

École Polytechnique Universitaire Pierre et Marie Curie

COURS D'INTRODUCTION

Électronique et informatique
Sciences de la Terre
Robotique

Probabilités

2014

Sommaire

1	Bases de la théorie des probabilités	7
1.1	Espaces probabilisables	7
1.1.1	Expérience aléatoire et événements	7
1.1.2	Les événements	9
1.2	Espace probabilisé $(\Omega, \mathcal{E}, \mathbf{P})$	13
1.2.1	L'axiomatique de KOLMOGOROV	13
1.2.2	Propriétés élémentaires	14
1.2.3	Équiprobabilité	14
1.3	Lois de probabilités conditionnelles, indépendance	16
1.3.1	Introduction et définitions	16
1.3.2	Indépendance (stochastique)	20
1.3.3	Formules de BAYES	22
2	Variables aléatoires réelles	23
2.1	Loi de probabilité et moments d'une variable aléatoire	23
2.1.1	Définition et fonction de répartition	23
	Le cas d'une variable aléatoire discrète	26
	Le cas d'une variable aléatoire continue	27
	Le cas général	30
2.1.2	Loi d'une fonction d'une variable aléatoire $\mathbf{Y} = \varphi(\mathbf{X})$	31
2.1.3	Indépendance de deux variables aléatoires	32
2.1.4	Espérance, variance, moments d'une variable aléatoire	33
	Espérance	33
	Moments, variance, écart-type	36
	Inégalités de MARKOV et BIENAYMÉ-TCHEBYSHEV	38
2.2	Lois de probabilité d'usage courant	39
2.2.1	Lois discrètes	39
	Loi discrète uniforme	39
	Loi de BERNOULLI	39
	Loi binomiale	40
	Loi de POISSON	41
	Lois géométrique et hypergéométrique	43
2.2.2	Lois continues	44
	Loi uniforme	44
	Loi exponentielle	44
	Loi normale (de LAPLACE-GAUSS)	45
	Loi Gamma	47
	Loi du χ^2	47
2.3	Convolution, loi d'une somme	48
2.3.1	Cas discret	48
2.3.2	Cas continu	49

2.4	Fonctions caractéristiques	51
2.4.1	Définition et principales propriétés	51
2.4.2	Fonctions caractéristiques et moments	53
2.4.3	Fonctions caractéristiques des lois usuelles	55
2.5	Convergences des suites de variables aléatoires	56
2.5.1	Différents type de convergence	56
	La convergence en probabilité	56
	La convergence presque sûre	57
	La convergence en moyenne quadratique	57
	La convergence en loi	57
2.5.2	Hierarchie des différentes convergences	58
2.6	Applications: loi des grands nombres, théorème central-limite	59
2.6.1	Lois des grands nombres	59
2.6.2	Autres types de théorèmes	61
2.6.3	Somme de variables aléatoires et théorème central-limite	64
3	Couples de variables aléatoires	69
3.1	Lois conjointes et lois marginales	69
3.1.1	Fonction de répartition conjointe et marginales	69
	Lois conjointes et lois marginales. Cas discret	70
	Lois conjointes et lois marginales. Cas continu	71
3.1.2	Covariance et coefficient de corrélation	74
	Espérances	74
	Variances, covariances et coefficient de corrélation	74
	Cas d'un vecteur gaussien	77
3.2	Lois conditionnelles	77
A		79
A.1	Rappels sur le dénombrements	79
A.2	Lois discrètes classiques	80
A.3	Lois continues classiques	81
A.4	Tables de probabilité et statistiques	82
	A.4.1 La loi de POISSON	82
	A.4.2 La loi normale	83
B	Compléments en probabilité	85
B.1	Lemme de BOREL-CANTELLI	85
B.2	Fonctions génératrices	85
B.3	Espérances conditionnelles et martingales	86
B.4	Chaînes de MARKOV	86
B.5	Un exemple d'inégalité de concentration	87
B.6	Entropie (de SHANNON)	87

Chapitre 1

Bases de la théorie des probabilités

L'objet de la théorie des probabilités est de fournir un formalisme mathématique précis, propre à décrire des situations dans lesquelles intervient le "hasard", c'est-à-dire des situations dans lesquelles un certain nombre de conditions étant réunies (les causes), plusieurs conséquences sont possibles (les effets) sans que l'on puisse a priori savoir laquelle sera réalisée. Une telle situation apparaît lors d'une expérience aléatoire ou stochastique (par opposition à une expérience déterministe pour laquelle l'issue est certaine).

1.1 Espaces probabilisables

On expose ici la formalisation d'une expérience aléatoire.

1.1.1 Expérience aléatoire et événements

Intuitivement, on a vu qu'une expérience est aléatoire si on ne peut prévoir par avance son résultat et si, répétée dans des conditions "parfaitement" identiques, elle peut donner lieu à des résultats différents. On veut modéliser des expériences telles que celles décrites par les exemples suivants :

- Exemple 1.1.**
1. On jette un dé et l'on observe le numéro de la face obtenue.
 2. On lance trois fois de suite la même pièce de monnaie. Si l'on désigne par P la sortie du côté "pile" et par F la sortie du côté "face", on peut distinguer huit cas possibles:

PPP, PPF, PFP, \dots , FFF.

3. Dans un lot de pièces fabriquées dans une usine et comprenant un certain nombre de pièces défectueuses, on prélève au hasard n pièces et l'on s'intéresse au nombre des pièces défectueuses contenues dans cet échantillon.
4. On joue aux fléchettes. On s'intéresse au nombre d'essais nécessaires pour atteindre la cible.
5. On mesure la durée du bon fonctionnement d'un dispositif technique choisi au hasard parmi un grand nombre de dispositifs identiques.
6. On observe, entre les instants 0 et T, l'intensité d'un signal électrique continu.

En conclusion, parler d'une expérience aléatoire, c'est en particulier donner les résultats possibles de cette expérience.

Définition 1 (Univers des possibles).

On représente le résultat de cette expérience comme un élément ω de l'ensemble Ω de tous les résultats possibles. On appelle l'ensemble Ω **l'espace des éventualités, ou l'univers des possibles**.

Avec les exemples 1.1, on constate que l'univers Ω peut être :

- fini (trois premiers exemples),
- infini dénombrable (quatrième exemple),
- infini et ayant la puissance du continu (cinquième et sixième exemples).

Il convient de noter ici qu'il n'y a pas unicité de l'ensemble Ω . Une des principales difficultés dans la modélisation d'un phénomène concret est de déterminer un univers Ω adapté au problème.

Exemple 1.2. Ainsi à l'expérience aléatoire qui consiste à lancer deux dés, on peut associer :

- l'ensemble Ω des couples (i.e avec ordre) formés par les deux chiffres

$$\Omega = \{(1, 1), (1, 2), \dots, (2, 1), \dots, (6, 5), (6, 6)\}$$

- l'ensemble Ω' des ensembles (i.e sans ordre) formés par les deux chiffres

$$\Omega' = \{\{1, 1\}, \{1, 2\}, \dots, \{5, 6\}, \{6, 6\}\}$$

- et, si l'on convient une fois pour toutes qu'on ne retiendra que la somme des points affichés, on peut associer l'ensemble des sommes possibles

$$\Omega'' = \{2, 3, \dots, 12\}$$

Intuitivement, un **événement aléatoire** (ou plus simplement un **événement**) est une assertion ou proposition logique relative au résultat de l'expérience (par exemple, la somme des points est paire). On dira qu'un événement est réalisé ou non suivant que la proposition est vraie ou fautive une fois l'expérience accomplie. Donc, à un événement, on peut associer la partie de Ω constituée de tous les résultats réalisant l'événement. Nous décidons :

Définition 2 (Événement).

Un **événement** A est une partie de Ω , c'est-à-dire $A \in \mathcal{P}(\Omega)$.

Si l'événement est réduit à un seul élément, on parle d'**événement élémentaire**.

Exemple 1.3. • Prenons le cas du jet d'un seul dé. L'espace des éventualités est défini par

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Considérons les événements suivants:

$$A = \{6\}, \quad B = \text{« numéro pair »} = \{2, 4, 6\} \quad \text{et} \quad C = \text{« numéro } \geq 4 \text{ »} = \{4, 5, 6\}$$

A est un événement élémentaire. On jette le dé, le « 6 » apparaît. Les événements A , B , C et Ω se trouvent réalisés. Dans ce cas, l'espace Ω est fini, un événement quelconque peut se définir par une simple énumération.

- Mais dans le cas où Ω n'est pas fini, on doit définir un événement par une propriété caractéristique. Ainsi dans le sixième exemple 1.1, l'espace Ω peut être l'ensemble des fonctions continues sur l'intervalle $[0, T]$. Cet ensemble peut être noté $\mathcal{C}^0([0, T])$ et si l'on s'intéresse à l'événement A défini de la manière suivante :

$A = \text{“le courant moyen est supérieur à 3 ampères”}$

On écrira $\Omega = \mathcal{C}^0([0, T])$ et $A = \left\{ I \in \mathcal{C}^0([0, T]) : \frac{1}{T} \int_0^T I(t) dt \geq 3 \right\}$

1.1.2 Les événements

Les événements A résultant d'une expérience aléatoire étant par définition des parties de Ω , l'ensemble des événements est un sous-ensemble de $\mathcal{P}(\Omega)$. Il est possible d'appliquer aux événements les opérations de la théorie des ensembles en employant une terminologie adaptée.

Terminologie des événements aléatoires

- Un événement A est **certain** si

$$A = \Omega$$

- Un événement A est **impossible** si

$$A = \emptyset$$

- \bar{A} est l'**événement contraire** d'un événement A si

$$\bar{A} = \Omega \setminus A$$

\bar{A} est réalisé si A ne l'est pas, et réciproquement :

$$\omega \in \bar{A} \quad \text{si et seulement si} \quad \omega \notin A$$

- Deux événements A_1 et A_2 sont **incompatibles** (ou mutuellement exclusifs) si

$$A_1 \cap A_2 = \emptyset$$

Ils n'ont aucune éventualité en commun.

- **L'intersection** (conjonction) de deux événements A_1 et A_2 est l'événement “ A_1 et A_2 ” qui est réalisé si A_1 et A_2 le sont. La conjonction de A_1 et A_2 est représentée par

$$A_1 \cap A_2 : \quad \omega \in A_1 \cap A_2 \quad \text{si et seulement si} \quad \omega \in A_1 \text{ et } \omega \in A_2$$

- **L'union** de deux événements A_1 et A_2 est l'événement “ A_1 ou A_2 ” qui est réalisé si l'un au moins de ces événements l'est. L'union de A_1 et A_2 est représentée par

$$A_1 \cup A_2 : \quad \omega \in A_1 \cup A_2 \quad \text{si et seulement si} \quad \omega \in A_1 \text{ ou } \omega \in A_2$$

- L'événement A_1 **implique** l'événement A_2 si

$$A_1 \subset A_2$$

A_1 étant réalisé, A_2 l'est aussi $\omega \in A_1 \Rightarrow \omega \in A_2$

- **Un système complet** (ou exhaustif) d'événements de Ω est une famille *finie ou dénombrable* $(A_n)_{n \in \mathbb{N}}$ d'événements différents de \emptyset et deux à deux incompatibles, telle que

$$\bigcup_n A_n = \Omega$$

Cette famille forme une partition de Ω .

Résumons le tout dans un tableau:

Langage probabiliste	Notation	Langage ensembliste
Issue ou résultat	ω ($\omega \in \Omega$)	élément de Ω
Événement A	$A \subset \Omega$	partie de Ω
A est réalisé	$\omega \in A$	appartenance
Événement contraire (non-A)	$\bar{A} = \Omega \setminus A$	complémentaire
A et B	$A \cap B$	intersection
A ou B	$A \cup B$	union
Événements incompatibles	$A \cap B = \emptyset$	disjoints
A implique l'événement B	$A \subset B$	inclusion
Événement impossible	\emptyset	ensemble vide
Événement certain	Ω	partie pleine
Système complet d'événements A_n	$\Omega = \bigcup_n A_n$ et $A_i \cap A_j = \emptyset$ pour $i \neq j$	partition

Exemple 1.4. Lors du jet d'un dé, définissons un événement $A_1 = \{2, 4, 6\}$ comme la sortie d'un numéro pair et $A_2 = \{4, 5, 6\}$ comme la sortie d'un numéro supérieur ou égal à 4. On a

$$A_1 \cap A_2 = \{4, 6\}, \quad A_1 \cup A_2 = \{2, 4, 5, 6\} \quad \text{et} \quad \bar{A}_1 = \{1, 3, 5\}$$

Notons que les événements A_1 et \bar{A}_1 forment un système exhaustif d'événements.

Exemple 1.5. Considérons le cas du fonctionnement de deux machines automatiques, chacune d'elles pouvant tomber en panne au cours d'un laps de temps donné. Sans expliciter Ω , considérons les événements suivants:

- A_1 : "la première machine fonctionne",
- A_2 : "la seconde machine fonctionne",
- B_k : "exactement k machines fonctionnent",
- C : "au moins une machine fonctionne".

On a $B_0 = \bar{A}_1 \cap \bar{A}_2$, $B_1 = (\bar{A}_1 \cap A_2) \cup (\bar{A}_2 \cap A_1)$, $B_2 = A_1 \cap A_2$ et $C = A_1 \cup A_2$

Opérations élémentaires sur les événements

Les opérations élémentaires sur les événements possèdent bien évidemment les propriétés fondamentales des opérations sur les ensembles

$$A \cap \emptyset = \emptyset, \quad A \cup \emptyset = A \quad \text{et} \quad \overline{\bar{A}} = A \tag{1.1}$$

$$\left. \begin{aligned} A_1 \cap (A_2 \cap A_3) &= (A_1 \cap A_2) \cap A_3 \\ A_1 \cup (A_2 \cup A_3) &= (A_1 \cup A_2) \cup A_3 \end{aligned} \right\} \text{ associativité} \quad (1.2)$$

$$\left. \begin{aligned} A_1 \cap A_2 &= A_2 \cap A_1 \\ A_1 \cup A_2 &= A_2 \cup A_1 \end{aligned} \right\} \text{ commutativité} \quad (1.3)$$

$$\left. \begin{aligned} A_1 \cap (A_2 \cup A_3) &= (A_1 \cap A_2) \cup (A_1 \cap A_3) \\ A_1 \cup (A_2 \cap A_3) &= (A_1 \cup A_2) \cap (A_1 \cup A_3) \end{aligned} \right\} \text{ distributivité} \quad (1.4)$$

$$\left. \begin{aligned} \overline{A_1 \cap A_2} &= \overline{A_1} \cup \overline{A_2} \\ \overline{A_1 \cup A_2} &= \overline{A_1} \cap \overline{A_2} \end{aligned} \right\} \text{ lois de MORGAN} \quad (1.5)$$

Afin de définir proprement la notion d'ensemble mesurable et donc la notion de probabilité, nous auront besoin de considérer un sous-ensemble des parties de Ω .

Définition 3 (Tribu).

Une **tribu** ou **σ -algèbre** sur Ω est un ensemble \mathcal{E} vérifiant les propriétés suivantes

1. $\Omega \in \mathcal{E}$
2. Stabilité par passage au complémentaire


$$\forall A \in \mathcal{E}, \quad \overline{A} = \Omega \setminus A \in \mathcal{E} \quad (1.6)$$

3. Stabilité par union finie ou dénombrable

$$\forall A_n \in \mathcal{E}, \quad \bigcup_n A_n \in \mathcal{E} \quad (1.7)$$

Il est facile de vérifier à partir des opérations élémentaires précédentes que l'ensemble \mathcal{E} des événements constitue une tribu de Ω . Ces trois propriétés ont deux conséquences immédiates

- Stabilité par intersection finie ou dénombrable.
- $\emptyset \in \mathcal{E}$

 Sur un même ensemble Ω des éventualités, on peut définir de nombreuses tribus \mathcal{E} d'événements différentes.

Exemple 1.6. • La tribu \mathcal{E} contenant le plus d'événements est la tribu $\mathcal{P}(\Omega)$ de toutes les parties de Ω . Toutes les autres tribus sur Ω sont des sous-tribus de $\mathcal{P}(\Omega)$.

- La tribu \mathcal{E} contenant le moins d'événements est la tribu réduite aux deux seules parties Ω et \emptyset . Elle est contenue dans toutes les autres.
- Si F est une famille de parties de Ω , on appelle **tribu engendrée** par F , la plus petite tribu contenant tous les éléments de F . Ainsi la tribu engendrée par une partie A de Ω , est formée des éléments

$$\{\emptyset, \Omega, A, \overline{A}\}$$

On peut maintenant donner la définition

Définition 4 (Espace probabilisable).

On appelle **espace probabilisable** le couple (Ω, \mathcal{E}) où \mathcal{E} constitue une tribu de parties de Ω .

1.2 Espace probabilisé $(\Omega, \mathcal{E}, \mathbf{P})$ **1.2.1 L'axiomatique de KOLMOGOROV****Définition 5** (Une probabilité).

Soit (Ω, \mathcal{E}) un espace probabilisable. Une **probabilité** est une application \mathbf{P} réelle définie sur \mathcal{E} vérifiant les trois axiomes suivants appelés axiomes de KOLMOGOROV

1. $\mathbf{P} : \mathcal{E} \rightarrow \mathbb{R}^+$
2. $\mathbf{P}(\Omega) = 1$
3. Pour toute suite finie ou dénombrable d'événements *deux à deux incompatibles*

$$\mathbf{P} \left(\bigcup_{n \in I} A_n \right) = \sum_{n \in I} \mathbf{P}(A_n) \quad (1.8)$$

Définition 6 (Espace probabilisé).

Un **espace probabilisé** est la donnée d'un triplet $(\Omega, \mathcal{E}, \mathbf{P})$ avec

- (Ω, \mathcal{E}) un espace probabilisable,
- \mathbf{P} une probabilité sur \mathcal{E} .

Remarque 1.7. La théorie des probabilités s'inscrit dans le cadre de la théorie de la mesure : une loi de probabilité n'est rien d'autre qu'une mesure positive de masse totale 1.

1.2.2 Propriétés élémentaires

Proposition 7.

On a les propriétés suivantes pour tout $(A, B) \in \mathcal{E}^2$

1. $\mathbf{P}(\Omega) = 1, \mathbf{P}(\emptyset) = 0,$
2. $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A),$
3. $\mathbf{P}(A) \in [0, 1],$
4. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B),$
5. $A \subset B$ implique $\mathbf{P}(A) \leq \mathbf{P}(B).$

Théorème des probabilités totales.

