



INSTITUT
Mines-Télécom

Statistique pour ingénieur

Thème 1 : Notions de probabilités

D. Pastor, F.-X. Socheleau & C. Garnier, 18 octobre 2016

Table des matières

1	Introduction	2
2	Modèles probabilistes	2
2.1	Axiomatique de Kolmogorov	2
2.2	Probabilité conditionnelle	5
2.3	Indépendance	6
3	Variables aléatoires	6
3.1	Définitions	6
3.2	Moments	11
3.3	Lois et variables aléatoires usuelles	12
3.3.1	Variables aléatoires discrètes	12
3.3.2	Variables aléatoires absolument continues	13
4	Couples de variables aléatoires	14
4.1	Loi conjointe et espérance	14
4.2	Lois marginales	15
4.3	Lois conditionnelles	17
4.4	Variables aléatoires indépendantes	18
4.5	Covariance et coefficient de corrélation	19
5	Convergences	20
5.1	Convergence en loi	20
5.2	Convergence en probabilité	22
5.3	Convergence en moyenne quadratique	23
5.4	Autres modes de convergence	23
6	Exercices	25
	Exercice 1 : Spams	25
	Exercice 2 : Durée de vie	25
	Exercice 3 : Consommation énergétique	25
	Exercice 4 : Transformation linéaire d'un vecteur gaussien	25
	Exercice 5 : Application non probabiliste du théorème central-limite	26

1 Introduction

Le hasard fait partie de notre vie quotidienne. Tout du moins en avons-nous l'impression. Peut-être que ce que nous appelons hasard n'est finalement que le reflet de notre inaptitude à appréhender toutes les causes et les effets qui détermineraient les événements que nous qualifions d'aléatoires. Quoi qu'il en soit, l'homme a cherché depuis très longtemps à modéliser le hasard, ne serait-ce que pour estimer ses chances de gagner à un jeu, ce qui a été une des premières motivations qui a conduit à la théorie des probabilités.¹

La théorie des probabilités propose un modèle mathématique permettant d'appréhender le monde, ses causes et ses effets. C'est une théorie très efficace, ce qui en explique la présence dans tous les domaines de la connaissance, de la science et de l'ingénierie. L'objet des chapitres suivants est d'en rappeler les définitions et résultats fondamentaux dans une perspective d'application à la statistique.

2 Expériences aléatoires et modèles probabilistes

Certaines expériences, même si elles ne produisent pas le même résultat à chaque réalisation, semblent exhiber des caractères de régularité lorsqu'elles sont répétées un grand nombre de fois dans des conditions identiques. De telles expériences sont dites aléatoires. Par exemple, un tirage à pile ou face, un jet de dé à 6 faces, le nombre de connexions à un serveur dans un intervalle de temps donné, la mesure de la tension aux bornes d'un composant électronique, sont autant d'expériences aléatoires. En effet, pour chacune de ces expériences, le résultat peut différer d'une réalisation à l'autre. Mais pour chacune d'entre elles, nous pouvons mettre en évidence un caractère de régularité à partir d'un grand nombre de réalisations. Par exemple, si nous effectuons N mesures de la tension aux bornes d'un composant électronique et que nous comptons le nombre de fois n où cette tension reste inférieure ou égale à x , nous pouvons calculer la fréquence n/N de l'événement « La tension est inférieure ou égale à x sur N mesures ». Lorsque N augmente et que les conditions expérimentales restent identiques, la fréquence d'occurrence semble tendre vers une valeur bien déterminée.

Puisqu'une expérience aléatoire peut produire des résultats différents à chaque réalisation, nous commençons par donner une définition et un nom à l'ensemble de tous les résultats possibles de cette expérience.

2.1 Axiomatique de Kolmogorov

Définition 1 (*Univers*)

*L'ensemble Ω de tous les résultats possibles d'une expérience spécifiée par un protocole expérimental donné est appelé **univers**. On dira aussi que Ω est l'espace des états ou espace des possibles de l'expérience aléatoire.*

Cette définition suppose évidemment l'univers Ω non vide car, sinon, elle n'a aucun intérêt. Insistons aussi sur le fait que le qualificatif de « possible » souligne que Ω ne

1. Rappelons à ce sujet que Blaise Pascal jette les bases de la théorie, telle que nous la connaissons aujourd'hui, dès 1654, suite à un problème posé par le Chevalier de Méré sur le partage des gains à un jeu.

contient que les résultats que l'on peut constater lors de l'expérience. Ainsi, lorsque nous lançons un dé à 6 faces, cet ensemble des possibles est $\Omega = \{1,2,3,4,5,6\}$ et n'est pas $\{1,2,3,4,5,6,7\}$ parce que 7 n'est pas une valeur que l'on peut obtenir en lançant le dé. L'espace des états n'est pas non plus \mathbb{N} , \mathbb{R} , etc. Dans le même ordre d'idée, notons aussi que dans le cas où le dé à 6 faces serait, sans que nous le sachions, lesté de telle manière que certaines valeurs ne puissent jamais sortir, il n'y a pas lieu de changer d'univers puisque nous ignorons *a priori* que ce dé est pipé. En d'autres termes, l'espace des états est l'ensemble des résultats possibles compte tenu de l'expérience : nous lançons un dé à 6 faces et chaque face est *a priori* possible.

Nous devons maintenant formaliser la notion d'événement avec, en arrière-pensée, le fait que nous cherchons une définition qui nous permette d'attacher une valeur de « probabilité » à un événement. Puisque tous les résultats possibles sont rassemblés dans l'univers Ω , il est naturel de considérer un événement comme un ensemble de possibilités et donc, comme un sous-ensemble de Ω . Tant que Ω est dénombrable², nous pouvons effectivement considérer toutes les parties de Ω comme autant d'événements possibles. Par contre, dès que Ω est \mathbb{R} , ou même un intervalle borné d'intérieur non vide de \mathbb{R} , Ω devient beaucoup trop grand pour considérer l'ensemble de ses parties comme l'ensemble des événements. Une bonne définition pour couvrir tous ces cas est la suivante.

Définition 2 (*Espace probabilisable*)

Un **espace probabilisable** (ou mesurable) est un couple (Ω, \mathcal{T}) où Ω est un ensemble et \mathcal{T} une tribu de Ω , c'est-à-dire un ensemble de parties de Ω vérifiant les propriétés suivantes :

- 1) $\Omega \in \mathcal{T}$.
- 2) Si $A \in \mathcal{T}$, alors $\bar{A} \in \mathcal{T}$ où \bar{A} est le complémentaire de A dans Ω .
- 3) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'éléments de \mathcal{T} , alors $\bigcup_{n=1}^{\infty} A_n \in \mathcal{T}$.

Les éléments de \mathcal{T} sont appelés événements. En particulier, pour tout $\omega \in \Omega$, le singleton $\{\omega\}$ est appelé événement élémentaire.

Les trois axiomes d'un espace probabilisable s'expliquent intuitivement comme suit. Puisque les éléments de \mathcal{T} sont des ensembles A pour lesquels $\omega \in A$ correspond à un événement possible de l'expérience et que nous avons trivialement $\omega \in \Omega$, Ω ne peut être qu'un élément de \mathcal{T} . On dit que Ω est l'événement certain.

En ce qui concerne le deuxième axiome (stabilité par complémentarité), $A \in \mathcal{T}$ signifie que $\omega \in A$ est un événement possible que peut produire l'expérience. L'événement contraire $\omega \in \bar{A}$ est donc lui aussi un événement possible pour l'expérience considérée.

Le troisième axiome (stabilité par union dénombrable) signifie simplement qu'un choix entre les différents événements $\omega \in A_n$ est encore un événement qui peut subvenir lors de la réalisation de l'expérience. Il faut noter que cet axiome considère une union dénombrable d'éléments de \mathcal{T} et l'on pourrait se demander si considérer une union finie ne suffirait pas en pratique. En fait, non. Il suffit pour s'en convaincre de considérer l'expérience aléatoire qui consiste à compter le nombre de requêtes auprès d'un serveur pendant dans une période de temps donné. Ce nombre ne peut pas être borné puisque nous pouvons

2. C'est-à-dire si Ω est fini ou en bijection avec \mathbb{N} . Dans ce second cas, nous disons que Ω est infini dénombrable.

avoir 1, 2, 3... requêtes.

Exemple 1

L'ensemble des parties de Ω est une tribu de Ω . C'est même la plus grande tribu que l'on peut construire sur Ω . Considérons l'expérience aléatoire qui consiste à tirer au hasard un objet sur une ligne de conditionnement sur laquelle circule trois types d'objets différents : a, b, c . L'univers associé à cette expérience est $\Omega = \{a, b, c\}$ et l'ensemble des parties de Ω est $\mathcal{T} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \Omega\}$ qui est une tribu de Ω .

Exemple 2

Si $A \subseteq \Omega$, alors $\{\emptyset, \Omega, A, \bar{A}\}$ est une tribu de Ω .

