SY 01

S. Bouzebda

ÉLEMENTS DE PROBABILITÉS

SY01C2P19

université de technologie de compiègne

SY 01

S. Bouzebda

ÉLEMENTS DE PROBABILITÉS

SY01C2P19

université de technologie de compiègne

UNIVERSITE DE TECHNOLOGIE DE COMPIEGNE

SY01
ELEMENTS DE PROBABILITES

Responsable de SY01 : Salim Bouzebda
Rédacteur de ce cours : Nikolaos Limnios
LMAC

2019

Table des matières

Chapitre 1

Expérience aléatoire et probabilité

1.1 Expérience aléatoire

Lors d'une expérience répétée sous des conditions identiques, nous observons des variations sur les résultats. C'est à dire que des conditions identiques ne permettent pas d'obtenir une seule valeur mais un ensemble des valeurs possibles.

♣ **Exemples 1.1.**

1. Le jeu de pile ou face, le jet de dés, les cartes, la loterie, etc., appelés jeux de "hasard".
2. L'observation du nombre d'appels passant par un central téléphonique.
3. L'observation de la croissance d'une population de bactéries.
4. L'observation de la désintégration des atomes d'une substance radioactive.
5. L'observation de la durée de fonctionnement sans panne d'un appareil.
6. L'observation dans l'intervalle de temps [s,t] de l'accélération d'un véhicule en mouvement.

1.2 Espace fondamental

Une expérience aléatoire se décrit mathématiquement par la donnée de l'ensemble des issues ou des résultats possibles de cette expérience.

Soit Ω l'espace des résultats possibles d'une expérience, appelé aussi *espace des issues* ou *espace des épreuves* ou *espace des éventualités* ou encore *espace fondamental*. Un élément ω de Ω est appelé *un résultat* ou *une issue* ou *une épreuve* ou *une éventualité*.

♣ **Exemples 1.2.**

1. L'espace fondamental du jeu de pile ou face est : $\Omega = \{P, F\}$
2. L'espace fondamental d'un jeu de pile ou face répété 3 fois est :
   $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$
   avec :
   $$\omega_1 = (P,P,P), \quad \omega_2 = (P,P,F), \quad \omega_3 = (P,F,P), \quad \omega_4 = (F,P,P)$$
   $$\omega_5 = (P,F,F), \quad \omega_6 = (F,P,F), \quad \omega_7 = (F,F,P), \quad \omega_8 = (F,F,F)$$
3. L'espace fondamental d'un jeu de pile ou face répété n fois est :
   $\Omega = \{\omega_1, ..., \omega_{2^n}\}$ avec $\omega_i = (P, F, P, ..., P)$ tous les n-uples possibles.
4. L'espace fondamental d'un jeu infini de pile ou face est : $\Omega = \{P, F\}^{\mathbb{N}}$
5. L'espace fondamental dans l'expérience aléatoire de durée de vie d'un appareil est : $\Omega = \mathbb{N}$ ou bien $\Omega = \mathbb{R}_+$ (selon l'axe des temps).
6. L'espace fondamental qui consiste à observer la vitesse d'un véhicule dans [s,t] est l'ensemble des fonctions réelles continues sur $]s, t[$, noté $C(]s, t[)$, $(s < t)$.

1.3 Evénement aléatoire

Un événement aléatoire est un événement lié à une expérience aléatoire. Nous considérons que la réalisation ou la non réalisation d'un événement dépend exclusivement du résultat de l'expérience aléatoire à laquelle il est lié.

Mathématiquement, un événement aléatoire est représenté par un sous ensemble de Ω contenant tous les résultats ω qui le réalisent.

♣ **Exemples 1.3.**

1. Dans l'exemple 2 du paragraphe précédent, obtenir au moins 2 fois pile est un événement, noté A et décrit par l'ensemble :
   $A = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
   Dans la même expérience, un autre événement, noté B, consistant à obtenir 3 fois face, est :
   $B = \{\omega_8\}$.
   Un événement comme B contenant un seul élément de Ω est dit *élémentaire*.
2. Dans l'exemple 6 du paragraphe précédent, l'événement "la vitesse ne dépasse pas la valeur v_o" se représente dans l'espace $\Omega = C[s,t]$ par la boule : $\{\omega : \sup_{u\in[s,t]} \omega(t) \leq v_o\}$

1.4 Opérations sur les événements

- **Evénement contraire.** A tout événement A est associé son *contraire*, noté A^c (ou $\overline{A}$ ou $C_E A$ ou $\neg A$ ou "non A"). Il est réalisé lorsque A n'est pas réalisé. Dans l'exemple 1 précédent, $A^c = \{\omega_5, \omega_6, \omega_7, \omega_8\}$ et $B^c = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$.
- **Evénement impossible.** C'est l'événement qui ne se réalise jamais. Il est noté $\emptyset$, comme l'ensemble vide.
- **Evénement certain.** Cet événement est toujours réalisé. Il est noté Ω, comme l'espace fondamental.
- **Conjonction de deux événements.** Pour tout couple d'événements A et B, la conjonction désigne leur réalisation simultanée. Elle est notée comme l'intersection des deux ensembles $A \cap B$ (ou $A \wedge B$ ou $''AB''$ ou $''A$ et B'').

  ♣ **Exemple 1.4.** (suite). $A^c \cap B^c = \{\omega_5, \omega_6, \omega_7\}$.

  La relation $A \cap B = \emptyset$ signifie que les événements A et B sont *incompatibles* (ou que les ensembles A et B sont disjoints).

  ♣ **Exemple 1.5.** (suite). $A \cap B = \emptyset$.
- **Disjonction de deux événements.** Pour tout couple d'événements A et B, la disjonction désigne la réalisation de l'un ou de l'autre ou des deux à la fois. Elle est notée comme la réunion des deux ensembles $A \cup B$ ou $A \vee B$ ou "A ou B" ou encore $A + B$, lorsque A et B sont incompatibles. Notons bien qu'il s'agit d'un "ou" non exclusif.

  ♣ **Exemple 1.6.** (suite). $A \cup B = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_8\}; \quad A \cup B^c = B^c = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$.
- **Disjonction exclusive de deux événements**. Pour tout couple d'événements A et B, la disjonction exclusive désigne la réalisation de l'un ou de l'autre mais pas des deux. Elle est notée comme la différence symétrique des deux ensembles $A \Delta B$.
- **Implication.** Lorsque l'événement A ne peut être réalisé sans que l'événement B ne le soit aussi, alors on dit que "l'événement A *implique* l'événement B" et on note : $A \subset B$.
- **Système exhaustif d'événements.** C'est une suite finie ou infinie d'événements $(A_n, n \in I)$ telle que $A_n \cap A_m = \emptyset$, $(n \neq m)$ et $\cup_n A_n = \Omega$.

- Conjonction et disjonction infinies. Soit $A_n, n \geq 1$, une suite infinie d'événements.

$$\begin{aligned}\bigcup_{n\geq 1} A_n &= \{\omega \in \Omega : \omega \in A_n \text{ pour au moins un certain } n\},\\ \bigcap_{n\geq 1} A_n &= \{\omega \in \Omega : \omega \in A_n \text{ pour tout } n\}.\end{aligned}$$

1.5 Probabilité et fréquence statistique

Une probabilité est définie de manière rigoureuse en tant qu'objet mathématique. Elle peut aussi être approchée soit par la fréquence statistique, soit par la notion de symétrie.

La notion de probabilité s'est dégagée historiquement par les fréquences statistiques des événements. Lorsque nous observons la réalisation ou non d'un événement, noté A, au cours de répétitions indépendantes d'une même expérience aléatoire, si nous notons $n(A)$ le nombre de ses réalisations et N le nombre total des expériences effectuées, la fréquence de cet événement est alors exprimée par le rapport : $n(A)/N$.

Lorsque N augmente, la fréquence de l'événement fluctue expérimentalement de moins en moins, et nous pouvons écrire [1]

$$\frac{n(A)}{N} \to p_A, \quad \text{lorsque } n \to \infty, \tag{1.1}$$

où p_A est la probabilité de l'événement observé. Ce résultat expérimental est appelé : *loi empirique des grands nombres* et il a conduit à abstraire de l'expérience la notion de probabilité.

L'application $A \to n(A)/N$, possède les deux propriétés suivantes :

1. $0 \leq n(A)/N \leq 1$
2. si $A_1, ..., A_k$ sont k événements incompatibles, liés à l'expérience Ω, alors :

$$n(A_1 \cup ... \cup A_k)/N = n(A_1)/N + \cdots + n(A_k)/N. \tag{1.2}$$

1.6 Probabilité combinatoire

Dans de nombreux problèmes nous avons considéré des espaces de probabilités où la mesure de probabilité est uniforme, c'est-à-dire que pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = 1/|\Omega|$ [2] et par conséquent la probabilité d'un événement A est $\mathbb{P}(A) = |A|/|\Omega|$. Cela suppose évidemment que Ω soit fini. Le problème consiste alors à identifier l'espace fondamental Ω et à calculer son cardinal.

1. Permutations. Supposons que nous disposons de n objets distinguables. Le nombre d'arrangements différents de ces objets, appelés permutations, est $n(n-1)\cdots 1$, car nous avons n choix différents pour le premier objet, $n-1$ choix pour le second, etc., et nous n'avons qu'un choix pour le dernier. Ainsi, on a montré que

Proposition 1.1. Le nombre de permutations de n objets distincts ($n \in \mathbb{N}^*$) est :

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

Par convention $0! = 1$.

♣ **Exemple 1.7.** Une urne contient 3 boules numérotées de 1 à 3. En tirant au hasard, et sans remise, les trois boules de l'urne, la probabilité d'obtenir le numéro 321 est égale à $1/6$, car le nombre de permutations de trois boules est $3\,! = 1 \cdot 2 \cdot 3 = 6$.

2. Arrangements. Soit un ensemble S de n éléments et m un entier tel que $1 \leq m \leq n$. Un arrangement de S, m à m est un m-uplet $(s_1, ..., s_m) \in S^m$ formé de m éléments de S deux à deux disjoints.

1. Le sens de cette convergence sera précisé dans le chapitre 7.
2. $|\Omega|$ désigne le cardinal de l'ensemble Ω.

Proposition 1.2. Le nombre d'arrangements m à m d'un ensemble S de $n \geq m$ éléments est

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

Démonstration. On a n choix pour le 1er élément de $(s_1, ..., s_m)$, $(n-1)$ choix pour le deuxième élément, etc. □

♣ **Exemple 1.8.** En tirant 2 boules, sans remise dans l'urne précédente, la probabilité d'obtenir le numéro 23 est égale à $1/6$, car le nombre d'arrangements de 2 à 2 des 3 boules est égal à $\frac{3\,!}{(3-2)\,!} = 6$.

3. Combinaisons. Les différentes façons de considérer m éléments d'un ensemble de n, $(n \geq m)$, éléments sont appelées combinaisons de m objets parmi n. Ici l'ordre n'a pas d'importance.

Proposition 1.3. Le nombre de combinaisons de m objets parmi n, est : $\binom{n}{m} := C_n^m := \frac{n\,!}{m\,! \cdot (n-m)\,!}$.

Démonstration. Nous avons clairement $A_n^m = \binom{n}{m} \cdot m!$, d'où le résultat. □

♣ **Exemple 1.9.** Dans une urne contenant m boules blanches et n boules noires, nous effectuons r $(r \leq m)$ tirages sans remise. La probabilité que les r tirages donnent tous des boules blanches est égale à $\binom{m}{r}/\binom{n+m}{r}$. En effet, il y a $\binom{m}{r}$ façons différentes de tirer r boules parmi m et $\binom{n+m}{r}$ façons de tirer r boules parmi $m+n$.

4. Combinaisons multinomiales. Lorsque les n objets sont partiellement indistinguables et forment r familles distinctes comptant $n_1, n_2, ..., n_r$ objets chacune, $(n_1 + n_2 + \cdots + n_r = n)$, alors

Proposition 1.4. Le nombre de permutations de n objets partiellement distincts est égal à

$$\frac{n\,!}{n_1! \cdots n_r!}.$$

Remarque 1.1. En fait, cette formule est le coefficient du développement multinomiale :

$$(a_1 + \cdots + a_r)^n = \sum_{n_1 + \cdots + n_r = n} \frac{n\,!}{n_1! \cdots n_r!} a_1^{n_1} \cdots a_r^{n_r}.$$

La même formule est obtenue lorsque n objets distincts doivent se placer dans r boîtes telles que la boîte 1 contient n_1 objets, la boîte 2 contient n_2 objets,..., la boîte r contient n_r objets. Le nombre des différents façons de réaliser cette répartition est donné par la proposition ci-dessus.

Remarque 1.2. Lorsque $r = 2$, nous avons le cas précédente de combinaisons avec $n_1 = m$ et $n_2 = n - m$.

♣ **Exemple 1.10.** Soit un jeu des 52 cartes : as, deux,..., dix, Roi, Dame, Valet, et pour chacune de ces cartes les quatre couleurs : coeur (♡), pique (♠), carreau (♢) et trèfle (♣).

Lors du jeu de bridge, les 13 cartes obtenues par chacun des 4 joueurs est un cas parmi

$$\frac{52\,!}{13!13!13!13!} = 5,364 \times 10^{28}$$

possibilités.

Considérons une urne contenant n boules numérotées de 1 à n $(n \geq 2)$. On tire k boules de cette urne $(1 \leq k \leq n)$, notées $b_1, ..., b_k$, avec $b_i = 1, ..., n$. Si l'ordre est d'importance, alors on notera le résultat $(b_1, ..., b_k)$ et on l'appelle *échantillon ordonné*, sinon (l'ordre n'a pas d'importance), on notera le résultat $\{b_1, ..., b_k\}$ et on l'appelle *échantillon (non ordonné)*.

Supposons maintenant que le tirage se fait selon les modes suivantes.

1. *Tirage avec remise*. On tire une boule et après avoir lu son numéro on la remet dans l'urne et on procède au tirage suivant.

(a) *Echantillon ordonné*. L'espace fondamental est $\Omega = \{\omega : \omega = (b_1, ..., b_k), b_i = 1, ..., n\}$, et donc $|\Omega| = n^k$.

(b) *Echantillon non ordonné*. L'espace fondamental est $\Omega = \{\omega : \omega = \{b_1, ..., b_k\}, b_i = 1, ..., n\}$, et donc $|\Omega| = \binom{n+k-1}{k}$.

2. *Tirage sans remise*. On tire la première boule et on la met de côté, en suite on tire la deuxième boule, etc.

(a) *Echantillon ordonné*. L'espace fondamental est $\Omega = \{\omega : \omega = (b_1, ..., b_k), b_i \neq b_j, i \neq j, , b_i = 1, ..., n\}$, et donc $|\Omega| = A_n^k$.

(b) *Echantillon non ordonné*. L'espace fondamental est $\Omega = \{\omega : \omega = \{b_1, ..., b_k\}, b_i \neq b_j, i \neq j, b_i = 1, ..., n\}$, et donc $|\Omega| = \binom{n}{k}$.

1.7 Probabilité géométrique

La probabilité géométrique concerne la disposition de points dans une région géométrique (i.e., un segment d'une droite, un domaine d'un plan ou de l'espace habituel tridimensionnel, etc.). Les mesures des régions géométriques (i.e., longueur d'un segment d'une droite, aire d'une surface, volume, etc.) lorsqu'elles sont normalisées, vérifient les trois axiomes de probabilité.

♣ **Exemple 1.11.** Soit $[a, b] \subset [0, 1]$. La probabilité qu'un point tiré au hasard dans $[0, 1]$ soit un point de $[a, b]$, est égale à $(b - a)$.

1.8 Définition axiomatique de la probabilité

Dans ce qui suit nous donnons les notions de base du *système axiomatique de Kolmogorov*, utiles pour la suite de ce cours.

Tout événement A défini sur une expérience aléatoire Ω est un sous ensemble de Ω, mais tout sous ensemble de Ω n'est pas *a priori* un événement. Il faut en outre qu'il fasse partie d'un "catalogue" d'événements, noté $\mathcal{F}$ et défini sur Ω, que nous appelons *σ-algèbre* ou *tribu*.

Définition 1.1. Une famille $\mathcal{F}$ de sous ensembles de Ω est appelée *σ-algèbre* ou *tribu* si elle vérifie les propriétés suivantes :

1. $\Omega \in \mathcal{F}$
2. Si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$
3. Si $A_1, A_2, \cdots \in \mathcal{F}$ alors $\cup_n A_n \in \mathcal{F}$

L'espace $(\Omega, \mathcal{F})$ est appelé *espace mesurable ou probabilisable*.

♣ **Exemple 1.12.** $\mathcal{F} = \{\emptyset, \Omega\}$ est une σ-algèbre car elle vérifie les 3 axiomes ; elle est appelée σ-algèbre triviale. C'est la plus petite σ-algèbre sur Ω.

♣ **Exemple 1.13.** $\mathcal{F} = \mathcal{P}(\Omega)$, est la σ-algèbre de tous les sous ensembles de Ω. C'est la plus grosse σ-algèbre sur Ω.

♣ **Exemple 1.14.** $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ est la σ-algèbre de Bernoulli.

Proposition 1.5. Soit $A, B \in \mathcal{F}$, alors $A \cap B \in \mathcal{F}$, $A \cup B \in \mathcal{F}$, $A - B = A \setminus B \in \mathcal{F}$ et $A \Delta B \in \mathcal{F}$.

Proposition 1.6. L'intersection de deux tribus est une tribu. La réunion de deux tribu n'est pas forcément une tribu.

Définition 1.2. Une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ est une application de $\mathcal{F}$ dans $\mathbb{R}$ qui vérifie les propriétés suivantes :

1. $\mathbb{P}(\Omega) = 1$
2. $\mathbb{P}(A) \geq 0$ pour tout $A \in \mathcal{F}$

3. Pour toute suite d'événements $A_n, n \geq 1$, deux à deux disjoints, nous avons

$$\mathbb{P}(\cup_n A_n) = \sum_n \mathbb{P}(A_n), \qquad (\sigma - \text{additivité}).$$

Remarque 1.3. La probabilité est un cas particulier des applications appelées *mesures* ; parfois nous disons *mesure de probabilité*.

Définition 1.3. L'espace $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé *espace probabilisé* ou *espace de probabilité* ou encore *modèle de probabilité*.

En conclusion, nous pouvons dire que toute expérience aléatoire se décrit mathématiquement par la donnée d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Remarque 1.4. Lorsque l'espace fondamental est dénombrable, nous pouvons nous affranchir de la définition de la tribu en l'identifiant à l'ensemble de tous les parties de l'espace fondamental. Dans ce cas tout sous ensemble de l'espace fondamental est un événement ; pour simplifier, l'espace probabilisé en question sera noté : $(\Omega, \mathbb{P})$.

Un événement A tel que $\mathbb{P}(A) = 1$ est dit *presque sûr* (p.s.) ou *de mesure pleine*. On dit aussi que $\mathbb{P}$ est portée par A. Nous verrons qu' il y a des événements de mesure pleine différents de Ω.

Un événement B tel que $\mathbb{P}(B) = 0$ est dit *négligeable*. Nous verrons également qu'il y a des événements négligeables différents de $\emptyset$.

Système des constituants. C'est un système exhaustif $(A_i, i \in I)$ ne contenant pas d'événements négligeables.

Remarque 1.5. Parfois, lorsque plusieurs probabilités sont en jeu, on note : $\mathbb{P} - p.s.$ ou $\mathbb{P}$-négligeable, afin de faire référence à la probabilité appropriée.

Soit un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, tel que pour tout $\omega \in \Omega$, $\{\omega\} \in \mathcal{F}$. Un élément $\omega \in \Omega$, tel que $\mathbb{P}(\{\omega\}) \neq 0$, est appelé *atome ponctuel* de $\mathbb{P}$.

La probabilité $\mathbb{P}$ est dite *discrète* ou *purement atomique*, s'il existe un événement A, (i.e. $A \in \mathcal{F}$), au plus infini dénombrable, de mesure pleine. $\mathbb{P}$ est dite *continue* ou *diffuse* si pour tout $\omega \in \Omega$ on a $\mathbb{P}(\{\omega\}) = 0$.

Les résultats suivants découlent directement de la définition de probabilité.

Proposition 1.7. Soient A et B deux événements d'une expérience Ω. Nous avons :

1. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
2. $\mathbb{P}(A) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B)$
3. si $A \subset B$ alors $\mathbb{P}(A) \leq \mathbb{P}(B)$
4. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
5. $\mathbb{P}(A \Delta B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B)$.

Chapitre 2

Calcul élémentaire de probabilités

2.1 Introduction

Le système axiomatique de Kolmogorov est complété par la définition de la probabilité conditionnelle et de l'indépendance de deux événements.

Les événements considérés dans ce chapitre sont supposés être définis sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

2.2 Probabilité conditionnelle

Définition 2.1. Pour tout couple d'événements A et B tel que $\mathbb{P}(B) > 0$, on définit la probabilité conditionnelle de A *sachant* B, notée $\mathbb{P}(A \mid B)$, par

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2.1}$$

C'est la probabilité de l'événement A sachant que l'événement B a été réalisé.

- Si A et B sont incompatibles (i.e. $A \cap B =$), alors $\mathbb{P}(A \cap B) = 0$ et par conséquent $\mathbb{P}(A \mid B) = 0$.
- Si $B \subset A$ alors $\mathbb{P}(A \cap B) = \mathbb{P}(B)$ et $\mathbb{P}(A \mid B) = 1$.
- Si $A \subset B$ alors $\mathbb{P}(A \mid B) = \mathbb{P}(A)/\mathbb{P}(B)$.

♣ **Exemple 2.1.** Dans l'expérience du jet de deux dés, si le premier est un 3, quelle est la probabilité que le total dépasse 6 ?

L'espace fondamental de cette expérience est : $\Omega = \{1, 2, 3, 4, 5, 6\}^2$. L'événement "le premier jet donne un 3" est : $A = \{(3, x) : x = 1, 2, 3, 4, 5, 6\}$. L'événement "le total dépasse 6" est : $B = \{(x, y) : x + y > 6\}$.

Par conséquent, nous avons :

$A \cap B = \{(3, 4), (3, 5), (3, 6)\}$ d'où $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{|A \cap B|}{|A|} = \frac{1}{2}$.

♣ **Exemple 2.2.** Quelle est la probabilité que les deux enfants d'une famille soient des garçons, sachant qu'au moins un est un garçon ?

L'espace fondamental est : $\Omega = \{FF, FG, GF, GG\}$ avec $\mathbb{P}(\{FF\}) = \mathbb{P}(\{FG\}) = \mathbb{P}(\{GF\}) = \mathbb{P}(\{GG\}) = 1/4$ (probabilité uniforme sur Ω).

Les événements concernés sont :

Deux garçons : $A = \{GG\}$

Au moins un garçon : $B = \{FG, GF, GG\}$

Nous observons que $A \subset B$. Par conséquent : $\mathbb{P}(A \mid B) = \mathbb{P}(A)/\mathbb{P}(B) = 1/3$.

La probabilité conditionnelle est aussi une probabilité car elle vérifie les trois axiomes de probabilité, i.e., pour tout événement B tel que $\mathbb{P}(B) > 0$,

1. $\mathbb{P}(A \mid B) \geq 0$
2. $\mathbb{P}(B \mid B) = 1$
3. $\mathbb{P}(\bigcup_{n\geq 0} A_n \mid B) = \sum_{n\geq 0} \mathbb{P}(A_n \mid B)$, $(A_i \cap A_j = , i \neq j)$

La démonstration est laissée en exercice.

2.3 Indépendance stochastique

Définition 2.2. Deux événements A et B sont dits stochastiquement indépendants ou indépendants en probabilité ou simplement indépendants, si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{2.2}$$

En comparant les relations (2.1) et (2.2) ci-dessus, on peut écrire

$$\mathbb{P}(A|B) = \mathbb{P}(A) \tag{2.3}$$

Par conséquent, la réalisation ou non de l'événement B n'a aucune influence sur la probabilité de l'événement A.

De manière générale, une famille finie ou infinie dénombrable d'événements $(A_i, i \in I)$ est dite indépendante si, pour tout sous ensemble fini J de I, on a

$$\mathbb{P}(\cap_{j\in J} A_j) = \prod_{j\in J} \mathbb{P}(A_j) \tag{2.4}$$

Remarque 2.1. Notons que dans le dernier cas ci-dessus, l'indépendance des événements deux à deux ne suffit pas pour que les événements soient indépendants.

Remarque 2.2. Le fait que deux événements A et B soient incompatibles n'implique pas qu'ils soient indépendants. Au contraire, pour deux événements incompatibles A et B tels que $\mathbb{P}(A) \times \mathbb{P}(B) \neq 0$, on montre par la relation (2.2) qu'ils ne peuvent pas être indépendants.

Proposition 2.1. Si A et B sont deux événements indépendants, alors :

1. A^c et B sont indépendants,
2. A et B^c sont indépendants,
3. A^c et B^c sont indépendants.

Démonstration. (1) $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B)$. Les points (2) et (3) peuvent être démontrés de la même manière. □

♣ **Exemple 2.3.** Soit une expérience aléatoire décrite par $(\Omega, \mathcal{F}, \mathbb{P})$ avec

$$\Omega = \{abc, acb, bac, bca, cab, cba, aaa, bbb, ccc\}$$

et $\mathbb{P}$ la probabilité uniforme sur Ω, i.e. pour tout $\omega \in \Omega$ on a $\mathbb{P}(\{\omega\}) = 1/|\Omega| = 1/9$. Considérons maintenant les trois événements suivants

$$A_k = \{ \text{ la } k\text{-ième lettre est un } a \}, k = 1, 2, 3.$$

Nous avons

$$\begin{aligned} \mathbb{P}(A_1) &= \mathbb{P}(\{abc, acb, aaa\}) = 3/9 \\ \mathbb{P}(A_2) &= \mathbb{P}(\{bac, cab, aaa\}) = 3/9 \\ \mathbb{P}(A_1 \cap A_2) &= \mathbb{P}(\{aaa\}) = 1/9 \end{aligned}$$

d'où

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$$

De même, nous avons : $\mathbb{P}(A_2 \cap A_3) = \mathbb{P}(A_3)\mathbb{P}(A_3)$ et $\mathbb{P}(A_3 \cap A_1) = \mathbb{P}(A_3)\mathbb{P}(A_1)$. D'autre part

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(\{aaa\}) = 1/9$$

Par conséquent

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$$

Conclusion : les événements A_1, A_2 et A_3 ne sont pas indépendants.

Définition 2.3 (Indépendance de deux familles d'événements)**.**
Soient deux familles d'événements $\mathcal{L}_1$ et $\mathcal{L}_2$ sur $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que ces deux familles sont indépendantes si pour tous $A \in \mathcal{L}_1$ et $B \in \mathcal{L}_2$, nous avons : $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Une notion également importante en probabilité est celle d'indépendance conditionnelle.

Définition 2.4 (Indépendance conditionnelle)**.**
Sur un espace probabilisé considérons l'événement B tel que $\mathbb{P}(B) > 0$ et les événements $A_1, ..., A_n$, $(n \geq 2)$. Alors si

$$\mathbb{P}(A_1 \cap \cdots \cap A_n | B) = \mathbb{P}(A_1|B) \cdots \mathbb{P}(A_n|B)$$

on dit que les événements $A_1, ..., A_n$ sont conditionnellement indépendants sachant B.

2.4 Suites monotones d'événements

Définition 2.5. -Proposition

1. Une suite d'événements $(A_n, n \in \mathbb{N})$ est appelée croissante si $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$, et nous avons
$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$
2. Une suite d'événements $(A_n, n \in \mathbb{N})$ est appelée décroissante si $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$, et nous avons
$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

Proposition 2.2 (Suites monotones d'événements)**.**
Si $(A_n, n \in \mathbb{N})$ est une suite monotone (i.e., croissante ou décroissante) d'événements, alors :

$$\mathbb{P}(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Démonstration. Soit la suite d'événements croissante $A_n, n \geq 1$, et la suite $B_n, n \geq 1$, définie par $B_1 = A_1$, et $B_n = A_{n-1}^c \cap A_n$, pour $n \geq 2$, deux à deux disjoints. Il est clair que $A_n = \cup_{i=1}^n B_i$, $n \geq 1$. Nous avons :

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(\cup_{k\geq 1} A_k) = \mathbb{P}(\cup_{k\geq 1} B_k) = \sum_{k\geq 1} \mathbb{P}(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(B_k) \\ &= \lim_{n\to\infty} \mathbb{P}(\cup_{k=1}^n B_k) = \lim_{n\to\infty} \mathbb{P}(A_n).\end{aligned}$$

2. Dans le cas où la suite $A_n, n \geq 1$, est décroissante, la suite $A_n^c, n \geq 1$, est croissante. On applique alors ce qui précède. □

2.5 Formules remarquables

Tous les événements considérés sont supposés être définis sur le même modèle de probabilité.

Proposition 2.3 (Théorème des probabilités totales).
Soit $(B_n, n \in I)$ un système exhaustif d'événements et A un événement. Nous avons

$$\mathbb{P}(A) = \sum_{n\in I} \mathbb{P}(A \cap B_n) = \sum_{n\in I} \mathbb{P}(A \mid B_n)\mathbb{P}(B_n)$$

La deuxième égalité est valable si : $\mathbb{P}(B_k) > 0$, pour tout $k \in I$.

Démonstration. Nous avons : $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap (\cup_n B_n)) = \mathbb{P}(\cup_n (A \cap B_n)) = \sum_n \mathbb{P}(A \cap B_n)$ car les événements $A \cap B_n$ sont deux à deux disjoints. □

Proposition 2.4 (Formule de Poincaré ou d'inclusion-exclusion).
Pour une suite finie d'événements : $A_1, A_2, \ldots, A_n$, nous avons

$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\leq i<j\leq n} \mathbb{P}(A_i \cap A_j) + \cdots + (-1)^{n-1}\mathbb{P}(\cap_{i=1}^{n} A_i)$$

Démonstration. Par récurrence sur $n \in \mathbb{N}^*$. Nous remarquons qu'elle est trivialement vérifiée pour $n = 1$. Supposons qu'elle est vérifiée pour n entier quelconque > 1. Nous allons montrer qu'elle est vérifiée pour $n + 1$. Nous avons $\mathbb{P}(\cup_{i=1}^{n+1} A_i) = \mathbb{P}(\cup_{i=1}^{n} A_i) + \mathbb{P}(A_{n+1}) - \mathbb{P}((\cup_{i=1}^{n} A_i) \cap A_{n+1})$. En utilisant l'hypothèse de récurrence et en arrangeant les termes du deuxième membre, nous obtenons le résultat énoncé. □

Proposition 2.5 (Formule des produits disjoints).
Pour une suite finie d'événements : $A_1, A_2, \ldots, A_n$, nous avons

$$\mathbb{P}(\cup_{i=1}^{n} A_i) = \mathbb{P}(A_1) + \mathbb{P}(A_1^c A_2) + \mathbb{P}(A_1^c A_2^c A_3) + \cdots + \mathbb{P}(A_1^c A_2^c \ldots A_{n-1}^c A_n)$$

Démonstration. Cela résult de la transformation de la disjonction de événements A_i en une disjonction d'événements disjoints comme suit : $A_1 \cap A_2 \cup \cdots \cup A_n = A_1 \cup A_1^c A_2 \cup A_1^c A_2^c A_3 \cup \cdots \cup A_1^c A_2^c \ldots A_{n-1}^c A_n$. □

Proposition 2.6 (Formule de multiplication).
Pour une suite finie d'événements : $A_1, A_2, \ldots, A_n$, tels que : $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$, nous avons

$$\mathbb{P}(\cap_{i=1}^{n} A_i) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_1 A_2) \cdots \mathbb{P}(A_n \mid A_1 A_2 \ldots A_{n-1})$$

Démonstration. Nous avons : $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_n | A_1 \cap \cdots \cap A_{n-1}) \times \mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) =$ etc. □

Proposition 2.7 (Formule de Bayes).
Soient $(B_n, n \in I)$ un système exhaustif d'événements avec $\mathbb{P}(B_k) > 0$, pour tout $k \in I$ et un événement A tel que $\mathbb{P}(A) > 0$, alors

$$\mathbb{P}(B_j \mid A) = \frac{\mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}{\sum_{k\in I} \mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}$$

Démonstration. Par la relation $\mathbb{P}(B_j|A) \times \mathbb{P}(A) = \mathbb{P}(A|B_j) \times \mathbb{P}(B_j)$ et le théorème des probabilités totales nous obtenons le résultat énoncé. □

Proposition 2.8 (Inégalité de Boole).
Soit une suite finie ou infinie d'événements $(A_n, n \in I)$, alors nous avons l'inégalité suivante

$$\mathbb{P}(\bigcup_{n\in I} A_n) \leq \sum_{n\in I} \mathbb{P}(A_n) \tag{2.5}$$

Démonstration. Cas fini. Pour $n = 1$ on a l'égalité. Pour $n = 2$ on a :

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Pour $n \geq 3$ on a

$$\mathbb{P}(A_1 \cup ... \cup A_n \cup A_{n+1}) \leq \mathbb{P}(A_1 \cup ... \cup A_n) + \mathbb{P}(A_{n+1}).$$

Maintenant, on applique l'hypothèse de la recurrence pour achever la démonstration.
Cas dénombrable. De ce qui précède on a :

$$\mathbb{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i) \leq \sum_{i=1}^\infty \mathbb{P}(A_i)$$

Donc, on utilise la proposition 2.2, et on obtient

$$\lim_{n\to\infty} \mathbb{P}(\cup_{i=1}^n A_i) = \mathbb{P}(\cup_{i=1}^\infty A_i) \leq \sum_{i=1}^\infty \mathbb{P}(A_i).$$

□

2.6 Expérience aléatoire composée

Lorsque nous voulons considérer des événements aléatoires liés à plusieurs expériences aléatoires, nous devons construire un espace probabilisé qui décrit l'ensemble des expériences en question. Nous allons donner une brève présentation des espaces produits en se basant sur des espaces probabilisés dénombrables.