Soit $\{B_i : i \in I\}$ un système complet d'événements, alors

$$\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \cap B_i) \quad (1.9)$$

Preuve. Elles découlent directement des axiomes de KOLMOGOROV. En effet, A et \bar{A} étant incompatibles, par l'axiome (1.8)

$$\mathbf{P}(\bar{A}) + \mathbf{P}(A) = \mathbf{P}(\Omega) = 1$$

En particulier cela démontre que la probabilité d'un événement est toujours comprise dans $[0, 1]$. Le résultat 4 provient du fait que si A et B ne sont pas disjoints, on peut toujours écrire $A \cup B = A \cup (B \setminus (A \cap B))$. Puis 4 implique 5. ■

1.2.3 Équiprobabilité

Un cas particulier important est le cas d'équiprobabilité quand Ω est de cardinal fini. On dit qu'il y a **équiprobabilité dans le cas où les événements élémentaires ont tous la même probabilité**. On a alors les résultats suivants, si $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$

$$\forall i \in \llbracket 1, n \rrbracket, \quad \mathbf{P}(\{\omega_i\}) = \frac{1}{n} \quad (1.10)$$

et plus généralement,

$$\forall A \in \mathcal{E}, \quad \mathbf{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} \quad (1.11)$$

Le calcul des probabilités peut, alors, se ramener à des calculs de dénombrement, il faut déterminer le cardinal d'ensembles. On pourra se reporter à l'annexe A.1.

Exemple 1.8. Si l'on jette un dé unique, on a

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

et si le dé n'est pas pipé, il est raisonnable de considérer tous les événements élémentaires comme équiprobables. On a alors

$$\mathbf{P}(\{1\}) = \mathbf{P}(\{2\}) = \dots = \mathbf{P}(\{6\}) = \frac{1}{6}$$

L'événement $A = \{1, 5\}$ a pour probabilité $1/3$.

Exemple 1.9. Si maintenant on jette deux dés (un bleu et un rouge pour fixer les idées) et que chaque éventualité est constituée par la somme k des deux numéros indiqués par les faces, on a au moins (voir l'exemple 1.2) trois univers possibles :

Le plus naturel $\Omega'' = \{2, 3, 4, \dots, 12\}$, a le désavantage que les événements élémentaires ne sont pas équiprobables.

En revanche, le fait de pouvoir distinguer les deux dés (par leur couleur, par leur ordre, par exemple) nous permet de considérer l'ensemble Ω des couples de deux chiffres

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$$

où il y a équiprobabilité. Dans l'ensemble

$$\Omega' = \{\{1, 1\}, \{1, 2\}, \dots, \{5, 6\}, \{6, 6\}\}$$

des ensembles formes par les deux chiffres, il n'y aurait pas non plus équiprobabilité. Il vient, à l'aide de la modélisation d'équiprobabilité dans Ω

$$\begin{aligned} \mathbf{P}(k = 2) = \mathbf{P}(k = 12) &= \frac{1}{36} & \mathbf{P}(k = 3) = \mathbf{P}(k = 11) &= \frac{2}{36} = \frac{1}{18} \\ \mathbf{P}(k = 4) = \mathbf{P}(k = 10) &= \frac{3}{36} = \frac{1}{12} & \mathbf{P}(k = 5) = \mathbf{P}(k = 9) &= \frac{4}{36} = \frac{1}{9} \\ \mathbf{P}(k = 6) = \mathbf{P}(k = 8) &= \frac{5}{36} & \text{et } \mathbf{P}(k = 7) &= \frac{6}{36} = \frac{1}{6} \end{aligned}$$

1.3 Lois de probabilités conditionnelles, indépendance

1.3.1 Introduction et définitions

Le but de ce paragraphe est de modéliser ce que l'on entend par

- deux événements sont indépendants.
- la réalisation d'un événement conditionne la réalisation d'un autre.

La notion de probabilité conditionnelle peut être nécessaire à chaque fois que pendant le déroulement d'une expérience aléatoire, on dispose d'une information partielle. Si on sait que l'événement A est réalisé, pour que l'événement B se réalise, on est amené à regarder l'événement $A \cap B$, puis à normaliser. Nous prenons la propriété-définition suivante :

Définition 8 (Probabilité conditionnelle).

Soit $(\Omega, \mathcal{E}, \mathbf{P})$ un espace probabilisé et A un événement **possible** ($\mathbf{P}(A) \neq 0$). L'application

$$\mathbf{P}_A : \begin{cases} \mathcal{E} & \rightarrow [0, 1] \\ B & \mapsto \mathbf{P}_A(B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} \end{cases}$$

est une probabilité sur (Ω, \mathcal{E}) appelée **probabilité conditionnelle sachant A** .

Preuve. Prouvons que \mathbf{P}_A définit bien une probabilité. On a $\mathbf{P}_A(\Omega) = 1$ et pour tout $B \in \mathcal{E}$

$$0 \leq \mathbf{P}_A(B) \leq 1$$

D'autre part l'additivité dénombrable est évidente, cela résulte de la distributivité de l'intersection par rapport à la réunion. Pour toute suite finie ou dénombrable $(B_i)_i$, d'événements deux à deux incompatibles, on a

$$\mathbf{P}_A \left(\bigcup_{i \in I} B_i \right) = \sum_{i \in I} \mathbf{P}_A(B_i)$$

Remarque 1.10. • $\mathbf{P}_A(B)$ peut se lire aussi “probabilité de B quand A” ou “probabilité de B si A”.

- Une autre notation est souvent utilisée : $\mathbf{P}(B|A)$. Elle a l'inconvénient que “B|A” n'est pas un événement et que \mathbf{P} est définie pour une partie de Ω ; mais cette notation a l'avantage d'être “parlante”. Dans la suite de ce polycopié, on utilisera dorénavant cette notation.
- Il sera nécessaire d'étendre ultérieurement la notion de probabilité conditionnelle lorsque A est de probabilité nulle.

Exemple 1.11. Reprenons l'exemple 1.9 du jet de deux dés (de couleurs différentes). Dans

$$\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\},$$

il y a équiprobabilité. Considérons les événements :

- A : “la somme des points obtenues est au moins égale à 10”
- B : “le dé bleu amène un 3”
- C : “le dé bleu amène un 6”

C'est-à-dire

- $A = \{(4, 6), (5, 5), (5, 6), (6, 4), (6, 5), (6, 6)\}$
- $B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$
- $C = \{(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$

D'après les calculs faits dans l'exemple 1.9, $\mathbf{P}(A) = 1/6$.

Mais si on connaît le nombre amené par le dé bleu, la probabilité que A soit réalisé dépend du résultat du dé bleu. Par exemple, on a :

- sachant que l'événement B est réalisé, l'événement A ne peut pas se réaliser, car $A \cap B = \emptyset$. La probabilité de A sachant que B est réalisé, est nulle. Par ailleurs,

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = 0$$

- sachant que l'événement C est réalisé, parmi les 6 éléments de C, 3 réalisent l'événement A, l'événement A a 3 chances sur 6 de se réaliser alors

$$\mathbf{P}(A|C) = \frac{\mathbf{P}(A \cap C)}{\mathbf{P}(C)} = \frac{1}{2}$$

Une conséquence immédiate de la définition est la formule des probabilités composées

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B|A) \quad (1.12)$$

C'est souvent sous cette forme que sera utilisé le conditionnement, ainsi que sous sa forme généralisée.

Proposition 9.

Soient A_1, A_2, \dots, A_n , des événements d'un espace probabilisé vérifiant

$$\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0 \quad (1.13)$$

On a alors

$$\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2|A_1) \cdot \mathbf{P}(A_3|A_1 \cap A_2) \cdot \dots \cdot \mathbf{P}(A_n|A_1 \cap \dots \cap A_{n-1})$$

Preuve. La démonstration se fait facilement par récurrence, en observant que l'hypothèse (1.13) implique que pour tout indice $i \in \llbracket 1, n-1 \rrbracket$, $\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_i) \neq 0$ par croissance de la probabilité. ■

Exemple 1.12. Une urne contient 15 boules : 6 sont rouges, 4 sont blanches et 5 sont bleues. On se propose de trouver la probabilité pour que, en tirant 3 boules successivement de l'urne, la première soit rouge, la seconde soit blanche et la troisième soit bleue. Introduisons les événements suivants

- R_i : { la i -ème boule est rouge },
- W_j : { la j -ème boule est blanche },
- B_k : { la k -ème boule est bleue }.

On cherche $\mathbf{P}(R_1 \cap W_2 \cap B_3)$. Si après chaque tirage, les boules sont remises dans l'urne et parfaitement mélangées, on a :

$$\mathbf{P}(R_1) = \frac{6}{15}, \quad \mathbf{P}(W_2|R_1) = \frac{4}{15} \quad \text{et} \quad \mathbf{P}(B_3|W_2 \cap R_1) = \frac{5}{15}$$

$$\mathbf{P}(R_1 \cap W_2 \cap B_3) = \mathbf{P}(R_1) \cdot \mathbf{P}(W_2|R_1) \cdot \mathbf{P}(B_3|R_1 \cap W_2) = \frac{6}{15} \times \frac{4}{15} \times \frac{5}{15} = \frac{8}{225}$$

Si après chaque tirage, les boules ne sont pas remises dans l'urne, on a

$$\mathbf{P}(R_1) = \frac{6}{15}, \quad \mathbf{P}(W_2|R_1) = \frac{4}{14}, \quad \mathbf{P}(B_3|R_1 \cap W_2) = \frac{5}{13}$$

et
$$\mathbf{P}(R_1 \cap W_2 \cap B_3) = \mathbf{P}(R_1) \cdot \mathbf{P}(W_2|R_1) \cdot \mathbf{P}(B_3|R_1 \cap W_2) = \frac{6}{15} \times \frac{4}{14} \times \frac{5}{13} = \frac{4}{91}$$

Une autre conséquence très importante est la formule des probabilités totales

Proposition 10 (Formule des probabilités totales).

Soient $(\Omega, \mathcal{E}, \mathbf{P})$ un espace probabilisé et $\{A_i\}$ un système complet d'événements tous de probabilité non nulle. Alors, pour tout événement B on a :

$$\mathbf{P}(B) = \sum_{i=1}^n \mathbf{P}(A_i) \mathbf{P}(B|A_i) = \mathbf{P}(A_1) \mathbf{P}(B|A_1) + \dots + \mathbf{P}(A_n) \mathbf{P}(B|A_n) \quad (1.14)$$

Preuve. On sait :

$$\mathbf{P}(B) = \mathbf{P}(B \cap A_1) + \mathbf{P}(B \cap A_2) + \cdots + \mathbf{P}(B \cap A_n)$$

et, en appliquant la formule des probabilités composées, on obtient le résultat énoncé. ■

Exemple 1.13. Trois usines fabriquent des pièces de même nature. Dans l'usine A, 1% des pièces sont défectueuses, dans l'usine B, 10% des pièces sont défectueuses et dans l'usine C, 3% des pièces sont défectueuses. On mélange en proportions égales, les productions en provenance de chacune des trois usines et l'on en tire une pièce au hasard. L'espace Ω est formé des pièces, c'est un ensemble fini, il est acceptable de supposer qu'il y a équiprobabilité. Introduisons les événements suivants

- D : { la pièce tirée est défectueuse },
- A_1 : { la pièce provient de l'usine A },
- A_2 : { la pièce provient de l'usine B },
- A_3 : { la pièce provient de l'usine C }.

$$\mathbf{P}(A_1) = \mathbf{P}(A_2) = \mathbf{P}(A_3) = \frac{1}{3}$$

$$\mathbf{P}(D|A_1) = \frac{1}{100}, \quad \mathbf{P}(D|A_2) = \frac{1}{10} \quad \text{et} \quad \mathbf{P}(D|A_3) = \frac{3}{100}$$

d'après la formule des probabilités totales, la probabilité pour que la pièce tirée soit défectueuse est

$$\mathbf{P}(D) = \mathbf{P}(D \cap A_1) + \mathbf{P}(D \cap A_2) + \mathbf{P}(D \cap A_3) = \frac{7}{150}$$

1.3.2 Indépendance (stochastique)

Si le fait que A est réalisé ne change pas la probabilité de B, autrement dit si

$$\mathbf{P}(B|A) = \mathbf{P}(B)$$

alors
$$\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \mathbf{P}(A)$$

D'où le choix de la définition pour l'indépendance de deux événements :

Définition 11 (indépendance deux à deux).

Soient A et B deux événements appartenant à la même algèbre \mathcal{E} , on dit que A et B sont deux événements (stochastiquement) **indépendants** pour la probabilité \mathbf{P} si

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \times \mathbf{P}(B) \tag{1.15}$$

Remarque 1.14. Si l'une des paires $\{A, \bar{B}\}$, $\{\bar{A}, B\}$, $\{\bar{A}, \bar{B}\}$, $\{A, B\}$ est constituée d'événements indépendants, alors les trois autres le sont aussi. En effet, supposons, par exemple que les événements A et B soient indépendants. Montrons que A et \bar{B} le sont aussi. Ecrivons :

$$A = (A \cap B) \cup (A \cap \bar{B}) \quad \text{et} \quad \emptyset = (A \cap B) \cap (A \cap \bar{B})$$

de l'additivité de la probabilité, il résulte que :

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

alors
$$\mathbf{P}(A \cap \bar{B}) = \mathbf{P}(A) - \mathbf{P}(A \cap B) = \mathbf{P}(A)(1 - \mathbf{P}(B)) = \mathbf{P}(A) \cdot \mathbf{P}(\bar{B})$$

 Deux événements peuvent être indépendants pour une certaine loi et ne pas l'être pour une autre...

Donnons maintenant une généralisation avec plusieurs événements.

Définition 12 (Indépendance mutuelle).

Soient des événements A_1, A_2, \dots, A_n . Ils sont dits **mutuellement indépendants** si pour toute partie non vide I de l'ensemble des indices allant de 1 à n on a :

$$\mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i) \quad (1.16)$$

Remarque 1.15. Cette condition est beaucoup plus forte que l'indépendance deux à deux, comme le montre l'exemple suivant.

Exemple 1.16. Sur l'exemple 1.2 du jet de deux dés, montrons que le fait que trois événements A, B, C soient deux à deux indépendants n'implique pas que l'on ait

$$\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C)$$

ni vice versa, le fait que l'égalité précédente se trouve vérifiée n'implique pas que les trois événements soient deux à deux indépendants. Ici, $\Omega = \{(1, 1), (1, 2), \dots, (6, 5), (6, 6)\}$, et il y a équiprobabilité. Soient i et j étant deux nombres entiers compris entre 1 et 6, on a

$$\mathbf{P}(\{i, j\}) = \frac{1}{36}$$

Considérons maintenant les trois événements suivants

$$A: \text{“}i \text{ est pair”}; \quad B: \text{“}j \text{ est impair”}; \quad C: \text{“}i + j \text{ est pair”}$$

On vérifie aisément (en comptant) que $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{1}{2}$

D'autre part, on a

- $A \cap B$: “ i est pair, j est impair”
- $B \cap C$: “ i est impair, j est impair”
- $A \cap C$: “ i est pair, j est pair”

et $\mathbf{P}(A \cap B) = \frac{9}{36} = \frac{1}{4} = \mathbf{P}(A) \cdot \mathbf{P}(B)$

De même $\mathbf{P}(B \cap C) = \mathbf{P}(B) \cdot \mathbf{P}(C)$ et $\mathbf{P}(A \cap C) = \mathbf{P}(A) \cdot \mathbf{P}(C)$. On en conclut que pris deux à deux les événements A, B, C sont indépendants (pour la probabilité \mathbf{P}). Cependant

$$\mathbf{P}(A \cap B \cap C) \neq \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(C)$$

Considérons maintenant l'événement

$$D = \text{“}1 < i \cdot j \leq 3\text{”} = \{(1, 2), (2, 1), (3, 1), (1, 3)\}$$

où $i \cdot j$ représente le produit des indices i et j . On a $\mathbf{P}(D) = \frac{4}{36} = \frac{1}{9}$. D'autre part,

$$A \cap B \cap D = \{(2, 1)\} \Rightarrow \mathbf{P}(A \cap B \cap D) = \frac{1}{36} = \mathbf{P}(A) \cdot \mathbf{P}(B) \cdot \mathbf{P}(D)$$

Cependant $B \cap D = \{(2, 1), (3, 1), (1, 3)\}$ et $\mathbf{P}(B \cap D) = \frac{1}{12} \neq \mathbf{P}(B) \cdot \mathbf{P}(D)$

Les événements B et D ne sont pas indépendants.

1.3.3 Formules de BAYES

Ces formules ont pour but d'exprimer $\mathbf{P}(A|B)$ en fonction de $\mathbf{P}(B|A)$. Étant donnés deux événements A et B avec $\mathbf{P}(A) \neq 0$ et $\mathbf{P}(B) \neq 0$, on a :

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B|A) = \mathbf{P}(B) \cdot \mathbf{P}(A|B)$$

Ce qui donne la **première formule de BAYES**

Proposition 13 (Formule de BAYES).

Si $\mathbf{P}(A) \neq 0$ et $\mathbf{P}(B) \neq 0$, alors

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(B|A) \cdot \mathbf{P}(A)}{\mathbf{P}(B)} \quad (1.17)$$

Plus généralement, en considérant $\{A_i\}$ un système complet d'événements tous de probabilité non nulle, et en utilisant la formule des probabilités totales on obtient une **deuxième formule de BAYES**:

Proposition 14 (Deuxième formule de BAYES).