Nous pouvons maintenant fournir la définition d'un espace probabilisé. Cette définition constitue l'axiomatique de Kolmogorov. Elle inclut la définition de la probabilité qu'un événement, au sens de la définition précédente, soit réalisé.

Définition 3 (*Espace probabilisé*)

Un **espace probabilisé** est un triplet $(\Omega, \mathcal{T}, \mathbb{P})$ où (Ω, \mathcal{T}) est un espace probabilisable et \mathbb{P} une mesure de probabilité sur \mathcal{T} , c'est-à-dire une application de \mathcal{T} dans $[0, 1]$ telle que :

- 1) $\mathbb{P}(\Omega) = 1$ [Condition de normalisation]
- 2) Soit $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements disjoints 2 à 2 :

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) \text{ } [\sigma\text{-additivité}].$$

Dans de nombreuses applications, Ω est \mathbb{R} . Dans ce cas, nous ne choisissons pas l'ensemble des parties de $\Omega = \mathbb{R}$ comme tribu. En effet, cet ensemble de parties est beaucoup trop grand pour y définir une probabilité \mathbb{P} . On utilisera la **tribu $\mathcal{B}(\mathbb{R})$ des boréliens** de \mathbb{R} , c'est-à-dire la plus petite tribu contenant tous les intervalles de \mathbb{R} . Cette tribu est déjà très grande et largement suffisante pour les applications pratiques. Dans les deux exemples suivants, Ω est dénombrable. On peut alors choisir l'ensemble des parties de Ω comme tribu.

Exemple 3 (*Univers dénombrable*)

Soit $\Omega = \{\omega_i : i \in I\}$ où $I \subseteq \mathbb{N}^*$ et les ω_i sont distincts 2 à 2 : $\omega_i \neq \omega_j$ pour $i \neq j$ dans I . On dit que Ω est dénombrable. Dans ce cas, la tribu \mathcal{T} sera l'ensemble des parties de Ω . On assignera ensuite à tout $\omega_i \in \Omega$ une valeur $p_i \in [0, 1]$ avec $\sum_{i \in I} p_i = 1$. Pour tout $A \subseteq \Omega$, on posera alors $\mathbb{P}(A) = \sum_{i \in K(A)} p_i$ où $K(A) \subseteq I$ est l'ensemble des indices des éléments ω_i de A .

Exemple 4 (*Probabilité uniforme ou équiprobabilité*)

Si $\Omega = \{\omega_1, \dots, \omega_n\}$ est fini, les ω_i étant distincts 2 à 2, la probabilité uniforme sur Ω est définie en posant $p_i = 1/n$.

2.2 Probabilité conditionnelle

Dans la pratique, il est très souvent utile de savoir calculer la probabilité d'un événement A , conditionnellement à ou sachant l'événement B . Par exemple, dans un jeu de dé à 6 faces, quelle est la probabilité que le résultat soit 6 sachant que ce résultat est pair ? Dans cette question, on cherche donc à calculer la probabilité de l'événement $A = \{6\}$ conditionnellement à l'événement $B = \{2, 4, 6\}$. Comme il y a équiprobabilité des tirages et qu'il n'y a qu'une seule chance sur 3 de tirer 6 parmi $\{2, 4, 6\}$, l'intuition nous dit que la probabilité conditionnelle de A sachant B est $1/3$. La définition générale qui permet de retrouver ce résultat est l'axiome de Bayes suivant.

Définition 4 (*Probabilité conditionnelle*)

Soit un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et un événement A tel que $\mathbb{P}(A) \neq 0$. Pour tout $B \in \mathcal{T}$, on définit la **probabilité conditionnelle** de B sachant A par :

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \quad [\text{Axiome de Bayes}].$$

Le résultat suivant illustre le bien-fondé de l'axiome de Bayes.

Proposition 1

Avec les notations précédentes, l'application qui associe à tout $B \in \mathcal{T}$ la valeur $\mathbb{P}(B|A)$ est une probabilité.

Proposition 2 (*Formule de Bayes*)

Supposons que $\Omega = \bigcup_{n=1}^{\infty} A_n$ où les A_n sont disjoints 2 à 2. La formule de Bayes permet de calculer les probabilités a posteriori $\mathbb{P}(A_n|B)$ à partir des valeurs $\mathbb{P}(A_n)$ et $\mathbb{P}(B|A_n)$:

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(A_n) \mathbb{P}(B|A_n)}{\sum_{k=1}^{\infty} \mathbb{P}(A_k) \mathbb{P}(B|A_k)}.$$

On notera que le résultat de la proposition précédente se démontre facilement après avoir remarqué que $\mathbb{P}(B) = \sum_{k=1}^{\infty} \mathbb{P}(A_k) \mathbb{P}(B|A_k)$ (formule des **probabilités totales**).

Exemple 5

35% des abonnés aux réseaux mobiles en France sont clients chez l'opérateur *SuperMobile*. Parmi les clients de cet opérateur, 25% ont un forfait 4G alors que pour les autres opérateurs ce chiffre est de seulement 15%. Quelle est la probabilité qu'un abonné quelconque ait un forfait 4G ?

Cette probabilité se détermine en appliquant la formule des probabilités totales. Soit $B = \{\text{l'abonné a un forfait 4G}\}$ et $A = \{\text{l'abonné est client chez SuperMobile}\}$, alors $\mathbb{P}(B) = \mathbb{P}(B|A) \mathbb{P}(A) + \mathbb{P}(B|\bar{A}) \mathbb{P}(\bar{A}) = 0,25 \times 0,35 + 0,15 \times 0,65 = 0,185$.

2.3 Indépendance

Définition 5 (*Événements indépendants*)

Soit un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$. On dit que deux événements A et B — c'est-à-dire, 2 éléments de \mathcal{T} — sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B)$.

Proposition 3 (*Indépendance et probabilité conditionnelle*)

Avec les notations de la définition précédente, supposons que $\mathbb{P}(B) \neq 0$. Les événements A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Autrement dit, deux événements sont indépendants (et en omettant le cas particulier où l'un est de probabilité nulle), si la réalisation de l'un n'influe pas sur la probabilité de réalisation de l'autre.

3 Variables aléatoires

3.1 Définitions

Le concept de **variable aléatoire** formalise la notion de « grandeur » associée au résultat d'une expérience aléatoire. Cette notion qui associe une grandeur à chaque élément de l'univers Ω d'une expérience aléatoire correspond à la notion d'application.

Définition 6 (*Variable aléatoire*)

Une variable aléatoire réelle définie sur un espace probabilisable (Ω, \mathcal{T}) est une application :

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto X(\omega) \end{aligned}$$

telle que l'image réciproque $X^{-1}(I)$ soit un élément de \mathcal{T} pour tout intervalle I de $\mathcal{B}(\mathbb{R})$. On dit que cette application est mesurable. L'ensemble $X(\Omega)$ des valeurs prises par la variable aléatoire X est appelé **domaine de variation** de X .

Exemple 6

Considérons une expérience où un objet manufacturé est tiré de façon aléatoire sur une chaîne de production. Une variable aléatoire peut être l'application qui à chaque objet va lui associer sa taille, son poids ou tout autre grandeur le définissant. C'est donc ici une application de $\Omega =$ « l'ensemble des objets possibles de la chaîne de production » vers \mathbb{R} .

Dans la suite, nous rencontrerons essentiellement :

- des **variables aléatoires discrètes** : $X(\Omega)$ est une partie finie ou dénombrable de \mathbb{R} ,
- des **variables aléatoires absolument continues** : $X(\Omega)$ est tout \mathbb{R} ou un intervalle d'intérieur non vide de \mathbb{R} et X admet une densité.

Définition 7 (*Loi de probabilité*)

La notion de variable aléatoire permet de transposer la structure abstraite du modèle probabiliste sur l'espace d'arrivée \mathbb{R} en définissant la **loi de probabilité** de X , notée

\mathbb{P}_X , qui est la probabilité image de \mathbb{P} par X :

$$\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B).$$

La loi d'une variable aléatoire réelle est entièrement caractérisée par la fonction de répartition de cette variable.

Définition 8 (*Fonction de répartition*)

La **fonction de répartition** \mathbb{F}_X d'une variable aléatoire réelle X est définie par :

$$\forall x \in \mathbb{R}, \mathbb{F}_X(x) = \mathbb{P}(X^{-1}(]-\infty, x])) = \mathbb{P}(X \leq x).$$

La fonction de répartition vérifie les propriétés suivantes :

- Elle existe toujours.
- \mathbb{F}_X est croissante.
- \mathbb{F}_X est continue à droite.
- $\lim_{x \rightarrow \infty} \mathbb{F}_X(x) = 1$ et $\lim_{x \rightarrow -\infty} \mathbb{F}_X(x) = 0$

Comme tous les intervalles de \mathbb{R} peuvent se construire par union, intersection finie ou dénombrable, complémentarités d'intervalles de la forme $] - \infty, x]$, toute probabilité $\mathbb{P}(X \in I)$, où I est un intervalle quelconque de \mathbb{R} , peut se déduire de la fonction de répartition. Des exemples de fonctions de répartition sont montrés à la [figure 1](#).