Soient deux espaces probabilisés dénombrables décrivant deux expériences aléatoires particulières $(\Omega_1, \mathbb{P}_1)$ et $(\Omega_2, \mathbb{P}_2)$ et un observateur qui observerait les deux expériences en même temps. L'issue d'une telle expérience se présente comme un couple de points : (ω_1, ω_2) avec $\omega_1 \in \Omega_1$ et $\omega_2 \in \Omega_2$. Autrement dit, l'espace fondamental de cette expérience est l'espace produit $\Omega_1 \times \Omega_2$, noté Ω. Sur cet espace nous pouvons définir une probabilité $\mathbb{P}$ par les relations suivantes

$$\mathbb{P}(\{\omega\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}), \quad pour \ \ tout \quad \omega = (\omega_1, \omega_2) \in \Omega$$
$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\}), \quad pour \ \ tout \quad A \subset \Omega.$$

Tout événement d'une expérience peut se représenter soit comme un sous-ensemble A de Ω_1, soit comme un sous-ensemble $A \times \Omega_2$ de Ω et il est facile de montrer que

$$\mathbb{P}(A \times \Omega_2) = \sum_{A_1}\sum_{\Omega_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) = \mathbb{P}_1(A).$$

Deux événements, A_1 lié à la première expérience et A_2 lié à la seconde expérience, sont indépendants relativement à $\mathbb{P}$, car on peut écrire

$$\begin{aligned}\mathbb{P}[(A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)] &= \mathbb{P}(A_1 \times A_2) = \sum_{A_1}\sum_{A_2} \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) \\ &= \mathbb{P}_1(A_1)\mathbb{P}_2(A_2) = \mathbb{P}(A_1 \times \Omega_2)\mathbb{P}(\Omega_1 \times A_2).\end{aligned}$$

Remarque 2.3. Dans le cas d'espaces probabilisés généraux, $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ et $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ l'espace produit est noté $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$. La tribu $\mathcal{F}_1 \otimes \mathcal{F}_2$ est engendrée par tous les ensembles $A_1 \times A_2$ avec $A_1 \in \mathcal{F}_1$ et $A_2 \in \mathcal{F}_2$. Et la probabilité produit $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ est définie de la même manière que ci-dessus.

Chapitre 3

Variable aléatoire discrète

3.1 Introduction

Une variable aléatoire (v.a.) est définie par référence à une expérience aléatoire comme une application, soit X, dont la valeur dépend du résultat ω de cette expérience. Les variables aléatoires que nous allons étudier dans ce cours prennent leurs valeurs dans l'ensemble $\mathbb{R}$ ou au plus dans $\mathbb{R}^d, (d > 1)$.

De manière plus précise, une variable aléatoire réelle (v.a.r.) est définie comme suit.

Définition 3.1. Une v.a. réelle, définie sur un espace mesurable $(\Omega, \mathcal{F})$, est une application X définie sur Ω et à valeurs dans $\mathbb{R}$ telle que, pour tout $x \in \mathbb{R}$,

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F} \tag{3.1}$$

Remarque 3.1. 1) Une application X sur $(\Omega, \mathcal{F})$ vérifiant (3.1) est dite *mesurable*.

2) L'événement $\{\omega : X(\omega) \leq x\}$ sera noté simplement : $\{X \leq x\}$.

♣ **Exemple 3.1.** Dans l'expérience du jet de deux dés, la somme, le produit, la différence, etc., des points amenés par les dés définissent autant de v.a., i.e.

$$X(\omega) = x + y, \ \ Y(\omega) = x \cdot y, \ \ Z(\omega) = x - y$$

où $\omega = (x, y)$.

Lorsque l'ensemble de valeurs E de la v.a. X est au plus dénombrable, la v.a. X est dite *discrète* (v.a.d.). Dans le cas où E est un ensemble fini, la v.a. X est dite *finie* ou *simple*. Lorsque $E = \overline{\mathbb{R}}$ la v.a. est dite *numérique* ou v.a.r.. Nous reviendrons plus tard, dans le cadre des v.a. à densité, sur le classement des v.a..

Remarque 3.2. Dans le cas où X est une v.a.d., comme c'est le cas dans ce chapitre, et on considère $\mathcal{F} = \mathcal{P}(\Omega)$, alors (3.1) est vérifiée pour toute fonction X. Ainsi toute fonction $X : \Omega \longrightarrow \mathbb{R}$ est une v.a.

Variable aléatoire complexe. C'est une v.a. $Z = X + iY$ où X, Y sont des v.a.r. $(i = \sqrt{-1})$.

Variable aléatoire indicatrice. Pour tout événement A de $\mathcal{F}$, nous définissons sa v.a. indicatrice, notée $\mathbf{1}_A$, par

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & si \quad \omega \in A \\ 0 & sinon \end{cases} \tag{3.2}$$

Quelques propriétés :

1. $\mathbf{1}_\Omega = 1$ et $\mathbf{1}_\emptyset = 0$
2. $\mathbf{1}_{A \cap B} = \mathbf{1}_A \cdot \mathbf{1}_B$

3. $\mathbf{1}_{A\cup B} = \mathbf{1}_A + \mathbf{1}_B - \mathbf{1}_A \cdot \mathbf{1}_B$
4. $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$
5. $A \subset B \iff \mathbf{1}_A \leq \mathbf{1}_B$ et $A = B \iff \mathbf{1}_A = \mathbf{1}_B$
6. $(\mathbf{1}_A)^{-1}(\{1\}) = A$ et $(\mathbf{1}_A)^{-1}(\{0\}) = A^c$.

Une v.a. discrète $X : \Omega \longrightarrow E = \{x_i : i \in I\}$, avec $I \subset \mathbb{N}$, peut être représentée comme suit

$$X = \sum_{i\in I} x_i \mathbf{1}_{A_i} \tag{3.3}$$

où la suite d'événements $(A_i,\ i \in \mathbb{N}^*)$ forme un système exhaustif de Ω. Nous avons

$$A_i = \{X = x_i\} = X^{-1}(\{x_i\}),$$

avec

$$\bigcup_{i\in I} \{X = x_i\} = \Omega.$$

On appellera la représentation (3.3) canonique.

3.2 Loi d'une variable aléatoire discrète

Soit $X : \Omega \longrightarrow E \subset \mathbb{R}$ une v.a. discrète. Pour tout $x \in E$ nous définissons l'événement $\{\omega : X(\omega) = x\}$, noté simplement $\{X = x\}$, et nous notons $p_X(x)$ sa probabilité, i.e. $p_X(x) = \mathbb{P}(X = x)$.

Définition 3.2. La famille des nombres $(p_X(x), x \in E)$ tels que $p_X(x) = \mathbb{P}(X = x)$ est appelée loi de la v.a. X.

Propriétés :

1. $p_X(x) \geq 0$
2. $\sum_{x\in E} p_X(x) = 1$

L'intérêt de la loi de probabilité d'une v.a. est de permettre de calculer directement, sans passer par l'espace de probabilité, les probabilités des événements définis par X, i.e., pour tout événement A, $A \subset E$, nous avons

$$\mathbb{P}(X \in A) = \sum_{x\in A} p_X(x). \tag{3.4}$$

♣ **Exemple 3.2.** Jet d'un dé (non pipé). La loi de la v.a. $X(\omega) = \omega$ est $p_X(\omega) = 1/6$ pour tout $\omega \in \Omega$, et elle est appelée *loi uniforme*.

♣ **Exemple 3.3.** Jet de deux dés.
La loi de la v.a. $X(\omega) = \omega_1 + \omega_2$ est donnée dans le tableau suivant

x	2	3	4	5	6	7	8	9	10	11	12
p(x)	1/36	2/ 36	3/ 36	4/ 36	5/ 36	6/ 36	5/ 36	4/ 36	3/ 36	2/ 36	1/ 36

Définition 3.3. La fonction $F_X(\cdot)$, définie sur $\mathbb{R}$ et à valeurs dans $[0, 1]$, par la formule

$$F_X(x) := \mathbb{P}(X \leq x) = \sum_{\{y\in E: y\leq x\}} p_X(y) \tag{3.5}$$

est appelée *fonction de répartition* de la v.a.d. X.

C'est une fonction croissante et constante par morceaux.
On définit également

$$F_X(x-) := \mathbb{P}(X < x) = \sum_{\{y \in E : y < x\}} p_X(y). \tag{3.6}$$

Considérons deux v.a.d. X_1, X_2 définies sur le même espace de probabilité et à valeurs dans E_1, E_2 respectivement.

Définition 3.4 (Indépendance de deux variables aléatoires)**.** .
Les v.a. X_1, X_2 sont dites indépendantes si, pour tout $x_1 \in E_1, x_2 \in E_2$, nous avons

$$\mathbb{P}(X_1 = x_1, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \tag{3.7}$$

Proposition 3.1. Soit X et Y deux v.a.d. indépendantes et $g : \mathbb{R} \to \mathbb{R}$ une application. Alors les v.a.d. $g \circ X$ et $g \circ Y$ sont indépendante.

Définition 3.5. On dit que les v.a. X et Y sont égales p.s., et on note $X \overset{p.s.}{=} Y$, si $\mathbb{P}(X = Y) = 1$.

On dit que les v.a. X et Y sont égales en loi, on note $X \overset{\mathcal{L}}{=} Y$, si $P_X = P_Y$ ou $F_X = F_Y$.

On dit que la v.a. X est plus grande que la v.a. Y p.s., et on note $X \overset{p.s.}{\geq} Y$, si $\mathbb{P}(X \geq Y) = 1$.

3.3 Espérance mathématique et moments

L'espérance d'une v.a. X est notée $\mathbb{E}X$ ou $\mathbb{E}[X]$. Si X est discrète à valeurs dans E (représentée par la relation (3.3)) et de loi $p = (p(x), x \in E)$ son *espérance mathématique* ou simplement *espérance* ou *valeur moyenne* est définie par

$$\mathbb{E}X = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \tag{3.8}$$

Proposition 3.2. Nous avons :

$$\mathbb{E}X = \sum_{x \in E} x p(x). \tag{3.9}$$

L'espérance de X existe et est finie si, et seulement si, $\mathbb{E}|X| = \sum_{x \in E} |x| p(x) < +\infty$.
Propriétés :

1. $\mathbb{E}[aX] = a\mathbb{E}[X]$, $(a \in \mathbb{R})$
2. $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$
3. Si $X \geq 0$ alors $\mathbb{E}[X] \geq 0$ et, si $X \geq Y$, alors $\mathbb{E}X \geq \mathbb{E}Y$.

Notons qu'on peut écrire l'espérance aussi à l'aide de la f.r., i.e.,

$$\mathbb{E}X = \sum_{x \in \mathbb{R}} x[F(x) - F(x-)] = \sum_{i \in I} x_i[F(x_i) - F(x_{i-1})].$$

♣ **Exemple 3.4.** Espérance d'une v.a. indicatrice.

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

♣ **Exemple 3.5.** Soit une v.a.d. X à valeurs dans $\mathbb{N}^*$ et de loi $p_X(n) = 6/\pi n^2$. Cette v.a. ne possède pas d'espérance car $\sum 1/n = \infty$.

Proposition 3.3. Soient une v.a.d. $X : \Omega \to E$ et une fonction $g : E \to F$, $(E, F \subset \mathbb{R})$, alors $g \circ X$ est une v.a.d. et

$$\mathbb{E}[g(X)] = \sum_{x \in E} g(x)\mathbb{P}(X = x),$$

à condition que $\mathbb{E}|g(X)| = \sum_{x \in E} |g(x)| p(x) < +\infty$.

Proposition 3.4. Si X et Y sont deux v.a.d. indépendantes, possédant des espérances finies, alors

$$\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y \tag{3.10}$$

Démonstration. On a

$$X = \sum_{i=1}^{\infty} x_i \mathbf{1}_{A_i} \quad \text{avec} \quad A_i = \{X = x_i\}$$
$$Y = \sum_{j=1}^{\infty} y_j \mathbf{1}_{B_j} \quad \text{avec} \quad B_j = \{Y = y_j\}$$

d'où

$$XY = \sum_{i,j} x_i y_j \mathbf{1}_{A_i} \mathbf{1}_{B_j} = \sum_{i,j} x_i y_j \mathbf{1}_{A_i \cap B_j}$$

donc XY est une v.a. discrète de valeur $x_i y_j$ sur $A_i \cap B_j$ et $\cup_{i,j} A_i \cap B_j = (\cup_i A_i) \cap (\cup_j B_j) = \Omega$.

Du fait que X et Y sont indépendantes, nous avons, pour tous i et j

$$\mathbb{P}(A_i \cap B_j) = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = \mathbb{P}(A_i)\mathbb{P}(B_j)$$

Par la définition de l'espérance d'une v.a.d., nous avons pour la v.a.d. XY

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{i,j} x_i y_j \mathbb{P}(A_i \cap B_j) \\
&= \sum_i \sum_j x_i y_j \mathbb{P}(A_i)\mathbb{P}(B_j) \\
&= \{\sum_i x_i \mathbb{P}(A_i)\}\{\sum_j y_j \mathbb{P}(B_j)\} \\
&= \mathbb{E}X\mathbb{E}Y.
\end{aligned}$$

□

Remarque 3.3. Nous verrons dans le chapitre 5 qu'une condition nécessaire et suffisante pour que (3.10) soit vérifiée, est que les v.a. X et Y soient non corrélées.

Le moment d'ordre k, ($k \in \mathbb{N}^*$), est défini par

$$\mu_k = \mathbb{E}[X^k] = \sum_{x \in E} x^k p(x) \tag{3.11}$$

si la série est absolument convergente.

Le moment centré d'ordre k, ($k \in \mathbb{N}^*$), est défini par

$$m_k = \mathbb{E}[(X - \mathbb{E}X)^k] = \sum_{x \in E} (x - \mathbb{E}X)^k p(x) \tag{3.12}$$

Le moment absolu d'ordre k est défini par

$$\mathbb{E}[|X|^k] = \sum_{x \in E} |x|^k p(x) \tag{3.13}$$

Le moment centré d'ordre 2, m_2, est noté $\sigma^2(X)$ ou $\mathrm{Var}(X)$ et il est appelé *variance* de la v.a. X. La variance mesure le degré de dispersion des valeurs de la v.a. autour de sa moyenne. La racine carrée de la variance, $\sigma \geq 0$, est appelée *écart-type*.

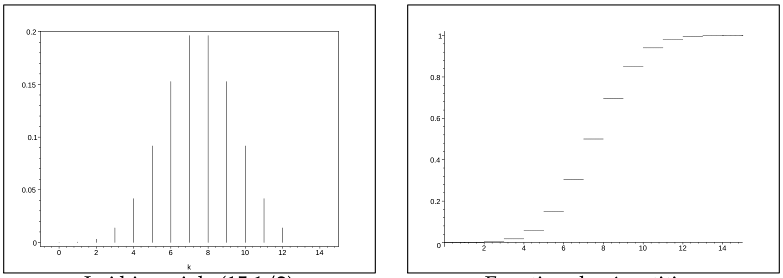
Proposition 3.5. Soit X une v.a. à valeurs dans $\mathbb{N}$, alors

$$\mathbb{E}X = \sum_{n \geq 1} \mathbb{P}(X \geq n)$$

Démonstration. Nous avons

$$\sum_{n \geq 1} \mathbb{P}(X \geq n) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(X = k) = \sum_{k \geq 1} \sum_{n=1}^{k} \mathbb{P}(X = k) = \sum_{k \geq 1} k\mathbb{P}(X = k) = \mathbb{E}X.$$

□



Loi binomiale (15,1/2) — Fonction de répartition

FIGURE 3.1 – Loi binomiale.

3.4 Quelques lois usuelles discrètes

3.4.1 Loi de Dirac

La v.a. X suit une loi de Dirac, si l'ensemble de ses valeurs se réduit à une seule valeur de probabilité non nulle, i.e. $E = \{a\}$, $a \in \mathbb{R}$. Cette loi est notée par δ_a, et elle est définie, pour tout $A \subset E$, par

$$\delta_a(A) = \begin{cases} 1 & si \quad a \in A \\ 0 & si \quad a \notin A. \end{cases}$$

Nous avons : $\mathbb{E}X = a$ et Var $X = 0$.

3.4.2 Loi de Bernoulli

La v.a. X suit une loi de Bernoulli de paramètre p ($0 < p < 1$), et on note : X$\sim B(p)$, si elle prend deux valeurs, soit 0 et 1 avec $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$.

L'événement $\{X = 1\}$ est appelé *succès* et l'événement $\{X = 0\}$ *échec*.

L'espérance de X est : $\mathbb{E}X = p$ et sa variance est : Var $X = \sigma^2 = p(1 - p)$.

3.4.3 Loi binomiale

Lorsque nous somme intéressés par le nombre des succès dans une suite finie d'expériences de Bernoulli, cela est décrit par une v.a. binomiale. Le nombre des succès dans une suite de n expériences de Bernoulli peut être : $0, 1, \ldots, n$. On dit qu'une v.a. Y suit une loi binomiale de paramètre $(n, p) \in \mathbb{N}^* \times]0, 1[$, et on note : $Y \sim b(n, p)$, si sa loi est donnée par

$$p(k) = \mathbb{P}(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad (k = 0, 1, \ldots, n) \tag{3.14}$$

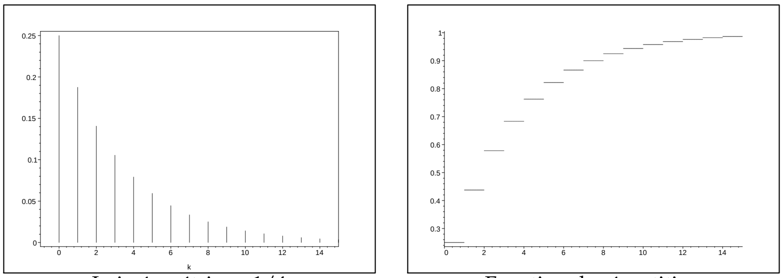
Si Y$\sim b(n, p)$, alors elle peut être représentée par la somme de n v.a. de Bernoulli de paramètre p indépendantes, i.e.

$$Y = X_1 + \cdots + X_n \tag{3.15}$$

avec X_i i.i.d.$\sim B(p)$.

Nous avons : $\mathbb{E}Y = np$ et Var $Y = np(1 - p)$.

Remarque 3.4. Lorsque n =1, on retrouve la loi de Bernoulli.



Loi géométrique 1/4 | Fonction de répartition

FIGURE 3.2 – Loi géométrique.

3.4.4 Loi géométrique

La loi géométrique est liée aux suites infinies d'expériences de Bernoulli indépendantes. La réalisation d'une v.a. géométrique, dans une telle suite, désigne le premier instant de réalisation d'un succès.

On dit que la v.a. Y suit une loi géométrique de paramètre p, $(0 < p < 1)$ et on note : $Y \sim G(p)$, si sa loi est donnée par

$$p(k) = \mathbb{P}(Y = k) = p(1-p)^{k-1}, \quad (k \in \mathbb{N}^*). \tag{3.16}$$

Considérons une suite infinie de v.a. de Bernoulli : $X_1, X_2, \cdots$ *i.i.d.* $\sim B(p)$. Alors une v.a. géométrique Y, définie relativement à cette suite, s'écrit comme suit

$$\{Y = k\} = \{X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1\}$$

L'espérance et la variance d'une v.a. géométrique sont : $\mathbb{E}Y = 1/p$ et $\operatorname{Var} Y = (1-p)/p^2$. La loi géométrique est caractérisée par l'absence de mémoire.

Nous avons également la loi géométrique sur $\mathbb{N}$, et on note $Y \sim G_0(p)$, si

$$p(k) = \mathbb{P}(Y = k) = p(1-p)^k, \quad (k \in \mathbb{N}),$$

avec $\mathbb{E}(Y) = (1-p)/p$.

3.4.5 Loi de Poisson

Une v.a. X suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$, et on note : $X \sim P(\lambda)$, si sa loi est donnée par

$$p(k) = \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad (k \in \mathbb{N}) \tag{3.17}$$

L'espérance et la variance sont : $\mathbb{E}X = \operatorname{Var} X = \lambda$.

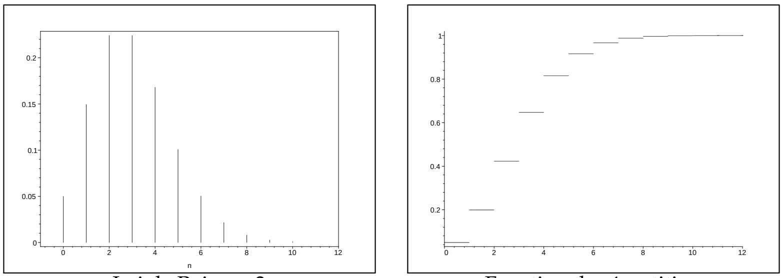
Proposition 3.6. [Théorème de Poisson] Soit une suite de probabilités $(p_n \;,\; n \in \mathbb{N})$, telle que $p_n \to 0$, et $np_n \to \lambda$, $(\lambda > 0)$ lorsque $n \to \infty$. Alors

$$b(n, p_n) \to P(\lambda)$$

i.e. pour tout $k = 0, 1, 2, \ldots$ nous avons : $b(n, p_n)(k) \to P(\lambda)(k)$, lorsque $n \to \infty$.

Démonstration. Pour $k = 0$ nous avons :

$$b(n, p_n)(0) = \binom{n}{0} p_n^0 (1-p_n)^{n-0} = (1-p_n)^n = (1 - \frac{\lambda}{n} + o(1/n))^n \to e^{-\lambda}$$



Loi de Poisson 3 — Fonction de répartition

FIGURE 3.3 – Loi de Poisson.

Formons le rapport

$$\frac{b(n,p_n)(k)}{b(n,p_n)(k-1)} = \frac{(n-k+1)p_n}{k(1-p_n)} \to \frac{\lambda}{k}, \quad n \to \infty$$

Par conséquent [1]

$$b(n,p_n)(1) \sim \frac{\lambda}{1} b(n,p_n)(0) \to \frac{\lambda}{1!}e^{-\lambda}$$
$$b(n,p_n)(2) \sim \frac{\lambda}{2} b(n,p_n)(1) \to \frac{\lambda^2}{2!}e^{-\lambda}$$

Et par récurrence, nous obtenons le résultat énoncé.

$$b(n,p_n)(k) \sim \frac{\lambda}{1} b(n,p_n)(k-1) \to \frac{\lambda^k}{k!}e^{-\lambda}.$$

□

Remarque 3.5. Cela définit la convergence en loi d'une suite de v.a. (Y_n) de lois (p_n) vers une v.a. Y de loi p, i.e., $p_n(k) \to p(k)$, $n \to \infty$, pour tout $k \in \mathbb{N}$. Ici $p_n = b(n,p_n)$ et $p = P(\lambda$. Nous allons présentée cette notion de convergence en loi, très importante dans les application, avec plus de précision, dans le chapitre 7.

3.4.6 Loi binomiale négative ou loi de Pascal

C'est la loi du nombre d'essais indépendants nécessaires pour obtenir m succès dans une suite de Bernoulli i.i.d.

On dit que X suit une loi binomiale négative de paramètre $(m,p) \in \mathbb{N}^* \times]0,1[$, et on note : $X \sim bn(m,p)$, si sa loi est donnée par

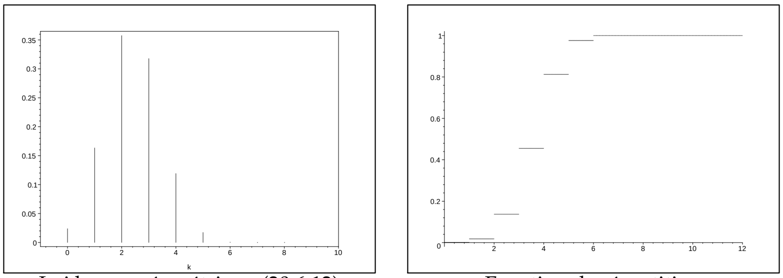
$$p(k) = \mathbb{P}(X=k) = \begin{cases} \binom{k-1}{m-1} p^m (1-p)^{k-m} & si \quad k \geq m \\ 0 & si \quad k < m \end{cases}$$

Nous avons : $\mathbb{E}X = m/p$ et $\mathrm{Var}(X) = m(1-p)/p^2$.

Remarque 3.6. Lorsque $m = 1$, on retrouve la loi géométrique $G(p)$.

Remarque 3.7. Une représentation intéressante de $X \sim bn(m,p)$ est la suivante : $X = Y_1 + \cdots + Y_m$ où Y_i i.i.d. $\sim G(p)$.

1. On note $u_n \sim v_n$ si $u_n/v_n \to 1$ lorsque $n \to \infty$.



Loi hyper-géométrique (20,6,12) Fonction de répartition

FIGURE 3.4 – Loi hyper-géométrique.

Remarque 3.8. Si $Z_1, \dots, Z_n$ sont n v.a. indépendantes et si $Z_i \sim bn(m_i, p)$ alors : $Z_1 + \cdots + Z_n \sim bn(\sum_{i=1}^n m_i, p)$.

3.4.7 Loi hypergéométrique

Dans une population de n éléments, dont n_1 rouges et $n_2 = n - n_1$ noirs, on choisit r éléments au hasard. Notons $p(k)$ la probabilité qu'exactement k éléments ainsi choisis soient rouges ($k = 0, 1, \dots, n_1 \wedge r$). Le groupe de r éléments contient k éléments rouges et $r - k$ éléments noirs. Les rouges peuvent être choisis de $\binom{n_1}{k}$ façons différentes, et les noirs de $\binom{n_2}{r-k}$ façons. Par conséquent, le nombre des choix possibles pour les rouges et les noirs est : $\binom{n_1}{r-k}\binom{n_2}{r-k}$. Le nombre de cas globalement possibles est : $\binom{n}{r}$. Soit X la v.a. désignant le nombre des éléments rouges parmi les éléments choisis.

Par conséquent, la loi hypergéométrique est

$$p(k) = \mathbb{P}(X = k) = \frac{\binom{n_1}{k}\binom{n_2}{r-k}}{\binom{n}{r}} \tag{3.18}$$

Les moments de cette loi sont : $\mathbb{E}X = rn_1/n$ et Var $X = r(n-r)n_1n_2/n^2(n-1)$.

3.5 Fonction génératrice

Pour une v.a. discrète X, à valeurs dans $\mathbb{N}$ et de loi $p = (p_i; i \in \mathbb{N})$, on définit la fonction génératrice g_X comme suit

$$g_X(s) = \mathbb{E}[s^X] = \sum_{i=0}^{\infty} p_i s^i \tag{3.19}$$

Cette série entière converge absolument au moins pour $-1 \le s \le 1$, car $g_X(1) = 1$.

Proposition 3.7. Nous avons : [2] La fonction g_X est analytique dans $[-1, +1]$ et définie de manière unique la loi de la v.a. X.

1. $\mathbb{E}[X] = g'_X(1)$.
2. $\mathrm{Var}(X) = g''_X(1) + g'_X(1) - [g'_X(1)]^2$.

Dans le cas d'une variance infinie $g''_X(s) \to \infty$ lorsque $s \to 1$.

2. Ici on fait usage du théorème suivant. Si la série entière $(a_n x^n)$ converge pour $|x| < r$, et elle a pour somme la fonction $S(x)$, alors la série entière $(na_n x^{n-1})$ converge également pour $|x| < r$ et elle est de somme $S'(x)$.

Lois	Fonctions génératrices
Bernoulli $B(p)$	$1-p+ps$
binomiale $b(n,p)$	$(1-p+ps)^n$
Poisson $P(\lambda)$	$e^{-\lambda(1-s)}$
géométrique $G(p)$	$\frac{ps}{1-(1-p)s}$
géométrique $G_0(p)$	$\frac{p}{1-(1-p)s}$
binomiale négative $bn(m,p)$	$(\frac{ps}{1-(1-p)s})^m$

Proposition 3.8 (Théorème de continuité des fonctions génératrices)**.** Soit une suite de lois de probabilités $(p_n(k), k \in \mathbb{N})$, $n \in \mathbb{N}$. Notons $g_n(s) := \sum_{k \geq 0} s^k p_n(k)$ la fonction génératrice de la loi p_n. Si $\lim_{n\to\infty} g_n(s) = g(s)$, pour tout $s \in]0,1[$, on a $\lim_{n\to\infty} p_n(k) = p(k)$ existe pour tout $k \geq 0$, et g est la fonction génératrice de la loi $p = (p(k), k \in \mathbb{N})$.

♣ **Exemple 3.6.** Le résultat de la proposition 3.6 peut être obtenu simplement en remarquant que la fonction génératrice de loi binomiale converge vers la fonction génératrice de la loi de Poisson. En effet,

$$(1-p_n+p_n s)^n = \Big(1 - \frac{\lambda(1-s)}{n}\Big)^n \to e^{-\lambda(1-s)}, \quad n \to \infty.$$

3.6 Convolution et somme de deux variables aléatoires

Considérons deux suites numériques $(a_n,\ n \geq 0)$ et $(b_n,\ n \geq 0)$, et définissons une troisième suite $(c_n,\ n \geq 0)$, dont le terme général est défini par

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \tag{3.20}$$

pour tout $n \in \mathbb{N}$.

Alors la suite (c_n) est appelée convolution de (a_n) et (b_n), et on note

$$(c_n) = (a_n) * (b_n)$$

Proposition 3.9. Soient deux v.a.d. X et Y, indépendantes, à valeurs dans $\mathbb{N}$ et de lois p_X et p_Y respectivement. Alors :

1. La loi de la somme $X+Y$ est donnée par la convolution des deux lois, i.e. $p_{X+Y} = p_X * p_Y$.
2. La fonction génératrice de la somme $X+Y$ est égale au produit des fonctions génératrices de X et Y.

Démonstration. (1) Nous avons :

$$\begin{aligned} p_{X+Y}(n) = \mathbb{P}(X+Y=n) &= \sum_{k=0}^{n} \mathbb{P}(X=k, Y=n-k) \\ &= \sum_{k=0}^{n} \mathbb{P}(X=k)\mathbb{P}(Y=n-k) \\ &= \sum_{k=0}^{n} p_X(k) p_Y(n-k) \\ &= (p_X * p_Y)(n) \end{aligned}$$

(2) Nous pouvons écrire :

$$g_{X+Y}(s) = \mathbb{E}[s^{X+Y}] = \mathbb{E}[s^X \cdot s^Y] = \mathbb{E}[s^X]\mathbb{E}[s^Y] = g_X(s) g_Y(s)$$

□

♣ **Exemple 3.7.** Soient deux v.a.d. $X \sim G(p)$ et $Y \sim G(p)$ indépendantes. Alors

$$\mathbb{P}(X+Y=n)=(n-1)p^2(1-p)^{n-2}, \quad n \geq 2.$$

Cette expression nous confirme que $X+Y \sim bn(2,p)$.

3.7 Entropie

L'entropie de Shannon joue un rôle important dans la théorie de l'information. Soient deux v.a.d. X et Y définies sur $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs dans $E=\{x_1,x_2,...,x_n\}$ et $F=\{y_1,y_2,...,y_m\}$ respectivement.

On appelle entropie (de Shannon) de la v.a.d. X (ou de sa loi P_X), et on note $H(X)$ (ou $H(P_X)$), le nombre

$$H(X)=-\sum_{i\geq 1}\mathbb{P}(X=x_i)\log_2 \mathbb{P}(X=x_i).$$

L'entropie de (X,Y) est définie par

$$H(X,Y)=-\sum_{i\geq 1}\sum_{j\geq 1}\mathbb{P}(X=x_i,Y=y_j)\log_2 \mathbb{P}(X=x_i,Y=y_j).$$

En théorie de l'information, si L est le nombre moyen de bits pour coder une information en tant que réalisation d'une v.a.d. X, elle vérifie les inégalités :

$$H(X)\leq L \leq H(X)+1.$$

Nous avons les propriétés suivantes.

Proposition 3.10. 1. Monter que $H(X)$ est maximum lorsque P_X est la loi uniforme sur E (incertitude maximale).

2. Si X et Y sont indépendantes, montrer que

$$H(X,Y)=H(X)+H(Y).$$

♣ **Exemple 3.8.** On jette un dé équilibré $n=10$ fois. Soit les v.a. $X_1,...,X_n$, telles que $X_i=1$ si le résultat du i-ème jeu est un 6, et $X_i=0$ sinon. On désire transmettre le résultat de cette expérience du point A au point B. Nous alons estimer le nombre moyen de bits nécessaire pour transmettre ce message.