Soit $\{A_j\}$ un système complet d'événements tous de probabilité non nulle, pour tout événement B de probabilité non nulle, on a :

$$\mathbf{P}(A_k|B) = \frac{\mathbf{P}(A_k \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_k) \cdot \mathbf{P}(B|A_k)}{\sum_{j=1}^n \mathbf{P}(A_j) \cdot \mathbf{P}(B|A_j)} \quad (1.18)$$

Exemple 1.17. Reprenons l'exemple 1.13. On tire au hasard une pièce, qui se trouve être défectueuse. La probabilité que cette pièce ait été fabriquée par l'usine A vaut $\mathbf{P}(A_1|D)$. Selon la formule de BAYES

$$\mathbf{P}(A_1|D) = \frac{\mathbf{P}(A_1) \cdot \mathbf{P}(D|A_1)}{\mathbf{P}(A_1) \cdot \mathbf{P}(D|A_1) + \mathbf{P}(A_2) \cdot \mathbf{P}(D|A_2) + \mathbf{P}(A_3) \cdot \mathbf{P}(D|A_3)} = \frac{1}{14}$$

Chapitre 2

Variables aléatoires réelles

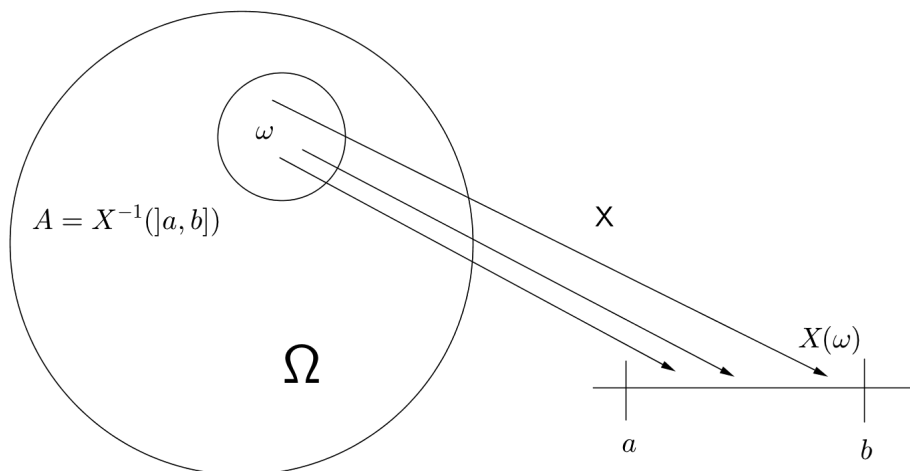
Dans de nombreuses situations, on ne s'intéresse pas directement aux événements aléatoires eux-mêmes, mais à des grandeurs numériques qui leur sont associées : nombre de particules émises par un élément radioactif, dans une direction et durant un intervalle de temps donné, puissance moyenne d'un "bruit" accompagnant la réception d'un signal radio, etc. Ainsi, si l'on cherche à définir la probabilité pour que, dans une population donnée, un individu ait une taille comprise entre 1 m 74 et 1 m 76, on est amené à introduire un espace des éventualités Ω dont les éléments - en l'occurrence, les tailles - varient continûment entre deux bornes précises: la taille du plus petit et celle du plus grand individu. Autrement dit, Ω est un intervalle de \mathbb{R} .

2.1 Loi de probabilité et moments d'une variable aléatoire

2.1.1 Définition et fonction de répartition

Définition 15 (Variable aléatoire réelle).

Étant donné un espace probabilisé $(\Omega, \mathcal{E}, \mathbf{P})$, on appelle **variable aléatoire réelle** une application X définie sur Ω à valeurs dans \mathbb{R} , telle que l'image réciproque par X de tout intervalle $]a, b]$ de \mathbb{R} soit un élément de la tribu \mathcal{E} .



Autrement dit,

$$X : \begin{cases} \Omega & \rightarrow \mathbb{R} \\ \omega & \mapsto X(\omega) \end{cases} \quad \text{et} \quad X^{-1}(]a, b]) \in \mathcal{E}$$

On dit aussi que X est une fonction *mesurable* de $(\Omega, \mathcal{E}, \mathbf{P})$ dans $(\mathbb{R}, \mathcal{B})$, où \mathcal{B} est la tribu des boréliens. Cette dernière étant la tribu engendrée par les intervalles $]a, b]$ (a et b étant des nombres réels quelconques). C'est aussi la tribu engendrée par les demi-droites $] - \infty, a]$ car

$$]a, b] =] - \infty, b] \setminus] - \infty, a] \quad (2.1)$$

Remarque 2.1. L'expression "variable aléatoire" remonte à une époque où sa définition précise n'était pas encore connue. D'après la définition donnée ci-dessus, une "variable aléatoire" n'est ni une variable, ni aléatoire: c'est une fonction définie sur un ensemble.

La définition de la variable aléatoire X permet de "probabiliser" l'espace probabilisable $(\mathbb{R}, \mathcal{B})$, c'est-à-dire de définir sur $(\mathbb{R}, \mathcal{B})$ une probabilité \mathbf{Q} . Pour cela, on pose

$$\forall B \in \mathcal{B}, \quad \mathbf{Q}(B) = \mathbf{P}(X^{-1}(B))$$

Définition 16 (Probabilité image).

On appelle loi de probabilité de X la probabilité \mathbf{Q} , probabilité image de \mathbf{P} par X .

En général la probabilité \mathbf{Q} est notée \mathbf{P}_X . D'après la remarque du paragraphe précédent, on a alors d'après (2.1)

$$\mathbf{P}_X(]a, b]) = \mathbf{P}(X^{-1}]a, b])$$

C'est-à-dire $\mathbf{P}_X(]a, b]) = \mathbf{P}_X(] - \infty, b] \setminus] - \infty, a]) = \mathbf{P}_X(] - \infty, b]) - \mathbf{P}_X(] - \infty, a])$

Si l'on pose maintenant $\forall y \in \mathbb{R}, \quad F_X(y) = \mathbf{P}_X(] - \infty, y])$

alors

$$\mathbf{P}_X(]a, b]) = F_X(b) - F_X(a) \quad (2.2)$$

Ce calcul amène la définition suivante

Définition 17 (Fonction de répartition).

La fonction F_X définie sur \mathbb{R} par

$$\forall y \in \mathbb{R}, \quad F_X(y) = \mathbf{P}_X(] - \infty, y]) \quad (2.3)$$

est la **fonction de répartition** de variable aléatoire X .

Pour simplifier les notations de ce polycopié, s'il n'y a pas d'ambiguïté, on notera \mathbf{P} et F au lieu de \mathbf{P}_X et F_X .

Propriétés de la fonction de répartition

La fonction F est monotone croissante. En tant que fonction monotone, elle admet un ensemble de points de discontinuité fini ou dénombrable. Elle est continue à droite par la propriété d'additivité (1.8). Réciproquement toute fonction monotone croissante continue à droite telle que $F(-\infty) = 0$ et $F(+\infty) = 1$ définit une loi de probabilité unique sur \mathbb{R} .

$F(x)$ est la probabilité pour que la variable aléatoire X prenne une valeur inférieure ou égale à x , ce qui se note :

$$F(x) = \mathbf{P}(X \leq x)$$

on a alors $\mathbf{P}(a < X \leq b) = F(b) - F(a)$. Résumons ces propriétés dans une proposition

Proposition 18 (Propriétés de la fonction de répartition).

- $F(-\infty) = 0$ et $F(+\infty) = 1$,
- F est croissante, c'est-à-dire $y \geq x \Rightarrow F(y) \geq F(x)$,
- F est continue à droite, c'est-à-dire $\lim_{y \rightarrow x^+} F(y) = F(x)$,
- Pour tout $(a, b) \in \mathbb{R}^2$,

$$\mathbf{P}(a < X \leq b) = F(b) - F(a). \quad (2.4)$$

Deux cas particuliers importants**Le cas d'une variable aléatoire discrète****Définition 19** (Variable aléatoire discrète).

Si $X(\Omega)$ est fini ou dénombrable, X définit une **variable aléatoire discrète**.

C'est en particulier le cas si l'espace des éventualités Ω est lui-même fini ou dénombrable. La loi de \mathbf{P} est constituée de masses ponctuelles :

$$X(\Omega) = \{x_1, x_2, \dots, x_n\} \quad \text{et} \quad \mathbf{P}(X = x_i) = p_i$$

Exemple 2.2. Considérons quatre épreuves répétées, identiques et indépendantes, telles qu'au cours de chacune d'elles, un événement A a une probabilité de se réaliser égale à $p = \mathbf{P}(A) = 0,3$ (et donc une probabilité de ne pas se réaliser égale à $q = \mathbf{P}(\bar{A}) = 1 - p = 0,7$). Désignons par X la variable aléatoire ayant pour valeur le nombre de réalisations de l'événement A au cours des quatre expériences aléatoires.¹ On a

$$\mathbf{P}(X = 0) = q^4 = (0,7)^4 = 0,2401, \quad \mathbf{P}(X = 1) = \binom{4}{1} p q^3 = 4 \cdot (0,3) \cdot (0,7)^3 = 0,4116$$

$$\mathbf{P}(X = 2) = \binom{4}{2} p^2 q^2 = 6 \cdot (0,3)^2 \cdot (0,7)^2 = 0,2646$$

$$\mathbf{P}(X = 3) = \binom{4}{3} p^3 q = 4 \cdot (0,3)^3 \cdot (0,7) = 0,0756, \quad \mathbf{P}(X = 4) = \binom{4}{0} p^4 q^0 = (0,3)^4 = 0,0081$$

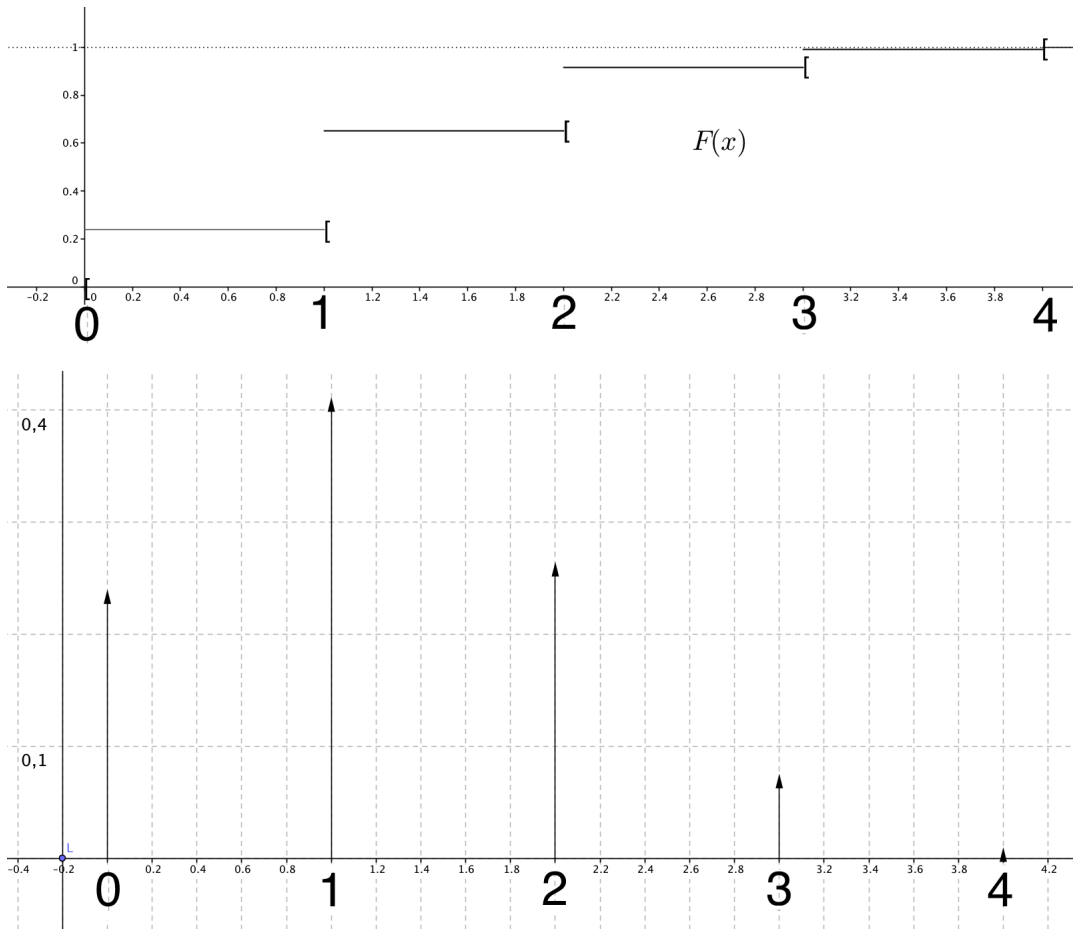
Notons que l'on a bien $\mathbf{P}(X = 0) + \mathbf{P}(X = 1) + \mathbf{P}(X = 2) + \mathbf{P}(X = 3) + \mathbf{P}(X = 4) = 1$.

La fonction de répartition F de la variable aléatoire X est définie par

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 0,2401 & \text{si } 0 \leq x < 1 \\ 0,6517 & \text{si } 1 \leq x < 2 \end{cases} \quad \text{et} \quad F(x) = \begin{cases} 0,9163 & \text{si } 2 \leq x < 3 \\ 0,9919 & \text{si } 3 \leq x < 4 \\ 1 & \text{si } 4 \leq x \end{cases}$$

La fonction de répartition F et la loi de probabilité \mathbf{P} admettent les représentations graphiques suivantes

1. Voir loi binomiale pour la justification des calculs



Le cas d’une variable aléatoire continue

Définition 20 (Densité de probabilité).

Si la fonction de répartition F d’une variable aléatoire X est continue et dérivable presque partout, on dit que X est continue (ou plus exactement absolument continue) et sa dérivée f est la **densité de probabilité**.

Remarque 2.3. L’expression “densité de probabilité” devient évidente si l’on considère les variations de la fonction de répartition d’une variable aléatoire continue entre x et $x + h$. On a alors

$$\Delta(F, x, h) = \frac{F(x + h) - F(x)}{h} = \frac{\mathbf{P}(x < X \leq x + h)}{h}$$

$\Delta(F, x, h)$ est la densité de probabilité de la variable aléatoire X dans l’intervalle $]x, x + h]$. Sa limite, quand h tend vers 0, est le nombre dérivé $F'(x) = f(x)$, c’est la densité “ponctuelle” de probabilité en x .

Proposition 21 (Propriétés de la densité de probabilité).

Si f est une densité de probabilité alors

- f est une fonction positive: $\forall x \geq 0, \quad f(x) \geq 0,$
- $F(x) = \int_{-\infty}^x f(t) dt,$
- $\int_{-\infty}^{+\infty} f(t) dt = 1.$

Ces propriétés découlent directement de celles de la fonction de répartition F . Par exemple pour la première : la densité de probabilité f , en tant que dérivée d'une fonction croissante, ne peut pas prendre de valeurs négatives.

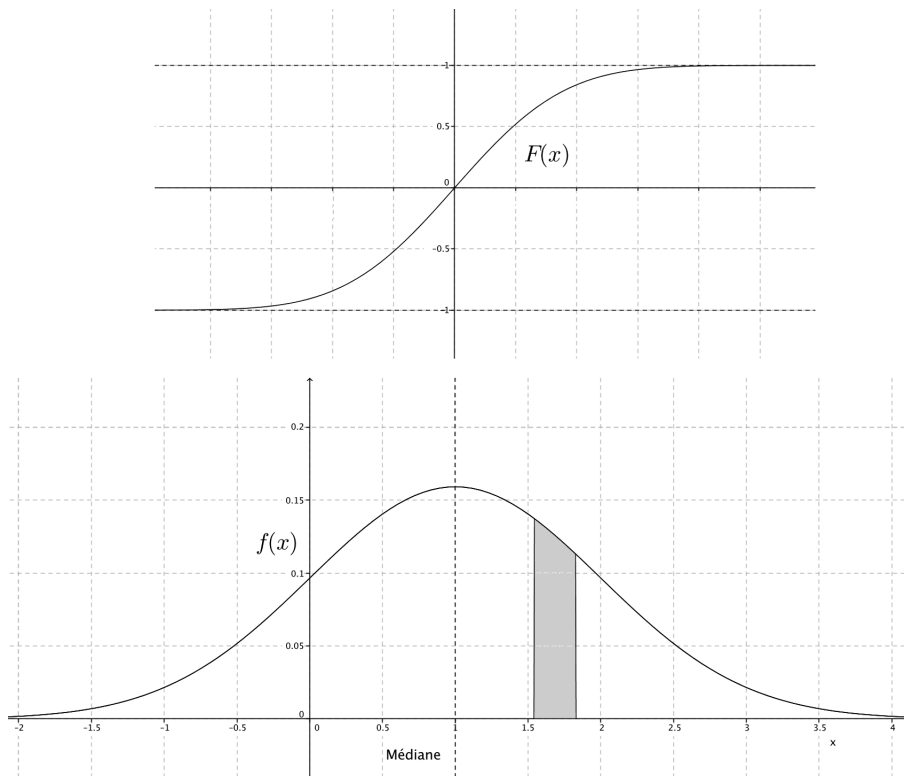
La première et la troisième conditions sont nécessaires et suffisantes pour qu'une fonction f quelconque soit la densité de probabilité d'une variable aléatoire continue. Pour une variable aléatoire continue, de densité de probabilité f , on a donc :

$$\mathbf{P}(a \leq X \leq b) = \int_a^b f(t) dt \tag{2.5}$$

En particulier pour une variable à densité

$$\forall x \in \mathbb{R}, \quad \mathbf{P}(X = x) = 0 \tag{2.6}$$

Les deux figures suivantes représentent les graphes types de la fonction de répartition d'une variable aléatoire continue et de la densité de probabilité correspondante. L'aire en pointillé représente la probabilité pour que X se trouve comprise entre x et $x + \Delta x$. L'aire totale de la surface limitée par la courbe de densité et l'axe des x est toujours égale à 1.



Donnons quelques définitions valables pour une variable aléatoire continue.

Définition 22 (modes, médiane et quartiles).

- L'abscisse x_M pour laquelle la loi où la densité de probabilité présente un maximum local est appelée **mode**. Il peut y avoir plusieurs modes.
- L'abscisse x_m pour laquelle on a $F(x_m) = \frac{1}{2}$, s'appelle la **médiane**.
- On appelle **quantiles** d'ordre p , pour $p \in \mathbb{N}^*$, les abscisses x_k avec $k \in \llbracket 1, p-1 \rrbracket$ pour lesquelles

$$F(x_k) = \frac{k}{p} \tag{2.7}$$

Pour une variable continue: l'aire de la surface limitée par la courbe de densité de probabilité et l'axe des x est divisée en deux parties égales au niveau de la médiane.

Par exemple, les quartiles sont les quantiles d'ordre 4. Le premier quartile est x_1 avec

$$F(x_1) = \frac{1}{4}$$

Le deuxième quartile est x_2 avec

$$F(x_2) = \frac{2}{4} = \frac{1}{2}, \quad \text{c'est la médiane.}$$

Le troisième quartile est x_3 avec

$$F(x_3) = \frac{3}{4}.$$

Dans le cas d'une variable discrète, ces définitions sont plus difficiles à mettre en oeuvre pour des questions d'unicité. On peut prendre comme définition de la médiane par exemple

$$x_m = \text{Inf}\{t : F(t) \geq 0,5\} \tag{2.8}$$

ou encore la médiane est tout nombre x tel que $\mathbf{P}(X \leq x) \geq 0,5$ et $\mathbf{P}(X \geq x) \geq 0,5$. Avec cette définition, la médiane n'est pas unique.