Proposition 4

La fonction de répartition \mathbb{F}_X d'une variable aléatoire réelle X possède les propriétés suivantes pour tout $a, b \in \mathbb{R}$ tels que $a < b$:

$$\begin{aligned} \mathbb{P}(a < X \leq b) &= \mathbb{F}_X(b) - \mathbb{F}_X(a) \\ \mathbb{P}(X = a) &= \mathbb{F}_X(a) - \mathbb{F}_X(a-) \\ \mathbb{P}(a < X) &= 1 - \mathbb{F}_X(a) \\ \mathbb{P}(X < a) &= \mathbb{F}_X(a-) \end{aligned}$$

où $\mathbb{F}_X(a-)$ est la limite à gauche de la fonction de répartition au point a .

Notez que pour une variable aléatoire discrète X , la fonction de répartition est peu utilisée car on connaît tout de X dès que l'on connaît les $\mathbb{P}(X = x_i)$ où $X(\Omega) = \{x_i : i \in I\}$ avec $I \subseteq \mathbb{N}^*$.

Définition 9 (*Densité de probabilité*)

Une variable aléatoire réelle X de fonction de répartition \mathbb{F}_X admet une **densité** s'il existe une fonction $f_X : \mathbb{R} \mapsto [0, +\infty[$ telle que :

$$\int_{\mathbb{R}} f_X(t) dt = 1 \text{ [Condition de normalisation]} \quad \text{et} \quad \mathbb{F}_X(x) = \int_{-\infty}^x f_X(t) dt.$$

Si une variable aléatoire X admet une densité f_X , on dit que la fonction de répartition \mathbb{F}_X est **absolument continue** de densité f_X . Par abus de langage, nous dirons aussi que

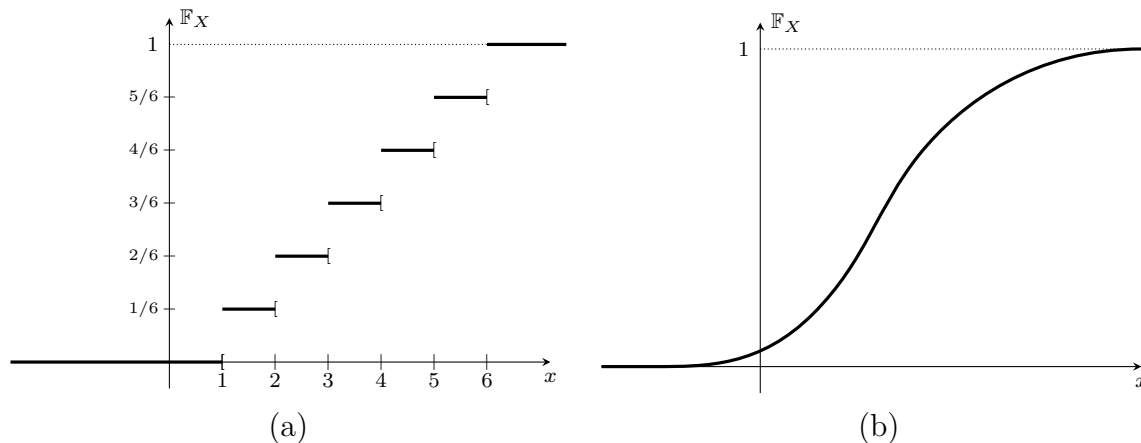


FIGURE 1 – (a) Fonction de répartition d'un jeu de dé équilibré. (b) Fonction de répartition d'une variable aléatoire réelle.

X est absolument continue. Des exemples de densités de probabilités sont tracés sur la [figure 2](#). A noter que toutes les fonctions de répartition ne sont pas absolument continues. Par conséquent, une variable aléatoire n'admet pas toujours de densité.

Proposition 5

Si une variable aléatoire réelle X est absolument continue, alors la fonction de répartition \mathbb{F}_X de X est continue et une densité f_X de X est obtenue par :

$$f_X(x) = \mathbb{F}'_X(x), \text{ presque partout (p.p).}$$

En pratique, (p.p) signifiera : partout sauf, peut-être, sur un sous-ensemble dénombrable de \mathbb{R} .

Notez que cette propriété implique qu'il n'existe pas qu'une seule densité de probabilité pour une variable aléatoire réelle absolument continue. En effet, la valeur de l'intégrale d'une densité de probabilité peut être égale à la valeur de l'intégrale d'une autre densité (elles ont donc la même fonction de répartition) si ces deux densités sont identiques sauf sur des sous-ensembles dénombrables de \mathbb{R} .

La proposition suivante est très utile en pratique pour savoir si une variable aléatoire est absolument continue et, si elle l'est, en calculer une densité.

Proposition 6

Soit une variable aléatoire X de fonction de répartition \mathbb{F}_X . Si l'ensemble $\mathcal{E} \subseteq \mathbb{R}$ des réels où \mathbb{F}_X n'est pas dérivable est vide ou fini, X est absolument continue et une densité f de X est $f(x) = \mathbb{F}'_X(x)$ pour tout $x \in \mathbb{R} \setminus \mathcal{E}$ et en choisissant arbitrairement la valeur $f(x)$ dans $[0, \infty[$ pour tout x de \mathcal{E} , si celui-ci est non vide.

Proposition 7

Soient a et b deux réels quelconques tels que $a \leq b$, toute variable aléatoire réelle X absolument continue de densité f_X et de fonction de répartition \mathbb{F}_X vérifie :

(1) $\mathbb{P}(X = a) = 0$

$$\begin{aligned}
(2) \quad \mathbb{P}(X \leq a) &= \mathbb{P}(X < a) = \int_{-\infty}^a f_X(x) dx \\
(3) \quad \mathbb{P}(a \leq X) &= \mathbb{P}(a < X) = \int_a^{\infty} f_X(x) dx \\
(4) \quad \mathbb{P}(a < X \leq b) &= \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b) \\
&= \int_a^b f(x) dx = F_X(b) - F_X(a)
\end{aligned}$$

Exemple 7

Pour dimensionner un réseau de téléphonie, on modélise la durée (en minutes) d'une conversation comme une variable aléatoire X de densité de probabilité $f_X(x) = \frac{1}{2}e^{-x/2}$ si x est positif et $f_X(x) = 0$ sinon. Quelle est la probabilité qu'une conversation dure entre une et deux minutes ?

Il suffit de calculer $\mathbb{P}(1 \leq X \leq 2) = \frac{1}{2} \int_1^2 e^{-x/2} dx = e^{-1/2} - e^{-1} \approx 0,24$.

Dans la pratique, nous rencontrons souvent des changements de variables où une variable aléatoire Y est obtenue en appliquant une fonction g à une variable aléatoire initiale X . La variable aléatoire Y est alors la composée $Y = g \circ X = g(X)$ de g et de X . Même lorsque X admet une densité, Y n'admet pas forcément de densité. Par exemple, si nous considérons un intervalle $[a, b]$ de \mathbb{R} et posons $g(x) = 1$ si $a \leq x \leq b$ et $g(x) = 0$, sinon³, $Y = g(X)$ est une variable aléatoire discrète et n'admet donc pas de densité.

Cependant, très souvent, les fonctions g vérifient de « bonnes » propriétés qui permettent d'en déduire que Y sera absolument continue dès que X l'est. Des résultats généraux permettent même de déduire de ces propriétés l'expression d'une densité de Y . Dans ce cours de statistique, ces résultats théoriques ne seront pas utiles et le lecteur intéressé pourra toujours se reporter à la bibliographie pour compléter ses connaissances sur le sujet. Par contre, il faut savoir faire quelques calculs élémentaires de changements de variables tels ceux proposés dans les exemples suivants.

Exemple 8 (*Transformation affine*)

Soit une variable aléatoire réelle X de densité f_X . Etant donnés deux réels $a \neq 0$ et b , considérons la variable aléatoire $Y = aX + b$ de X .

