes manipulées dans le cas des variables aléatoires discrètes sont remplacées par des intégrales. Et lorsque l'on s'intéresse à un vecteur aléatoire à densité en dimension n, il est nécessaire de manipuler des intégrales multiples posées sur $\mathbb{R}^n$.

3.1 Manipulation d'intégrales multiples

3.1.1 Théorème de Fubini

Théorème 3.1.1. *Soit $f : \mathbb{R}^2 \to \mathbb{R}$.*

- *Si f est* **positive** *,*

$$\int_{\mathbb{R}^2} f(x,y)dxdy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dy \right) dx = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)dx \right) dy.$$

  Cela signifie que les trois termes sont soit simultanément finis et égaux soit simultanément égaux à $+\infty$.
- *Si f est* **intégrable** *au sens où $\int_{\mathbb{R}^2} |f(x,y)|dxdy < +\infty$ alors l'égalité ci-dessus reste vraie.*

Remarque 3.1.2.
- Pour vérifier l'intégrabilité de f, on peut bien sûr appliquer le théorème à la fonction positive $|f|$.
- Ce résultat se généralise en dimension supérieure i.e. pour le calcul de l'intégrale d'une fonction de n variables positive ou intégrable, l'ordre dans lequel on effectue les intégrations sur chacune de ces variables est sans importance.

[1] Soit en effet θ t.q. $\mathbb{P}(\Theta = \theta) > 0$. Pour $\varepsilon \leq \pi\mathbb{P}(\Theta = \theta)/2$, on a $\mathbb{P}(\Theta \in [\theta - \varepsilon, \theta + \varepsilon]) \geq \mathbb{P}(\Theta = \theta) > 2\varepsilon/2\pi$.

Exemple 3.1.3. Soit $f : [0, +\infty[\to [0, +\infty[$.

$$\begin{aligned}
\int_{[0,+\infty[\times[0,+\infty[} \frac{f(x+y)}{x+y}\, dxdy &= \int_0^{+\infty} \left(\int_0^{+\infty} \frac{f(x+y)}{x+y} dy \right) dx \\
&= \int_0^{+\infty} \left(\int_x^{+\infty} \frac{f(z)}{z} dz \right) dx \\
&= \int_0^{+\infty} \left(\int_0^{+\infty} 1_{\{z \geq x\}} \frac{f(z)}{z} dz \right) dx \\
&= \int_0^{+\infty} \frac{f(z)}{z} \left(\int_0^{+\infty} 1_{\{z \geq x\}} dx \right) dz \\
&= \int_0^{+\infty} f(z) dz.
\end{aligned}$$

Notons au passage l'intérêt d'utiliser la fonction indicatrice $1_{\{z \geq x\}}$ pour éviter de faire des erreurs sur le domaine d'intégration lors de l'échange des intégrales en x et en z.

3.1.2 Changement de variables

Soit φ une bijection continuement différentiable ainsi que son inverse φ^{-1} d'un ouvert O de $\mathbb{R}^d$ sur un ouvert O' de $\mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ bornée et $g : \mathbb{R}^d \to \mathbb{R}$ intégrable.

$$\boxed{\int_O f(\varphi(x)) g(x)\, dx = \int_{O'} f(y) g(\varphi^{-1}(y))\, |\text{Jac } \varphi^{-1}(y)|\, dy,}$$

où

$$\text{Jac } \varphi^{-1}(y) = \text{Det } \left(\frac{\partial (\varphi^{-1})_i}{\partial y_j}(y),\ 1 \leq i, j \leq d \right) = 1 \Big/ \text{Det } \left(\frac{\partial \varphi_i}{\partial x_j}(\varphi^{-1}(y)),\ 1 \leq i, j \leq d \right).$$

Remarque 3.1.4. En dimension $d = 1$, lorsque $O =]a, b[$ avec $a < b$ et φ est strictement décroissante alors $O' =]\varphi(b), \varphi(a)[$ et φ^{-1} est décroissante si bien que $|\text{Jac } \varphi^{-1}(y)| = -(\varphi^{-1})'(y)$. Ainsi le second membre s'écrit

$$\int_{\varphi(b)}^{\varphi(a)} f(y)\, g(\varphi^{-1}(y))\, (-(\varphi^{-1})'(y)) dy = \int_{\varphi(a)}^{\varphi(b)} f(y)\, g(\varphi^{-1}(y))\, (\varphi^{-1})'(y) dy,$$

et on retrouve bien le changement de variables usuel dans $\mathbb{R}$.

Exercice résolu 3.1.5. *Calculer* $I = \int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx$.

On va utiliser la formule de changement de variables pour calculer

$$I^2 = \int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}}\, dx\, dy.$$

On utilise pour cela le changement de variables

$$\varphi : (x, y) \in O = \mathbb{R}^2 \setminus \{(x, 0) : x \leq 0\} \to (\rho, \theta) = \left(x^2 + y^2, 2 \arctan \left(\frac{y}{x + \sqrt{x^2 + y^2}} \right) \right).$$

Notons que si θ désigne l'angle polaire de (x, y),

$$\frac{y/\sqrt{x^2+y^2}}{1+x/\sqrt{x^2+y^2}} = \frac{\sin\theta}{1+\cos\theta} = \tan\frac{\theta}{2}, \quad \text{d'où } 2\arctan\left(\frac{y}{x+\sqrt{x^2+y^2}}\right) = \theta.$$

Ainsi φ est une bijection C^1 ainsi que son inverse de O sur $O' =]0, +\infty[\times] - \pi, \pi[$ et $\varphi^{-1}(\rho, \theta) = (\sqrt{\rho}\ \cos\theta, \sqrt{\rho}\ \sin\theta)$. Pour ne pas se tromper entre le jacobien de φ et de φ^{-1} dans le changement d'élément différentiel, il est éclairant d'utiliser la notation suivante pour la matrice jacobienne

$$\frac{D(x,y)}{D(\rho,\theta)} = \begin{pmatrix} \cos(\theta)/(2\sqrt{\rho}) & -\sqrt{\rho}\ \sin\theta \\ \sin(\theta)/(2\sqrt{\rho}) & \sqrt{\rho}\ \cos\theta \end{pmatrix}.$$

En prenant la valeur absolue du déterminant du second membre on obtient formellement que $\frac{dx\,dy}{d\rho\,d\theta} = \frac{\cos^2\theta+\sin^2\theta}{2} = \frac{1}{2}$ i.e. qu'il faut remplacer $dx\ dy$ par $\frac{1}{2}\ d\rho\ d\theta$.
Comme le volume (ou plutôt la surface) dans $\mathbb{R}^2$ de la demi-droite $\{(x, 0) : x \leq 0\}$ est nul, on ne change rien à la valeur de I^2 en restreignant l'intégration à O : $I^2 = \int_O e^{-\frac{x^2+y^2}{2}}\ dx\ dy$. Par la formule de changement de variable on en déduit que

$$I^2 = \int_{]0,+\infty[\times]-\pi,\pi[} \exp\left(-\frac{\rho}{2}\right)\ \frac{1}{2}\ d\rho\ d\theta = 2\pi.$$

Comme I est l'intégrale d'une fonction positive, on conclut que $I = \sqrt{I^2} = \sqrt{2\pi}$.

En général, dans les problèmes de probabilités, on connaît O et φ et on souhaite transformer une intégrale comme celle du premier membre en une intégrale comme celle du second. Il faut faire attention aux difficultés suivantes :

- la fonction φ n'est pas nécessairement injective sur le domaine O de départ (ex : $O = \mathbb{R}$, $\varphi(x) = x^2$). Pour surmonter cette difficulté, on peut essayer de découper O en sous-domaines sur lesquels φ est injective.
- lorsque φ est injective sur O, il faut bien raisonner par conditions nécessaires et suffisantes pour obtenir le domaine image O' et ne pas se contenter de conditions nécessaires.

  Exemple 3.1.6. Si $O =]0, +\infty[\times]0, +\infty[$ et $\varphi(x, y) = (x+y, x-y)$. Un raisonnement hâtif pourrait laisser penser que $O' = \varphi(O)$ est égal à $]0, +\infty[\times\mathbb{R}$ ce qui est faux. Pour déterminer O', on commence par déterminer φ^{-1} en résolvant le système

  $$\varphi^{-1}(z, w) = (x, y) \Leftrightarrow (z, w) = \varphi(x, y) \Leftrightarrow \begin{cases} z = x + y \\ w = x - y \end{cases} \Leftrightarrow \begin{cases} x = \frac{z+w}{2} \\ y = \frac{z-w}{2} \end{cases}$$

  Ainsi $\varphi^{-1}(z, w) = (\frac{z+w}{2}, \frac{z-w}{2})$. Ensuite on raisonne par conditions nécessaires et suffisantes :

  $$(z, w) \in O' \Leftrightarrow \varphi^{-1}(z, w) \in O \Leftrightarrow \begin{cases} \frac{z+w}{2} > 0 \\ \frac{z-w}{2} > 0 \end{cases} \Leftrightarrow z > |w|.$$

  Ainsi $O' = \{(z, w) \in \mathbb{R}^2 :\ z > |w|\}$.

Remarque 3.1.7. Pour le calcul de certaines intégrales en dimension d petite, le recours au théorème de Fubini et à des changements de variable en dimension 1 peut constituer

une alternative à un changement de variables dans $\mathbb{R}^d$. Ainsi pour $f : \mathbb{R}^2 \to \mathbb{R}$ positive,

$$\begin{aligned}
\int_0^{+\infty}\int_0^{+\infty} f(x+y,x-y)dxdy &= \int_0^{+\infty}\left(\int_0^{+\infty} f(z,z-2y)1_{\{z\geq y\}}dz\right)dy \text{ où } z=x+y,\\
&= \int_0^{+\infty}\left(\int_0^{+\infty} f(z,z-2y)1_{\{z\geq y\}}dy\right)dz \text{ par Fubini},\\
&= \int_0^{+\infty}\left(\int_0^{z} f(z,z-2y)dy\right)dz\\
&= \int_0^{+\infty}\left(\frac{1}{2}\int_{-z}^{z} f(z,w)dw\right)dz \text{ où } w=z-2y,\\
&= \frac{1}{2}\int_{\mathbb{R}^2} 1_{\{z>|w|\}}f(z,w)dzdw.
\end{aligned}$$

Ce résultat s'obtient également par le changement de variables $\varphi(x,y) = (x+y, x-y)$: le domaine image O' a été déterminé dans l'exemple 3.1.6 ci-dessus et un rapide calcul permet de vérifier que $|\text{Jac}\,\varphi^{-1}| = 1/2$.
Mais dès que la dimension d devient grande ou la transformation φ compliquée, la formule de changement de variables ci-dessus devient un outil incontournable.

3.2 Variables aléatoires réelles à densité

3.2.1 Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Définition 3.2.1. *On dit que la variable aléatoire $X : \Omega \to \mathbb{R}$ possède la densité $p : \mathbb{R} \to \mathbb{R}$ si*

$$\boxed{\forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a < X \leq b) = \int_a^b p(x)dx.} \tag{3.1}$$

où $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$.

Remarque 3.2.2.
- En toute rigueur pour pouvoir considérer $\mathbb{P}(a < X \leq b)$ pour $a < b \in \bar{\mathbb{R}}$, il faut demander que $\forall a < b \in \bar{\mathbb{R}}$, $\{a < X < b\} \in \mathcal{A}$.
- La positivité de $\mathbb{P}$ entraîne qu'une densité de probabilité p est une fonction positive. En outre $\mathbb{P}(X \in \mathbb{R}) = \int_{\mathbb{R}} p(x)dx$ implique $\int_{\mathbb{R}} p(x)dx = 1$. Ainsi une densité de probabilité est une fonction positive d'intégrale 1.
- Soit $x \in \mathbb{R}$. Comme $\{X < x\} = \bigcup_{n\geq 0}\{x - \frac{1}{n} < X \leq x - \frac{1}{n+1}\}$ avec la convention $x - \frac{1}{0} = -\infty$, la propriété de σ-additivité implique que

$$\mathbb{P}(X < x) = \sum_{n\geq 0}\mathbb{P}\left(x - \frac{1}{n} < X \leq x - \frac{1}{n+1}\right) = \sum_{n\geq 0}\int_{x-\frac{1}{n}}^{x-\frac{1}{n+1}} p(y)dy = \int_{-\infty}^{x} p(y)dy.$$

  Comme $\{X > x\} = \{x < X \leq +\infty\}$, $\mathbb{P}(X > x) = \int_x^{+\infty} p(y)dy$. Donc

$$\mathbb{P}(X = x) = 1 - \mathbb{P}(X < x) - \mathbb{P}(X > x) = 1 - \int_{\mathbb{R}} p(y)dy = 0.$$

On en déduit que (3.1) est équivalente à

$$\begin{aligned} &\forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a < X < b) = \int_a^b p(x)dx \\ \iff &\forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a \leq X < b) = \int_a^b p(x)dx \\ \iff &\forall a < b \in \bar{\mathbb{R}},\ \mathbb{P}(a \leq X \leq b) = \int_a^b p(x)dx. \end{aligned}$$

Le fait que $\forall x \in \mathbb{R},\ \mathbb{P}(X = x) = 0$ entraîne aussi par σ-additivité que pour tout sous-ensemble F de $\mathbb{R}$ dénombrable, $\mathbb{P}(X \in F) = \sum_{x \in F} \mathbb{P}(X = x) = 0$, ce qui montre la différence de nature entre variables aléatoires discrètes et variables aléatoires à densité.

– D'un point de vue infinitésimal, on a $\mathbb{P}(X \in [x, x + dx]) \simeq p(x)dx$.

3.2.2 Densités réelles usuelles

On dit que X suit la

- **loi uniforme sur** $[a, b]$ **où** $a < b$ et on note $X \sim \mathcal{U}[a, b]$, si X possède la densité

$$\frac{1}{b-a} 1_{[a,b]}(x).$$

- **loi exponentielle de paramètre** $\lambda > 0$ et on note $X \sim \mathcal{E}(\lambda)$ si X possède la densité

$$\lambda e^{-\lambda x} 1_{\{x>0\}}.$$

- **loi gaussienne (ou normale) de paramètres** $m \in \mathbb{R}$ **et** $\sigma^2 > 0$ et on note $X \sim \mathcal{N}_1(m, \sigma^2)$ si X possède la densité

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

  Dans le cas où $m = 0$ et $\sigma^2 = 1$, on dit aussi que X suit la loi normale centrée réduite.

- **loi de Cauchy de paramètre** $a > 0$ et on note $X \sim \mathcal{C}(a)$ si X possède la densité

$$\frac{a}{\pi} \frac{1}{x^2 + a^2}.$$

3.2.3 Espérance, variance

Définition 3.2.3. *La variable aléatoire $X : \Omega \to \mathbb{R}$ qui possède la densité p est dite*

– *intégrable si $\int_{\mathbb{R}} |x| p(x) dx < +\infty$ et alors on définit son espérance par*

$$\boxed{\mathbb{E}(X) = \int_{\mathbb{R}} x\ p(x)\ dx.}$$

– *de carré intégrable si* $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2\, p(x)\, dx < +\infty$ *et alors on définit sa variance par*

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

Les propriétés de l'espérance et de la variance sont les mêmes que dans le cas des variables aléatoires discrètes :

Propriétés 3.2.4. *1. Linéarité :* $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda\mathbb{E}(Y)$.

2. C.S. d'intégrabilité : si $\mathbb{P}(|X| \leq Y) = 1$ *et* Y *est intégrable alors* X *l'est aussi.*

3. Croissance : si X *et* Y *sont intégrables,* $\mathbb{P}(X \geq Y) = 1 \Longrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$.

Exercice 3.2.5. Calculer l'espérance et la variance d'une variable uniforme sur $[a, b]$, d'une variable exponentielle de paramètre λ, d'une variable de Cauchy de paramètre a et d'une variable normale centrée réduite.

3.2.4 Fonction de répartition

Définition 3.2.6. *Soit* $X : \Omega \to \mathbb{R}$ *une variable aléatoire réelle (qui ne possède pas nécessairement une densité). On appelle fonction de répartition de* X *la fonction* $F_X : x \in \mathbb{R} \to \mathbb{P}(X \leq x)$.

Exemple 3.2.7. – Si $X \sim \mathcal{B}(p)$, $F_X(x) = (1-p)1_{\{x\geq 0\}} + p1_{\{x\geq 1\}}$.
– Si $T \sim \mathcal{E}(\lambda)$, $F_T(t) = 1_{\{t\geq 0\}}(1 - \exp(-\lambda t))$.

Proposition 3.2.8. *Si la fonction de répartition* F_X *de la variable aléatoire réelle* X *est globalement continue et* C^1 *par morceaux (au sens où il existe un nombre fini de points* $x_1 < x_2 < \ldots < x_n$ *t.q.* F_X *est* C^1 *sur* $]-\infty, x_1[$, $]x_1, x_2[$,..., $]x_{n-1}, x_n[$, $]x_n, +\infty[$*) alors* X *possède la densité* F'_X.

Démonstration : Soit $a < b \in \bar{R}$. Comme $\{X \leq b\}$ est l'union disjointe de $\{X \leq a\}$ et de $\{a < X \leq b\}$,

$$\begin{aligned}\mathbb{P}(a < X \leq b) &= \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a)\\ &= F_X(b) - F_X(a)\\ &= \int_a^b F'_X(x)\, dx.\end{aligned}$$

D'après la définition 3.2.1, on conclut que X possède la densité $F'_X(x)$. □

Exercice 3.2.9. On considère un système constitué de n composants. On suppose que les durées de vie des composants sont des variables exponentielles $T_1, \ldots, T_n$ de paramètres respectifs $\lambda_1, \ldots, \lambda_n > 0$ et qu'elles sont indépendantes ce qui implique en particulier que $\forall (t_1, \ldots, t_n) \in \mathbb{R}^n$, les événements $\{T_1 \leq t_1\}, \ldots, \{T_n \leq t_n\}$ sont indépendants ainsi que les événements $\{T_1 > t_1\}, \ldots, \{T_n > t_n\}$.

1. On suppose que le système est en parallèle i.e. qu'il fonctionne lorsqu'un au moins des composants fonctionne. Exprimer sa durée de vie T en fonction des T_i. Déterminer la fonction de répartition de T et en déduire sa loi.
2. Même question dans le cas où le système est en série i.e. où il fonctionne seulement lorsque tous les composants fonctionnent.

3.3 Vecteurs aléatoires à densité

3.3.1 Définition

Définition 3.3.1. *On dit que le vecteur aléatoire $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ possède la densité $p : \mathbb{R}^d \to \mathbb{R}$ si*

$$\boxed{\forall D \text{ ouvert de } \mathbb{R}^d,\ \mathbb{P}(X \in D) = \int_D p(x)\, dx = \int_{\mathbb{R}^d} 1_D(x_1, \ldots, x_d)\, p(x_1, \ldots, x_d)\, dx_1 \ldots dx_d.}$$

Remarque 3.3.2.
- En toute rigueur pour pouvoir considérer $\mathbb{P}(X \in D)$ pour D ouvert, il faut demander que $\{X \in D\}$ soit dans la tribu $\mathcal{A}$.
- Une densité de probabilité p sur $\mathbb{R}^d$ est une fonction positive d'intégrale 1.

Exemple 3.3.3. Si $p_1, \ldots, p_d$ sont des densités de probabilité sur $\mathbb{R}$ alors $p(x) = p_1(x_1) \times p_2(x_2) \times \ldots \times p_d(x_d)$ est une densité de probabilité sur $\mathbb{R}^d$.

La caractérisation suivante par la *méthode de la fonction muette* est très utile en pratique :

Théorème 3.3.4. *Le vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ possède la densité p si et seulement si*

$$\boxed{\forall f : \mathbb{R}^d \to \mathbb{R} \text{ bornée},\ \mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\, p(x)\, dx.}$$

Il est possible et souvent utile de généraliser la condition nécessaire à des fonctions non bornées.

Remarque 3.3.5. Soit $X : \Omega \to \mathbb{R}^d$ de densité p et $f : \mathbb{R}^d \to \mathbb{R}$. Alors $f(X)$ est intégrable si et seulement si $\int_{\mathbb{R}^d} |f(x)|\, p(x)\, dx < +\infty$ et alors $\mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\, p(x)\, dx$.

3.3.2 Densité marginale

Soit $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ un vecteur aléatoire de densité p et $k < d$. Si D_k est un ouvert de $\mathbb{R}^k$,

$$\begin{aligned}
\mathbb{P}((X_1, \ldots, X_k) \in D_k) &= \mathbb{P}\left(X \in D_k \times \mathbb{R}^{d-k}\right) \\
&= \int_{D_k \times \mathbb{R}^{d-k}} p(x)\, dx \\
&= \int_{D_k} \left(\int_{\mathbb{R}^{d-k}} p(x_1, \ldots, x_d)\, dx_{k+1} \ldots dx_d \right) dx_1 \ldots dx_k.
\end{aligned}$$

On en déduit que le sous-vecteur $(X_1, \ldots, X_k)$ possède la densité

$$q(x_1, \ldots, x_k) = \int_{\mathbb{R}^{d-k}} p(x_1, \ldots, x_d)\, dx_{k+1} \ldots dx_d.$$

Proposition 3.3.6. *Soit X un vecteur aléatoire qui possède une densité. Alors tout sous-vecteur Y possède la densité* **marginale** *obtenue en intégrant celle de X sur les composantes qui ne figurent pas dans Y.*

3.3.3 Changement de variables

Avant d'énoncer un résultat abstrait, nous allons illustrer cette technique sur un exemple particulier. En effet, dans la pratique, pour éviter les erreurs, il vaut mieux reprendre la démarche explicitée sur cet exemple que d'appliquer le résultat abstrait.

Exercice résolu 3.3.7. *Soit* (X, Y) *un couple aléatoire de densité* $\lambda^2 \exp(-\lambda(x + y))1_{\{x\geq 0\}}1_{\{y\geq 0\}}$. *Quelle est la loi de* $(Z, W) = (X + Y, X - Y)$ *?*

Pour cela, on utilise la *méthode de la fonction muette* i.e. on se donne une fonction $f : \mathbb{R}^2 \to \mathbb{R}$ bornée et on cherche à mettre $\mathbb{E}(f(Z, W))$ sous la forme $\int_{\mathbb{R}^2} f(z, w)p(z, w)\, dz\, dw$, ce qui permettra de conclure par la condition suffisante du théorème 3.3.4 que (Z, W) possède la densité p. Mais bien sûr, il faut utiliser ce que l'on connaît c'est à dire la loi de (X, Y) :

$$\mathbb{E}(f(Z, W)) = \mathbb{E}(f(X + Y, X - Y)).$$

Soit $\varphi : (x, y) \in \mathbb{R}^2 \to (x+y, x-y) \in \mathbb{R}^2$. La fonction $g(x, y) = f \circ \varphi(x, y) = f(x+y, x-y)$ est une fonction continue bornée sur $\mathbb{R}^2$. Donc d'après la condition nécessaire du théorème 3.3.4,

$$\mathbb{E}(g(X, Y)) = \int_{\mathbb{R}^2} g(x, y)\lambda^2 \exp(-\lambda(x + y))1_{\{x\geq 0\}}1_{\{y\geq 0\}}\, dx\, dy,$$

égalité qui se récrit en explicitant g

$$\mathbb{E}(f(X + Y, X - Y)) = \int_{\mathbb{R}^2} f(x + y, x - y)\lambda^2 \exp(-\lambda(x + y))1_{\{x\geq 0\}}1_{\{y\geq 0\}}\, dx\, dy.$$

Ainsi

$$\mathbb{E}(f(Z, W)) = \int_0^{+\infty} \int_0^{+\infty} f(x + y, x - y)\lambda^2 \exp(-\lambda(x + y))dxdy.$$

Pour écrire le second membre comme une intégrale de $f(z, w)$ contre une fonction des variables (z, w) le bon outil est la formule de changement de variables vue au début de ce chapître. La fonction φ est une bijection C^1 ainsi que son inverse $(x, y) = \varphi^{-1}(z, w) = \left(\frac{z+w}{2}, \frac{z-w}{2}\right)$ de $O =]0, +\infty[\times]0, +\infty[$ sur $O' = \{(z, w) \in \mathbb{R}^2 : \ z > |w|\}$ (voir exemple 3.1.6). Par ailleurs,

$$\frac{D(x, y)}{D(z, w)} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

En prenant la valeur absolue du déterminant du second membre on obtient formellement que $\frac{dx\, dy}{dz\, dw} = \frac{1}{2}$ i.e. qu'il faut remplacer $dx\, dy$ par $\frac{1}{2}dz\, dw$. Ainsi

$$\begin{aligned}\mathbb{E}(f(Z, W)) &\int_{\{(z,w):z>|w|\}} f(z, w)\lambda^2 \exp\left(-\lambda\left(\frac{z + w}{2} + \frac{z - w}{2}\right)\right)\frac{1}{2}dzdw \\ &= \int_{\mathbb{R}^2} f(z, w)\frac{\lambda^2}{2} \exp(-\lambda z)1_{\{z>|w|\}}dzdw.\end{aligned}$$

On conclut par la condition suffisante du théorème 3.3.4 que (Z, W) possède la densité $\frac{\lambda^2}{2} \exp(-\lambda z)1_{\{z>|w|\}}$.
Par la formule des densité marginales, on en déduit que Z possède la densité

$$\int_{\mathbb{R}} \frac{\lambda^2}{2} \exp(-\lambda z)1_{\{z>|w|\}}dw = 1_{\{z>0\}}\lambda^2 z \exp(-\lambda z),$$

et W la densité

$$\int_{\mathbb{R}} \frac{\lambda^2}{2} \exp(-\lambda z)1_{\{z>|w|\}}dz = \frac{\lambda}{2} \exp(-\lambda|w|).$$

Le résultat abstrait qui se démontre en reprenant la démarche ci-dessus est le suivant

Proposition 3.3.8. *Soit $X : \Omega \to \mathbb{R}^d$ qui possède la densité $p(x)$ portée par un ouvert O de $\mathbb{R}^d$ au sens où $\int_O p(x)dx = 1$ et φ une bijection de O sur O' C^1 ainsi que son inverse φ^{-1}. Alors le vecteur $Y = \varphi(X)$ possède la densité*

$$q(y) = 1_{O'}(y)\, p(\varphi^{-1}(y))\, |\text{Jac } \varphi^{-1}(y)|.$$

3.3.4 Indépendance

Définition 3.3.9. *Les vecteurs aléatoires $X_1 : \Omega \to \mathbb{R}^{d_1}, \ldots, X_n : \Omega \to \mathbb{R}^{d_n}$ qui possèdent respectivement les densités p_1,...,p_n sont dits indépendants si $X = (X_1, \ldots, X_n)$ possède la densité produit $p_1(x_1) \times \ldots \times p_n(x_n)$.*

Remarque 3.3.10. Si $X = (X_1, \ldots, X_n)$ possède la densité $p(x) = cg_1(x_1) \times \ldots \times g_n(x_n)$ alors les vecteurs X_i sont indépendants et possèdent les densités

$$p_i(x_i) = \frac{1}{\int_{\mathbb{R}^{d_i}} g_i(y_i)dy_i}\ g_i(x_i).$$

En effet, comme l'intégrale de la densité p sur $\mathbb{R}^{d_1+\ldots+d_n}$ vaut 1,

$$1/c = \left(\int_{\mathbb{R}^{d_1}} g_1(y_1)dy_1\right) \times \ldots \times \left(\int_{\mathbb{R}^{d_n}} g_n(y_n)dy_n\right).$$

La formule des densités marginales permet de vérifier que la densité de X_i est p_i. Et comme $p(x) = p_1(x_1) \times \ldots \times p_n(x_n)$, les X_i sont indépendants.

La caractérisation suivante de l'indépendance de vecteurs aléatoires qui ne possèdent pas nécessairement des densités est parfois utile :

Proposition 3.3.11. *Les vecteurs aléatoires $X_1 : \Omega \to \mathbb{R}^{d_1}, \ldots, X_n : \Omega \to \mathbb{R}^{d_n}$ sont indépendants si et seulement si pour toutes fonctions $f_1 : \mathbb{R}^{d_1} \to \mathbb{R}, \ldots, f_n : \mathbb{R}^{d_n} \to \mathbb{R}$ telles que $f_1(X_1), \ldots, f_n(X_n)$ sont intégrables alors $f_1(X_1) \times \ldots \times f_n(X_n)$ est intégrable et*

$$\mathbb{E}(f_1(X_1) \times \ldots \times f_n(X_n)) = \prod_{i=1}^{n} \mathbb{E}(f_i(X_i)).$$

Remarque 3.3.12. Lorsque les vecteurs aléatoires $X_1, \ldots, X_n$ sont indépendants, alors $\forall m \in [1, n]$, $\forall 1 \leq d_1 < d_2 < \ldots < d_m \leq n$, les vecteurs $(X_1, X_2, \ldots, X_{d_1})$, $(X_{d_1+1}, \ldots, X_{d_2})$,...,$(X_{d_{m-1}+1}, \ldots, X_{d_m})$ et $(X_{d_m+1}, \ldots, X_n)$ sont indépendants, propriété démontrée dans la remarque 2.2.7 dans le cas discret.