Le cas général

Plus généralement, en utilisant la théorie des distributions, on peut considérer des variables aléatoires ni discrètes ni continues. Voici quelques notions sur ce sujet qui peuvent être réservées à une seconde lecture. Lorsque la variable aléatoire X n'est pas continue, F possède au moins une discontinuité en un point d'abscisse x_0 . Ses propriétés font qu'en ce point F effectue un saut fini σ défini par $\sigma = F(x_0^+) - F(x_0^-)$.

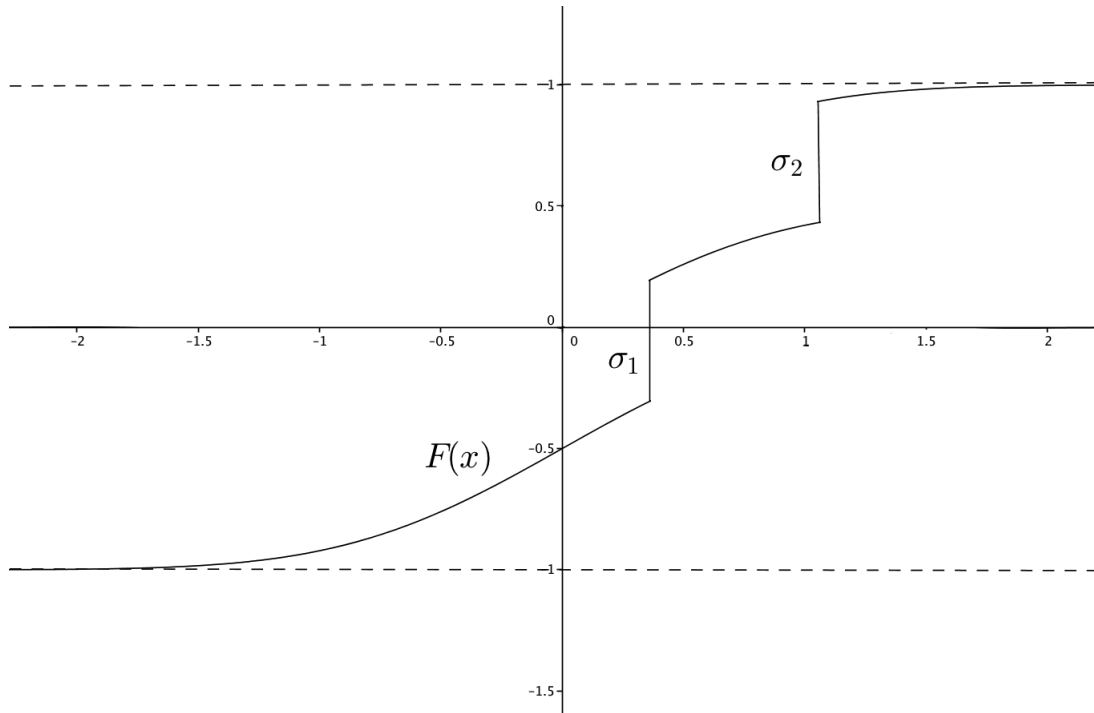
On peut encore définir une densité de probabilité à condition de dériver la fonction de répartition F au sens des distributions ce qui fait apparaître la distribution de DIRAC δ_{x_0} . La densité de probabilité est alors une distribution singulière. On peut généraliser au cas où il existe un nombre dénombrable de points de discontinuités x_k où F effectue des sauts σ_k (finis positifs). On écrit :

$$F(x) = \check{F}(x) + \sum_k \sigma_k H(x - x_k) \quad \text{avec} \quad H(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

où H est la fonction de HEAVISIDE. On en déduit par la formule des sauts que

$$f = \check{f} + \sum_k \sigma_k \delta_{x_k}$$

\check{F} et \check{f} représentent respectivement les "parties continues" de F et de f .



D'une manière générale, toute distribution S de la forme

$$S = f + \sum_k c_k \delta_{x_k} \quad \text{avec} \quad c_k \geq 0 \quad \text{et} \quad f \geq 0$$

vérifiant $\langle S, 1 \rangle = 1$ définit une densité de probabilité. Ceci implique la sommabilité de f et la convergence de la série de terme général c_k , car

$$\langle S, 1 \rangle = \|f\|_{L^1(\mathbb{R})} + \sum_k c_k = 1$$

Ainsi, l'emploi de la théorie des distributions permet de traiter le cas discret et le cas continu dans un seul et même formalisme.

2.1.2 Loi d'une fonction d'une variable aléatoire $Y = \varphi(X)$

On suppose X continue avec une densité f et une fonction de répartition F . Soit φ une fonction dérivable. On cherche à expliciter g et G , densité et fonction de répartition de $Y = \varphi(X)$.

Si φ est bijective. φ est monotone, par exemple si φ est croissante, on a :

$$Y \leq y \quad \text{si et seulement si} \quad \varphi(X) \leq y \quad \text{si et seulement si} \quad X \leq \varphi^{-1}(y)$$

D'où $G(y) = F(\varphi^{-1}(y))$, ou encore $G(\varphi(x)) = F(x)$. En dérivant, on obtient :

$$g(\varphi(x)) \cdot \varphi'(x) = f(x)$$

soit
$$g(y) = \frac{f(x)}{\varphi'(x)} = \frac{f(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}$$

Si φ est décroissante, on a de même

$$g(y) = -\frac{f(x)}{\varphi'(x)} = -\frac{f(\varphi^{-1}(y))}{\varphi'(\varphi^{-1}(y))}$$

Puisque φ est décroissante, $\varphi' < 0$, et on a la formule générale pour une application bijective φ quelconque:

$$g(y) = \frac{f(\varphi^{-1}(y))}{|\varphi'(\varphi^{-1}(y))|} \tag{2.9}$$

Cas général. Le principe consiste toujours à identifier la fonction de répartition G , en recherchant l'antécédent pour X de l'événement " $Y \leq y$ ".

Exemple 2.4. Considérons le cas $Y = X^2$

$$Y \leq y \quad \text{si et seulement si} \quad |X| \leq \sqrt{y}$$

On obtient

$$G(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ F(\sqrt{y}) - F(-\sqrt{y}) & \text{si } y > 0 \end{cases}$$

et enfin

$$g(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ \frac{1}{2\sqrt{y}}(f(\sqrt{y}) - f(-\sqrt{y})) & \text{si } y > 0 \end{cases}$$

Si de plus, on suppose que X suit une loi normale centrée réduite (i.e $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$) alors pour $y > 0$, $g(y) = \frac{1}{\sqrt{2\pi y}}e^{-y/2}$. Cette loi est très importante en statistique, c'est la *loi du χ^2* .

2.1.3 Indépendance de deux variables aléatoires

Donnons à présent une définition de l'indépendance de deux variables aléatoires.

Définition 23 (Indépendance de deux variables aléatoires).

Soient X et Y deux variables aléatoires réelles définies sur un même espace probabilisé. X et Y sont **indépendantes** si pour tout couple d'intervalles I et J de \mathbb{R} , on a

$$\mathbf{P}(X \in I \text{ et } Y \in J) = \mathbf{P}(X \in I) \cdot \mathbf{P}(Y \in J) \quad (2.10)$$

Exemple 2.5. Considérons un jet de deux dés. Soit X et Y les variables aléatoires représentant respectivement la valeur du premier dé et leur somme. On constate que

$$0 = \mathbf{P}(Y = 12 \text{ et } X = 1) \neq \mathbf{P}(Y = 12) \cdot \mathbf{P}(X = 1) = \frac{1}{36}$$

Les deux variables ne sont pas indépendantes.

2.1.4 Espérance, variance, moments d'une variable aléatoire

Espérance

Dans de nombreux cas, il n'est pas nécessaire de connaître une propriété de la variable aléatoire aussi précise que sa fonction de répartition. Certains paramètres numériques caractérisent parfois cette variable aléatoire. L'étude de quelques-uns de ces paramètres fait l'objet du présent paragraphe.

Soit X une variable aléatoire discrète pouvant prendre un nombre fini de valeurs x_1, x_2, \dots, x_n , avec les probabilités respectives p_1, p_2, \dots, p_n . On définit une moyenne arithmétique des différentes valeurs de X pondérées par leurs probabilités p_k . Cette moyenne pondérée est appelée espérance mathématique (ou moyenne stochastique) et elle est notée $\mathbf{E}(X)$. On a :

$$\mathbf{E}(X) = \frac{p_1 x_1 + \dots + p_n x_n}{p_1 + \dots + p_n}$$

Or $p_1 + \dots + p_n = 1$. Il vient

$$\mathbf{E}(X) = p_1 x_1 + \dots + p_n x_n = \sum_{i=1}^n p_i x_i$$

La formule s'étend aux variables aléatoires non finies.

Définition 24 (Espérance).

X étant une variable aléatoire réelle, on appelle **espérance mathématique** de X le nombre $\mathbf{E}(X)$, si il existe, défini par

- si X est une variable aléatoire finie

$$\mathbf{E}(X) = \sum_{k=1}^n x_k \mathbf{P}(X = x_k) \quad (2.11)$$

- si X est une variable aléatoire discrète dénombrable (sous réserve de convergence)

$$\mathbf{E}(X) = \sum_{k=1}^{+\infty} x_k \mathbf{P}(X = x_k) \quad (2.12)$$

- si X est une variable aléatoire continue, f étant la densité de probabilité, on a (sous réserve de convergence)

$$\mathbf{E}(X) = \int_{\mathbb{R}} t \cdot f(t) dt \quad (2.13)$$

Remarque 2.6. Pour une variable aléatoire discrète, on peut aussi revenir à l'ensemble Ω pour faire la sommation

$$\mathbf{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbf{P}(\{\omega\}) \quad (2.14)$$

Exemple 2.7. • Au cours d'un jeu de dés, un joueur gagne 20 euros s'il tire le 4, gagne 40 euros s'il tire l'as, perd 30 euros s'il tire le 6 et ne gagne ni ne perd dans les autres cas. On a

$$\mathbf{E}(X) = \frac{1}{6} \times 40 + \frac{1}{6} \times 0 + \frac{1}{6} \times 20 + \frac{1}{6} \times 30 = 5$$

Ainsi, le joueur peut espérer gagner 5 euros en moyenne à chaque lancer de dé.

- La variable aléatoire X ayant pour densité la fonction f définie par

$$f(t) = \frac{1}{\pi(1+t^2)}$$

ne possède pas de moyenne car la fonction $t \mapsto t \cdot f(t)$ n'est pas intégrable sur \mathbb{R} .

- La variable aléatoire X ayant pour densité la fonction f définie par

$$f(t) = \begin{cases} t/2 & \text{si } 0 \leq t < 2 \\ 0 & \text{sinon} \end{cases}$$

a pour espérance mathématique

$$\mathbf{E}(X) = \int_0^2 \frac{t}{2} dt = \frac{4}{3}$$

Propriétés élémentaires de l'espérance mathématique

Proposition 25.

Soient X et Y deux variables aléatoires réelles d'espérances finies et $\lambda \in \mathbb{R}$, alors

- Linéarité de l'espérance

$$\mathbf{E}(X + \lambda Y) = \mathbf{E}(X) + \lambda \mathbf{E}(Y) \quad (2.15)$$

- Positivité : $|\mathbf{E}(X)| \leq \mathbf{E}(|X|)$
- Si X et Y sont indépendantes

$$\mathbf{E}(X \times Y) = \mathbf{E}(X) \times \mathbf{E}(Y) \quad (2.16)$$

- Pour toute fonction $h : \mathbb{R} \mapsto \mathbb{R}$ avec $\mathbf{E}(|h(X)|) < +\infty$:

$$\mathbf{E}(f(X)) = \begin{cases} \int_{\mathbb{R}} h(t) f_X(t) dt & \text{en continu} \\ \sum_{\omega \in \Omega} h(X(\omega)) \cdot \mathbf{P}_X(\{\omega\}) & \text{en discret} \end{cases} \quad (2.17)$$

Où f_X désigne la densité de probabilité de X . C'est la **formule de transfert**.

- Pour tout $A \in \mathcal{E}$

$$\mathbf{P}(X \in A) = \mathbf{P}_X(A) = \mathbf{E}(\mathbf{1}_A(X)) \quad (2.18)$$

Preuve. Donnons la preuve de la linéarité pour une variable aléatoire discrète

$$\begin{aligned} \mathbf{E}(X + \lambda Y) &= \sum_{\omega \in \Omega} (X + \lambda Y)(\omega) \mathbf{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} (X(\omega) + \lambda Y(\omega)) \mathbf{P}(\{\omega\}) \\ &= \sum_{\omega \in \Omega} X(\omega) \mathbf{P}(\{\omega\}) + \lambda \sum_{\omega \in \Omega} Y(\omega) \mathbf{P}(\{\omega\}) \\ \mathbf{E}(X + \lambda Y) &= \mathbf{E}(X) + \lambda \mathbf{E}(Y) \end{aligned}$$

La propriété de positivité est immédiate à partir de la définition. On verra la preuve de la propriété (2.16) au chapitre 3, quand on parlera de loi conjointe d'un couple de variables aléatoires (indépendantes ou non) et de la covariance. Quant à la formule de transfert, elle se montre avec la définition de l'espérance. La dernière assertion découle de la formule de transfert en considérant la fonction indicatrice $f = \mathbf{1}_A$. ■

La formule de transfert permet le constat suivant:

Il n'est pas nécessaire de déterminer la loi de probabilité de $Y = \varphi(X)$ pour calculer son espérance: la connaissance de φ et de la loi de probabilité de X peut suffire.

Remarque 2.8. Étant données une variable aléatoire X et son espérance mathématique $\mathbf{E}(X)$, on appelle **variable aléatoire centrée**, la variable aléatoire

$$Y = X - \mathbf{E}(X) \quad (2.19)$$

En effet, il est facile de vérifier que l'espérance mathématique d'une variable aléatoire centrée est nulle.

Moments, variance, écart-type

Pour tout entier naturel s , on appelle moment d'ordre s d'une variable aléatoire X et l'on note $m_s(X)$, l'espérance mathématique de X à la puissance s .

$$m_s(X) = \mathbf{E}(X^s) \quad (2.20)$$

En particulier pour $s = 1$, on retrouve l'espérance $m_1(X) = \mathbf{E}(X)$. Comme on l'a remarqué, on n'a pas besoin de connaître la loi de probabilité de X^s pour calculer son espérance, la loi de X seule suffit.

- dans le cas d'une variable aléatoire discrète, on a

$$m_s(X) = \sum_k x_k^s \mathbf{P}(X = x_k) \quad (2.21)$$

- dans le cas d'une variable aléatoire continue, on a

$$m_s(X) = \int_{\mathbb{R}} t^s \cdot f(t) dt \quad (2.22)$$

On appelle moment centré d'ordre s d'une variable aléatoire X , l'espérance mathématique de la $s^{\text{ième}}$ puissance de la variable centrée associée à X . On écrit :

$$\mu_s(X) = \mathbf{E}((X - \mathbf{E}(X))^s) \quad (2.23)$$

Définition 26.

On appelle variance d'une variable aléatoire, son moment centré d'ordre 2. On la note $\mathbf{Var}(X)$.

$$\mathbf{Var}(X) = \mu_2(X) = \mathbf{E}((X - \mathbf{E}(X))^2) \quad (2.24)$$

La variance est un nombre positif. Sa racine carrée définit l'**écart-type** dont l'expression est

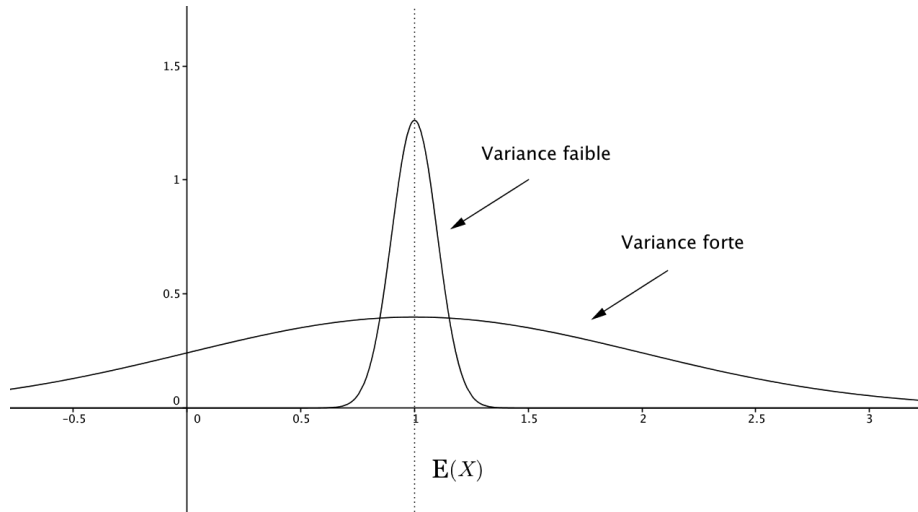
$$\sigma_X = \sqrt{\mathbf{Var}(X)} \quad (2.25)$$

Exemple 2.9. La variance et l'écart-type de la variable aléatoire du troisième exemple de 2.7 sont données par

$$\mathbf{Var}(X) = \mathbf{E}\left(X - \frac{4}{3}\right)^2 = \int_{\mathbb{R}} \left(t - \frac{4}{3}\right)^2 f(t) dt = \int_0^2 \left(t - \frac{4}{3}\right)^2 dt = \frac{2}{9}$$

L'écart type vaut alors $\frac{\sqrt{2}}{3}$.

La variance (comme l'écart-type) est une mesure de la dispersion (nom parfois donné à la variance) des valeurs de la variable aléatoire autour de son espérance mathématique $\mathbf{E}(X) = \mu$. Si ces valeurs tendent à se concentrer au voisinage de μ , la variance est faible. Au contraire, si ces valeurs ont une tendance à la dispersion, la variance est grande. Ces situations sont représentées par la figure suivante, dans le cas de deux densités de probabilité ayant la même espérance mathématique



Propriétés élémentaires de la variance

Proposition 27.