(i) Si $a > 0$, pour tout $y \in \mathbb{R}$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right).$$

La **proposition 6** implique que Y est absolument continue de densité

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

3. g est appelée fonction indicatrice de $[a, b]$

(ii) Si $a < 0$, pour tout $y \in \mathbb{R}$,

$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(X \geq \frac{y-b}{a}\right) = 1 - \mathbb{F}_X\left(\frac{y-b}{a}\right).$$

La **proposition 6** implique que, dans ce cas, Y est absolument continue de densité

$$f_Y(y) = -\frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

Nous concluons que Y est absolument continue de densité :

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Exemple 9

Soit une variable aléatoire X absolument continue de densité f_X . On suppose que \mathbb{F}_X est dérivable sur tout \mathbb{R} . Soit la variable aléatoire $Y = X^2$ et \mathbb{F}_Y , la fonction de répartition de Y . Pour $y < 0$, $\mathbb{F}_Y(y) = 0$ puisque Y ne prend que des valeurs positives ou nulles. Pour $y \geq 0$, nous avons :

$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \mathbb{F}_X(\sqrt{y}) - \mathbb{F}_X(-\sqrt{y}).$$

où \mathbb{F}_X est la fonction de répartition de X . En résumé, nous avons :

$$\mathbb{F}_Y(y) = \begin{cases} \mathbb{F}_X(\sqrt{y}) - \mathbb{F}_X(-\sqrt{y}) & \text{si } y \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

La fonction de répartition de Y est dérivable sur $\mathbb{R} \setminus \{0\}$. La **proposition 6** nous permet de déduire que la variable aléatoire Y est absolument continue de densité :

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) & \text{si } y > 0 \\ 0 & \text{sinon.} \end{cases}$$

En $y = 0$, nous avons choisi de forcer f_Y à 0. Mais nous aurions très bien pu choisir $f_Y(0) = 316.9874254$, par exemple, si nous l'avions voulu. Mais cela n'a pas d'intérêt d'introduire une valeur supplémentaire pour f_Y . Donc autant garder la valeur nulle dans le prolongement de ce qu'il se passe pour $y < 0$.

Exemple 10

Soient une variable aléatoire X de densité f_X et la variable aléatoire $|X|$. Nous supposons que la fonction de répartition \mathbb{F}_X est dérivable sur tout \mathbb{R} . Pour $y < 0$, $\mathbb{F}_{|X|}(y) = 0$, puisque Y ne prend que des valeurs positives ou nulles. Pour $y \geq 0$, nous avons :

$$\mathbb{F}_{|X|}(y) = \mathbb{P}(|X| \leq y) = \mathbb{P}(-y \leq X \leq y) = \mathbb{F}_X(y) - \mathbb{F}_X(-y)$$

où \mathbb{F}_X est la fonction de répartition de X . Finalement, nous avons :

$$\mathbb{F}_{|X|}(y) = \begin{cases} \mathbb{F}_X(y) - \mathbb{F}_X(-y) & \text{si } y \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

La fonction de répartition de $|Y|$ est dérivable sur \mathbb{R} . Il s'ensuit que :

$$\mathbb{F}'_{|X|}(y) = f_X(y) + f_X(-y)$$

pour tout $y \geq 0$. Comme $\mathbb{F}_X(y) = 0$ pour $y < 0$, la fonction de répartition $\mathbb{F}_{|X|}$ est dérivable pour tout $y < 0$. Il s'ensuit que $\mathbb{F}_{|X|}$ est dérivable sur tout \mathbb{R} . Par application de la **proposition 6**, $|Y|$ est absolument continue et une densité de cette variable aléatoire est :

$$f_{|X|}(y) = \begin{cases} f_X(y) + f_X(-y) & \text{si } y > 0 \\ 0 & \text{sinon.} \end{cases}$$

3.2 Moments

La loi d'une variable aléatoire n'est pas toujours connue. Même lorsque cette loi est connue, il est souvent suffisant dans la pratique de manipuler des paramètres qui la décrivent de manière globale. Ces paramètres sont appelés moments.

Définition 10 (*Moments*)

Soit X une variable aléatoire. On appelle **moment** d'ordre k de X la grandeur $\mathbb{E}(X^k)$ définie, si elle existe, par :

- **Cas discret** : $\mathbb{E}(X^k) = \sum_i x_i^k \mathbb{P}(X = x_i)$.
- **Cas continu** : $\mathbb{E}(X^k) = \int_{-\infty}^{+\infty} x^k f_X(x) dx$.

Le moment d'ordre k existe lorsque $\mathbb{E}(|X^k|) < +\infty$. Deux paramètres, calculés à partir des moments d'ordre 1 et 2, sont particulièrement importants en statistique : l'**espérance** et la **variance**.

Définition 11 (*Espérance*)

L'espérance est le moment d'ordre $k = 1$ et est un paramètre de centralité de la variable aléatoire X .

Exemple 11

Considérons le même scénario que celui énoncé à l'**exemple 7**. Quelle est la durée moyenne d'un appel téléphonique ?

La durée moyenne est donnée par le calcul de $\mathbb{E}(X) = \int_0^{+\infty} \frac{x}{2} e^{-x/2} dx$. À l'aide d'une intégration par parties, on trouve $\mathbb{E}(X) = 2$ min.

Définition 12 (Variance)

La variance est un paramètre de dispersion de X et est définie comme

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

Lorsque $\mathbb{E}(X) = 0$, on dit que la variable aléatoire est **centrée** et, lorsque $\mathbb{V}(X) = 1$, on dit qu'elle est **réduite**.

Théorème 8 (Théorème de transfert)

Soit une variable aléatoire réelle X et g une application telle que $\mathbb{E}(|g(X)|) < +\infty$, alors

- **Cas discret** : $\mathbb{E}(g(X)) = \sum_i g(x_i) \mathbb{P}(X = x_i)$.
- **Cas continu** : $\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$.

Exemple 12

Soit R une variable aléatoire représentant le rayon de bulles créées par un phénomène physique quelconque. On suppose que ce rayon est uniformément distribué entre 4 et 6 mm (cf. [exemple 17](#)). Quel est le volume moyen de ces bulles ?

Soit V la variable aléatoire représentant le volume des bulles. V et R sont liées par la relation suivante : $V = \frac{4}{3}\pi R^3$. En appliquant le théorème du transfert on en déduit que $\mathbb{E}(V) = \int_{-\infty}^{+\infty} \frac{4}{3}\pi r^3 f_R(r) dr$. R étant une variable aléatoire uniforme comprise entre 4 et 6 mm, alors $\mathbb{E}(V) = \int_4^6 \frac{4}{3}\pi r^3 \times \frac{1}{2} dr \approx 544,5 \text{ mm}^3$.

3.3 Lois et variables aléatoires usuelles

Dans cette section, quelques exemples de lois de variables aléatoires discrètes et continues sont présentés. Cette liste n'est pas exhaustive et d'autres exemples de lois seront présentés au fur et à mesure de l'avancement du cours de statistique.

3.3.1 Variables aléatoires discrètes**Exemple 13 (Loi de Bernoulli)**

Une variable aléatoire X est dite de Bernoulli s'il existe un nombre $p \in [0,1]$ tel que la loi de probabilité de X soit donnée par $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$. On notera $\mathcal{B}(1,p)$ la loi de Bernoulli. L'espérance et la variance de X sont données respectivement par $\mathbb{E}(X) = p$ et $\mathbb{V}(X) = p(1 - p)$.

La loi de Bernoulli modélise les expériences de type « succès » ou « échec » comme le jeu de pile ou face.

Exemple 14 (Loi géométrique)

Une variable aléatoire X qui prend ses valeurs dans \mathbb{N}^* est géométrique de paramètre $p \in [0,1]$, si sa loi de probabilité s'écrit : $\mathbb{P}(N = k) = (1 - p)^{k-1}p$. L'espérance de X

est $\mathbb{E}(X) = 1/p$ et la variance de X est $\mathbb{V}(X) = (1-p)/p^2$.

Cette loi modélise le rang du premier succès d'une série d'épreuves de Bernoulli renouvelées de façon indépendante (où la probabilité de succès individuel est p).

Exemple 15 (*Loi binomiale*)

Une variable aléatoire X est binomiale de paramètres (n, p) avec $n \in \mathbb{N}$ et $p \in [0, 1]$ si sa loi de probabilité s'écrit : $\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}$, $k \in \{0, 1, \dots, n\}$. On notera $\mathcal{B}(n, p)$ la loi binomiale. L'espérance et la variance de X sont respectivement par $\mathbb{E}(X) = np$ et $\mathbb{V}(X) = np(1-p)$.

La loi binomiale modélise le nombre de succès d'une série de n épreuves aléatoires de Bernoulli indépendantes (où la probabilité de succès individuel est p).

Exemple 16 (*Loi de Poisson*)

Une variable aléatoire X suit une loi de Poisson de paramètre $\lambda > 0$, si sa loi de probabilité s'écrit : $\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ pour $k = 0, 1, 2, \dots$. On notera $\mathcal{P}(\lambda)$ la loi de Poisson. On a $\mathbb{E}(X) = \mathbb{V}(X) = \lambda$.

La loi de Poisson modélise le comportement du nombre d'événements se produisant dans un laps de temps fixé lorsqu'un événement se produit indépendamment du temps écoulé depuis l'événement précédent (ex : nombre de personnes entrant dans un magasin dans une journée, nombre de connexions à un serveur dans une journée, etc.).