En effet, on a $X_1,\ldots,X_{10}$ des v.a. indépendantes identiquement distribuées de loi de Bernouilli $\mathcal{B}(p)$, avec $p=1/6$. Pour $i=1,\ldots,10$, on a

$$\begin{aligned} H(X_i) &= -p\log_2 p-(1-p)\log_2(1-p)=\frac{1}{6}\log_2 6-\frac{5}{6}\log_2 5+\frac{5}{6}\log_2 6=\log_2 6-\frac{5}{6}\log_2 5.\\ &\approx 0.65. \end{aligned}$$

L'entropie recherchée est

$$H(X_1,\ldots,X_{10})=\sum_{i=1}^{10}H(X_i)\approx 6.5,$$

d'où le nombre de bit $L=7$.

Chapitre 4

Variable aléatoire réelle et variable aléatoire à densité

4.1 Introduction

La théorie de la v.a. discrète que nous venons d'exposer ne suffit pas à traiter rigoureusement les deux cas suivants :

– les suites infinies des v.a. telles que le jeu de pile ou face,
– le choix au hasard d'un point sur un segment.

En effet, dans ces deux cas l'espace fondamental n'est pas dénombrable.

Par ailleurs beaucoup de phénomènes physiques sont mesurés par de nombres réels.

4.2 Loi d'une variable aléatoire réelle

Soit une v.a. réelle (v.a.r.) X, i.e. $X : \Omega \to \mathbb{R}$. On appelle loi de la v.a.r. X, et on note P_X, l'image de $\mathbb{P}$ par X, i.e., pour tout intervalle B de $\mathbb{R}$, on a :

$$P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) \tag{4.1}$$

où $X^{-1}(B) = \{\omega : X(\omega) = x \in B\}$.

Remarque 4.1. La loi P_X est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, où $\mathcal{B}(\mathbb{R})$ est la σ-algèbre engendrée par les intervalles de $\mathbb{R}$, appelée *σ-algèbre borélienne*. C'est la plus petite σ-algèbre contenant les intervalles de $\mathbb{R}$. Une application mesurable définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite *borélienne*.

La loi d'une v.a.r. peut être précisée par la fonction de répartition (unique) ou par la fonction de densité de probabilité (non unique) lorsqu'elle existe que nous présentons dans ce qui suit.

4.3 Fonction de répartition

Elle est définie, pour tout $x \in \mathbb{R}$, par

$$F_X(x) = P_X(]-\infty, x]) = \mathbb{P}(X \leq x) \tag{4.2}$$

Par conséquent, pour tout intervalle $]a, b]$ de $\mathbb{R}$, nous avons :

$$\mathbb{P}(X \in]a, b]) = F_X(b) - F_X(a). \tag{4.3}$$

Elle vérifie les propriétés suivantes :

1. F_X est croissante sur $\mathbb{R}$;
2. $F_X(-\infty) = \lim_{x\to-\infty} F_X(x) = 0$, et $F_X(\infty) = \lim_{x\to\infty} F_X(x) = 1$;
3. $F_X(\cdot)$ est continue à droite, i.e. $F_X(x) = F_X(x+0)$ et possède une limite à gauche en tout point $x \in \mathbb{R}$, notée $F_X(x-)$, i.e., $F_X(x-) = \lim_{\varepsilon\downarrow 0} F_X(x-\varepsilon)$.

Démonstration. Les deux premières propriétés étant évidentes, nous allons démontrer la troisième. Soit une suite monotone de nombres réels positifs (u_n) telle que $u_n \downarrow 0$. Nous avons $F_X(x+u_n) - F_X(x) = P_X(]x, x+u_n])$. D'autre part, en posant $A_n =]x, x+u_n]$, nous avons, pour tous $m, n \in \mathbb{N}$ tels que $m > n$, $A_m \subset A_n$, et par le théorème des suites d'événements monotones, $\lim_{n\to\infty} P_X(A_n) = P_X(\lim_{n\to\infty} A_n) = P_X(\emptyset) = 0$. □

Remarque 4.2. Dans certains ouvrages la fonction de répartition est définie par $F_X(x) = \mathbb{P}(X < x)$, ce qui implique sa continuité à gauche.

Proposition 4.1. Soit une v.a.r. X de f.r. F, alors :

1. pour tout $x \in \mathbb{R}$ on a : $\mathbb{P}(X = x) = F(x) - F(x-)$;
2. si $x \in \mathbb{R}$ est un point de continuité de F, $\mathbb{P}(X = x) = 0$;
3. si $a, b \in \mathbb{R}$, $a \le b$ sont deux point de continuité de F,

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a < X \le b) = \mathbb{P}(a \le X \le b).$$

Démonstration. 1. On a $\{X = x\} = \cap_{n\ge 1}\{x - 1/n < X \le x + 1/n\}$, d'où

$$\mathbb{P}(X = x) = \lim_{n\to\infty}\{F(x+1/n) - F(x-1/n)\} = F(x) - F(x-).$$

2. On a $F(x) = F(x-)$ et de 1. on a le résultat.
3. C'est une conséquence immédiate de (1) car

$$\mathbb{P}(a \le X \le b) = \mathbb{P}(a < X \le b) + \mathbb{P}(X = a) = \mathbb{P}(a < X \le b),$$

etc. □

La f.r. inverse (généralisée), F^{-1}, est définie comme suit

$$F^{-1}(y) = \inf\{x : F(x) > y\}, \quad 0 < y < 1$$

Aux points extrêmes des valeurs de y, on peut définir F^{-1} par $F^{-1}(1) = +\infty$, si $F(x) < 1$ pour tout x, et pour tout F, $F^{-1}(0) = -\infty$.

♣ **Exemple 4.1.** Soit une fonction $F : \mathbb{R} \to [0, 1]$, définie par :

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ \frac{1+x}{5}, & \text{si } 0 \le x < 1 \\ \frac{3+x}{10}, & \text{si } 1 \le x < 2 \\ \frac{3x}{10}, & \text{si } 2 \le x < 3 \\ 1, & \text{si } 3 \le x \end{cases}$$

La représentation graphique de la fonction F est donnée dans la fig. 4.1.

On peut observer que F est une fonction :
- croissante ;
- continue à droite ;
- $F(-\infty) = 0$ et $F(+\infty) = 1$.

et donc F vérifie les propriétés d'une fonction de répartition et en conséquence, il existe une et une seule probabilité Q dont la fonction de répartition est F (i.e. $Q(]-\infty, x]) = F(x),\ x \in \mathbb{R}$).

L'ensemble des atomes ponctuels de Q est $A = \{0, 2, 3\}$, identifiés par les sauts de F :

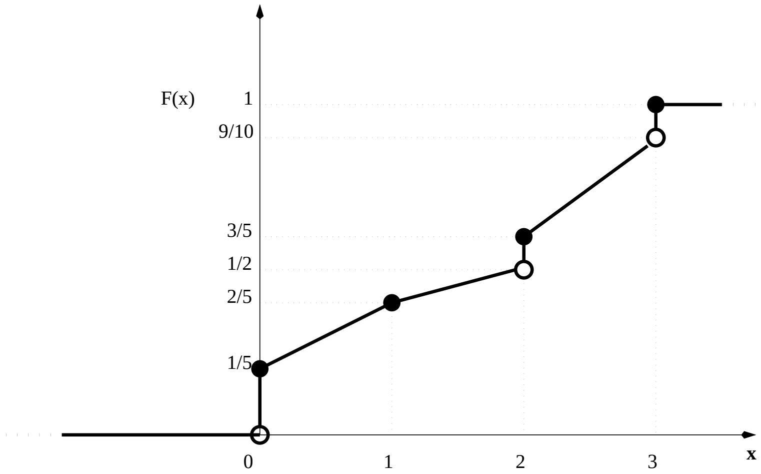


FIGURE 4.1 – La représentation graphique de la fonction F.

- $Q(\{0\}) = \frac{1}{5} > 0$;
- $Q(\{2\}) = \frac{3}{5} - \frac{1}{2} = \frac{1}{10} > 0$;
- $Q(\{3\}) = 1 - \frac{9}{10} = \frac{1}{10} > 0$;

Proposition 4.2. Soit $F : \mathbb{R} \to [0,1]$ une fonction croissante continue à droite ayant de limites à gauche et telle que $F(-\infty) = 0$, $F(+\infty) = 1$. Alors il existe une probabilité unique sur $(\mathbb{R}, \mathcal{B})$ telle que

$$P(]a,b]) = F(b) - F(a),$$

pour $\infty \leq a \leq b \leq \infty$.

4.4 Variable aléatoire absolument continue

Définition 4.1. On dit qu'une v.a.r. X (ou de manière équivalente, sa loi P_X) est absolument continue, s'il existe une fonction $f_X : \mathbb{R} \longrightarrow \mathbb{R}_+$ telle que la loi P_X soit donnée par la formule suivante

$$P_X(B) = \int_B f_X(x)dx \tag{4.4}$$

pour tout intervalle B de $\mathbb{R}$.

Remarque 4.3. Pour plus de précision, il faut souligner que f_X est une application mesurable de $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ dans $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$. La densité d'une v.a. est définie p.s., donc il peut y avoir plusieurs fonctions qui peuvent être la densité d'une même v.a..

La fonction f_X s'appelle la *densité de probabilité* de P_X ou de la v.a. X. Elle vérifie les propriétés suivantes :

1. Elle est positive, i.e. $f_X(x) \geq 0$ pour tout $x \in \mathbb{R}$;
2. $\int_{\mathbb{R}} f_X(x)dx = 1$.

Remarque 4.4. Il faut noter que f_X n'est pas une probabilité. En fait, elle peut prendre des valeurs bien supérieures à 1.

Une interprétation de la densité est la suivante [1]

$$\mathbb{P}(x < X \leq x + \Delta x) = f_X(x)\Delta x + o(\Delta x), \quad \Delta x \to 0, \tag{4.5}$$

1. $o(\cdot)$ désigne une fonction telle que $o(h)/h \to 0$ lorsque $h \to 0$.

ou

$$f_X(x) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \mathbb{P}(x < X \leq x + \Delta x) \tag{4.6}$$

Remarque 4.5. Nous verrons qu'à part les v.a. discrètes, il existe d'autres v.a. qui ne sont pas absolument continues par rapport à la mesure de longueur (de Lebesgue).

Si la v.a. X est absolument continue, alors

$$F_X(x) = \int_{]-\infty,x]} f_X(u)du \tag{4.7}$$

Si x est un point de continuité de f_X, alors nous avons :

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{4.8}$$

♣ **Exemple 4.2.** Loi uniforme sur $[0, 1]$.

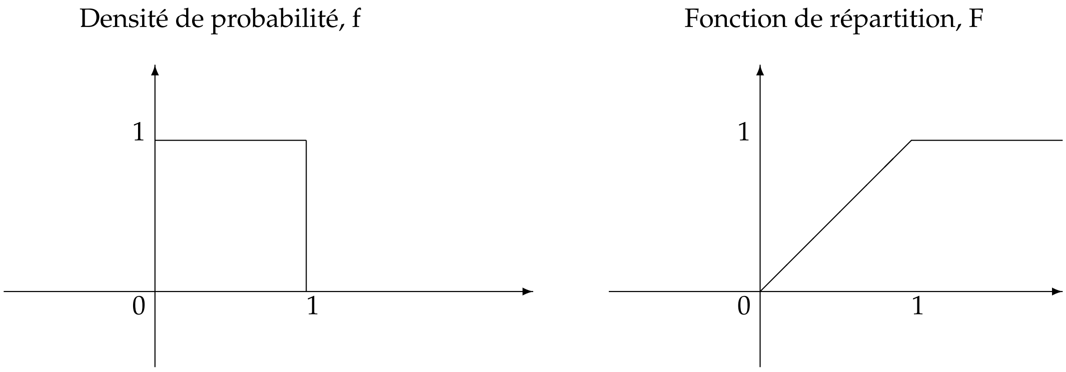


FIGURE 4.2 – La loi uniforme sur $[0, 1]$

La densité de cette loi est : $f(x) = \mathbf{1}_{[0,1]}(x)$. A l'aide de la relation (4.7), la fonction de répartition s'écrit comme suit

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \quad 0 \leq x \leq 1 \\ 1 & si \quad x > 1 \end{cases}$$

Notons aussi que, par exemple, les fonctions : $f_1(x) = \mathbf{1}_{]0,1[}(x)$, $f_2(x) = \mathbf{1}_{]0,1]}(x)$, $f_3(x) = \mathbf{1}_{[0,1[}(x)$, etc., sont aussi de densités de la même f.r. F.

4.5 Intégrale de Riemann-Stieltjes

L'intégrale de Riemann-Stieltjes est une généralisation de l'intégrale de Riemann [2]. Considérons une fonction α croissante définie sur $[a, b] \subset \mathbb{R}$ et f une fonction définie sur le même intervalle. Soit une partition de l'intervalle [a,b], $\mathcal{P}_n = \{a = x_0 \leq x_1 \leq \cdots \leq x_{n-1} \leq x_n = b\}$.

Soient les deux sommes suivantes :

$$S(\mathcal{P}_n) = \sum_{i=1}^{n} M_i \Delta\alpha_i, \; (M_i = \sup\{f(x), x \in [x_{i-1}, x_i]\}) \tag{4.9}$$

2. Cf. Cours MT22

$$s(\mathcal{P}_n) = \sum_{i=1}^{n} m_i \Delta\alpha_i, \ \ (m_i = \inf\{f(x), x \in [x_{i-1}, x_i]\}) \tag{4.10}$$

avec $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$

Si, lorsque la flèche de la partition $\max_i(x_i - x_{i-1})$ tend vers 0, lorsque n tend vers l'infini, les deux sommes ci-dessus tendent vers le même nombre I, alors on note

$$I = \int_{[a,b]} f d\alpha, \tag{4.11}$$

que l'on appelle *intégrale de Riemann-Stieltjes*. On dit que f est α-intégrable sur $[a, b]$.

L'intégrale de Riemann-Stieltjes possède les propriétés habituelles (linéarité, additivité, conservation des inégalités, etc.) de l'intégrale. Si $\alpha(x) = x$ pour tout $x \in [a, b]$, nous obtenons l'intégrale de Riemann. Si α possède une dérivée continue α', alors nous avons

$$\int_{[a,b]} f d\alpha = \int_{[a,b]} f(x)\alpha'(x) dx \tag{4.12}$$

L'intégrale de Riemann-Stieltjes suivante, lorsqu'elle existe,

$$\int_{-\infty}^{+\infty} f d\alpha \tag{4.13}$$

est définie par

$$\lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f d\alpha \tag{4.14}$$

La proposition suivante est très importante pour le calcul des espérances.

Proposition 4.3. Soient g une fonction bornée sur $[a, b]$, ξ un point de $]a, b[$ et $\alpha(x) = 1(x - \xi)$. Si g est continue au point ξ, alors

$$\int_a^b g \, d\alpha = g(\xi) \tag{4.15}$$

Proposition 4.4. Soient α une fonction monotone croissante sur $[a, b]$, f_n une suite de fonctions Riemann-Stieltjes intégrables sur $[a, b]$ pour tous $n = 1, 2, ...$ telles que $f_n \to f$ uniformément sur $[a, b]$. Alors f est Riemann-Stieltjes intégrable sur $[a, b]$ et

$$\int_a^b f_n d\alpha \to \int_a^b f d\alpha, \ \ \textit{lorsque} \ \ n \to \infty.$$

4.6 Espérance mathématique

Si F est la fonction de répartition de la v.a.r. X, alors l'espérance de X est définie par l'intégrale de Riemann-Stieltjes suivante

$$\mathbb{E}X = \int_{\mathbb{R}} x dF(x) \tag{4.16}$$

Proposition 4.5. 1. Dans le cas d'une v.a. à densité, l'intégrale ci-dessus s'écrit

$$\mathbb{E}X = \int_{\mathbb{R}} x f(x) dx \tag{4.17}$$

2. Dans le cas d'une v.a. discrète, nous obtenons la formule déjà connue

$$\mathbb{E}X = \sum_{x} xp(x) \tag{4.18}$$

3. Dans le cas d'une v.a. mixte, de f.r. F de points de discontinuité $x_1, x_2, ..., x_r$ de sauts $p_1, p_2, ..., p_r$ respectivement, nous avons

$$\mathbb{E}X = \sum_{i=1}^{r} x_i p_i + \sum_{i=0}^{r} \int_{x_i}^{x_{i+1}} xF'_{ac}(x)dx \tag{4.19}$$

avec $x_0 = -\infty$, $x_{r+1} = \infty$, et où $p(x)$ est la loi de X.

Pour que l'espérance d'une v.a. à densité existe, il faut et il suffit que $\int_{\mathbb{R}} |x|f(x)dx < \infty$.

♣ **Exemple 4.3.** Espérance d'une v.a. uniforme sur $[0, 1]$.
On a : $f(x) = \mathbf{1}_{[0,1]}(x)$ et

$$\mathbb{E}X = \int_{\mathbb{R}} x\mathbf{1}_{[0,1]}(x)dx = \int_0^1 x\,dx = \frac{1}{2}$$

Soit X une v.a.r. et φ une application continue de $\mathbb{R}$ dans $\mathbb{R}$, alors $\varphi \circ X$ (i.e. pour tout $\omega \in \Omega$, $(\varphi \circ X)(\omega) = \varphi(X(\omega))$) est une v.a., et nous avons

$$\mathbb{E}[\varphi \circ X] = \int_{\mathbb{R}} \varphi(x)dF(x) \tag{4.20}$$

♣ **Exemple 4.4.** Soit $X \sim U[0, 1]$ (loi uniforme sur $[0, 1]$) et $\varphi(x) = x^2, (x \in \mathbb{R})$. Alors

$$\mathbb{E}[\varphi \circ X] = \int_{\mathbb{R}} x^2\mathbf{1}_{[0,1]}(x)dx = \int_0^1 x^2dx = \frac{1}{3}.$$

♣ **Exemple 4.5.** Calcul de l'espérance d'une v.a. X de f.r. donnée dans l'exemple 4.1.

$$\begin{aligned}\mathbb{E}(X) \quad &= \quad \int_{-\infty}^{0} x \cdot 0dx + \int_0^1 x \cdot \frac{1}{5}dx + \int_1^2 x \cdot \frac{1}{10}dx + \int_2^3 x \cdot \frac{3}{10}dx \\ &\quad + \int_3^{+\infty} x \cdot 0dx + 0 \cdot \frac{1}{5} + 2 \cdot \frac{1}{10} + 3 \cdot \frac{1}{10} = \frac{3}{2}.\end{aligned}$$

Proposition 4.6 (Théorème fondamental)**.**
Pour que X ait pour densité de probabilité la fonction f, il faut et il suffit que, pour toute application continue et bornée φ

$$\mathbb{E}[\varphi \circ X] = \int_{\mathbb{R}} \varphi(x)f(x)dx$$

4.7 Variance et moments

La variance (ou dispersion) est définie, comme dans le cas de la v.a. discrète, par

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2] = \int_{\mathbb{R}} (x - \mathbb{E}X)^2 dF(x) \tag{4.21}$$

La racine carrée de la variance, i.e., $\sqrt{\mathrm{Var}(X)}$ est appelée *écart-type* de X.
De cette définition, nous déduisons la relation suivante qui est beaucoup plus utile en pratique

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 \tag{4.22}$$

Nous en déduisons que $\mathrm{Var}(X) < \infty$ si, et seulement si, $\mathbb{E}[X^2] < \infty$. Les v.a. possédant des variances sont appelées *v.a. du deuxième ordre*.

♣ **Exemple 4.6.** Variance de $X \sim U[0,1]$. On a déjà calculé : $\mathbb{E}X = 1/2$ et $\mathbb{E}[X^2] = 1/3$, donc $\mathrm{Var}(X) = 1/12$.

Quelques propriétés de la variance :

1. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, $(a, b \in \mathbb{R})$;
2. si X et Y sont indépendantes [3] alors $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$;
3. si $a \leq X \leq b$, alors $\mathrm{Var}(X) \leq \frac{(b-a)^2}{4}$.

Proposition 4.7. L'espérance d'une v.a.r. X réalise le minimum de la fonction : $x \mapsto \mathbb{E}(X - x)^2$, c'est-à-dire que $\mathrm{Var}(X) = \min_{x\in\mathbb{R}} \mathbb{E}(X - x)^2$.

Démonstration. De l'identité $(X - x)^2 - (\mathbb{E}X - x)^2 = (X - \mathbb{E}X)^2 + 2(\mathbb{E}X - x) \cdot (X - \mathbb{E}X)$, on déduit $\mathbb{E}(X - x)^2 = (\mathbb{E}X - x)^2 + \mathrm{Var}(X) \geq 0$, d'où le résultat. □

Les moments sont définis de la même façon que pour les v.a. discrètes :
Le moment d'ordre k, lorsque $|x|^k f(x)$ est sommable, est donné par

$$\mu_k = \mathbb{E}[X^k] = \int_{\mathbb{R}} x^k dF(x) = \int_{\mathbb{R}} x^k f(x)dx \tag{4.23}$$

Le moment centré d'ordre k est

$$m_k = \mathbb{E}[(X - \mathbb{E}X)^k] = \int_{\mathbb{R}} (x - \mathbb{E}X)^k dF(x) = \int_{\mathbb{R}} (x - \mathbb{E}X)^k f(x)dx \tag{4.24}$$

Proposition 4.8. Si la v.a. X admet un moment d'ordre n, $(n > 1)$, alors elle admet également tous les moments d'ordre $\leq n$.

Démonstration. Ce résultat est une conséquence directe de l'inégalité : $|X|^{n-1} \leq |X|^n + 1$.

♣ **Exemple 4.7.** Calcul de la variance d'une v.a. X de f.r. donnée dans l'exemple 4.1. Pour le moment d'ordre 2, on a :

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{0} x^2 \cdot 0dx + \int_0^1 x^2 \cdot \frac{1}{5}dx + \int_1^2 x^2 \cdot \frac{1}{10}dx + \int_2^3 x^2 \cdot \frac{3}{10}dx \\ &\quad + \int_3^{+\infty} x^2 \cdot 0dx + 0^2 \cdot \frac{1}{5} + 2^2 \cdot \frac{1}{10} + 3^2 \cdot \frac{1}{10} = \frac{7}{2}.\end{aligned}$$

La variance de X s'écrit comme suit :

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{7}{2} - \left(\frac{3}{2}\right)^2 = \frac{5}{4}.$$

4.8 Médiane

Définition 4.2. On appelle médiane d'une v.a.r. X un nombre M réel tel que

$$\mathbb{P}(X \leq M) \geq 1/2 \ \ et \ \ \mathbb{P}(X \geq M) \geq 1/2.$$

Toute v.a.r. possède une médiane (au moins).

Une v.a.r. peut avoir plusieurs médianes. Si la f.r. F de X est continue et strictement croissante, alors X admet une médiane unique M, et on a $F(M) = 1/2$.

Proposition 4.9. -Définition. Si $\mathbb{E} \mid X \mid < \infty$, alors l'expression $\mathbb{E} \mid X - M \mid$ prend la même valeur pour toute médiane M, et elle est appelée écart moyen minimum ou écart médian de X.

3. On verra dans le ch.5 qu'il suffit que X et Y soient non corrélées

4.9 Inégalités de Markov et de Bienaymé-Tchebichev

Ces inégalités sont d'une grande importance en probabilités tant au point de vue théorique que pratique.

Proposition 4.10. 1. *Inégalité de Markov*. Soit une v.a.r. X possédant le moment d'ordre $p \geq 1$. Alors pour tout $\varepsilon > 0$, nous avons :

$$P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}(|X|^p)}{\varepsilon^p}. \tag{4.25}$$

2. *Inégalité de Bienaymé-Tchebichev*. Soit une v.a.r. X de carré intégrable. Alors pour tout $\varepsilon > 0$, nous avons :

$$P(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}. \tag{4.26}$$

Démonstration. Pour toute v.a.r. Y, possédant le moment d'ordre $p \geq 1$, et tout $\varepsilon > 0$, nous avons clairement

$$\mathbf{1}_{\{|Y| \geq \varepsilon\}} \leq \frac{|Y|^p}{\varepsilon^p}.$$

En intégrant les deux membres de cette inégalité, on obtient :

$$\mathbb{P}(|Y| \geq \varepsilon) \leq \frac{\mathbb{E}\,|Y|^p}{\varepsilon^p}.$$

En posant $Y = X - \mathbb{E}X$ et $p = 2$ dans cette dernière inégalité, on obtient la seconde inégalité. □

Interprétation. Pour $\epsilon = k\sigma$, (σ est l'écart-type de la v.a. X), en appliquant l'inégalité de Bienaymé-Tchebichev, on a

$$\mathbb{P}(\mid X - \mu \mid \geq k\sigma) \leq \frac{1}{k^2} \tag{4.27}$$

Il y a donc une probabilité inférieure à $1/4$ pour que X s'écarte de 2σ de sa moyenne ou une probabilité inférieure à $1/9$ pour que X s'écarte de 3σ de sa moyenne, etc.

4.10 Quelques lois particulières

4.10.1 Loi uniforme

Une v.a. X suit une loi uniforme sur $[a, b] \subset \mathbb{R}$, et on note $X \sim U[a, b]$, si sa densité est donnée par

$$f(x) = \begin{cases} \frac{1}{b-a} & si \quad x \in [a, b] \\ 0 & sinon \end{cases}$$

La fonction de répartition est

$$F(x) = \begin{cases} 0 & si \quad x \leq a \\ \frac{x-a}{b-a} & si \quad a \leq x \leq b \\ 1 & si \quad x \geq b \end{cases}$$

L'espérance et la variance sont : $\mathbb{E}X = (a + b)/2$ et Var $X = (b - a)^2/12$.

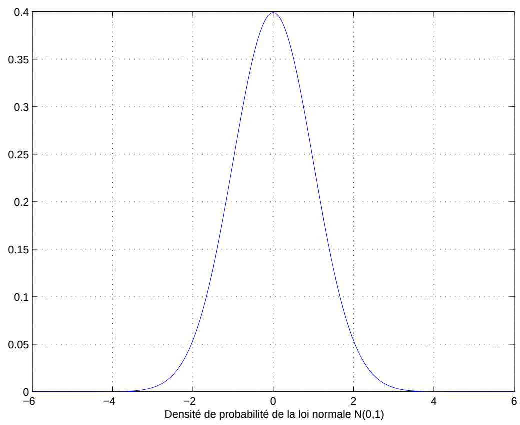


FIGURE 4.3 – Densité de la loi normale $\mathcal{N}(0,1)$.-

4.10.2 Loi normale ou loi de Gauss

a) *Loi normale centrée réduite.*
On dit qu'une v.a. Z suit une loi normale centrée réduite et on note : $Z \sim N(0,1)$, si sa densité de probabilité est donnée, pour tout $z \in \mathbb{R}$, par

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}). \tag{4.28}$$

Sa fonction de répartition est

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du. \tag{4.29}$$

L'espérance et la variance sont : $\mathbb{E}Z = 0$ et $\mathrm{Var}(Z) = 1$.
b)*Loi normale générale.*
Soit $Z \sim N(0,1)$, $(\mu,\sigma) \in \mathbb{R} \times \mathbb{R}_+$ et $X = \sigma Z + \mu$. Alors $\mathbb{E}X = \mu$ et $\mathrm{Var}\, X = \sigma^2$, et

$$\mathbb{P}(X \le x) = \mathbb{P}(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}) \tag{4.30}$$

$$= \mathbb{P}(Z \le \frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{u^2}{2}} du. \tag{4.31}$$

Par le changement de variable : $u = \frac{z-\mu}{\sigma}$, on obtient

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du, \tag{4.32}$$

et la densité est

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{4.33}$$

On note alors : $X \sim N(\mu, \sigma^2)$.

4.10.3 Loi exponentielle

La v.a. X suit une loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$, et on note $X \sim E(\lambda)$, si sa densité est donnée par

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x). \tag{4.34}$$

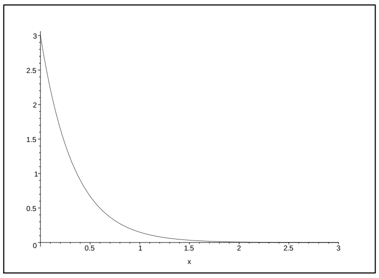


FIGURE 4.4 – Loi exponentielle de paramètre 3.

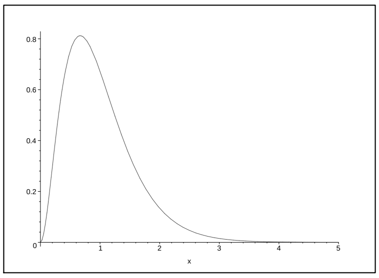


FIGURE 4.5 – Loi d'Erlang de paramètres 3 et 3.

Fonction de répartition

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0. \end{cases}$$

L'espérance et la variance de X sont : $\mathbb{E}X = 1/\lambda$ et Var $X = 1/\lambda^2$.

La propriété : $\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t)$ met en évidence l'absence de mémoire de la loi exponentielle et joue un rôle très important en probabilité car elle est liée aux systèmes aléatoires markoviens.

4.10.4 Loi gamma

La v.a. X suit une loi gamma de paramètre $(\alpha, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, et on note $X \sim \gamma(\alpha, \lambda)$, si sa densité est

$$f(x) = \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x) \tag{4.35}$$

où $\Gamma(\cdot)$ est la fonction eulérienne gamma définie par la formule suivante

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \qquad (a \in \mathbb{C}). \tag{4.36}$$

Remarque 4.6. Lorsque $a \in \mathbb{N}^*$, $\Gamma(a) = (a-1)!$.

L'espérance et la variance de X sont : $\mathbb{E}X = \alpha/\lambda$ et Var $X = \alpha/\lambda^2$.

Si $\alpha \in \mathbb{N}^*$, alors $\gamma(\alpha, \lambda)$ est appelée *loi d'Erlang*. Pour $\alpha = 1$, nous retrouvons la loi exponentielle de paramètre λ.

4.10.5 Loi du chi-deux

La v.a. X suit une loi du chi-deux, à n degrés de libertés, $(n \in \mathbb{N}^*)$, et on note $X \sim \chi^2(n)$, si $X \sim \gamma(n/2, 1/2)$.

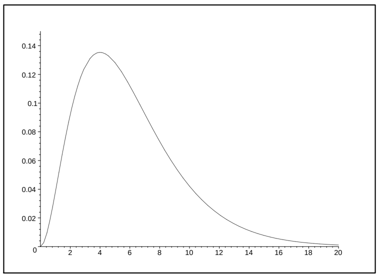


FIGURE 4.6 – Loi de chi-deux à 3 degrés de liberté.

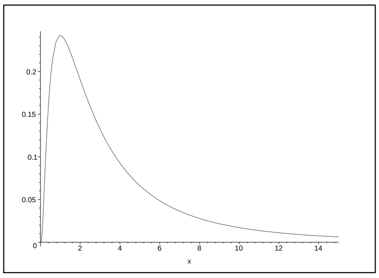


FIGURE 4.7 – Loi log-normale de paramètres 1 et 1.

Une v.a. $X \sim \chi^2(n)$ est la somme des carrés de n v.a. normales centrées réduites et indépendantes, i.e., si $X \sim \chi^2(n)$, on a

$$X = Z_1^2 + \cdots + Z_n^2 \qquad Z_i \quad i.i.d. \sim N(0,1)$$

L'espérance et la variance sont : $\mathbb{E}X = n$ et Var $X = 2n$.

4.10.6 Loi log-normale

La v.a. X suit une loi log-normale de paramètre $(\mu, \sigma^2) \in \mathbb{R}_+ \times \mathbb{R}_+$, notée $X \sim LN(\mu, \sigma^2)$, si $X = e^Y$ et $Y \sim N(\mu, \sigma^2)$, et sa densité est

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \mathbf{1}_{\mathbb{R}_+}(x) \tag{4.37}$$

L'espérance et la variance sont : $\mathbb{E}X = e^{\mu+\sigma^2/2}$ et Var $X = e^{2(\mu+\sigma^2/2)} - e^{2\mu+\sigma^2}$.

4.10.7 Loi de Weibull

La v.a. X suit une loi de Weibull de paramètre $(\beta, \eta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$, et on note $X \sim W(\beta, \eta)$, si sa densité est

$$f(x) = \frac{\beta}{\eta}(\frac{x}{\eta})^{\beta-1} e^{-(\frac{x}{\eta})^\beta} \mathbf{1}_{\mathbb{R}_+}(x) \tag{4.38}$$

La fonction de répartition est

$$F(x) = \begin{cases} 1 - e^{-(\frac{x}{\eta})^\beta} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

L'espérance et la variance sont : $\mathbb{E}X = \eta\Gamma(1 + \frac{1}{\beta})$ et Var $X = \eta^2[\Gamma(1 + \frac{2}{\beta}) - (\Gamma(1 + \frac{1}{\beta}))^2]$.

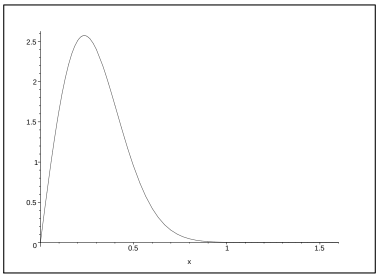


FIGURE 4.8 – Loi de Weibull de paramètres 2 et 3.