- La variance est toujours positive

$$\mathbf{Var}(X) \geq 0 \tag{2.26}$$

- Pour tout $a, b \in \mathbb{R}$, on a

$$\mathbf{Var}(aX + b) = a^2 \mathbf{Var}(X) \tag{2.27}$$

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

- Si X_1 et X_2 sont deux variables aléatoires *indépendantes*

$$\mathbf{Var}(X_1 + X_2) = \mathbf{Var}(X_1) + \mathbf{Var}(X_2) \tag{2.28}$$

Preuve. Les deux premières propriétés résultent directement de la définition de la variance. Pour la troisième, on a

$$\mathbf{Var}(X) = \mathbf{E} \left((X - \mathbf{E}(X))^2 \right) = \mathbf{E} \left(X^2 - 2\mathbf{E}(X) \cdot X + \mathbf{E}(X)^2 \right) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$$

Cette formule est aussi la formulation en termes de probabilité du théorème de cinétique de KOENIG-HUYGENS. Elle s'écrit avec les notations des moments : $\mu_2 = m_2 - m_1$. Pour la quatrième, on peut supposer que $\mathbf{E}(X_1) = \mathbf{E}(X_2) = 0$ et

$$\mathbf{Var}(X_1 + X_2) = \mathbf{E} \left((X_1 + X_2)^2 \right) - \underbrace{\left(\mathbf{E}(X_1 + X_2) \right)^2}_0 = \mathbf{E}(X_1^2) + \mathbf{E}(X_2^2) + \mathbf{E}(X_1 \times X_2)$$

Or par indépendance $\mathbf{E}(X_1 \times X_2) = \mathbf{E}(X_1) \times \mathbf{E}(X_2) = 0$. D'où le résultat. ■

Inégalités de MARKOV et BIENAYMÉ-TCHEBYSHEV

L'espérance $\mathbf{E}(X)$ et l'écart-type σ sont liés par l'inégalité de BIENAYMÉ- TCHEBYSHEV

Proposition 28 (Inégalité de MARKOV et de BIENAYMÉ-TCHEBYSHEV).

- Soit Z une v.a positive, alors

$$\forall \varepsilon > 0, \quad \mathbf{P}(Z \geq \varepsilon) \leq \frac{\mathbf{E}(Z)}{\varepsilon} \quad (2.29)$$

- Soit X une v.a.r admettant un moment d'ordre 2 ($\mathbf{E}(X^2) < +\infty$) alors

$$\forall \varepsilon > 0, \quad \mathbf{P}(|X - \mathbf{E}(X)| > \varepsilon) \leq \frac{\sigma_X^2}{\varepsilon^2} \quad (2.30)$$

Preuve. Notons que pour toute variable aléatoire positive $Z \geq \varepsilon \mathbf{1}_{\{Z \geq \varepsilon\}}$. On en déduit la première inégalité en utilisant la relation (2.18). La seconde inégalité en découle directement en considérant $Z = (X - \mathbf{E}(X))^2$. ■

Exemple 2.10. Si $\mathbf{E}(X) = 10$ et $\sigma = 0.1$ sont les valeurs exactes de la moyenne et de l'écart-type d'une variable aléatoire X , sans aucune indication particulière sur la loi de X , on sait que X prendra des valeurs entre 9.7 et 10.3 avec une probabilité supérieure à 88%. Si on connaît la loi de X , on a des estimations beaucoup plus précises.

Remarque 2.11. Ces inégalités ont peu d'applications pratiques, car la majoration qu'elles fournissent est la plupart du temps excessive, mais elles sont valables quelle que soit la loi de X pourvu que l'on puisse définir une espérance et une variance... L'inégalité de BIENAYMÉ-TCHEBYSHEV permet toutefois de démontrer la loi faible des grands nombres (voir théorème 42).

2.2 Lois de probabilité d'usage courant

2.2.1 Lois discrètes

Loi discrète uniforme

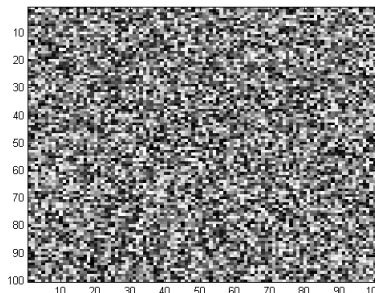
C'est la loi d'une variable aléatoire X équiprobable sur l'ensemble $\Omega = \{1, 2, \dots, n\}$

$$\begin{cases} X(\Omega) &= \llbracket 1, n \rrbracket \\ \mathbf{P}(X = k) &= \frac{1}{n} \end{cases}$$

Un calcul direct donne

$$\mathbf{E}(X) = \frac{n+1}{2} \quad \text{et} \quad \mathbf{Var}(X) = \frac{n^2-1}{12} \quad (2.31)$$

Exemple 2.12. Soit X suivant une loi uniforme sur $\mathcal{U}_{\llbracket 0, 255 \rrbracket}$. En chaque point $(i, j) \in \llbracket 1, 100 \rrbracket^2$, on représente une réalisation $X(\omega)$ en niveau de gris



Loi de BERNOULLI

C'est la loi d'une variable aléatoire X ne pouvant prendre que les deux valeurs 1 ou 0 avec les probabilités p et $1 - p$.

$$\begin{cases} X(\Omega) = \{0, 1\} \\ P(X = 1) = p \\ P(X = 0) = 1 - p = q \end{cases}$$

De même, un calcul direct donne

$$\boxed{\mathbf{E}(X) = p \quad \text{et} \quad \mathbf{Var}(X) = pq} \quad (2.32)$$

Loi binomiale

Soient n un entier non nul et p un réel tel que $0 \leq p \leq 1$. On dit que X suit la loi binomiale de paramètres n et p notée $\mathcal{B}(n, p)$ si

$$\begin{cases} X(\Omega) = \llbracket 0, n \rrbracket \\ P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \end{cases}$$

On modélise ainsi le cas où une même expérience aléatoire est répétée plusieurs fois. Chaque expérience peut réaliser ou ne pas réaliser un événement. Les expériences sont **indépendantes**, c'est à dire que le résultat obtenu au cours de chacune d'elles, A ou \bar{A} , ne dépend pas des résultats des autres.

On envisage donc n expériences aléatoires identiques et indépendantes appelées épreuves répétées et pouvant chacune soit réaliser l'événement avec la probabilité p , soit ne pas le réaliser avec la probabilité $q = 1 - p$. La probabilité pour que l'événement A se réalise k fois au cours de n expériences aléatoires indépendantes, est donnée par

$$p_{n,k} = \binom{n}{k} p^k q^{n-k} \quad (2.33)$$

En effet, désignons par B_k l'événement correspondant à la réalisation de A exactement k fois au cours de n épreuves répétées et par A_i (resp. \bar{A}_i) la réalisation (resp. la non réalisation) de l'événement au cours de la i -ème épreuve. Il vient

$$B_k = \bigcup_{i_1 < i_2 < \dots < i_k} \left(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k} \cap \left(\bigcap_{j \neq i_\ell} \bar{A}_j \right) \right)$$

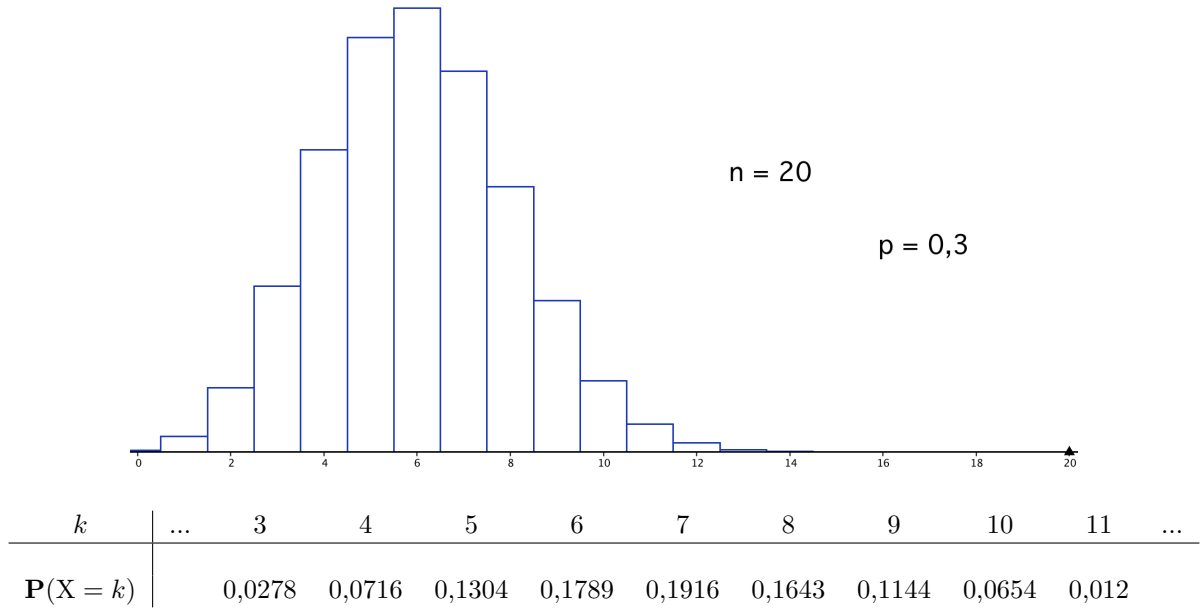
Dans chaque terme, la réalisation de A doit figurer k fois et sa non-réalisation $n - k$ fois, le nombre de combinaison de ce genre est $\binom{n}{k}$. C'est le nombre de manières de choisir les k épreuves parmi n expériences répétées. Or par indépendance

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \cdots \mathbf{P}(A_{i_k}) = p^k \quad \text{et} \quad \mathbf{P}\left(\bigcap_{j \neq i_\ell} \bar{A}_j\right) = q^{n-k}$$

Finalement, on retrouve bien l'expression (2.33).

Remarque 2.13. La loi de BERNOULLI est la loi binomiale $\mathcal{B}(1, p)$.

Exemple 2.14. Donnons l'histogramme de la loi $\mathcal{B}(20; 0, 3)$, c'est-à-dire, pour chaque valeur $k \in \llbracket 0, 20 \rrbracket$, la valeur $\mathbf{P}(X = k)$



Remarque 2.15. La loi est appelée loi binomiale dans la mesure où les probabilités correspondent aux termes du développement du binôme de NEWTON.

Calculons l'espérance

$$\begin{aligned}
 \mathbf{E}(X) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} \\
 &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} \quad \text{car } n \binom{n-1}{k-1} = k \binom{n}{k} \\
 \mathbf{E}(X) &= np(p+q)^{n-1} = np
 \end{aligned}$$

Pour le calcul de la variance $\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$

$$\begin{aligned}
 \mathbf{E}(X^2) &= \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} \quad \text{car } k^2 = k(k-1) + k \\
 &= np \sum_{k=1}^n (k-1) \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} \\
 &\quad + np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{n-1-(k-1)} \\
 \mathbf{E}(X^2) &= np(n-1)p + np
 \end{aligned}$$

Finalement, on trouve

$\mathbf{E}(X) = np \quad \text{et} \quad \mathbf{Var}(X) = npq$

(2.34)

Loi de POISSON

Soit λ un réel positif. On dit que X suit la loi de POISSON de paramètre λ notée $\mathcal{P}(\lambda)$ si

$$\begin{cases} X(\Omega) = \mathbb{N} \\ \mathbf{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \end{cases}$$

Rapelons le développement en série de l'exponentielle

$$\forall x \in \mathbb{R}, \quad e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!} \quad (2.35)$$

De sorte que

$$\sum_{n=0}^{+\infty} \mathbf{P}(X = n) = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1$$

L'espérance mathématique de X est donnée par

$$\mathbf{E}(X) = \sum_{n=0}^{+\infty} n \times e^{-\lambda} \frac{\lambda^n}{n!} = \lambda \sum_{n=1}^{+\infty} e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Pour calculer la variance, commençons par calculer $\mathbf{E}(X(X-1))$

$$\mathbf{E}(X(X-1)) = \sum_{n=0}^{+\infty} n(n-1) \times e^{-\lambda} \frac{\lambda^n}{n!} = \lambda^2 \sum_{n=2}^{+\infty} e^{-\lambda} \frac{\lambda^{n-2}}{(n-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2$$

Or

$$\mathbf{E}(X(X-1)) = \mathbf{E}(X^2) - \mathbf{E}(X) \quad \Rightarrow \quad \mathbf{E}(X^2) = \lambda^2 + \lambda$$

D'où

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

En résumé

$$\boxed{\mathbf{E}(X) = \lambda \quad \text{et} \quad \mathbf{Var}(X) = \lambda} \quad (2.36)$$

La loi de POISSON modélise des comptages qui suivent un processus de POISSON. Par exemple, le nombre d'appels à un central téléphonique pendant une période donnée, le nombre de voitures qui passent à un carrefour en un temps donné... On parle "**d'événements rares**". De plus, l'utilisation la plus fréquente de la loi de POISSON est comme approximation de la loi binomiale (voir théorème 46).

Exemple 2.16. Un central téléphonique reçoit en moyenne 100 appels par heure. En supposant que le nombre d'appels durant un intervalle de temps quelconque suit une loi de POISSON,

1. quelle est la probabilité que le central reçoive trois appels en deux minutes?
2. quelle est la probabilité pour qu'en deux minutes, il reçoive au moins un appel?

En deux minutes, on peut s'attendre à ce que le nombre d'appel soit

$$\lambda = \mathbf{E}(X) = \frac{2 \times 100}{60} = \frac{10}{3}$$

Ainsi, la probabilité d'obtenir trois appels en deux minutes est

$$\mathbf{P}(X = 3) = \frac{\left(\frac{10}{3}\right)^3}{3!} \exp\left(-\frac{10}{3}\right) \simeq 0,220$$

et la probabilité d'obtenir au moins un appel en deux minutes est

$$\mathbf{P}(X \geq 1) = 1 - \exp\left(-\frac{10}{3}\right) \simeq 0,964$$

Lois géométrique et hypergéométrique

Pour $p \in [0, 1]$, $A \in \mathbb{N}$ et $m \leq A$, on définit la loi géométrique, notée $\mathcal{G}(p)$, par

$$\begin{cases} X(\Omega) = \mathbb{N}^* \\ P(X = k) = (1 - p)^{k-1}p = q^{k-1}p \end{cases}$$

On montre que

$$\boxed{\mathbf{E}(X) = \frac{1}{p} \quad \text{et} \quad \mathbf{Var}(X) = \frac{q}{p^2}} \quad (2.37)$$

Par exemple, la probabilité $q^{k-1}p$ correspond à la probabilité d'obtenir dans une succession de k épreuves de BERNOULLI indépendantes, $k - 1$ échecs suivis d'un succès. De plus, la loi géométrique est le premier modèle discret de la mort d'une particule radioactive. En effet, la durée de vie de la particule radioactive, notée T , suit la loi de probabilité pour $k \in \mathbb{N}$

$$\mathbf{P}(T = k) = q^k p$$

On définit aussi la loi hypergéométrique, notée $\mathcal{H}(n, p, A)$ par

$$\begin{cases} X(\Omega) \subset \llbracket 0, n \rrbracket \\ P(X = k) = \frac{\binom{pA}{k} \binom{qA}{n-k}}{\binom{A}{n}} \quad \text{avec} \quad q = 1 - p \end{cases}$$

On prouve que

$$\boxed{\mathbf{E}(X) = np \quad \text{et} \quad \mathbf{Var}(X) = npq \frac{(A - n)}{(A - 1)}} \quad (2.38)$$

Exemple 2.17. On tire n boules (sans remise) dans une urne contenant pA boules rouges et qA boules bleues, soit un nombre total de boules de $A = pA + qA$. Alors la variable aléatoire donnant le nombre de boules rouges suit une loi hypergéométrique $\mathcal{H}(n, p, A)$.

2.2.2 Lois continues

Loi uniforme

On dit que X suit la loi uniforme sur $[a, b]$, notée $\mathbb{U}_{[a,b]}$, si

$$\begin{cases} X(\Omega) = [a, b] \\ f(x) = \frac{1}{b - a} \mathbf{1}_{[a,b]}(x) \end{cases}$$

On a

$$\boxed{\mathbf{E}(X) = \frac{a + b}{2} \quad \text{et} \quad \mathbf{Var}(X) = \frac{(b - a)^2}{12}} \quad (2.39)$$

En effet, il suffit de calculer les intégrales

$$\mathbf{E}(X) = \frac{1}{b - a} \int_a^b t \, dt \quad \text{et} \quad \mathbf{Var}(X) = \frac{1}{b - a} \int_a^b \left(t - \frac{a + b}{2} \right)^2 \, dt$$

Remarque 2.18. Le facteur $\frac{1}{b - a}$ est imposé par le fait qu'une fonction de densité est toujours d'intégrale 1 ($\int_{\mathbb{R}} f = 1$).

Loi exponentielle

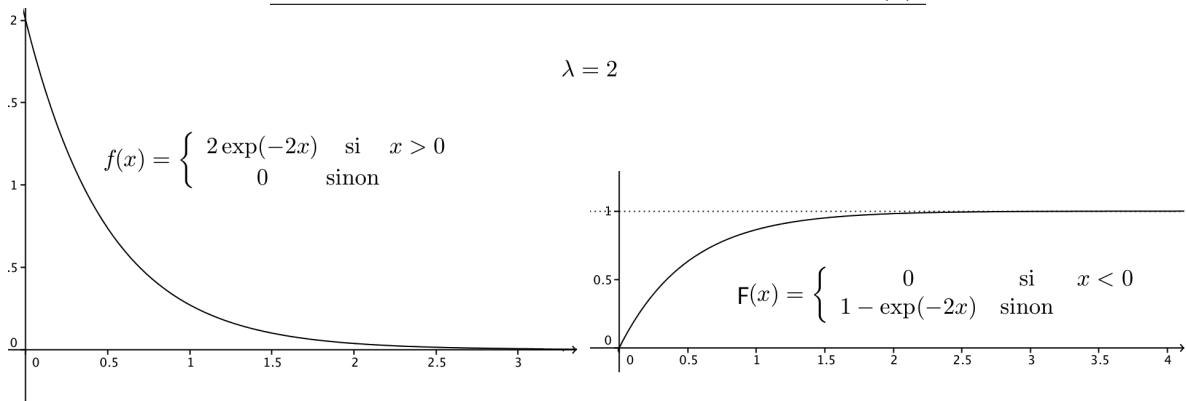
On dit que X suit la loi exponentielle de paramètre λ , notée $\mathcal{E}(\lambda)$, si

$$\begin{cases} X(\Omega) &= \mathbb{R}^+ \\ f_\lambda(x) &= \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}^+}(x) \end{cases}$$

On a par intégration par parties

$$\boxed{\mathbf{E}(X) = \frac{1}{\lambda} \quad \text{et} \quad \mathbf{Var}(X) = \frac{1}{\lambda^2}} \tag{2.40}$$

La densité et la fonction de répartition de $\mathcal{P}(2)$.