3.3.5 Covariance

On rappelle que si Y et Z sont deux variables aléatoires de carré intégrable, comme $|YZ| \leq (Y^2 + Z^2)/2$, leur produit YZ est intégrable.

Définition 3.3.13. – *Soit Y et Z deux variables aléatoires réelles de carré intégrable (non nécessairement à densité). On appelle covariance de Y et Z le nombre* Cov (Y, Z) *défini par*

$$\text{Cov } (Y, Z) = \mathbb{E}((Y - \mathbb{E}(Y))(Z - \mathbb{E}(Z))).$$

– *Soit $X = (X_1, \ldots, X_d)$ un vecteur aléatoire à valeurs $\mathbb{R}^d$ (non nécessairement à densité) dont les composantes sont de carré intégrable. On appelle matrice de covariance du vecteur X la matrice K^X de dimension $d \times d$ et d'élément courant*

$$K^X_{i,j} = \text{Cov}\,(X_i, X_j) = \mathbb{E}((X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j)),\ 1 \leq i, j \leq d.$$

Propriétés 3.3.14. *1.* $\text{Cov}\,(Y, Y) = \text{Var}\,(Y)$ *et les éléments diagonaux de la matrice de covariance K^X sont les variances des composantes de X.*

2. $\text{Cov}\,(Y, Z) = \mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)$.

3. Si Y et Z sont indépendantes $\text{Cov}\,(Y, Z) = 0$.

4. La matrice de covariance K^X est symétrique et positive. En outre,

$$\forall u \in \mathbb{R}^d,\ \text{Var}\,(u.X) = u.K^X u \quad \text{où } u.v \text{ désigne le produit scalaire de } u \text{ et } v \text{ dans } \mathbb{R}^d.$$

En utilisant la propriété 4. pour $u = (1, 1, \ldots, 1)$, on obtient que $\text{Var}\,(X_1 + \ldots + X_d) = \sum_{i,j=1}^d \text{Cov}\,(X_i, X_j)$. Avec les propriétés 1. et 3., on en déduit que

Corollaire 3.3.15. *Si $X_1, \ldots, X_d$ sont des variables aléatoires réelles de carré intégrable* **indépendantes**, $\text{Var}\,(X_1 + \ldots + X_d) = \sum_{i=1}^d \text{Var}\,(X_i)$.

Démonstration : La preuve des trois premiers est laissée à titre d'exercice.
Soit $u = (u_1, \ldots, u_d) \in \mathbb{R}^d$.

$$\begin{aligned}
\text{Var}\,(u.X) &= \mathbb{E}\left(\left(\sum_{i=1}^d u_i X_i - \mathbb{E}\left(\sum_{i=1}^d u_i X_i\right)\right)^2\right) \\
&= \mathbb{E}\left(\sum_{i,j=1}^d u_i u_j (X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))\right) \\
&= \sum_{i,j=1}^d u_i u_j \text{Cov}\,(X_i, X_j) \\
&= u.K^X u.
\end{aligned}$$

Comme la variance d'une variable aléatoire est toujours positive, on conclut que $\forall u \in \mathbb{R}^d$, $u.K^X u \geq 0$. □

Remarque 3.3.16. – Notons que lorsque K^X est dégénérée, il existe u_0 non nul tel que $K^X u_0 = 0$. Alors $\text{Var}\,(u_0.X) = 0$, ce qui entraîne que $u_0.X$ est une constante i.e. X prend ses valeurs dans un sous espace affine strict de $\mathbb{R}^d$.

– La covariance de deux variables aléatoires indépendantes est nulle mais il ne suffit pas que $\text{Cov}\,(Y, Z) = 0$ pour que Y et Z soient indépendantes.
En effet si $Y \sim \mathcal{U}[-1, 1]$ et $Z = \varepsilon Y$ où ε est une variable aléatoire indépendante de Y t.q. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = \frac{1}{2}$ alors $\mathbb{E}(Y) = 0$ et $\mathbb{E}(YZ) = \mathbb{E}(\varepsilon Y^2) = \mathbb{E}(\varepsilon)\mathbb{E}(Y^2) =$

$0 \times \mathbb{E}(Y^2)$ si bien que Cov $(Y, Z) = 0$. Mais comme $\varepsilon^2 = 1$,

$$E(Y^2Z^2) = \mathbb{E}(\varepsilon^2 Y^4) = \mathbb{E}(Y^4) = \frac{1}{2}\int_{-1}^{1} y^4 dy = \frac{1}{5}$$

$$E(Y^2) = \frac{1}{2}\int_{-1}^{1} y^2 dy = \frac{1}{3} \text{ et } \mathbb{E}(Z^2) = \mathbb{E}(\varepsilon^2 Y^2) = \mathbb{E}(Y^2) = \frac{1}{3}$$

$$\text{si bien que } \mathbb{E}(Y^2Z^2) = \frac{1}{5} \neq \frac{1}{9} = \mathbb{E}(Y^2)\mathbb{E}(Z^2),$$

et les variables Y et Z ne sont donc pas indépendantes.

3.4 Lois béta, gamma et du khi 2

On note Γ la fonction gamma d'Euler :

$$a > 0 \to \Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

On vérifie facilement que cette fonction est t.q.

$$\forall a > 0,\ \Gamma(a+1) = a\Gamma(a) \quad \text{et} \quad \forall n \in \mathbb{N}^*,\ \Gamma(n) = (n-1)!$$

On dit que X suit la

- **loi gamma de paramètres** a **et** θ **où** $a, \theta > 0$ et on note $X \sim \Gamma(a, \theta)$ si X possède la densité
$$\frac{\theta^a}{\Gamma(a)}\ x^{a-1}\ e^{-\theta x}\ 1_{\{x>0\}}.$$
- **loi béta de paramètres** a **et** b **où** $a, b > 0$ et on note $X \sim \beta(a, b)$, si X possède la densité
$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ x^{a-1}(1-x)^{b-1} 1_{\{0<x<1\}}.$$

Exemple 3.4.1. – La loi uniforme sur $[0, 1]$ est la loi $\beta(1, 1)$.
– La loi exponentielle de paramètre θ est la loi $\Gamma(1, \theta)$.

Remarque 3.4.2. Le changement de variables $y = \theta x$ entraîne que

$$\int_0^{+\infty} \theta^a\ x^{a-1}\ e^{-\theta x}\ dx = \int_0^{+\infty} y^{a-1}\ e^{-y}\ dy = \Gamma(a).$$

Le fait que $\int_0^1 z^{a-1}(1-z)^{b-1} dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ sera vérifié au cours de la preuve de la proposition 3.4.4 ci-dessous.

Exercice résolu 3.4.3. *Soit $X_1, \ldots, X_n$ des variables aléatoires I.I.D. suivant la loi exponentielle de paramètre $\theta > 0$. Quelle est la loi de $S_n = X_1 + \ldots + X_n$?*

On se donne une fonction muette $f : \mathbb{R} \to \mathbb{R}$ bornée.

$$\begin{aligned}\mathbb{E}(f(S_n)) &= \mathbb{E}(f(X_1 + \ldots + X_n)) \\ &= \int_0^{+\infty} \ldots \int_0^{+\infty} f(x_1 + \ldots + x_n)\theta^n \exp(-\theta(x_1 + \ldots + x_n)) dx_1 \ldots dx_n,\end{aligned}$$

car par indépendance des X_i la densité de $(X_1, \ldots, X_n)$ est le produit des densités de ces variables. On effectue le changement de variables

$$(y_1, \ldots, y_n) = \varphi(x_1, \ldots, x_n) = (x_1, x_1 + x_2, \ldots, x_1 + \ldots + x_n).$$

La matrice jacobienne de φ est triangulaire avec des 1 sur la diagonale. Donc $dx_1 \ldots dx_n = dy_1 \ldots dy_n$ et

$$\begin{aligned}
\mathbb{E}(f(S_n)) &= \int_{0<y_1<y_2<\ldots<y_n} f(y_n)\theta^n \exp(-\theta y_n) dy_1 \ldots dy_n \\
&= \int_0^{+\infty} f(y_n)\theta^n \exp(-\theta y_n) \left(\int_{0<y_1<\ldots<y_n} dy_1 \ldots dy_{n-1} \right) dy_n \\
&= \int_0^{+\infty} f(y_n)\theta^n \exp(-\theta y_n) \frac{y_n^{(n-1)}}{(n-1)!} dy_n.
\end{aligned}$$

Donc $S_n \sim \Gamma(n, \theta)$.

Ainsi la somme de n-variables de loi $\Gamma(1, \theta)$ indépendantes suit la loi $\Gamma(n, \theta)$.

Ce résultat peut-être déduit de la proposition suivante par un raisonnement de récurrence :

Proposition 3.4.4. *Soit $X \sim \Gamma(a, \theta)$ et $Y \sim \Gamma(b, \theta)$ indépendantes. Alors $S = X + Y$ et $Z = \frac{X}{X+Y}$ sont deux variables aléatoires indépendantes distribuées respectivement suivant les lois $\Gamma(a + b, \theta)$ et $\beta(a, b)$.*

Démonstration : On utilise la méthode de la fonction muette : soit $f : \mathbb{R}^2 \to \mathbb{R}$ bornée.

$$\begin{aligned}
\mathbb{E}(f(S, Z)) &= \mathbb{E}\left(f\left(X + Y, \frac{X}{X+Y} \right) \right) \\
&= \int_0^{+\infty} \int_0^{+\infty} f\left(x + y, \frac{x}{x+y} \right) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} y^{b-1} e^{-\theta(x+y)} dx\, dy.
\end{aligned}$$

On effectue le changement de variables $(s, z) = \varphi(x, y) = \left(x + y, \frac{x}{x+y} \right)$. On a

$$\left\{ \begin{array}{l} s = x + y \\ z = \frac{x}{x+y} \end{array} \right. \iff \left\{ \begin{array}{l} x = sz \\ y = s(1-z) \end{array} \right. \quad \text{et} \quad \left\{ \begin{array}{l} x > 0 \\ y > 0 \end{array} \right. \iff \left\{ \begin{array}{l} sz > 0 \\ s(1-z) > 0 \end{array} \right. \iff \left\{ \begin{array}{l} s > 0 \\ 0 < z < 1 \end{array} \right.$$

$$\frac{D(x, y)}{D(s, z)} = \begin{pmatrix} z & 1-z \\ s & -s \end{pmatrix}, \text{ d'où } dx\, dy = s\, ds\, dz.$$

On en déduit par la formule de changement de variables que

$$\mathbb{E}(f(S, Z)) = \int_0^{+\infty} \int_0^1 f(s, z) \frac{\theta^{a+b}}{\Gamma(a)\Gamma(b)} (sz)^{a-1} (s(1-z))^{b-1} e^{-\theta s} s\, dz\, ds.$$

Ainsi (S, Z) possède la densité produit

$$\frac{\theta^{a+b}}{\Gamma(a+b)} s^{a+b-1} e^{-\theta s} 1_{\{s>0\}} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1} (1-z)^{b-1} 1_{\{a<z<b\}}.$$

Notons que cela démontre au passage que $\int_0^1 z^{a-1}(1-z)^{b-1} dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. □

Exercice résolu 3.4.5. *On appelle loi du khi 2 à n degrés de liberté et on note $\chi^2(n)$ la loi de $X_1^2+\ldots+X_n^2$ où $X_1,\ldots,X_n$ sont n variables normales centrées réduites indépendantes. Quelle est la densité de cette loi ?*

On commence par traiter le cas $n=1$. Pour $f:\mathbb{R}\to\mathbb{R}$ bornée,

$$\begin{aligned}\mathbb{E}(f(X_1^2)) &= \int_{\mathbb{R}} f(x^2)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx = 2\int_0^{+\infty} f(x^2)\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx \\ &= \int_0^{+\infty} f(y)\frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}dy, \quad \text{en posant } y=x^2.\end{aligned}$$

Ainsi $\chi^2(1)\equiv\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$.
Avec la proposition 3.4.4, on conclut par récurrence que $\chi^2(n)\equiv\Gamma\left(\frac{n}{2},\frac{1}{2}\right)$ est la loi de densité

$$\frac{1}{2^{n/2}\Gamma(\frac{n}{2})}\, x^{\frac{n}{2}-1}\, e^{-\frac{x}{2}}1_{\{x>0\}}.$$

3.5 Exercices

Exercice 3.5.1. Une cerise est placée sur la circonférence d'un gateau rond que l'on partage en 2 au hasard en pratiquant deux découpes suivant des rayons.
Si on prend la position de la cerise comme origine des angles, les positions U et V des deux coups de couteau sont des variables uniformément réparties sur $[0,2\pi]$ indépendantes. Exprimer la taille T de la part contenant la cerise, calculer son espérance puis déterminer la probabilité pour cette part soit plus grosse que l'autre. Quelle doit être la décision d'un gourmand qui doit choisir entre la part avec la cerise et la part sans la cerise avant le découpage ?

Exercice 3.5.2. On suppose que la durée de vie d'un équipement A est une variable aléatoire réelle S possédant la densité $\alpha e^{-\alpha s}1_{\{s\geq 0\}}$, que celle de l'équipement B est une variable T possédant la densité $\beta^2 te^{-\beta t}1_{\{t\geq 0\}}$ avec $\alpha,\beta>0$ et que S et T sont indépendantes.

1. Que vaut $P(S\geq t+s \mid S\geq t)$ pour $s,t\geq 0$? Est-il judicieux de changer l'équipement A au temps t s'il fonctionne toujours ?
2. Quelle est la probabilité pour que l'équipement A tombe en panne avant l'équipement B ?

Exercice 3.5.3. Soit X une variable qui suit la loi de Cauchy de paramètre $a>0$. Quelle est la loi de $Y=1/X$?

Exercice 3.5.4. Soit (X,Y) un couple de variables normales centrées réduites indépendantes. Déterminer la loi de $(Z,W)=(X/Y,X+Y)$ et en déduire la loi de Z.

Exercice 3.5.5. Soit X et Y deux variables aléatoires réelles exponentielles de paramètre λ ($\lambda>0$) indépendantes.

1. Quelle est la loi de X^2 ?
2. Quelle est celle de

$$V=\begin{cases} X & si\ \ 0\leq X\leq 1 \\ 2X & si\ \ X>1\end{cases}$$

3. Déterminer la loi de
$$(Z,S)=\left(\frac{X}{X+Y},X+Y\right).$$
Les variables Z et S sont-elles indépendantes ?

Exercice 3.5.6. Les durées de vie S et T de deux machines suivent des lois exponentielles de paramètres respectifs α et β et sont indépendantes. Quelle est la loi du couple $(\min(S,T),|T-S|)$? (On rappelle que $|T-S|=\max(S,T)-\min(S,T)$.) Les variables $\min(S,T)$ et $|T-S|$ sont-elles indépendantes ?

Exercice 3.5.7. On coupe un bâton de longueur 1 au hasard en trois morceaux : les abcisses U et V des découpes sont supposées indépendantes et uniformément réparties sur $[0,1]$. Calculer la probabilité pour que l'on puisse faire un triangle avec les trois morceaux (on peut faire un triangle avec trois segments de longueur l_1, l_2 et l_3 ssi $l_1\leq l_2+l_3$, $l_2\leq l_3+l_1$ et $l_3\leq l_1+l_2$).

Exercice 3.5.8. Le cycle d'un feu de circulation est le suivant : le feu est vert sur l'intervalle $[0,v]$ et orange ou rouge sur $[v,v+r]$ avec $v,r>0$. L'instant d'arrivée U d'un automobiliste est supposé uniformément réparti sur le cycle $[0,r+v]$.

1. Exprimer en fonction de U le temps T d'attente de cet automobiliste au feu dans le cas où aucun véhicule ne se trouve devant le feu à l'instant ou il arrive.
2. Montrer que pour $f:\mathbb{R}\to\mathbb{R}$ bornée,
$$\mathbb{E}(f(T))=\frac{v}{r+v}f(0)+\frac{1}{r+v}\int_0^r f(s)ds.$$
3. Déterminer la probabilité $\mathbb{P}(T=0)$ pour que l'automobiliste passe immédiatement.
4. Montrer que si $t\neq 0$, $\mathbb{P}(T=t)=0$.
5. Conclure que le temps d'attente T n'est ni une variable aléatoire discrète ni une variable aléatoire à densité.

Exercice 3.5.9. Montrer que la probabilité de tomber sur une fève qui a la forme d'un disque de rayon r lors de la découpe d'un rayon d'une galette des rois de rayon $R>2r$ est égale à
$$\frac{r^2}{2(R-r)^2}+\frac{r\sqrt{(R-r)^2-r^2}}{\pi(R-r)^2}+\frac{1}{\pi}\arcsin\left(\frac{r}{R-r}\right).$$
<u>Indication :</u> cette probabilité est égale à celle pour que la fève dont le centre est uniformément réparti sur le disque de rayon $R-r$ touche un rayon fixé de la galette.

Exercice 3.5.10. Au cours d'un jeu télévisé, un candidat tire au sort entre 2 enveloppes contenant respectivement les sommes x_1 et x_2 ($x_1>x_2>0$), qu'il ne connaît pas. Après avoir examiné le contenu X de son enveloppe, il a le choix entre conserver ce contenu et effectuer un échange avec l'autre enveloppe. On veut étudier la stratégie suivante : on se donne T une variable exponentielle de paramètre 1 indépendante du tirage au sort et on change d'enveloppe seulement lorsque $T>X$. Calculer la probabilité d'obtenir la somme la plus élevée x_1 en suivant cette stratégie. Est-elle supérieure ou inférieure à $1/2$? Trouvez-vous ce résultat intuitif ?

Exercice 3.5.11. Soit R et Θ, deux variables aléatoires indépendantes uniformément réparties respectivement sur $[0,1]$ et $[0,2\pi]$. Le vecteur

$$(X,Y) = (R\cos\Theta, R\sin\Theta)$$

est-il uniformément réparti sur le disque unité $D = \{(x,y) : x^2 + y^2 \leq 1\}$ i.e. possède-t-il la densité $\frac{1}{\pi}1_D(x,y)$?

3.6 Résumé

- $X : \Omega \to \mathbb{R}^d$ **possède la densité** $p : \mathbb{R}^d \to \mathbb{R}_+$ si

$$\forall D \text{ ouvert de } \mathbb{R}^d,\ \mathbb{P}(X \in D) = \int_D p(x)\ dx = \int_{\mathbb{R}^d} 1_D(x)\ p(x)\ dx.$$

- **Densité marginale :** Si $(X, Y) : \Omega \to \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ possède la densité $p(x, y)$ alors X a pour densité

$$\boxed{q(x) = \int_{\mathbb{R}^{d_2}} p(x, y)\ dy.}$$

- **Espérance, variance :** $X : \Omega \to \mathbb{R}$ de densité p est dite
  - intégrable si $\int_{\mathbb{R}} |x| p(x) dx < +\infty$ et alors $\boxed{\mathbb{E}(X) = \int_{\mathbb{R}} x\ p(x)\ dx.}$
  - de carré intégrable si $\mathbb{E}(X^2) = \int_{\mathbb{R}} x^2\ p(x)\ dx < +\infty$ et alors

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

  Propriétés :
  1. Linéarité : $\mathbb{E}(X + \lambda Y) = \mathbb{E}(X) + \lambda \mathbb{E}(Y)$.
  2. Croissance : $\mathbb{P}(X \geq Y) = 1 \Longrightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$.
  3. $X_1, \ldots, X_n$ **indépendantes** $\Longrightarrow$ Var $(X_1 + \ldots + X_n) = \sum_{i=1}^n$ Var (X_i).

- **Fonction muette :** $X : \Omega \to \mathbb{R}^d$ possède la densité p

$$\Longleftrightarrow \boxed{\forall f : \mathbb{R}^d \to \mathbb{R} \text{ bornée }, \ \mathbb{E}(f(X)) = \int_{\mathbb{R}^d} f(x)\ p(x)\ dx.}$$

- **Densités usuelles sur $\mathbb{R}$:**

Nom	**Densité**	$\mathbb{E}(X)$	Var (X)
uniforme $\mathcal{U}[a, b]$	$\frac{1}{b-a}\ 1_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
exponentielle $\mathcal{E}(\lambda)$	$\lambda \exp(-\lambda x) 1_{\{x \geq 0\}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
normale $\mathcal{N}_1(m, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}}\ \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$	m	σ^2
Cauchy $\mathcal{C}(a)$	$\frac{a}{\pi}\ \frac{1}{x^2+a^2}$	non intég	-

- **Covariance :** si X et Y sont deux variables réelles de carré intégrable,

$$\boxed{\text{Cov }(X, Y) \stackrel{\text{def}}{=} \mathbb{E}\left((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).}$$

Chapitre 4

Simulation

Pour effectuer des simulations probabilistes sur ordinateur, on utilise un *générateur de nombres pseudo-aléatoires*. Un tel générateur retourne une suite $(x_n)_n$ de nombres réels compris entre 0 et 1. Ces réels sont calculés par un algorithme déterministe mais imitent une réalisation d'une suite de variables aléatoires indépendantes et indentiquement distribuées suivant la loi uniforme sur $[0,1]$. Le bon comportement de la suite est vérifié à l'aide de tests statistiques.
Une méthode couramment utilisée pour construire la suite $(x_n)_n$ est la *congruence* : $x_n = y_n/N$ où les y_n sont des entiers compris entre 0 et $N-1$ calculés grâce à la relation de récurrence

$$y_{n+1} = (ay_n + b) \mod (N).$$

Le choix des entiers a, b, N est fait de façon à ce que la période du générateur (toujours plus petite que $N-1$) soit aussi grande que possible et que les propriétés de la suite $(x_n)_n$ soient proches de celles d'une réalisation d'une suite de variables I.I.D. suivant la loi uniforme sur $[0,1]$.

Dans tout ce chapitre, le point de vue et le suivant : on suppose que l'on dispose d'un bon générateur de nombres pseudo-aléatoires et on se demande comment à partir d'une suite $(U_i)_{i\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant la loi uniforme sur $[0,1]$ construire une variable aléatoire de loi donnée, avec une attention particulière pour les lois usuelles introduites dans les chapitres précédents.

Remarque 4.0.1. Il existe des dispositifs qui à partir de phénomènes physiques aléatoires comme les temps de désintégration atomique de matériaux radioactifs permettent de transmettre à l'ordinateur des nombres aléatoires. Mais le coût de ces dispositifs reste prohibitif.

4.1 Simulation de variables aléatoires discrètes

4.1.1 Loi de Bernoulli de paramètre $p \in [0,1]$

Si $U \sim \mathcal{U}([0,1])$ alors

$$\boxed{X = 1_{\{U \leq p\}} \sim \mathcal{B}(p).}$$

En effet X prend les valeurs 0 ou 1 et

$$\mathbb{P}(X = 1) = \mathbb{P}(U \leq p) = \int_0^1 1_{\{u \leq p\}} du = p.$$

4.1.2 Loi binômiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$

Si $U_1, \ldots, U_n$ sont n variables uniformes sur $[0, 1]$ indépendantes alors

$$\boxed{X = 1_{\{U_1 \leq p\}} + \ldots + 1_{\{U_n \leq p\}} = \sum_{i=1}^{n} 1_{\{U_i \leq p\}} \sim \mathcal{B}(n, p).}$$

D'après ce qui précède, les variables $1_{\{U_i \leq p\}}$, $1 \leq i \leq n$ sont des variables de Bernoulli de paramètre p indépendantes. La variable aléatoire X, somme de ces n variables suit donc la loi binômiale de paramètres n et p.

4.1.3 Loi géométrique de paramètre $p \in]0, 1]$

C'est la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes avec propabilité de succès p. Ainsi, si les $(U_i)_{i \geq 1}$ sont des variables uniformes sur $[0, 1]$ indépendantes

$$\boxed{N = \inf\{i \geq 1 : \ U_i \leq p\} \sim \mathcal{G}eo(p).}$$

Remarque 4.1.1. Nous verrons dans l'exercice 4.3.1 une méthode qui permet de simuler une variable géométrique de paramètre p à l'aide d'une seule variable uniforme sur $[0, 1]$. Cette méthode qui repose sur l'interprétation de la loi géométrique comme version discrète de la loi exponentielle est beaucoup rapide en terme de temps de calcul.

4.1.4 Simulation suivant une loi discrète quelconque

Nous venons de voir des méthodes spécifiques pour simuler des variables aléatoires suivant des lois discrètes usuelles.
Il est cependant toujours possible d'obtenir une variable qui prend les valeurs $(x_i)_{i \in \mathbb{N}^*}$ avec probabilités respectives $(p_i)_{i \in \mathbb{N}^*}$ (avec les $p_i \geq 0$ t.q. $\sum_{i \in \mathbb{N}^*} p_i = 1$) à l'aide d'une seule variable U uniforme sur $[0, 1]$ en posant

$$\boxed{X = x_1 1_{\{U \leq p_1\}} + x_2 1_{\{p_1 < U \leq p_1 + p_2\}} + \ldots + x_i 1_{\{p_1 + \ldots + p_{i-1} < U \leq p_1 + \ldots + p_i\}} + \ldots.}$$

Remarque 4.1.2. Pour implémenter cette méthode très générale, il faut programmer une boucle sur i avec comme test d'arrêt $p_1 + \ldots + p_i \geq U$. Cela peut s'avérer coûteux en temps de calcul lorsque la série de terme général p_i converge lentement vers 1.

Remarque 4.1.3. Outre cette méthode générale, pour simuler une variable qui suit la loi de Poisson de paramètre $\lambda > 0$, on peut recourir à l'approche présentée dans l'exercice 4.3.2 qui repose sur les liens entre variables de Poisson et variables exponentielles de même paramètre λ.

4.2 Simulation de variables aléatoires à densité

4.2.1 Loi uniforme sur $[a, b]$ avec $a < b \in \mathbb{R}$

Si U est une variable uniforme sur $[0, 1]$ alors

$$\boxed{X = a + (b - a)U \sim \mathcal{U}[a, b].}$$

4.2.2 Méthode d'inversion de la fonction de répartition

Soit p une densité de probabilité sur $\mathbb{R}$ strictement positive et $F(x) = \int_{-\infty}^{x} p(y)dy$ la fonction de répartition associée. Comme F est continue et strictement croissante sur $\mathbb{R}$ et vérifie $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$, elle admet une fonction inverse $F^{-1} :]0, 1[\to \mathbb{R}$.

Proposition 4.2.1. *Si $U \sim \mathcal{U}[0, 1]$, alors $X = F^{-1}(U)$ possède la densité p.*

Démonstration : Soit $a < b \in \mathbb{R}$. Comme F est strictement croissante, $\{a < X \leq b\} = \{a < F^{-1}(U) \leq b\} = \{F(a) < U \leq F(b)\}$. Donc

$$\mathbb{P}(a < X \leq b) = \int_0^1 1_{\{F(a)<u\leq F(b)\}} du = F(b) - F(a) = \int_a^b p(y)dy,$$

ce qui d'après la définition 3.2.1 signifie que X possède la densité p. □

Exemple 4.2.2. La loi de Cauchy de paramètre $a > 0$ est la loi de densité $p(y) = \frac{a}{\pi(y^2+a^2)}$ sur $\mathbb{R}$. La fonction de répartition associée est

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} p(y)dy = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{(y/a)^2 + 1} \frac{dy}{a} \\ &= \frac{1}{\pi}\int_{-\infty}^{x/a} \frac{1}{1 + z^2} dz = \frac{1}{\pi}[\arctan(z)]_{-\infty}^{x/a} = \frac{1}{\pi}\arctan(x/a) + 1/2. \end{aligned}$$

Pour inverser la fonction de répartition, on se donne $y \in]0, 1[$ et on cherche x t.q. $F(x) = y$:

$$F(x) = y \Leftrightarrow x = a\tan(\pi(y - 1/2)).$$

On conclut donc que si $U \sim \mathcal{U}[0, 1]$,

$$\boxed{X = a\tan(\pi(U - 1/2)) \sim \mathcal{C}(a).}$$

Exemple 4.2.3. La loi exponentielle de paramètre $\lambda > 0$ est la loi de densité $\lambda e^{-\lambda z} 1_{\{z>0\}}$. La fonction de répartition associée est

$$F(x) = \int_{-\infty}^{x} \lambda e^{-\lambda z} 1_{\{z>0\}} dz = 1_{\{x>0\}}(1 - e^{-\lambda x}).$$

Pour $y \in]0, 1[$, on a

$$F(x) = y \Leftrightarrow x = -\frac{1}{\lambda}\ln(1 - y).$$

On en déduit que si $U \sim \mathcal{U}[0,1]$, $-\frac{1}{\lambda}\ln(1-U) \sim \mathcal{E}(\lambda)$.
On peut apporter une simplification en remarquant que $1-U \overset{\mathcal{L}}{=} U$ ($1-U$ a même loi que U), ce qui entraîne que $-\frac{1}{\lambda}\ln(1-U) \overset{\mathcal{L}}{=} -\frac{1}{\lambda}\ln(U)$ et permet de conclure que

$$\boxed{S = -\frac{1}{\lambda}\ln(U) \sim \mathcal{E}(\lambda).}$$

4.2.3 Méthode polaire pour la loi normale centrée réduite

Proposition 4.2.4. *Soit R de loi exponentielle de paramètre $1/2$ et Θ de loi uniforme sur $[0,2\pi]$ indépendantes. Alors*

$$X = \sqrt{R}\,\cos(\Theta) \quad \textit{et} \quad Y = \sqrt{R}\,\sin(\Theta)$$

sont des variables normales centrées réduites (densité $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$) indépendantes.