3.3.2 Variables aléatoires absolument continues

Exemple 17 (*Loi uniforme*)

Une variable aléatoire X suit une loi uniforme de paramètre (a, b) , avec $-\infty < a < b < +\infty$ si elle admet pour densité de probabilité la fonction f définie par

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

On notera $\mathcal{U}([a, b])$ la loi uniforme. L'espérance et la variance de X sont respectivement $\mathbb{E}(X) = (a+b)/2$ et $\mathbb{V}(X) = (b-a)^2/12$.

Cette loi est une extension au domaine continu de la notion d'équiprobabilité définie dans le domaine discret.

Exemple 18 (*Loi normale ou gaussienne*)

Une variable aléatoire X suit une loi normale ou gaussienne de paramètres (μ, σ^2) , avec $\mu \in \mathbb{R}$ et $\sigma \in]0, +\infty[$ si elle admet pour densité de probabilité la fonction f définie par

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

On notera $\mathcal{N}(\mu, \sigma^2)$ la loi normale. L'espérance de X est $\mathbb{E}(X) = \mu$ et la variance de X est $\mathbb{V}(X) = \sigma^2$.

Cette loi est omniprésente en statistique du fait, notamment, du théorème central-limite présenté à la [section 5](#).

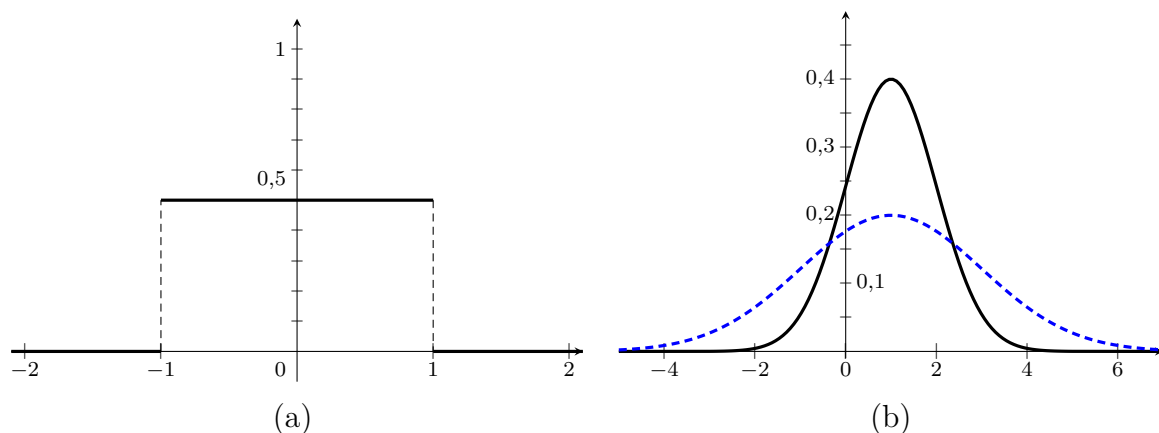


FIGURE 2 – (a) Densité de probabilité de la loi Uniforme $\mathcal{U}([-1, +1])$. (b) Densité de probabilité de la loi Normale, trait plein : $\mathcal{N}(1,1)$, pointillés : $\mathcal{N}(1,4)$.

4 Couple de variables aléatoires

On définit la tribu $\mathcal{B}(\mathbb{R}^2)$ des boréliens de \mathbb{R}^2 comme la plus petite tribu de \mathbb{R}^2 contenant tous les pavés $]a,b[\times]c,d[$ où a,b,c,d sont réels. C'est encore la plus petite tribu contenant les ensembles $] -\infty, x] \times] -\infty, y]$ lorsque (x,y) parcourt \mathbb{R}^2 .

Pour une bonne compréhension de la suite, il faut noter dès à présent que si $(\Omega, \mathcal{T}, \mathbb{P})$ est un espace probabilisé et X et Y sont des variables aléatoires définies sur l'espace probabilisable (Ω, \mathcal{T}) , alors $\{(X,Y) \in C\} \in \mathcal{T}$ pour tout $C \in \mathcal{B}(\mathbb{R}^2)$.

4.1 Loi conjointe et espérance

Définition 13 (*Loi conjointe*)

Soient un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et deux variables aléatoires réelles définies sur l'espace probabilisable (Ω, \mathcal{T}) . La **loi conjointe** du couple (X,Y) est la mesure de probabilité définie sur $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ par :

$$\mathbb{P}_{X,Y}(C) = \mathbb{P}((X,Y) \in C)$$

pour tout $C \in \mathcal{B}(\mathbb{R}^2)$.

Avec les notations précédentes, $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mathbb{P}_{X,Y})$ est un espace probabilisé.

Remarque 1

Pour connaître $\mathbb{P}_{X,Y}$, il suffit de connaître les valeurs de probabilité $\mathbb{P}_{X,Y}(A \times B) = \mathbb{P}((X,Y) \in A \times B)$ pour tous les événements A et B de \mathcal{T} . On démontre aussi que la connaissance de $\mathbb{P}_{X,Y}$ équivaut à celle de la **fonction de répartition conjointe**

définie pour tout $(x, y) \in \mathbb{R}^2$ par :

$$\mathbb{F}_{X,Y}(x, y) = \mathbb{P}((X, Y) \in]-\infty, x] \times]-\infty, y]) = \mathbb{P}(X \leq x, Y \leq y).$$

Dans la pratique, comme dans le cas des variables aléatoires réelles, nous allons souvent rencontrer des couples de variables aléatoires discrètes et des couples absolument continus de variables aléatoires.

1) Soient deux variables aléatoires réelles X et Y à valeurs discrètes. Supposons que le domaine de variation de X soit $\{x_i : i \in I\}$, où $I \subseteq \mathbb{N}^*$ et les valeurs x_i sont distinctes 2 à 2. De même, soit $\{y_j : j \in J\}$, le domaine de variation de Y où $J \subseteq \mathbb{N}^*$ et les valeurs y_j sont distinctes 2 à 2. La loi conjointe de (X, Y) est donnée par les valeurs $\mathbb{P}(X = x_i, Y = y_j)$ et, d'après ce qui précède,

$$\mathbb{P}((X, Y) \in C) = \sum_{(i,j) \in K(C)} \mathbb{P}(X = x_i, Y = y_j)$$

où $K(C) \subseteq I \times J$ est l'ensemble des indices $(i, j) \in I \times J$ tels que $(x_i, y_j) \in C$.

2) Soit un couple (X, Y) de variables aléatoires réelles. On dit que (X, Y) est absolument continu ou que X et Y sont conjointement continues s'il existe une fonction $f_{X,Y}$ positive ou nulle telle que, pour tout couple (x, y) de réels :

$$\mathbb{F}_{X,Y}(x, y) = \iint_{]-\infty, x] \times]-\infty, y]} f_{X,Y}(u, v) du dv.$$

La fonction $f_{X,Y}$ est appelée **densité de probabilité conjointe** du couple (X, Y) . On a alors, pour tout $C \in \mathcal{B}(\mathbb{R}^2)$:

$$\mathbb{P}((X, Y) \in C) = \mathbb{P}_{X,Y}(C) = \iint_C f_{X,Y}(x, y) dx dy.$$

A partir des définitions précédentes, on peut déterminer l'espérance d'une fonction de deux variables aléatoires en appliquant le théorème de transfert énoncé ci-après.

Théorème 9 (Théorème de transfert pour les couples de variables aléatoires)

Soit (X, Y) un couple de variables aléatoires réelles et g une application telle que $\mathbb{E}(|g(X, Y)|) < +\infty$, alors

- **Cas discret** : $\mathbb{E}(g(X, Y)) = \sum_{(i,j) \in K(C)} g(x_i, y_j) \mathbb{P}(X = x_i, Y = y_j).$
- **Cas continu** : $\mathbb{E}(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) dx dy.$

4.2 Lois marginales

Soient un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et un couple (X, Y) de variables aléatoires réelles définies sur l'espace probabilisable (Ω, \mathcal{T}) . Les lois de X et de Y s'appellent les lois marginales du couple (X, Y) . La question est de savoir si ces **lois marginales** \mathbb{P}_X et \mathbb{P}_Y peuvent se retrouver à partir de $\mathbb{P}_{X,Y}$. La réponse à cette question est oui puisque, pour tout $A \in \mathcal{T}$,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A, Y \in \mathbb{R}) = \mathbb{P}((X, Y) \in A \times \mathbb{R}),$$

ce qui permet de démontrer que pour tout $x \in \mathbb{R}$:

$$\mathbb{F}_X(x) = \lim_{y \rightarrow \infty} \mathbb{F}_{X,Y}(x,y)$$

De la même façon, nous avons pour tout $B \in \mathcal{T}$:

$$\mathbb{P}_Y(B) = \mathbb{P}(X \in \mathbb{R}, Y \in B) = \mathbb{P}((X,Y) \in \mathbb{R} \times B),$$

ce qui permet de démontrer que pour tout $y \in \mathbb{R}$:

$$\mathbb{F}_Y(y) = \lim_{x \rightarrow \infty} \mathbb{F}_{X,Y}(x,y).$$

Proposition 10

Soient deux variables aléatoires réelles X et Y à valeurs discrètes. Comme précédemment, on note $\{x_i : i \in I\}$ et $\{y_j : j \in J\}$ les domaines de variation respectifs de ces variables aléatoires, où $I \subseteq \mathbb{N}^*$ et $J \subseteq \mathbb{N}^*$. On suppose que les valeurs x_i sont distinctes 2 à 2 ainsi que les valeurs y_j . Sinon, une valeur pourrait être comptée 2 fois dans la formule suivante ! En vertu de ce qui précède, on a :

$$\mathbb{P}(X = x_i) = \sum_{j \in J} \mathbb{P}(X = x_i, Y = y_j).$$

Proposition 11

Dans le cas d'un couple (X,Y) absolument continu, les variables aléatoires X et Y sont elles aussi absolument continues. Une densité de X est alors :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$$

et une densité de Y est :

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$$

Ces densités sont appelées **densités marginales** respectives de X et de Y .