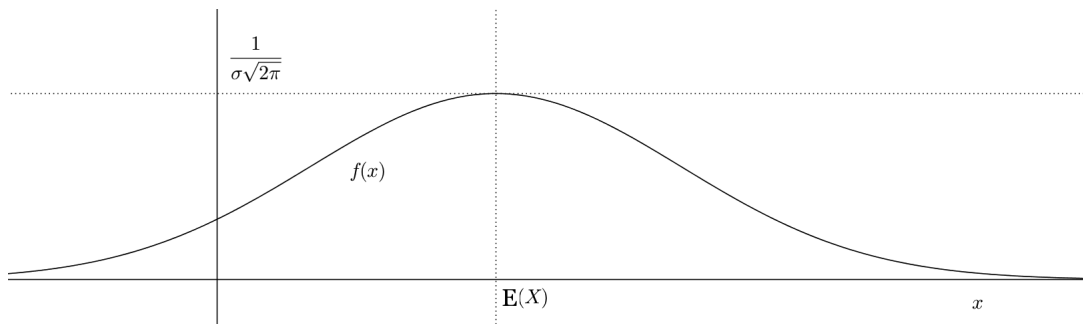
La loi exponentielle intervient dans les processus continus sans mémoire comme la désintégration d'un noyau atomique, l'émission d'un électron... Elle généralise au cas continu la loi géométrique.

Loi normale (de LAPLACE-GAUSS)

On dit que X suit la loi normale de paramètre (μ, σ) , notée $\mathcal{N}(\mu, \sigma)$, si

$$\begin{cases} X(\Omega) &= \mathbb{R} \\ f_{\mu, \sigma}(t) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) \end{cases}$$

Ci-dessous, l'allure caractéristique en cloche de la densité de la loi normale



Remarque 2.19. Notons que f est bien une densité puisque

$$\int_{\mathbb{R}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt = \sigma\sqrt{2\pi}$$

En effet, par le changement de variable $u = \frac{(t - \mu)}{\sqrt{2}\sigma}$, il suffit de retrouver la valeur de l'intégrale de GAUSS

$$I = \int_{\mathbb{R}} \exp(-u^2) du = \sqrt{\pi} \quad (2.41)$$

Pour cela, effectuons un changement de variable en coordonnées polaires

$$\begin{aligned} I^2 &= \int_{\mathbb{R}} \exp(-x^2) dx \times \int_{\mathbb{R}} \exp(-y^2) dy = \int_{\mathbb{R}^2} \exp(-(x+y)^2) dx dy \\ &= \int_{\theta=0}^{2\pi} \int_{r=0}^{+\infty} \exp(-r^2) r dr d\theta = 2\pi \int_0^{+\infty} \exp(-r^2) r dr \\ &= -\pi [\exp(-r^2)]_0^{+\infty} = \pi \end{aligned}$$

Calculons l'espérance et la variance par le changement de variable $u = \frac{(t - \mu)}{\sqrt{2}\sigma}$

$$\begin{aligned} \mathbf{E}(X) &= \int_{\mathbb{R}} t \cdot f_{\mu,\sigma}(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} t \cdot \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt \\ &= \int_{\mathbb{R}} \frac{\sqrt{2}\sigma u + \mu}{\sqrt{\pi}} e^{-u^2} du = \underbrace{\frac{\sqrt{2}\sigma}{\sqrt{\pi}} \int_{\mathbb{R}} u e^{-u^2} du}_0 \text{ par imparité} + \underbrace{\frac{\mu}{\sqrt{\pi}} \int_{\mathbb{R}} e^{-u^2} du}_{\mu \text{ d'après (2.41)}} = \mu \end{aligned}$$

$$\begin{aligned} \text{et } \mathbf{Var}(X) &= \int_{\mathbb{R}} (t - \mu)^2 \cdot f_{\mu,\sigma}(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} (t - \mu)^2 \cdot \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{\mathbb{R}} u^2 e^{-u^2} du = \frac{2\sigma^2}{\sqrt{\pi}} \left(\frac{1}{2} [-u e^{-u^2}]_{\mathbb{R}} + \frac{1}{2} \int_{\mathbb{R}} e^{-u^2} du \right) = \sigma^2 \end{aligned}$$

En résumé

$$\boxed{\mathbf{E}(X) = \mu \quad \text{et} \quad \mathbf{Var}(X) = \sigma^2} \quad (2.42)$$

En utilisant les résultats du paragraphe 2.1.2 avec $\varphi(x) = \sigma x + \mu$ et $\sigma \neq 0$, on obtient

Proposition 29 (Loi normale et transformation affine).

Si X suit une loi normale $\mathcal{N}(0; 1)$, alors pour $\sigma \neq 0$, la variable aléatoire $Y = \sigma X + \mu$ suit une la loi normale $\mathcal{N}(\mu; \sigma)$.

Plus généralement, toute transformation affine d'une loi normale est encore une loi normale.

Cette propriété justifie l'intérêt particulier donné à la **la loi normale centrée réduite** $\mathcal{N}(0; 1)$. Son espérance est nulle ("elle est centrée") et son écart type est de 1 ("elle est réduite"), sa densité est donnée par

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Elle admet des moments à tout ordre. Par parité, ils sont nuls pour les ordres impairs et

$$m_{2k} = \frac{2^k}{\sqrt{\pi}} \Gamma\left(k + \frac{1}{2}\right) = \frac{(2k)!}{2^k k!}$$

Remarque 2.20. Afin de calculer les probabilité d'un événement pour une variable aléatoire suivant une loi normale, on utilisera des tables (voir l'annexe A.4.2) ou sa calculatrice.

Loi Gamma

On dit que X suit la loi Gamma de paramètre (α, β) , notée $\Gamma(\alpha, \beta)$, si

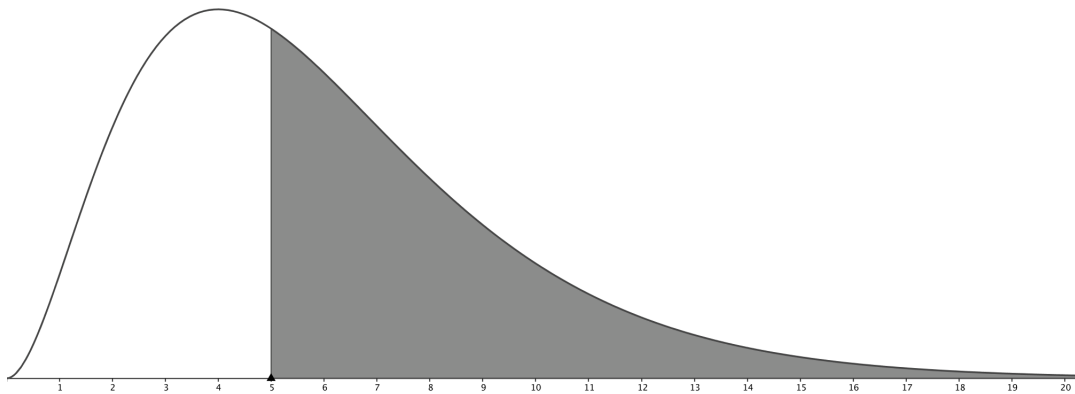
$$\begin{cases} X(\Omega) &= \mathbb{R}^+ \\ f_{\alpha, \beta}(t) &= t^{\alpha-1} \frac{\beta^\alpha e^{-\beta t}}{\Gamma(\alpha)} \text{ pour } t > 0. \end{cases}$$

La loi Gamma généralise la loi exponentielle. Un calcul d'espérance et variance donne

$$\mathbf{E}(X) = \alpha\beta \quad \text{et} \quad \mathbf{Var}(X) = \alpha\beta^2 \quad (2.43)$$

Traçons la densité pour $\alpha = 3$ et $\beta = 2$.

La densité de $\Gamma(3, 2)$



Remarque 2.21. On rappelle que la *fonction* Gamma est définie par

$$\Gamma : z \mapsto \int_0^{+\infty} t^{z-1} e^{-t} dt \quad (2.44)$$

C'est un prolongement de la factorielle. Par exemple, on a les relations

$$\forall n \in \mathbb{N}, \Gamma(n+1) = n! \quad \text{et} \quad \Gamma(1/2) = \sqrt{\pi}$$

Loi du χ^2

On dit que X suit la loi du χ^2 ("khi-deux") à k degré(s) de liberté si

$$\begin{cases} X(\Omega) &= \mathbb{R}^+ \\ f_k(t) &= \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} t^{\frac{k}{2}-1} e^{-\frac{t}{2}} \text{ pour } t > 0. \end{cases}$$

La loi du χ^2 est une loi très classique en statistique. Elle est liée au **test du χ^2** qui permet, par exemple, de savoir si un échantillon donné est en adéquation avec une loi de probabilité définie *a priori*. De plus,

$$\mathbf{E}(X) = k \quad \text{et} \quad \mathbf{Var}(X) = 2k \quad (2.45)$$

Exemple 2.22. Nous avons vu à l'exemple 2.4 que la loi du χ^2 à 1 degré de liberté est la loi du carré d'une loi normale centrée réduite. Plus généralement, la somme de k lois normales centrées réduites indépendantes suit une loi du χ^2 à k degré de liberté (voir exemple 2.26).

2.3 Convolution, loi d'une somme

Un problème courant consiste à trouver la loi de probabilité d'une somme de deux variables indépendantes $Z = X + Y$.

2.3.1 Cas discret

Le théorème des probabilités totales donne la solution du problème

$$\mathbf{P}(Z = z) = \sum_x \mathbf{P}(X = x \text{ et } Y = z - x) = \sum_y \mathbf{P}(X = z - y \text{ et } Y = y)$$

Lorsque X et Y sont indépendantes, on a

$$\mathbf{P}(Z = z) = \sum_x \mathbf{P}(X = x) \cdot \mathbf{P}(Y = z - x) = \sum_y \mathbf{P}(X = z - y) \cdot \mathbf{P}(Y = y) \quad (2.46)$$

Remarque 2.23. Certains termes peuvent être nuls. x ne prend pas nécessairement toutes les valeurs possibles de X mais uniquement celles compatibles avec l'événement $Z = z$ (de même pour y valeurs de Y).

Proposition 30.

La somme de deux variables aléatoires indépendantes de lois de POISSON $\mathcal{P}(\lambda)$ et $\mathcal{P}(\mu)$ suit la loi de POISSON $\mathcal{P}(\lambda + \mu)$.

Preuve. Soient X une variable aléatoire de loi $\mathcal{P}(\lambda)$, Y une variable aléatoire de loi $\mathcal{P}(\mu)$, avec X et Y indépendantes. Soit $Z = X + Y$, la variable aléatoire somme. X et Y prennent toutes les valeurs entières positives, Z prend donc aussi toutes les valeurs entières positives. Pour tout entier positif k

$$\begin{aligned} \mathbf{P}(Z = k) &= \sum_{i \in \mathbb{N}} \mathbf{P}(X = i) \cdot \mathbf{P}(Y = k - i) \\ &= \sum_{i=0}^k \left(e^{-\lambda} \frac{\lambda^i}{i!} \right) \times \left(e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} \right) \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \lambda^i \mu^{k-i} \\ &= \frac{e^{-(\lambda+\mu)}}{k!} \sum_{i=0}^k \binom{k}{i} \lambda^i \mu^{k-i} \\ \mathbf{P}(Z = k) &= \frac{e^{-(\lambda+\mu)}}{k!} (\lambda + \mu)^k \quad \text{par le binôme de NEWTON} \end{aligned}$$

Donc Z suit la loi de POISSON $\mathcal{P}(\lambda + \mu)$. ■

On a un résultat analogue pour la somme de loi binomiale. On démontrerait

Proposition 31.

La somme de deux variables aléatoires indépendantes de lois binomiale $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ suit une loi binomiale $\mathcal{B}(n + m, p)$.

La loi $\mathcal{B}(n, p)$ peut être considérée comme la somme de n variables aléatoires mutuellement indépendantes de BERNOULLI $\mathcal{B}(1, p)$. Dès lors, si X_i désigne n variables aléatoires indépendantes, alors $S_n = \sum_{i=1}^n X_i$ suit une loi binomiale et on retrouve

$$\mathbf{E}(S_n) = \sum_{i=1}^n \mathbf{E}(X_i) = \sum_{i=1}^n p = np \quad \text{et} \quad \mathbf{Var}(S_n) = \sum_{i=1}^n \mathbf{Var}(X_i) = npq$$

2.3.2 Cas continu

Avant de poursuivre, rappelons la définition d'un produit de convolution $*$. Sous réserve de convergence

$$(f * g)(x) = \int_{\mathbb{R}} f(y - x)g(y) \, dy \tag{2.47}$$

Proposition 32.

La densité de probabilité de la somme de deux variables aléatoires à densité indépendantes est le produit de convolution des densités des deux variables aléatoires. C'est à dire

$$f_{X+Y} = f_X * f_Y \tag{2.48}$$

La démonstration complète utilise la théorie de la mesure, donc n'est pas exposée dans le cadre de ce cours. Nous admettrons ce résultat.

Remarque 2.24. Cette formule est la traduction "continue" de la formule du cas discret. Pour la loi (fonction de densité ou probabilité ponctuelle) de Z en z , on ajoute (intégrer ou sommer) tous les produits (des fonctions de densité ou des probabilités ponctuelles) pour les valeurs x et y de X et Y telles que $x + y = z$.

Exemple 2.25. Soient X et Y deux variables aléatoires de loi respectives $\Gamma(r, \lambda)$ et $\Gamma(s, \lambda)$ avec X et Y indépendantes. Soit $Z = X + Y$, la variable aléatoire somme. X et Y prennent toutes les valeurs réelles positives, Z prend donc aussi toutes les valeurs réelles positives. Pour tout réel z

$$f_Z(z) = (f_X * f_Y)(z) = \begin{cases} 0 & \text{si } z < 0 \\ \int_0^z f_X(t) \cdot f_Y(z - t) \, dt & \text{sinon} \end{cases}$$

La fonction f_Z n'est pas forcément définie en 0, on peut choisir $f_Z(0) = 0$ et si $z > 0$

$$\begin{aligned} f_Z(z) &= \frac{\lambda^{r+s}}{\Gamma(r) \cdot \Gamma(s)} \int_0^z t^{r-1} e^{-\lambda t} \cdot (z - t)^{s-1} e^{-\lambda(z-t)} \, dt \\ &= \frac{\lambda^{r+s} e^{-\lambda z}}{\Gamma(r) \cdot \Gamma(s)} \int_0^z t^{r-1} (z - t)^{s-1} \, dt \\ &= \frac{\lambda^{r+s} e^{-\lambda z}}{\Gamma(r) \cdot \Gamma(s)} \int_0^1 (uz)^{r-1} (z - uz)^{s-1} z \, du && \text{par } t = uz \\ f_Z(z) &= \frac{\lambda^{r+s} z^{r+s-1} e^{-\lambda z}}{\Gamma(r) \cdot \Gamma(s)} \beta(r, s) && \text{où } \beta(r, s) = \int_0^1 u^{r-1} (1 - u)^{s-1} \, du \end{aligned}$$

f_Z est une densité, la valeur du réel $\beta(r, s)$ est telle que $\int_{\mathbb{R}} f_Z = 1$, donc **la fonction bêta** est

$$\beta(r, s) = \frac{\Gamma(r) \cdot \Gamma(s)}{\Gamma(r + s)} \tag{2.49}$$

En conclusion

$$f_Z(z) = (f_X * f_Y)(z) = \begin{cases} 0 & \text{si } z < 0 \\ \lambda^{r+s} \frac{z^{r+s-1} e^{-\lambda z}}{\Gamma(r+s)} & \text{sinon} \end{cases}$$

La loi de Z est donc la loi $\Gamma(r+s, \lambda)$. On retiendra donc

La somme de deux lois Gamma indépendantes $\Gamma(r, \lambda)$ et $\Gamma(s, \lambda)$ suit la loi Gamma $\Gamma(r+s, \lambda)$.

Exemple 2.26. Soient Y_1, \dots, Y_n des variables aléatoires indépendantes suivant des lois normales centrées réduites et Z définie par

$$Z = Y_1^2 + \dots + Y_k^2$$

Nous avons vu à l'exemple 2.4 que Y_i suit une loi du χ^2 à 1 degré de liberté. Or cette dernière est aussi une loi Gamma de paramètre $(\frac{1}{2}; \frac{1}{2})$. D'après ce qui précède, Z suit donc une loi $\Gamma(\frac{k}{2}; \frac{1}{2})$. C'est l'expression de la loi du χ^2 à k degré de liberté. En particulier, on en déduit l'expression de l'espérance et de la variance pour la loi du χ^2

$$\mathbf{E}(Z) = k\mathbf{E}(Y_1^2) = k \quad \text{et} \quad \mathbf{Var}(Z) = k\mathbf{Var}(Y_1^2) = 2k$$

De même que pour la somme de loi Γ , on démontrerait le résultat fondamental suivant

Proposition 33.

La somme de deux lois normales indépendantes $\mathcal{N}(\mu, \sigma)$ et $\mathcal{N}(\mu', \sigma')$ suit la loi normale $\mathcal{N}(\mu + \mu', \sqrt{\sigma^2 + \sigma'^2})$.

Ce résultat se démontre en faisant le produit de convolution de deux gaussiennes, ou plus simplement en utilisant les résultats connus sur la transformation de FOURIER (voir le paragraphe suivant).

2.4 Fonctions caractéristiques

2.4.1 Définition et principales propriétés

Tout d'abord la notion de variable aléatoire réelle s'étend au cas complexe en considérant X à valeurs dans \mathbb{C} et non plus seulement dans \mathbb{R} . Ainsi si X est une variable aléatoire réelle et t un paramètre réel, e^{itX} définit une variable aléatoire *complexe* définie sur (Ω, \mathcal{E}) .

Définition 34 (Fonction caractéristique).

On appelle **fonction caractéristique** de la variable aléatoire X la fonction ρ_X de la variable t définie par l'espérance mathématique de la variable aléatoire e^{itX} .

$$\rho_X(t) = \mathbf{E}(e^{itX}) \tag{2.50}$$

- si X est une variable aléatoire discrète, alors

$$\rho_X(t) = \sum_k e^{itx_k} \mathbf{P}(X = x_k) \tag{2.51}$$

- si X est une variable aléatoire continue ayant pour densité f , alors

$$\rho_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx \tag{2.52}$$

Remarque 2.27. Dans le cas continu, ρ_X coïncide avec la transformée de FOURIER (inverse) de la densité f pour un choix particulier des paramètres. Pour rappel, la transformée de f est

$$\mathcal{F}(f)(u) = \int_{-\infty}^{+\infty} e^{-iut} f(t) dt \quad (2.53)$$

et elle transforme le produit de convolution en produit $\mathcal{F}(f * g) = \mathcal{F}(f) \times \mathcal{F}(g)$.