Démonstration : On applique la méthode de la fonction muette. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ bornée. On a

$$\begin{aligned}\mathbb{E}(f(X,Y)) &= \mathbb{E}(f(\sqrt{R}\,\cos(\Theta), \sqrt{R}\,\sin(\Theta))) \\ &= \frac{1}{4\pi}\int_0^{+\infty}\int_0^{2\pi} f(\sqrt{r}\,\cos(\theta), \sqrt{r}\,\sin(\theta))e^{-r/2}drd\theta.\end{aligned}$$

Le changement de variable $(x,y) = \varphi(r,\theta) = (\sqrt{r}\,\cos(\theta), \sqrt{r}\,\sin(\theta))$ est une bijection C^1 ainsi que son inverse de $]0,+\infty[\times]0,2\pi[$ sur $\mathbb{R}^2 \setminus \{(x,0) : x \geq 0\}$. Sa matrice jacobienne est

$$\frac{D(x,y)}{D(r,\theta)} = \begin{pmatrix} \cos(\theta)/(2\sqrt{r}) & -\sqrt{r}\,\sin(\theta) \\ \sin(\theta)/(2\sqrt{r}) & \sqrt{r}\,\cos(\theta) \end{pmatrix},$$

ce qui entraîne que $dxdy = \frac{1}{2}drd\theta$. On conclut par la formule de changement de variables que

$$\mathbb{E}(f(X,Y)) = \frac{1}{2\pi}\int_{\mathbb{R}^2} f(x,y)e^{-\frac{x^2+y^2}{2}}dxdy.$$

□

D'après ce qui précède si (U_1,U_2) est un couple de variables uniformes sur $[0,1]$ indépendantes : $(-2\ln(U_1), 2\pi U_2) \overset{\mathcal{L}}{=} (R,\Theta)$, ce qui entraîne que

$$\left(\sqrt{-2\ln(U_1)}\,\cos(2\pi U_2), \sqrt{-2\ln(U_1)}\,\sin(2\pi U_2))\right) \overset{\mathcal{L}}{=} (X,Y).$$

On conclut donc que

$$\boxed{\begin{cases} \sqrt{-2\ln(U_1)}\,\cos(2\pi U_2) \\ \sqrt{-2\ln(U_1)}\,\sin(2\pi U_2) \end{cases} \text{ sont deux variables normales centrées réduites indépendantes.}}$$

On en déduit facilement par changement de variable que

Corollaire 4.2.5. *Soit $m,\sigma \in \mathbb{R}$ et U_1,U_2 deux variables uniformes sur $[0,1]$ indépendantes. Alors $m + \sigma\sqrt{-2\ln(U_1)}\cos(2\pi U_2) \sim \mathcal{N}_1(m,\sigma^2)$.*

4.2.4 Méthode du rejet

On souhaite simuler une variable aléatoire qui possède la densité p sur $\mathbb{R}^d$ dans le cas où il existe une densité q sur $\mathbb{R}^d$ suivant laquelle on sait simuler et une constante $k > 0$ t.q.

$$\forall x \in \mathbb{R}^d,\ p(x) \leq kq(x).$$

On se donne une suite $(Y_i, U_i)_{i\geq 1}$ I.I.D. avec Y_1 de densité q et U_1 uniforme sur $[0,1]$ indépendantes.

Le principe de la méthode du rejet est le suivant :

1. Tant que $kU_i > \frac{p(Y_i)}{q(Y_i)}$ cela signifie que le rapport au point Y_i entre la densité p suivant laquelle on souhaite simuler et la densité q de Y_i est en certain sens trop petit et on rejette le couple d'indice i (d'où le nom de la méthode) et on passe au couple d'indice $i+1$.
2. Si $kq(Y_i)U_i \leq p(Y_i)$, on pose $X = Y_i$.

D'un point de vue plus formalisé on pose

$$N = \inf\{i \geq 1 :\ kq(Y_i)U_i \leq p(Y_i)\} \ \text{ et } \ X = Y_N.$$

Il faut noter que N est une variable aléatoire : $X(\omega) = Y_{N(\omega)}(\omega)$.

Théorème 4.2.6. *La variable aléatoire N suit la loi géométrique de paramètre $1/k$ (elle est donc finie avec probabilité 1). Elle est indépendante du couple $(Y_N, kq(Y_N)U_N)$ qui est uniformément réparti sur $\{(x,z) \in \mathbb{R}^d \times \mathbb{R} :\ 0 \leq z \leq p(x)\}$.*
En particulier $X = Y_N$ possède la densité p.

Démonstration : Commençons par déterminer la loi de $(Y_i, kq(Y_i)U_i)$. Pour cela on se donne une fonction muette $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ bornée :

$$\begin{aligned}\mathbb{E}(f(Y_i, kq(Y_i)U_i)) &= \int_{\mathbb{R}^d}\left(\int_0^1 f(y, kq(y)u)q(y)du\right)dy = \int_{\mathbb{R}^d}\left(\frac{1}{k}\int_0^{kq(y)} f(y,z)dz\right)dy \\ &= \frac{1}{k}\int_{\mathbb{R}^d}\int_{\mathbb{R}} f(y,z)1_{\{0\leq z\leq kq(y)\}}dydz. \end{aligned} \tag{4.1}$$

Ainsi $(Y_i, kq(Y_i)U_i))$ est une variable aléatoire uniformément répartie sur

$$D_{kq} = \{(y,z) \in \mathbb{R}^d \times \mathbb{R} :\ 0 \leq z \leq kq(y)\}$$

qui a pour volume k ($\int_{\mathbb{R}^d} q(y)dy = 1$).
L'événement $kq(Y_i)U_i \leq p(Y_i)$ s'écrit également $(Y_i, kq(Y_i)U_i) \in D_p$ où

$$D_p = \{(x,z) \in \mathbb{R}^d \times \mathbb{R} :\ 0 \leq z \leq p(x)\} \subset D_{kq}$$

a pour volume 1.
On en déduit que la probabilité pour que $kq(Y_i)U_i \leq p(Y_i)$ est égale au rapport des volumes soit $1/k$. Ainsi la variable aléatoire N est le temps de premier succès dans une

suite d'expériences indépendantes avec probabilité de succès $1/k$ et suit la loi géométrique de paramètre $1/k$. En particulier, elle est finie avec probabilité 1 c'est-à-dire que

$$\mathbb{P}\left(\sum_{n\in\mathbb{N}^*} 1_{\{N=n\}} = 1\right) = 1. \tag{4.2}$$

On se donne maintenant des fonctions muettes bornées $g : \mathbb{N}^* \to \mathbb{R}$ et $f : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$. Pour traiter le caractère aléatoire de l'indice N dans le calcul de $\mathbb{E}(g(N)f(Y_N, kq(Y_N)U_N))$, on utilise (4.2) qui permet de décomposer l'espérance suivant les valeurs prises par N :

$$\begin{aligned}
\mathbb{E}\left(g(N)f(Y_N, kq(Y_N)U_N)\right) &= \mathbb{E}\left(\sum_{n\in\mathbb{N}^*} 1_{\{N=n\}} g(N) f(Y_N, kq(Y_N)U_N)\right) \\
&= \sum_{n\in\mathbb{N}^*} \mathbb{E}\left(1_{\{N=n\}} g(n) f(Y_n, kq(Y_n)U_n)\right) \quad \text{par linéarité de l'espérance} \\
&= \sum_{n\in\mathbb{N}^*} g(n)\mathbb{E}\left(1_{\{kq(Y_1)U_1>p(Y_1)\}} \dots 1_{\{kq(Y_{n-1})U_{n-1}>p(Y_{n-1})\}} 1_{\{kq(Y_n)U_n\le p(Y_n)\}} f(Y_n, kq(Y_n)U_n)\right) \\
&= \sum_{n\in\mathbb{N}^*} g(n)\mathbb{P}(kq(Y_1)U_1 > p(Y_1))^{n-1}\mathbb{E}\left(1_{\{kq(Y_n)U_n\le p(Y_n)\}} f(Y_n, kq(Y_n)U_n)\right) \quad \text{car } (Y_i, U_i) \text{ I.I.D.} \\
&= \sum_{n\in\mathbb{N}^*} g(n)\left(1-\frac{1}{k}\right)^{n-1}\frac{1}{k}\int_{\mathbb{R}^d}\int_{\mathbb{R}} 1_{\{0\le z\le p(y)\}} f(y,z) 1_{\{0\le z\le kq(y)\}} dz dy \quad \text{d'après (4.1)} \\
&= \sum_{n\in\mathbb{N}^*} g(n)\frac{1}{k}\left(1-\frac{1}{k}\right)^{n-1}\int_{\mathbb{R}^d}\int_{\mathbb{R}} f(y,z) 1_{\{0\le z\le p(y)\}} dz dy.
\end{aligned}$$

Ainsi le couple $(Y_N, kq(Y_N)U_N)$ est uniformément réparti sur D_p et indépendant de N. Pour conclure, on calcule la densité de $X = Y_N$ par la formule des densités marginales :

$$\int_{\mathbb{R}} 1_{\{0\le z\le p(x)\}} dz = p(x).$$

□

Remarque 4.2.7. Le nombre moyen de tirages (Y_i, U_i) nécessaires pour générer X est $\mathbb{E}(N) = k$. Pour diminuer le temps de calcul, on a donc bien sûr intérêt à privilégier le choix d'une densité q telle $\sup_{x\in\mathbb{R}^d} \frac{p(x)}{q(x)}$ soit aussi petit que possible et à poser $k = \sup_{x\in\mathbb{R}^d} \frac{p(x)}{q(x)}$.

Exercice résolu 4.2.8. *On souhaite simuler une variable aléatoire qui possède la loi gamma de paramètres* $3/2$ *et* 1 *i.e. qui possède la densité*

$$p(x) = \frac{2}{\sqrt{\pi}}\sqrt{x}\, e^{-x} 1_{\{x\ge 0\}}$$

par la méthode du rejet avec $q(x) = \lambda e^{-\lambda x} 1_{\{x\ge 0\}}$. *Quel est le choix de* λ *optimal ? Que vaut alors* $\mathbb{E}(N)$ *?*

On veut trouver $\lambda^* \in]0, +\infty[$ qui minimise $f(\lambda) = \sup_{x\ge 0} \frac{\sqrt{x}e^{(\lambda-1)x}}{\lambda}$.
Pour $\lambda \ge 1$, $f(\lambda) = +\infty$. Déterminons maintenant $f(\lambda)$ pour $\lambda \in]0, 1[$. Le supremum en x est réalisé en $x(\lambda)$ qui annule la dérivée i.e. qui vérifie

$$\left(\frac{1}{2\sqrt{x}} + (\lambda-1)\sqrt{x}\right) e^{(\lambda-1)x} = 0,$$

ce qui entraîne $x(\lambda) = \frac{1}{2(1-\lambda)}$.
Ainsi $\forall \lambda \in]0,1[,\ f(\lambda) = 1/\sqrt{2e\lambda^2(1-\lambda)}$ et pour minimiser $f(\lambda)$ il suffit de maximiser $\lambda^2(1-\lambda)$ ce qui conduit à $\lambda^* = 2/3$.
En prenant $k = \frac{2}{\sqrt{\pi}} f(2/3) = 3^{3/2}/\sqrt{2\pi e}$, on a $\mathbb{E}(N) = k \sim 1,257$.

4.3 Exercices

Exercice 4.3.1. Pour $x \in \mathbb{R}$, on note $[x]$ la partie entière de x.
Soit X une variable exponentielle de paramètre $\lambda > 0$. Quelle est la loi de $N = 1 + [X]$?
Pour $p \in]0,1[$, en déduire λ t.q.

$$1 + \left[-\frac{1}{\lambda}\ln(U)\right] \sim \mathcal{G}eo(p).$$

Exercice 4.3.2.

1. Soit $(X_i)_{i\geq 1}$ une suite de variables exponentielles de paramètre $\lambda > 0$ indépendantes. Pour $n \in \mathbb{N}^*$, calculer $P(X_1 + \ldots + X_n \leq 1 < X_1 + \ldots + X_{n+1})$. Quelle est la loi de

$$N = \sum_{n\in\mathbb{N}^*} n 1_{\{X_1+\ldots+X_n \leq 1 < X_1+\ldots+X_{n+1}\}} ?$$

2. En déduire que si $(U_i)_{i\geq 1}$ est une suite de variables I.I.D. suivant la loi uniforme sur $[0,1]$,

$$\inf\{n \in \mathbb{N} :\ U_1 \times U_2 \times \ldots \times U_{n+1} \leq e^{-\lambda}\}$$

suit la loi de Poisson de paramètre λ.

Exercice 4.3.3. Soit (X,Y) un vecteur uniformément réparti sur le disque D centré en 0 et de rayon 1 (densité $1_{\{x^2+y^2\leq 1\}}/\pi$).

1. Quelle est la loi de

$$(R,\Theta) = \left(X^2 + Y^2\ ,\ 2\arctan\left(\frac{Y}{X + \sqrt{X^2+Y^2}}\right)\right) ?$$

2. Soit

$$(Z,W) = \left(X\sqrt{\frac{-2\ln(X^2+Y^2)}{X^2+Y^2}}, Y\sqrt{\frac{-2\ln(X^2+Y^2)}{X^2+Y^2}}\right).$$

Exprimer (Z,W) en fonction de (R,Θ) et en déduire sans calcul la loi de (Z,W).

3. Comment simuler un couple uniformément réparti sur le disque D à partir d'une suite I.I.D. de couples uniformément répartis sur le carré $[-1,1]^2$. En déduire une méthode pour simuler des variables normales centrées réduites qui ne fait pas appel aux fonctions trigonométriques.

Exercice 4.3.4. Soit $a, b > 0$. On appelle loi béta de paramètres (a,b) la loi de densité

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ x^{a-1}\ (1-x)^{b-1} 1_{\{0<x<1\}} \text{ où } \forall c > 0,\ \Gamma(c) = \int_0^{+\infty} e^{-x} x^{c-1} dx.$$

1. Montrer que si (U, V) suit la loi uniforme sur le domaine

$$D = \{(u, v) \in \mathbb{R}^2,\ u > 0,\ v > 0,\ u^{\frac{1}{a}} + v^{\frac{1}{b}} < 1\},$$

(i.e. (U, V) possède la densité $1_{\{(u,v)\in D\}}/|D|$ où $|D|$ est la surface de D) alors

$$X = \frac{U^{\frac{1}{a}}}{U^{\frac{1}{a}} + V^{\frac{1}{b}}}$$

suit la loi béta de paramètres (a, b). (Il suffit de montrer que X possède une densité proportionnelle à $x^{a-1}(1-x)^{b-1}1_{\{0<x<1\}}$.)

2. En déduire une méthode pour simuler une variable aléatoire de loi béta.

Exercice 4.3.5. Soit $((X_i, Y_i))_{i\geq 1}$ une suite I.I.D. avec X_1 et Y_1 variables exponentielles de paramètre 1 indépendantes. On se donne également indépendamment de cette suite une variable ϵ t.q. $\mathbb{P}(\epsilon = 1) = \mathbb{P}(\epsilon = -1) = \frac{1}{2}$ et on pose

$$N = \inf\{i \geq 1 :\ 2Y_i \geq (1 - X_i)^2\} \ \text{ et } \ Z = \varepsilon X_N.$$

1. Quelle est la loi de N ? Donner $\mathbb{E}(N)$.
2. Quelle est la loi de Z ?
3. En déduire une méthode pour simuler une variable distribuée suivant la loi normale centrée réduite.

4.4 Résumé

Dans tout ce résumé, on note U une variable de loi uniforme sur $[0,1]$ et $(U_i)_{i\geq 1}$ une suite de variables I.I.D. suivant cette loi.

Loi	**Méthode de simulation**
Bernoulli $\mathcal{B}(p)$	$1_{\{U\leq p\}}$
binômiale $\mathcal{B}(n,p)$	$\sum_{i=1}^{n} 1_{\{U_i\leq p\}}$
géométrique $\mathcal{G}eo(p)$	$1+\left[\frac{\ln(U)}{\ln(1-p)}\right]$ où $[x]$ désigne la partie entière de x
Poisson $\mathcal{P}(\lambda)$	$\inf\{n\in\mathbb{N}:\ U_1\times U_2\times\ldots\times U_{n+1}\leq e^{-\lambda}\}$
Uniforme $\mathcal{U}[a,b]$	$a+(b-a)U$
exponentielle $\mathcal{E}(\lambda)$	$-\frac{1}{\lambda}\ln(U)$
Cauchy $\mathcal{C}(a)$	$a\tan(\pi(U-1/2))$
Normale $\mathcal{N}_1(m,\sigma^2)$	$m+\sigma\sqrt{-2\ln(U_1)}\ \cos(2\pi U_2)$

Chapitre 5

Convergence et théorèmes limites

Dans ce chapitre, nous allons introduire différentes notions de convergence pour une suite de variables aléatoires. Après avoir étudié les liens entre ces différentes notions, nous énoncerons deux théorèmes limites très importants des probabilités :

- la loi forte des grands nombres qui énonce la convergence de la moyenne empirique $\frac{1}{n}\sum_{j=1}^{n} X_j$ d'une suite $(X_j)_{j\geq 1}$ de variables aléatoires réelles indépendantes, identiquement distribuées et intégrables vers $\mathbb{E}(X_1)$ lorsque $n \to +\infty$.

- le théorème de la limite centrale qui indique à quelle vitesse cette convergence a lieu sous l'hypothèse supplémentaire que les X_j sont de carré intégrable.

5.1 Convergence

Définition 5.1.1. *Pour $n \to +\infty$, on dit qu'une suite $(X_n)_{n\geq 1}$ de variables aléatoires à valeurs $\mathbb{R}^d$ converge vers la variable aléatoire X à valeurs $\mathbb{R}^d$*

- *presque sûrement si $\mathbb{P}(X_n \longrightarrow X) = \mathbb{P}(\{\omega : X_n(\omega) \longrightarrow X(\omega)\}) = 1$.*
- *en probabilité si $\forall \varepsilon > 0,\ \mathbb{P}(|X_n - X| \geq \varepsilon) \longrightarrow_{n\to+\infty} 0$.*
- *dans L^1 si les variables X_n, X sont intégrables et $\mathbb{E}|X_n - X| \longrightarrow_{n\to+\infty} 0$.*
- *dans L^2 (ou en moyenne quadratique) si les variables X_n, X sont de carré intégrable et $\mathbb{E}(|X_n - X|^2) \longrightarrow_{n\to+\infty} 0$.*

Remarque 5.1.2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge dans L^1 vers X. Alors $\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$. En effet, comme $X_n - X \leq |X_n - X|$, par linéarité et croissance de l'espérance, $\mathbb{E}(X_n) - \mathbb{E}(X) = \mathbb{E}(X_n - X) \leq \mathbb{E}|X_n - X|$. Par symétrie, $\mathbb{E}(X_n) - \mathbb{E}(X) \leq \mathbb{E}|X_n - X|$. Ainsi $|\mathbb{E}(X_n) - \mathbb{E}(X)| \leq \mathbb{E}|X_n - X|$, ce qui permet de conclure.

Pour établir les liens entre ces différents types de convergence, nous utiliserons les deux résultats suivants. Le premier qui porte le nom de théorème de convergence dominée et que nous ne démontrerons pas est l'analogue du théorème de convergence dominée pour les intégrales.

Théorème 5.1.3 (de convergence dominée). *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge presque sûrement vers X. On suppose de plus que la suite est dominée au sens où il existe une variable aléatoire Y* **intégrable** *telle que*

$$\forall n \geq 1,\ \mathbb{P}(|X_n| \leq Y) = 1.$$

Alors X est intégrable et $(X_n)_n$ converge dans L^1 vers X ce qui entraîne en particulier que

$$\boxed{\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X).}$$

Proposition 5.1.4 (Quelques inégalités). *Soit X et Y deux variables aléatoires réelles.*

Inégalité de Markov : *Si $\mathbb{E}|X| < +\infty$ alors* $\boxed{\forall a > 0,\ \mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}|X|}{a}.}$

Inégalité de Bienaymé-Tchebychev : *Si $\mathbb{E}(X^2) < +\infty$,* $\boxed{\forall a > 0,\ \mathbb{P}(|X| \geq a) \leq \dfrac{\mathbb{E}(X^2)}{a^2}.}$

Inégalité de Cauchy-Schwarz : *Si les variables X et Y sont de carré intégrable,* $\boxed{|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)}\ \sqrt{\mathbb{E}(Y^2)}.}$

Démonstration : Comme $\forall x \in \mathbb{R},\ 1_{\{|x|\geq a\}} \leq \frac{|x|}{a}$,

$$\mathbb{P}\left(1_{\{|X|\geq a\}} \leq \frac{|X|}{a}\right) = 1$$

et on obtient l'inégalité de Markov en utilisant la propriété de croissance de l'espérance et l'égalité $\mathbb{E}(1_{\{|X|\geq a\}}) = \mathbb{P}(|X| \geq a)$.
L'inégalité de Bienaymé-Tchebychev s'obtient par la même méthode en remarquant que $\forall x \in \mathbb{R},\ 1_{\{|x|\geq a\}} \leq \frac{x^2}{a^2}$.
Pour montrer l'inégalité de Cauchy-Schwarz, on constate que pour tout $\lambda \in \mathbb{R}$,

$$\mathbb{E}(X^2) + 2\lambda\mathbb{E}(XY) + \lambda^2\mathbb{E}(Y^2) = \mathbb{E}((X + \lambda Y)^2) \geq 0.$$

On en déduit que le discriminant $4(\mathbb{E}(XY))^2 - 4\mathbb{E}(X^2)\mathbb{E}(Y^2)$ de ce polynôme du second degré en λ est négatif ou nul, ce qui fournit le résultat souhaité. □

Proposition 5.1.5.
- *La convergence L^2 implique la convergence L^1 qui elle-même implique la convergence en probabilité.*
- *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles qui converge dans L^2 vers X. Alors $\mathbb{E}(X_n)$, $\mathbb{E}(X_n^2)$ et $\mathrm{Var}\,(X_n)$ convergent resp. vers $\mathbb{E}(X)$, $\mathbb{E}(X^2)$ et $\mathrm{Var}\,(X)$.*
- *La convergence presque sûre entraîne la convergence en probabilité.*

Démonstration : *Convergence L^2 $\Rightarrow$ convergence L^1 :*
Cela découle du fait que toute variable aléatoire de carré intégrable est intégrable et de l'inégalité $\mathbb{E}|X_n - X| \leq \sqrt{\mathbb{E}(|X_n - X|^2)}$ (qui se déduit elle-même par exemple de la positivité de la variance de $|X_n - X|$).

Convergence L^1 $\Rightarrow$ convergence en probabilité :
Cela découle de l'inégalité de Markov $\mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{|X_n - X|}{\varepsilon}$ pour $\varepsilon > 0$.

Convergence L^2 $\Rightarrow$ convergence des espérances et variances :
D'après le premier point et la remarque 5.1.2, il suffit de vérifier que $\mathbb{E}(X_n^2)$ converge vers $\mathbb{E}(X^2)$.

Par l'inégalité de Cauchy-Schwarz, $\mathbb{E}(X_n X) \leq \sqrt{\mathbb{E}(X_n^2)}\sqrt{\mathbb{E}(X^2)}$.
Donc

$$\mathbb{E}((X_n - X)^2) = \mathbb{E}(X_n^2) - 2\mathbb{E}(X_n X) + \mathbb{E}(X^2) \geq (\sqrt{\mathbb{E}(X_n^2)} - \sqrt{\mathbb{E}(X^2)})^2.$$

Ainsi $\sqrt{\mathbb{E}(X_n^2)}$ converge vers $\sqrt{\mathbb{E}(X^2)}$ et on conclut par continuité de $x \to x^2$.

Convergence presque sûre $\Rightarrow$ convergence en probabilité :
Soit $(X_n)_n$ une suite qui converge presque sûrement vers X et $\varepsilon > 0$. La suite $|X_n - X|$ converge presque sûrement vers 0.
Comme la fonction $1_{\{|x| \geq \varepsilon\}}$ est continue en 0, on en déduit que $Y_n = 1_{\{|X_n - X| \geq \varepsilon\}}$ converge presque sûrement vers 0. Les variables Y_n sont dominées par la constante 1 qui est bien sûr intégrable. Donc le théorème de convergence dominée 5.1.3 implique que

$$\lim_{n \to +\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

□

5.2 Lois des grands nombres

Soit $(X_j)_{j \geq 1}$ une suite de variables aléatoires réelles indépendantes indentiquement distribuées (ce que l'on note aussi I.I.D.). Les lois des grands nombres portent sur le comportement de la moyenne empirique $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ lorsque $n \to +\infty$.

5.2.1 Loi faible des grands nombres

On suppose que les X_j sont de carré intégrable (i.e. que $\mathbb{E}(X_1^2) < +\infty$).

$$\begin{aligned}
\mathbb{E}\left((M_n - \mathbb{E}(X_1))^2\right) &= \frac{1}{n^2}\mathbb{E}\left(\left(\sum_{j=1}^n X_j - \mathbb{E}\left(\sum_{j=1}^n X_j\right)\right)^2\right) = \frac{1}{n^2}\mathrm{Var}\,(\sum_{j=1}^n X_j) \\
&= \frac{1}{n^2}\sum_{j=1}^n \mathrm{Var}\,(X_j) \quad \text{par indépendance des } X_j, \\
&= \frac{\mathrm{Var}\,(X_1)}{n} \longrightarrow_{n \to +\infty} 0
\end{aligned}$$

On en déduit

Proposition 5.2.1. *La moyenne empirique $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ d'une suite $(X_j)_{j \geq 1}$ de variables aléatoires réelles I.I.D. et de carré intégrable converge dans L^2 (et donc dans L^1 et en probabilité) vers l'espérance commune des X_j lorsque $n \to +\infty$.*

5.2.2 Loi forte des grands nombres

Le résultat précédent est supplanté par la loi forte des grands nombres qui indique que la convergence a lieu presque sûrement :

Théorème 5.2.2 (Loi forte des grands nombres). *La moyenne empirique* $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ *d'une suite* $(X_j)_{j\geq 1}$ *de variables aléatoires réelles I.I.D.* **intégrables** *converge presque sûrement et dans* L^1 *vers l'espérance commune des* X_j *lorsque* $n \to +\infty$.

$$\boxed{\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^n X_j \to \mathbb{E}(X_1)\right) = 1.}$$

Démonstration : La démonstration de ce résultat est délicate et nous nous contenterons de montrer la convergence presque sûre sous l'hypothèse forte $\mathbb{E}(X_1^4) < +\infty$.
Par linéarité de l'espérance,

$$\mathbb{E}\Big((M_n - \mathbb{E}(X_1))^4\Big)$$
$$= \frac{1}{n^4}\sum_{i_1,i_2,i_3,i_4=1}^n \mathbb{E}\left((X_{i_1} - \mathbb{E}(X_{i_1}))(X_{i_2} - \mathbb{E}(X_{i_2}))(X_{i_3} - \mathbb{E}(X_{i_3}))(X_{i_4} - \mathbb{E}(X_{i_4}))\right).$$

Parmi les n^4 espérances qui figurent dans la somme au second membre beaucoup sont nulles : en effet dès que l'un des quatre indices i_k est différent des trois autres, par indépendance des X_j, l'espérance est égale à l'espérance du produit des termes correspondant aux trois autres indices par $\mathbb{E}(X_{i_k} - \mathbb{E}(X_{i_k})) = 0$.
Il reste donc seulement les contributions des cas où

- les 4 indices sont égaux soit n termes en $\mathbb{E}((X_1 - \mathbb{E}(X_1))^4)$
- 2 des indices prennent une valeur et les 2 autres une autre valeur soit $3n(n-1)$ termes en $\mathbb{E}\left((X_1 - \mathbb{E}(X_1))^2(X_2 - \mathbb{E}(X_2))^2\right) = (\mathrm{Var}\,(X_1))^2$ (3 pour le choix de l'indice i_k, $k = 2, 3$ ou 4, égal au premier indice i_1, $n(n-1)$ pour le choix de deux valeurs différentes parmi n).

Ainsi

$$\mathbb{E}((M_n - \mathbb{E}(X_1))^4 = \frac{1}{n^3}\mathbb{E}((X_1 - \mathbb{E}(X_1))^4) + \frac{3(n-1)}{n^3}(\mathrm{Var}\,(X_1))^2.$$

On en déduit qu'il existe $C > 0$ t.q. $\forall n \geq 1$, $\mathbb{E}\Big((M_n - \mathbb{E}(X_1))^4\Big) \leq \frac{C}{n^2}$ puis que

$$\mathbb{E}\left(\sum_{n\geq 1}((M_n - \mathbb{E}(X_1))^4\right) = \sum_{n\geq 1}\mathbb{E}\left(((M_n - \mathbb{E}(X_1))^4\right) < +\infty.$$

$$\text{Donc } \mathbb{P}\left(\sum_{n\geq 1}(M_n - \mathbb{E}(X_1))^4 < +\infty\right) = 1.$$

Ainsi avec probabilité 1, $((M_n - \mathbb{E}(X_1))^4)_n$ converge vers 0 comme terme général d'une série convergente. □

Exercice 5.2.3. Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles I.I.D. et de carré intégrable, $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ et $V_n = \frac{1}{n}\sum_{j=1}^n X_j^2 - (M_n)^2$.