4.10.8 Loi de Cauchy

Soit $a \in \mathbb{R}_+^*$. On dit que la v.a.r. X suit une loi de Cauchy de paramètre a, et on note $X \sim C(a)$, si sa densité est

$$f(x) = \frac{a}{\pi}\frac{1}{x^2+a^2}, \quad x \in \mathbb{R}.$$

4.11 Fonction génératrice des moments

La fonction génératrice des moments d'une v.a. réelle X, notée $M(t)$, est définie par la relation suivante

$$M(t) = \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} dF(x) = \begin{cases} \sum_x e^{tx} p(x) & si\ X\ discrte \\ \int_{\mathbb{R}} e^{tx} f(x) dx & si\ X\ à\ densit \end{cases} \tag{4.39}$$

pour tout $t \in \mathbb{R}$ tel que $M(t) < \infty$. Elle est définie dans un certain intervalle contenant l'origine. Son nom est dû à la formule suivante

$$\mathbb{E}[X^n] = M^{(n)}(0) = \frac{d^n}{dt^n} M(t)|_{t=0}, \tag{4.40}$$

que nous pouvons vérifier facilement.

♣ **Exemple 4.8.** Si $X \sim b(n,p)$, alors :

$$M(t) = \mathbb{E}[e^{tX}] = \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = (1-p+pe^t)^n \tag{4.41}$$

Par la formule (4.40), on a :

$$\mathbb{E}X = M'(t)|_{t=0} = npe^t(1-p+pe^t)^{n-1}|_{t=0} = np.$$

Lois	Fonctions génératrices des moments
Dirac δ_0	1
binomiale $b(n,p)$	$(1-p+p\,e^t)^n$
Poisson $P(\lambda)$	$e^{-\lambda+\lambda\,e^t}$
géométrique $G(p)$	$\frac{pe^t}{1-(1-p)e^t}$
binomiale négative $bn(m,p)$	$(\frac{pe^t}{1-(1-p)e^t})^m$
uniforme $U(a,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
normale $N(\mu,\sigma^2)$	$e^{\mu t+\frac{1}{2}\sigma^2 t^2}$
exponentielle $E(\lambda)$	$\frac{\lambda}{\lambda-t}$
gamma $\gamma(\alpha,\lambda)$	$(\frac{\lambda}{\lambda-t})^\alpha$

Proposition 4.11. Si $M_X(t)$ est définie pour tout t dans $[-t_0, t_0]$, $(t_0 > 0)$, alors :

1. X possède des moments finis de tous ordres ;
2. $M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k]$.

Proposition 4.12. Soient deux v.a.r. X et Y. S'il existe une constante $h > 0$ telle que $M_X(t) = M_Y(t)$ pour tout $|t| < h$, alors $F_X \equiv F_Y$.

Problème de moments : Sous quelles conditions une distribution de probabilité est-elle définie de manière unique par ses moments ? Il n'existe pas de condition nécessaire et suffisante. En revanche, nous avons différentes conditions suffisantes dont la plus simple exige que la fonction génératrice des moments soit finie dans un voisinage de 0.

Transformée de Laplace. Pour une v.a. X positive, on définit sa transformée de Laplace comme suit :

$$\Psi_X(s) := \mathbb{E}[e^{-sX}] = \int_0^{\infty} e^{-sx} dF(x),$$

pour $t \geq 0$. On a $\Psi_X(t) = M_X(-t)$. Ses propriétés sont analogues à celles de la fonction génératrice des moments.

4.12 Décomposition des lois

A part les v.a. discrètes et à densité, il existe un troisième type de v.a. appelées *singulières* et qui n'ont pas de densité bien qu'elles puissent être continues. Un exposé sur ce type des v.a. sort du cadre de ce cours car il ferait appel à des outils mathématiques avancés et qui ne sont pas enseignés dans les cours de Mathématiques en TC.

Néanmoins, nous allons énoncer la proposition suivante, combinant le théorème de décomposition de Jordan et celui de Lebesgue, qui nous permettra de faire face à certains problèmes pratiques.

Proposition 4.13 (Théorème de décomposition de Jordan-Lebesgue)**.** .
Toute distribution de probabilité F admet la représentation

$$F = pF_a + qF_s + rF_{ac} \tag{4.42}$$

où F_a est une distribution atomique (discrète), F_s est une distribution singulière, F_{ac} est une distribution absolument continue, avec $p, q, r \in [0, 1]$ et $p + q + r = 1$.

♣ **Exemple 4.9.** Décomposition de la fonction de répartition de l'exemple 4.3. Puisque on a déjà trouvé la partie discrète représentée par ses atomes, il suffit de calculer la partie continue de F, qui est $(c = 6/10)$:

$$cF_c(x) = \begin{cases} 0, & \text{si } x < 0 \\ \frac{1+x}{5} - \frac{1}{5} = \frac{x}{5}, & \text{si } 0 \leq x < 1 \\ \frac{3+x}{10} - \frac{1}{5} = \frac{x+1}{10} = \frac{x+1}{10}, & \text{si } 1 \leq x < 2 \\ \frac{3x}{10} - (\frac{1}{5} + \frac{1}{10}) = \frac{3(x-1)}{10}, & \text{si } 2 \leq x < 3 \\ 1 - (\frac{1}{5} + \frac{1}{10} + \frac{1}{10}) = \frac{3}{5}, & \text{si } 3 \leq x \end{cases}$$

4.13 Davantage d'inégalités

A part l'inégalité de Bienaymé-Tchebichev que nous avons déjà présentée, il y a beaucoup d'autres inégalités dont l'utilité pratique ou théorique est variable. Dans ce paragraphe, nous allons présenter quatre inégalités : celles de Cauchy-Schwarz, de Hölder, de Jensen et de Kolmogorov, très importantes en probabilités.

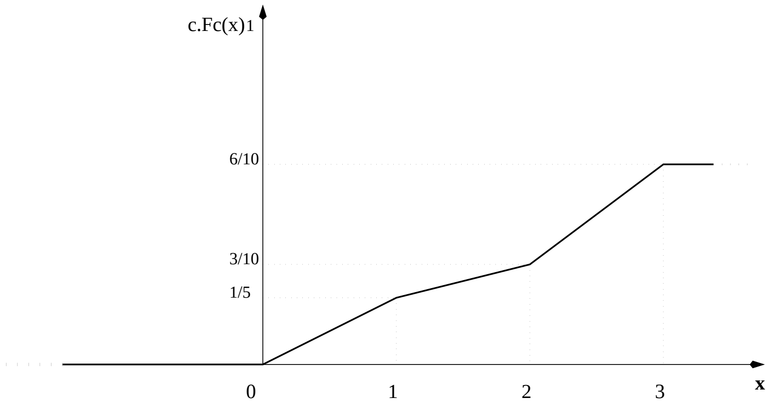


FIGURE 4.9 – La représentation graphique de la partie continue de F.

4.13.1 Inégalité de Cauchy-Schwarz

Proposition 4.14. Si les v.a.r. X et Y possèdent des moments d'ordre 2, alors :

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2) \tag{4.43}$$

avec égalité ssi $Y = cX$ avec probabilité 1.

Démonstration. Nous avons, pour tout $t \in \mathbb{R}$, $0 \leq \mathbb{E}[(X+tY)^2] = \mathbb{E}X^2 + 2t\mathbb{E}XY + t^2\mathbb{E}Y^2$, par conséquent : $\Delta = 4(\mathbb{E}XY)^2 - 4\mathbb{E}X^2\mathbb{E}Y^2 < 0$. □

4.13.2 Inégalité de Jensen

Soit I un intervalle de $\mathbb{R}$. Une fonction $\varphi : I \longrightarrow \mathbb{R}$ est dite *convexe*, si pour tout $x, y \in I$ et $0 \leq t \leq 1$, nous avons

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y). \tag{4.44}$$

Une propriété intéressante d'une telle fonction est que pour tout point $(\xi, \varphi(\xi))$ de son graphe $(\xi \in I)$, il existe une droite (ℓ), passant par ce point et laissant le graphe de φ dans le demi plan défini par cette droite, i.e., pour tout $x \in I$,

$$\varphi(x) \geq \varphi(\xi) + \lambda(x - \xi) \tag{4.45}$$

où λ est le coefficient directeur de la droite (ℓ).

Pour les fonctions convexes, nous avons en probabilités l'inégalité suivante.

Proposition 4.15. Soient une v.a. X et une fonction convexe φ sur l'ensemble de valeurs de X. Si X et $\varphi(X)$ possèdent des espérances finies, alors

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]. \tag{4.46}$$

Démonstration. Dans l'inégalité (4.45) ci-dessus, en posant $\xi = \mathbb{E}X$ et X au lieu de x et en prenant l'espérance de deux membres nous obtenons le résultat énoncé. □

Remarque 4.7. Si φ est concave, alors l'inégalité (4.46) est inversée.

♣ **Exemple 4.10.** L'application $x \mapsto x^2$ étant convexe, nous avons : $(\mathbb{E}X)^2 \leq \mathbb{E}(X^2)$.

♣ **Exemple 4.11.** L'application $x \mapsto |x|$ étant convexe, nous avons : $|\mathbb{E}X| \leq \mathbb{E}|X|$.

4.13.3 Inégalité de Hölder

Proposition 4.16. Soient deux v.a. X et Y. Supposons que X et Y possèdent des moments d'ordres p et q respectivement, où p et q sont des nombres positifs tels que : $\frac{1}{p} + \frac{1}{q} = 1$. Alors XY est intégrable et nous avons :

$$\mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q}. \tag{4.47}$$

L'égalité est vérifiée ssi $|Y|^q = c|X|^p$.

4.13.4 Inégalité de Kolmogorov

Proposition 4.17. Soient n v.a. $X_1, X_2, ..., X_n$ indépendantes, centrées ayant des variances finies, σ_i^2. Pour tout $\varepsilon > 0$, nous avons

$$\mathbb{P}(\max_{1\leq i\leq n} |X_1 + ... + X_i| > \varepsilon) \leq \sum_{i=1}^{n} \frac{\sigma_i^2}{\varepsilon^2} \tag{4.48}$$

4.14 Fonction du taux de hasard

Pour une v.a.r. X positive absolument continue de f.r. F et de densité f, nous pouvons définir la fonction de hasard $\lambda : \mathbb{R}_+ \longrightarrow \bar{\mathbb{R}}_+$, par la formule suivante

$$\lambda(x) = \frac{f(x)}{1 - F(x)} \tag{4.49}$$

Lorsque X mesure la durée de vie d'un système ou d'un individu, la fonction de hasard s'appelle fonction du taux de défaillance ou fonction du taux de mortalité ou fonction de risque, et lorsque X mesure la durée de réparation d'un système, elle s'appelle fonction de taux de réparation.

D'après la définition ci-dessus, on peut écrire

$$\lambda(x) = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{P}(x < X \leq x + h|X > x) \tag{4.50}$$

ou

$$\mathbb{P}(x < X \leq x + h|X > x) = \lambda(x)h + o(h) \tag{4.51}$$

c'est-à-dire que pour un système mis en fonctionnement au temps $x = 0$, la quantité $\lambda(x)h$, lorsque $h \to 0$, donne approximativement la probabilité que le système tombe en panne dans l'intervalle de temps $]x, x + h]$ sachant qu'il n'est pas tombé en panne dans $[0, x]$.

Dans le cas où λ est une constante, alors X suit une loi exponentielle de paramètre λ et réciproquement.

Pour les v.a.r. positives la fonction $1 - F(\cdot)$, notée $S(\cdot)$ ou $R(\cdot)$, est appelée fonction de survie ou fonction de fiabilité et joue un role très important dans les applications (Biostatistique, Fiabilité, Maintenance,...).

De (4.50) nous déduisons que la fonction du taux de hasard et la fonction de fiabilité sont liées par la relation suivante

$$R(x) = R(0)e^{-\int_0^x \lambda(u)du}, \quad (x \geq 0). \tag{4.52}$$

4.15 Notions sur les relations d'ordre stochastiques

Définition 4.3. La v.a. X est stochastiquement plus grande que la v.a. Y, et on note $X \overset{st}{\geq} Y$, si, pour tout réel c,

$$\mathbb{P}(X > c) \geq \mathbb{P}(Y > c) \tag{4.53}$$

Proposition 4.18. $X \overset{st}{\geq} Y$ ssi $\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)]$ pour toute fonction f croissante.

Corollaire 4.1. Si $X \overset{st}{\geq} Y$, alors $\mathbb{E}X \geq \mathbb{E}Y$.

Chapitre 5

Variable aléatoire vectorielle

5.1 Covariance de deux variables aléatoires

Dans les chapitres précédents, nous avons parlé exclusivement de v.a. scalaires. Dans ce chapitre nous allons présenter d'abord deux v.a.r. jointes ou simultanées X et Y et ensuite des vecteurs [1] $\mathbf{X} = (X_1, X_2, \dots, X_d)'$ à valeurs dans $\mathbb{R}^d$.

Considérons, par exemple, deux v.a. jointes discrètes. On dit que les v.a. X et Y à valeurs dans E et F (au plus dénombrables) sont jointes si

$$p_{XY}(x,y) = \mathbb{P}(X = x, Y = y) \qquad (x \in E, y \in F) \tag{5.1}$$

satisfait :

1. $p_{XY}(x,y) \geq 0$ pour tout $(x,y) \in E \times F$,
2. $\sum_{(x,y)\in E\times F} p_{XY}(x,y) = 1$.

La fonction $p_{XY}(\cdot,\cdot)$ est appelée *la loi jointe* de X et Y. Les probabilités définies par

$$\mathbb{P}(Y = y) = p_Y(y) = \sum_{x\in E} p_{XY}(x,y) \tag{5.2}$$

$$\mathbb{P}(X = x) = p_X(x) = \sum_{y\in F} p_{XY}(x,y) \tag{5.3}$$

sont appelées *lois marginales* de Y et X respectivement. En fait, ce sont les lois de Y et de X respectivement.

Si X et Y sont indépendantes alors pour tout $x \in E$ et $y \in F$ on a

$$p_{XY}(x,y) = p_X(x)p_Y(y), \tag{5.4}$$

et on note $p_{XY} = p_X \otimes p_Y$.

Fonction de répartition jointe. Elle est définie comme suit

$$F_{XY}(x,y) = \mathbb{P}(X \leq x, Y \leq y). \tag{5.5}$$

Définition 5.1 (Covariance)**.**
Soient deux v.a.r. X et Y du deuxième ordre (i.e. $\mathbb{E}X^2 < \infty$ et $\mathbb{E}Y^2 < \infty$). La covariance de X et Y est définie par la relation suivante

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]. \tag{5.6}$$

Propriétés :

1. $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

1. $\mathbf{X}'$ signifie transposé.

2. $\text{Cov}(X, X) = \text{Var}(X)$
3. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
4. $\text{Cov}(aX + bY, Z) = a\,\text{Cov}(X, Z) + b\,\text{Cov}(Y, Z), \qquad (a, b \in \mathbb{R})$

Remarque 5.1. De la première propriété nous déduisons que l'égalité $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$ est vérifiée, si, et seulement si, $\text{Cov}(X, Y) = 0$. Les v.a. X et Y sont alors appelées *non corrélées*.

Proposition 5.1. Si les v.a. X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.

Remarque 5.2. Il faut noter que la réciproque de la proposition précédente est fausse.

♣ **Exemple 5.1.** Soient deux v.a. X et Y. De la relation $\text{Var}(X + Y) = \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2$, nous déduisons $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X, Y)$.

Définition 5.2 (Coefficient de corrélation linéaire)**.** .
Le coefficient de corrélation linéaire ρ_{XY} de deux v.a. X et Y est défini par la relation suivante

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \tag{5.7}$$

De l'inégalité de Cauchy-Schwarz, nous déduisons directement que : $|\rho_{XY}| \leq 1$. Si $| \rho_{XY} |= 1$, alors X et Y sont linéairement dépendantes avec probabilité 1.

5.2 Variable aléatoire vectorielle

Considérons l'espace euclidien $\mathbb{R}^d$, $(d \geq 1)$ muni de sa base canonique, et soit $\mathbf{X} : \Omega \to \mathbb{R}^d$. Si $X_i, i = 1, ..., d$ sont des v.a.r., alors $\mathbf{X} = (X_1, ..., X_d)' \in \mathbb{R}^d$ est appelée v.a. vectorielle ou vecteur aléatoire. Il est à noter que toute coordonnée X_i est une v.a. réelle.

5.2.1 Loi, densité et fonction de répartition

La loi de probabilité de la v.a.v. $\mathbf{X}$, notée $P_{\mathbf{X}}$, est la loi image de $\mathbb{P}$ par $\mathbf{X}$, i.e.

$$P_{\mathbf{X}}(B) = \mathbb{P}(\mathbf{X}^{-1}(B)) = \mathbb{P}(\{\omega : \mathbf{X}(\omega) \in B\}) \tag{5.8}$$

en particulier, pour tout $B = \prod_{i=1}^{d}]-\infty, x_i]$ de $\mathbb{R}^d$.

S'il existe une fonction f, sur $\mathbb{R}^d$, telle que pour tout événement B de $\mathbb{R}^d$ nous avons

$$P_{\mathbf{X}}(B) = \int \cdots \int_B f(x_1, ..., x_d)dx_1 \cdots dx_d \tag{5.9}$$

alors elle est appelée densité de probabilité de la v.a.v. $\mathbf{X}$ (ou de sa loi $P_{\mathbf{X}}$).

La fonction de répartition de la v.a.v. $\mathbf{X}$, notée F, est définie comme suit

$$F(x_1, ..., x_d) = \mathbb{P}(X_1 \leq x_1, ..., X_d \leq x_d) \tag{5.10}$$

pour tout $\mathbf{x} = (x_1, ..., x_d)' \in \mathbb{R}^d$.

La fonction de répartition d'une v.a.v. vérifie les quatre propriétés suivantes :

1. $F(x_1, ..., x_d)$ est croissante pour chacun de ses arguments ;
2. $F(x_1, ..., x_d)$ est continue à droite pour chacun de ses arguments ;
3. $F(x_1, ..., x_d) \to 0$ lorsque $x_j \to -\infty$ pour au moins un j, et $F(x_1, ..., x_d) \to 1$ lorsque $x_j \to +\infty$ pour tout $j = 1, ..., d$;
4. si $a_i \leq b_i, i = 1, ..., d$ et
$\Delta_{a_i b_i}F(x_1, ..., x_d) = F(x_1, ..., x_{i-1}, b_i, x_{i+1}, ..., x_d) - F(x_1, ..., x_{i-1}, a_i, x_{i+1}, ..., x_d)$,
alors $\Delta_{a_1 b_1}\Delta_{a_2 b_2}...\Delta_{a_d b_d}F(x_1, ..., x_d) \geq 0$.

Toute fonction vérifiant les propriétés 1., 2., 3. et 4., ci-dessus, est appelée *fonction de répartition d-dimensionnelle*.

Remarque 5.3. Dans le cas où $d = 1$, la première condition ci-dessus implique la quatrième ; ce n'est plus vrai pour $d > 1$.

La densité f (lorsqu'elle existe) et la fonction de répartition F, d'une v.a.v., sont liées par les relations suivantes :

$$F(x_1, ..., x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, ..., u_d) du_1 \cdots du_d \tag{5.11}$$

$$\frac{\partial^d F}{\partial x_1 \cdots \partial x_d}(x_1, ..., x_d) = f(x_1, ..., x_d). \tag{5.12}$$

La deuxième relation ci-dessus est vérifiée lorsque $(x_1, ..., x_n)$ est un point de continuité de f. Les fonctions de répartition suivantes

$$F_i(x) = F(\infty, ..., \infty, x, \infty, ..., \infty) \tag{5.13}$$

où x est le ième argument, pour $i = 1, ..., d$, sont appelées *fonctions de répartition marginales*. En fait, ces sont les fonctions de répartition des v.a. X_i.

De même, la fonction $f_i(x)$ définit comme suit

$$f_i(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, ..., u_{i-1}, x, u_{i+1}, ..., u_d) du_1 \cdots du_{i-1} du_{i+1} \cdots du_d, \tag{5.14}$$

est la i-ème densité marginale ou densité de la v.a. X_i.

5.2.2 Espérance et matrice de variances-covariances

L'espérance d'un vecteur aléatoire $\mathbf{X}$ à valeurs dans $\mathbb{R}^d$, lorsqu'elle existe [2], est un point de $\mathbb{R}^d$:

$$\mathbb{E}\mathbf{X} = (\mathbb{E}X_1, ..., \mathbb{E}X_d)' \tag{5.15}$$

avec

$$\mathbb{E}X_i = \int_{\mathbb{R}} x dF_i(x) \tag{5.16}$$

pour $i = 1, ..., d$.

Lorsque $\mathbb{E}\mathbf{X} = \mathbf{0}$, l'origine de $\mathbb{R}^d$, alors le vecteur $\mathbf{X}$ est dit *centré*.

Soit une v.a.v. $\mathbf{X}$ à valeurs dans l'espace euclidien $\mathbb{R}^d$. La matrice de variances-covariances de $\mathbf{X}$, notée K est : $K = (K_{ij}; i, j = 1, ..., d)$, avec

$$K_{ij} = \mathrm{Cov}(X_i, X_j) \tag{5.17}$$

Propriétés :

1. K est symétrique ;
2. K est définie positive, i.e. pour tout $\mathbf{a} \in \mathbb{R}^d$ on a : $\mathbf{a}'K\mathbf{a} \geq 0$.

Remarque 5.4. La matrice de variances-covariances d'un vecteur dont les coordonnées sont non corrélées (c'est le cas, par exemple, des v.a. indépendantes) est une matrice diagonale.

Proposition 5.2. Soit une v.a.v. $\mathbf{X} = (X_1, ..., X_d)'$ de matrice de variances-covariances K. Alors nous avons :

1. $K = \mathbb{E}[(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))'] = \mathbb{E}(\mathbf{X}\mathbf{X}') - \mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{X})$.
2. $\mathrm{Var}(\sum_{i=1}^d X_i) = \mathbf{1}'K\mathbf{1}$, où $\mathbf{1} = (1, 1, ..., 1)' \in \mathbb{R}^d$.

2. Une v.a.v. à valeurs dans l'espace euclidien $\mathbb{R}^d$ est dite intégrable si $\mathbb{E}\|\mathbf{X}\| < \infty$, où $\|\cdot\|$ est la norme euclidienne.

5.3 Loi multinomiale

Considérons une urne contenant d boules, , $d \in \mathbb{N}^*$, de couleurs différentes en proportion : $p_1 : p_2 : \cdots : p_d$ avec $p_1 + p_2 + \cdots + p_d = 1$.

On tire n boules l'une après l'autre avec remise. Notons N_i la v.a. désignant le nombre de boules de couleur i tirées lors de cette expérience, $1 \leq i \leq d$. Alors nous avons

$$\mathbb{P}(N_1 = n_1, N_2 = n_2, ..., N_d = n_d) = \frac{n!}{n_1! n_2! ... n_d!} p_1^{n_1} p_2^{n_2} ... p_d^{n_d},$$

avec $0 \leq n_i \leq n$ et $n_1 + n_2 + ... + n_d = n$.

C'est la loi multinomiale, notée $M(n; p_1, p_2, ..., p_d)$. Pour $r = 2$ on a la loi binomiale $b(n, p_1)$.

Nous avons $\mathbb{E}N_i = np_i$, $\text{Var}(N_i) = np_i(1 - p_i)$ et $\text{Cov}(N_i, N_j) = n(n-1)p_i p_j$.

5.4 Vecteur aléatoire normal ou gaussien

Définition 5.3. Une v.a.v. $\mathbf{X} = (X_1, ..., X_d)'$ à valeurs dans $\mathbb{R}^d$ est dite gaussienne, si pour tout vecteur fixé $\mathbf{a} = (a_1, ..., a_d)' \in \mathbb{R}^d$, la v.a. réelle $a_1 X_1 + \cdots + a_d X_d$ est gaussienne.

Proposition 5.3. Soit $\mathbf{X}$ une v.a.v. gaussienne à valeurs dans $\mathbb{R}^d$. Alors

1. chacune de ses coordonnées X_i est une v.a. gaussienne ;
2. si $\mathbf{u} : \mathbb{R}^d \to \mathbb{R}^m$ est une application affine, alors $\mathbf{u} \circ \mathbf{X}$ est gaussienne. Si, de plus, $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, alors $K_{\mathbf{Y}} = BK_{\mathbf{X}}B'$.

Remarque 5.5. La réciproque de 1. est fausse.

♣ **Exemple 5.2.** Soit $\mathbf{X} \sim N(\mu, K)$ avec

$$\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

et

$$K = \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$$

Si $Y_1 = X_1 + 3X_2 + 1$ et $Y_2 = 2X_1 - 5X_2 - 1$ alors suivant la proposition 5.3(2), nous avons $Y = (Y_1, Y_2)' \sim N(\bar{}_{\mathbf{Y}}, K_{\mathbf{Y}})$ avec

$$\mu_Y = \begin{bmatrix} 8 \\ -9 \end{bmatrix}$$

et

$$K_{\mathbf{Y}} = \begin{bmatrix} 34 & -75 \\ -75 & 169 \end{bmatrix}$$

Proposition 5.4. Si les v.a. $X_i, i = 1, ..., d$ (réelles ou vectorielles) sont gaussiennes et indépendantes, alors le vecteur $\mathbf{X} = (X_1, ..., X_d)'$ est gaussien.

Proposition 5.5. Si la matrice de variances-covariances d'un vecteur gaussien est diagonale, alors ses coordonnées sont indépendantes.

Proposition 5.6. Si $\mathbf{X}$ est un vecteur gaussien, à valeurs dans $\mathbb{R}^d$, de moyenne μ, et de matrice de variances-covariances K, pour que la loi de $\mathbf{X}$ admette une densité, il faut et il suffit que la matrice K soit non singulière. La densité de $\mathbf{X}$ est alors

$$f(x_1, ..., x_d) = (2\pi)^{-d/2}(detK)^{-1/2} \exp\{-\frac{1}{2}(\mathbf{x} - \mu)' K^{-1} (\mathbf{x} - \mu)\} \tag{5.18}$$

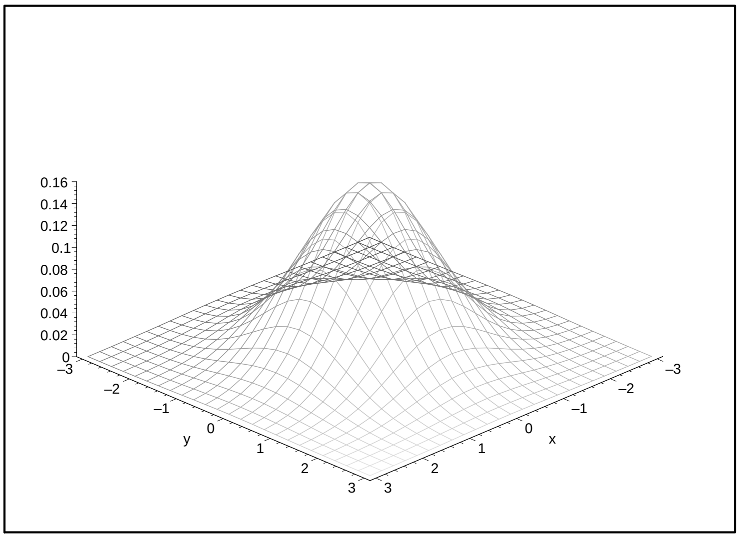


FIGURE 5.1 – Densité de la loi normale $\mathcal{N}_2(0, I)$.-

Notation Lorsque **X** est un vecteur gaussien dans $\mathbb{R}^d$, de vecteur moyen μ et de matrice de variances-covariances K, on note : $\mathbf{X} \sim N_d(\mu, K)$.

Vecteur gaussien bidimensionnelle. Soit $(X_1, X_2)' \sim N_2(\mu, K)$, avec $\mu = (\mu_1, \mu_2)'$ et

$$K = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Alors, nous obtenons $det(K) = \sigma_1^2\sigma_2^2(1-\rho^2)$, et

$$K^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.$$

Et donc la densité de $(X_1, X_2)'$ s'écrit

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\Big\{ -\frac{1}{2(1-\rho^2)}[(\frac{x_1-\mu_1}{\sigma_1})^2 + (\frac{x_2-\mu_2}{\sigma_2})^2 - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}]\Big\}.$$

5.5 Fonction d'une variable aléatoire vectorielle

Considérons une v.a.v. **X** à valeurs dans $\mathbb{R}^d$ de densité f, une application (mesurable) g : $\mathbb{R}^d \longrightarrow \mathbb{R}$ et la v.a. $Y = g \circ \mathbf{X}$, i.e., pour tout $\omega \in \Omega$, $Y(\omega) = g(X_1(\omega), ..., X_d(\omega))$. La fonction de répartition de Y est donnée par

$$F_Y(y) = \int \cdots \int_B f(t_1, ..., t_d) dt_1 ... dt_d \tag{5.19}$$

où $B = \{(t_1, ..., t_d) \in \mathbb{R}^d : g(t_1, ..., t_d) \leq y\}$.

L'espérance de Y est

$$\mathbb{E}Y = \int \cdots \int_{\mathbb{R}^d} g(t_1, ..., t_d) f(t_1, ..., t_d) dt_1 \cdots dt_d. \tag{5.20}$$

Soient E, F deux ouverts de $\mathbb{R}^d$ et un difféomorphisme $\mathbf{g} = (g_1, \ldots, g_d)' : E \to F$ (i.e., sa réciproque $\mathbf{g}^{-1}$ existe et $\mathbf{g}$ et $\mathbf{g}^{-1}$ sont continûment différentiables). Alors

$$\mathbf{Y} = \mathbf{g} \circ \mathbf{X} \tag{5.21}$$

est une v.a.v. à valeurs dans F. Les coordonnées de $\mathbf{Y}$ sont :

$$\begin{aligned} Y_1 &= g_1(X_1, ..., X_d) \\ &\vdots \\ Y_d &= g_d(X_1, ..., X_d) \end{aligned}$$

Proposition 5.7. La densité h de la v.a.v. $\mathbf{Y}$ est donnée par

$$h(y) = f \circ \mathbf{g}^{-1}(y) |DJ_{\mathbf{g}^{-1}}(y)| \mathbf{1}_F(y), \tag{5.22}$$

où $DJ_{\mathbf{g}^{-1}}$ ($DJ_{\mathbf{g}}$) est le jacobien de $\mathbf{g}^{-1}$ ($\mathbf{g}$), et nous avons : $DJ_{\mathbf{g}^{-1}} = (DJ_{\mathbf{g}})^{-1}$, et

$$DJ_{\mathbf{g}} = det \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_d} \\ \vdots & & \vdots \\ \frac{\partial g_d}{\partial x_1} & \cdots & \frac{\partial g_d}{\partial x_d} \end{bmatrix} \tag{5.23}$$

Démonstration. Par la formule de changement de variables dans les intégrales multiples [3], nous pouvons écrire, pour toute fonction $\psi : \mathbb{R}^d \longrightarrow \mathbb{R}$ continue

$$\int_E (\psi \circ \mathbf{g})(x) f(x) dx = \int_F \psi(y)(f \circ \mathbf{g}^{-1})(y) |DJ_{\mathbf{g}^{-1}}(y)| dy. \tag{5.24}$$

Et, par le théorème fondamental (cf. proposition 4.6) écrit pour des v.a.v., nous obtenons le résultat énoncé. $\square$

♣ **Exemple 5.3.** Soit la v.a.v. $(X, Y)'$ à valeurs dans $\mathbb{R}^2$ de densité f, et la transformation linéaire $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ avec $u = g_1(x, y) = ax + by$ et $v = g_2(x, y) = cx + dy$, $(a, b, c, d \in \mathbb{R})$ et $ad - bc \neq 0$. Alors la v.a. $(U, V)' = \mathbf{g}(X, Y)$ a pour densité la fonction h donnée par

$$h(u, v) = \frac{1}{|ad - bc|} f\left(\frac{du - bv}{ad - bc}, \frac{-bu + av}{ad - bc}\right) \tag{5.25}$$

Pour $a = b = 1$ et $c = -d = 1$, nous avons : $h(u, v) = \frac{1}{2} f(\frac{1}{2}(u + v), \frac{1}{2}(u - v))$. La densité de la somme de deux v.a. $X + Y$ est la densité marginale de U, i.e.

$$h_U(u) = \int_{\mathbb{R}} h(u, v) dv = \int_{\mathbb{R}} f(y, u - y) dy \tag{5.26}$$

Si, de plus, les v.a. X et Y sont indépendantes, alors $f(x, y) = f_X(x) f_Y(y)$ et par conséquent la relation (5.26) s'écrit comme suit

$$h_U(u) = \int_{\mathbb{R}} f_X(y) f_Y(u - y) dy = (f_X * f_Y)(u) \tag{5.27}$$

Lorsque $\mathbf{g}$ n'est pas un difféomorphisme sur E tout entier, et si E peut être partitionné en m parties disjointes $E_1, ..., E_m$ telles que les restrictions de $\mathbf{g}$ sur ces parties soient des difféomorphismes, alors

$$\mathbb{P}(Y \in F) = \mathbb{P}(X \in E) = \sum_{k=1}^{m} \mathbb{P}(X \in E_k)$$

En appliquant maintenant m fois la proposition précédente on obtient

$$h(y) = \sum_{k=1}^{m} f \circ \mathbf{g}_k^{-1}(y) \left| DJ_{\mathbf{g}_k^{-1}}(y) \right| \mathbf{1}_{F_k}(y) \tag{5.28}$$

où $\mathbf{g}_k$ est la restriction de $\mathbf{g}$ sur E_k ; $\mathbf{g}_k^{-1}$ est la réciproque de $\mathbf{g}_k$ et $DJ_{\mathbf{g}_k^{-1}}$ son jacobien.