Nous savons donc comment remonter aux lois de X et de Y à partir de la loi du couple (X,Y) . Par contre, le contraire n'est pas toujours vrai. Plus précisément, sans information supplémentaire, la connaissance seule des lois de X et de Y ne suffit pas en déduire celle de (X,Y) . En voici un exemple très simple.

Exemple 19

Soit $0 < \alpha \leq 1/2$. Considérons un couple (X,Y) dont la loi est :

$\mathbb{P}(X = 0, Y = 0) = \alpha$	$\mathbb{P}(X = 1, Y = 0) = 1/2 - \alpha$
$\mathbb{P}(X = 0, Y = 1) = 1/2 - \alpha$	$\mathbb{P}(X = 1, Y = 1) = \alpha$

D'après l'exemple 10, nous calculons la loi marginale de X par la formule :

$$\begin{aligned} \mathbb{P}(X = 0) &= \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 0, Y = 1) = 1/2 \\ \mathbb{P}(X = 1) &= \mathbb{P}(X = 1, Y = 0) + \mathbb{P}(X = 1, Y = 1) = 1/2. \end{aligned}$$

De même, nous avons :

$$\begin{aligned}\mathbb{P}(Y = 0) &= \mathbb{P}(X = 0, Y = 0) + \mathbb{P}(X = 1, Y = 0) = 1/2 \\ \mathbb{P}(Y = 1) &= \mathbb{P}(X = 0, Y = 1) + \mathbb{P}(X = 1, Y = 1) = 1/2.\end{aligned}$$

Nous constatons que les lois de X et Y ne dépendent pas de α , alors que celle de (X, Y) en dépend ! Il est donc impossible de remonter à la loi conjointe du couple (X, Y) à partir des lois marginales de X et de Y .

4.3 Lois conditionnelles

La notion de **loi conditionnelle** permet de décrire l'influence d'une variable aléatoire sur une autre. Plus précisément, soient un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et deux variables aléatoires X et Y définies sur l'espace probabilisable (Ω, \mathcal{T}) , nous aimerions donner un sens et calculer la loi conditionnelle $\mathbb{P}_Y(\cdot | X = x)$ de Y sachant l'événement $\{X = x\}$. Tant que $\mathbb{P}(X = x) \neq 0$, l'axiome de Bayes suffit et nous permet de définir cette loi conditionnelle de Y sachant $X = x$ en posant :

$$\mathbb{P}_Y(A | X = x) = \frac{\mathbb{P}(\{Y \in A\} \cap \{X = x\})}{\mathbb{P}(X = x)}$$

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

Exemple 20

Avec les notations précédentes, supposons que X soit discrète, de domaine de variation $\{x_i : i \in I\}$ où $I \subseteq \mathbb{N}^*$ et les x_i sont distincts 2 à 2. La loi conditionnelle de Y sachant $\{X = x\}$ est donc :

$$\mathbb{P}_Y(A | X = x_i) = \frac{\mathbb{P}(\{Y \in A\} \cap \{X = x_i\})}{\mathbb{P}(X = x_i)}$$

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

La situation se complique si X est absolument continue. En effet, dans ce cas, $\mathbb{P}(X = x) = 0$! L'axiome de Bayes ne fonctionne plus et il faut trouver autre chose. Une solution simple existe lorsque (X, Y) est absolument continu. Dans ce cas, $\mathbb{P}(y_1 < Y \leq y_2 | x < X \leq x + h)$ pour $h > 0$ peut être calculé en employant l'axiome de Bayes puisque $\mathbb{P}(x < X \leq x + h) \neq 0$. Si $f_X(x) \neq 0$, le lecteur vérifiera que pour h suffisamment petit :

$$\mathbb{P}(y_1 < Y \leq y_2 | x < X \leq x + h) \approx \int_{y_1}^{y_2} f_{Y|X=x}(y) dy \text{ avec } f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

En utilisant le résultat de l'exemple 11, il est facile de vérifier que $f_{Y|X=x}$ est une densité de probabilité. D'où la définition suivante.

Définition 14 (*Densité conditionnelle*)

Soient un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$ et un couple (X, Y) absolument continu de variables aléatoires définies sur l'espace probabilisable (Ω, \mathcal{T}) . Si f_X est la densité mar-

ginale de X , la loi conditionnelle de Y sachant $\{X = x\}$ est définie pour tout x où $f_X(x) \neq 0$ par sa densité :

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Puisque $f_{Y|X=x}$ est une densité de probabilité, on peut lui associer une notion d'espérance. Celle-ci sera alors l'espérance conditionnelle de Y sachant $\{X = x\}$. La définition de cette espérance conditionnelle est la suivante :

Définition 15 (*Espérance conditionnelle*)

Avec les notations et les hypothèses de la **définition 14**, on définit l'**espérance conditionnelle** de Y sachant $\{X = x\}$ par :

$$\mathbb{E}(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy.$$

Le résultat suivant est souvent très utile.

Proposition 12 (*Règle des espérances itérées*)

Avec les notations et les hypothèses de la **définition 14** :

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y|X = x) f_X(x) dx.$$

4.4 Variables aléatoires indépendantes

L'indépendance entre deux variables aléatoires X et Y est une hypothèse très commode pour en calcul des probabilités et en statistique. L'indépendance entre X et Y signifie que chacune de ces variables aléatoires varie indépendamment l'une de l'autre. C'est équivalent à ce que la loi du couple (X,Y) s'exprime en séparant la loi de X et celle de Y .

Définition 16 (*Variables aléatoires indépendantes*)

Soit un espace probabilisé $(\Omega, \mathcal{T}, \mathbb{P})$. On dit que deux variables aléatoires X et Y sur l'espace probabilisable (Ω, \mathcal{T}) sont **indépendantes** si pour tous A et B de $\mathcal{B}(\mathbb{R})$, les événements $X \in A$ et $Y \in B$ sont indépendants.

L'indépendance de X et de Y se traduit donc par :

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \mathbb{P}(Y \in B),$$

que l'on peut encore écrire :

$$\mathbb{P}_{X,Y}(A \times B) = \mathbb{P}_X(A) \mathbb{P}_Y(B).$$

Théorème 13

Avec les notations précédentes :

1. X et Y sont indépendantes si et seulement si $\mathbb{F}_{X,Y}(x,y) = \mathbb{F}_X(x) \mathbb{F}_Y(y)$ pour

tout $(x, y) \in \mathbb{R}^2$, où $\mathbb{F}_X, \mathbb{F}_Y, \mathbb{F}_{X,Y}$ sont les fonctions de répartition respectives de X, Y et (X, Y) .

2. Si X et Y sont discrètes de domaines de variation respectifs $\{x_i : i \in I\}$ et $\{y_j : j \in J\}$, où $I \subseteq \mathbb{N}^*, J \subseteq \mathbb{N}^*$ et les x_i (resp. y_j) sont distincts, X et Y sont indépendantes si et seulement si $\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j)$ pour tout $(i, j) \in I \times J$.
3. Si le couple (X, Y) est absolument continu, X et Y sont indépendantes si et seulement si $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ est une densité de probabilité de (X, Y) pour toute densité f_X de X et toute densité f_Y de Y .

Proposition 14

Soit la somme $X + Y$ de deux variables aléatoires indépendantes X et Y . On a :

- (1) si X et Y sont discrètes et ont pour domaine de variation commun \mathbb{N}^* , alors :

$$\mathbb{P}(X + Y = k) = (\mathbb{P}_X * \mathbb{P}_Y)(k) = \sum_{i=1}^{\infty} \mathbb{P}(X = k - i) \mathbb{P}(Y = i)$$

- (2) si X et Y sont absolument continues de densités respectives f_X et f_Y , alors $X + Y$ est elle aussi absolument continue et a pour densité la convolution $f_{X+Y} = f_X * f_Y$ des densités de X et de Y . On a donc :

$$\forall s \in \mathbb{R}, f_{X+Y}(s) = \int_{-\infty}^{\infty} f_X(s - y) f_Y(y) dy = \int_{-\infty}^{\infty} f_X(x) f_Y(s - x) dx.$$

4.5 Covariance et coefficient de corrélation

L'indépendance entre variables aléatoires est une hypothèse très forte. En effet, on démontre que deux variables aléatoires X et Y sont indépendantes si $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$ pour tout couple (f, g) de fonctions mesurables et pour lesquelles les espérances mises en jeu existent, c'est-à-dire, pour une classe très grande d'applications f et g . L'indépendance n'est donc pas toujours vérifiée, ni même vérifiable en pratique. Une notion plus faible est celle de **décorrélacion** qu'introduit la définition suivante.