1. Montrer que V_n converge presque sûrement vers Var (X_1).
2. Montrer que $\mathbb{E}(V_n) = \frac{n-1}{n}\mathrm{Var}\,(X_1)$. Comme l'espérance de V_n est différente de Var (X_1), on dit que V_n est un estimateur biaisé de Var (X_1). En ravanche, $\frac{n}{n-1}V_n$ est un estimateur sans biais.

5.3 Fonction caractéristique et convergence en loi

5.3.1 Fonction caractéristique

Définition 5.3.1. *Soit X un vecteur aléatoire à valeurs $\mathbb{R}^d$. On appelle fonction caractéristique de X la fonction*

$$\boxed{\Phi_X : u \in \mathbb{R}^d \to \Phi_X(u) = \mathbb{E}\left(e^{iu.X}\right) \in \mathbb{C}.}$$

Remarque 5.3.2.
- $\Phi_X(0,\ldots,0) = 1$.
- $\forall u \in \mathbb{R}^d,\ \Phi_X(-u) = \overline{\Phi_X(u)}$.
- La fonction caractéristique de X ne dépend que de la loi de X :

$$X \overset{\mathcal{L}}{=} Y \Longrightarrow \Phi_X \equiv \Phi_Y.$$

- Si Φ_X est intégrable alors X possède la densité obtenue par inversion de Fourier :

$$x \in \mathbb{R}^d \to p(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-iu.x} \Phi_X(u) du.$$

Le tableau suivant donne les fonctions caractéristiques associées aux lois réelles usuelles :

Loi	**fonction caractéristique**
Bernoulli $\mathcal{B}(p)$	$(1-p) + pe^{iu}$
binômiale $\mathcal{B}(n,p)$	$((1-p) + pe^{iu})^n$
géométrique $\mathcal{G}eo(p)$	$\frac{pe^{iu}}{1-(1-p)e^{iu}}$
Poisson $\mathcal{P}(\lambda)$	$\exp(\lambda(e^{iu}-1))$
Uniforme $\mathcal{U}[a,b]$	$\frac{\sin((b-a)u/2)}{(b-a)u/2}\ e^{iu\frac{a+b}{2}}$
exponentielle $\mathcal{E}(\lambda)$	$\frac{\lambda}{\lambda - iu}$
Cauchy $\mathcal{C}(a)$	$e^{-a\lvert u\rvert}$
Gaussienne $\mathcal{N}_1(m,\sigma^2)$	$e^{ium - \frac{\sigma^2 u^2}{2}}$

Pour les six premières lois, le calcul est un exercice sans difficulté (le faire). Le cas de la loi de Cauchy est traité dans l'exercice suivant :

Exercice 5.3.3. Soit T une variable aléatoire exponentielle de paramètre $a > 0$ et ε une variable indépendante t.q. $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. On pose

$$X = \varepsilon T.$$

1. Déterminer la loi de X (on pourra décomposer suivant les valeurs prises par ε).
2. Calculer sa fonction caractéristique Φ_X.
3. En appliquant la formule d'inversion donnée dans la remarque 5.3.2, en déduire la fonction caractéristique d'une variable aléatoire qui suit la loi de Cauchy $\mathcal{C}(a)$.

Calcul de Φ_X pour $X \sim \mathcal{N}_1(m, \sigma^2)$:

On commence par vérifier par changement de variable que $G = \frac{1}{\sigma}(X - m) \sim \mathcal{N}_1(0, 1)$. Comme

$$\Phi_X(u) = \mathbb{E}\left(e^{iuX}\right) = \mathbb{E}\left(e^{iu(\sigma G+m)}\right) = e^{ium}\mathbb{E}\left(e^{iu\sigma G}\right) = e^{ium}\Phi_G(u\sigma),$$

il suffit de calculer Φ_G pour en déduire Φ_X.
On a

$$\Phi_G(u) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2}} dx.$$

Par dérivation sous le signe intégral, on en déduit que

$$\begin{aligned}
\Phi'_G(u) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} ixe^{iux} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(-i \int_{\mathbb{R}} (iu - x) e^{iux} e^{-\frac{x^2}{2}} dx - u \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2}} dx \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\left[e^{iux - \frac{x^2}{2}} \right]_{-\infty}^{+\infty} - u \int_{\mathbb{R}} e^{iux} e^{-\frac{x^2}{2}} dx \right) \\
&= -u\Phi_G(u).
\end{aligned}$$

On résout donc l'équation différentielle $\Phi'_G(u) = -u\Phi_G(u)$ avec la condition initiale $\Phi_G(0) = 1$ (cf remarque 5.3.2). On obtient $\Phi_G(u) = e^{-\frac{u^2}{2}}$ puis on en déduit Φ_X.

Nous avons remarqué que la fonction caractéristique d'une variable aléatoire ne dépend que de sa loi. Il s'avère en fait que comme son nom le laisse penser :

Proposition 5.3.4. *La fonction caractéristique caractérise la loi c'est à dire que*

$$\forall u \in \mathbb{R}^d,\ \Phi_X(u) = \Phi_Y(u) \Longrightarrow X \overset{\mathcal{L}}{=} Y \text{ i.e. } \forall f : \mathbb{R}^d \to \mathbb{R} \text{ bornée } \mathbb{E}(f(X)) = \mathbb{E}(f(Y)).$$

Nous ne démontrerons pas ce résultat qui intuitivement vient du fait que les exponentielles complexes $x \to e^{iu.x}$ forment un ensemble *suffisamment gros* de fonctions bornées pour que lorsque l'égalité $\mathbb{E}(f(X)) = \mathbb{E}(f(Y))$ est vérifiée pour toute fonction f dans cet ensemble alors elle l'est pour toute fonction f bornée. Dans le cas où Φ_X est intégrable, c'est une conséquence de la formule d'inversion donnée dans la remarque 5.3.2.

Le résultat suivant qui illustre le fait que régularité de la fonction caractéristique d'une variable aléatoire est reliée à l'intégrabilité de cette variable, nous servira pour établir le théorème de la limite centrale.

Lemme 5.3.5. *Soit X une variable aléatoire réelle de carré intégrale (i.e. telle que $\mathbb{E}(X^2) < +\infty$. Alors sa fonction caractéristique Φ_X est C^2 et admet le développement limité suivant en 0 :*

$$\Phi_X(u) = 1 + iu\mathbb{E}(X) - \frac{u^2}{2}\mathbb{E}(X^2) + o(u^2).$$

Démonstration : Soit $u \in \mathbb{R}$.

$$\frac{1}{h}\left(e^{i(u+h)X} - e^{iuX}\right) \xrightarrow{h\to 0} iXe^{iuX} \text{ presque sûrement.}$$

Comme $\forall a, b \in \mathbb{R},\ |e^{ia} - e^{ib}| \leq |b-a|$,

$$\left|\frac{1}{h}\left(e^{i(u+h)X} - e^{iuX}\right)\right| \leq |X|.$$

La variable aléatoire X étant de carré intégrable, $|X|$ est intégrable et par le théorème de convergence dominée 5.1.3, on obtient que

$$\lim_{h\to 0} \mathbb{E}\left(\frac{1}{h}\left(e^{i(u+h)X} - e^{iuX}\right)\right) = \mathbb{E}(iXe^{iuX}).$$

Ainsi $\lim_{h\to 0} \frac{1}{h}(\Phi_X(u+h) - \Phi_X(u)) = \mathbb{E}(iXe^{iuX})$ et la fonction Φ_X est dérivable de dérivée $\Phi'_X(u) = \mathbb{E}(iXe^{iuX})$. En appliquant à nouveau le théorème de convergence dominée, on peut montrer que Φ'_X est continue et même C^1 avec $\Phi''_X(u) = -\mathbb{E}(X^2 e^{iuX})$.
Ainsi Φ_X est C^2 avec $\Phi'_X(0) = i\mathbb{E}(X)$ et $\Phi''_X(0) = -\mathbb{E}(X^2)$. Le développement limité en 0 s'en déduit par exemple en utilisant la formule de Taylor avec reste intégral. □

5.3.2 Convergence en loi

Définition 5.3.6. *On dit que la suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X à valeurs $\mathbb{R}^d$ et on note $X_n \xrightarrow{\mathcal{L}} X$ si*

$$\boxed{\forall f : \mathbb{R}^d \to \mathbb{R} \ \textbf{continue bornée},\ \mathbb{E}(f(X_n)) \longrightarrow_{n\to+\infty} \mathbb{E}(f(X)).}$$

Exemple 5.3.7.
- Pour $n \in \mathbb{N}^*$, on suppose que $\forall 1 \leq k \leq n$, $\mathbb{P}(U_n = k/n) = 1/n$. Soit $f : \mathbb{R} \to \mathbb{R}$ continue bornée. La convergence des sommes de Riemann vers l'intégrale entraîne que

$$\mathbb{E}(f(U_n)) = \frac{1}{n}\sum_{k=1}^{n} f(k/n) \to_{n\to+\infty} \int_0^1 f(u)du = \mathbb{E}(f(U)),$$

  où U est une variable uniforme sur $[0,1]$. Ainsi la suite $(U_n)_n$ converge en loi vers $U \sim \mathcal{U}[0,1]$.
- Soit maintenant pour $n \in \mathbb{N}^*$, X_n une variable aléatoire uniformément répartie sur $[0, 1/n]$. Alors pour f continue bornée,

$$\mathbb{E}(f(X_n)) = n\int_0^{1/n} f(x)dx \to_{n\to+\infty} f(0).$$

  Donc la suite $(X_n)_n$ converge en loi vers X telle que $\mathbb{P}(X = 0) = 1$.

– Pour $n \in \mathbb{N}$, soit T_n une variable exponentielle de paramètre $\lambda_n > 0$. On suppose que la suite $(\lambda_n)_n$ converge vers $\lambda > 0$. Alors comme pour $f : \mathbb{R} \to \mathbb{R}$ continue bornée,

$$\forall n \in \mathbb{N},\ \forall x \geq 0,\ |f(x)\lambda_n e^{-\lambda_n x}| \leq g(x) = |f(x)|(\sup_n \lambda_n)e^{-(\inf_n \lambda_n)x},$$

où la fonction g est intégrable sur $[0, +\infty[$, par le théorème de convergence dominée,

$$\mathbb{E}(f(T_n)) = \int_0^{+\infty} f(x)\lambda_n e^{-\lambda_n x} dx$$

converge lorsque $n \to +\infty$ vers $\int_0^{+\infty} f(x)\lambda e^{-\lambda x} dx = \mathbb{E}(f(T))$ où T suit la loi exponentielle de paramètre λ. Ainsi $(T_n)_n$ converge en loi vers $T \sim \mathcal{E}(\lambda)$.

Propriétés 5.3.8. *Soit $(X_n)_n$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ qui converge en loi vers X et $\varphi : \mathbb{R}^d \to \mathbb{R}^q$ une fonction continue. Alors la suite $(\varphi(X_n))_n$ converge en loi vers $\varphi(X)$.*

Démonstration : Soit $g : \mathbb{R}^q \to \mathbb{R}$ continue bornée. La fonction $g \circ \varphi : \mathbb{R}^d \to \mathbb{R}$ est continue bornée. Donc la convergence en loi de $(X_n)_n$ vers X entraîne que

$$\lim_{n \to +\infty} \mathbb{E}(g(\varphi(X_n)) = \mathbb{E}(g(\varphi(X))).$$

□

Du fait que les exponentielles complexes constituent à nouveau un sous-ensemble assez gros de l'ensemble des fonctions continues bornées (pour se ramener à des fonctions à valeurs réelles, il suffit de séparer partie réelle et partie imaginaire), on a la caractérisation suivante de la convergence en loi

Théorème 5.3.9. *La suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X à valeurs $\mathbb{R}^d$ si et seulement si la fonction caractéristique de X_n converge ponctuellement vers la fonction caractéristique de X i.e.*

$$\boxed{X_n \xrightarrow{\mathcal{L}} X \Longleftrightarrow \forall u \in \mathbb{R}^d,\ \Phi_{X_n}(u) \to \Phi_X(u).}$$

Corollaire 5.3.10. *Si la suite $(X_n)_n$ converge en probabilité vers X alors elle converge en loi vers X.*

Démonstration : Soit $u \in \mathbb{R}^d$ et $\varepsilon > 0$. On a

$$\begin{aligned}\left|e^{iu.X_n} - e^{iu.X}\right| &\leq \left|e^{iu.X_n} - e^{iu.X}\right| 1_{\{|X_n - X| \geq \varepsilon\}} + \left|e^{iu.X_n} - e^{iu.X}\right| 1_{\{|X_n - X| < \varepsilon\}} \\ &\leq 2 \times 1_{\{|X_n - X| \geq \varepsilon\}} + \left|e^{iu.X_n} - e^{iu.X}\right| 1_{\{|X_n - X| < \varepsilon\}}.\end{aligned}$$

Comme $\forall a, b \in \mathbb{R},\ |e^{ia} - e^{ib}| \leq |b - a|$, on en déduit que

$$\left|e^{iu.X_n} - e^{iu.X}\right| \leq 2 \times 1_{\{|X_n - X| \geq \varepsilon\}} + |u|\varepsilon 1_{\{|X_n - X| < \varepsilon\}} \leq 2 \times 1_{\{|X_n - X| \geq \varepsilon\}} + |u|\varepsilon$$

puis que

$$|\Phi_{X_n}(u) - \Phi_X(u)| = |\mathbb{E}(e^{iu.X_n} - e^{iu.X})| \leq \mathbb{E}|e^{iu.X_n} - e^{iu.X}| \leq |u|\varepsilon + 2\mathbb{P}(|X_n - X| \geq \varepsilon).$$

Le premier terme du second membre est arbitrairement petit (uniformément en n) pour ε petit tandis qu'à ε fixé le second terme converge vers 0 lors $n \to +\infty$ par définition de la convergence en probabilité. Ainsi $\forall u \in \mathbb{R}^d$, $\lim_{n\to+\infty} \Phi_{X_n}(u) = \Phi_X(u)$ et on conclut en utilisant le théorème 5.3.9. □

Remarque 5.3.11. Si la suite $(X_n)_n$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers la variable aléatoire X, alors

$$\mathbb{E}(f(X_n)) \longrightarrow_{n\to+\infty} \mathbb{E}(f(X))$$

pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ bornée dont l'ensemble des points de discontinuité $\mathcal{D}_f$ vérifie $\mathbb{P}(X \in \mathcal{D}_f) = 0$.

Exercice résolu 5.3.12. *Soit $\lambda > 0$ et pour tout $n \geq \lambda$, X_n une variable géométrique de paramètre λ/n. Montrer que les variables $Y_n = X_n/n$ convergent en loi vers une limite que l'on précisera.*
On se donne $U \sim \mathcal{U}[0,1]$. D'après l'exercice 4.3.1, $1 + [\frac{\ln(U)}{\ln(1-\lambda/n)}] \stackrel{\mathcal{L}}{=} X_n$.
Donc $\frac{1}{n}(1 + [\frac{\ln(U)}{\ln(1-\lambda/n)}]) \stackrel{\mathcal{L}}{=} Y_n$. Pour $n \to +\infty$, $\frac{\ln(U)}{\ln(1-\lambda/n)}$ est équivalent à $-\frac{n}{\lambda}\ln(U)$. On en déduit que pour $n \to +\infty$, $\frac{1}{n}(1 + [\frac{\ln(U)}{\ln(1-\lambda/n)}])$ converge presque sûrement et donc en probabilité et en loi (voir proposition 5.1.5 et corollaire 5.3.10) vers $-\frac{1}{\lambda}\ln(U)$. Pour $f : \mathbb{R} \to \mathbb{R}$ continue bornée, on a donc

$$\mathbb{E}(f(Y_n)) = \mathbb{E}\left(f\left(\frac{1}{n}\left(1 + \left[\frac{\ln(U)}{\ln(1-\lambda/n)}\right]\right)\right)\right) \longrightarrow_{n\to+\infty} \mathbb{E}\left(f\left(-\frac{1}{\lambda}\ln(U)\right)\right).$$

D'après l'exemple 4.2.3, $-\frac{1}{\lambda}\ln(U) \sim \mathcal{E}(\lambda)$. On conclut donc que la suite $(Y_n)_n$ converge en loi vers Y qui suit la loi exponentielle de paramètre λ.
Il faut du recul pour pouvoir trouver la solution qui précède mais on peut parvenir à la même conclusion sans astuce en calculant la fonction carcatéristique de Y_n et en utilisant le théorème 5.3.9 (le faire).

5.4 Le théorème de la limite centrale

5.4.1 Enoncé et preuve du résultat

Le théorème de la limite centrale donne la vitesse à laquelle la convergence a lieu dans loi forte des grands nombres, sous l'hypothèse supplémentaire d'intégrabilité des X_j^2 :

Théorème 5.4.1 (de la limite centrale). *Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles indépendantes et identiquement distribuées telles que $\mathbb{E}(X_1^2) < +\infty$ et $\sigma = \sqrt{\mathrm{Var}\,(X_1)} > 0$. On note $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ la moyenne empirique.*
Alors pour $n \to +\infty$,

$$\boxed{\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Y \sim \mathcal{N}_1(0,1).}$$

Remarque 5.4.2. – Formellement, on récrit $M_n \overset{\mathcal{L}}{\simeq} \mathbb{E}(X_1) + \frac{\sigma}{\sqrt{n}}\, Y$, ce qui indique que la convergence dans la loi forte des grands nombres a lieu à la vitesse $\frac{\sigma}{\sqrt{n}}$.
– Ce résultat explique la place fondamentale des variables aléatoires gaussiennes (ou normales) en probabilités et en statistiques.
– Notons que la renormalisation effectuée est telle que

$$\forall n \in \mathbb{N}^*,\ \mathrm{Var}\left(\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}(X_1))\right) = \frac{n}{\sigma^2}\mathrm{Var}\left((M_n - \mathbb{E}(X_1)\right) = \frac{1}{n\sigma^2}\sum_{j=1}^{n}\mathrm{Var}\,(X_j) = 1.$$

Démonstration : On utilise la fonction caractéristique. Soit $u \in \mathbb{R}$.

$$\begin{aligned}
\Phi_{\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}(X_1))}(u) &= \mathbb{E}\left(e^{iu\frac{1}{\sigma\sqrt{n}}\sum_{j=1}^{n}(X_j - \mathbb{E}(X_j))}\right) \\
&= \prod_{j=1}^{n}\mathbb{E}\left(e^{iu\frac{1}{\sigma\sqrt{n}}(X_j - \mathbb{E}(X_j))}\right) \quad \text{par indépendance des } X_j, \\
&= \left(\mathbb{E}\left(e^{iu\frac{1}{\sigma\sqrt{n}}(X_1 - \mathbb{E}(X_1))}\right)\right)^n \quad \text{car les } X_j \text{ ont même loi}, \\
&= \left(\Phi_{X_1 - \mathbb{E}(X_1)}\left(\frac{u}{\sigma\sqrt{n}}\right)\right)^n.
\end{aligned}$$

Comme $\mathbb{E}(X_1 - \mathbb{E}(X_1)) = 0$ et $\mathbb{E}((X_1 - \mathbb{E}(X_1))^2) = \sigma^2$, d'après le lemme 5.3.5, pour v au voisinage de 0, $\Phi_{X_1 - \mathbb{E}(X_1)}(v) = 1 - \frac{\sigma^2}{2}v^2 + o(v^2)$. Donc pour n grand,

$$\Phi_{X_1 - \mathbb{E}(X_1)}\left(\frac{u}{\sigma\sqrt{n}}\right) = 1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right).$$

Ainsi

$$\Phi_{\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}(X_1))}(u) = \left(1 - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \longrightarrow_{n\to+\infty} e^{-\frac{u^2}{2}} = \Phi_Y(u).$$

On conclut par le théorème 5.3.9. □

5.4.2 Intervalle de confiance dans la méthode de Monte-Carlo

Le principe de la méthode de Monte-Carlo est le suivant : pour évaluer numériquement l'espérance d'une variable aléatoire réelle X intégrable, on génère une réalisation d'une suite $(X_j)_{j\geq 1}$ de variables aléatoires indépendantes et de même loi que X (par exemple en utilisant les techniques de simulation présentées dans le chapitre 4) et on approche $\mathbb{E}(X)$ par la moyenne emprique $M_n = \frac{1}{n}\sum_{j=1}^{n} X_j$.

La loi forte des grands nombres justifie la convergence de cette méthode. Comme nous allons le voir, si X est de carré intégrable, le théorème de la limite centrale permet de se faire une idée de la précision avec laquelle $\mathbb{E}(X)$ est évaluée.

Si on se donne $a > 0$, la fonction $f(x) = 1_{\{|x|\leq a\}}$ est bornée et continue en dehors des points $-a$ et a. Malgré ces discontinuités, comme pour $Y \sim \mathcal{N}_1(0,1)$, $\mathbb{P}(Y = -a) =$

$\mathbb{P}(Y = a) = 0$, d'après la remarque 5.3.11, la convergence en loi de $\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}(X))$ vers Y entraîne la convergence de

$$\mathbb{P}\left(M_n - \frac{\sigma a}{\sqrt{n}} \leq \mathbb{E}(X) \leq M_n + \frac{\sigma a}{\sqrt{n}}\right) = \mathbb{E}\left(f\left(\frac{\sqrt{n}}{\sigma}(M_n - \mathbb{E}(X))\right)\right)$$

vers $\mathbb{E}(f(Y))$ pour $n \to +\infty$.
C'est pourquoi dès que n est supérieur à 20, on fait l'approximation

$$\mathbb{P}\left(M_n - \frac{\sigma a}{\sqrt{n}} \leq \mathbb{E}(X) \leq M_n + \frac{\sigma a}{\sqrt{n}}\right) \simeq \mathbb{E}(f(Y)) = \mathbb{E}(1_{\{|Y| \leq a\}}) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{-\frac{x^2}{2}} dx.$$

On utilise le plus couramment les valeurs de a pour lesquelles l'intégrale vaut $0,95$ ou $0,99$ soit respectivement $a = 1,96$ ou $a = 2,58$.

Ainsi la probabilité pour que la valeur $\mathbb{E}(X)$ que l'on cherche à calculer se trouve dans l'intervalle $\left[M_n - \frac{1,96\sigma}{\sqrt{n}}, M_n + \frac{1,96\sigma}{\sqrt{n}}\right]$ (resp. $\left[M_n - \frac{2,58\sigma}{\sqrt{n}}, M_n + \frac{2,58\sigma}{\sqrt{n}}\right]$) est de l'ordre de $0,95$ (resp. $0,99$). On parle alors d'intervalle de confiance à 95% (resp. 99%).

Dans la pratique, le plus souvent, la valeur de l'écart type σ est inconnue. En effet calculer Var $(X) = \sigma^2$ est un problème au moins aussi compliqué que de calculer $\mathbb{E}(X)$. On remplace donc σ par $\sqrt{V_n}$ où $V_n = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - (M_n)^2$ converge presque sûrement vers Var (X) d'après l'exercice 5.2.3. Pour évaluer V_n, il est simplement nécessaire de stocker la somme des carré des X_i en plus de la somme des X_i au cours de la simulation : cela ne demande que très peu de calculs supplémentaires.

On obtient alors l'intervalle de confiance à 95% :

$$\boxed{\left[M_n - \frac{1,96\sqrt{V_n}}{\sqrt{n}}, M_n + \frac{1,96\sqrt{V_n}}{\sqrt{n}}\right]}.$$

Lorsque l'on emploie la méthode de Monte-Carlo, il faut systématiquement déterminer un intervalle de confiance puisque cela permet de se faire une idée de la précision de la valeur obtenue tout en nécessitant très peu de calculs supplémentaires. Bien sûr, dans le cas où la largeur de l'intervalle $\frac{2a\sqrt{V_n}}{\sqrt{n}}$ est grande par rapport à $|M_n|$ cela signifie qu'il ne faut pas accorder beaucoup de crédit à la valeur M_n calculée. On peut alors augmenter le nombre n de tirages pour augmenter la précision.

Remarque 5.4.3. On peut vérifier que, sous les hypothèses du théorème 5.4.1, $\sqrt{\frac{n}{V_n}}(M_n - \mathbb{E}(X_1))$ converge en loi vers $Y \sim \mathcal{N}_1(0,1)$, ce qui justifie l'approximation de σ par $\sqrt{V_n}$ faite plus haut dans la donnée de l'intervalle de confiance.

5.5 Exercices

Exercice 5.5.1. Soit Y une variable uniformément répartie sur $[-1,1]$, Z une variable normale centrée réduite et $(X_n)_{n\geq 1}$ une suite de variables indépendantes t.q.

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

$$\text{On pose} \qquad Y_n = \sum_{k=1}^{n} \frac{X_k}{2^k} \qquad Z_n = \frac{1}{n}\sum_{k=1}^{n} \sqrt{2k}\, X_k.$$

1. Calculer la fonction caractéristique de Y_n. En remarquant que

$$\sin(\lambda/2^n)\prod_{k=1}^{n}\cos(\lambda/2^k) = \sin(\lambda)/2^n,$$

en déduire que la suite Y_n converge en loi vers Y.

2. Calculer la fonction caractéristique Φ_{Z_n} de Z_n. Donner le développement limité à l'ordre 2 à l'origine de la fonction $x \in]-\frac{\pi}{2}, \frac{\pi}{2}[\to \ln(\cos(x))$. En déduire pour $u \in \mathbb{R}$ la limite de $\ln(\Phi_{Z_n}(u))$ lorsque n tend vers l'infini et conclure que la suite Z_n converge en loi vers Z.

Exercice 5.5.2. Soit U_i une suite de variables IID suivant la loi uniforme sur $[0,1]$ et $\alpha > 0$.

1. Montrer que la suite $X_n = (U_1 \dots U_n)^{\frac{\alpha}{n}}$ converge presque sûrement et donner sa limite (on pourra s'intéresser à $\ln(X_n)$).
2. On pose $Y_n = e^{\alpha\sqrt{n}}(U_1 ... U_n)^{\alpha/\sqrt{n}}$.
   Quel est le comportement de la suite $\frac{1}{\alpha}\ln(Y_n)$ pour $n \to +\infty$?
   En déduire que la suite Y_n converge en loi et déterminer la loi de sa limite.

Exercice 5.5.3.

1. Soit X une variable de Bernoulli de paramètre $p \in [0,1]$. Pour quelle valeur de p la variance de X est-elle maximale ? Que vaut-elle alors ?
2. Aux Etats-Unis sur 4 000 000 de naissance annuelles, on observe un ratio de 1 048 garçons pour 1 000 filles. Donner un intervalle de confiance à 99% pour la probabilité p qu'un bébé soit un garçon. Que pensez-vous de l'hypothèse d'équilibre des naissances : $p = 0.5$?
3. Lors du second tour de l'élection présidentielle, un sondage est effectué à la sortie des urnes sur un échantillon de 1 000 personnes. Le candidat A recueille a% (a proche de 50) des suffrages des personnes interrogées. A l'aide du théorème de la limite centrale, donner un intervalle de confiance à 95% pour le score S_A réalisé par ce candidat. A partir de quelle valeur de $|a - 50|$, peut-on se baser sur le sondage pour connaître le résultat de l'élection ?

Exercice 5.5.4. Soit $(Y_k)_{k\in\mathbb{N}^*}$ une suite de variables indépendantes et identiquement distribuées suivant la loi de Cauchy de paramètre $a > 0$. On pose $M_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$.

1. Quelle est la loi de M_n ? La suite $(M_n)_{n\in\mathbb{N}^*}$ converge-t-elle en loi ?
2. Quelle est la loi de $M_{n+p} - M_n$? Quel est le comportement de $P(|M_{2n} - M_n| \geq 1)$ lorsque $n \to +\infty$?
3. Montrer que si la suite $(X_n)_{n\in\mathbb{N}^*}$ converge en probabilité vers X alors $\forall\epsilon > 0$, $\lim_{n,m\to+\infty} \mathbb{P}(|X_n - X_m| \geq \epsilon) = 0$. La suite $(M_n)_{n\in\mathbb{N}^*}$ converge-t-elle en probabilité ?

Exercice 5.5.5. Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes d'espérance 1 qui prennent deux valeurs a et b $(0 < a < 1 < b)$ avec probabilité p et $1-p$ respectivement $(0 < p < 1)$. On pose $Y_n = \prod_{i=1}^{n} X_i$.

1. Vérifier que pour $x \in]0,1[\cup]1,+\infty[$, $\ln(x) < x - 1$ et en déduire que $\mathbb{E}(\ln(X_1)) < 0$. En étudiant le comportement de $\frac{1}{n}\ln(Y_n)$ lorsque $n \to +\infty$, montrer que la suite Y_n converge presque sûrement vers une limite que l'on précisera. Un joueur joue tous les jours le dixième de sa fortune à pile ou face avec un ami. Quelle est l'évolution de cette fortune au bout d'un temps très long ?