3. Cf. Cours MT22.

Chapitre 6

Indépendance et conditionnement

6.1 Indépendance des variables aléatoires

Considérons deux v.a.r. X et Y définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

La notion d'indépendance en probabilité entre les v.a. X et Y a pour objectif de donner un sens précis à la situation suivante : le fait de connaître dans l'expérience considérée la valeur de X n'apporte aucune information supplémentaire sur la valeur prise par Y.

Lorsque X et Y sont discrètes à valeurs dans E et F respectivement, leur indépendance s'exprime comme suit

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \textit{pour tout} \;\; (x, y) \in E \times F. \tag{6.1}$$

Définition 6.1. Les v.a.r. $X_1, ..., X_n$ définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sont dites indépendantes si, pour tous $A_1, ..., A_n \in \mathcal{B}$, les événements $\{X_1 \in A_1\}, ..., \{X_n \in A_n\}$ sont indépendants.

Proposition 6.1. Les v.a. $X_1, ..., X_n$ sont indépendantes ssi pour tous $A_1, ..., A_n \in \mathcal{B}$, nous avons

$$\mathbb{P}(X_1 \in A_1, ..., X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \cdots \mathbb{P}(X_n \in A_n). \tag{6.2}$$

Corollaire 6.1 (Variables aléatoires discrètes)**.**
Soient n v.a. discrètes $X_1, ..., X_n$ à valeurs dans $E_1, ..., E_n$ de lois marginales $p_1, ..., p_n$ respectivement et p leur loi jointe. Les propriétés suivantes sont équivalentes :

1. les v.a. $X_1, ..., X_n$ sont indépendantes ;
2. pour tout $(x_1, ..., x_n) \in E_1 \times \cdots \times E_n$ nous avons : $p(x_1, ..., x_n) = p_1(x_1) \cdots p_n(x_n)$.

Corollaire 6.2 (Variables aléatoires réelles)**.**
Les v.a. $X_1, ..., X_n$ de f.r. $F_1(\cdot), ..., F_n(\cdot)$ respectivement, sont indépendantes ssi leur f.r. jointe F s'écrit, pour tout $(x_1, ..., x_n) \in \mathbb{R}^n$,

$$F(x_1, ..., x_n) = F_1(x_1) \cdots F_n(x_n) \tag{6.3}$$

Corollaire 6.3 (Variables aléatoires à densité)**.**
Les v.a. $X_1, ..., X_n$ de densités $f_1(\cdot), ..., f_n(\cdot)$ respectivement, sont indépendantes ssi leur densité jointe f s'écrit, pour tout $(x_1, ..., x_n) \in \mathbb{R}^n$,

$$f(x_1, ..., x_n) = f_1(x_1) \cdots f_n(x_n). \tag{6.4}$$

Corollaire 6.4. Si la densité jointe de $X_1, ..., X_n$ s'écrit : $f(x_1, ..., x_n) = \prod_{i=1}^{n} h_i(x_i)$, $(x_1, ..., x_n) \in \mathbb{R}^n$, alors les v.a. $X_1, ..., X_n$ sont indépendantes et la densité de X_i est $c_i h_i(\cdot)$ avec c_i une constante réelle.

Nous pouvons regrouper les résultats ci-dessus sous la forme :

Proposition 6.2. Les v.a. $X_1, ..., X_n$ sont indépendantes ssi, pour toutes fonctions $h_1, ..., h_n$, pour lesquelles l'expression ci-dessous a un sens, on a

$$\mathbb{E}[h_1(X_1) \cdots h_n(X_n)] = \prod_{i=1}^{n} \mathbb{E}[h_i(X_i)]. \tag{6.5}$$

♣ **Exemple 6.1.** Les v.a. X et Y de densité jointe $f(x,y) = \mathbf{1}_{[0,1]\times[0,1]}(x,y)$ sont indépendantes car on peut écrire : $f(x,y) = \mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,1]}(y)$. Et leurs densités marginales sont identiques et égales à $\mathbf{1}_{[0,1]}(\cdot)$.

Proposition 6.3. L'indépendance d'une suite de n v.a. entraîne l'indépendance :

1. de toute sous suite $(X_{i_1}, \ldots, X_{i_k}), 1 \leq i_1 < \cdots < i_k \leq n$,
2. de toute suite des vecteurs $(X_1, \ldots, X_{n_1}), (X_{n_1+1}, \ldots, X_{n_2}), \ldots, (X_{n_k+1}, \ldots, X_n)$ obtenue par regroupement d'ensembles disjoints de coordonnées,
3. de toute suite des fonctions $f_1(X_1, \ldots, X_{n_1}), f_2(X_{n_1+1}, \ldots, X_{n_2}), \ldots, f_k(X_{n_k+1}, \ldots, X_n)$.

Définition 6.2. Dans une suite infinie, les v.a. sont indépendantes si toute sous suite finie est formée de v.a. indépendantes.

6.2 Somme de deux variables aléatoires

Nous avons déjà vu la somme de deux v.a.d. (cf. section 3.6) et la somme de deux v.a. absolument continues (cf. section 5.3). Ici nous allons voir la somme de deux v.a.r. (non nécessairement absolument continues).

Soient deux v.a.r. X et Y indépendantes, de fonctions de répartitions F_X et F_Y respectivement.

Proposition 6.4. La fonction de répartition F_Z de la v.a. $Z = X + Y$ est donnée par la convolution de Stieltjes de F_X et F_Y, c'est-à-dire

$$F_Z(u) = (F_X * F_Y)(u) = \int_{\mathbb{R}} F_X(u-y)dF_Y(y) = \int_{\mathbb{R}} F_Y(u-x)dF_X(x). \tag{6.6}$$

Démonstration. Nous avons :

$$\begin{aligned} F_Z(u) = \mathbb{P}(X+Y \leq u) &= \iint_{\mathbb{R}^2} \mathbf{1}_{]-\infty,u]}(x+y)dF_{(X,Y)}(x,y) \\ &= \iint_{\mathbb{R}^2} \mathbf{1}_{]-\infty,u]}(x+y)dF_X(x)dF_Y(y) \\ &= \int_{\mathbb{R}} dF_X(x) \int_{-\infty}^{u-x} dF_Y(y) \\ &= \int_{\mathbb{R}} F_Y(u-x)dF_X(x). \end{aligned}$$

□

Si les v.a. X et Y possèdent de densités, alors on retrouve les résultats de la section 5.5.

6.3 Loi conditionnelle

6.3.1 Variable aléatoire discrète

Soient les v.a.d. X et Y de loi jointe $p_{(X,Y)}(k,m)$, $(k,m) \in E \times F$. Pour tout $k \in E$ tel que $\mathbb{P}(X = k) > 0$, la fonction : $p_{Y|X}(\cdot \mid k) : n \mapsto \mathbb{P}(Y = n \mid X = k)$ définit une loi de probabilité sur F appelée *loi conditionnelle* de Y en $\{X = k\}$. Alors on a :

$$p_k(m) = \frac{p_{(X,Y)}(k,m)}{p_X(m)} =: p_{Y|X}(m \mid k).$$

Définition 6.3. Nous appelons loi conditionnelle de Y sachant X la famille des lois $(p_k, k \in E)$ sur F.

Espérance conditionnelle. Espérance conditionnelle de Y en $\{X = x\}$.

$$\mathbb{E}[Y \mid X = x] = \sum_{y \in F} y \mathbb{P}(Y = y \mid X = x) = \sum_{y \in V} y p_{Y|X}(y \mid x). \tag{6.7}$$

Variance conditionnelle. Variance conditionnelle de Y en $\{X = x\}$.

$$\mathrm{Var}(Y \mid X = x) = \mathbb{E}[(Y - \mathbb{E}[Y \mid X = x])^2 \mid X = x] = \sum_{y \in F} y^2 \mathbb{P}(Y = y \mid X = x) - (\mathbb{E}[Y \mid X = x])^2. \tag{6.8}$$

♣ **Exemple 6.2.** Soient X et Y deux v.a. jointes à valeurs dans $\{1, 2\}$ et $\{2, 3, 4, 9\}$ respectivement,

avec :

$$\begin{array}{ll} \mathbb{P}(Y = 2 \mid X = 1) = 0.4; & \mathbb{P}(Y = 3 \mid X = 1) = 0.6 \\ \mathbb{P}(Y = 4 \mid X = 2) = 0.4; & \mathbb{P}(Y = 9 \mid X = 2) = 0.6 \end{array}$$

Alors nous obtenons :

$$\begin{array}{c} \mathbb{E}[Y \mid X = 1] = 2 \times 0.4 + 3 \times 0.6 = 2.6 \\ \mathbb{E}[Y \mid X = 2] = 4 \times 0.4 + 9 \times 0.6 = 7 \\ \mathrm{Var}(Y|X = 1) = 2^2 \times 0.4 + 3^2 \times 0.6 - (2.6)^2 = 0.24 \\ \mathrm{Var}(Y|X = 2) = 4^2 \times 0.4 + 9^2 \times 0.6 - 7^2 = 6. \end{array}$$

6.3.2 Variable aléatoire à densité

Pour une v.a. X à valeurs dans $\mathbb{R}$ et de densité f, nous avons $\mathbb{P}(X = x) = 0$ quel que soit $x \in \mathbb{R}$. Par conséquent, la démarche précédente concernant la loi conditionnelle de Y en $\{X = x\}$ n'a plus de sens. Néanmoins, nous allons lui donner un sens analogue en considérant l'événement $\{x < X \leq x + h\}$ au lieu de $\{X = x\}$ et en prenant la limite lorsque $h \to 0$.

Soient deux v.a.r. X et Y de densité jointe f continue. Nous allons définir la fonction de répartition et la densité conditionnelle comme suit. Nous avons

$$\mathbb{P}(Y \leq y \mid X \in]x, x + h]) = \frac{\mathbb{P}(Y \leq y, x < X \leq x + h)}{\mathbb{P}(x < X \leq x + h)} \tag{6.9}$$

$$= \frac{\int_x^{x+h} du \int_{-\infty}^{y} f(u, v) dv}{\int_x^{x+h} du \int_{-\infty}^{\infty} f(u, v) dv}. \tag{6.10}$$

En divisant le numérateur et le dénominateur de (6.10) par h, et lorsque $h \to 0$, on obtient

$$F_{Y|X}(y \mid x) = \frac{1}{f_X(x)} \int_{-\infty}^{y} f(x, v) dv \tag{6.11}$$

d'où

$$f_{Y|X}(y \mid x) = \frac{\partial}{\partial y} F_{Y|X}(y \mid x) = \frac{f(x, y)}{\int_{\mathbb{R}} f(x, u) du} = \frac{f(x, y)}{f_X(x)} \tag{6.12}$$

Par conséquent, la densité conditionnelle de Y en $\{X = x\}$ est l'application

$$f_{Y|X}(\cdot|x) : y \mapsto \frac{f(x, y)}{\int_{\mathbb{R}} f(x, u) du} \tag{6.13}$$

♣ **Exemple 6.3.** Soit la loi uniforme sur le disque unité : $D = \{(x, y) : x^2 + y^2 < 1\}$ de densité

$$f(x, y) = \frac{1}{\pi} \mathbf{1}_D(x, y). \tag{6.14}$$

Nous avons :

$$f_X(x) = \int_{\mathbb{R}} f(x,u)du = \frac{2}{\pi}\sqrt{1-x^2}, \qquad -1 < x < 1. \tag{6.15}$$

Et

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{2\sqrt{1-x^2}}\mathbb{1}_D(x,y). \tag{6.16}$$

Espérance conditionnelle. Espérance conditionnelle de Y en $\{X = x\}$.

$$\mathbb{E}[Y|X = x] = \int_{\mathbb{R}} y f_{Y|X}(y|x)dy = 0. \tag{6.17}$$

Variance conditionnelle. Variance conditionnelle de Y en $\{X = x\}$.

$$\begin{aligned} \mathrm{Var}(Y|X = x) &= \int_{\mathbb{R}} y^2 f_{Y|X}(y|x)dy - (\mathbb{E}[Y|X = x])^2 \\ &= (1-x^2)/3. \end{aligned}$$

6.4 Espérance conditionnelle

Considérons les v.a. X et Y et supposons que $\mathbb{E} \mid Y \mid < \infty$.

Définition 6.4. L'espérance conditionnelle de Y sachant X, notée $\mathbb{E}[Y|X]$ est une v.a. $g(X)$, avec g définie par la formule $g(x) = \mathbb{E}[Y \mid X = x]$.

Remarque 6.1. En statistique la fonction g est appelée *fonction de régression*.

♣ **Exemple 6.4.** (Suite de l'exemple de la section 6.3.). Dans cet exemple nous avions obtenu :

$$\mathbb{E}[Y \mid X = 1] = 2 \times 0.4 + 3 \times 0.6$$
$$\mathbb{E}[Y \mid X = 2] = 4 \times 0.4 + 9 \times 0.6.$$

Par conséquent : $\mathbb{E}[Y \mid X] = 2^X \times 0.4 + 3^X \times 0.6$.

Proposition 6.5. (Cette proposition est souvent prise comme définition de l'espérance conditionnelle).
L'espérance conditionnelle $\mathbb{E}[Y|X]$ est l'unique (p.s.) v.a. de la forme $g(X)$, pour g une application bornée[1], satisfaisant la condition suivante. Pour toute fonction bornée [1] h,

$$\mathbb{E}[h(X)\mathbb{E}(Y|X)] = \mathbb{E}[h(X)Y] \tag{6.18}$$

En posant $h \equiv 1$ dans la relation (6.18), nous obtenons

Corollaire 6.5 (Formule de l'espérance totale)**.**

$$\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}(Y). \tag{6.19}$$

Quelques propriétés :

1. Linéarité : $\mathbb{E}[\sum_{i=1}^n a_i Y_i|X] = \sum_{i=1}^n a_i \mathbb{E}[Y_i|X]$ (les Y_i possèdent des espérances finies).
2. Si Y a une espérance finie et $Y \geq 0$, alors : $\mathbb{E}(Y|X) \geq 0$. En particulier, si Y_1 et Y_2 ont des espérances finies et $Y_1 \leq Y_2$, alors : $\mathbb{E}[Y_1|X] \leq \mathbb{E}[Y_2|X]$.
3. Si Y a une espérance finie et si X et Y sont indépendantes, alors : $\mathbb{E}[Y|X] = \mathbb{E}[Y]$. (La réciproque est fausse).

1. En fait, il s'agit de fonctions boréliennes bornées.

4. Si Y et $h(X)Y$ ont des espérances finies, alors : $\mathbb{E}[h(X)Y|X] = h(X)\mathbb{E}[Y|X]$.

Donnons une application de l'espérance conditionnelle.

Proposition 6.6 (Lemme de Wald)**.**
Soit une suite de v.a. $X_1, X_2, ...$ i.i.d. et une v.a. N à valeurs entières positives et indépendante de la suite considérée. Si $S_N = X_1 + X_2 + \cdots + X_N$, alors $\mathbb{E}[S_N] = \mathbb{E}[X_1]\mathbb{E}[N]$.

Démonstration. Nous avons

$$\mathbb{E}[S_N|N = n] = \mathbb{E}[X_1 + \cdots + X_n] = n\mathbb{E}[X_1].$$

Par conséquent

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N|N]] = \mathbb{E}[N\mathbb{E}X_1] = \mathbb{E}[N]\mathbb{E}[X_1].$$

□

Proposition 6.7 (Régression de la moyenne)**.**
Soit X une v.a.r. telle que $\mathbb{E}X^2 < \infty$ et Y une v.a.v. à valeurs dans $\mathbb{R}^d$. Le minimum de $\mathbb{E}(X - \varphi(Y))^2$ est réalisé lorsque $\varphi : y \mapsto \mathbb{E}[X|Y = y]$.
La fonction φ s'appelle régression de la moyenne de X en Y.

Démonstration. Par un calcul simple, nous pouvons vérifier facilement que

$$\mathbb{E}[(X - \varphi(y))^2|Y = y] \geq \mathbb{E}[(X - \mathbb{E}(X|Y = y))^2|Y = y]. \tag{6.20}$$

En intégrant les deux membres de cette inégalité par rapport à la loi marginale de Y, nous obtenons le résultat. □

Proposition 6.8 (Formule de la variance totale)**.**
Soit X une v.a.r. telle que $\mathbb{E}Y^2 < \infty$ et X une v.a.r. Alors nous avons la formule suivante

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}[\mathbb{E}(Y \mid X)].$$

Démonstration. Nous avons :

$$\begin{aligned}\mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}(Y))^2] &= \mathbb{E}[(Y - \mathbb{E}(Y \mid X) + \mathbb{E}(Y \mid X) - \mathbb{E}(Y))^2] \\ &= \mathbb{E}[(Y - \mathbb{E}(Y \mid X))^2] + \mathbb{E}[(\mathbb{E}(Y \mid X) - \mathbb{E}(Y))^2] \\ &\quad +2\mathbb{E}[(Y - \mathbb{E}(Y \mid X))(\mathbb{E}(Y \mid X) - \mathbb{E}(Y))].\end{aligned}$$

Le premier terme à droite, s'écrit

$$\mathbb{E}[(Y - \mathbb{E}(Y \mid X))^2] = \mathbb{E}\{\mathbb{E}[(Y - \mathbb{E}(Y \mid X))^2 \mid X]\} = \mathbb{E}[\mathrm{Var}(Y \mid X)].$$

Le second terme est la $\mathrm{Var}[\mathbb{E}(Y \mid X)]$. Nous allons montrer maintenant que le troixième terme est nul ce qui conlut la démonstration. En effet,

$$\begin{aligned}\mathbb{E}[(Y - \mathbb{E}(Y \mid X))(\mathbb{E}(Y \mid X) - \mathbb{E}(Y))] &= \mathbb{E}\{\mathbb{E}[(Y - \mathbb{E}(Y \mid X))(\mathbb{E}(Y \mid X) - \mathbb{E}(Y)) \mid X]\} \\ &= \mathbb{E}\{(\mathbb{E}(Y \mid X) - \mathbb{E}(Y))\mathbb{E}[(Y - \mathbb{E}(Y \mid X)) \mid X]\} \\ &= \mathbb{E}\{(\mathbb{E}(Y \mid X) - \mathbb{E}(Y))\mathbb{E}[(\mathbb{E}(Y \mid X) - \mathbb{E}(Y \mid X))]\} \\ &= 0.\end{aligned}$$

□

Chapitre 7

Convergences stochastiques et théorèmes limites

7.1 Introduction

Le problème dans ce chapitre est le suivant : étant donné un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel nous définissons :

– une v.a. X ;

– une suite de v.a. $(X_n, n \in \mathbb{N})$

dire si X_n, lorsque $n \to \infty$, converge vers la v.a. X dans un sens à préciser.

Il y a plusieurs modes de convergence. Nous allons exposer les quatre les plus usuels. Par la suite, ces quelques éléments de la topologie stochastique nous permettrons d'exposer les résultats les plus remarquables de la théorie de probabilités que sont les lois des grands nombres et le théorème de la limite centrale. Les lois des grands nombres sont liées à la théorie de l'ergodicité qui étudie les conditions sous lesquelles la moyenne arithmétique des v.a. a une limite. L'étude d'un grand nombre de phénomènes physiques est basée sur cette théorie : mélange de deux gaz parfaits, turbulence, encéphalogramme d'un homme au repos, etc.

7.2 Fonction caractéristique

Définition 7.1. La fonction caractéristique d'une v.a. X est la fonction $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ définie par

$$\varphi(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} dF(x) = \int_{\mathbb{R}} \cos(tx) dF(x) + i \int_{\mathbb{R}} \sin(tx) dF(x) \tag{7.1}$$

avec $i = \sqrt{-1}, t \in \mathbb{R}$.

Elle existe toujours (car $|e^{itx}| = 1$) et dans le cas où X possède une densité f, alors $\varphi(\cdot)$ n'est autre que la transformée de Fourier de f.

Si X est discrète de loi p, on a

$$\varphi(t) = \sum_x e^{itx} p(x) \tag{7.2}$$

Propriétés :

1. $\varphi(0) = 1$, et pour tout t, $|\varphi(t)| \leq 1$.
2. φ est uniformément continue sur $\mathbb{R}$.
3. φ est définie positive, i.e. pour tout $t_1, ..., t_n \in \mathbb{R}$ et pour tout $z_1, ..., z_n \in \mathbb{C}$, nous avons : $\sum_{j,k} \varphi(t_j - t_k) z_j \bar{z}_k \geq 0$.
4. $\varphi(-t) = \overline{\varphi}(t)$.

5. Si $Y = aX + b$, alors $\varphi_Y(t) \equiv e^{itb}\varphi_X(t)$

Les trois propriétés 1), 2) et 3) ci-dessus, caractérisent la f.c. dans le sens où φ est une f.c. ssi elle vérifie ces trois propriétés (Théorème de Bochner).

Inversion.

Par la théorie de la transformée de Fourier, nous avons : si $\int_{\mathbb{R}} |\varphi(t)|dt < \infty$, alors

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi(t)dt \tag{7.3}$$

Lois	Fonctions caractéristiques
binomiale $b(n,p)$	$(1-p-p\,e^{it})^n$
Poisson $P(\lambda)$	$e^{-\lambda+\lambda\,e^{it}}$
géométrique $G(p)$	$\frac{pe^{it}}{1-(1-p)e^{it}}$
exponentielle $E(\lambda)$	$\frac{\lambda}{\lambda-it}$
normale $N(\mu,\sigma^2)$	$e^{i\mu t-\frac{1}{2}\sigma^2 t^2}$

Proposition 7.1 (Développement de la fonction caractéristique)**.**
Si la v.a. X vérifie $\mathbb{E}|X|^k < \infty$, alors sa f.c. admet le développement limité

$$\varphi(t) = 1 + it\mu_1 + \frac{(it)^2}{2}\mu_2 + \cdots + \frac{(it)^k}{k!}\mu_k + o(t^k) \tag{7.4}$$

où $\mu_k = \mathbb{E}X^k$.

Proposition 7.2. Si les v.a. X et Y sont indépendantes, alors : $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$. Une relation analogue évidente est valable pour n v.a. indépendantes.

♣ **Exemple 7.1.** Soit $X \sim b(n,p)$. Alors on peut représenter X par la somme $Y_1 + \cdots + Y_n$, avec Y_i, i.i.d. $\sim B(p)$. Par conséquent, à l'aide de la Proposition (7.2), on obtient $\varphi_X(t) = (\varphi_Y(t))^n = (1-p+pe^{it})^n$.

♣ **Exemple 7.2.** Soit $Y \sim N(\mu,\sigma^2)$. Nous avons la représentation $Y = \sigma Z + \mu$, avec $Z \sim N(0,1)$. Alors

$$\varphi_Y(t) = e^{it\mu}\varphi_X(\sigma t) = e^{it\mu}e^{-\sigma^2 t^2/2} = e^{it\mu-\frac{1}{2}\sigma^2 t^2}$$

Fonction caractéristique d'une v.a. vectorielle $\mathbf{X}$ $= (X_1, ..., X_d)'$.

Elle est définie par la relation suivante

$$\varphi_{\mathbf{X}}(t_1, ..., t_d) = \mathbb{E}[e^{i(t_1X_1+\cdots+t_dX_d)}] \tag{7.5}$$

Si $\mathbf{X}$ possède une densité, f, alors

$$\varphi_{\mathbf{X}}(t_1, ..., t_d) = \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} e^{i(t_1x_1+\cdots+t_nx_n)}f(x_1, ..., x_d)dx_1\cdots dx_d \tag{7.6}$$

La fonction caractéristique de X_i est donnée par $\varphi_{X_i}(t) = \varphi_{\mathbf{X}}(0, ..., 0, t, 0, ..., 0)$, avec l'argument t comme i-ème argument.

Proposition 7.3. Les v.a. $X_1, ..., X_d$ sont indépendantes ssi $\varphi_{\mathbf{X}}(t_1, ..., t_n) = \prod_{i=1}^n \varphi_{X_i}(t_i)$, pour tout $(t_1, ..., t_d) \in \mathbb{R}^d$.

Proposition 7.4. Lorsque $\mathbb{E}[|X_1|^{k_1}\cdots|X_d|^{k_d}] < \infty$, alors

$$\frac{\partial^k\varphi_{\mathbf{X}}}{\partial t_1^{k_1}\cdots\partial t_d^{k_d}}(0, ..., 0) = i^k\mathbb{E}[X_1^{k_1}\cdots X_d^{k_d}] \tag{7.7}$$

avec $k = k_1 + \cdots + k_d$ et $k_i \in \mathbb{N}$.

7.3 Types de convergences

Ici les v.a. $X_n, n = 1, 2, ...$ et X sont définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

7.3.1 Convergence presque sûre

Définition 7.2. On dit que X_n, converge, lorsque $n \to \infty$, vers la v.a. X presque sûrement (p.s.) ou avec probabilité 1, si : $\mathbb{P}(\{\omega : X_n(\omega) \longrightarrow X(\omega) \text{ lorsque } n \to \infty\}) = 1$, et on note :

$$X_n \xrightarrow{p.s.} X.$$

7.3.2 Convergence en moyenne quadratique

Définition 7.3. On dit que X_n converge en moyenne quadratique, lorsque $n \to \infty$, vers X, si $\mathbb{E}[\mid X_n \mid^2] < \infty$ et $\mathbb{E}[\mid X_n - X \mid^2] \longrightarrow 0$, lorsque $n \to \infty$, et on note

$$X_n \xrightarrow{m.q.} X.$$

Remarque 7.1. La convergence en moyenne quadratique est un cas particulier de la convergence dans L^p pour $p = 2$. Dans le cas général, on a $\mathbb{E}[\mid X_n - X \mid^p] \longrightarrow 0$ et on note

$$X_n \xrightarrow{L^p} X.$$

7.3.3 Convergence en probabilité

Définition 7.4. On dit que X_n converge en probabilité, lorsque $n \to \infty$, vers X, si pour tout $\epsilon > 0$

$$\mathbb{P}(\mid X_n - X \mid > \epsilon) \longrightarrow 0 \text{ lorsque } n \to \infty$$

et on note

$$X_n \xrightarrow{p} X$$

7.3.4 Convergence en loi

Notons F_n la fonction de répartition de X_n et F la fonction de répartition de X.

Définition 7.5. On dit que X_n converge en loi, lorsque $n \to \infty$, vers X, si $F_n \longrightarrow F$ en tout point de continuité de F, et on note

$$X_n \xrightarrow{\mathcal{L}} X.$$

Proposition 7.5. Entre les quatre modes de convergence, ci-dessus, nous avons les implications suivantes :

- La convergence presque sûre implique la convergence en probabilité qui implique la convergence en loi.
- La convergence dans L^q implique la convergence dans L^p, si $q > p \geq 1$ qui implique la convergence en probabilité.

Remarque 7.2. Nous n'avons pas d'autres implications générales.

Proposition 7.6 (Théorème de la continuité de P. Lévy)**.**
Soit $(X_n; n \in \mathbb{N})$ une suite de v.a. de fonctions caractéristiques $(\varphi_n; n \in \mathbb{N})$. Si la suite (φ_n) converge simplement vers une fonction φ de $\mathbb{R}$ dans $\mathbb{C}$, continue en 0, alors φ est une fonction caractéristique et, de plus, (X_n) converge en loi vers une v.a. dont la fonction caractéristique est φ.

Proposition 7.7. Si $X_n \xrightarrow{\lambda} X$ et $Y_n \xrightarrow{\lambda} Y$, alors $X_n + Y_n \xrightarrow{\lambda} X + Y$ pour $\lambda = p.s., L^p$ et p.

Ce n'est pas vrai pour la convergence en loi.

Proposition 7.8. Si $X_n \xrightarrow{\mathcal{L}} X$ et si g est une application continue de $\mathbb{R}$ dans $\mathbb{R}$, alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

7.4 Lois des grands nombres et théorème de la limite centrale

Les suites de v.a. que nous allons considérer ici sont formées de v.a. indépendantes et identiquement distribuées (équidistribuées), désignées par l'abréviation i.i.d.. Dans le paragraphe précédent, nous avons défini des suites finies de type i.i.d. Nous définissons maintenant la même notion mais pour les suites infinies.

Définition 7.6 (Suites i.i.d.)**.**
Une suite de v.a. $(X_n, n \geq 1)$ définie sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est dite i.i.d. si toute suite finie extraite de cette suite est de type i.i.d..

Notons $S_n = X_1 + \cdots + X_n, \geq 1$.

Proposition 7.9 (Loi faible des grands nombres-Théorème de Khintchine)**.**
Soit une suite de v.a. $(X_n, n \geq 1)$ i.i.d. avec $\mathbb{E}|X_1| < \infty$. Lorsque $n \to \infty$, nous avons

$$\frac{S_n}{n} \xrightarrow{p} \mathbb{E}X_1. \tag{7.8}$$

Proposition 7.10 (Loi forte des grands nombres-Théorème de Kolmogorov)**.**
Soit une suite de v.a. $(X_n, n \geq 1)$ i.i.d. Alors $\mathbb{E}|X_1| < \infty$ est une condition nécessaire et suffisante pour que la suite (X_n) vérifie la loi forte des grands nombres, i.e., lorsque $n \to \infty$, nous avons

$$\frac{S_n}{n} \xrightarrow{p.s.} \mathbb{E}X_1. \tag{7.9}$$

♣ **Exemple 7.3.** Pour une suite d'événements $(A_n, n \geq 1)$ indépendants et de même probabilité p, la loi forte des grands nombres dit que les fréquences $\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{A_k}$ convergent p.s. vers p lorsque $n \to \infty$. Ce résultat justifie l'estimation des probabilités par les fréquences.

♣ **Exemple 7.4.** La fonction de répartition empirique d'un n-échantillon est définie par

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}. \tag{7.10}$$

Par la loi forte des grands nombres nous obtenons directement, pour tout $x \in \mathbb{R}$, lorsque $n \to \infty$

$$F_n(x) \xrightarrow{p.s.} \mathbb{E}\mathbf{1}_{\{X_1 \leq x\}} = F(x). \tag{7.11}$$

Proposition 7.11 (Théorème de la limite centrale (TLC))**.**
Soit une suite de v.a. $(X_n, n \geq 1)$, i.i.d. avec espérance commune μ et variance commune σ^2 telle que $0 < \sigma^2 < \infty$. Alors, lorsque $n \to \infty$, nous avons

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1) \tag{7.12}$$

Démonstration. Sans perdre la généralité, nous pouvons supposer que $\mu = 0$. Soit φ la f.c. de X_1, i.e., $\varphi(t) = \mathbb{E}[e^{itX_1}]$. Alors la f.c. de la v.a. $T_n = S_n/\sigma\sqrt{n}$ est $\varphi_{T_n}(t) = [\varphi(t/\sigma\sqrt{n})]^n$. D'autre part, étant donné que $\varphi(0) = 1$, $\varphi'(0) = \mathbb{E}X_1 = 0$ et $\varphi''(0) = -\sigma^2$, le développement de Taylor de φ en 0, d'ordre 2, donne

$$\varphi(\frac{t}{\sigma\sqrt{n}}) = 1 - \frac{1}{2n} + o(\frac{1}{n}).$$

Par conséquent,

$$\lim_{n\to\infty}[\varphi(\frac{t}{\sigma\sqrt{n}})]^n = \lim_{n\to\infty}(1-\frac{1}{2n}+o(\frac{1}{n}))^n = e^{-t^2/2} = \varphi_Z(t),$$

où $Z \sim N(0,1)$, et la démonstration est terminée. □

♣ **Exemple 7.5.** (Théorème de De Moivre-Laplace). Soit $(X_n, n \geq 1)$, une suite de v.a. i.i.d. avec $X_n \sim B(p)$, $(0 < p < 1)$, alors $\mu = p$ et $\sigma^2 = p(1-p) > 0$. Suivant la proposition 7.11, nous avons

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{7.13}$$

♣ **Exemple 7.6.** L'application du TLC à la suite des fonctions empiriques $F_n(\cdot)$, lorsque $n \to \infty$, donne, pour tout $x \in \mathbb{R}$,

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{\mathcal{L}} N(0, \sigma^2(x)) \tag{7.14}$$

avec $\sigma^2(x) = F(x)(1 - F(x))$.

Proposition 7.12 (TLC pour des v.a. vectorielles)**.**
Soit une suite de vecteurs aléatoires $(X_n, n \geq 1)$, i.i.d. à valeurs dans $\mathbb{R}^d$ de vecteur moyen $\mu \in \mathbb{R}^d$, et de matrice de variances-covariances commune K. Alors, lorsque $n \to \infty$, nous avons

$$\frac{1}{\sqrt{n}}(S_n - n\mu) \xrightarrow{\mathcal{L}} N_d(0, K). \tag{7.15}$$

7.5 Méthode de Monte Carlo

La méthode de Monte Carlo ou simulation stochastique constitue un des outils majeurs dans les sciences appliquées et en particulier dans la science de l'ingénieur.

Nous allons présenter le principe de cette méthode en montrant comment calculer approximativement la valeur d'une intégrale multiple d'une fonction continue $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ $(d \geq 1)$.