Propriétés élémentaires de la fonction caractéristique

Proposition 35.

- ρ_X est continue sur \mathbb{R} et bornée par $\rho_X(0) = 1$.
- Pour tout couple $(a, b) \in \mathbb{R}^2$, on a $\rho_{aX+b}(t) = e^{ibt} \rho_X(at)$.
- Si X et Y sont indépendantes alors

$$\rho_{X+Y} = \rho_X \rho_Y \quad (2.54)$$

Preuve. La première propriété provient du fait que X étant une variable aléatoire réelle alors

$$|\mathbf{E}(e^{itX})| \leq \mathbf{E}(|e^{itX}|) = \mathbf{E}(1) = \mathbf{P}(\Omega) = 1 = \rho_X(0)$$

Les autres propriétés sont immédiates, en utilisant les résultats connus sur la transformée de FOURIER et celui sur la somme de variables aléatoires indépendantes (voir la propriété 32). ■

2.4.2 Fonctions caractéristiques et moments

Proposition 36.

- Si $\mathbf{E}(X^s) < +\infty$ pour $s \in \mathbb{N}^*$ alors

$$\rho_X^{(s)}(0) = i^s \mathbf{E}(X^s) \quad (2.55)$$

- En particulier

$$\mathbf{E}(X) = -i\rho_X'(0) \quad \text{et} \quad \mathbf{Var}(X) = \rho_X'^2(0) - \rho_X''(0) \quad (2.56)$$

Preuve. La condition $\mathbf{E}(X^s) < +\infty$ permet de dériver s fois la fonction caractéristique. Par exemple, dans le cas continu, on a

$$\rho_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx$$

D'où
$$\rho_X^{(s)}(t) = \int_{\mathbb{R}} \frac{\partial^s}{\partial t^s} (e^{itx} f(x)) dx = i^s \int_{\mathbb{R}} x^s \cdot e^{itx} f(x) dx$$

En particulier
$$\rho_X^{(s)}(0) = i^s \int_{\mathbb{R}} x^s \cdot f(x) dx = i^s \mathbf{E}(X^s)$$

La seconde relation découle des définitions. ■

Remarque 2.28. Si la fonction caractéristique est développable en série entière, alors par la formule de TAYLOR

$$\rho_X(t) = \sum_{s=0}^{+\infty} \frac{t^s}{s!} \cdot i^s \mathbf{E}(X^s) \quad (2.57)$$

En particulier

$$\rho_X(t) = 1 + it\mathbf{E}(X) - \frac{t^2}{2}\mathbf{E}(X^2) + t^2\varepsilon(t) \quad \text{avec} \quad \varepsilon(t) \xrightarrow[t \rightarrow 0]{} 0 \quad (2.58)$$

Réciproquement, par injectivité de la transformée de FOURIER, deux variables aléatoires ayant la même fonction caractéristique, ont même loi de probabilité: on retiendra donc que

la fonction caractéristique détermine la loi de la variable aléatoire X.

En pratique, pour montrer que deux variables aléatoires ont même loi, on peut tenter de prouver qu'elles ont même fonction caractéristique. C'est ainsi que l'on démontrera le théorème central-limite. De plus, il existe des formules permettant connaissant ρ_X de déterminer la loi de X. Elles ne seront pas utilisées dans ce cours.

2.4.3 Fonctions caractéristiques des lois usuelles

Loi Binomiale	$\mathcal{B}(n; p)$	$\rho_X(t) = (q + pe^{it})^n$
Loi de POISSON	$\mathcal{P}(\lambda)$	$\exp(\lambda(e^{it} - 1))$
Loi géométrique	$\mathcal{G}(p)$	$\frac{pe^{it}}{1 - qe^{it}}$
Loi uniforme discrete	$\mathcal{U}_{[a,b]}$	$\frac{e^{iat}}{b - a + 1} \sum_{k=0}^{b-a} e^{ikt}$
Loi uniforme continue	$\mathcal{U}_{[a,b]}$	$\frac{e^{itb} - e^{ita}}{it(b - a)}$
Loi exponentielle	$\mathcal{E}(\lambda)$	$\left(1 - \frac{it}{\lambda}\right)^{-1}$
Loi normale	$\mathcal{N}(\mu, \sigma)$	$\exp\left(\mu it - \frac{\sigma^2 t^2}{2}\right)$

Justifions rapidement ces valeurs. Pour la loi binomiale

$$\rho_X(t) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{itk} = \sum_{k=0}^n \binom{n}{k} (e^{it}p)^k q^{n-k} = (q + pe^{it})^n$$

Pour la loi de POISSON,

$$\rho_X(t) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} e^{itk} = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{(e^{it}\lambda)^k}{k!} = \exp(\lambda(e^{it} - 1))$$

Pour la loi géométrique,

$$\rho_X(t) = \sum_{k=1}^{+\infty} pq^{k-1} e^{itk} = pe^{it} \sum_{k=0}^{+\infty} (e^{it})^k = \frac{pe^{it}}{1 - qe^{it}}$$

Pour la loi uniforme continue,

$$\rho_X(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

Pour la loi exponentielle,

$$\rho_X(t) = \lambda \int_0^{+\infty} e^{-(\lambda-it)x} dx = \frac{\lambda}{\lambda - it}$$

Terminons par la loi normale centrée réduite. Nous savons que la transformée de FOURIER d'une fonction gaussienne est encore de type gaussien (voir (2.53) pour la définition). Plus précisément, on prouve que la fonction caractéristique de la loi normale centrée réduite est

$$\rho_X(t) = \exp\left(-\frac{t^2}{2}\right) \quad (2.59)$$

En utilisant les propositions 29 et 35, on obtient la fonction caractéristique d'une loi normale $\mathcal{N}(\mu, \sigma)$.

Remarque 2.29. En utilisant les fonctions caractéristiques, on retrouve les résultats suivants.

- La somme de deux lois binomiales $\mathcal{B}(n, p)$ et $\mathcal{B}(n', p)$ indépendantes est une loi binomiale $\mathcal{B}(n + n', p)$.
- La somme de deux lois de POISSON $\mathcal{P}(\lambda)$ et $\mathcal{P}(\lambda')$ indépendantes est une loi de POISSON $\mathcal{P}(\lambda + \lambda')$.
- La somme de deux lois normales indépendantes est encore une loi normale.

2.5 Convergences des suites de variables aléatoires

Soit $(X_n)_n$ une suite de variables aléatoires, c'est une suite de fonctions de Ω dans \mathbb{R} : le caractère aléatoire des résultats des expériences rend nécessaire l'introduction de nouvelles notions de convergence et de limite, différentes de celle de l'analyse mathématique classique.

2.5.1 Différents type de convergence

La convergence en probabilité

Définition 37 (Convergence en probabilité).

La suite $(X_n)_n$ converge en probabilité vers la variable aléatoire X si pour tout nombre réel $\varepsilon > 0$, on a

$$\lim_{n \rightarrow +\infty} \mathbf{P}(|X_n - X| \geq \varepsilon) = 0 \quad (2.60)$$

On note $X_n \xrightarrow{\mathbf{P}} X$

Exemple 2.30. Par une application directe de l'inégalité de BIENAYMÉ-TCHEBYSHEV, on montre que si $\mathbf{E}(X_n) \rightarrow a \in \mathbb{R}$ et $\mathbf{Var}(X_n) \rightarrow 0$ alors

$$X_n \xrightarrow{\mathbf{P}} a \quad (2.61)$$

La convergence presque sûre

Définissons d'abord l'égalité presque sûre de deux variables aléatoires :

$$X \text{ et } Y \text{ sont égales presque sûrement si } \mathbf{P}(\{\omega : X(\omega) \neq Y(\omega)\}) = 0$$

C'est l'égalité presque partout des fonctions mesurables. On définit alors la convergence presque-sûre.

Définition 38 (Convergence presque-sûre).

La suite $(X_n)_n$ converge presque sûrement vers la variable aléatoire X si

$$\mathbf{P}\left(\{\omega : \lim_{n \rightarrow +\infty} X_n(\omega) \neq X(\omega)\}\right) = 0 \quad (2.62)$$

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

La convergence en moyenne quadratique

Définition 39 (La convergence en moyenne quadratique).

La suite $(X_n)_n$ converge en moyenne quadratique vers la variable aléatoire X si

$$\lim_{n \rightarrow +\infty} \mathbf{E} |X_n - X|^2 = 0 \quad (2.63)$$

On note $X_n \xrightarrow{L^2} X$

La convergence en loi

Définition 40 (Convergence en loi).

La suite $(X_n)_n$ converge en loi vers la variable aléatoire X de fonction de répartition F , si, en tout point de continuité de F , la suite $(F_n)_n$ des fonctions de répartition des X_n converge vers F .

On note $X_n \xrightarrow{Loi} X$

Exemple 2.31. Pour les lois discrètes, on montre que la convergence en loi vers une variable discrète est équivalente à

$$\forall x \in \mathbb{R}, \quad \mathbf{P}(X_n = x) \rightarrow \mathbf{P}(X = x) \quad (2.64)$$

Mais une variable discrète peut converger vers une loi continue... Et, pour les lois à densité, on montre que la convergence en loi implique la convergence simple de la suite des fonctions de densité

$$\forall x \in \mathbb{R}, \quad f_n(x) \rightarrow f(x) \quad (2.65)$$

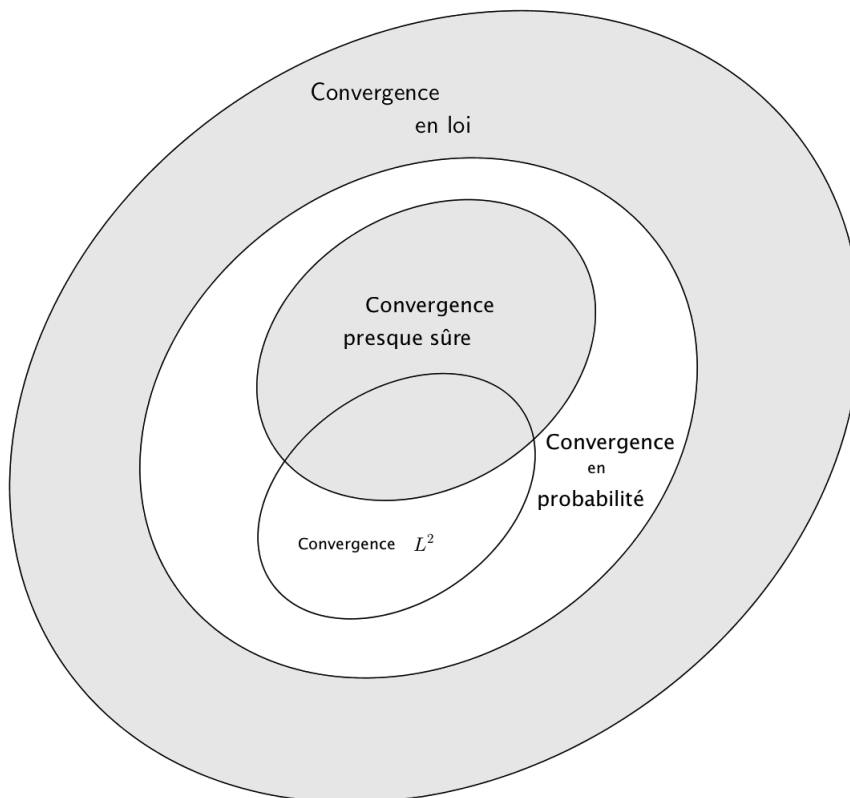
La convergence en loi est intimement liée à la convergence des fonctions caractéristiques comme le précise le théorème suivant qui nous sera très utile pour prouver le théorème central-limite.

Théorème 41 (de LÉVY).

Soit $(X_n)_n$ une suite de variable aléatoire réelle, alors les fonctions caractéristiques ρ_{X_n} convergent simplement vers ρ_X la fonction caractéristique de X si et seulement si $(X_n)_n$ converge en loi vers la variable aléatoire X .

2.5.2 Hiérarchie des différentes convergences

La figure suivante résume les implications des différents modes de convergence probabilistes. La convergence en loi est la plus faible, c'est la plus utilisée.

**2.6 Applications: loi des grands nombres, théorème central-limite****2.6.1 Lois des grands nombres**

Il existe plusieurs théorèmes appelés “lois des grands nombres”.

Théorème 42 (Loi faible des grands nombres).

Si X admet un moment d'ordre 2, la variable aléatoire \bar{X}_n , moyenne arithmétique de n variables aléatoires indépendantes X_1, X_2, \dots, X_n de même loi que la variable aléatoire X converge en probabilité vers son espérance mathématique $\mathbf{E}(X)$. Autrement dit

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{\mathbf{P}} \mathbf{E}(X) \quad (2.66)$$

Preuve. Calculons l'espérance mathématique et la variance de la moyenne arithmétique de \overline{X}_n . On a par linéarité de l'espérance sachant que les X_i ayant même loi, ils ont même espérance $\mathbf{E}(X)$.

$$\mathbf{E}(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}(X_i) = \frac{n\mathbf{E}(X)}{n} = \mathbf{E}(X)$$

Par indépendance
$$\mathbf{Var}(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \mathbf{Var}(X_i) = \frac{\mathbf{Var}(X)}{n}$$

Donc la variable aléatoire X_n vérifie $\mathbf{E}(X_n) \xrightarrow{n \rightarrow +\infty} \mathbf{E}(X)$ et $\mathbf{Var}(X_n) \xrightarrow{n \rightarrow +\infty} 0$, ce qui suffit à établir le résultat d'après l'inégalité de BIENAYMÉ-TCHEBYSHEV. ■

Considérons une expérience aléatoire, et un événement A de probabilité théorique p . Répétons n fois l'expérience et désignons par X_k , la variable aléatoire qui prend la valeur 1 si A se réalise à la k -ième épreuve, 0 sinon. X_k suit la loi de BERNOULLI $\mathcal{B}(1, p)$. On a alors

$$F_n(A) = \frac{1}{n} \sum_{i=1}^n X_i = \overline{X}_n \quad \text{et} \quad \mathbf{E}(\overline{X}_n) = p$$

Par la loi des grands nombres, on en déduit l'énoncé suivant

Corollaire 43 (Théorème de BERNOULLI).

Lorsque le nombre d'expériences aléatoires augmente indéfiniment, la fréquence d'apparition $F_n(A)$ d'un événement A converge en probabilité vers sa probabilité théorique p . Autrement dit

$$\forall \varepsilon > 0, \quad \mathbf{P}(|F_n(A) - p| > \varepsilon) \xrightarrow{n \rightarrow +\infty} 0$$

Donnons maintenant une version plus précise de la loi des grands nombres mais sans preuve.

Théorème 44 (Loi forte des grands nombres).

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite v.a.r indépendantes et de même loi. Si de plus $\mathbf{E}(X_1^2) < +\infty$, alors on a

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbf{E}(X_1)$$

On en déduit donc

Corollaire 45.

$$F_n(A) \xrightarrow[n \rightarrow +\infty]{p.s.} p$$

Ce résultat prouve la pertinence du formalisme des variables aléatoires et de l'axiomatique de KOLMOGOROV puisqu'elle fait le lien avec l'approche fréquentielle d'une probabilité d'un événement!

Application 2.32. La méthode de **Monte-Carlo** est une application directe de la loi des grands nombres pour le calcul d'intégrales ayant un très grand nombre de paramètres.

2.6.2 Autres types de théorèmes

Théorème 46 (Convergence de la loi binomiale vers la loi de POISSON).

Soit $(X_n)_n$ une suite de variables aléatoires de lois binomiales $\mathcal{B}(n, p_n)$ telles que

$$np_n \xrightarrow{n \rightarrow +\infty} \lambda \in \mathbb{R}^+ \quad (2.67)$$

Alors la suite $(X_n)_n$ converge en loi vers une variable de POISSON $\mathcal{P}(\lambda)$.

$$\mathcal{B}(n, p_n) \xrightarrow[n \rightarrow +\infty]{\text{Loi}} \mathcal{P}(\lambda) \quad (2.68)$$

Preuve. Soient k un entier naturel fixé et n un entier supérieur à k

$$\begin{aligned} \mathbf{P}(X_n = k) &= \binom{n}{k} p^k q^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} \cdot p^k q^{n-k} \\ \mathbf{P}(X_n = k) &= 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{(pn)^k q^{n-k}}{k!} \end{aligned}$$

Or lorsque n tend vers $+\infty$,

$$1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \rightarrow 1 \quad \text{et} \quad (1-p)^{n-k} \rightarrow e^{-\lambda}$$

D'où
$$\mathbf{P}(X_n = k) \xrightarrow[n \rightarrow +\infty]{} \lambda^k \frac{e^{-\lambda}}{k!}$$

En pratique l'approximation d'une loi binomiale par une loi de POISSON est "acceptable" si " n est grand, p petit et np pas trop grand", la loi de POISSON reste associée à des phénomènes rares. Les valeurs limites qui permettent l'approximation diffèrent suivant les utilisateurs, les branches professionnelles, etc. Fréquemment, elles sont :

$$n \text{ grand} : n > 30 ; p \text{ petit} : p < 0,1 ; \text{ et } np \text{ pas trop grand} : np < 10.$$

Théorème 47 (Convergence de la loi hypergéométrique vers la loi binomiale).

Soit $(X_N)_N$ une suite de variables aléatoires hypergéométriques $\mathcal{H}(N, n, p)$. Alors la suite de variables aléatoires $(X_N)_N$ converge en loi vers une variable binomiale $\mathcal{B}(n, p)$ lorsque N tend vers $+\infty$. Ce qui se note

$$\mathcal{H}(N, n, p) \xrightarrow[N \rightarrow +\infty]{\text{Loi}} \mathcal{B}(n, p) \quad (2.69)$$

La démonstration se fait sans difficulté avec les probabilités ponctuelles. En pratique, l'approximation est "acceptable" dès que $\frac{n}{N} < 10\%$, c'est à dire dès que la population est 10 fois plus grande que l'échantillon, ce qui arrive fréquemment (sondages, contrôle qualité, etc). Par exemple, un échantillon de 2 000 individus conviendra aussi bien pour faire un sondage dans un population de 200 000 habitants, que dans un population de 2 millions d'habitants.