2. Calculer $\mathbb{E}(Y_n)$. La suite Y_n converge-t-elle dans L^1 ?

Exercice 5.5.6. Soit $(X_i)_{i\geq 1}$ une suite de variables I.I.D. suivant la loi uniforme sur $[0,\theta]$ où $\theta > 0$. On pose $Y_n = \max(X_1, X_2, \ldots, X_n)$.

1. Pour $n \geq 1$, calculer la fonction de répartition de Y_n.
2. En déduire la loi de Y_n et calculer $\mathbb{E}(Y_n)$ et $\mathrm{Var}(Y_n)$.
3. Montrer à l'aide de 1. que la suite Y_n converge en probabilité vers une limite que l'on précisera.
   Converge-t-elle presque sûrement ? (on pourra s'intéresser à la monotonie de $(Y_n)_n$.)

Exercice 5.5.7. Soient $X_1, Y_1, \ldots, X_n, Y_n, \ldots$ des variables aléatoires réelles indépendantes de carré intégrable. On suppose que les X_i sont identiquement distribuées et que les Y_i le sont également. On note $\mathbb{E}(X_1) = \mu$, $\mathbb{E}(Y_1) = \nu$, $Var(X_1) = \sigma^2$ et $Var(Y_1) = \tau^2$. Soit $S_n = \sum_{m=1}^n X_m$, $T_n = \sum_{m=1}^n X_m Y_m$. Quel est le comportement asymptotique des variables

$$\frac{S_n}{n}, \quad \frac{T_n}{n}, \quad \text{et} \quad \frac{T_n - \nu S_n}{\sqrt{n}} \ ?$$

Exercice 5.5.8. Soit $f : [0,1] \to \mathbb{R}$ une fonction continue et S_n^x une variable distribuée suivant la loi binomiale de paramètres (n, x). Montrer que

$$\lim_{n\to+\infty} \sup_{x\in[0,1]} |\mathbb{E}(f(x) - f(S_n^x/n))| = 0.$$

(On pourra séparer l'espérance en deux termes suivant que $|S_n^x/n - x| > \alpha$ ou non, utiliser l'uniforme continuité de f et majorer $P(|S_n^x/n - x| > \alpha)$ uniformément en x en utilisant l'inégalité de Bienaymé-Tchebychev.)
Comme $\mathbb{E}(f(S_n^x/n)) = \sum_{k=0}^n f(k/n) C_k^n x^k (1-x)^{n-k}$, toute fonction continue sur $[0,1]$ est limite uniforme d'une suite de polynômes (théorème de Weierstrass).

Exercice 5.5.9. Soit $(X_j)_{j\geq 1}$ une suite de variables aléatoires réelles I.I.D. et de carré intégrable t.q. $\frac{X_1+X_2}{\sqrt{2}} \overset{\mathcal{L}}{=} X_1$.

1. Quelle est l'espérance commune des X_i ?
2. Pour $k \in \mathbb{N}^*$, que peut-on dire de la loi $\frac{X_1+\ldots+X_{2^k}}{2^{k/2}}$?
3. En déduire que les X_j sont des variables gaussiennes.

Exercice 5.5.10. Soit $(X_k)_{k\geq 1}$ une suite de variables aléatoires réelles indépendantes et indentiquement distribuées suivant une densité $p(x)$ paire, bornée, continue en 0 et telle que $p(0) > 0$

1. Quelle est la loi de $Y_k = \frac{1}{X_k}$?
2. Exprimer à l'aide de p la fonction caractéristique commune des Y_k.
3. On rappelle que $\int_0^{+\infty} \frac{1-\cos(z)}{z^2}\, dz = \frac{\pi}{2}$. Montrer que la suite $\frac{1}{n}\sum_{k=1}^n Y_k$ converge en loi vers Y qui suit la loi de Cauchy de paramètre $\pi p(0)$.
4. En déduire à l'aide de la remarque 5.3.11 que la moyenne harmonique $n/\sum_{k=1}^n \frac{1}{X_k}$ des X_k converge en loi vers une limite que l'on précisera.

5.6 Résumé

• **Diagramme des convergences :**

$$\boxed{\begin{array}{ccccc}
\begin{array}{c}L^2 : \\ \mathbb{E}(|X_n - X|^2) \to 0\end{array} & & & & \\
& \searrow & & & \\
& & \begin{array}{c}L^1 : \\ \mathbb{E}|X_n - X| \to 0\end{array} & & \begin{array}{c}\text{presque sûre :} \\ \mathbb{P}(X_n \to X) = 1\end{array} \\
& & & \searrow \quad \swarrow & \\
& & & \begin{array}{c}\text{Proba : } \forall \varepsilon > 0, \\ \mathbb{P}(|X_n - X| \geq \varepsilon) \to 0 \\ \downarrow \\ \text{Loi : } \forall f \textbf{continue} \text{ bornée,} \\ \mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))\end{array} &
\end{array}}$$

• **Fonction caractéristique :** $\boxed{\Phi_X(u) = \mathbb{E}(e^{iu.X}).}$

– caractérise la loi :

$$\boxed{\Phi_X \equiv \Phi_Y \Longleftrightarrow X \overset{\mathcal{L}}{=} Y \Longleftrightarrow \forall f \text{ bornée, } \mathbb{E}(f(X)) = \mathbb{E}(f(Y)).}$$

– caractérise la convergence en loi :

$$\boxed{\forall u,\ \Phi_{X_n}(u) \longrightarrow \Phi_X(u) \Longleftrightarrow X_n \xrightarrow{\mathcal{L}} X \Longleftrightarrow \forall f \textbf{continue} \text{ bornée, } \mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X)).}$$

• **Comportement de la moyenne empirique** $M_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ d'une suite de variables aléatoires réelles I.I.D. $(X_j)_{j\geq 1}$:

Loi forte des grands nombres : $\boxed{\mathbb{E}|X_1| < +\infty \Longrightarrow \mathbb{P}(M_n \longrightarrow \mathbb{E}(X_1)) = 1}$.

Théorème de la limite centrale :

$$\boxed{\mathbb{E}(X_1^2) < +\infty \text{ et } \mathrm{Var}\,(X_1) \neq 0 \Longrightarrow \sqrt{\frac{n}{\mathrm{Var}\,(X_1)}}\,(M_n - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} Y \sim \mathcal{N}_1(0,1).}$$

Chapitre 6

Vecteurs gaussiens

6.1 Définition, construction

6.1.1 Définition

Définition 6.1.1. *On dit qu'un vecteur aléatoire $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ est un vecteur gaussien si toute combinaison linéaire de ses coordonnées est une variable aléatoire gaussienne (ou normale) réelle i.e. si pour tout $u \in \mathbb{R}^d$, la variable aléatoire réelle $u.X$ est une variable aléatoire gaussienne.*

Soit $X = (X_1, \ldots, X_d)$ un vecteur gaussien à valeurs $\mathbb{R}^d$. Les X_i étant des variables gaussiennes réelles, elles sont de carré intégrable. Notons Γ avec $\Gamma_{ij} = \text{Cov}\ (X_i, X_j)$, la matrice de covariance du vecteur X et m, avec $m_i = \mathbb{E}(X_i)$, son vecteur espérance. Par linéarité de l'espérance on a :

$$\mathbb{E}(u.X) = u_1\mathbb{E}(X_1) + \cdots + u_d\mathbb{E}(X_n) = u.\mathbb{E}(X),$$

et d'après les propriétés 3.3.14,

$$\text{Var}\ (u.X) = \sum_{i=1}^{d}\sum_{j=1}^{d} u_i u_j \text{Cov}\ (X_i, X_j) = u.\Gamma u.$$

Donc $u.X \sim \mathcal{N}_1(u.\mathbb{E}(X), u.\Gamma u)$ et

$$\Phi_X(u) = \mathbb{E}\left(e^{iu.X}\right) = \Phi_{u.X}(1) = e^{i\mathbb{E}(u.X) - \frac{1}{2}\text{Var}\ (u.X)} = e^{iu.m - \frac{1}{2}u.\Gamma u}.$$

On constate que la fonction caractéristique (et donc la loi) ne dépend que du couple (m, Γ). On dit que X suit la loi gaussienne en dimension d d'espérance m et de matrice de covariance Γ, ce que l'on note $X \sim \mathcal{N}_d(m, \Gamma)$.

Inversement, si le vecteur $X = (X_1, \ldots, X_d)$ a pour fonction caractéristique $v \in \mathbb{R}^d \to \Phi_X(v) = e^{iv.m - \frac{1}{2}v.\Gamma v}$ où $m \in \mathbb{R}^d$ et Γ est une matrice $d \times d$ symétrique et positive, alors pour $u \in \mathbb{R}^d$,

$$\forall t \in \mathbb{R},\ \Phi_{u.X}(t) = \mathbb{E}(e^{it(u.X)}) = \mathbb{E}(e^{i(tu).X}) = \Phi_X(tu) = e^{it(u.m) - \frac{t^2}{2}u.\Gamma u}.$$

Donc $\forall u \in \mathbb{R}^d$, $u.X \sim \mathcal{N}_1(u.m, u.\Gamma u)$. Ainsi X est un vecteur gaussien.
En laissant u décrire la base canonique de $\mathbb{R}^d$, on vérifie que $\forall 1 \leq j \leq d$, $\mathbb{E}(X_j) = m_j$.
Puis en utilisant la bilinéarité de la covariance, on obtient que pour $1 \leq i, j \leq d$

$$2\text{Cov}\,(X_i, X_j) = \text{Var}\,(X_i + X_j) - \text{Var}\,(X_i) - \text{Var}\,(X_j) = (\Gamma_{ii} + 2\Gamma_{ij} + \Gamma_{jj}) - \Gamma_{ii} - \Gamma_{jj} = 2\Gamma_{ij}.$$

Ainsi m et Γ sont respectivement le vecteur espérance et la matrice de covariance de X.

Exemple 6.1.2. Si pour $1 \leq j \leq d$, $G_j \sim \mathcal{N}_1(m_j, \sigma_j^2)$ et ces variables aléatoires sont indépendantes, alors pour $G = (G_1, \ldots, G_d)$ et $u \in \mathbb{R}^d$,

$$\Phi_G(u) = \mathbb{E}(e^{iu.G}) = \mathbb{E}\left(\prod_{j=1}^{d} e^{iu_j G_j}\right) = \prod_{j=1}^{d} \mathbb{E}\left(e^{iu_j G_j}\right) = \prod_{j=1}^{d} e^{iu_j m_j - \frac{1}{2}u_j^2 \sigma_j^2} = e^{iu.m - \frac{1}{2}u.\text{Diag}(\sigma^2)u},$$

où $\text{Diag}(\sigma^2)$ est la matrice diagonale d'éléments diagonaux $\text{Diag}(\sigma^2)_{jj} = \sigma_j^2$ et m le vecteur de coordonnées m_j. Ainsi $G \sim \mathcal{N}_d(m, \text{Diag}(\sigma^2))$.
Dans le cas particulier ou les G_j sont centrées réduites, $G \sim \mathcal{N}_d(0, I_d)$ où $I_d \in \mathbb{R}^{d \times d}$ est la matrice identité.

Remarque 6.1.3. Il ne suffit pas que X_1 et X_2 soient des variables gaussiennes réelles pour que (X_1, X_2) soit un vecteur gaussien. Ce résultat est vrai dans le cas particulier où X_1 et X_2 sont indépendantes mais pas en toute généralité comme l'illustre l'exercice 6.3.2.

6.1.2 Stabilité du caractère gaussien par transformation linéaire

Si X est un vecteur aléatoire à valeurs $\mathbb{R}^d$ et $Y = a + MX$ pour un vecteur (constant) $a \in \mathbb{R}^n$ et une matrice M de taille $n \times d$, alors toute combinaison linéaire des coordonnées de Y est combinaison linéaire des coordonnées de X à une constante près : pour $v \in \mathbb{R}^n$, $v.Y = v.a + (M^* v).X$ où M^* désigne la transposée de la matrice M. Donc si X est gaussien, Y l'est aussi.

Exercice 6.1.4. Vérifier que $\mathbb{E}(Y) = a + M.\mathbb{E}(X)$ et que la matrice de covariance Λ de Y s'exprime en fonction de celle Γ de X par la relation $\Lambda = M\Gamma M^*$.

Exemple 6.1.5. Si $X = (X_1, \ldots, X_d)$ est un vecteur gaussien à valeurs $\mathbb{R}^d$ alors pour $k \leq d$, le vecteur $(X_1, \ldots, X_k)$ est gaussien (de même que tout vecteur obtenu à partir de X en enlevant certaines coordonnées). La moyenne empirique $\frac{1}{d}\sum_{i=1}^{d} X_i$ est une gaussienne réelle.

6.1.3 Construction d'un vecteur gaussien de loi $\mathcal{N}_n(m, \Gamma)$

Soit $m \in \mathbb{R}^n$ et Γ une matrice de covariance $n \times n$ i.e. une matrice symétrique positive. Il existe une matrice $A \in \mathbb{R}^{n \times d}$ telle que :

$$AA^* = \Gamma.$$

On peut par exemple choisir $d = n$ et construire A comme la racine carrée de la matrice symétrique et positive Γ en la diagonalisant.

D'après ce qui précède, si $G = (G_1, \ldots, G_d)$ est un vecteur de variables aléatoires gaussiennes centrées réduites indépendantes alors $m + AG$ suit la loi $\mathcal{N}_n(m, \Gamma)$.

On débouche ainsi sur l'algorithme suivant pour simuler un vecteur gaussien m, Γ :

Simulation d'un vecteur gaussien :
- Calculer $A \in \mathbb{R}^{n \times d}$ t.q. $AA^* = \Gamma$,
- tirer un vecteur de gaussiennes centrées réduites indépendantes $G = (G_1, \ldots, G_d)$,
- retourner le vecteur $m + AG$.

6.2 Propriétés des vecteurs gaussiens

6.2.1 Vecteurs gaussiens et indépendance

Si les coordonnées du vecteur aléatoire $X = (X_1, \ldots, X_d)$ sont indépendantes et de carré intégrable, alors sa matrice de covariance est diagonale. En effet pour $i \neq j$, $\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\mathbb{E}(X_j)$ i.e. Cov $(X_i, X_j) = 0$.
Dans le cas où X est un vecteur gaussien, le caractère diagonal de la matrice de covariance s'avère une condition suffisante d'indépendance. Cela vient du fait que la loi d'un vecteur gaussien ne dépend que de son espérance et de sa matrice de covariance.

Proposition 6.2.1. *Les coordonnées d'un vecteur gaussien $X = (X_1, \ldots, X_d)$ sont indépendantes si et seulement si sa matrice de covariance Γ est diagonale.*

Démonstration : Nous avons déjà démontré la condition nécessaire. Supposons que la matrice de covariance Γ est diagonale et notons m le vecteur espérance de X. D'après l'exemple 6.1.2, si $Y_1, \ldots, Y_d$ sont des variables gaussiennes indépendantes d'espérance et variance respectives $m_1, \Gamma_{11}, \ldots, m_d, \Gamma_{dd}$ alors le vecteur $Y = (Y_1, \ldots, Y_d)$ suit la loi $\mathcal{N}_d(m, \Gamma)$. Donc X et Y ont même loi, et en particulier pour tout $1 \leq j \leq d$, X_j a même loi que Y_j.
Si les fonctions $f_j : \mathbb{R} \to \mathbb{R}$, $1 \leq j \leq d$ sont telles que les variables $f_j(X_j)$ sont intégrables, alors les variables $f_j(Y_j)$ sont intégrables et ont les mêmes espérances. D'après la condition nécessaire de la proposition 3.3.11, par indépendance des Y_j, $\prod_{j=1}^d f_j(Y_j)$ est intégrable et $\mathbb{E}\left(\prod_{j=1}^d f_j(Y_j)\right) = \prod_{j=1}^d \mathbb{E}\left(f_j(Y_j)\right)$. La variable $\prod_{j=1}^d f_j(X_j)$ qui a même loi que $\prod_{j=1}^d f_j(Y_j)$ est également intégrable et a même espérance. Ainsi,

$$\mathbb{E}\left(\prod_{j=1}^d f_j(X_j)\right) = \mathbb{E}\left(\prod_{j=1}^d f_j(Y_j)\right) = \prod_{j=1}^d \mathbb{E}\left(f_j(Y_j)\right) = \prod_{j=1}^d \mathbb{E}\left(f_j(X_j)\right).$$

D'après la condition suffisante de la proposition 3.3.11, les coordonnées X_j sont des variables indépendantes. $\square$

Exercice 6.2.2. Soit $Y_1, \ldots, Y_n$ des vecteurs aléatoires à valeurs respectivement dans $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_n}$ t.q. le vecteur $Y = (Y_1, \ldots, Y_n)$ est gaussien (cela implique en particulier que les Y_i sont des vecteurs gaussiens). Montrer que les vecteurs $Y_1, \ldots, Y_n$ sont indépendants si et seulement si la matrice de covariance Γ de Y est diagonale par blocs au sens où :

$$\forall (j,k) \notin \bigcup_{i=1}^{n} [d_1 + \ldots + d_{i-1} + 1, d_1 + \ldots + d_i]^2,\ \Gamma_{jk} = 0.$$

6.2.2 Stabilité par convergence en loi

Le résultat suivant qui assure que toute limite en loi (à fortiori presque sûre, en probabilité, dans L^1 ou dans L^2 puisque ces convergences impliquent la convergence en loi) de variables gaussiennes réelles est gaussienne, est très utile pour obtenir le caractère gaussien de variables obtenues par un procédé de passage à la limite.

Proposition 6.2.3. *Soit $(X_n)_n$ une suite de variables aléatoires gaussiennes réelles qui converge en loi vers X. Alors X est gaussienne.*

Démonstration : Nous nous contenterons de faire la preuve dans le cas beaucoup plus facile où X_n converge vers X dans L^2 i.e. où $\mathbb{E}((X_n - X)^2) \to 0$.
D'après la Proposition 5.1.5, on a alors convergence de $\mathbb{E}(X_n)$ et Var (X_n) respectivement vers $\mathbb{E}(X)$ et Var (X).
On en déduit que pour $u \in \mathbb{R}$, $\mathbb{E}(e^{iuX_n}) = e^{iu\mathbb{E}(X_n) - \frac{u^2}{2}\text{Var }(X_n)} \to e^{iu\mathbb{E}(X) - \frac{u^2}{2}\text{Var }(X)}$. La convergence dans L^2 de X_n vers X entraîne la convergence en loi de X_n vers X et donc d'après le théorème 5.3.9, la convergence de $\mathbb{E}(e^{iuX_n})$ vers $\mathbb{E}(e^{iuX})$. On conclut donc que

$$\forall u \in \mathbb{R},\ \mathbb{E}(e^{iuX}) = e^{iu\mathbb{E}(X) - \frac{u^2}{2}\text{Var }(X)},$$

ce qui implique que X est gaussienne. □

Exercice 6.2.4. Soit $(X_n)_n$ une suite de vecteurs gaussiens à valeurs $\mathbb{R}^d$ qui converge en loi vers un vecteur X.

1. Soit $u \in \mathbb{R}^d$. Montrer que $u.X_n$ converge en loi vers $u.X$. Que peut-on en déduire sur la loi de $u.X$?
2. Conclure que X est gaussien.

6.3 Exercices

Exercice 6.3.1. Soit X et Y des variables gaussiennes réelles indépendantes. Donner une condition nécessaire et suffisante pour que les variables $X + Y$ et $X - Y$ soient indépendantes.

Exercice 6.3.2. Soit X une variable gaussienne centrée réduite et ε une variable vérifiant $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$ indépendantes.

1. Quelle est la loi de $Y = \epsilon X$?
2. Donner la matrice de covariance du vecteur (X, Y).
3. Comparer $\mathbb{E}(X^2Y^2)$ à $\mathbb{E}(X^2)\mathbb{E}(Y^2)$. Les variables X et Y sont-elles indépendantes ?
4. Que peut-on conclure sur le couple (X, Y) ?

Exercice 6.3.3. Soit (X, Y) un couple gaussien de matrice de covariance

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

où $\rho \in [-1, 1]$. Que peut-on dire des variables aléatoires X et $Z = Y - \rho X$?

Exercice 6.3.4. Soit $X_1, \ldots, X_n$ des variables gaussiennes centrées réduites indépendantes et a un vecteur non nul de $\mathbb{R}^d$.

1. Donner une condition nécessaire et suffisante sur a pour que le vecteur aléatoire $X - (a.X)a$ et la variable aléatoire $a.X$ soient indépendants.
2. On note $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ et $V_n = \frac{1}{n-1}\sum_{j=1}^n (X_j - M_n)^2$ la moyenne et la variance empirique.
   (a) Par un bon choix de a, montrer que $\sqrt{n}\ M_n$ est indépendante du vecteur $(X_1 - M_n, X_2 - M_n, \ldots, X_n - M_n)$.
   (b) Conclure que M_n et V_n sont indépendantes.

Exercice 6.3.5. Soit $X_1, \ldots, X_n$ des variables aléatoires réelles I.I.D. et de carré intégrable telles que la moyenne empirique $M_n = \frac{1}{n}\sum_{j=1}^n X_j$ et la variance empirique $V_n = \frac{1}{n-1}\sum_{j=1}^n (X_j - M_n)^2$ sont indépendantes.

1. Vérifier que $V_n = \dfrac{1}{n}\displaystyle\sum_{j=1}^n X_j^2 - \dfrac{1}{n(n-1)}\sum_{\substack{j,k=1 \\ k\neq j}}^n X_jX_k$ et en déduire $\mathbb{E}(V_n)$.
2. En utilisant l'hypothèse d'indépendance, exprimer $\mathbb{E}(V_n e^{iunM_n})$ en fonction de $\mathrm{Var}\,(X_1)$ et de la fonction caractéristique commune des X_j que l'on notera Φ_X.
3. Montrer par ailleurs que
$$\mathbb{E}(V_n e^{iunM_n}) = \mathbb{E}(X_1^2 e^{iuX_1})\Phi_X(u)^{n-1} - (\mathbb{E}(X_1 e^{iuX_1}))^2\Phi_X(u)^{n-2}.$$
4. Remarquer que $\mathbb{E}(X_1 e^{iuX_1})$ et $\mathbb{E}(X_1^2 e^{iuX_1})$ s'exprime simplement à l'aide des deux premières dérivées de Φ_X et en déduire une équation différentielle satisfaite par cette fonction.
5. Poser $f(u) = \Phi'_X(u)/\Phi_X(u)$ et calculer $f'(u)$. En déduire Φ_X puis la loi commune des X_j.

Exercice 6.3.6. Théorème de la limite centrale multidimensionnel
Soit $(X_i, i \geq 1)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ indépendantes suivant toutes la même loi que X. On suppose que $\mathbb{E}\left(|X|^2\right) < +\infty$ (où $|x|$ désigne la norme euclidienne d'un vecteur de $\mathbb{R}^d$). On note m le vecteur espérance de X et Γ sa matrice de covariance. Montrer que

$$\sqrt{n}\left(\frac{1}{n}(X_1 + \cdots + X_n) - m\right)$$

converge en loi vers un vecteur gaussien centré de matrice de covariance Γ.
(<u>Indication :</u> on pourra appliquer le théorème de la limite centrale usuel aux variables aléatoires réelles $(u.X_i)_{i\geq 1}$ où $u \in \mathbb{R}^d$).

6.4 Résumé

- $X = (X_1, \ldots, X_d) : \Omega \to \mathbb{R}^d$ est **un vecteur gaussien** si toute combinaison linéaire de ses coordonnées est une variable gaussienne réelle

  $\iff \Phi_X(u) = e^{iu.m - \frac{1}{2}u.\Gamma u}$ où $m \in \mathbb{R}^d$ et Γ est une matrice $d \times d$ symétrique positive.

  Alors $m_j = \mathbb{E}(X_j)$ et $\Gamma_{ij} = \text{Cov}\ (X_i, X_j)$ et on note $X \sim \mathcal{N}_d(m, \Gamma)$.

- **Stabilité par transformation affine :** si $X \sim \mathcal{N}_d(m, \Gamma)$ et $a \in \mathbb{R}^n$, $M \in \mathbb{R}^{n \times d}$ alors $a + MX \sim \mathcal{N}_n(a + Mm, M\Gamma M^*)$.

- Les coordonnées d'un vecteur gaussien X sont **indépendantes** si et seulement si sa matrice de covariance Γ est **diagonale**.

- Toute limite en loi d'une suite de vecteurs gaussiens est un vecteur gaussien.

Chapitre 7

Quelques exercices corrigés

7.1 Énoncés

7.1.1 Variables aléatoires discrètes

Exercice 7.1.1. Un composant électronique a une durée de vie X qu'on mesure en nombre entier d'unités de temps. On fait l'hypothèse que, à chaque unité de temps, ce composant a une probabilité $p \in]0,1[$ de tomber en panne, de sorte que X suit une loi géométrique $\mathcal{G}(p)$ sur $\{1, 2, \ldots\}$:

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

On considère un autre composant dont la durée de vie Y est indépendante de X et de même loi. On pose

$$S = \min(X, Y) \text{ et } T = |X - Y|.$$

1. Que représentent S et T ?
2. Calculer $\mathbb{P}(S = s \text{ et } T = t)$ pour $s \geq 1$ et $t \geq 0$ (distinguer $t = 0$ de $t \geq 1$).
3. En déduire les lois de S et T puis $\mathbb{E}(T)$. Quel est le nom de la loi de S ?
4. Les variables aléatoires S et T sont-elles indépendantes ?

Problème 7.1.2. Soit $A_1, \cdots, A_n$ une suite d'événements.

1. Montrer que $1_{A_1 \cup \ldots \cup A_n} = 1 - \prod_{i=1}^{n}(1 - 1_{A_i})$.
2. En déduire la formule de Poincaré :

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}).$$

Application : Une personne écrit à n correspondants des lettres personnelles, met chaque lettre dans une enveloppe, ferme les enveloppes, puis écrit les adresses au hasard. On s'intéresse au nombre X_n de lettres qui parviennent à leur destinataire. Pour $1 \leq i \leq n$, on note A_i l'événement : la lettre i arrive à son destinataire.

3. Préciser l'espace fini Ω choisi pour modéliser le problème ainsi que la probabilité $\mathbb{P}$ dont il est muni.
   Pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, calculer $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$.

4. En déduire $\mathbb{P}(X_n > 0)$ puis $\mathbb{P}(X_n = 0)$. Quel est le nombre de permutations σ de $\{1, \ldots, n\}$ sans point fixe i.e. telles que $\forall i \in \{1, \ldots, n\}$, $\sigma(i) \neq i$?
5. En déduire que pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$, $\mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \dfrac{(n-k)!}{n!} \displaystyle\sum_{l=0}^{n-k} \dfrac{(-1)^l}{l!}$ puis donner la loi de X_n.
6. Vérifier que $k\mathbb{P}(X_n = k) = \mathbb{P}(X_n = k-1) - \dfrac{(-1)^{n-(k-1)}}{(k-1)!(n-(k-1))!}$ pour $k \in \{1, \ldots, n\}$. En déduire que $\mathbb{E}(X_n) = 1$. Retrouver ce résultat en remarquant que $X_n = \sum_{i=1}^n 1_{A_i}$.
7. Soit X une variable de loi de Poisson de paramètre 1. Vérifier que pour tout entier k, $\lim_{n \to +\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$. On dit que $(X_n)_n$ converge en loi vers X.

Problème 7.1.3. Pour fidéliser ses clients, une marque de chocolat place dans chaque tablette une pièce d'un puzzle. Le puzzle est composé de n morceaux distincts. Le morceau qui se trouve dans une tablette est supposé suivre une loi uniforme sur les n morceaux possibles. Le client est supposé acheter les différentes tablettes au hasard. On s'intéresse au nombre N d'achats à réaliser pour obtenir l'ensemble des morceaux du puzzle.

Événements

1. Pour $0 \leq m \leq n-1$, que vaut $\mathbb{P}(N > m)$?
2. On suppose maintenant $m \geq n$.
   (a) Pour $1 \leq k \leq n$, on désigne par A_k^m l'événement "la pièce k n'a toujours pas été obtenue au bout de m achats". Calculer pour $k_1, \ldots, k_r \in \{1, \ldots, n\}$ des entiers distincts, la probabilité $\mathbb{P}(A_{k_1}^m \cap \ldots \cap A_{k_r}^m)$ (préciser l'espace fini Ω choisi pour modéliser les pièces obtenues dans les m tablettes et la probabilité dont il est muni).
   (b) Montrer que $\{N > m\} = \displaystyle\bigcup_{k=1}^n A_k^m$.
   (c) En déduire que
   $$\mathbb{P}(N > m) = \sum_{r=1}^n (-1)^{r+1} C_n^r \left(1 - \frac{r}{n}\right)^m.$$
3. En remarquant que pour tout $l \in \mathbb{N}$, $l = \displaystyle\sum_{m=0}^{+\infty} 1_{\{l > m\}}$, montrer que
   $$\mathbb{E}(N) = \sum_{m=0}^{+\infty} \mathbb{P}(N > m).$$
   En déduire une expression pour cette espérance.
4. Donner la loi de N.

Variables aléatoires

La première pièce ayant été découverte dans la première tablette ($N_1 = 1$), on note N_2 le nombre d'achats supplémentaires nécessaires à l'obtention d'une seconde pièce, puis N_3 le nombre d'achats supplémentaires nécessaires à l'obtention d'une troisième, et ainsi de suite.