Définition 17 (Covariance et coefficient de corrélation)

Soient deux variables aléatoires X et Y .

1. La **covariance** de X et Y est définie par :

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

2. On dit que X et Y sont **décorrélées** si $\text{Cov}(X, Y) = 0$.
3. Le **coefficient de corrélation** de X et Y de variances non nulles est défini par :

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}(X)\mathbb{V}(Y)}}.$$

Proposition 15

Avec les notations de la définition précédente :

1. Si X et Y sont indépendantes, $\text{Cov}(X, Y) = \rho(X, Y) = 0$. L'indépendance est une hypothèse plus forte que la décorrélation, car l'indépendance entraîne la décorrélation. La réciproque est fausse en général.
2. Le coefficient de corrélation prend ses valeurs dans l'intervalle $[-1, 1]$:

$$|\rho(X, Y)| \leq 1.$$

3. X et Y sont presque sûrement liées par une relation affine de la forme $Y = aX + b$ si et seulement si $|\rho(X, Y)| = 1$.

Proposition 16

Considérons la somme $X + Y$ de deux variables aléatoires décorrélées X et Y . La variance de cette somme est la somme des variances de X et de Y :

$$\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y).$$

De manière générale la variance d'une somme de n variables aléatoires décorrélées est la somme des variances de ces variables aléatoires.

5 Convergences de suites de variables aléatoires réelles

Les convergences de suites de variables aléatoires jouent un rôle extrêmement important en théorie des probabilités pour au moins deux raisons fondamentales. Premièrement, un résultat de convergence important connu sous le nom de « loi des grands nombres » permet d'établir un lien formel entre la notion théorique de probabilité et son interprétation intuitive de fréquence d'apparition d'un événement, lorsqu'une expérience aléatoire est répétée à l'infini de façon indépendante⁴. La deuxième raison est, qu'au travers du célèbre théorème central-limite, la notion de convergence en loi permet de justifier l'omniprésence de la loi de probabilité gaussienne en statistique.

Il existe plusieurs critères de convergence et il existe aussi des relations d'implication entre ces différents critères. Dans ce chapitre, nous nous limitons aux critères de convergence les plus usuels en statistique. Ci-dessous, nous travaillerons avec des suites $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires. D'autre part, nous rencontrerons beaucoup de suites $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes et identiquement distribuées (i.i.d.). Cela signifie que les variables aléatoires X_n ont toutes la même loi et que k variables $X_{n_1}, X_{n_2}, \dots, X_{n_k}$ distinctes sont mutuellement indépendantes.

5.1 Convergence en loi

La convergence en loi est le mode de convergence le plus faible dans le sens où il n'implique aucun des autres modes de convergences présentés dans ce chapitre. Il définit

4. D'ailleurs, un estimateur important qui sera étudié dans le thème 2 est modelé par la loi faible des grands nombres

une relation de voisinage non pas entre les variables aléatoires elles-mêmes mais entre leurs fonctions de répartition.

Définition 18 (*Convergence en loi*)

Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers X si

$$\lim_{n \rightarrow +\infty} \mathbb{F}_{X_n}(x) = \mathbb{F}_X(x)$$

en tout point x où \mathbb{F}_X est continue.

Cette convergence est notée : $X_n \xrightarrow{\mathcal{L}} X$.

Exemple 21

Un exemple célèbre de convergence en loi est donné par le théorème de Moivre-Laplace : Soit $X_1, X_2, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{B}(1, p)$ alors la moyenne empirique $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ vérifie la convergence suivante :

$$\frac{\bar{X} - p}{\sqrt{p(1-p)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Notez que la variable aléatoire limite X n'est pas précisée ici, car elle n'est ni unique ni utile dans l'expression de cette convergence où seule la loi intervient. Il est aussi intéressant de remarquer que dans cet exemple une suite de variable aléatoires discrètes converge en loi vers une variable aléatoire absolument continue.

L'exemple précédent est un cas particulier d'un théorème majeur en théorie des probabilités : le **théorème central-limite**, aussi appelé théorème-limite central ou encore théorème de la limite centrale. Dans sa forme la plus simple, ce théorème s'énonce comme suit :

Théorème 17 (*théorème central-limite*)

Si $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires i.i.d. de même espérance μ et de même écart-type $\sigma > 0$ alors, en posant $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, on a :

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Autrement dit, la somme centrée réduite de variables aléatoires i.i.d., de variance finie converge en loi vers la loi normale centrée réduite $\mathcal{N}(0,1)$. A noter qu'il existe d'autres formulations de ce théorème qui sont moins restrictives et qui n'exigent pas que les variables aléatoires soient i.i.d. mais qui imposent des conditions sur les moments de ces variables aléatoires.

Dans la pratique ce théorème est très utile car il nous permet de dire que, pour n suffisamment grand, une somme de variables aléatoires i.i.d. suit approximativement une loi normale.

Exemple 22

Pour n grand :

(1) $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ suit approximativement la loi $\mathcal{N}(\mu, \sigma^2/n)$.

(2) $S_n = \sum_{i=1}^n X_i$ suit approximativement la loi $\mathcal{N}(n\mu, n\sigma^2)$.

Le régime asymptotique à partir duquel n est jugé suffisamment grand pour que l'approximation soit valide dépend très fortement de la loi de probabilité des X_i .

La convergence en loi de la suite $(X_n)_{n \in \mathbb{N}^*}$ vers X ne nous dit rien sur l'écart qui peut exister ou non entre X_n et X lorsque n est grand. Autrement dit, ce n'est pas parce que la suite $(X_n)_{n \in \mathbb{N}^*}$ converge en loi vers X que les valeurs de X_n sont proches de celles de X . Pour décrire l'écart entre X_n et X , il faut définir d'autres modes de convergence.

5.2 Convergence en probabilité

La convergence en probabilité représente très bien la notion de voisinage entre variables aléatoires. Elle est définie comme suit :

Définition 19 (Convergence en probabilité)

Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers X si pour tout réel $\varepsilon > 0$,

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

Cette convergence est notée : $X_n \xrightarrow{\mathbb{P}} X$.

La convergence en probabilité est dite plus forte que la convergence en loi car il y a une relation d'implication entre les deux, à savoir : $(X_n \xrightarrow{\mathbb{P}} X) \Rightarrow (X_n \xrightarrow{\mathcal{L}} X)$. Un très bon exemple d'application de ce mode de convergence est la loi faible des grands nombres.

Théorème 18 (loi faible des grands nombres)

Si $(X_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires i.i.d. de même espérance μ et de même écart-type $\sigma > 0$ alors, en posant $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$:

$$\bar{X} \xrightarrow{\mathbb{P}} \mu.$$

Notez que dans ce cas particulier la variable aléatoire limite est une variable certaine μ . Cette loi est dite faible car il existe une autre formulation plus générale qui repose sur un critère de convergence plus fort que celui de convergence en probabilité.

Ce théorème joue un rôle très important en statistique car il permet de justifier l'utilisation de la moyenne empirique comme un « estimateur » de l'espérance d'une variable aléatoire. Comme indiqué dans l'introduction, cette loi permet aussi de justifier a posteriori la relation qui existe entre la probabilité d'occurrence d'un événement et la fréquence d'apparition de ce même événement dans une suite d'expériences aléatoires indépendantes.

5.3 Convergence en moyenne quadratique

En statistique, la convergence en moyenne quadratique intervient régulièrement dans la justification du choix d'un estimateur. Elle est définie comme suit :

Définition 20 (*Convergence en moyenne quadratique*)

Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ converge en moyenne quadratique vers X si,

$$\lim_{n \rightarrow +\infty} \mathbb{E}((X_n - X)^2) = 0.$$

Cette convergence est notée : $X_n \xrightarrow{m.q.} X$.

Cette convergence est plus forte que la convergence en probabilité car $(X_n \xrightarrow{m.q.} X) \Rightarrow (X_n \xrightarrow{\mathbb{P}} X)$. Cette relation d'implication découle de l'**inégalité de Bienaymé-Tchebychev** qui s'énonce comme suit : pour tout réel strictement positif α , et toute variable aléatoire X d'espérance μ et de variance σ^2 , $\mathbb{P}(|X - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$.