L'intégrale considérée est la suivante

$$I = \int_0^1 \cdots \int_0^1 f(x_1, ..., x_d) dx_1 \cdots dx_d \tag{7.16}$$

Une valeur approchée de cette intégrale peut être obtenue à l'aide de la loi forte des grands nombres. En effet, si on pose : $z_i = f(\mathbf{x}^i)$, où $\mathbf{x}^i$ est la i-ème réalisation d'un vecteur aléatoire $\mathbf{X} = (X_1, ..., X_d)'$ uniformément distribué sur le cube unité d-dimensionnel $[0,1]^d$, alors pour un entier N assez grand, et lorsque les réalisations $\mathbf{x}^1, ..., \mathbf{x}^N$ sont indépendantes, nous avons

$$I \approx \frac{1}{N}\sum_{i=1}^{N} z_i \tag{7.17}$$

Soit $R_N = \frac{1}{N}\sum_{i=1}^{N} Z_i - I$, alors par le TLC nous avons $\mathbb{E}|R_N| \approx \sigma(Z_i)/\sqrt{N}$, $(N \to \infty)$. Par conséquent la vitesse de convergence de Monte Carlo est en $O(1/\sqrt{N})$ en moyenne.

En fait, la méthode de Monte Carlo consiste à mettre en oeuvre la loi forte des grands nombres et à étudier numériquement ce genre d'approximations dans le cadre du calcul des espérances mathématiques.

Exercices 8

EXERCICES AVEC SOLUTIONS

▷EXERCICE 1. Quatre urnes identiques contients :

Urne	Boules blanches	boules rouges
1	25	10
2	20	30
3	10	45
4	70	30

Une boule est tirée au hasard et elle est blanche. Quelle est la probabilité que cette boule soit tirée de l'urne 1 ?

SOLUTION. Soit les événements :

$U_i = \{$ la boule est tirée de l'urne $i\}$, $i = 1, 2, 3, 4$,

$B =$ {la boule tirée est blanche}.

La probabilité demandée est $\mathbb{P}(U_1 \mid B)$. Nous avons $\mathbb{P}(B \mid U_1) = 25/35$, $\mathbb{P}(B \mid U_2) = 20/50$, $\mathbb{P}(B \mid U_3) = 10/55$ et $\mathbb{P}(B \mid U_4) = 70/100$. En utilisant la Formule de Bayes on obtient :

$$\begin{aligned}
\mathbb{P}(U_1 \mid B) &= \frac{\mathbb{P}(B \mid U_1)\mathbb{P}(U_1)}{\mathbb{P}(B \mid U_1)\mathbb{P}(U_1) + \mathbb{P}(B \mid U_2)\mathbb{P}(U_2) + \mathbb{P}(B \mid U_3)\mathbb{P}(U_3) + \mathbb{P}(B \mid U_4)\mathbb{P}(U_4)} \\
&= \frac{1/4 \times 5/7}{1/4 \times (5/7 + 2/5 + 2/11 + 7/10)} \\
&= 0,3578.
\end{aligned}$$

On a utilisé le fait que les urnes sont équiprobables, c'est à dire $\mathbb{P}(U_1) = \mathbb{P}(U_2) = \mathbb{P}(U_3) = \mathbb{P}(U_4) = 1/4$.

▷EXERCICE 2. Un juge d'instruction chargé d'une enquête criminelle est convaincu à 70% de la culpabilité d'un suspect donné à un certain stade de l'enquête. Peu du temps après, il arrive un nouvel indice précisant que le criminel possédait un certain attribut. Cet attribut se trouve à 30% dans la population. Quelle est la probabilité de culpabilité du suspect si il possède cet attribut ?

SOLUTION. Considérons les événements :

$A =$ {le suspect possède l'attribut}

$C =$ {le suspect est coupable}.

La probabilité demandée représente $\mathbb{P}(C \mid A)$ et on la calcule en utilisant la formule de Bayes :

$$\mathbb{P}(C \mid A) = \frac{\mathbb{P}(A \mid C)\mathbb{P}(C)}{\mathbb{P}(A \mid C)\mathbb{P}(C) + \mathbb{P}(A \mid C^C)\mathbb{P}(C^C)} = \frac{1 \times 0,7}{1 \times 0,7 + 0,3 \times 0,3} \approx 0,8860.$$

▷EXERCICE 3. Un système comporte cinq composants : $i, 1 \leq i \leq 5$, dont les fonctionnements ou défaillances sont indépendants les uns des autres. Le système tombe en panne si au moins une de cinq combinaisons de pannes suivantes a lieu : (c_1) ou $(c_2$ et $c_4)$ ou $(c_2$ et $c_5)$ ou $(c_3$ et $c_4)$ ou $(c_3$ et $c_5)$.

La probabilité qu'un composant tombe en panne est p. Soit P_S la probabilité que le système tombe en panne.

1. Calculer P_S en fonction de p. (Utiliser la formule de Poincaré).
2. Sachant que le composant c_3 est en panne, quelle est la probabilité que le système tombe en panne.
3. Application numérique. Soit $p = 10^{-2}$. Encadrer la valeur de P_S par deux valeurs a et b dans $]0, 1[$, c'est-à-dire $a \leq P_S \leq b$.

SOLUTION.

1. Considérons les événements $K_i := \{$ le composant c_i est en panne $\}, 1 \leq i \leq 5$. On a $\mathbb{P}(K_i) = p, 1 \leq i \leq 5$. En utilisant la formule de Poincaré et en regroupant les termes, on calcule la probabilité que le système tombe en panne :

$$\begin{aligned} P(\text{ panne}) &= \mathbb{P}(K_1 \cup (K_2 \cap K_4) \cup (K_2 \cap K_5) \cup (K_3 \cap K_4) \cup (K_3 \cap K_5)) \\ &= (p + 4p^2) - (8p^3 + 2p^4) + (8p^4 + 2p^5) - (p^4 + 4p^5) + p^5 \\ &= p - p^5 + p^4 - 8p^3 + 4p^2. \end{aligned}$$

2. La probabilité que le système tombe en panne peut être écrite :

$$\begin{aligned} \mathbb{P}(\text{ panne}) &= \mathbb{P}(K_1 \cup [K_2 \cap (K_4 \cup K_5)] \cup [K_3 \cap (K_4 \cup K_5)]) \\ &= \mathbb{P}(K_1 \cup [(K_2 \cup K_3) \cap (K_4 \cup K_5)]). \end{aligned}$$

En utilisant cette expression de $\mathbb{P}(\text{ panne})$ et l'indépendance des K_i on obtient

$$\begin{aligned} \mathbb{P}(panne \mid K_3) &= \frac{\mathbb{P}(panne \cap K_3)}{\mathbb{P}(K_3)} \\ &= (1/p)\mathbb{P}([K_1 \cup ((K_2 \cup K_3) \cap (K_4 \cup K_5))] \cap K_3) \\ &= (1/p)\mathbb{P}(K_3 \cap (K_1 \cup K_4 \cup K_5)) \\ &= \mathbb{P}(K_1 \cup K_4 \cup K_5) \\ &= 3p - 3p^2 + p^3. \end{aligned}$$

3.

$$(p + 4p^2) - (8p^3 + 2p^4) \leq P_S \leq (p + 4p^2).$$

En remplacent $p = 0,01$ on obtient

$$0,0103 \leq P_S \leq 0,0104.$$

$\triangleright$EXERCICE 4. On lance une pièce de monnaie équilibrée n fois. Soit les v.a. discrètes X et Y, dont $X =$ "nombre de piles" et $Y =$ "nombre de faces" observés, et on pose $Z = X - Y$.

1. Préciser l'ensemble des valeurs de la v.a. Z et donner sa loi. (Prendre en compte la parité de n).
2. Calculer $\mathbb{E}(Z)$.

SOLUTION. Soit $\epsilon_i, i = 1, \ldots, n$, les variables aléatoires définies par

$$\epsilon_i := \begin{cases} 1 & \text{si on a pile au } i^{eme} \text{ coup} \\ 0 & \text{si on a face au } i^{eme} \text{ coup.} \end{cases}$$

Les $\epsilon_i, i = 1, \ldots, n$, sont des variables de Bernoulli i.i.d., $\mathbb{P}(\epsilon_i = 1) = 1/2, \mathbb{P}(\epsilon_i = 0) = 1/2$.

1. L'ensemble des valeurs de de la v.a. Z est $\{-n, -n + 1, \ldots, n - 1, n\}$.

– $n = 2m, m \in \mathbb{N}$ (n pair)
Pour $k \in \mathbb{Z}, 0 \leq k \leq m$ on a

$$\begin{aligned} \mathbb{P}(Z = 2k) &= \mathbb{P}(X - Y = 2k) \\ &= \mathbb{P}(\{\epsilon_i = 1 \text{ pour m+k parmi les 2m coups}\,;\, \epsilon_i = 0 \text{ pour les autres m-k coups}\}) \\ &= \binom{2m}{m+k}(1/2)^{m+k}(1/2)^{m-k} = \binom{2m}{m+k}(1/2)^{2m}. \end{aligned}$$

Pour $k \in \mathbb{Z}, -m \leq k \leq -1$ on obtient le même résultat

$$\mathbb{P}(Z = 2k) = \binom{2m}{m+k}(1/2)^{2m}.$$

Pour $k \in \mathbb{Z}, k \leq -m-1$ ou $k \geq m+1$, on a évidement

$$\mathbb{P}(Z = 2k) = \mathbb{P}(X - Y = 2k) = 0.$$

De même, pour $k \in \mathbb{Z}$, on a

$$\mathbb{P}(Z = 2k+1) = \mathbb{P}(X - Y = 2k+1) = 0.$$

– $n = 2m+1, m \in \mathbb{N}^*$ (n impair)
Pour $k \in \mathbb{Z}, 0 \leq k \leq m$ on a

$$\begin{aligned} \mathbb{P}(Z = 2k+1) &= \mathbb{P}(X - Y = 2k+1) \\ &= \mathbb{P}(\{\epsilon_i = 1 \text{ pour m+k+1 parmi les 2m+1 coups}\,;\, \epsilon_i = 0 \text{ pour les autres m-k }\}) \\ &= \binom{2m+1}{m+k+1}(1/2)^{m+k+1}(1/2)^{m-k} = \binom{2m+1}{m+k+1}(1/2)^{2m+1}. \end{aligned}$$

Pour $k \in \mathbb{Z}, -m-1 \leq k \leq -1$ on obtient le même résultat

$$\mathbb{P}(Z = 2k+1) = \binom{2m+1}{m+k+1}(1/2)^{2m+1}.$$

Pour $k \in \mathbb{Z}, k \leq -m-2$ ou $k \geq m+1$, on a évidement

$$\mathbb{P}(Z = 2k+1) = \mathbb{P}(X - Y = 2k+1) = 0.$$

De même, pour $k \in \mathbb{Z}$, on a

$$\mathbb{P}(Z = 2k) = \mathbb{P}(X - Y = 2k) = 0.$$

En conclusion, on a obtenu
– Pour $n = 2m, m \in \mathbb{N}$ (n pair),

$$\begin{aligned} \mathbb{P}(Z = 2k) &= \begin{cases} \binom{2m}{m+k}(1/2)^{2m} & \text{si } -m \leq k \leq m, k \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases} \\ \mathbb{P}(Z = 2k+1) &= 0, k \in \mathbb{Z}. \end{aligned}$$

– Pour $n = 2m+1, m \in \mathbb{N}$ (n impair),

$$\begin{aligned} \mathbb{P}(Z = 2k) &= 0, k \in \mathbb{Z} \\ \mathbb{P}(Z = 2k+1) &= \begin{cases} \binom{2m+1}{m+k+1}(1/2)^{2m+1} & \text{si } -m-1 \leq k \leq m, k \in \mathbb{Z} \\ 0 & \text{sinon} \end{cases}. \end{aligned}$$

2.

– $n = 2m, m \in \mathbb{N}$ (n pair)

$$\begin{aligned}\mathbb{E}(Z) &= \sum_{k=-m}^{m} 2k\binom{2m}{m+k}(1/2)^{2m} \\ &= (1/2)^{2m}[\sum_{k=1}^{m} 2k\binom{2m}{m+k} - \sum_{k=1}^{m} 2k\binom{2m}{m-k}] \\ &= 0.\end{aligned}$$

– $n = 2m+1, m \in \mathbb{N}$ (n impair)

$$\begin{aligned}\mathbb{E}(Z) &= \sum_{k=-m-1}^{m} (2k+1)\binom{2m+1}{m+k+1}(1/2)^{2m+1} \\ &= (1/2)^{2m+1}[\sum_{k=0}^{m}(2k+1)\binom{2m+1}{m+k+1} - \sum_{k=1}^{m+1}(2k-1)\binom{2m+1}{m-k+1}] \\ &= (1/2)^{2m+1}[\sum_{k=1}^{m+1}(2k-1)\binom{2m+1}{m+k} - \sum_{k=1}^{m+1}(2k-1)\binom{2m+1}{m-k+1}] \\ &= 0.\end{aligned}$$

Alors, $\mathbb{E}(Z) = 0$.

$\triangleright$EXERCICE 5. Considérons l'axe des entiers relatifs $\mathbb{Z}$, et une particule se déplaçant sur les entiers en partant de zéro. A chaque unité de temps il effectue un mouvement soit vers la droite de $(+1)$ avec probabilité p, soit vers la gauche de (-1) avec probabilité $1-p$, où $p \in]0,1[$.

Soit S_n, pour $n \in \mathbb{N}$, la position de la particule au temps n. Soit $X_1, ..., X_n$ une suite de v.a. i.i.d. à valeurs $\{-1,+1\}$ et de loi commune $P_{X_1}(-1) = 1-p$ et $P_{X_1}(+1) = p$.

Pour un $n \geq 1$ fixé, calculer :

1. S_n en fonction de $X_1, ..., X_n$;
2. la position moyenne de la particule au temps $n \geq 1$ et sa variance, c'est-à-dire $\mathbb{E}[S_n]$ et $\mathrm{Var}(S_n)$;
3. la probabilité de retour en 0, au temps n, $\mathbb{P}(S_n = 0)$;
4. la probabilité : $\mathbb{P}(X_k \neq X_{k+1})$ pour $1 \leq k \leq n-1$;
5. le nombre de changements de direction, noté N, en fonction des v.a. $\mathbf{1}_{\{X_k \neq X_{k+1}\}}$, pour $1 \leq k \leq n-1$;
6. $\mathbb{E}(N)$.

SOLUTION. Soit $X_i, i = 1, \ldots, n$, les variables aléatoires définies par

$$X_i := \begin{cases} 1 & \text{si à l'instant } i-1 \text{ la particule se déplace vers la droite (avec } +1) \\ -1 & \text{si à l'instant } i-1 \text{ la particule se déplace vers la gauche (avec } -1). \end{cases}$$

Les $X_i, i = 1, \ldots, n$, sont des variables de Bernoulli i.i.d. avec $\mathbb{P}(X_i = 1) = p, \mathbb{P}(X_i = -1) = 1-p$.

1. $S_n = \sum_{i=1}^n X_i$.

2. Comme $\mathbb{P}(X_i^2 = 1) = 1$, on a

$$\begin{aligned}\mathbb{E}(X_i) &= 1 \cdot p + (-1) \cdot (1-p) = 2p-1, \\ \mathrm{Var}(X_i) &= \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = 1-(2p-1)^2 = 4p(1-p).\end{aligned}$$

Les X_i etant independantes, on obtient

$$\begin{aligned}\mathbb{E}(S_n) &= \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n (2p-1) = n(2p-1), \\ \mathrm{Var}(S_n) &= \mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i) = \sum_{i=1}^n 4p(1-p) = 4np(1-p).\end{aligned}$$

3.
– $n = 2m, m \in \mathbb{N}$ (n pair)

$$\begin{aligned} & \mathbb{P}(S_n = 0) \\ = \;& \mathbb{P}(\{X_i = 1 \text{ pour m parmi les 2m mouvements}\,;\, X_i = -1 \text{ pour les autres m mouvements}\}) \\ = \;& \binom{2m}{m} p^m (1-p)^m. \end{aligned}$$

– $n = 2m + 1, m \in \mathbb{N}^*$ (n impair)
$\mathbb{P}(S_n = 0) = 0$

4. En utilisant l'indépendance des v.a. $X_k, k = 1, \ldots, n$, et l'incompatibilité des événements $\{X_k = -1; X_{k+1} = 1\}$ et $\{X_k = 1; X_{k+1} = -1\}, k = 1, \ldots, n-1$, on obtient

$$\begin{aligned} \mathbb{P}(X_k \neq X_{k+1}) &= \mathbb{P}(\{X_k = 1; X_{k+1} = -1\} \cup \{X_k = -1; X_{k+1} = 1\}) \\ &= \mathbb{P}(X_k = 1)\mathbb{P}(X_{k+1} = -1) + \mathbb{P}(X_k = -1)\mathbb{P}(X_{k+1} = 1) \\ &= 2p(1-p). \end{aligned}$$

5. Le nombre de changements de direction est donné par

$$N = \sum_{k=1}^{n-1} \mathbf{1}_{\{X_k \neq X_{k+1}\}}.$$

6.

$$\begin{aligned} \mathbb{E}(N) &= \sum_{k=1}^{n-1} \mathbb{E}(\mathbf{1}_{\{X_k \neq X_{k+1}\}}) = \sum_{k=1}^{n-1} \mathbb{P}(\{X_k \neq X_{k+1}\}) \\ &= \sum_{k=1}^{n-1} 2p(1-p) = 2p(1-p)(n-1). \end{aligned}$$

$\triangleright$EXERCICE 6. Un système à 4 composants $c_i, i = 1, 2, 3, 4$ tobme en panne si une des combinaisons suivantes est réalisée : (c_1 et c_2) ou (c_3 et c_4).

Soit $X_i, i = 1, \ldots, 4$, les variables aléatoires définies par

$$X_i := \begin{cases} 1, & \text{si la composante } c_i \text{ est en bon état} \\ 0, & \text{sinon.} \end{cases}$$

Pour chaque variable X_i on connaît son espérance, $\mathbb{E}(X_i) = p_i$, avec $p_i \in (0, 1), i = 1, \ldots, 4$. Considérons l'événement $S = \{$ Système en bon état$\}$.

1. Exprimer S en fonction des événements $\{X_i = 1\}, i = 1, \ldots, 4$.
2. Calculer $\mathbb{P}(S)$ en fonction de $p_i, i = 1, \ldots, 4$.
3. Calculer $\mathbb{P}(X_i = 0 \mid S^C)$.

SOLUTION.

1.

$$S = \Big(\{X_1 = 1\} \cup \{X_2 = 1\}\Big) \cap \Big(\{X_3 = 1\} \cup \{X_4 = 1\}\Big)$$

2. L'espérance $\mathbb{E}(X_i)$ est donnée par

$$\mathbb{E}(X_i) = 1 \times \mathbb{P}(X_i = 1) + 0 \times \mathbb{P}(Xi = 0) = \mathbb{P}(X_i = 1).$$

Donc $\mathbb{P}(X_i = 1) = p_i$ et on note $\mathbb{P}(Xi = 0) = 1 - p_i =: q_i, i = 1, \ldots, 4$. En tenant compte de l'indépendance des quatre composants, on obtient :

$$\begin{aligned} \mathbb{P}(S) &= \mathbb{P}\Big(\{X_1 = 1\} \cup \{X_2 = 1\}\Big)\mathbb{P}\Big(\{X_3 = 1\} \cup \{X_4 = 1\}\Big) \\ &= \Big(\mathbb{P}(X_1 = 1) + \mathbb{P}(X_2 = 1) - \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)\Big) \\ &\quad \times \Big(\mathbb{P}(X_3 = 1) + \mathbb{P}(X_4 = 1) - \mathbb{P}(X_3 = 1)\mathbb{P}(X_4 = 1)\Big) \\ &= (p_1 + p_2 - p_1p_2)(p_3 + p_4 - p_3p_4). \end{aligned}$$

3. Pour i fixé, en écrivant la formule de Bayes on a :

$$\mathbb{P}(X_i = 0 \mid S^C) = \frac{\mathbb{P}(S^C \mid X_i = 0)\mathbb{P}(X_i = 0)}{\mathbb{P}(S^C \mid X_i = 0)\mathbb{P}(X_i = 0) + \mathbb{P}(S^C \mid X_i = 1)\mathbb{P}(X_i = 1)}$$

Pour $i = 1$ on a :

$$\begin{aligned} \mathbb{P}(S^C \mid X_1 = 0) &= \mathbb{P}\Big[\{X_2 = 0\} \cup \Big(\{X_3 = 0\} \cap \{X_4 = 0\}\Big)\Big] \\ &= q_2 + q_3q_4 - q_2q_3q_4. \\ \mathbb{P}(S^C \mid X_1 = 1) &= \mathbb{P}\Big(\{X_3 = 0\} \cap \{X_4 = 0\}\Big) = q_3q_4 \end{aligned}$$

Alors,

$$\mathbb{P}(X_1 = 0 \mid S^C) = \frac{(q_2 + q_3q_4 - q_2q_3q_4)q_1}{(q_2 + q_3q_4 - q_2q_3q_4)q_1 + q_3q_4p_1}.$$

De même, pour $i = 2, 3, 4$ on obtient

$$\begin{aligned} \mathbb{P}(X_2 = 0 \mid S^C) &= \frac{(q_1 + q_3q_4 - q_1q_3q_4)q_2}{(q_1 + q_3q_4 - q_1q_3q_4)q_2 + q_3q_4p_2} \\ \mathbb{P}(X_3 = 0 \mid S^C) &= \frac{(q_4 + q_1q_2 - q_4q_1q_2)q_3}{(q_4 + q_1q_2 - q_4q_1q_2)q_3 + q_1q_2p_3} \\ \mathbb{P}(X_4 = 0 \mid S^C) &= \frac{(q_3 + q_1q_2 - q_3q_1q_2)q_4}{(q_3 + q_1q_2 - q_3q_1q_2)q_4 + q_1q_2p_4}. \end{aligned}$$

▷EXERCICE 7. Dans un jeu de pile ou face, un joueur gagne ou perd un euro à chaque lancer. Lorsque le résultat est pile le joueur gagne un euro, et lorsqu'il est face le joueur perd un euro. Le joueur entre au jeu avec une fortune initiale de x euro. Il quitte le jeu lorsque sa fortune atteint a euro ou lorsqu'il perd sa fortune. Nous avons $x, a \in \mathbb{N}^*$.

Soit $p(x)$ la probabilité que le joueur se ruine.

1. Montrer que
$$p(x) = \frac{1}{2}[p(x+1) + p(x-1)], \quad 0 < x < a. \tag{8.1}$$
(Utiliser le théorème de probabilités totales).

2. On admet que la solution de l'équation (8.1) est de la forme suivante
$$p(x) = A + Bx.$$
Déterminer les constantes A et B. (Utiliser les conditions aux limites : $p(0) = 1$ et $p(a) = 0$).

3. Quelle est l'espérance de gains du joueur ? Expliquer.

4. APPLICATION NUMÉRIQUE. Soit $a = 60$ et $x = 15$. Calculer la probabilité de ruine et l'espérance de gains du joueur.

SOLUTION.

1. Considérons les événements :
   $A =$ {le joueur se ruine}
   $B =$ {le joueur gagne lors du premier lancer}.
   Nous avons $\mathbb{P}(A) = p(x)$ et $\mathbb{P}(B) = 1/2$. Alors :
$$\begin{aligned} p(x) &= \mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^C) \\ &= \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^C)\mathbb{P}(B^C) \\ &= p(x+1)\frac{1}{2} + p(x-1)\frac{1}{2}. \end{aligned}$$
2. En utilisant les conditions aux limites $p(0) = 1$ et $p(a) = 0$ dans l'équation $p(x) = A + Bx$ nous avons le système
$$\left\{ \begin{array}{l} 1 = A + B \cdot 0 \\ 0 = A + B \cdot a \end{array} \right. .$$
   Nous obtenons $A = 1$ et $B = -1/a$, donc $p(x) = 1 - x/a$.
3.
$$\mathbb{E}(\text{gains}) = (a - x)(1 - p(x)) - x \cdot p(x) = 0.$$
   Ce résultat est dû au fait que le jeu est équitable.
4. Nous avons
$$\begin{aligned} p(15) &= 1 - \frac{15}{60} = \frac{3}{4}, \\ \mathbb{E}(\text{gains}) &= (60 - 15)(1 - \frac{3}{4}) - 15 \cdot \frac{3}{4} = 0. \end{aligned}$$

▷EXERCICE 8. Dans un stock la proportion des composants défectueux est égal à α, $(0 \leq \alpha \leq 1)$. La durée de vie T d'un composant non défectueux suit une loi géométrique de paramètre p, sur $\mathbb{N}^*$, $(0 \leq p \leq 1)$. La durée de vie d'un composant défectueux est égale à 0.

I. On prend un composant au hasard dans ce stock dont on note T_1 sa durée de vie.

I-1) Donnez la loi de T_1.

I-2) Calculez son espérance et sa variance.

II. On choisit un composant avec probabilité β du stock précédent ou avec probabilité $1 - \beta$ $(0 \leq \beta \leq 1)$ un composant garanti (c'est-à-dire sans défaut, de durée de vie T. Notons T_2 la durée de vie de ce composant.

II-1) Donnez la loi de T_2.

II-2) Donner la condition sur β, en fonction de α et p, pour que T_2 suit une loi géométrique sur $\mathbb{N}$.

III. Montrer que :

III-1. $\mathbb{E}(T_2) = (1 - \alpha\beta)\mathbb{E}(T)$.

III-2. $\text{Var}(T_2) = (1 - \alpha\beta)\text{Var}(T) + \alpha\beta(1 - \alpha\beta)/p^2$.

SOLUTION.

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

I.

I-1)
$$\mathbb{P}(T_1 = k) = \left\{ \begin{array}{ll} \alpha & \text{si } k = 0 \\ (1 - \alpha)p(1 - p)^{k-1} & \text{si } k > 0 \end{array} \right. .$$

I-2)

$$
\begin{aligned}
\mathbb{E}(T_1) &= \sum_{i\geq 0} i\cdot \mathbb{P}(T_1=i) = \sum_{i\geq 1} i\cdot(1-\alpha)p(1-p)^{i-1} = (1-\alpha)p\sum_{i\geq 1} i\cdot(1-p)^{i-1} \\
&= -(1-\alpha)p\left(\sum_{i\geq 0}(1-p)^i\right)' = -(1-\alpha)p\left(\frac{1}{p}\right)' = \frac{1-\alpha}{p}.
\end{aligned}
$$

II.
II-1)

$$
\begin{aligned}
\mathbb{P}(T_2=k) &= \mathbb{P}(T_2=k \mid \text{ composant garanti })\mathbb{P}(\text{ composant garanti }) \\
&\quad +\mathbb{P}(T_2=k \mid \text{ composant du stock })\mathbb{P}(\text{ composant du stock })
\end{aligned}
$$

Nous avons

$$
\mathbb{P}(T_2=k \mid \text{ composant garanti }) = \begin{cases} 0 & \text{si } k=0 \\ p(1-p)^{k-1} & \text{si } k>0 \end{cases}
$$

et

$$
\mathbb{P}(T_2=k \mid \text{ composant du stock }) = \mathbb{P}(T_1=k).
$$

Par conséquent, nous obtenons

$$
\begin{aligned}
\mathbb{P}(T_2=k) &= \begin{cases} \alpha\beta & \text{si } k=0 \\ \big((1-\alpha)\beta+(1-\alpha)(1-\beta)\big)p(1-p)^{k-1} & \text{si } k>0 \end{cases} \\
&= \begin{cases} \alpha\beta & \text{si } k=0 \\ (1-\alpha\beta)p(1-p)^{k-1} & \text{si } k>0 \end{cases}
\end{aligned}
$$

II-2) Il faut que $\alpha\beta = p$.
III.
III-1

$$
\mathbb{E}(T_2) = \sum_{i\geq 0} i\cdot\mathbb{P}(T_2=i) = \sum_{i\geq 1} i(1-\alpha\beta)p(1-p)^{i-1} = (1-\alpha\beta)\sum_{i\geq 1} ip(1-p)^{i-1} = (1-\alpha\beta)\mathbb{E}(T).
$$

III-2

$$
\begin{aligned}
\mathbb{E}(T_2^2) &= \sum_{i\geq 0} i^2\cdot\mathbb{P}(T_2=i) = (1-\alpha\beta)\sum_{i\geq 1} i^2p(1-p)^{i-1} = (1-\alpha\beta)\mathbb{E}(T^2). \\
\mathrm{Var}(T_2) &= \mathbb{E}(T_2^2)-[\mathbb{E}(T_2)]^2 = (1-\alpha\beta)\mathbb{E}(T^2)-(1-\alpha\beta)^2[\mathbb{E}(T)]^2 \\
&= (1-\alpha\beta)[\mathrm{Var}(T)+(\mathbb{E}T)^2]-(1-\alpha\beta)^2[\mathbb{E}(T)]^2 = (1-\alpha\beta)\mathrm{Var}(T)+\alpha\beta(1-\alpha\beta)/p^2,
\end{aligned}
$$

car $[\mathbb{E}(T)]^2 = 1/p^2$.

▷EXERCICE 9. Un coureur se deplace régulièrement d'un point A à un point B, par deux chemins différents, soit $C1$ et $C2$. A chaque début de parcours, il choisit un chemin C où C est le chemin $C1$ avec probabilité p et le chemin $C2$ avec probabilité $1-p$, $p\in]0,1[$. On estime (par expérience passée), que la durée de parcours du chemin $C1$ suit une loi exponentielle de paramètre $\lambda > 0$ et que la loi du chemin $C2$ est aussi exponentielle, mais de paramètre $\mu > 0$.

1. Donner la loi de probabilité de la durée d'un parcours C du coureur.
2. Calculer la durée moyenne et la variance d'un parcours.

SOLUTION. Soit X_1 la variable aléatoire désignant la durée du chemin $C1$ et X_2 la variable aléatoire désignant la durée du chemin $C2$.

1. Notons X la durée d'un parcours quelconque C. La loi de X est un mélange de la loi de X_1 et de X_2, avec les probabilités $p,\ 1-p$ respectivement. Calculons la fonction de répartition de X :

$$\begin{aligned} F_X(t) = \mathbb{P}(X \le t) &= \mathbb{P}(X \le t, C = C1) + \mathbb{P}(X \le t, C = C2) \\ &= \mathbb{P}(X \le t \mid C = C1)\mathbb{P}(C = C1) + \mathbb{P}(X \le t \mid C = C2)\mathbb{P}(C = C2) \\ &= \mathbb{P}(X_1 \le t)\mathbb{P}(C = C1) + \mathbb{P}(X_2 \le t)\mathbb{P}(C = C2) \\ &= pF_{X_1}(t) + (1-p)F_{X_2}(t). \end{aligned}$$

   Alors la densité de X est donnée par

$$f_X(t) = pf_{X_1}(t) + (1-p)f_{X_2}(t),$$

   où $f_{X_1}(t)$, respectivement $f_{X_2}(t)$, est la densité de X_1, respectivement de X_2. Comme $X_1 \sim \mathcal{E}(\lambda)$ et $X_2 \sim \mathcal{E}(\mu)$, on obtient la densité de X :

$$f_X(t) = p\lambda \exp(-\lambda t)\mathbf{1}_{\mathbb{R}_+}(t) + (1-p)\mu \exp(-\mu t)\mathbf{1}_{\mathbb{R}_+}(t).$$

2. On va utiliser le fait que pour une v.a. $X \sim \mathcal{E}(\lambda)$, on a

$$\mathbb{E}(X) = \frac{1}{\lambda},\ \ \mathbb{E}(X^2) = \frac{2}{\lambda^2},\ \ \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

   - La durée moyenne d'un parcours est donnée par

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} t(pf_{X_1}(t) + (1-p)f_{X_2}(t))dt = p\int_{\mathbb{R}} tf_{X_1}(t)dt + (1-p)\int_{\mathbb{R}} tf_{X_2}(t)dt \\ &= p\mathbb{E}(X_1) + (1-p)\mathbb{E}(X_2) = p\frac{1}{\lambda} + (1-p)\frac{1}{\mu}. \end{aligned}$$

   - Pour la variance d'un parcours on calcule d'abord $\mathbb{E}(X^2)$:

$$\begin{aligned} \mathbb{E}(X^2) &= \int_{\mathbb{R}} t^2(pf_{X_1}(t) + (1-p)f_{X_2}(t))dt = p\int_{\mathbb{R}} t^2f_{X_1}(t)dt + (1-p)\int_{\mathbb{R}} t^2f_{X_2}(t)dt \\ &= p\mathbb{E}(X_1^2) + (1-p)\mathbb{E}(X_2^2) = p\frac{2}{\lambda^2} + (1-p)\frac{2}{\mu^2}. \end{aligned}$$

   On obtient ensuite la variance d'un parcours :

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = p\frac{2}{\lambda^2} + (1-p)\frac{2}{\mu^2} - [p\frac{1}{\lambda} + (1-p)\frac{1}{\mu}]^2 \\ &= p(2-p)\frac{1}{\lambda^2} + (1+p)(1-p)\frac{1}{\mu^2} - 2p(1-p)\frac{1}{\lambda\mu}. \end{aligned}$$

▷EXERCICE 10. Soit deux fonctions f_1 et $f_2 : \mathbb{R}_+ \to \mathbb{R}_+$, définies par :

$$\begin{aligned} &(a) \quad f_1(x) = \frac{1}{x\sqrt{2\pi}}e^{-(\log x)^2/2}, x \ge 0, \\ &(b) \quad f_2(x) = f_1(x)[1 + sin(2\pi \log x)], x \ge 0. \end{aligned}$$

1. Montrer que f_1 et f_2 sont des densités de probabilité.
   Indication : On peut faire un changement de variables (a) $y = \log x$.
2. Tracer les deux courbes sur ordinateur, pour $0 \le x \le 5$.
   Vous pouvez utiliser le logiciel SCILAB que vous pouvez télécharger à l'adresse
   `http://scilabsoft.inria.fr/`

3. Soit les variables aléatoires X_1, X_2 ayant comme densités les fonctions f_1, f_2 respectivement. Montrer que les moments de tous ordre sont donnés par :
$$\mathbb{E}[X_1^k] = \mathbb{E}[X_2^k] = e^{k^2/2},\ k = 1, 2, \ldots.$$
4. Tracer les fonctions de répartition de X_1 et X_2.
5. Donner une conclusion concernant les deux variables aléatoires.