1. Exprimer N en fonction de N_1,$N_2, \ldots, N_n$.
2. On note X_i le numéro de la i-ème pièce de puzzle obtenue. Pour $m \geq 1$, montrer que
$$\{N_2 = m\} = \bigcup_{i=1}^{n} \{X_1 = i, X_2 = i, \ldots, X_m = i, X_{m+1} \neq i\}.$$
3. Par hypothèse, les X_i sont des variables aléatoires indépendantes, de loi uniforme sur $\{1, \ldots, n\}$. En déduire la loi de N_2.
4. Pour $m_2, m_3 \geq 1$, exprimer l'événement $\{N_2 = m_2, N_3 = m_3\}$ en fonction des X_i. Calculer sa probabilité et en déduire l'indépendance de N_2 et N_3.
5. Justifier intuitivement que les N_i sont des variables de loi géométrique et donner leurs paramètres.
6. En déduire une autre expression de l'espérance de N.
7. Conclure que
$$\sum_{r=1}^{n-1} (-1)^{r+1} C_n^r \frac{1}{r} \left(1 - \frac{r}{n}\right)^n = \sum_{i=2}^{n} \frac{1}{i}.$$

7.1.2 Variables aléatoires à densité

Exercice 7.1.4. On coupe un bâton de longueur 1 au hasard en trois morceaux : les abscisses U et V des découpes sont supposées indépendantes et uniformément réparties sur $[0, 1]$. On veut calculer la probabilité p pour que l'on puisse faire un triangle avec les trois morceaux (on peut faire un triangle avec trois segments de longueur l_1, l_2 et l_3 si et seulement si $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$).

1. Exprimer en fonction de U et V les longueurs respectives L_1, L_2 et L_3 du morceau le plus à gauche, du morceau du milieu et du morceau le plus à droite.
2. Montrer que
$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$
3. Calculer p.

Exercice 7.1.5. Soient X et Y deux variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi de Pareto de paramètre 2 i.e. qui possèdent la densité $p(x) = \dfrac{1_{\{x>1\}}}{x^2}$. On pose
$$(Z, W) = \left(\ln(X), 1 + \frac{\ln(Y)}{\ln(X)}\right).$$

1. Quelle est la loi de (Z, W) ? Les variables Z et W sont-elles indépendantes ?
2. Quelle est la loi de W ?
3. Quelle est la loi de Z ?

Exercice 7.1.6. Soient X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\Gamma(p, 1)$ et $\Gamma(q, 1)$, $p > 0$ et $q > 0$. On rappelle qu'une variable aléatoire de loi $\Gamma(a, 1)$, $a > 0$, a pour densité
$$f_a(x) = \frac{1}{\Gamma(a)} x^{a-1} \exp(-x) 1_{]0,\infty[}(x)$$

où $\Gamma(a) = \displaystyle\int_0^\infty x^{a-1}\exp(-x)dx$ vérifie $\forall a > 0,\ \Gamma(a+1) = a\Gamma(a)$.

1. Déterminer l'espérance et la variance de X.
2. On veut calculer la loi de la variable aléatoire $Z = \dfrac{X}{Y}$. On dit que Z suit la loi bêta de seconde espèce de paramètres p et q.
   (a) On pose $W = X - Y$. Exprimer en fonction de X et Y l'événement $\{Z < 1, W > 0\}$. En déduire sans calcul sa probabilité. Les variables Z et W sont-elles indépendantes ?
   (b) Déterminer la loi du vecteur (Z, W).
   (c) En déduire la loi de la variable aléatoire Z.
3. Déterminer la loi de $T = \dfrac{Z}{1+Z}$.

Problème 7.1.7. Soient $\Theta \sim \mathcal{U}[-\pi, \pi]$ et R une variable aléatoire réelle indépendante de Θ possédant une densité p nulle sur $\mathbb{R}_-$. On pose

$$(X, Y) = (R\cos\Theta, R\sin\Theta).$$

1. Déterminer la loi du vecteur (X, Y). Que peut-on dire des lois des vecteurs (X, Y) et (Y, X) ? En déduire que X et Y ont la même densité.
2. Pour $a > 0$, quelle est la loi de $Z = aX/Y$ (on pourra exprimer Z en fonction de Θ) ? En déduire sans calcul la loi de l'inverse d'une variable de Cauchy de paramètre a.
3. Quelle doit être la loi de R pour que le couple (X, Y) soit uniformément distribué sur le disque $D(0, \rho)$ où $\rho > 0$ (i.e. possède une densité constante sur ce disque et nulle en dehors) ?
4. On suppose la densité p de R continue sur $\mathbb{R}_+$ et strictement positive sur $\mathbb{R}_+^*$. On souhaite trouver à quelle condition sur p les variables X et Y sont indépendantes.
   (a) Dans le cas où X et Y sont indépendantes, si q désigne leur densité commune, exprimer $q(x)q(y)$ en fonction de p.
   En commençant par choisir $y = 0$ dans cette relation, établir que pour $x, y \in \mathbb{R}^*$,
   $$\frac{p(|x|)}{2\pi q(0)|x|} \times \frac{p(|y|)}{2\pi q(0)|y|} = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}.$$
   (b) Pour $s > 0$, on pose $f(s) = \ln\left(\frac{p(\sqrt{s})}{2\pi q^2(0)\sqrt{s}}\right)$. Vérifier que f est une fonction continue sur $\mathbb{R}_+^*$ telle que pour $s, t > 0$, $f(t) + f(s) = f(t+s)$.
   En déduire qu'il existe une constante $\sigma > 0$ telle que $p(r) = \frac{r}{\sigma^2}e^{-\frac{r^2}{2\sigma^2}}1_{\{r>0\}}$ i.e. que R suit la loi de Rayleigh de paramètre σ^2. Quelle est alors la loi de (X, Y) ?
   La loi de Rayleigh intervient notamment dans l'analyse d'erreur de positionnement du système GPS.

7.1.3 Convergence des variables aléatoires

Exercice 7.1.8. On considère une machine qui tourne à une cadence de n cycles par unité de temps et qui a une probabilité $p \in]0, 1[$ de tomber en panne lors de chaque cycle et ce

indépendamment des autres cycles. Soit N le numéro du cycle où se produit la première panne.

1. Quelle est la loi de N ? Que représente $T = N/n$?
   Soit τ l'espérance de T. Exprimer τ en fonction de p et n.
2. Dans le cas où n est grand, on souhaite approcher la variable discrète T par une variable aléatoire à densité. Pour cela on va effectuer le passage à la limite $n \to +\infty$ en choisissant $p(n)$ de façon à ce que l'espérance de $T(n)$ reste égale à τ.
   (a) Que vaut $p(n)$?
   (b) Calculer la fonction caractéristique de $T(n)$.
   (c) En déduire que cette variable converge en loi vers une limite que l'on précisera.

Exercice 7.1.9. Chaque jour, dans une certaine ville, 100 personnes ont besoin d'un examen radioscopique. Pour préserver le libre choix, n centres d'imagerie sont installés dans cette ville. On admet que les patients choisissent indifféremment l'un ou l'autre centre d'imagerie.
Soit N le nombre de clients journaliers dans un centre d'imagerie.

1. Quelle est la probabilité qu'un client choisisse le centre d'imagerie considéré ?
2. Montrer que N peut s'écrire $N = \sum_{i=1}^{i=100} X_i$ où les $(X_i)_{1\leq i\leq n}$ sont n variables aléatoires indépendantes et distribuées suivant la loi de Bernoulli de même paramètre p que l'on déterminera.
3. Quelle est la loi de N ?
4. On donne que si Y suit la loi normale centrée réduite, alors $\mathbb{P}(Y \leq 2) = 98\%$.
   En utilisant le théorème de la limite centrale, déterminer quelle capacité $c(n)$ chaque centre d'imagerie doit avoir pour être capable de répondre à la demande avec une probabilité de 98% ? Cas où $n = 2$, $n = 3$, $n = 4$?
5. Quel est le coût de la concurrence : quelle surcapacité $s(n)$ la concurrence entraîne-t-elle par rapport à une situation où chaque centre se verrait affecter un même nombre de clients ? Cas où $n = 2$, $n = 3$, $n = 4$?

Exercice 7.1.10.
1. Calculer la fonction caractéristique d'une variable X suivant la loi de Poisson $\mathcal{P}(\lambda)$. Calculer $\mathbb{E}(X)$ et Var (X).
2. Soit $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ deux variables de Poisson indépendantes. Quelle est la loi de leur somme ?
3. Soit $X_n \sim \mathcal{P}(n)$. Montrer que pour $n \to +\infty$, $X_n^* = (X_n - n)/\sqrt{n}$ converge en loi vers une limite que l'on précisera. Commenter ce résultat à la lumière du théorème de la limite centrale.

Problème 7.1.11. Un automobiliste emprunte tous les jours le même trajet qui comporte un feu tricolore pour se rendre à son travail. Comme le trajet est peu encombré, lorsque le feu est rouge, l'automobiliste peut redémarrer dès que le feu passe au vert. Mais pour faire passer le temps, il se demande quelle est la durée θ pendant laquelle le feu reste rouge. On note $(X_i)_{i\geq 1}$ ses durées d'attente successives au feu lorsque celui-ci est rouge et on suppose ces variables i.i.d. suivant la loi uniforme $\mathcal{U}[0, \theta]$. Pour $n \geq 1$, on note $Z_n = \max_{1\leq i\leq n} X_i$.

1. Montrer que $\mathbb{P}(X_1 = X_2) = 0$.

2. Montrer que plus généralement $\mathbb{P}(\exists 1 \leq i \neq j \leq n, X_i = X_j) = 0$. En déduire que pour toute fonction f bornée sur $\mathbb{R}$, on a :

$$\mathbb{E}[f(Z_n)] = \sum_{i=1}^{n} \mathbb{E}\left[f(X_i)1_{\{\forall j \neq i,\, X_j < X_i\}}\right] = n\mathbb{E}\left[f(X_1)1_{\{X_2 < X_1, \ldots, X_n < X_1\}}\right].$$

3. En déduire que Z_n suit la densité $\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1} 1_{\{0 \leq z \leq \theta\}}$.
4. Calculer $\mathbb{E}(Z_n)$ et Var (Z_n). En conclure que Z_n converge presque sûrement vers θ (*indication* : on rappelle (voir preuve du théorème 5.2.2) que si $(\tilde{Z}_n)_{n\geq 1}$ est une suite de variables aléatoires telle que $\sum_{n\geq 1} \mathbb{E}(\tilde{Z}_n^2) < +\infty$, alors la suite $(\tilde{Z}_n)_{n\geq 1}$ converge presque sûrement vers 0).
5. Soit f une fonction bornée sur $\mathbb{R}$. Montrer que :

$$\mathbb{E}\left[f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right] = \int_0^n f(u)\left(1 - \frac{u}{n}\right)^{n-1} du$$

On rappelle que pour $0 < v < 1$, $\ln(1-v) \leq -v$. En déduire que pour $n \geq 2$, $\left(1 - \frac{u}{n}\right)^{n-1} 1_{\{0<u\leq n\}} \leq e^{-u/2} 1_{\{u>0\}}$ et conclure que :

$$\lim_{n\to\infty} \mathbb{E}\left[f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right] = \int_0^\infty f(u)e^{-u}\, du$$

6. En déduire que $\frac{n}{\theta}(\theta - Z_n)$ converge en loi et identifier la limite.
7. Montrer que $M_n = \frac{2}{n}(X_1 + \cdots + X_n)$ converge presque sûrement vers θ. À quelle vitesse a lieu cette convergence ?
8. L'automobiliste doit-il préférer Z_n ou M_n pour estimer rapidement θ ?

7.2 Corrigés

7.2.1 Variables aléatoires discrètes

Corrigé de l'exercice 7.1.1. Un composant électronique a une durée de vie X qu'on mesure en nombre entier d'unités de temps. On fait l'hypothèse que, à chaque unité de temps, ce composant a une probabilité $p \in]0,1[$ de tomber en panne, de sorte que X suit une loi géométrique $\mathcal{G}(p)$ sur $\{1, 2, \cdots\}$:

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

On considère un autre composant dont la durée de vie Y est indépendante de X et de même loi. On pose

$$S = \min(X, Y) \text{ et } T = |X - Y|.$$

1. **Que représentent S et T ?**
   S représente le premier temps de panne et T la durée qui sépare les deux temps de panne.
2. **Calculer $\mathbb{P}(S = s$ et $T = t)$ pour $s \geq 1$ et $t \geq 0$ (distinguer $t = 0$ et $t \geq 1$).**
   Pour $s \geq 1$ et $t \geq 0$, on a toujours

$$\{S = s, T = t\} = \{X = s, Y = s + t\} \cup \{X = s + t, Y = s\}$$

mais l'union n'est disjointe que si $t \geq 1$ (sinon les deux événements sont égaux à $\{X = Y = s\}$). Donc pour $s, t \geq 1$, on a, en utilisant l'indépendance de X et Y,

$$\begin{aligned}
\mathbb{P}(S = s \text{ et } T = t) &= \mathbb{P}(X = s \text{ et } Y = s + t) + \mathbb{P}(Y = s \text{ et } X = s + t) \\
&= \mathbb{P}(X = s)\mathbb{P}(Y = s + t) + \mathbb{P}(Y = s)\mathbb{P}(X = s + t) \\
&= 2p^2(1 - p)^{2s+t-2}.
\end{aligned}$$

Pour $s \geq 1$ et $t = 0$, on a, toujours par indépendance de X et Y,

$$\begin{aligned}
\mathbb{P}(S = s \text{ et } T = 0) &= \mathbb{P}(X = s \text{ et } Y = s) \\
&= \mathbb{P}(X = s)\mathbb{P}(Y = s) = p^2(1 - p)^{2s-2}.
\end{aligned}$$

On conclut que

$$\forall s \geq 1,\ \forall t \geq 0,\ \mathbb{P}(S = s, T = t) = p^2 \left(1 + 1_{\{t>0\}}\right)(1 - p)^t(1 - p)^{2(s-1)}. \tag{7.1}$$

3. **En déduire les lois de S et T puis $\mathbb{E}(T)$. Quel est le nom de la loi de S ?**
Pour $s \geq 1$, on a par la formule des lois marginales

$$\begin{aligned}
\mathbb{P}(S = s) &= \sum_{t=0}^{\infty} \mathbb{P}(S = s \text{ et } T = t) \\
&= p^2(1 - p)^{2s-2} + \sum_{t=1}^{\infty} 2p^2(1 - p)^{2s+t-2} \\
&= (1 - p)^{2(s-1)} \left(p^2 + 2p^2(1 - p)\sum_{u=0}^{\infty}(1 - p)^u\right) \text{ où } u = t - 1 \\
&= \left(p^2 + 2p^2(1 - p)\frac{1}{p}\right)(1 - p(2 - p))^{s-1} = p(2 - p)(1 - p(2 - p))^{s-1}.
\end{aligned}$$

On reconnaît la loi géométrique de paramètre $p(2 - p)$.
Maintenant, toujours par la formule des lois marginales, pour $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(T = t) &= \sum_{s=1}^{\infty} \mathbb{P}(S = s \text{ et } T = t) = \sum_{s=1}^{\infty} \left(1 + 1_{\{t>0\}}\right) p^2(1 - p)^{2(s-1)+t} \\
&= \left(1 + 1_{\{t>0\}}\right) p^2(1 - p)^t \sum_{u=0}^{\infty}(1 - p(2 - p))^u = \left(1 + 1_{\{t>0\}}\right) \frac{p(1 - p)^t}{2 - p}.
\end{aligned}$$

D'où

$$\mathbb{E}(T) = \sum_{t=0}^{\infty} t\mathbb{P}(T = t) = \sum_{t=1}^{\infty} t\frac{2p(1 - p)^t}{2 - p} = \frac{2p(1 - p)}{2 - p}\sum_{t=1}^{\infty} t(1 - p)^{t-1} = \frac{2p(1 - p)}{2 - p}f'(1 - p).$$

où $f(x) = \sum_{t=0}^{\infty} x^t = \frac{1}{1-x}$. Comme $f'(x) = \frac{1}{(1-x)^2}$, $f'(1 - p) = 1/p^2$ et $\mathbb{E}(T) = \frac{2(1-p)}{p(2-p)}$.

4. **Les variables aléatoires S et T sont-elles indépendantes ?**
On peut montrer l'indépendance en utilisant la définition. En effet, on a pour $s \geq 1$ et $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(S = s)\mathbb{P}(T = t) &= p(2 - p)[(1 - p)^2]^{s-1} \left(1 + 1_{\{t>0\}}\right) \frac{p(1 - p)^t}{2 - p} \\
&= p^2 \left(1 + 1_{\{t>0\}}\right)(1 - p)^{t+2(s-1)} = \mathbb{P}(S = s \text{ et } T = t),
\end{aligned}$$

d'après (7.1). On peut aussi remarquer sur (7.1) que la loi du couple (S, T) se met sous forme produit $c\mu(s)\nu(t)$ et conclure que ces variables sont indépendantes en utilisant la remarque 2.2.12 du polycopié.

Corrigé du problème 7.1.2. Soit $A_1, \cdots, A_n$ une suite d'événements.

1. **Montrer que $1_{A_1 \cup \ldots \cup A_n} = 1 - \prod_{i=1}^n (1 - 1_{A_i})$.**

$$1_{A_1 \cup \ldots \cup A_n} = 1 - 1_{(A_1 \cup \ldots \cup A_n)^c} = 1 - 1_{A_1^c \cap \ldots \cap A_n^c} = 1 - \prod_{i=1}^n 1_{A_i^c} = 1 - \prod_{i=1}^n (1 - 1_{A_i}).$$

2. **En déduire la formule de Poincaré :**

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}).$$

En écrivant,

$$\prod_{i=1}^n (1 - 1_{A_i}) = \sum_{k=0}^n (-1)^k \sum_{1 \le i_1 < \ldots < i_k \le n} 1_{A_{i_1}} \ldots 1_{A_{i_k}},$$

et en prenant l'espérance, on obtient la formule de Poincaré.

Application : Une personne écrit à n correspondants des lettres personnelles, met chaque lettre dans une enveloppe, ferme les enveloppe puis écrit les adresses au hasard. On s'intéresse au nombre X_n de lettres qui parviennent à leur destinataire. Pour $1 \le i \le n$, on note A_i l'événement : la lettre i arrive à son destinataire.

3. **Préciser l'espace fini Ω choisi pour modéliser le problème ainsi que la probabilité $\mathbb{P}$ dont il est muni.** On suppose que pour $i \in \{1, \ldots, n\}$, le i-ième correspondant est celui pour lequel i-ième lettre a été écrite. On considère $\Omega = \mathcal{S}_n$, l'espace des permutations de $\{1, \cdots, n\}$. Pour $\sigma \in \Omega$ et $i \in \{1, \ldots, n\}$, $\sigma(i)$ désigne le numéro de la lettre reçue par le i-ième correspondant. On munit Ω de la probabilité uniforme. On a Card $(\Omega) = n!$.
Pour $1 \le k \le n$ et $1 \le i_1 < i_2 < \ldots < i_k \le n$, calculer $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})$.
$A_{i_1} \cap \ldots \cap A_{i_k}$ représente l'ensemble des permutations σ telles que pour $1 \le j \le k$, $\sigma(i_j) = i_j$. On a Card $(A_{i_1} \cap \ldots \cap A_{i_k}) = (n-k)!$ et $\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \dfrac{(n-k)!}{n!}$.

4. **En déduire $\mathbb{P}(X_n > 0)$ puis $\mathbb{P}(X_n = 0)$. Quel est le nombre de permutations σ de $\{1, \cdots, n\}$ sans point fixe i.e. telles que $i \in \{1, \cdots, n\}$, $\sigma(i) \neq i$?**
L'événement au moins une lettre parvient à son destinataire est $\{X_n > 0\} = A_1 \cup \ldots \cup A_n$. En utilisant la formule de Poincaré, on obtient

$$\begin{aligned}\mathbb{P}(X_n > 0) = \mathbb{P}(A_1 \cup \ldots \cup A_n) &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{1 \le i_1 < \ldots < i_k \le n} \frac{(n-k)!}{n!} = \sum_{k=1}^n \frac{(-1)^{k+1}}{k!},\end{aligned}$$

puisque Card $(\{(i_1, \ldots, i_k) : 1 \le i_1 < \ldots < i_k \le n\}) = C_n^k$. Donc

$$\mathbb{P}(X_n = 0) = 1 - \mathbb{P}(X_n > 0) = \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Le nombre de permutations σ de $\{1, \cdots, n\}$ sans point fixe est égal à

$$\text{Card } (\{X_n = 0\}) = n! \mathbb{P}(X_n = 0) = n! \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

5. **En déduire que pour $1 \leq k \leq n$ et $1 \leq i_1 < i_2 < \ldots < i_k \leq n$,**
$\mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \dfrac{(n-k)!}{n!} \displaystyle\sum_{l=0}^{n-k} \frac{(-1)^l}{l!}$ **puis donner la loi de X_n.**
L'application qui à $\sigma \in \mathcal{S}_n$ associe sa restriction à $\mathcal{T} = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$ établit une bijection entre $A_{i_1} \cap \ldots \cap A_{i_k}$ et l'ensemble des permutations de $\mathcal{T}$. L'image de $\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}$ par cette restriction est l'ensemble des permutations sans point fixe de $\mathcal{T}$ dont le cardinal est $(n-k)! \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}$ d'après la question précédente. Donc

$$\text{Card}\,(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = (n-k)! \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}$$

$$\text{puis} \quad \mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) = \frac{(n-k)!}{n!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}.$$

Comme $\{X_n = k\} = \displaystyle\bigcup_{1 \leq i_1 < \ldots < i_k \leq n} \{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}$ où l'union est disjointe,

$$\begin{aligned}\mathbb{P}(X_n = k) &= \sum_{1 \leq i_1 < \ldots < i_k \leq n} \mathbb{P}(\{X_n = k\} \cap A_{i_1} \cap \ldots \cap A_{i_k}) \\ &= C_n^k \frac{(n-k)!}{n!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!} = \frac{1}{k!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!}.\end{aligned}$$

6. **Vérifier que $k\mathbb{P}(X_n = k) = \mathbb{P}(X_n = k-1) - \dfrac{(-1)^{n-(k-1)}}{(k-1)!(n-(k-1))!}$ pour $k \in \{1, \ldots, n\}$.**
En déduire que $\mathbb{E}(X_n) = 1$.
Il est facile de vérifier que

$$\begin{aligned}k\mathbb{P}(X_n = k) - \mathbb{P}(X_n = k-1) &= \frac{1}{(k-1)!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!} - \frac{1}{(k-1)!} \sum_{l=0}^{n-k+1} \frac{(-1)^l}{l!} \\ &= -\frac{(-1)^{n-k+1}}{(k-1)!(n-k+1)!}.\end{aligned}$$

Pour l'espérance de X_n, on a

$$\begin{aligned}\mathbb{E}(X_n) &= \sum_{k=0}^{n} k\mathbb{P}(X_n = k) = \sum_{k=1}^{n} k\mathbb{P}(X_n = k) \\ &= \sum_{k=1}^{n} \mathbb{P}(X_n = k-1) - \sum_{k=1}^{n} \frac{(-1)^{n-(k-1)}}{(k-1)!(n-(k-1))!} \\ &= 1 - \mathbb{P}(X_n = n) - \sum_{l=0}^{n-1} \frac{(-1)^{n-l}}{l!(n-l)!} = 1 - \frac{1}{n!} \sum_{l=0}^{n} C_n^l (-1)^{n-l} = 1.\end{aligned}$$

Comme $X_n = \sum_{i=1}^{n} 1_{A_i}$, par linéarité de l'espérance,

$$\mathbb{E}(X_n) = \sum_{i=1}^{n} \mathbb{P}(A_i) = n \times \frac{1}{n} = 1.$$

7. **Soit X une variable de loi de Poisson de paramètre 1. Vérifier que pour tout entier k, $\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$. On dit que $(X_n)_n$ converge en loi vers X.**
Soit $k \in \mathbb{N}$,

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \lim_{n\to\infty} \frac{1}{k!} \sum_{l=0}^{n-k} \frac{(-1)^l}{l!} = \frac{e^{-1}}{k!} = \mathbb{P}(X = k).$$

Corrigé du problème 7.1.3. Pour fidéliser ses clients, une marque de chocolat place dans chaque tablette une pièce d'un puzzle. Le puzzle est composé de n morceaux distincts. Le morceau qui se trouve dans une tablette est supposé suivre une loi uniforme sur les n morceaux possibles. Le client est supposé acheter les différentes tablettes au hasard. On s'intéresse au nombre N d'achats à réaliser pour obtenir l'ensemble des morceaux du puzzle.
Événements

1. **Pour $0 \leq m \leq n-1$, que vaut $\mathbb{P}(N > m)$?**
   Il est clair qu'il faut au minimum n achats pour obtenir l'ensemble des pièces du puzzle. Ainsi, nécessairement $N \geq n$ et donc pour $m \leq n-1$, $\mathbb{P}(N > m) = 1$.

2. **On suppose maintenant $m \geq n$.**

   (a) **Pour $1 \leq k \leq n$, on désigne par A_k^m l'événement "la pièce k n'a toujours pas été obtenue au bout de m achats". Calculer pour $k_1, \ldots, k_r \in \{1, \ldots, n\}$ des entiers distincts, la probabilité $\mathbb{P}(A_{k_1}^m \cap \ldots \cap A_{k_r}^m)$.**
   On modélise l'achat des m tablettes par la suite $\omega = (\omega_1, \ldots, \omega_m) \in \{1, \ldots, n\}^m$ des numéros des tablettes obtenues. On ne dispose d'aucun élément pour affirmer qu'une configuration ω particulière est plus probable qu'une autre ; on munit donc $\Omega = \{1, \ldots, n\}^m$ de la probabilité uniforme $\mathbb{P}$. Une configuration ω est dans $A_{k_1}^m \cap \ldots \cap A_{k_r}^m$ si et seulement si au bout de m achats, on n'a toujours pas obtenu les pièces $k_1, \ldots, k_r$. Il est donc clair que $A_{k_1}^m \cap \ldots \cap A_{k_r}^m = (\{1, \ldots, n\} \setminus \{k_1, \ldots, k_r\})^m$. Par conséquent :

$$\begin{aligned} \mathbb{P}\left(A_{k_1}^m \cap \ldots \cap A_{k_r}^m\right) &= \frac{\text{Card } (\{1, \ldots, n\} \setminus \{k_1, \ldots, k_r\})^m}{\text{Card } (\Omega)} \\ &= \frac{(n-r)^m}{n^m} = \left(1 - \frac{r}{n}\right)^m \end{aligned}$$

   (b) **Montrer que $\{N > m\} = \bigcup_{k=1}^{n} A_k^m$.**
   Soit maintenant $\omega \in \{N > m\}$. Pour cette configuration ω particulière, toutes les pièces n'ont pas encore été obtenues au bout de m achats. Ceci signifie également qu'au bout de m achats, il manque encore au moins une pièce.

   (c) **En déduire que $\mathbb{P}(N > m) = \sum_{r=1}^{n} (-1)^{r+1} C_n^r \left(1 - \frac{r}{n}\right)^m$.** Par application de la formule de Poincaré, on peut alors calculer :

$$\begin{aligned} \mathbb{P}(N > m) &= \mathbb{P}\left(\bigcup_{k=1}^{n} A_k^m\right) \\ &= \sum_{r=1}^{n} (-1)^{r+1} \sum_{\mathcal{K} \subset \{1, \ldots, n\}, \text{Card } (\mathcal{K}) = r} \mathbb{P}\left(\cap_{j \in \mathcal{K}} A_j^m\right). \end{aligned}$$

   Or, d'après 2.(a), $\mathbb{P}\left(\cap_{j \in \mathcal{K}} A_j^m\right) = (1 - r/n)^m$ dès lors que Card $(\mathcal{K}) = r$. Comme il y a exactement C_n^r parties $\mathcal{K}$ de cardinal r dans $\{1, \ldots, n\}$, la formule proposée est démontrée.