Exemple 23

Si l'on considère toujours la moyenne empirique telle que présentée dans l'énoncé de la loi faible des grands nombres, on peut montrer que $\mathbb{E}(\bar{X}) = \mu$ et que $\mathbb{V}(\bar{X}) = \sigma^2/n$. On peut alors en déduire que

$$\lim_{n \rightarrow +\infty} \mathbb{E}((\bar{X} - \mu)^2) = \lim_{n \rightarrow +\infty} \mathbb{V}(\bar{X}) = 0.$$

Par conséquent, \bar{X} converge en moyenne quadratique vers μ .

5.4 Autres modes de convergence

Il existe d'autres modes de convergence en probabilité qui sont un peu plus anecdotiques dans le contexte de ce cours sur la statistique pour l'ingénieur. On peut citer par exemple la convergence en moyenne d'ordre $q > 0$ quelconque qui est une généralisation de la convergence en moyenne quadratique, à savoir :

Définition 21 (*Convergence en moyenne d'ordre q*)

Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ converge en moyenne d'ordre q vers X si,

$$\lim_{n \rightarrow +\infty} \mathbb{E}((X_n - X)^q) = 0.$$

Il y a également la notion de convergence presque sûre qui est définie comme suit :

Définition 22 (*Convergence presque sûre*)

Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}^*}$ converge presque sûrement vers X si,

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} X_n = X\right) = 1.$$

On peut montrer que la convergence presque sûre implique la convergence en probabilité. Notez que la loi des grands nombres dite forte s'énonce comme la loi faible mais avec une convergence presque sûre et non plus avec une convergence en probabilité.

Les relations d'implication entre toutes les notions de convergence présentées dans ce chapitre sont résumées par la [figure 3](#).

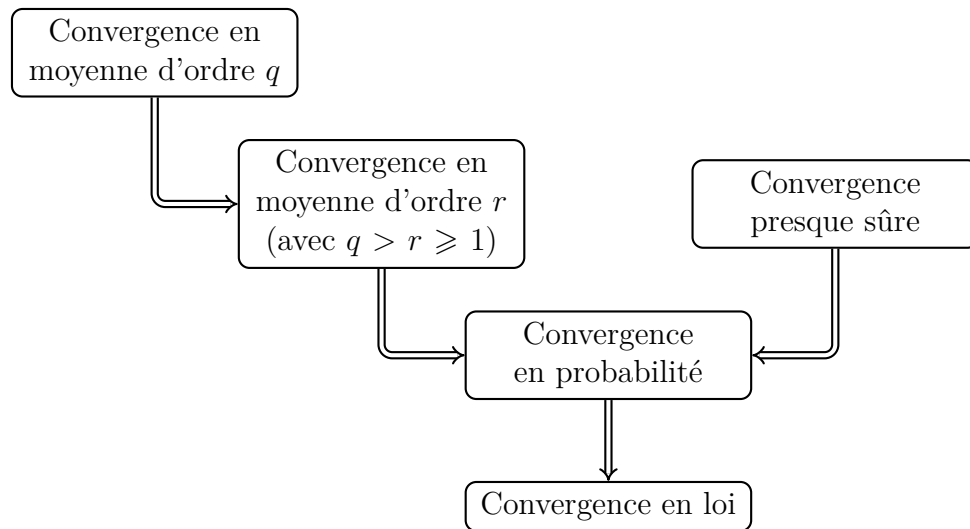


FIGURE 3 – Relations d'implication entre les modes de convergence.

6 Exercices

Exercice 1 : Spams

Dans une entreprise, 40% des courriers électroniques reçus sont des spams. Parmi tous les spams reçus, 50% contiennent le mot «order» et, parmi tous les courriels légitimes, 30% contiennent le mot «order».

1. Quelle est la probabilité qu'un mail arrivant contienne le mot «order» ?
2. Un mail arrivant contient le mot «order», quelle est la probabilité que ce soit un spam ?

Exercice 2 : Durée de vie

La durée de vie d'un certain type de lampes de vidéoprojecteur suit une loi exponentielle de densité de probabilité

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{sinon.} \end{cases}$$

1. Une variable aléatoire X est dite « sans mémoire », ou « sans vieillissement », si elle vérifie la propriété suivante :

$$\mathbb{P}(X > x + x_0 | X > x_0) = \mathbb{P}(X > x),$$

pour tout x, x_0 positifs.

Montrer que la durée de vie de ces lampes de vidéoprojecteur est sans mémoire.

2. Déterminer la durée de vie moyenne d'une lampe en fonction de λ ?
3. Un fabricant de lampes annonce à son client que la durée de vie moyenne des lampes qu'il produit est d'au moins 10000 heures alors que, dans les faits, une lampe sur deux a une durée de vie de moins de 7000 heures. Est-il honnête ? Justifier.

Exercice 3 : Consommation énergétique

Pour effectuer un comparatif entre deux modèles de smartphones, une association de consommateurs s'intéresse à leurs consommations électriques respectives. Pour cette étude, elle dispose de données constructeur qui fournissent des informations sur la variabilité de consommation en fonction de l'usage qui est fait du téléphone.

Pour le modèle 1, le fabricant indique que sa consommation est une variable aléatoire X_1 de fonction de répartition \mathbb{F}_{X_1} et de densité f_{X_1} . Pour le modèle 2, la consommation est une variable aléatoire X_2 de fonction de répartition \mathbb{F}_{X_2} et de densité f_{X_2} . On suppose que les consommations des deux modèles sont indépendantes. L'association de consommateur ne dispose que d'un exemplaire de chaque modèle.

1. Pour un même usage, exprimer la probabilité que l'exemplaire du modèle 1 consomme plus que l'exemplaire du modèle 2 à l'aide de \mathbb{F}_{X_1} et de f_{X_2} .
2. Dans le cas où $\mathbb{F}_{X_1} = \mathbb{F}_{X_2}$, quelle est la valeur de cette probabilité ?

Exercice 4 : Transformation linéaire d'un vecteur gaussien

Soit $X = (X_1, X_2)^T$, un vecteur aléatoire⁵ réel centré, à deux dimensions, de loi gaussienne et de matrice de covariance

$$\Gamma_X = \begin{pmatrix} 3 & \rho\sqrt{3} \\ \rho\sqrt{3} & 1 \end{pmatrix},$$

avec $|\rho| < 1$. On définit un nouveau vecteur aléatoire $(Y_1, Y_2)^T$ avec :

$$\begin{cases} Y_1 &= \frac{X_1}{\sqrt{3}} - X_2, \\ Y_2 &= \frac{X_1}{\sqrt{3}} + X_2. \end{cases}$$

- 1.** Calculer la variance $\mathbb{V}(X_1)$ de X_1 et la covariance $\mathbb{Cov}(X_1, X_2)$ de (X_1, X_2) ?
- 2.** Dans quel cas, les deux variables X_1 et X_2 sont-elles indépendantes ?
- 3.** Calculer la covariance $\mathbb{Cov}(Y_1, Y_2)$ du couple (Y_1, Y_2) . Calculer les variances des variables aléatoires Y_1 et Y_2 .
- 4.** Déterminer les densités de probabilités marginales de Y_1 et Y_2 .
- 5.** Les deux variables Y_1 et Y_2 sont-elles indépendantes ? Justifier clairement votre réponse.

Exercice 5 : Application non probabiliste du théorème central-limite

On considère une suite $(X_n)_{n \in \mathbb{N}^*}$ de variables aléatoires indépendantes, qui suivent des lois de Poisson de paramètre unité : pour tout $n = 1, 2, \dots$, $X_n \sim \mathcal{P}(1)$. On pose $S_n = X_1 + X_2 + \dots + X_n$.

Pour l'exercice, on admettra que la somme de deux variables aléatoires de Poisson indépendantes X_1 et X_2 , telles que $X_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 \sim \mathcal{P}(\lambda_2)$ suit une loi de Poisson : $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

- 1.** Quelle est la loi de S_n ? Soit \mathbb{F}_n la fonction de répartition de la variable aléatoire S_n . Montrer que $\mathbb{F}_n(n) = e^{-n} \sum_{k=0}^n \frac{n^k}{k!}$.
- 2.** Montrer que $Z_n = \frac{S_n - n}{\sqrt{n}}$ converge en loi vers la loi normale centrée réduite.
- 3.** Dédurre des questions précédentes que $\lim_{n \rightarrow +\infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!} = 1/2$.

5. Reportez-vous à la vidéo 2 du thème 1 de ce MOOC pour la représentation d'un couple de variables aléatoires sous forme vectorielle.

Références

- [1] P. Billingsley, *Probability and Measure*, 3^e édition, Wiley, Hoboken, 1995.
- [2] D. Pastor and C. Sintes, *Probabilités pour l'ingénieur*, Lavoisier, Paris, 2014.
- [3] J. Jacod and P. Protter, *L'essentiel en théorie des probabilités*, Cassini, Paris, 2003.
- [4] D. Ghorbanzadeh, *Probabilités, exercices corrigés*, Technip, Paris, 1998.