SOLUTION.

1. Avec le changement de variable $y = \log x$ on obtient
$$\int_{-\infty}^{\infty} f_1(x)dx = \int_0^{\infty} \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = 1,$$
car on reconnaît la densité d'une variable aléatoire normale centrée réduite $\mathcal{N}(0,1)$. Alors, f_1 est bien une densité. Ensuite,
$$\int_{-\infty}^{\infty} f_2(x)dx = \int_{-\infty}^{\infty} f_1(x)dx + \int_{-\infty}^{\infty} f_1(x)\sin(2\pi \log x)dx.$$
La première intégrale est égale à 1, car f_1 est une densité. Pour la deuxième intégrale, en faisant le changement de variable $y = \log x$, on a
$$\begin{aligned}\int_0^{\infty} f_1(x)\sin(2\pi\log x)dx &= \int_0^{\infty} \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}\sin(2\pi\log x)dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\sin(2\pi y)dy = 0,\end{aligned}$$
car la fonction à intégrer est impaire ($\sin(2\pi y)$ impaire, $e^{-y^2/2}$ paire, alors le produit est une fonction impaire). Donc f_2 est bien une densité.
2. Les courbes des f_1 et f_2 sont données dans la Figure **??**.
3. Avec le changement de variable $y = \log x$ on a
$$\begin{aligned}\mathbb{E}(X_1^k) &= \int_0^{\infty} x^k \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2} dx = \int_{-\infty}^{\infty} e^{ky} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(y^2-2ky+k^2-k^2)/2} dy = e^{k^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(y-k)^2/2} dy.\end{aligned}$$
Avec le changement de variable $z = y - k$, on obtient
$$\mathbb{E}(X_1^k) = e^{k^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = e^{k^2/2},$$
car on reconnaît la densité d'une variable $\mathcal{N}(0,1)$.
$$\begin{aligned}\mathbb{E}(X_2^k) &= \int_{-\infty}^{\infty} x^k f_1(x)dx + \int_{-\infty}^{\infty} x^k f_1(x)\sin(2\pi\log x)dx \\ &= \mathbb{E}(X_1^k) + \int_{-\infty}^{\infty} x^k f_1(x)\sin(2\pi\log x)dx.\end{aligned}$$
En faisant le changement de variable $y = \log x - k$, la deuxième intégrale devient
$$\begin{aligned}\int_0^{\infty} x^k f_1(x)\sin(2\pi\log x)dx &= \int_0^{\infty} x^k \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}\sin(2\pi\log x)dx \\ &= \int_{-\infty}^{\infty} e^{k(y+k)} \frac{1}{\sqrt{2\pi}} e^{-(y+k)^2/2}\sin(2\pi(y+k))dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(y^2-k^2)/2}\sin(2\pi y)dy = 0,\end{aligned}$$
car la fonction à intégrer est impaire.

4. Avec le changement de variable $z = \log x$, la fonction de répartition de X_1 s'écrit

$$F_1(y) = \int_0^y \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2} dx = \int_{-\infty}^{\log y} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \Phi(\log y), y > 0,$$

où $\Phi(\cdot)$ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$. Pour $y > 0$, la fonction de répartition de X_2 est

$$F_2(y) = F_1(y) + \int_0^y \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2} \sin(2\pi \log x) dx = F_1(y) + \int_{-\infty}^{\log y} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \sin(2\pi z) dz$$

Pour la deuxième intégrale on ne peut pas obtenir une expression analytique. Les graphes des fonctions de répartition F_1 et F_2 sont données dans la figure **??**.

5. On observe dans la figure **??** que les densités des variables aléatoires X_1 et X_2 sont très différentes. En revanche, à partir de la figure **??**, on se rend compte que les fonctions de répartition correspondantes sont assez proches, donc les variables X_1 et X_2 ont des lois similaires. En conclusion, pour comparer deux lois de probabilité absolument continues, il ne suffit pas de comparer les densités, mais il faut analyser les fonctions de répartition.

▷EXERCICE 11. Soit le système à 4 composants de l'exercice 6. Les durées de vie $T_i, i = 1,2,3,4$ des quatre composants sont des variables aléatoires i.i.d. (indépendantes identiquement distribuées) de loi exponentielle de paramètre λ, $\lambda > 0$.

1. Donner la loi de la durée de vie T du système.
2. Calculer la fonction du taux de défaillance.
3. Calculer $\mathbb{E}(T)$ et $\mathrm{Var}(T)$.
4. Calculer $\mathbb{P}(T > \mathbb{E}(T))$.

SOLUTION.

1. En tenant compte du fait que les variables $T_i, i = 1,2,3,4$ sont indépendantes, on a, pour $x \geq 0$:

$$\begin{aligned}
\mathbb{P}(T > x) &= \mathbb{P}\Big[\Big(\{T_1 > x\} \cup \{T_2 > x\}\Big) \cap \Big(\{T_3 > x\} \cup \{T_4 > x\}\Big)\Big] \\
&= \mathbb{P}\Big(\{T_1 > x\} \cup \{T_2 > x\}\Big)\mathbb{P}\Big(\{T_3 > x\} \cup \{T_4 > x\}\Big) \\
&= \Big(\mathbb{P}(T_1 > x) + \mathbb{P}(T_2 > x) - \mathbb{P}(T_1 > x)\mathbb{P}(T_2 > x)\Big) \\
&\quad \times \Big(\mathbb{P}(T_3 > x) + \mathbb{P}(T_4 > x) - \mathbb{P}(T_3 > x)\mathbb{P}(T_4 > x)\Big).
\end{aligned}$$

En utilisant le fait que la fonction de répartition d'une v.a. $\mathcal{E}(\lambda)$ est donnée par $F_{\mathcal{E}(\lambda)}(x) = 1 - e^{-\lambda x}, x \geq 0$, on obtient

$$\mathbb{P}(T_i > x) = 1 - \mathbb{P}(T_i \leq x) = e^{-\lambda x}, i = 1,2,3,4, x \geq 0.$$

$$\mathbb{P}(T > x) = e^{-2\lambda x}(2 - e^{-\lambda x})^2 = 4e^{-2\lambda x} - 4e^{-3\lambda x} + e^{-4\lambda x}.$$

La fonction de répartition de la durée de vie T du système est donnée par

$$F_T(x) = \mathbb{P}(T \leq x) = 1 - \mathbb{P}(T > x) = 1 - 4e^{-2\lambda x} + 4e^{-3\lambda x} - e^{-4\lambda x}, \quad x > 0$$

et on obtient la densité

$$f_T(x) = \frac{d}{dx} F_T(x) = 8\lambda e^{-2\lambda x} - 12\lambda e^{-3\lambda x} + 4\lambda e^{-4\lambda x}, \quad x > 0.$$

2. Pour tout $x > 0$, le taux de défaillance est donné par
$$\lambda_T(x) = \frac{f_T(x)}{1 - F_T(x)} = \frac{8\lambda e^{-2\lambda x} - 12\lambda e^{-3\lambda x} + 4\lambda e^{-4\lambda x}}{4e^{-2\lambda x} - 4e^{-3\lambda x} + e^{-4\lambda x}}.$$

3. On utilise le fait que pour une v.a. $X \sim \mathcal{E}(\mu)$, on a $\mathbb{E}(X) = 1/\mu$, $\mathbb{E}(X^2) = 2/\mu^2$, et $\mathrm{Var}(X) = 1/\mu^2$.
$$\begin{aligned}\mathbb{E}(T) &= \int_0^\infty x(8\lambda e^{-2\lambda x} - 12\lambda e^{-3\lambda x} + 4\lambda e^{-4\lambda x})dx \\ &= 4\int_0^\infty x2\lambda e^{-2\lambda x}dx - 4\int_0^\infty x3\lambda e^{-3\lambda x}dx + \int_0^\infty x4\lambda e^{-4\lambda x}dx \\ &= 4\frac{1}{2\lambda} - 4\frac{1}{3\lambda} + \frac{1}{4\lambda} = \frac{11}{12\lambda} \approx 0,91\frac{1}{\lambda}.\end{aligned}$$
$$\begin{aligned}\mathbb{E}(T^2) &= \int_0^\infty x^2(8\lambda e^{-2\lambda x} - 12\lambda e^{-3\lambda x} + 4\lambda e^{-4\lambda x})dx \\ &= 4\int_0^\infty x^2 2\lambda e^{-2\lambda x}dx - 4\int_0^\infty x^2 3\lambda e^{-3\lambda x}dx + \int_0^\infty x^2 4\lambda e^{-4\lambda x}dx \\ &= 4\frac{2}{4\lambda^2} - 4\frac{2}{9\lambda^2} + \frac{2}{16\lambda^2} = \frac{89}{72\lambda^2} \approx 1,23\frac{1}{\lambda^2}.\end{aligned}$$
Alors,
$$\mathrm{Var}(T) = \mathbb{E}(T^2) - \mathbb{E}(T)^2 = \frac{19}{48}\frac{1}{\lambda^2} \approx 0,39\frac{1}{\lambda^2}.$$

4. En utilisant les résultats de la question 1, on a
$$\mathbb{P}(T > \mathbb{E}(T)) = \mathbb{P}(T > \frac{11}{12\lambda}) = 4e^{-11/6} - 4e^{-11/4} + e^{-11/3} \approx 0,4094.$$

▷EXERCICE 12. La fonction de répartition de la durée de vie d'un appareil électrique est la suivante :
$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - e^{-x\cdot 0{,}002} & \text{si } 0 \le x < 100 \\ 1 - e^{-x\cdot 0{,}005} & \text{si } 100 \le x < 200 \\ 1 - e^{-x\cdot 0{,}01} & \text{si } x \ge 200 \end{cases}$$
Calculer la durée de vie moyenne de l'appareil.

SOLUTION. Soit X une variable aléatoire ayant comme fonction de répartition F.

La fonction de répartition F est représentée dans la figure **??**. Les atomes de F, identifiés par ses points de discontinuité, sont 100 et 200 et nous avons
$$\begin{aligned}\mathbb{P}(X = 100) &= F_X(100) - F_X(100-) = e^{-0.2} - e^{-0{,}5} \approx 0,2122, \\ \mathbb{P}(X = 200) &= F_X(200) - F_X(200-) = e^{-1} - e^{-2} \approx 0,2325.\end{aligned}$$
Alors, la partie continue de F est
$$c \cdot F_c(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - e^{-x\cdot 0{,}002} & \text{si } 0 \le x < 100, \\ 1 - e^{-x\cdot 0{,}005} - (e^{-0.2} - e^{-0{,}5}) & \text{si } 100 \le x < 200, \\ 1 - e^{-x\cdot 0{,}01} - (e^{-0.2} - e^{-0{,}5} + e^{-1} - e^{-2}) & \text{si } x \ge 200, \end{cases}$$
où $c = 1 - (e^{-0.2} - e^{-0{,}5} + e^{-1} - e^{-2})$. La dérivé correspondante est
$$F_c'(x) = \begin{cases} 0 & \text{si } x < 0, \\ 0,002 \cdot e^{-x\cdot 0{,}002} & \text{si } 0 \le x < 100, \\ 0,005 \cdot e^{-x\cdot 0{,}005} & \text{si } 100 \le x < 200, \\ 0,01 \cdot e^{-x\cdot 0{,}01} & \text{si } x \ge 200. \end{cases}$$

Nous obtenons (cf.formule (4.19) Chapitre 4) que la durée de vie moyenne de l'appareil est donnée par

$$\mathbb{E}(X) = 100(e^{-0.2} - e^{-0,5}) + 200(e^{-1} - e^{-2}) + \int_{-\infty}^{\infty} xF_c'(x)dx.$$

En utilisant

$$\int xe^{-\lambda x}dx = -\frac{e^{-\lambda x}}{\lambda}(x + \frac{1}{\lambda}),$$

nous calculons la durée de vie moyenne :

$$\begin{aligned}
\mathbb{E}(X) &= 100(e^{-0.2} - e^{-0,5}) + 200(e^{-1} - e^{-2}) \\
&\quad + \int_0^{100} x \cdot 0,002e^{-x\cdot 0,002}dx + \int_{100}^{200} x \cdot 0,005e^{-x\cdot 0,005}dx + \int_{200}^{\infty} x \cdot 0,01e^{-x\cdot 0,01}dx \\
&= 100(e^{-0.2} - e^{-0,5}) + 200(e^{-1} - e^{-2}) \\
&\quad - e^{-0,002x}(x + \frac{1}{0,002})\Big|_0^{100} - e^{-0,005x}(x + \frac{1}{0,005})\Big|_{100}^{200} - e^{-0,01x}(x + \frac{1}{0,01})\Big|_{200}^{\infty} \\
&\approx 151,8984.
\end{aligned}$$

$\triangleright$EXERCICE 13. Soit (X, Y) un vecteur aléatoire de densité :

$$f_{(X,Y)}(x, y) = \frac{3}{2}(x^2 + y^2)\mathbf{1}_{[0,1]^2}(x, y).$$

1. Calculer la fonction de répartition du vecteur (X, Y).
2. Calculer les densités marginales de X et de Y.
3. Les variables aléatoires X et Y sont-elles indépendantes ? Expliquer.
4. Calculer la matrice de variances-covariances du vecteur (X, Y).
5. Calculer l'espérance et la variance de la variable aléatoire $X + 2Y$.
6. Calculer la densité conditionnelle de X en $\{Y = y\}$, $y \in [0, 1]$.
7. Calculer l'espérance conditionnelle de X en $\{Y = 1/2\}$.
8. On observe X et on souhaite estimer Y. Donner la meilleure prédiction de Y sachant $\{X = 1/2\}$, par rapport à l'erreur quadratique.
9. Soit $V = X + 2Y$ et $W = X - Y$. Calculer la densité du vecteur (V, W).

SOLUTION.

1. La fonction de répartition du vecteur (X, Y) est donnée par

$$F_{(X,Y)}(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{(X,Y)}(t, v)dtdv.$$

- Pour $x < 0$ ou $y < 0$ on a $f_{(X,Y)}(t, v) = 0$, alors $F_{(X,Y)}(x, y) = 0$.
- Pour $(x, y) \in [0, 1]^2$ on a

$$F_{(X,Y)}(x, y) = \frac{3}{2}\int_0^x [\int_0^y (t^2 + v^2)dv]dt = \frac{3}{2}\int_0^x (t^2 y + \frac{y^3}{3})dt = \frac{1}{2}(x^3 y + xy^3).$$

- Pour $x > 1$ et $y \in [0, 1]$ on a

$$F_{(X,Y)}(x, y) = \frac{3}{2}\int_0^1 [\int_0^y (t^2 + v^2)dv]dt = \frac{3}{2}\int_0^1 (t^2 y + \frac{y^3}{3})dt = \frac{1}{2}(y + y^3).$$

- Pour $y > 1$ et $x \in [0,1]$, comme $f_{(X,Y)}(x,y)$ est symétrique en x et y on a

$$F_{(X,Y)}(x,y) = \frac{1}{2}(x + x^3).$$

- Pour $x > 1$ et $y > 1$ on a

$$F_{(X,Y)}(x,y) = \frac{3}{2}\int_0^1 [\int_0^1 (t^2 + v^2)dv]dt = 1,$$

car $f_{(X,Y)}(x,y)$ densité.
En conclusion, la fonction de répartition du vecteur (X,Y) est

$$F_{(X,Y)}(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ où } y < 0 \\ \frac{1}{2}(x^3y + xy^3) & \text{si } (x,y) \in [0,1]^2 \\ \frac{1}{2}(y + y^3) & \text{si } x > 1 \text{ et } y \in [0,1] \\ \frac{1}{2}(x + x^3) & \text{si } y > 1 \text{ et } x \in [0,1] \\ 1 & \text{si } x > 1 \text{ et } y > 1. \end{cases}$$

2. La densité marginale de X est donnée par

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f_{(X,Y)}(x,y)dy = \int_{\mathbb{R}} \frac{3}{2}(x^2 + y^2)\mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,1]}(y)dy \\ &= \frac{3}{2}\mathbf{1}_{[0,1]}(x)\int_0^1 (x^2 + y^2)dy = \frac{3}{2}(x^2 + \frac{1}{3})\mathbf{1}_{[0,1]}(x). \end{aligned}$$

Comme $f_{(X,Y)}(x,y)$ est symétrique en x et y, la densité marginale de Y est

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dx = \frac{3}{2}(y^2 + \frac{1}{3})\mathbf{1}_{[0,1]}(y).$$

3. Les variables aléatoires X et Y ne sont pas indépendantes car

$$f_{(X,Y)}(x,y) \neq f_X(x) \cdot f_Y(y).$$

4. La matrice de variances-covariances du vecteur (X,Y) est la matrice K donnée par

$$K = \begin{pmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{pmatrix}.$$

Comme les variables X et Y ont la même densité, elle sont identiquement distribuées, alors $\mathbb{E}(X) = \mathbb{E}(Y), \mathrm{Var}(X) = \mathrm{Var}(Y)$.
- Calculons $\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$.

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} x f_X(x)dx = \frac{3}{2}\int_0^1 (x^3 + \frac{1}{3}x)dx = \frac{3}{2}(\frac{1}{4} + \frac{1}{6}) = \frac{5}{8}. \\ \mathbb{E}(X^2) &= \int_{\mathbb{R}} x^2 f_X(x)dx = \frac{3}{2}\int_0^1 (x^4 + \frac{1}{3}x^2)dx = \frac{3}{2}(\frac{1}{5} + \frac{1}{9}) = \frac{7}{15}. \end{aligned}$$

On obtient la variance de X et de Y :

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{7}{15} - \frac{25}{64} = \frac{73}{960}.$$

- Calculons la covariance $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.
On sait que $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{5}{8}$.

Comme $XY = h(X,Y)$, avec $h : \mathbf{R}^2 \to \mathbf{R}$ définie par $h(x,y) = xy$, on obtient

$$\begin{aligned}
\mathbb{E}(XY) &= \int\int_{\mathbb{R}^2} h(x,y) f_{(X,Y)}(x,y)dxdy = \frac{3}{2}\int_0^1 [\int_0^1 (x^3y + xy^3)dx]dy \\
&= \frac{3}{2}\int_0^1 (\frac{1}{4}y + \frac{1}{2}y^3)dy = \frac{3}{2}(\frac{1}{8} + \frac{1}{8}) = \frac{3}{8}.
\end{aligned}$$

Alors

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X) = \frac{3}{8} - \frac{25}{64} = -\frac{1}{64}$$

et la matrice de variances-covariances du vecteur (X,Y) est

$$K = \begin{pmatrix} \frac{73}{960} & -\frac{1}{64} \\ -\frac{1}{64} & \frac{73}{960} \end{pmatrix}.$$

5.

$$\begin{aligned}
\mathbb{E}(X+2Y) &= \mathbb{E}(X) + 2\mathbb{E}(Y) = 3\mathbb{E}(X) = \frac{15}{8}. \\
\mathrm{Var}(X+2Y) &= \mathrm{Var}(X) + \mathrm{Var}(2Y) + 2\mathrm{Cov}(X,2Y) \\
&= \mathrm{Var}(X) + 4\mathrm{Var}(Y) + 4\mathrm{Cov}(X,Y) \\
&= 5\frac{73}{960} - 4\frac{1}{64} = \frac{61}{192}.
\end{aligned}$$

6. La densité conditionnelle de X en $\{Y = y\}, y \in [0,1]$ est

$$f_{X|Y}(x \mid y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \frac{\frac{3}{2}(x^2+y^2)\mathbf{1}_{[0,1]}(x)}{\frac{3}{2}(y^2+\frac{1}{3})} = \frac{(x^2+y^2)}{y^2+\frac{1}{3}}\mathbf{1}_{[0,1]}(x).$$

7.

$$\begin{aligned}
\mathbb{E}(X \mid Y = \frac{1}{2}) &= \int_{\mathbb{R}} x f_{X|Y}(x \mid \frac{1}{2})dx \\
&= \int_0^1 x\frac{(x^2+\frac{1}{4})}{\frac{1}{4}+\frac{1}{3}}dx \\
&= \frac{12}{7}\int_0^1 (x^3 + \frac{1}{4}x)dx \\
&= \frac{12}{7}\frac{3}{8} = \frac{9}{14}.
\end{aligned}$$

8. La meilleure prédiction de Y sachant $\{X = 1/2\}$, par rapport à l'erreur quadratique est l'espérance conditionnelle de Y en $\{X = 1/2\}$. En raison de la symétrie entre X et Y on a

$$\mathbb{E}(Y \mid X = \frac{1}{2}) = \mathbb{E}(X \mid Y = \frac{1}{2}) = \frac{9}{14}.$$

9. Soit l'application $g : [0,1]^2 \to \mathbf{R}^2$ définie par

$$g(x,y) = (x+2y, x-y) = (v,w).$$

Nous avons $g(X,Y) = (V,W)$. Pour calculer la fonction inverse de g, on exprime x et y en fonction des v et de w et on obtient

$$g^{-1}(v,w) = (\frac{v+2w}{3}, \frac{v-w}{3}).$$

Le système des inéquations $0 \leq \frac{v+2w}{3} \leq 1, \quad 0 \leq \frac{v-w}{3} \leq 1$ nous donne le domaine de définition de g^{-1},

$$g^{-1} : D \to [0,1]^2,$$

où $D := \{(v,w) \in \mathbf{R}^2 \mid 0 \leq \frac{v+2w}{3} \leq 1, \ 0 \leq \frac{v-w}{3} \leq 1\}$.
Le jacobien de g^{-1} est

$$DJ_{g^{-1}}(v,w) = \text{ dét } \begin{vmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{vmatrix} = -\frac{1}{3}.$$

Alors $g : [0,1]^2 \to D$ est un difféomorphisme et on obtient que la densité du vecteur $(V,W) = g(X,Y)$ est donnée par

$$\begin{aligned} f_{(V,W)}(v,w) &= f_{(X,Y)}(g^{-1}(v,w)) \mid DJ_{g^{-1}}(v,w) \mid \\ &= \frac{3}{2}[(\frac{v+2w}{3})^2 + (\frac{v-w}{3})^2]\mathbf{1}_{[0,1]^2}(\frac{v+2w}{3}, \frac{v-w}{3})\frac{1}{3} \\ &= \frac{1}{18}(2v^2 + 5w^2 + 2vw)\mathbf{1}_D(v,w). \end{aligned}$$

$\triangleright$EXERCICE 14. Soit X un vecteur aléatoire dans $\mathbf{R}^d$ et soit $E \subset \mathbf{R}^d$ tel que $\mathbf{P}(X \in E) = 1$ et $\mu = \mathbb{E}(X), \mathbb{E}\,\|X\| < \infty$. Soit $g : E \to \mathbf{R}$ continûment différentiable et considérons la variable aléatoire $g(X)$. En utilisant un développement de Taylor d'ordre 1, donner une approximation de $\mathbb{E}(g(X))$ et de $\mathrm{Var}(g(X))$.
APPLICATION. Soit le vecteur aléatoire $(X,Y)'$ de loi $\mathcal{N}_2(0, I\sigma^2)$, et $g(x,y) = x^2 + y^2 - xy$. Calculer $\mathbb{E}(g(X,Y))$ et $\mathrm{Var}(g(X,Y))$.

SOLUTION. Pour $x \in E$, on a le développement de Taylor d'ordre 1 en μ pour la fonction g

$$g(x) = g(\mu) + \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)(x_i - \mu_i) + o(\|x - \mu\|),$$

où $o : \mathbf{R} \to \mathbf{R}$ est une fonction qui vérifie $\lim_{t\to 0} \frac{o(t)}{t} = 0$. On obtient la valeur approchée de $g(x)$

$$g(x) \approx g(\mu) + \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)(x_i - \mu_i), x \in E.$$

En remplaçant $x \in E$ avec la variable aléatoire $X \quad (\mathbf{P}(X \in E) = 1)$, nous avons une valeur approchée de $g(X)$

$$g(X) \approx g(\mu) + \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)(X_i - \mu_i). \tag{8.2}$$

En appliquant l'espérance sur l'équation (8.2) on obtient une valeur approchée de $\mathbb{E}(g(X))$:

$$\begin{aligned} \mathbb{E}(g(X)) &\approx \mathbb{E}(g(\mu)) + \sum_{i=1}^{d} \mathbb{E}(\frac{\partial g}{\partial x_i}(\mu)(X_i - \mu_i)) \\ &\approx g(\mu) + \sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)\mathbb{E}(X_i - \mu_i) \\ &\approx g(\mu), \end{aligned}$$

car $\mathbb{E}(\cdot)$ linéaire et $\mathbb{E}(X_i) = \mu_i$.

En appliquant la variance sur l'équation (8.2) on obtient une valeur approchée de $\text{Var}(g(X))$:

$$\begin{aligned}
\text{Var}(g(X)) &\approx \text{Var}(g(\mu)) + \text{Var}(\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)(X_i - \mu_i)) \\
&\approx 0 + \mathbb{E}[(\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)(X_i - \mu_i))^2] - [\mathbb{E}(\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)(X_i - \mu_i))]^2 \\
&\approx 0 + \mathbb{E}[\sum_{i=1}^{d}\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)\frac{\partial g}{\partial x_j}(\mu)(X_i - \mu_i)(X_j - \mu_j)] - 0 \\
&\approx \sum_{i=1}^{d}\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)\frac{\partial g}{\partial x_j}(\mu)\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \\
&\approx \sum_{i=1}^{d}\sum_{i=1}^{d} \frac{\partial g}{\partial x_i}(\mu)\frac{\partial g}{\partial x_j}(\mu)\text{Cov}(X_i, X_j).
\end{aligned}$$

APPLICATION

Comme $(X,Y) \sim \mathcal{N}_2(0, I\sigma^2)$, on a $\mu = \mathbb{E}(X,Y) = (0,0)$. Les dérivées partielles de g sont

$$\frac{\partial g}{\partial x}(x,y) = (2x - y), \quad \frac{\partial g}{\partial y}(x,y) = (2y - x),$$

$$\frac{\partial g}{\partial x}(0,0) = 0, \quad \frac{\partial g}{\partial y}(0,0) = 0.$$

En utilisant les formules générales obtenues, on a

$$\mathbb{E}(g(X)) \approx 0, \quad \text{Var}(g(X)) \approx 0.$$

▷EXERCICE 15. (Loi de Dirichlet). Soit $X_1, \ldots, X_{n+1}$ de variables aléatoires indépendantes telles que $X_i \sim \gamma(\alpha_i, \beta)$, pour $1 \leq i \leq n+1$. Soit les variables aléatoires $S_{n+1} := X_1 + \ldots + X_{n+1}$ et $V_i := X_i / S_{n+1}, 1 \leq i \leq n, V_{n+1} := S_{n+1}$.

1. Trouver l'application $g : \mathbf{R}^{n+1} \to \mathbf{R}^{n+1}$ qui donne les variables aléatoires V_i en fonction de X_i et calculer le jacobien de g^{-1}.
2. Montrer que $(V_1, \ldots, V_n)$, et S_{n+1} sont indépendantes.
3. Montrer que la densité de $V = (V_1, \ldots, V_n)$ est

$$f_V(v_1, \ldots, v_n) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_{n+1})}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_{n+1})} v_1^{\alpha_1 - 1} \ldots v_n^{\alpha_n - 1}(1 - \sum_{i=1}^{n} v_i)^{\alpha_{n+1}-1}\mathbf{1}_{A_n}((v_1, \ldots, v_n)),$$

où $A_n := \{(v_1, \ldots, v_n) \in \mathbf{R}^n \mid v_i \geq 0, i = 1, \ldots, n, \sum_{i=1}^{n} v_i \leq 1\}$.

SOLUTION.

1. Soit l'application $g : [0; +\infty)^{n+1} \to \mathbf{R}^{n+1}$ définie par

$$g(x_1, \ldots, x_{n+1}) = (\frac{x_1}{x_1 + \ldots + x_{n+1}}, \ldots, \frac{x_n}{x_1 + \ldots + x_{n+1}}, x_1 + \ldots + x_{n+1}) = (v_1, \ldots, v_{n+1}).$$

Nous avons $g(X_1, \ldots, X_{n+1}) = (V_1, \ldots, V_{n+1})$. Pour calculer la fonction inverse de g, on exprime les x_i en fonction des $v_i, 1 \leq i \leq n+1$, et on obtient

$$g^{-1}(v_1, \ldots, v_{n+1}) = (v_1 v_{n+1}, \ldots, v_n v_{n+1}, (1 - v_1 - \ldots - v_n)v_{n+1}).$$

En résolvant le système des inéquations $v_1 v_{n+1} \geq 0, \ldots, v_n v_{n+1} \geq 0, (1 - v_1 - \ldots - v_n)v_{n+1} \geq 0$ on trouve le domaine de définition de g^{-1},

$$g^{-1} : A_n \times [0; +\infty[\to [0; +\infty[^{n+1},$$

où $A_n := \{(v_1, \ldots, v_n) \in \mathbf{R}^n \mid v_i \geq 0, i = 1, \ldots, n, \sum_{i=1}^n v_i \leq 1\}$.

Le jacobien de g^{-1} est

$$DJg^{-1} = \text{dét} \begin{vmatrix} v_{n+1} & 0 & 0 & \ldots & 0 & v_1 \\ 0 & v_{n+1} & 0 & \ldots & 0 & v_2 \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ 0 & \ldots & \ldots & 0 & v_{n+1} & v_n \\ -v_{n+1} & \ldots & \ldots & \ldots & -v_{n+1} & 1 - v_1 - \ldots - v_n \end{vmatrix}.$$

En ajoutant les premières n lignes à la dernière ligne, on obtient

$$DJg^{-1} = \text{dét} \begin{vmatrix} v_{n+1} & 0 & 0 & \ldots & 0 & v_1 \\ 0 & v_{n+1} & 0 & \ldots & 0 & v_2 \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ 0 & \ldots & \ldots & 0 & v_{n+1} & v_n \\ 0 & \ldots & \ldots & \ldots & 0 & 1 \end{vmatrix} = v_{n+1}^n.$$

2. Comme les variables aléatoires $X_i, 1 \leq i \leq n+1$, sont indépendantes et $X_i \sim \gamma(\alpha_i, \beta)$, la densité du vecteur $(X_1, \ldots, X_{n+1})$ est

$$\begin{aligned} f_{(X_1,\ldots,X_{n+1})}(x_1, \ldots, x_{n+1}) &= f_{X_1}(x_1) \ldots f_{X_{n+1}}(x_{n+1}) \\ &= \prod_{i=1}^{n+1} \frac{\beta^{\alpha_i}}{\Gamma(\alpha_i)} e^{-\beta x_i} x_i^{\alpha_i - 1} \mathbf{1}_{[0;+\infty)}(x_i) \\ &= \frac{\beta^{\alpha_1 + \ldots + \alpha_{n+1}}}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_{n+1})} e^{-\beta(x_1 + \ldots + x_{n+1})} \prod_{i=1}^{n+1} x_i^{\alpha_i - 1} \mathbf{1}_{[0;+\infty)}(x_i). \end{aligned}$$

$g : [0; +\infty)^{n+1} \to A_n \times [0; +\infty)$ est un difféomorphisme et on obtient que la densité du vecteur $(V_1, \ldots, V_n, S_{n+1}) = (V_1, \ldots, V_{n+1})$ est donnée par

$$\begin{aligned} & f_{(V_1,\ldots,V_n,S_{n+1})}(v_1, \ldots, v_{n+1}) \\ = \; & f_{(X_1,\ldots,X_{n+1})}(g^{-1}(v_1, \ldots, v_{n+1})) \left| DJg^{-1}(v_1, \ldots, v_{n+1}) \right| \\ = \; & \frac{\beta^{\alpha_1 + \ldots + \alpha_{n+1}}}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_{n+1})} e^{-\beta(v_1 v_{n+1} + \ldots + v_n v_{n+1} + (1 - \sum_{i=1}^n v_i) v_{n+1})} \\ & \times \prod_{i=1}^n (v_i v_{n+1})^{\alpha_i - 1} [(1 - \sum_{i=1}^n v_i) v_{n+1}]^{\alpha_{n+1} - 1} \mathbf{1}_{A_n}((v_1, \ldots, v_n)) \mathbf{1}_{[0;+\infty)}(v_{n+1}) \\ = \; & \frac{1}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_{n+1})} \prod_{i=1}^n v_i^{\alpha_i - 1} (1 - \sum_{i=1}^n v_i)^{\alpha_{n+1} - 1} \mathbf{1}_{A_n}((v_1, \ldots, v_n)) \\ & \times \beta^{\alpha_1 + \ldots + \alpha_{n+1}} e^{-\beta v_{n+1}} v_{n+1}^{\alpha_1 + \ldots + \alpha_{n+1} - 1} \mathbf{1}_{[0;+\infty)}(v_{n+1}). \end{aligned}$$

Comme la densité du vecteur $(V_1, \ldots, V_n, S_{n+1})$ s'écrit sous la forme

$$f_{(V_1,\ldots,V_n,S_{n+1})}(v_1, \ldots, v_{n+1}) = h_1(v_1, \ldots, v_n) h_2(v_{n+1}),$$

on en déduit que $(V_1, \ldots, V_n)$, et S_{n+1} sont indépendantes.