3. **En remarquant que $\forall l \in \mathbb{N}$, $l = \sum_{m=0}^{+\infty} 1_{\{l > m\}}$, m.q. $\mathbb{E}(N) = \sum_{m=0}^{+\infty} \mathbb{P}(N > m)$. En déduire une expression pour cette espérance.**
   En appliquant à $N(\omega)$ la décomposition de l'énoncé, il vient :

$$N(\omega) = \sum_{m=0}^{+\infty} 1_{\{N(\omega) > m\}}$$

Or pour tout ω, les deux expressions $1_{\{N(\omega)>m\}}$ et $1_{\{N>m\}}(\omega)$ sont égales et ainsi :

$$N(\omega) = \sum_{m=0}^{+\infty} 1_{\{N>m\}}(\omega)$$

On calcule l'espérance des deux membres et, tout les termes étant positifs, on permute l'espérance et la somme :

$$\mathbb{E}(N) = \mathbb{E}\left(\sum_{m=0}^{+\infty} 1_{\{N>m\}}\right) = \sum_{m=0}^{+\infty} \mathbb{E}\left(1_{\{N>m\}}\right) = \sum_{m=0}^{+\infty} \mathbb{P}\left(N > m\right).$$

En utilisant les deux expressions de $\mathbb{P}\left(N > m\right)$ suivant que $m \leq n-1$ (voir 1.) ou que $m \geq n$ (voir 2.(c)), on obtient :

$$\begin{aligned}\mathbb{E}(N) &= n + \sum_{m=n}^{+\infty}\left[\sum_{r=1}^{n}(-1)^{r+1}C_n^r\left(1-\frac{r}{n}\right)^m\right] = n + \sum_{r=1}^{n}(-1)^{r+1}C_n^r\left[\sum_{m=n}^{+\infty}\left(1-\frac{r}{n}\right)^m\right] \\ &= n + \sum_{r=1}^{n}(-1)^{r+1}C_n^r\frac{n}{r}\left(1-\frac{r}{n}\right)^n = n + n\sum_{r=1}^{n-1}(-1)^{r+1}\frac{1}{r}\left(1-\frac{r}{n}\right)^n.\end{aligned}$$

4. **Donner la loi de N.**
La variable aléatoire N prend ses valeurs dans $\{n, n+1, \ldots\}$. Pour $m \geq n$, on a :

$$\mathbb{P}\left(N = m\right) = \mathbb{P}\left(N > m-1\right) - \mathbb{P}\left(N > m\right).$$

On calcule cette valeur en distinguant suivant que $m \geq n+1$ ou $m = n$. Dans le premier cas, on utilise l'expression en 2.(c) pour obtenir :

$$\mathbb{P}\left(N = m\right) = \sum_{r=1}^{n}(-1)^{r+1}C_n^r\frac{r}{n}\left(1-\frac{r}{n}\right)^{m-1}.$$

Dans le cas où $m = n$, on utilise 1. pour écrire que $\mathbb{P}\left(N > n-1\right) = 1$ et donc $\mathbb{P}\left(N = n\right) = 1 - \mathbb{P}\left(N > n\right)$. Ainsi,

$$\mathbb{P}\left(N = n\right) = -\sum_{r=0}^{n}(-1)^{r+1}C_n^r\left(1-\frac{r}{n}\right)^n.$$

Remarques

- La première remarque que l'on peut faire est que $\{N = n\} = \mathcal{S}_n$, ensemble des permutations de $\{1, \ldots, n\}$. Aussi, $\mathbb{P}\left(N = n\right) = n!/n^n$.
- On vérifie ensuite sur l'expression 2.(c) que $\lim_{m\to\infty}\mathbb{P}\left(N > m\right) = 0$. Ainsi $\mathbb{P}\left(N < +\infty\right) = 1$ et on finit toujours par obtenir toutes les pièces.

Variables aléatoires

La première pièce ayant été découverte dans la première tablette ($N_1 = 1$), on note N_2 le nombre d'achats supplémentaires nécessaires à l'obtention d'une seconde pièce, puis N_3 le nombre d'achats supplémentaires nécessaires à l'obtention d'une troisième, et ainsi de suite.

1. **Exprimer N en fonction de $N_1, N_2, \ldots, N_n$.**
Il est clair que pour obtenir toutes les pièces du puzzle, il faut obtenir la première, puis la seconde, puis la troisième, etc. Ainsi, $N = N_1 + N_2 + \cdots + N_n$.
2. **On note X_i le numéro de la i-ème pièce de puzzle obtenue. Pour $m \geq 1$, montrer que $\{N_2 = m\} = \bigcup_{i=1}^{n}\{X_1 = i, X_2 = i, \ldots, X_m = i, X_{m+1} \neq i\}$.**
Soit $\omega \in \{N_2 = m\}$ et soit i le numéro de la première pièce obtenue i.e. $X_1(\omega)$. Dire que $N_2(\omega) = m$, c'est dire qu'il a fallu m pièces supplémentaires pour obtenir une pièce différente de la première. C'est donc dire que $X_2(\omega) = \ldots = X_m(\omega) = i$ mais que $X_{m+1}(\omega) \neq i$. D'où le résultat.

3. **Par hypothèse, les X_i sont des variables aléatoires indépendantes, de loi uniforme sur $\{1, \ldots, n\}$. En déduire la loi de N_2.**
Les différents éléments dans la réunion ci-dessus sont disjoints car ils correspondent à des valeurs distinctes de X_1. Aussi,

$$\begin{aligned}\mathbb{P}(N_2 = m) &= \sum_{i=1}^{n} \mathbb{P}(X_1 = i, X_2 = i, \ldots, X_m = i, X_{m+1} \neq i) \\ &= \sum_{i=1}^{n} \mathbb{P}(X_1 = i)\,\mathbb{P}(X_2 = i) \ldots \mathbb{P}(X_m = i)\,\mathbb{P}(X_{m+1} \neq i)\end{aligned}$$

On a utilisé dans la dernière égalité l'indépendance des variables X_i. Comme elles suivent une même loi uniforme sur $\{1, \ldots, n\}$, il est clair que :

$$\mathbb{P}(X_1 = i) = \mathbb{P}(X_2 = i) = \ldots = \mathbb{P}(X_m = i) = \frac{1}{n}, \qquad \mathbb{P}(X_{m+1} \neq i) = \frac{n-1}{n}.$$

Par conséquent,

$$\mathbb{P}(N_2 = m) = \sum_{i=1}^{n} \frac{1}{n^m} \times \frac{n-1}{n} = \left(1 - \frac{1}{n}\right) \times \left(\frac{1}{n}\right)^{m-1}.$$

D'après le cours, ceci signifie que N_2 suit la loi géométrique de paramètre $1 - 1/n$.

4. **Pour $m_2, m_3 \geq 1$, exprimer l'événement $\{N_2 = m_2, N_3 = m_3\}$ en fonction des X_i. Calculer sa probabilité et en déduire l'indépendance de N_2 et N_3.**
En examinant les numéros des deux premières pièces différentes obtenues comme ci-dessus, on peut écrire :

$$\begin{aligned}\{N_2 = m_2, N_3 = m_3\} = \bigcup_{1 \leq i \neq j \leq n} \{&X_1 = i, X_2 = i, \ldots, X_{m_2} = i, X_{m_2+1} = j, \\ &X_{m_2+2} \in \{i, j\}, \ldots, X_{m_2+m_3} \in \{i, j\}, X_{m_2+m_3+1} \notin \{i, j\}\}\end{aligned}$$

On en déduit en utilisant le fait que les événements ci-dessus sont disjoints puis que les X_i sont indépendants que :

$$\begin{aligned}\mathbb{P}(N_2 = m_2, N_3 = m_3) &= n(n-1)\frac{1}{n^{m_2+1}} \left(\frac{2}{n}\right)^{m_3-1} \left(\frac{n-2}{n}\right) \\ &= \left(1 - \frac{1}{n}\right) \times \left(\frac{1}{n}\right)^{m_2-1} \times \left(1 - \frac{2}{n}\right) \times \left(\frac{2}{n}\right)^{m_3-1}\end{aligned}$$

Ceci étant vrai pour tous $m_2, m_3 \geq 1$, cette égalité prouve d'une part l'indépendance de N_2 et N_3 (l'expression est le produit d'une fonction de m_2 par une fonction de m_3) et d'autre part que N_3 suit la loi géométrique de paramètre $1 - 2/n$.

5. **Justifier intuitivement que les N_i sont des variables de loi géométrique et donner leurs paramètres.**
Le raisonnement intuitif est le suivant : une fois que les $(k-1)$ premières pièces ont été obtenues, on achète des tablettes jusqu'à obtenir une nouvelle pièce. À chaque achat, on a une probabilité $(k-1)/n$ d'obtenir une pièce que l'on possédait déjà et une probabilité $(1 - (k-1)/n)$ d'en obtenir une que l'on ne possédait pas. Les achats successifs sont indépendants. On reconnaît ainsi le temps de premier succès dans une suite d'expériences indépendantes avec probabilité de succès $(1 - (k-1)/n)$, qui suit la loi géométrique de paramètre cette probabilité. Ainsi, N_k est une variable de loi géométrique de paramètre $(1 - (k-1)/n)$.

6. **En déduire une autre expression de l'espérance de N.**
On sait que l'espérance d'une variable géométrique est l'inverse de son paramètre. Ainsi, $\mathbb{E}(N_k) = \frac{n}{n-k+1}$ et donc d'après 1., il vient :

$$\mathbb{E}(N) = \sum_{k=1}^{n} \frac{n}{n-k+1} = n \sum_{i=1}^{n} \frac{1}{i}.$$

7. **Conclure que $\sum_{r=1}^{n-1}(-1)^{r+1} C_n^r \frac{1}{r}\left(1-\frac{r}{n}\right)^n = \sum_{i=2}^{n} \frac{1}{i}$.**
En confrontant l'expression obtenue à la question précédente avec celle obtenue au 3. de la première partie, on peut écrire :

$$n + n\sum_{r=1}^{n-1}(-1)^{r+1} C_n^r \frac{1}{r}\left(1-\frac{1}{n}\right)^n = n\sum_{i=1}^{n}\frac{1}{i} = n + n\sum_{i=2}^{n}\frac{1}{i}.$$

D'où le résultat annoncé.

7.2.2 Variables aléatoires à densité

Corrigé de l'exercice 7.1.4. On coupe un bâton de longueur 1 au hasard en trois morceaux : les abscisses U et V des découpes sont supposées indépendantes et uniformément réparties sur $[0,1]$. On veut calculer la probabilité p pour que l'on puisse faire un triangle avec les trois morceaux (on peut faire un triangle avec trois segments de longueur l_1, l_2 et l_3 si et seulement si $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$).

1. **Exprimer en fonction de U et V les longueurs respectives L_1, L_2 et L_3 du morceau le plus à gauche, du morceau du milieu et du morceau le plus à droite.**
Il est clair que $L_1 = \inf(U,V)$, $L_2 = \sup(U,V) - \inf(U,V)$ et $L_3 = 1 - \sup(U,V)$.

2. **Montrer que**

$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$

En remarquant que lorsque $l_1+l_2+l_3 = 1$, la condition $l_1 \leq l_2+l_3$, $l_2 \leq l_3+l_1$ et $l_3 \leq l_1+l_2$ est équivalente à $l_1 \leq \frac{1}{2}$, $l_2 \leq \frac{1}{2}$ et $l_3 \leq \frac{1}{2}$, on en déduit que p est égale à la probabilité de $\left\{L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right\}$. En décomposant cet événement sur la partition $\{U < V\}$, $\{V < U\}$ et $\{U = V\}$ et en utilisant l'égalité $\mathbb{P}(U = V) = 0$, on obtient

$$p = \mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right) + \mathbb{P}\left(V \leq U, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$

Comme les couples des variables (U,V) et (V,U) ont même loi, les deux probabilités à droite de l'égalité sont égales. Donc

$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$

3. **Calculer p.**

$$\begin{aligned} p &= 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right) \\ &= 2\mathbb{P}\left(U \leq V, U \leq \frac{1}{2}, V - U \leq \frac{1}{2}, V \geq \frac{1}{2}\right) \\ &= 2\int\int 1_{\{x\leq y, x\leq \frac{1}{2}, y-x\leq \frac{1}{2}, y\geq \frac{1}{2}\}} 1_{[0,1]}(x) 1_{[0,1]}(y) dxdy \\ &= 2\int_0^{\frac{1}{2}} \left(\int_{\frac{1}{2}}^{x+\frac{1}{2}} dy\right) dx = 2\int_0^{\frac{1}{2}} xdx = \frac{1}{4}. \end{aligned}$$

Corrigé de l'exercice 7.1.5. Soient X et Y deux variables aléatoires réelles indépendantes et identiquement distribuées suivant la loi de Pareto de paramètre 2 i.e. qui possèdent la densité $p(x) = \dfrac{1_{\{x>1\}}}{x^2}$. On pose

$$(Z, W) = \left(\ln(X), 1 + \frac{\ln(Y)}{\ln(X)}\right).$$

1. **Quelle est la loi de (Z, W) ? Les variables Z et W sont-elles indépendantes ?**
On se donne une fonction muette $f : \mathbb{R}^2 \to \mathbb{R}$ bornée. On a

$$\mathbb{E}(f(Z, W)) = \mathbb{E}(f(\ln(X), 1 + \ln(Y)/\ln(X))).$$

Comme X et Y sont indépendantes, (X, Y) a pour densité $p(x)p(y)$. Donc

$$\mathbb{E}(f(Z, W)) = \int_{]1,+\infty[^2} f(\ln(x), 1 + \ln(y)/\ln(x)) \frac{dx}{x^2}\frac{dy}{y^2}.$$

Pour $(x, y) \in]1, +\infty[^2$, on pose $\varphi(x, y) = (z, w) = (\ln(x), 1 + \ln(y)/\ln(x))$. Comme ln est une fonction injective et positive sur $]1, +\infty[$, φ est injective sur $]1, +\infty[^2$ et $\varphi(]1, +\infty[^2) \subset]0, +\infty[\times]1, +\infty[$. Soit $(z, w) \in]0, +\infty[\times]1, +\infty[$,

$$\begin{cases} z = \ln(x) \\ w = 1 + \ln(y)/\ln(x) \end{cases} \Leftrightarrow \begin{cases} \ln(x) = z \\ \ln(y) = z(w-1) \end{cases} \Leftrightarrow \begin{cases} x = e^z \\ y = e^{z(w-1)} \end{cases}.$$

Comme $(e^z, e^{z(w-1)}) \in]1, +\infty[^2$ on en déduit que $\varphi(]1, +\infty[^2) =]0, +\infty[\times]1, +\infty[$ et que $(x, y) = \varphi^{-1}(z, w) = (e^z, e^{z(w-1)})$. Comme

$$\frac{D(x, y)}{D(z, w)} = \begin{pmatrix} e^z & 0 \\ (w-1)e^{z(w-1)} & ze^{z(w-1)} \end{pmatrix},$$

en calculant la valeur absolue du déterminant de cette matrice, il vient $dxdy = ze^{zw}dzdw$. Donc en appliquant la formule de changement de variables, on obtient

$$\mathbb{E}(f(Z, W)) = \int_{]0,+\infty[\times]1,+\infty[} f(z, w) \frac{ze^{zw}}{e^{2z}e^{2z(w-1)}} dzdw.$$

Donc (Z, W) possède la densité $1_{\{z>0\}}1_{\{w>1\}}ze^{-zw}$. Comme cette densité ne peut pas s'écrire comme le produit d'une fonction de z et d'une fonction de w, les variables Z et W ne sont pas indépendantes.

2. **Quelle est la loi de W ?**
Par la formule des densités marginales, W a pour densité

$$\begin{aligned} \int_{\mathbb{R}} 1_{\{z>0\}}1_{\{w>1\}}ze^{-zw}dz &= 1_{\{w>1\}} \int_0^{+\infty} ze^{-zw}dz \\ &= 1_{\{w>1\}} \left[z \times \frac{-1}{w}e^{-zw}\right]_0^{+\infty} + \frac{1_{\{w>1\}}}{w} \int_0^{+\infty} e^{-zw}dz \\ &= 0 + \frac{1_{\{w>1\}}}{w} \left[\frac{-1}{w}e^{-zw}\right]_0^{+\infty} = \frac{1_{\{w>1\}}}{w^2}. \end{aligned}$$

Donc W suit la loi de Pareto de paramètre 2.

3. **Quelle est la loi de Z ?**
Toujours par la formule des densités marginales, Z a pour densité

$$\int_{\mathbb{R}} 1_{\{z>0\}} 1_{\{w>1\}} z e^{-zw} dw = 1_{\{z>0\}} \left[-e^{-zw}\right]_1^{+\infty} = 1_{\{z>0\}} e^{-z}.$$

Donc $Z \sim \mathcal{E}(1)$.

Corrigé de l'exercice 7.1.6. Soient X et Y deux variables aléatoires indépendantes suivant respectivement les lois $\Gamma(p,1)$ et $\Gamma(q,1)$, $p>0$ et $q>0$. On rappelle qu'une variable aléatoire de loi $\Gamma(a,1)$, $a>0$, a pour densité

$$f_a(x) = \frac{1}{\Gamma(a)} x^{a-1} \exp(-x) 1_{]0,\infty[}(x)$$

où $\Gamma(a) = \displaystyle\int_0^\infty x^{a-1}\exp(-x)dx$ vérifie $\forall a>0,\ \Gamma(a+1) = a\Gamma(a)$.

1. **Déterminer l'espérance et la variance de X.**

$$\mathbb{E}(X) = \frac{1}{\Gamma(p)} \int_0^\infty x^p \exp(-x)dx = \frac{\Gamma(p+1)}{\Gamma(p)} = p.$$

$$\mathbb{E}(X^2) = \frac{1}{\Gamma(p)} \int_0^\infty x^{p+1} \exp(-x)dx = \frac{\Gamma(p+2)}{\Gamma(p)} = p(p+1).$$

Donc $\mathrm{Var}\,(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p$.

2. On veut calculer la loi de la variable aléatoire $Z = \dfrac{X}{Y}$. On dit que Z suit la loi bêta de seconde espèce de paramètres p et q.

   (a) **On pose $W = X - Y$. Exprimer en fonction de X et Y l'événement $\{Z < 1, W > 0\}$. En déduire sans calcul sa probabilité. Les variables Z et W sont-elles indépendantes ?**
   Il est clair que $\{Z<1, W>0\} = \{X<Y, X>Y\} = \emptyset$. La probabilité de cet événement est nulle. Comme il est facile de vérifier que $\mathbb{P}(Z<1) = \mathbb{P}(X<Y)$ et $\mathbb{P}(W>0) = \mathbb{P}(Y<X)$ sont strictement positives, on en déduit que

$$\mathbb{P}(Z<1, W>0) \neq \mathbb{P}(Z<1)\mathbb{P}(W>0).$$

   Ceci montre que les deux variables ne sont pas indépendantes,

   (b) **Déterminer la loi du vecteur (Z,W).**
   Soient $U = \mathbb{R}_+^* \times \mathbb{R}_+^* \setminus \{(x,y) \text{ t.q. } x=y\}$ et $V =]0,1[\times]-\infty,0[\cup]1,+\infty[\times]0,+\infty[$. U et V sont deux ouverts de $\mathbb{R}^2$. On note par φ la l'application de U dans V définie par $\varphi(x,y) = (z,w) = \left(\dfrac{x}{y}, x-y\right)$. En résolvant le système d'équations

$$\left\{\begin{array}{lcl} z &=& \frac{x}{y} \\ w &=& x-y \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} x &=& yz \\ w &=& y(z-1) \end{array}\right. \Leftrightarrow \left\{\begin{array}{lcl} x &=& \frac{wz}{z-1} \\ y &=& \frac{w}{z-1} \end{array}\right.$$

   on déduit que φ est bijective et que $\varphi^{-1}(z,w) = (x,y) = \left(\dfrac{wz}{z-1}, \dfrac{w}{z-1}\right)$. On remarque que φ et φ^{-1} sont de classe C^1. La matrice jacobienne de φ au point (x,y) est donnée par

$$J_\varphi(x,y) = \begin{pmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ 1 & -1 \end{pmatrix},$$

le jacobien de φ au point (x,y), déterminant de la matrice jacobienne de φ au point (x,y), est égal à $\dfrac{x-y}{y^2}$. Maintenant

$$j_{\varphi^{-1}}(z,w) = \frac{1}{j_\varphi(\varphi^{-1}(z,w))} = \frac{(\frac{w}{z-1})^2}{\frac{wz}{z-1} - \frac{w}{z-1}} = \frac{w}{(z-1)^2}.$$

Pour appliquer la méthode de la fonction muette, on considère h une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ bornée,

$$\mathbb{E}\left(h(Z,W)\right) = \mathbb{E}\left(h\left(\frac{X}{Y}, X-Y\right)\right) = \int\int_{\mathbb{R}^2} h\left(\frac{x}{y}, x-y\right) f_{X,Y}(x,y)dxdy,$$

avec $f_{X,Y}$ la densité du couple (X,Y). Comme les variables X et Y sont indépendantes, on déduit que $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, avec f_X la densité de X et f_Y la densité de Y. Maintenant on écrit

$$\begin{aligned}
\mathbb{E}\left(h(Z,W)\right) &= \int\int_U h\circ\varphi(x,y) f_{X,Y}(x,y)dxdy \\
&= \int\int_V h(z,w) f_{X,Y}\circ\varphi^{-1}(z,w)\left|j_{\varphi^{-1}}(z,w)\right| dzdw \\
&= \int\int_V h(z,w) f_{X,Y}\left(\frac{wz}{z-1}, \frac{w}{z-1}\right)\frac{|w|}{(z-1)^2}dzdw \\
&= \frac{1}{\Gamma(p)\Gamma(q)}\int\int_V h(z,w)\left(\frac{wz}{z-1}\right)^{p-1}\left(\frac{w}{z-1}\right)^{q-1} e^{-w\frac{z+1}{z-1}}\frac{|w|}{(z-1)^2}dzdw \\
&= \frac{1}{\Gamma(p)\Gamma(q)}\int\int_V h(z,w)\frac{|w|^{p+q-1}\, z^{p-1}}{|z-1|^{p+q}} e^{-|w|\frac{z+1}{|z-1|}}dzdw.
\end{aligned}$$

Ainsi (Z,W) a pour densité

$$f_{Z,W}(z,w) = \frac{1}{\Gamma(p)\Gamma(q)}\frac{|w|^{p+q-1}\, z^{p-1}}{|z-1|^{p+q}} e^{-|w|\frac{z+1}{|z-1|}}\mathbf{1}_V(z,w).$$

(c) **En déduire la loi de la variable aléatoire Z.**
Par la formule des densités marginales, on obtient :

$$\begin{aligned}
f_Z(z) &= \int_{\mathbb{R}} f_{Z,W}(z,w)dw \\
&= \mathbf{1}_{]0,1[}(z)\frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{|z-1|^{p+q}}\int_{-\infty}^0 |w|^{p+q-1}\, e^{-|w|\frac{z+1}{|z-1|}}dw \\
&\quad + \mathbf{1}_{]1,+\infty[}(z)\frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{|z-1|^{p+q}}\int_0^{+\infty} |w|^{p+q-1}\, e^{-|w|\frac{z+1}{|z-1|}}dw.
\end{aligned}$$

Par la parité de la fonction valeur absolue, on en déduit que les deux intégrales sont égales, et que pour tout $z \in]0,1[\cup]1,+\infty[$,

$$\begin{aligned}
f_Z(z) &= \frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{|z-1|^{p+q}}\int_0^{+\infty} w^{p+q-1}e^{-w\frac{z+1}{|z-1|}}dw \quad \left(\text{on pose } a = w\frac{z+1}{|z-1|}\right) \\
&= \frac{1}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{(z+1)^{p+q}}\int_0^{+\infty} a^{p+q-1}e^{-a}da = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\frac{z^{p-1}}{(z+1)^{p+q}}.
\end{aligned}$$

Donc Z possède la densité $f_Z(z) = \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\dfrac{z^{p-1}}{(z+1)^{p+q}}1_{]0,+\infty[}(z)$.

3. **Déterminer la loi de $T = \dfrac{Z}{1+Z}$.**
On applique la méthode de la fonction muette, on considère h une fonction bornée de $\mathbb{R}$ dans $\mathbb{R}$.

$$\mathbb{E}\left(h(T)\right) = \mathbb{E}\left(h\left(\frac{Z}{1+Z}\right)\right) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^{+\infty} h\left(\frac{z}{1+z}\right)\frac{z^{p-1}}{(z+1)^{p+q}}dz.$$

On pose $t = \frac{z}{1+z}$, ce qui donne $z = \frac{t}{1-t} = -1 + \frac{1}{1-t}$ et $dz = \frac{dt}{(1-t)^2}$. Donc $\mathbb{E}\left(h(T)\right) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^1 h(t)t^{p-1}(1-t)^{q-1}dt$ et T suit la loi béta de paramètres p et q.

Corrigé du problème 7.1.7. Soient $\Theta \sim \mathcal{U}[-\pi,\pi]$ et R une variable aléatoire réelle indépendante de Θ possédant une densité p nulle sur $\mathbb{R}_-$. On pose

$$(X,Y) = (R\cos\Theta, R\sin\Theta).$$

1. **Déterminer la loi du vecteur (X,Y). Que peut-on dire des lois des vecteurs (X,Y) et (Y,X) ? En déduire que X et Y ont la même densité.**
Comme R et Θ sont indépendantes, le couple (R,Θ) a pour densité $\frac{1}{2\pi}p(r)1_{[-\pi,\pi]}(\theta)$ fonction qui est nulle pour $r \leq 0$. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ bornée.

$$\mathbb{E}(f(X,Y)) = \mathbb{E}(f(R\cos\Theta, R\sin\Theta)) = \int_{]0,+\infty[\times]-\pi,\pi[} f(r\cos\theta, r\sin\theta)\frac{p(r)}{2\pi}drd\theta.$$

A l'aide de l'exercice résolu 3.1.5, on vérifie que le changement de variables $(x,y) = \varphi(r,\theta) = (r\cos\theta, r\sin\theta)$ est une bijection de $]0,+\infty[\times]-\pi,\pi[$ sur $\mathbb{R}^2 \setminus \{\mathbb{R}_- \times \{0\}\}$ de classe C^1 ainsi que son inverse

$$(r,\theta) = \varphi^{-1}(x,y) = \left(\sqrt{x^2+y^2}, 2\arctan\left(\frac{x}{x+\sqrt{x^2+y^2}}\right)\right).$$

$$\text{Comme } \frac{D(x,y)}{D(r,\theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix},$$

$dxdy = r(\cos^2\theta + \sin^2\theta)drd\theta$, ce qui s'écrit aussi $drd\theta = \frac{dxdy}{\sqrt{x^2+y^2}}$. Donc par la formule de changement de variable,

$$\mathbb{E}(f(X,Y)) = \int_{\mathbb{R}^2\setminus\{\mathbb{R}_-\times\{0\}\}} f(x,y)\frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}dxdy.$$

Comme on ne change pas la valeur du membre de droite en rajoutant la demi-droite $\mathbb{R}_- \times \{0\}$ au domaine d'intégration, on conclut que (X,Y) a pour densité $p_{(X,Y)}(x,y) = p(\sqrt{x^2+y^2})/2\pi\sqrt{x^2+y^2}$.
Puisque cette densité est symétrique, pour f de $\mathbb{R}^2$ dans $\mathbb{R}$ bornée,

$$\mathbb{E}(f(Y,X)) = \int_{\mathbb{R}^2} f(y,x)p_{X,Y}(x,y)dxdy = \int_{\mathbb{R}^2} f(y,x)p_{X,Y}(y,x)dxdy.$$

Donc (Y,X) admet $p_{X,Y}$ pour densité et $(Y,X) \overset{\mathcal{L}}{=} (X,Y)$. Les images respectives X et Y de ces deux vecteurs par la projection $(z,w) \in \mathbb{R}^2 \to z \in \mathbb{R}$ ont également même loi. Comme d'après la formule des densités marginales X et Y ont une densité, leurs densités sont égales.

2. **Pour $a > 0$, quelle est la loi de $Z = aX/Y$ (on pourra exprimer Z en fonction de Θ) ? En déduire sans calcul la loi de l'inverse d'une variable de Cauchy de**

paramètre a.
Comme $Z = a\cot\Theta$, pour f de $\mathbb{R}$ dans $\mathbb{R}$ bornée,

$$\mathbb{E}(f(Z)) = \mathbb{E}(f(a\cot\Theta)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(a\cot\theta)d\theta = \frac{1}{\pi}\int_0^{\pi} f(a\cot\theta)d\theta,$$

en utilisant la périodicité de la fonction cotangente. Si on pose $z = a\cot\theta$, $\frac{dz}{d\theta} = -a(1+\cot^2\theta)$ ce qui implique que $d\theta = \frac{dz}{-a(1+z^2/a^2)}$. On en déduit que

$$\mathbb{E}(f(Z)) = \int_{+\infty}^{-\infty} f(z)\frac{dz}{-a\pi(1+z^2/a^2)} = \int_{\mathbb{R}} f(z)\frac{adz}{\pi(a^2+z^2)}.$$

Donc la variable $Z = aX/Y$ suit la loi de Cauchy de paramètre a. Comme $(X,Y) \overset{\mathcal{L}}{=} (Y,X)$, $Y/aX \overset{\mathcal{L}}{=} X/aY$. Donc $1/Z = Y/aX \overset{\mathcal{L}}{=} \frac{1}{a}X/Y$. D'après ce qui précède, $\frac{1}{a}X/Y \sim \mathcal{C}(1/a)$. Donc la loi de l'inverse d'une variable de Cauchy de paramètre a est la loi de Cauchy de paramètre $1/a$.

3. **Quelle doit être la loi de R pour que le couple (X,Y) soit uniformément distribué sur le disque $D(0,\rho)$ où $\rho > 0$ (i.e. possède une densité constante sur ce disque et nulle en dehors) ?**
Pour que (X,Y) soit uniformément réparti sur le disque $D(0,\rho)$ i.e. possède la densité $1_{\{\sqrt{x^2+y^2}<\rho\}}/\pi\rho^2$, il faut et il suffit que

$$\frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}} = 1_{\{\sqrt{x^2+y^2}<\rho\}}/\pi\rho^2$$

c'est-à-dire que $p(r) = 1_{]0,\rho[}(r)2r/\rho^2$.

4. **On suppose la densité p de R continue sur $\mathbb{R}_+$ et strictement positive sur $\mathbb{R}_+^*$. On souhaite trouver à quelle condition sur p les variables X et Y sont indépendantes.**
   (a) **Dans le cas où X et Y sont indépendantes, si q désigne leur densité commune, exprimer $q(x)q(y)$ en fonction de p.**
   En commençant par choisir $y = 0$ dans cette relation, établir que pour $x, y \in \mathbb{R}^*$,

$$\frac{p(|x|)}{2\pi q(0)|x|} \times \frac{p(|y|)}{2\pi q(0)|y|} = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}. \tag{7.2}$$

   Si X et Y sont indépendantes, $p_{X,Y}(x,y) = q(x)q(y)$, ce qui s'écrit aussi

$$q(x)q(y) = \frac{p(\sqrt{x^2+y^2})}{2\pi\sqrt{x^2+y^2}}.$$

   Le choix $y = 0$ implique alors que $q(x) = p(|x|)/2\pi q(0)|x|$ et on obtient (7.2) en reportant cette expression dans l'égalité précédente.

   (b) **Pour $s > 0$, on pose $f(s) = \ln\left(\frac{p(\sqrt{s})}{2\pi q^2(0)\sqrt{s}}\right)$. Vérifier que f est une fonction continue sur $\mathbb{R}_+^*$ telle que pour $s, t > 0$, $f(s) + f(t) = f(s+t)$.**
   Pour $s, t > 0$, en posant $x = \sqrt{s}$, $y = \sqrt{t}$ et en divisant (7.2) par $q^2(0)$, on obtient

$$\frac{p(\sqrt{s})}{2\pi q^2(0)\sqrt{s}} \times \frac{p(\sqrt{t})}{2\pi q(0)^2\sqrt{t}} = \frac{p(\sqrt{t+s})}{2\pi q^2(0)\sqrt{t+s}}.$$

   Par passage au logarithme, il vient $f(s) + f(t) = f(s+t)$, la continuité de f se déduisant de celle de p.

En déduire qu'il existe une constante $\sigma > 0$ telle que $p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} 1_{\{r>0\}}$ i.e. que R suit la loi de Rayleigh de paramètre σ^2.
Par récurrence, on en déduit que pour $k \in \mathbb{N}^*$ et $s_1, \ldots, s_k > 0$, $\sum_{j=1}^k f(s_j) = f(\sum_{j=1}^k s_j)$. Soit $n \in \mathbb{N}^*$. Pour le choix $s_j = \frac{1}{n}$ pour tout $j \geq 1$, on en déduit que pour tout $k \in \mathbb{N}^*$, $kf(1/n) = f(k/n)$. Comme en particulier pour $k = n$, cette relation s'écrit $f(1/n) = f(1)/n$, on en déduit que pour tout $k, n \in \mathbb{N}^*$, $f(k/n) = kf(1)/n$. La densité des rationnels dans $\mathbb{R}_+^*$ et la continuité de f impliquent que pour tout $s > 0$, $f(s) = f(1)s$.
En utilisant la définition de f, on obtient qu'il existe deux constantes c et a telles que $p(r) = cre^{ar^2} 1_{\{r>0\}}$. La condition de sommabilité de p impose que $a < 0$ i.e. qu'il existe une constante $\sigma > 0$ telle que $a = -1/2\sigma^2$. Puis la condition de normalisation de p impose que $c = 1/\sigma^2$. Ainsi R suit la loi de Rayleigh de paramètre σ^2.
Quelle est alors la loi de (X, Y) ? On a

$$p_{X,Y}(x, y) = \frac{p(\sqrt{x^2 + y^2})}{2\pi\sqrt{x^2 + y^2}} = \frac{e^{-x^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \times \frac{e^{-y^2/2\sigma^2}}{\sigma\sqrt{2\pi}},$$

c'est-à-dire que X et Y sont indépendantes et suivent la loi normale $\mathcal{N}_1(0, \sigma^2)$.

7.2.3 Convergence des variables aléatoires

Corrigé de l'exercice 7.1.8. On considère une machine qui tourne à une cadence de n cycles par unité de temps et qui a une probabilité $p \in]0, 1[$ de tomber en panne lors de chaque cycle et ce indépendamment des autres cycles. Soit N le numéro du cycle où se produit la première panne.

1. **Quelle est la loi de N ? Que représente $T = N/n$?**
   N suit la loi géomètrique de paramètre p. La variable T représente le temps avant la première panne.
   Soit τ l'espérance de T. Exprimer τ en fonction de p et n.
   $\tau = \mathbb{E}(N)/n = 1/pn$.
2. Dans le cas où n est grand, on souhaite approcher la variable discrète T par une variable aléatoire à densité. Pour cela on va effectuer le passage à la limite $n \to +\infty$ en choisissant $p(n)$ de façon à ce que l'espérance de $T(n)$ reste égale à τ.
   (a) **Que vaut $p(n)$?**
   $p(n) = 1/n\tau$.
   (b) **Calculer la fonction caractéristique de $T(n)$.**

$$\Phi_{T(n)}(u) = \sum_{k \geq 1} p(n)(1 - p(n))^{k-1} \exp(iuk/n) = \frac{p(n)\exp(iu/n)}{1 - (1 - p(n))\exp(iu/n)}.$$

   En remplaçant $p(n)$ par sa valeur $1/n\tau$ on obtient

$$\Phi_{T(n)}(u) = \frac{1}{1 + n\tau(\exp(-iu/n) - 1)}.$$

   (c) **En déduire que cette variable converge en loi vers une limite que l'on précisera.**
   Comme $\lim_{n\to+\infty} n(\exp(-iu/n) - 1) = -iu$, on a

$$\forall u \in \mathbb{R}, \lim_{n\to+\infty} \Phi_{T(n)}(u) = \frac{1}{1 - iu\tau} = \frac{1/\tau}{1/\tau - iu}.$$

   On reconnaît la fonction caractéristique de la loi exponentielle de paramètre $1/\tau$. Donc les variables $T(n)$ convergent en loi vers la loi exponentielle de paramètre $1/\tau$.

Corrigé de l'exercice 7.1.9. Chaque jour, dans une certaine ville, 100 personnes ont besoin d'un examen radioscopique. Pour préserver le libre choix, n centres d'imagerie sont installés dans cette ville. On admet que les patients choisissent indifféremment l'un ou l'autre centre d'imagerie.
Soit N le nombre de clients journaliers dans un centre d'imagerie .

1. **Quelle est la probabilité qu'un client choisisse le centre d'imagerie considéré ?**
   Comme on suppose que les patients choisissent indifféremment l'un ou l'autre des n centres d'imagerie, la probabilité qu'un client choisisse le centre d'imagerie considéré est $\frac{1}{n}$.
2. **Montrer que N peut s'écrire $N = \sum_{i=1}^{i=100} X_i$ où les $(X_i)_{1\leq i\leq n}$ sont n variables indépendantes distribuées suivant la loi de Bernoulli de même paramètre p que l'on déterminera.**
   Soit X_i la variable aléatoire discrète réelle qui vaut 1 si le client i va au centre d'imagerie étudié et 0 sinon. Alors X_i suit la loi de Bernoulli de paramètre p tel que
   $$p = \mathbb{P}(X_i = 1) = \frac{1}{n}.$$
3. **Quelle est la loi de N ?**
   N suit la loi binomiale de paramètres 100 et $\frac{1}{n}$. Son espérance est donc : $\mathbb{E}(N) = \frac{100}{n}$, et sa variance : $V(N) = 100\frac{1}{n}(1 - \frac{1}{n})$.
4. **On donne que si Y suit la loi normale centrée réduite, alors $\mathbb{P}(Y \leq 2) = 98\%$. En utilisant le théorème de la limite centrale, déterminer quelle capacité $c(n)$ chaque centre d'imagerie doit avoir pour être capable de répondre à la demande avec une probabilité de 98% ? Cas où $n = 2$, $n = 3$, $n = 4$?**
   Le Théorème Central Limite dit que
   $$X = \frac{N - \mathbb{E}(N)}{\sqrt{V(N)}}$$
   peut être considérée comme une variable normale centrée réduite.
   Or
   $$\{N \leq c\} = \left\{X \leq \frac{c - \mathbb{E}(N)}{\sqrt{V(N)}}\right\}.$$
   et les tables de la loi normale centrée réduite donnent que $\mathbb{P}(X \leq 2) = 98\%$. Donc pour pouvoir faire face à demande avec une probalité de 98%, le centre doit avoir la capacité $c(n)$ donnée par
   $$\frac{c(n) - \mathbb{E}(N)}{\sqrt{V(N)}} = 2 \text{ i.e. } c(n) = \mathbb{E}(N) + 2\sqrt{V(N)} = \frac{100}{n} + 20\sqrt{\frac{1}{n}\left(1 - \frac{1}{n}\right)}.$$
   On a les résultats suivants :
   $$c(1) = 100 \text{ (heureusement !)}, \quad c(2) = 60, \quad c(3) = 42,8, \quad c(4) = 33,7.$$
5. **Quel est le coût de la concurrence : quelle surcapacité $s(n)$ la concurrence entraîne-t-elle par rapport à une situation où chaque centre se verrait affecter un même nombre de clients ? Cas où $n = 2$, $n = 3$, $n = 4$?**
   La surcapacité est $s(n) = n \times c(n) - 100 = 20 \times \sqrt{n-1}$ soit
   $$s(1) = 0, \quad s(2) = 20, \quad s(3) = 28,3, \quad s(4) = 34,6.$$

Corrigé de l'exercice 7.1.10.

1. **Calculer la fonction caractéristique d'une variable X suivant la loi de Poisson $\mathcal{P}(\lambda)$. Calculer $\mathbb{E}(X)$ et Var (X).**
Soit $u \in \mathbb{R}$. Par définition, la fonction caractéristique ϕ_X de X vaut :
$$\phi_X(u) = \mathbb{E}\left[e^{iuX}\right] = \sum_{k=0}^{\infty} e^{iuk}\mathbb{P}\{X = k\} = \sum_{k=0}^{\infty} e^{iuk}e^{-\lambda}\frac{\lambda^k}{k!} = \exp(\lambda(e^{iu} - 1)).$$
Développons cette expression au voisinage de $u = 0$:
$$e^{iu} - 1 = iu - \frac{1}{2}u^2 + o(u^2).$$
Par conséquent,
$$\begin{aligned}\phi_X(u) = \exp\{\lambda(e^{iu} - 1)\} &= \exp\{iu\lambda - \frac{\lambda}{2}u^2 + o(u^2)\}\\ &= 1 + \left(iu\lambda - \frac{\lambda}{2}u^2\right) - \frac{\lambda^2}{2}u^2 + o(u^2)\\ &= 1 + iu\lambda - \frac{\lambda + \lambda^2}{2}u^2 + o(u^2).\end{aligned}$$
En identifiant avec le développement limité en $u = 0$ de ϕ_X ($\phi_X(u) = 1 + i\mathbb{E}(X)u - \frac{\mathbb{E}(X^2)}{2}u^2 + o(u^2)$), il vient : $\mathbb{E}(X) = \lambda$ et $\mathbb{E}(X^2) = \lambda + \lambda^2$ donc $\mathrm{Var}(X) = \lambda + \lambda^2 - \lambda^2 = \lambda$.
2. **Soit $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ deux variables de Poisson indépendantes. Quelle est la loi de leur somme ?**
Les deux variables étant indépendantes, la fonction caractéristique de leur somme est le produit de leurs fonctions caractéristiques. Ainsi, pour tout $u \in \mathbb{R}$, on a :
$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u) = \exp\{\lambda(e^{iu} - 1)\}\exp\{\mu(e^{iu} - 1)\} = \exp\{(\lambda + \mu)(e^{iu} - 1)\}.$$
On reconnaît la fonction caractéristique d'une variable de Poisson de paramètre $\lambda + \mu$. Comme la fonction caractéristique caractérise la loi, on conclut que $X + Y \sim \mathcal{P}(\lambda + \mu)$.
3. **Soit $X_n \sim \mathcal{P}(n)$. Montrer que pour $n \to +\infty$, $X_n^* = (X_n - n)/\sqrt{n}$ converge en loi vers une limite que l'on précisera. Commenter ce résultat à la lumière du théorème de la limite centrale.**
Soit $u \in \mathbb{R}$.
$$\begin{aligned}\phi_{X_n^*}(u) &= \exp\left(-iu\sqrt{n}\right)\phi_{X_n}\left(\frac{u}{\sqrt{n}}\right)\\ &= \exp\left(-iu\sqrt{n}\right)\exp\{n(e^{i\frac{u}{\sqrt{n}}} - 1)\}\\ &= \exp\left(-iu\sqrt{n}\right)\exp\left[\left(i\frac{u}{\sqrt{n}} - \frac{u^2}{2n} + o\left(\frac{1}{n}\right)\right)\right] = \exp\left(-\frac{u^2}{2} + o(1)\right).\end{aligned}$$
Ainsi, $\forall u \in \mathbb{R}$, $\lim_{n\to\infty}\phi_{X_n^*}(u) = \exp\{-\frac{u^2}{2}\} = \phi_G(u)$ où $G \sim \mathcal{N}(0, 1)$. Par conséquent, X_n^* converge en loi vers G qui suit la loi normale centrée réduite.
Pour expliquer ce résultat, il suffit de considérer une suite $(Y_n)_{n\geq 1}$ de variables i.i.d. de loi $\mathcal{P}(1)$. Il résulte de la deuxième question que $Y_1 + \ldots + Y_n$ suit la loi $\mathcal{P}(n)$ et par conséquent X_n^* a même loi que $Z_n = \sqrt{n}(\frac{Y_1+\cdots+Y_n}{n} - 1)$. Comme $\mathbb{E}(Y_1) = 1$ et Var $(Y_1) = 1$, le théorème de la limite centrale implique que la suite $(Z_n)_{n\geq 1}$ converge en loi vers G.

Corrigé du problème 7.1.10. Un automobiliste emprunte tous les jours le même trajet qui comporte un feu tricolore pour se rendre à son travail. Comme le trajet est peu encombré, lorsque le feu est rouge, l'automobiliste peut redémarrer dès que le feu passe au vert. Mais pour faire passer le temps, il se demande quelle est la durée θ pendant laquelle le feu reste rouge. On note $(X_i)_{i\geq 1}$ ses durées d'attente successives au feu lorsque celui-ci est rouge et on suppose ces variables i.i.d. suivant la loi uniforme $\mathcal{U}[0, \theta]$. Pour $n \geq 1$, on note $Z_n = \max_{1\leq i\leq n} X_i$.

1. **Montrer que $\mathbb{P}(X_1 = X_2) = 0$.**
Les variables X_1 et X_2 étant indépendantes et à densité, le couple (X_1, X_2) admet une densité qui est le produit des densités de X_1 et de X_2. Ainsi,

$$\mathbb{P}(X_1 = X_2) = \frac{1}{\theta^2} \int_{[0,\theta]^2} 1_{\{x_1 = x_2\}} dx_1 dx_2 = 0$$

car la droite d'équation $x_1 = x_2$ est de mesure nulle dans le plan.

2. **Montrer que plus généralement $\mathbb{P}(\exists 1 \leq i \neq j \leq n, X_i = X_j) = 0$. En déduire que pour toute fonction f bornée sur $\mathbb{R}$, on a :**

$$\mathbb{E}[f(Z_n)] = \sum_{i=1}^{n} \mathbb{E}\left[f(X_i) 1_{\{\forall j \neq i,\, X_j < X_i\}}\right] = n\mathbb{E}\left[f(X_1) 1_{\{X_2 < X_1, \ldots, X_n < X_1\}}\right].$$

Il suffit de remarquer que comme pour $i \neq j$, $\mathbb{P}(X_i = X_j) = \mathbb{P}(X_1 = X_2) = 0$,

$$\mathbb{P}(\exists 1 \leq i \neq j \leq n, X_i = X_j) = \mathbb{P}\left(\cup_{1 \leq i \neq j \leq n} \{X_i = X_j\}\right) \leq \sum_{1 \leq i \neq j \leq n} \mathbb{P}(X_i = X_j) = 0.$$

Par suite, comme $\{\exists 1 \leq i \neq j \leq n, X_i = X_j\}^c = \{\forall 1 \leq i \neq j \leq n, X_i \neq X_j\}$ est un événement de probabilité 1, on peut écrire :

$$\mathbb{E}[f(Z_n)] = \mathbb{E}[f(Z_n) 1_{\{\forall 1 \leq i \neq j \leq n, X_i \neq X_j\}}] = \sum_{k=1}^{n} \mathbb{E}[f(Z_n) 1_{\{\forall 1 \leq i \neq j \leq n, X_i \neq X_j\}} 1_{\{Z_n = X_k\}}]$$

où on a décomposé suivant l'indice k tel que X_k est le maximum des valeurs distinctes $X_1, \ldots, X_n$. Examinons l'un des termes de cette somme :

$$\mathbb{E}[f(Z_n) 1_{\{\forall 1 \leq i \neq j \leq n, X_i \neq X_j\}} 1_{\{Z_n = X_k\}}] = \mathbb{E}[f(X_k) 1_{\{\forall i \neq k, X_i < X_k\}}]$$

car l'événement $\{\exists i \neq j \in \{1, \ldots, k-1, k+1, \ldots, n\} : X_i = X_j\}$ que l'on rajoute dans l'indicatrice est de probabilité nulle. Il ne reste plus qu'à remarquer que l'expression ne dépend pas de l'indice k choisi car pour toute permutation σ de $\{1, \ldots, n\}$, le vecteur $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ est constitué de n variables indépendantes de loi uniforme sur $[0, \theta]$ et a même loi que le vecteur $(X_1, \ldots, X_n)$. Ainsi,

$$\mathbb{E}[f(X_k) 1_{\{\forall i \neq k, X_i < X_k\}}] = \mathbb{E}[f(X_1) 1_{\{\forall i \neq 1, X_i < X_1\}}].$$

Le résultat en découle immédiatement.

3. **En déduire que Z_n suit la densité $\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1} 1_{\{0 \leq z \leq \theta\}}$.**
Les variables $X_1, \ldots, X_n$ étant indépendantes et à densité, le n-uplet $(X_1, \ldots, X_n)$ admet une densité, produit des densités marginales. Ainsi,

$$\begin{aligned}
\mathbb{E}[f(X_1) 1_{\{\forall i \neq 1, X_i < X_1\}}] &= \int_{[0,\theta]^n} 1_{\{\forall i \neq 1, x_i < x_1\}} \frac{dx_1 \ldots dx_n}{\theta^n} \\
&= \int_{x_1=0}^{\theta} f(x_1) \left[\int_{x_2=0}^{x_1} \frac{dx_2}{\theta}\right] \cdots \left[\int_{x_n=0}^{x_1} \frac{dx_n}{\theta}\right] \frac{dx_1}{\theta} \\
&= \int_0^{\theta} f(x_1) \left(\frac{x_1}{\theta}\right)^{n-1} \frac{dx_1}{\theta}
\end{aligned}$$

Il en résulte que $\mathbb{E}[f(Z_n)] = n \int_0^{\theta} f(x_1) \left(\frac{x_1}{\theta}\right)^{n-1} \frac{dx_1}{\theta}$. D'où le résultat.

4. **Calculer $\mathbb{E}(Z_n)$ et $\mathrm{Var}\,(Z_n)$. En conclure que Z_n converge presque sûrement vers θ (*indication* : on rappelle (voir preuve du théorème 5.2.2) que si $(\tilde{Z}_n)_{n\geq 1}$ est une suite de variables aléatoires telle que $\sum_{n\geq 1}\mathbb{E}(\tilde{Z}_n^2) < +\infty$, alors la suite $(\tilde{Z}_n)_{n\geq 1}$ converge presque sûrement vers 0).**
On a :
$$\mathbb{E}(Z_n) = \int_0^\theta z\frac{n}{\theta^n}z^{n-1}\,dz = \frac{n}{\theta^n}\times\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta.$$
De même,
$$\mathbb{E}(Z_n^2) = \frac{n}{n+2}\theta^2$$
puis
$$\mathrm{Var}\,(Z_n) = \mathbb{E}(Z_n^2) - \mathbb{E}(Z_n)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2.$$
Comme $\mathrm{Var}\,(Z_n) = O(1/n^2)$, la série de terme général $\mathbb{E}[(Z_n - \mathbb{E}(Z_n))^2]$ est convergente et donc $(\tilde{Z}_n = Z_n - \mathbb{E}(Z_n))_{n\geq 1}$ converge presque sûrement vers 0. Or la suite (déterministe) $(\mathbb{E}(Z_n))_{n\geq 1}$ converge vers θ. Par conséquent, $(Z_n)_{n\geq 1}$ converge presque sûrement vers θ.

5. **Soit f une fonction bornée sur $\mathbb{R}$. Montrer que :**
$$\mathbb{E}\left[f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right] = \int_0^n f(u)\left(1-\frac{u}{n}\right)^{n-1}\,du.$$
On rappelle que pour $0 < v < 1$, $\ln(1-v) \leq -v$. En déduire que pour $n \geq 2$, $\left(1-\frac{u}{n}\right)^{n-1}1_{\{0<u\leq n\}} \leq e^{-u/2}1_{\{u>0\}}$ et conclure que :
$$\lim_{n\to\infty}\mathbb{E}\left[f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right] = \int_0^\infty f(u)e^{-u}\,du$$
En utilisant la densité de Z_n, il est clair que :
$$\mathbb{E}\left[f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right] = \int_0^\theta f\left(\frac{n}{\theta}(\theta - z)\right)\frac{n}{\theta}\left(\frac{z}{\theta}\right)^{n-1}\,dz.$$
Le changement de variable $u = \frac{n}{\theta}(\theta - z)$ donne immédiatement le résultat souhaité.
Soit $u \in]0, n[$. En appliquant l'inégalité donnée avec $v = u/n$, il vient
$$\left(1-\frac{u}{n}\right)^{n-1} = \exp\left((n-1)\ln\left(1-\frac{u}{n}\right)\right) \leq \exp\left(-\frac{n-1}{n}u\right).$$
Or, dès lors que $n \geq 2$, on a $(n-1)/n \geq 1/2$. D'où pour $u \in]0, n[$,
$$\left(1-\frac{u}{n}\right)^{n-1}1_{\{0<u\leq n\}} \leq e^{-u/2}1_{\{u>0\}}.$$
Pour $u \geq n$, le membre de gauche de l'inégalité vaut 0 tandis que celui de droite est strictement positif. Elle est donc encore vérifiée.
Ainsi
$$\forall u \in \mathbb{R},\ |f(u)|\left(1-\frac{u}{n}\right)^{n-1}1_{\{0<u\leq n\}} \leq \sup_x |f(x)|e^{-u/2}1_{\{u>0\}},\ \text{avec}\ \int_0^\infty e^{-u/2}1_{\{u>0\}} < \infty$$
et
$$\forall u \in \mathbb{R},\ \lim_{n\to\infty} f(u)\left(1-\frac{u}{n}\right)^{n-1}1_{\{0<u\leq n\}} = f(u)e^{-u}1_{\{u>0\}}$$
Le résultat annoncé découle du théorème de convergence dominée.

6. **En déduire que $\frac{n}{\theta}(\theta - Z_n)$ converge en loi et identifier la limite.**
Ainsi pour toute fonction f bornée, si X suit la loi exponentielle de paramètre 1,

$$\lim_{n\to\infty} \mathbb{E}\left[f\left(\frac{n}{\theta}(\theta - Z_n)\right)\right] = \mathbb{E}[f(X)].$$

Ce résultat est a fortiori vrai pour les fonctions f continues bornées, ce qui signifie par définition que $\frac{n}{\theta}(\theta - Z_n)$ converge en loi vers $X \sim \mathcal{E}(1)$.

7. **Montrer que $M_n = \frac{2}{n}(X_1 + \cdots + X_n)$ converge presque sûrement vers θ. À quelle vitesse à lieu cette convergence ?**
Par application de la loi forte des grands nombres aux variables X_i i.i.d. et intégrables (car bornées) :

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{p.s.}_{n\to\infty} \mathbb{E}(X_1) = \int_0^\theta x\frac{dx}{\theta} = \frac{\theta}{2}$$

Ainsi M_n converge presque sûrement vers θ. Pour la vitesse de convergence, on utilise le théorème de la limite centrale. Les variables X_i étant de carré intégrable et de variance commune $\theta^2/12$,

$$\frac{\sqrt{12n}}{\theta}\left(\frac{X_1 + \cdots + X_n}{n} - \frac{\theta}{2}\right) \xrightarrow{\mathcal{L}}_{n\to\infty} G \sim \mathcal{N}_1(0,1).$$

Comme la fonction $\varphi(x) = \theta x/\sqrt{3}$ est continue, on en déduit que

$$\sqrt{n}(M_n - \theta) = \varphi\left(\frac{\sqrt{12n}}{\theta}\left(\frac{X_1 + \cdots + X_n}{n} - \frac{\theta}{2}\right)\right) \xrightarrow{\mathcal{L}}_{n\to\infty} \varphi(G) = \frac{\theta G}{\sqrt{3}} \sim \mathcal{N}_1(0, \theta^2/3).$$

La convergence de M_n vers θ a donc lieu à la vitesse $1/\sqrt{n}$.

8. **L'automobiliste doit-il préférer Z_n ou M_n pour estimer rapidement θ ?**
La convergence de Z_n étant en $1/n$ tandis que celle de M_n est en $1/\sqrt{n}$, l'utilisation de Z_n est préférable.

Bibliographie

[1] N. Bouleau. *Probabilités de l'ingénieur. Variables aléatoires et simulation.* Hermann, 1986.

[2] L. Breiman. *Probability.* Number 7 in Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992 edition, 1968.

[3] P. Brémaud. *Introduction aux probabilités.* Springer Verlag, 1984.

[4] D. Dacunha-Castelle and M. Duflo. *Probabilités et statistiques : Problèmes à temps fixe.* Masson, 1982.

[5] W. Feller. *An introduction to probability theory and its applications*, volume 1. Wiley and Sons, Inc., 3rd edition, 1968.

[6] G. Grimmett and D. Stirzaker. *Probability and random processes.* Oxford University Press, 2nd edition, 1992.

[7] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson, 1964.

[8] J. Neveu. *Introduction aux probabilités.* École Polytechnique, 1988.

[9] S. Port. *Theoretical probability for applications.* Wiley & Sons, 1993.

[10] D. Revuz. *Mesure et Intégration.* Hermann, 1994.

[11] Y. Sinai. *Probability theory : an introductory course.* Springer Verlag, 1992.

[12] D. Stirzaker. *Elementary probability.* Cambridge Univ. Press, 1994.

Index

Changement de variables, 28, 34
Convergence
  dans L^1, 53, 54
  dans L^2, 53, 54
  en loi, 59, 60, 70
  en probabilité, 53, 54, 60
  presque sûre, 53, 54
Covariance, 35, 67, 69, 70

Densité marginale, 33

Ecart type, 21
Espérance
  d'une variable à densité, 31
  d'une variable discrète, 17

Fonction
  caractéristique, 57, 58, 60
  caractéristique des lois usuelles, 57
  de répartition, 32, 45
  indicatrice, 3, 13, 17
  muette, 33
Formule du crible, 8, 25

Inégalité
  de Bienaymé-Tchebychev, 54
  de Cauchy-Schwarz, 54
  de Markov, 54
Indépendance
  cas gaussien, 69
  d'événements, 7
  de variables à densité, 35
  des variables discrètes, 13, 20, 22
Intervalle de confiance, 62

Loi
  béta $\beta(a,b)$, 37
  binômiale $\mathcal{B}(n,p)$, 15, 44
  conditionnelle, 22
  de Bernoulli $\mathcal{B}(p)$, 14, 43
  de Cauchy $\mathcal{C}(a)$, 31, 45
  de Poisson $\mathcal{P}(\lambda)$, 15, 49
  discrète, 12
  du khi 2 $\chi^2(n)$, 39
  exponentielle $\mathcal{E}(\lambda)$, 31, 45
  géométrique $\mathcal{G}eo(p)$, 15, 44, 49
  gamma $\Gamma(a,\theta)$, 37
  gaussienne $\mathcal{N}_1(m,\sigma^2)$, 31
  gaussienne $\mathcal{N}_1(m,\sigma^2)$, 46, 67
  gaussienne $\mathcal{N}_d(m,\Gamma)$, 67
  marginale, 16
  normale $\mathcal{N}_1(m,\sigma^2)$, 31, 46, 67
  normale centrée réduite, 31, 46, 61
  uniforme $\mathcal{U}[a,b]$, 31, 45
Loi forte des grands nombres, 56

Méthode
  d'inversion, 45
  de Monte-Carlo, 62
  du rejet, 47

Probabilité
  conditionnelle, 5
  définition, 11
  sur un espace fini, 1
  uniforme, 4

Théorème
  de convergence dominée, 53
  de Fubini, 27
  de la limite centrale, 61
Tribu, 11

Variable
  à densité, 30
  discrète, 12
Variance
  d'une somme, 21, 36
  d'une variable discrète, 20
  d'une variable à densité, 32
Vecteur
  à densité, 33
  gaussien, 67