3. La densité de $(V_1, \ldots, V_n)$ est donnée par

$$\begin{aligned} f_{(V_1,\ldots,V_n)}(v_1, \ldots, v_n) &= \int_{\mathbf{R}} f_{(V_1,\ldots,V_n,S_{n+1})}(v_1, \ldots, v_{n+1}) dv_{n+1} \\ = & \frac{1}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_{n+1})} \prod_{i=1}^{n} v_i^{\alpha_i - 1} (1 - \sum_{i=1}^{n} v_i)^{\alpha_{n+1}-1} \mathbf{1}_{A_n}((v_1, \ldots, v_n)) \\ & \times \int_0^{+\infty} \beta^{\alpha_1+\ldots+\alpha_{n+1}} e^{-\beta v_{n+1}} v_{n+1}^{\alpha_1+\ldots+\alpha_{n+1}-1} dv_{n+1}. \end{aligned}$$

En faisant le changement de variable $\beta v_{n+1} = t$ et en utilisant la définition de la fonction gamma, $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$, on obtient la densité de $(V_1, \ldots, V_n)$

$$f_{(V_1,\ldots,V_n)}(v_1, \ldots, v_n) = \frac{\Gamma(\alpha_1 + \ldots + \alpha_{n+1})}{\Gamma(\alpha_1)\ldots\Gamma(\alpha_{n+1})} \prod_{i=1}^{n} v_i^{\alpha_i - 1} (1 - \sum_{i=1}^{n} v_i)^{\alpha_{n+1}-1} \mathbf{1}_{A_n}((v_1, \ldots, v_n)).$$

$\triangleright$EXERCICE 16. Une liste de n éléments, notés $1, 2, ..., n$, sont stockés dans la mémoire d'un ordinateur. La consultation de l'élément j $(1 \leq j \leq n)$ s'effectue en parcourant la liste - commençant par le premier élément à chaque fois - , et le temps d'accès est j unités de temps. L'élément i est consulté indépendamment du passé avec probabilité p_i, $\sum_{i=1}^n p_i = 1$. Ces probabilités sont supposées connues. On cherche à déterminer le stockage qui minimise le temps moyen de recherche d'un élément.

Soient σ_0 la permutation $(1, 2, ..., n)$ correspondant à $p_1 \geq p_2 \geq \cdots \geq p_n$, et σ une permutation $(i_1, i_2, ..., i_n)$ quelconque. Notons N la variable aléatoire désignant la position d'un élément quelconque consulté, et $\sum$ la variable aléatoire désignant la permutation choisie pour ranger les n éléments.

1. Comparer les probabilités $\mathbb{P}(N \geq k \mid \sum = \sigma_0)$ et $\mathbb{P}(N \geq k \mid \sum = \sigma)$.
2. En déduire une comparaison entre $\mathbb{E}(N \mid \sum = \sigma_0)$ et $\mathbb{E}(N \mid \sum = \sigma)$, et donner la permutation optimale.

Remarque. On pourra utiliser la formule : $\mathbb{E}[X \mid A] = \sum_{k \geq 1} \mathbb{P}(X \geq k \mid A)$, où X est une v.a. à valeurs dans $\mathbb{N}$ et A est un événement.

SOLUTION.

1.

$$\mathbb{P}(N \geq k \mid \sum = \sigma_0) = \mathbb{P}(\bigcup_{j=k}^{n} \{N = j\} \mid \sum = \sigma_0) = \sum_{j=k}^{n} \mathbb{P}(N = j \mid \sum = \sigma_0) = \sum_{j=k}^{n} p_j$$

D'autre part, on a

$$\mathbb{P}(N \geq k \mid \sum = \sigma) = \mathbb{P}(\bigcup_{j=k}^{n} \{N = j\} \mid \sum = \sigma) = \sum_{j=k}^{n} \mathbb{P}(N = j \mid \sum = \sigma) = \sum_{j=k}^{n} p_{\sigma_{i_j}}$$

Comme $p_1 \geq p_2 \geq \cdots \geq p_n$, on obtient

$$\sum_{j=k}^{n} p_j \leq \sum_{j=k}^{n} p_{\sigma_{i_j}}$$

pour toute permutation $\sigma = (i_1, i_2, ..., i_n)$ et on déduit

$$\mathbb{P}(N \geq k \mid \sum = \sigma_0) \leq \mathbb{P}(N \geq k \mid \sum = \sigma).$$

2. En utilisant la formule : $\mathbb{E}[X \mid A] = \sum_{k\geq 1} \mathbb{P}(X \geq k \mid A)$, pour toute permutation σ on a

$$\sum_{k\geq 1} \mathbb{P}(N \geq k \mid \sum = \sigma_0) \leq \sum_{k\geq 1} \mathbb{P}(N \geq k \mid \sum = \sigma)$$
$$\mathbb{E}(N \mid \sum = \sigma_0) \leq \mathbb{E}(N \mid \sum = \sigma).$$

La permutation est optimale si le temps moyen de recherche est minimum. L'inégalité qu'on vient de démontrer montre que $\mathbb{E}(N \mid \sum = \sigma)$ est minimale pour $\sigma = \sigma_0$, donc la permutation optimale est σ_0.

▷EXERCICE 17. On souhaite, à l'aide de Monte Carlo, calculer une "bonne" approximation de l'aire de la surface délimitée par la parabole $y = x^2 - 1$ et la droite $y = 3$. Appelons S cette surface.

1. La fonction RAND de votre ordinateur vous fournit, à chaque appel, une réalisation d'un nombre aléatoire dans $[0,1]$. Utiliser cette fonction pour générer un nombre aléatoire dans $[-2,2]$.

2. A l'aide de 1., écrire un algorithme qui génère des points aléatoires dans le réctangle $[-2,2] \times [-1,3]$.

3. A l'aide de 2. et de la *loi empirique des grands nombres*, écrire un algorithme pour calculer l'aire de S. Executer l'algorithme pour $N = 10000$ points.

4. Calculer l'aire de S analytiquement.

5. Représenter sur un graphe la courbe de l'évolution de la fréquence et sa convergence vers la valeur exacte.

SOLUTION.

1. Comme $0 \leq RAND \leq 1$ fournit un nombre aléatoire dans $[0,1]$, $4 * RAND$ fournit un nombre aléatoire dans $[0,4]$ et ensuite $4 * RAND - 2$ c'est la fonction cherchée, qui fournit un nombre aléatoire dans $[-2,2]$.
2.
```
x ← 4 * RAND - 2
y ← 4 * RAND - 1
Afficher (x,y)
```
3.
```
Initialisation: s=compteur; N=nombre d'iterations
Pour i=1 jusqu'à N faire
    x=4 * RAND - 2;
    y=4 * RAND - 1;
    Si y >=x^2 -1 faire
        s=s+1;
    Fin Si
Fin Pour
Afficher Aire = 16*s/N
```
4. L'aire de S est donnée par

$$A(S) = \int_{-1}^{3} \Big[\int_{-\sqrt{y+1}}^{\sqrt{y+1}} dx \Big] dy = \int_{-1}^{3} 2\sqrt{y+1} dy = 32/3.$$

5. La courbe demandée est donnée dans la Figure **??**. On remarque la convergence de l'aire simulée vers la valeur exacte.

Exercices 9

EXERCICES A RESOUDRE

▷EXERCICE 1. Dans une urne on dispose de trois boules : Une boule bleue, une rouge et une autre verte. Donnez l'espace fondamental pour les expériences suivantes :

(a) On tire une seule boule.

(b) On tire deux boules sans remise.

(c) On tire trois boules sans remise.

(d) On tire des boules avec remise, jusqu'à ce qu'on obtienne la boule rouge.

▷EXERCICE 2. Soit Ω l'espace fondamental de lancement de deux dés distinguables et soit la tribu $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$, sur Ω, avec $A = \{(1,2),(2,1)\}$. Considérons l'application X sur Ω définie par la formule

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2, \quad (\omega_1, \omega_2) \in \Omega.$$

Montrer que X n'est pas une v.a. sur $(\Omega, \mathcal{F})$.

▷EXERCICE 3. Deux personnes A et B jouent aux échecs, au maximum 7 tours. La personne qui gagne quatre tours, gagnera le jeu. On suppose qu'il y a un gagnant à chaque tour. Si la personne A joue en premier, la probabilité qu'elle gagne est de 0.7, dans le cas contraire elle est de 0.5. Les deux personnes se sont mises d'accord que ce soit B qui joue en premier pour les deux premiers tours, ensuite c'est A qui jouera en premier pour les trois tours suivants et enfin, pour les tours qui restent, c'est B qui commencera. On suppose que les joueurs ne sont pas influencés par leur victoire ou celle de leur adversaire et donc les jeux sont indépendants. Calculer alors la probabilité que :

(a) Le tournoi se termine au quatrième tour et que le gagnant soit A ;

(b) A rapporte le tournoi au cinquième tour ;

(c) A gagne le jeu avant le sixième tour ;

(d) A gagne le tournoi.

▷EXERCICE 4. Une chaîne Hi-Fi comprend un tuner, un lecteur de disques, un amplificateur et deux enceintes. La chaîne Hi-Fi marche quand le tuner ou le lecteur de disques marche, l'amplificateur marche, et au moins l'une des deux enceintes marche. Tous les composants sont supposés fonctionner indépendemment les uns des autres. La marche (ou la panne) de la chaîne est considérée dans le sens qu'on puisse (ou non) entendre du son.

1. Exprimer l'événement A = "La Hi-Fi marche", en fonction des événements C_i = "Le composant i marche".

2. Calculer la probabilité $\mathbb{P}(A)$.

3. Application numérique : $\mathbb{P}(C_1) = 0.8, \mathbb{P}(C_2) = 0.85, \mathbb{P}(C_3) = 0.7$ et $\mathbb{P}(C_4) = \mathbb{P}(C_5) = 0.6$.

$\triangleright$EXERCICE 5. Soient deux variables aléatoires X et Y, indépendantes et telles que : X $\sim$ P(λ) et Y $\sim$ P(μ).

(a) Calculer la fonction génératrice de la v.a. X.

(b) Calculer l'espérance mathématique et la variance de X à l'aide du résultat obtenu en (a).

(c) Montrer que si Z = X + Y, alors Z $\sim$ P($\lambda + \mu$).

$\triangleright$EXERCICE 6. Un système est formé de 2 composants en série. La durée de vie de chaque composant, soit $T_i, i = 1, 2$, suit une loi géometrique sur $\mathbb{N}$ de paramètre $p_i, i = 1, 2$. Les temps de bon fonctionnement sont indépendants entre eux. La fiabilité d'un système au temps $n \geq 0$ est définie par

$$R(n) = P(T > n),$$

où T est sa durée de vie.

(a) Calculer la fiabilité au temps n.

(b) Application numérique : $p_i = 0,01, i = 1, 2$. Donner le graphe de la fiabilité pour $0 \leq n \leq 100$.

$\triangleright$EXERCICE 7. Soit X une v.a., $X \sim U[-1, 1]$. Trouver une fonction $g : [-1, 1] \to \mathbb{R}_+$, telle que la v.a. $g(X) \sim \mathcal{E}(\lambda), \lambda > 0$.

$\triangleright$EXERCICE 8. Soit X une variable aléatoire de densité f définie par $f(x) = xe^{-\frac{x^2}{2}}$ si $x \geq 0$ et $f(x) = 0$ sinon.

(a) Vérifier que f est une densité de probabilité.

(b) Montrer que $Y = X^2$ est une variable aléatoire dont on donnera une densité.

(c) Calculer l'espérance et la variance de Y.

$\triangleright$EXERCICE 9. Un homme tirant sur une cible reçoit 10 points si son coup est à moins de 1 cm du centre de la cible, 5 points s'il s'en éloigne de 1 à 3 cm et 3 points s'il s'en éloigne de 3 à 5 cm. Trouver l'espérance du nombre de points si la distance au centre de la cible est uniformément distribué entre 0 et 10.

$\triangleright$EXERCICE 10.

(a) Une caserne de pompiers doit être construite sur une route de longueur A, $A < \infty$. Si un incendie se déclare en des points uniformément distribués sur $[0, A]$, où doit être située la caserne pour minimiser l'espérance de la distance jusqu'au feu ? Autrement dit, trouver a tel que $\mathbb{E}(\mid X - a \mid)$ soit minimisée lorsque X est distribué uniformément sur $[0, A]$.

(b) Supposer à présent que la route soit de longueur infinie - partant d'un point 0 vers $+\infty$. Si la distance d'un incendie au point 0 est distribuée selon une loi exponentielle de paramètre λ, où doit se trouver la caserne ? Ici, on cherche à minimiser $\mathbb{E}(\mid X - a \mid)$, où X est exponentielle de paramètre λ.

$\triangleright$EXERCICE 11. Soit X un vecteur aléatoire dans $\mathbb{R}^d$ et soit $E \subset \mathbb{R}^d$ tel que $\mathbb{P}(X \in E) = 1$ et $\mu = \mathbb{E}(X), \mathbb{E}\,\|X\| < \infty$. Soit $g : E \to \mathbb{R}$ continûment différentiable et considérons la variable aléatoire $g(X)$. En utilisant un développement de Taylor d'ordre 1, donner une approximation de $\mathbb{E}(g(X))$ et de $\mathrm{Var}(g(X))$.

APPLICATION. Soit le vecteur aléatoire $(X, Y)'$ de loi $\mathcal{N}_2(0, I\sigma^2)$, et $g(x, y) = x^2 + y^2 - xy$. Calculer $\mathbb{E}(g(X, Y))$ et $\mathrm{Var}(g(X, Y))$.

$\triangleright$EXERCICE 12. **Estimation de la valeur de π par la méthode de** *Monte Carlo*. On considère le carré $\mathcal{C}$ ayant comme sommets les points $A(0,0), B(1,0), C(1,1), D(0,1)$ et le quart de cercle de centre $A(0,0)$ et de rayon 1, inscrit dans le carré. On fait N tirages successifs des points uniformément distribués dans $\mathcal{C}$. Notons A_i l'événement "le i-ème point est tiré à l'intérieur du quart de cercle", $i = 1, \dots, N$ et (X_i, Y_i), $i = 1, \dots, N$, les coordonnées du i-ème point tiré.

(a) Donnez à l'aide des A_i une variable aléatoire S_N qui décrit le nombre de point à l'intérieur du cercle sur le nombre total N de tirages.

(b) Montrer que S_N/N converge *p.s.* vers $\pi/4$, quand N tend vers l'infini.

(c) A l'aide de l'inégalité de Tchebichev, calculer le nombre minimum de tirages N que l'on doit effectuer, pour que la probabilité de s'écarter de la valeur de π de plus de 10^{-2}, soit inférieure à 10^{-2}.

(d) Ecrire un programme et effectuer le calcul de π.

$\triangleright$EXERCICE 13. (Médian A03). On lance une pièce de monnaie équilibrée n fois. Soit les v.a. discrètes X et Y, dont $X =$ "nombre de piles" et $Y =$ "nombre de faces" observés, et on pose $Z = X - Y$.

(a) Préciser l'ensemble des valeurs de de la v.a. Z et donner sa loi. (Prendre en compte la parité de n).

(b) Calculer $\mathbb{E}(Z)$.

$\triangleright$EXERCICE 14. (Médian A03). Considérons l'axe des entiers relatifs $\mathbb{Z}$, et une particule se déplaçant sur les entiers en partant de zéro. A chaque unité de temps il effectue un mouvement soit vers la droite de $(+1)$ avec probabilité p, soit vers la gauche de (-1) avec probabilité $1 - p$, où $p \in]0,1[$.
Soit S_n, pour $n \in \mathbb{N}$, la position de la particule au temps n. Soit $X_1, ..., X_n$ une suite de v.a. i.i.d. à valeurs $\{-1, +1\}$ et de loi commune $P_{X_1}(-1) = 1 - p$ et $P_{X_1}(+1) = p$.
Pour un $n \geq 1$ fixé, calculer :

(a) S_n en fonction de $X_1, ..., X_n$;

(b) la position moyenne de la particule au temps $n \geq 1$ et sa variance, c'est-à-dire $\mathbb{E}[S_n]$ et $\mathrm{Var}(S_n)$;

(c) la probabilité de retour en 0, au temps n, $\mathbb{P}(S_n = 0)$;

(d) la probabilité : $\mathbb{P}(X_k \neq X_{k+1})$ pour $1 \leq k \leq n - 1$;

(e) le nombre de changements de direction, noté N, en fonction des v.a. $\mathbf{1}_{\{X_k \neq X_{k+1}\}}$, pour $1 \leq k \leq n - 1$;

(f) $\mathbb{E}(N)$.

$\triangleright$EXERCICE 15. (Final A03). Une liste de n éléments, notés $1, 2, ..., n$, sont stockés dans la mémoire d'un ordinateur. La consultation de l'élément j ($1 \leq j \leq n$) s'effectue en parcourant la liste - commençant par le premier élément à chaque fois - , et le temps d'accès est j unités de temps. L'élément i est consulté indépendamment du passé avec probabilité p_i, $\sum_{i=1}^{n} p_i = 1$. Ces probabilités sont supposées connues. On cherche à déterminer le stockage qui minimise le temps moyen de recherche d'un élément.
Soient σ_0 la permutation $(1, 2, ..., n)$ correspondant à $p_1 \geq p_2 \geq \cdots \geq p_n$, et σ une permutation $(i_1, i_2, ..., i_n)$ quelconque. Notons N la variable aléatoire désignant la position d'un élément quelconque consulté, et $\sum$ la variable aléatoire désignant la permutation choisie pour ranger les n éléments.

(a) Comparer les probabilités $\mathbb{P}(N \geq k \mid \sum = \sigma_0)$ et $\mathbb{P}(N \geq k \mid \sum = \sigma)$.

(b) En déduire une comparaison entre $\mathbb{E}(N \mid \sum = \sigma_0)$ et $\mathbb{E}(N \mid \sum = \sigma)$, et donner la permutation optimale.

Remarque. On pourra utiliser la formule : $\mathbb{E}[X \mid A] = \sum_{k\geq 1} \mathbb{P}(X \geq k \mid A)$, où X est une v.a. à valeurs dans $\mathbb{N}$ et A est un événement.

$\triangleright$EXERCICE 16. (Final A04). L'écart X de la longueur nominale d'une barre suit une loi normale $\mathcal{N}(\mu, \sigma^2)$. Les barres dont l'écart est compris entre a et b ($a < b$) sont dites conformes à la norme française (NF) et sont retenues, les autres sont rejetées pour être recyclées.

(a) Donner la fonction de répartition de l'écart d'une barre conforme à la NF.

(b) A.N. Si $\mu = 0$ et $\sigma = 1$, avec $a = -1$, et $b = 1$, donner la probabilité que l'écart d'une barre conforme à la NF soit supérieur à $0,71$. (Utiliser la table de la loi normale).

$\triangleright$EXERCICE 17. (Final A04). Les v.a. X et Y sont i.i.d. de loi commune exponentielle de paramètre λ, ($\lambda > 0$), i.e. de densité f :

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Soit la v.a. $Z := X + Y$.

(a) Calculer la densité de Z.

(b) Si U est une v.a. uniforme sur $(0, 1)$, donner la f.r. de la v.a.

$$V := -\frac{1}{\lambda} \ln U,$$

(c) Proposer une v.a. W comme fonction de v.a. uniformes indépendantes ayant la même loi que Z.

$\triangleright$EXERCICE 18. (Final A04). Soit un vecteur aléatoire (X, Y) dans $\mathbb{R}^2$ désignant les coordonnées d'un point de loi uniforme sur l'ellipse :

$$D = \{(x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\},$$

où $a, b > 0$. L'aire de l'ellipse est égale à πab.

(a) Donner la densité jointe de (X, Y), notée $f_{(X,Y)}$.

(b) On pose :

$$X = aR\cos\Theta, \quad Y = bR\sin\Theta,$$

avec $-\pi < \Theta \leq \pi$ et $R > 0$.
Déterminer la densité de probabilité jointe $f_{(R,\Theta)}$ du vecteur aléatoire (R, Θ)

(c) Déterminer les densités marginales respectives f_R et f_Θ des v.a. R et Θ.

(d) Déterminer les densités conditionnelles $f_{R|\Theta}(r \mid \theta)$ et $f_{\Theta|R}(\theta \mid r)$.

(e) Calculer $\mathbb{E}[R \mid \Theta = \theta]$ et $\mathbb{E}[R \mid \Theta]$.

Annexes 10

Annexes

Annexe A

Annexe B

Annexe C

TABLE 10.1 – Notations

$\mathbb{E}$	espérance mathématique
$g(s)$	fonction génératrice
$M(t)$	fonction génératrice des moments
$\varphi(t)$	fonction caractéristique
Ω	espace fondamental d'une expérience aléatoire
ω	issue d'une expérience aléatoire
$\mathcal{F}$	tribu ou σ-algèbre d'événements
$\mathcal{P}(\Omega)$	l'ensemble de tous les sous ensembles de Ω
$\mathbb{P}$	probabilité
$\mathbf{1}_A$	fonction indicatrice de l'événement A
μ_k	moment d'ordre k non-centré
m_k	moment d'ordre k centré
P_X	loi de probabilité de la v.a. X
F	fonction de répartition d'une v.a.
f	fonction de densité de probabilité d'une v.a.
Var ou σ^2	variance d'une v.a.
$\mathrm{Cov}(X,Y)$	covariance de deux v.a. X,Y
K	matrice de variances-covariances
ρ	coefficient de corrélation linéaire
$f_{Y\|X}(\cdot\|x)$	densité conditionnelle de la v.a. Y en $\{X = x\}$

TABLE 10.2 – Lois de probabilités

$B(p)$	loi de Bernoulli de paramètre p
$b(n,p)$	loi binomiale de paramètres n et p
$bn(m,p)$	loi binomiale négative de paramètres m et p
$E(a)$	loi exponentielle de paramètre a
$G(p)$	loi géométrique de paramètre p
$N(0,1)$	loi normale ou de Gauss centrée réduite
$N(a,b)$	loi normale de moyenne a et de variance b
$P(\lambda)$	loi de Poisson de paramètre λ
$U[a,b]$	loi uniforme sur l'intervalle $[a,b]$
$W(a,b)$	loi de Weibull à deux paramètres a et b
$\chi^2(n)$	loi de chi-2 à n degrés de liberté
$\gamma(a,b)$	loi gamma de paramètres a et b

TABLE 10.3 – Convergences stochastiques

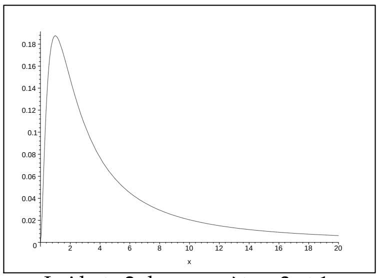
$\overset{p}{\to}$	convergence en probabilité
$\overset{p.s.}{\to}$	convergence presque sûre
$\overset{m.q.}{\to}$	convergence en moyenne quadratique
$\overset{L^p}{\to}$	convergence au sense de l'espace L^p
$\overset{\mathcal{L}}{\to}$	convergence en loi

TABLE 10.4 – Sigles

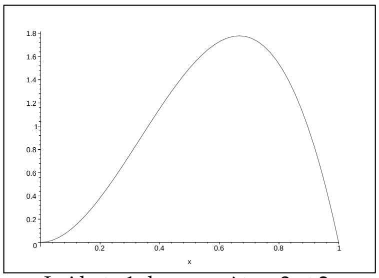
i.i.d.	indépendant(e)s et identiquement distribué(e)s
p.s.	presque sûr(e)
ssi	si et seulement si
v.a.	variable aléatoire
v.a.d.	variable aléatoire discrète
v.a.r.	variable aléatoire réelle
v.a.v.	variable aléatoire vectorielle

TABLE 10.5 – Alphabet grec

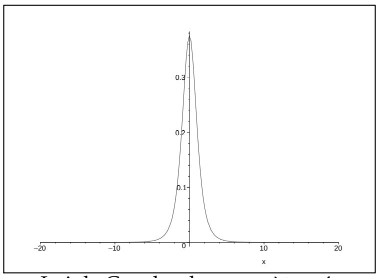
A	α	Alpha
B	β	Beta
Γ	γ	Gamma
Δ	δ	Delta
E	ε (ϵ)	Epsilon
Z	ζ	Zeta
H	η	Eta
Θ	ϑ (θ)	Theta
I	ι	Iota
K	κ	Kappa
Λ	λ	Lambda
M	μ	Mu
N	ν	Nu
Ξ	ξ	Ksi
O	o	Omicron
Π	π	Pi
P	ρ (ϱ)	Rho
Σ	σ (ς)	Sigma
T	τ	Tau
Υ	υ	Upsilon
Φ	ϕ (φ)	Phi
X	χ	Chi
Ψ	ψ	Psi
Ω	ω	Omega



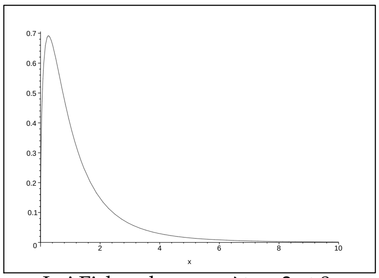
Loi beta 2 de paramètres 3 et 1.



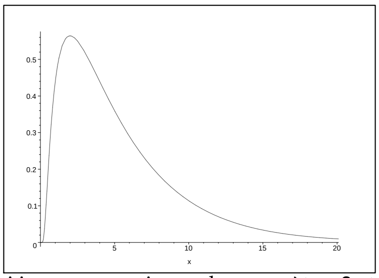
Loi beta 1 de paramètres 3 et 2.



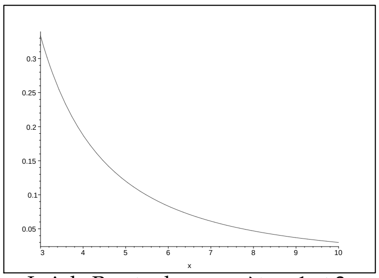
Loi de Cauchy de paramètres 4.



Loi Fisher de paramètres 3 et 8.



Loi inverse gaussienne de paramètres 2 et 2.



Loi de Pareto de paramètres 1 et 3.

FIGURE 10.1 – Densités de différentes lois continues.-

Bibliographie

[1] Bass, J. (1974), *Eléments de calcul des probabilités*, 3e edition, Masson, Paris.

[2] Boccara, N. (1995), *Probabilités*, collection : Mathématiques pour l'ingénieur, Ellipses, Paris.

[3] Bouleau, N. (1986), *Probabilités de l'ingénieur : variables aléatoires et simulation*, Hermann, Paris.

[4] Brémaud, P. (1988), *An introduction to probabilistic modeling*, Springer-Verlag.

[5] Dacunha-Castelle, D., Duflo, M. (1990), *Probabilités et statistiques : 1-Problèmes à temps fixe*, Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris.

[6] Deheuvels, P. (2008), *La probabilité, le hasard et la certitude*, Que sais-je ?, PUF, Paris.

[7] Feller, W. (1971), *An introduction to probability theory and its applications*, Vol.1, J.Wiley, N.Y.

[8] Foata, D., Fuchs, A. (1996), Calcul des probabilités, Masson, Paris.

[9] Girardin, V., Limnios, N. (2001, 2008, 2014), *Probabilités*, Vol. 1, Vuibert, Paris.

[10] Métivier, M. (1987), *Probabilités : dix leçons d'introduction*, Ecole Polytechnique, Ellipses, Paris.

[11] Neveu, J. (1988), *Introduction aux probabilités*, Ecole Polytechnique, Paris.

[12] Ross, S.M. (1987), *Initiation aux probabilités*,(traduit de l'américain), Presses Polytechniques Romandes, Lausanne.

[13] Saporta, G. (1990), *Probabilités analyse des données et statistique*, Technip, Paris.

[14] Shiryaev, A.N. (1996). *Probability*, Springer, 2nd Edition, N.Y.

TABLE

Fonction de répartition de la loi normale centrée réduite : $\mathcal{N}(0,1)$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\Big(-\frac{t^2}{2}\Big) dt$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500 00	0.503 99	0.507 98	0.511 97	0.515 95	0.519 94	0.523 92	0.527 90	0.531 88	0.535 86
0.1	0.539 83	0.543 80	0.547 76	0.551 72	0.555 67	0.559 62	0.563 56	0.567 50	0.571 42	0.575 35
0.2	0.579 26	0.583 17	0.587 06	0.590 95	0.594 84	0.598 71	0.602 57	0.606 42	0.610 26	0.614 09
0.3	0.617 91	0.621 72	0.625 52	0.629 30	0.633 07	0.636 83	0.640 58	0.644 31	0.648 03	0.651 73
0.4	0.655 42	0.659 10	0.662 76	0.666 40	0.670 03	0.673 65	0.677 24	0.680 82	0.684 39	0.687 93
0.5	0.691 46	0.694 97	0.698 47	0.701 94	0.705 40	0.708 84	0.712 26	0.715 66	0.719 04	0.722 40
0.6	0.725 75	0.729 07	0.732 37	0.735 65	0.738 91	0.742 15	0.745 37	0.748 57	0.751 75	0.754 90
0.7	0.758 04	0.761 15	0.764 24	0.767 31	0.770 35	0.773 37	0.776 37	0.779 35	0.782 30	0.785 24
0.8	0.788 14	0.791 03	0.793 89	0.796 73	0.799 55	0.802 34	0.805 11	0.807 85	0.810 57	0.813 27
0.9	0.815 94	0.818 59	0.821 21	0.823 81	0.826 39	0.828 94	0.831 47	0.833 98	0.836 46	0.838 91
1.0	0.841 34	0.843 75	0.846 14	0.848 50	0.850 83	0.853 14	0.855 43	0.857 69	0.859 93	0.862 14
1.1	0.864 33	0.866 50	0.868 64	0.870 76	0.872 86	0.874 93	0.876 98	0.879 00	0.881 00	0.882 98
1.2	0.884 93	0.886 86	0.888 77	0.890 65	0.892 51	0.894 35	0.896 17	0.897 96	0.899 73	0.901 47
1.3	0.903 20	0.904 90	0.906 58	0.908 24	0.909 88	0.911 49	0.913 09	0.914 66	0.916 21	0.917 74
1.4	0.919 24	0.920 73	0.922 20	0.923 64	0.925 07	0.926 47	0.927 86	0.929 22	0.930 56	0.931 89
1.5	0.933 19	0.934 48	0.935 74	0.936 99	0.938 22	0.939 43	0.940 62	0.941 79	0.942 95	0.944 08
1.6	0.945 20	0.946 30	0.947 38	0.948 45	0.949 50	0.950 53	0.951 54	0.952 54	0.953 52	0.954 49
1.7	0.955 43	0.956 37	0.957 28	0.958 19	0.959 07	0.959 94	0.960 80	0.961 64	0.962 46	0.963 27
1.8	0.964 07	0.964 85	0.965 62	0.966 38	0.967 12	0.967 84	0.968 56	0.969 26	0.969 95	0.970 62
1.9	0.971 28	0.971 93	0.972 57	0.973 20	0.973 81	0.974 41	0.975 00	0.975 58	0.976 15	0.976 70
2.0	0.977 25	0.977 78	0.978 31	0.978 82	0.979 32	0.979 82	0.980 30	0.980 77	0.981 24	0.981 69
2.1	0.982 14	0.982 57	0.983 00	0.983 41	0.983 82	0.984 22	0.984 61	0.985 00	0.985 37	0.985 74
2.2	0.986 11	0.986 45	0.986 79	0.987 13	0.987 45	0.987 78	0.988 09	0.988 40	0.988 70	0.988 99
2.3	0.989 28	0.989 56	0.989 83	0.990 10	0.990 36	0.990 61	0.990 86	0.991 11	0.991 34	0.991 58
2.4	0.991 80	0.992 02	0.992 24	0.992 45	0.992 66	0.992 86	0.993 05	0.993 24	0.993 43	0.993 61
2.5	0.993 79	0.993 96	0.994 13	0.994 30	0.994 46	0.994 61	0.994 77	0.994 92	0.995 06	0.995 20
2.6	0.995 34	0.995 47	0.995 60	0.995 73	0.995 85	0.995 98	0.996 09	0.996 21	0.996 32	0.996 43
2.7	0.996 53	0.996 64	0.996 74	0.996 83	0.996 93	0.997 02	0.997 11	0.997 20	0.997 28	0.997 36
2.8	0.997 44	0.997 52	0.997 60	0.997 67	0.997 74	0.997 81	0.997 88	0.997 95	0.998 01	0.998 07
2.9	0.998 13	0.998 19	0.998 25	0.998 31	0.998 36	0.998 41	0.998 46	0.998 51	0.998 56	0.998 61

Queue de la loi : $1 - \Phi(x)$

x	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3.	135 10e-5	968 10e-6	687 10e-6	483 10e-6	337 10e-6	233 10e-6	159 10e-6	108 10e-6	723 10e-7	481 10e-7
4.	317 10e-7	207 10e-7	133 10e-7	85 10e-7	54 10e-7	34 10e-7	21 10e-7	13 10e-7	79 10e-8	48 10e-8
5.	29 10e-8	17 10e-8	10 10e-8	58 10e-8	33 10e-8	19 10e-8	11 10e-8	60 10e-10	33 10e-10	18 10e-10