Théorème 48 (De MOIVRE-LAPLACE).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires binomiales $\mathcal{B}(n, p)$, alors

$$\frac{X_n - np}{npq} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0; 1) \quad (2.70)$$

Preuve. On passe par les fonctions caractéristiques. Posons

$$Y_n = \frac{X_n - np}{npq}$$

On sait que $\rho_{X_n}(t) = (pe^{it} + q)^n$, donc

$$\rho_{Y_n} = \exp\left(i \frac{np \cdot t}{\sqrt{npq}}\right) \cdot \left(pe^{\frac{i \cdot t}{\sqrt{npq}}} + q\right)^n \rightarrow e^{-t^2/2}$$

On reconnaît la fonction caractéristique de la loi normale centrée réduite, ce qui termine la démonstration par le théorème de LÉVY. ■

Application en pratique. Lorsque n est assez grand, on peut approximer la loi binomiale $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np; npq)$. On donne parfois comme condition np et $nq > 5$. Plus généralement les conditions d'approximation sont

$$n \ll \text{grand} \gg : n > 30, p \text{ pas } \ll \text{trop petit} \gg : p > 0,1 \text{ et } np \ll \text{assez grand} \gg : np > 10.$$

Correction de continuité. Il convient cependant d'effectuer ce que l'on appelle une correction de continuité: la convergence de la loi binomiale vers la loi normale se traduit par le fait que les extrémités supérieures des bâtons du diagramme de la loi binomiale $\mathcal{B}(n, p)$ sont voisines de la courbe de densité de la loi $\mathcal{N}(np; npq)$. On obtient donc une valeur approchée de $\mathbf{P}(X = k)$ par la surface sous la courbe de densité comprise entre les droites d'abscisse $k - \frac{1}{2}$ et $k + \frac{1}{2}$. Si U suit une loi $\mathcal{N}(0; 1)$, on a

$$\mathbf{P}(X = x) \simeq \mathbf{P}\left(\frac{x - \frac{1}{2} - np}{\sqrt{npq}} < U < \frac{x + \frac{1}{2} - np}{\sqrt{npq}}\right) \quad (2.71)$$

On prendra garde aux inégalités strictes et aux $\pm \frac{1}{2}$

$$\mathbf{P}(X \leq x) \simeq \mathbf{P}\left(U < \frac{x + \frac{1}{2} - np}{\sqrt{npq}}\right) \quad \text{et} \quad \mathbf{P}(X < x) \simeq \mathbf{P}\left(U < \frac{x - \frac{1}{2} - np}{\sqrt{npq}}\right) \quad (2.72)$$

Exemple 2.33. Approximation de la loi binomiale $\mathcal{B}(50, \frac{1}{2})$ par la loi normale $\mathcal{N}(25; \frac{5}{2})$. Le calcul de $\mathbf{P}(24 \geq X \geq 26)$ par la loi binomiale exacte donne à peu près 0,3282. A comparer avec l'approximation par la loi normale

- sans correction de continuité 0,223,
- avec correction de continuité 0,328.

Théorème 49 (Convergence en loi de la loi de POISSON vers la loi normale).

Soit $(X_\lambda)_\lambda$ une suite de variables aléatoires de POISSON $\mathcal{P}(\lambda)$, quand $\lambda \rightarrow +\infty$,

$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow[\lambda \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0; 1) \quad (2.73)$$

La démonstration passe par les fonctions caractéristiques. Pour l'application en pratique, λ doit être assez grand, afin d'approximer la loi de POISSON $\mathcal{P}(\lambda)$, par la loi normale $\mathcal{N}(\lambda; \lambda)$. L'approximation est en général acceptée dès que $\lambda > 10$, elle est très satisfaisante pour $\lambda > 18$. Il faut encore faire une correction de continuité car on passe d'une loi discrète à une loi continue. L'approximation par une loi de POISSON permet ainsi d'éviter le calcul pénible des coefficients binomiaux.

2.6.3 Somme de variables aléatoires et théorème central-limite

Le théorème suivant est connu sous le nom de "théorème central-limite" parfois noté en abrégé TCL, aussi appelé à tort théorème de la limite centrée. Il établit la convergence en loi d'une somme de variables aléatoires indépendantes vers la loi normale. Ce théorème d'une très grande généralité a un rôle capital en statistique.

Théorème 50 (Théorème central-limite).

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes de même loi, d'espérance μ et d'écart-type σ . Alors

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0; 1) \quad (2.74)$$

Autrement dit, la variable aléatoire $\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ suit approximativement la loi normale $\mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$.

Preuve. Posons pour tout entier k , $Y_k = \frac{X_k - \mu}{\sigma\sqrt{n}}$, de sorte que $\mathbf{E}(Y_k) = 0$ et $\mathbf{Var}(Y_k) = \frac{1}{n}$.

Considérons de plus Z_n une variable centrée réduite

$$Z_n = \sum_{k=1}^n Y_k = \sum_{k=1}^n \frac{X_k - \mu}{\sigma\sqrt{n}} \Rightarrow \mathbf{E}(Z_n) = 0 \quad \text{et} \quad \mathbf{Var}(Z_n) = 1$$

Il faut montrer que Z_n tend en loi vers la loi normale. Les variables X_1, X_2, \cdots ayant même loi, elles ont même fonction caractéristique ρ_X . D'après la proposition 35

$$\rho_{Y_k}(t) = \exp\left(-it \frac{\mu}{\sigma\sqrt{n}}\right) \cdot \rho_X\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Comme X admet un moment d'ordre 2, on obtient un développement limité à l'ordre 2 en 0 (voir (2.58)) lorsque $n \rightarrow +\infty$:

$$\begin{aligned} \rho_X\left(\frac{t}{\sigma\sqrt{n}}\right) &= 1 + i \left(\frac{t}{\sigma\sqrt{n}}\right) \mathbf{E}(X) - \frac{1}{2} \left(\frac{t}{\sigma\sqrt{n}}\right)^2 \mathbf{E}(X^2) + o\left(\frac{1}{n}\right) \\ &= 1 + it \frac{\mu}{\sigma\sqrt{n}} - \frac{t^2(\sigma^2 + \mu^2)}{2n\sigma^2} + o\left(\frac{1}{n}\right) \end{aligned}$$

En effectuant un développement limité de $\exp\left(-it\frac{\mu}{\sigma\sqrt{n}}\right)$, on a

$$\exp\left(-it\frac{\mu}{\sigma\sqrt{n}}\right) = 1 - it\frac{\mu}{\sigma\sqrt{n}} - \frac{t^2\mu^2}{2n\sigma^2} + o\left(\frac{1}{n}\right)$$

On en déduit celui de ρ_{Y_k}

$$\begin{aligned} \rho_{Y_k}(t) &= \left(1 + it\frac{\mu}{\sigma\sqrt{n}} - \frac{t^2(\sigma^2 + \mu^2)}{2n\sigma^2} + o\left(\frac{1}{n}\right)\right) \cdot \left(1 - it\frac{\mu}{\sigma\sqrt{n}} - \frac{t^2\mu^2}{2n\sigma^2} + o\left(\frac{1}{n}\right)\right) \\ &= \left(1 - it\frac{\mu}{\sigma\sqrt{n}} - \frac{t^2\mu^2}{2n\sigma^2}\right) + it\frac{\mu}{\sigma\sqrt{n}} \cdot \left(1 - it\frac{\mu}{\sigma\sqrt{n}}\right) - \frac{t^2(\sigma^2 + \mu^2)}{2n\sigma^2} + o\left(\frac{1}{n}\right) \\ &= 1 - \frac{t^2\mu^2}{2n\sigma^2} + \frac{t^2\mu^2}{n\sigma^2} - \frac{t^2(\sigma^2 + \mu^2)}{2n\sigma^2} + o\left(\frac{1}{n}\right) \\ \rho_{Y_k}(t) &= 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \end{aligned}$$

La fonction caractéristique de Z_n est, comme les variables Y_k sont indépendantes, donnée par

$$\rho_{Z_n}(t) = \prod_{k=1}^n \rho_{Y_k}(t) = (\rho_{Y_n}(t))^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n$$

Afin d'appliquer le théorème de LÉVY, il faut étudier la différence

$$\rho_{Z_n}(t) - \exp\left(-\frac{t^2}{2}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n - \left(\exp\left(-\frac{t^2}{2n}\right)\right)^n \quad (2.75)$$

On va utiliser le lemme suivant

Lemme 51.

Si $z_1, \dots, z_n, y_1, \dots, y_n$ sont des complexes de modules inférieur ou égal à 1, on a

$$\left| \prod_{i=1}^n z_i - \prod_{i=1}^n y_i \right| \leq \sum_{i=1}^n |z_i - y_i| \quad (2.76)$$

La preuve se fait par récurrence, en écrivant que

$$\begin{aligned} \left(\prod_{i=1}^n z_i - \prod_{i=1}^n y_i \right) &= \left(\prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} y_i \right) \times z_n + \prod_{i=1}^{n-1} y_i (z_n - y_n) \\ \Rightarrow \left| \prod_{i=1}^n z_i - \prod_{i=1}^n y_i \right| &\leq \left| \prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} y_i \right| \times \underbrace{|z_n|}_{\leq 1} + \underbrace{\left| \prod_{i=1}^{n-1} y_i \right|}_{\leq 1} \times |z_n - y_n| \end{aligned}$$

Appliquons ce lemme à la relation (2.75)

$$\begin{aligned} \left| \rho_{Z_n}(t) - \exp\left(-\frac{t^2}{2}\right) \right| &\leq \left| \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n - \left(\exp\left(-\frac{t^2}{2n}\right)\right)^n \right| \\ &\leq \sum_{i=1}^n \left| \exp\left(-\frac{t^2}{2n}\right) - 1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right| \\ \left| \rho_{Z_n}(t) - \exp\left(-\frac{t^2}{2}\right) \right| &\leq C^{\text{te}} n \left| \frac{t^2}{2n} \right|^2 + o(1) \end{aligned}$$

Car pour tout réel x inférieur à 1,

$$|\exp(x) - 1 - x| \leq C^{\text{te}} x^2 \quad (2.77)$$

On démontre ainsi que ρ_{Z_n} converge simplement vers la fonction caractéristique de la loi normale centrée réduite. Ce qui conclut la preuve par le théorème de LÉVY. ■

Remarque 2.34. On remarque si les variables aléatoires sont des variables de BERNOULLI, on retrouve le cas particulier du théorème de MOIVRE-LAPLACE et la convergence de la loi binomiale vers la loi normale.

Application 2.35. Lorsque n est assez grand, on peut approximer la loi de la somme $X_1 + X_2 + \dots + X_n$ par la loi normale $\mathcal{N}(n\mu; n\sigma)$. L'approximation est en général acceptée pour $n > 30$.

Exemple 2.36. 120 personnes s'adressent à un même guichet de Sécurité Sociale pour se faire rembourser des frais de maladie. Soit X_k la somme versée à la k -ième personne. On sait, d'après des études statistiques, que l'espérance mathématique de X_k est $m_{X_k} = 50$ euros avec un écart-type $\sigma_{X_k} = 30$ euros. Le guichet dispose au total de 6500 euros, calculons la probabilité pour que cette somme d'argent suffise à rembourser les 120 personnes. Posons $X = \sum_{k=1}^{120} X_k$ et supposons les X_k indépendantes, on a

$$\mathbf{E}(X) = 120 \times 50 = 6000; \quad \mathbf{Var}(X) = 120 \times (30)^2 = 10800; \quad \text{et} \quad \sigma_X \simeq 329$$

Sans aucune autre information sur la loi des X_k , le nombre de termes étant grand, le théorème central-limite s'applique et la loi de X est voisine de la loi normale. La probabilité cherchée est alors donnée par

$$\mathbf{P}(X < 6500) \simeq \mathbf{P}\left(U < \frac{6500 - 6000}{329}\right) \simeq 0,935$$

On peut démontrer un théorème plus général dû à LINDEBERG.

Théorème 52 (de LINDELBERG).

Soit (X_n) une suite de variables aléatoires indépendantes d'espérance μ_n et d'écart-type σ_n .

Soit $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Si pour tout $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbf{E} \left((X_i - \mu_i)^2 \cdot \mathbf{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}} \right) = 0 \quad (2.78)$$

alors on a

$$\frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n} \xrightarrow[n \rightarrow +\infty]{\text{Loi}} \mathcal{N}(0; 1) \quad (2.79)$$

La condition de LINDEBERG (2.78) exprime que les variables sont "uniformément proches de l'espérance" avec une grande probabilité. Le théorème dit qu'à force d'ajouter de telles variables, on finit par obtenir une loi normale.

Exemple 2.37. L'existence des moments $\mathbf{E}(X)$ et $\mathbf{Var}(X)$, espérance et variance de X , sont indispensables pour ces théorèmes. La loi de CAUCHY de densité

$$\forall t \in \mathbb{R}, \quad f(t) = \frac{1}{\pi(1+t^2)}$$

n'admet aucun moment et on peut prouver que $\overline{X_n}$ suit la loi de CAUCHY pour tout n et non une loi normale approximativement.

Application 2.38. La mesure d'une grandeur physique (pression, température..) est soumise à un très grand nombre d'erreurs accidentelles provenant de causes diverses et variées. On peut faire l'hypothèse raisonnable que l'erreur totale est la somme d'un très grand nombre de petites erreurs élémentaires et indépendantes. Cette erreur se représente donc comme une variable aléatoire, somme de variables aléatoires indépendantes, dont la loi suit approximativement une loi normale d'après les deux derniers théorèmes. Ce que l'on vérifie expérimentalement...

Chapitre 3

Couples de variables aléatoires

Soient X et Y deux variables aléatoires réelles définies sur un même espace probabilisé $(\Omega, \mathcal{E}, \mathbf{P})$. À chaque événement ω de Ω , on associe le couple $(X(\omega), Y(\omega)) \in \mathbb{R}^2$. On a défini ainsi un couple (X, Y) de variables aléatoires réelles, aussi appelé vecteur aléatoire.

3.1 Lois conjointes et lois marginales

3.1.1 Fonction de répartition conjointe et marginales

Définition 53 (Fonction de répartition conjointe).

Soient X et Y deux variables aléatoires réelles définies sur un même espace probabilisé $(\Omega, \mathcal{E}, \mathbf{P})$. On appelle **fonction de répartition conjointe** de X et Y , ou fonction de répartition du couple (X, Y) , la fonction $F_{X,Y}$ de deux variables définie par

$$F_{X,Y}(x, y) = \mathbf{P}(X \leq x \text{ et } Y \leq y) \quad (3.1)$$

La fonction $F_{X,Y}$ possède les propriétés suivantes

$$F_{X,Y}(x, -\infty) = F_{X,Y}(y, -\infty) = 0 \quad \text{et} \quad F_{X,Y}(+\infty, +\infty) = 1 \quad (3.2)$$

En faisant tendre x ou y indépendamment vers l'infini:

Définition 54 (Fonctions de répartition marginales).

On définit les **fonctions de répartition marginales** de X et Y , notées F_X et F_Y :

$$F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) = \mathbf{P}(X \leq x) \quad \text{et} \quad F_Y(y) = \lim_{x \rightarrow +\infty} F_{X,Y}(x, y) = \mathbf{P}(Y \leq y) \quad (3.3)$$

Connaissant $F_{X,Y}$, on peut déterminer F_X et F_Y par les formules ci dessus. La réciproque est fautive, il faut avoir des informations sur le conditionnement des deux variables. De plus, si les deux variables aléatoires sont indépendantes alors :

$$\mathbf{P}(X \leq x \text{ et } Y \leq y) = \mathbf{P}(X \leq x) \cdot \mathbf{P}(Y \leq y)$$

Cela donne la propriété

Proposition 55.

Deux variables aléatoires X et Y sont indépendantes si et seulement si pour tout couple de réels (x, y)

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y) \tag{3.4}$$

L’objet de ce chapitre est d’étudier le cas des variables “liées”. Tout d’abord, il faut définir des notions de probabilité ponctuelle d’un couple dans le cas discret, ou de fonction de densité d’un couple dans le cas continu. Il est possible de considérer un couple composé d’une variable aléatoire discrète et d’une variable aléatoire continue, mais ce n’est pas étudié dans le cadre de ce cours.

Lois conjointes et lois marginales. Cas discret

Définition 56 (Loi conjointe).

Si X et Y prennent des valeurs x_i et y_j en nombre fini ou dénombrable, la loi conjointe du couple (X, Y) , notée $\mathbf{P}_{X,Y}$, est entièrement définie par l’ensemble des nombres

$$\mathbf{P}_{X,Y}(x_i, y_j) = \mathbf{P}(X = x_i \text{ et } Y = y_j) \tag{3.5}$$

On note $p_{i,j} = \mathbf{P}(X = x_i \text{ et } Y = y_j)$.

On a bien sûr

$$\sum_{i,j} p_{i,j} = 1 \tag{3.6}$$

Dans le cas fini, cette loi de probabilité conjointe peut se mettre sous la forme d’un tableau. Par exemple

	X			
Y		$x_1 = 0$	$x_2 = 1$	$x_3 = 2$
$y_1 = 1$	0,1	0,2	0,2	
$y_2 = 2$	0,1	0,3	0,1	

Définition 57 (Lois marginales).

On appelle lois marginales les lois de probabilité de X et de Y prises séparément.

- Loi marginale de X : $\mathbf{P}(X = x_i) = \sum_j p_{i,j} := p_i$,
- Loi marginale de Y : $\mathbf{P}(Y = y_i) = \sum_i p_{i,j} := p_j$.

Dans l’exemple ci-dessus, on les trouve dans les “marges” du tableau :

	X				
Y		$x_1 = 0$	$x_2 = 1$	$x_3 = 2$	Total
$y_1 = 1$	0,1	0,2	0,2	0,5	Loi
$y_2 = 2$	0,1	0,3	0,1	0,5	de
Total	0,2	0,5	0,3	1	X

Loi marginale de Y

Cas particulier des variables aléatoires indépendantes.

Proposition 58.

Deux variables aléatoires discrètes X et Y sont indépendantes si et seulement si pour tout couple (i, j), on a $p_{ij} = p_i \cdot p_j$.

Dans l'exemple précédent, si X et Y sont deux variables aléatoires indépendantes avec les mêmes lois marginales, la loi conjointe du couple est

	X	$x_1 = 0$	$x_2 = 1$	$x_3 = 2$	Total
Y	$y_1 = 1$	0,1	0,25	0,15	0,5
$y_2 = 2$		0,1	0,25	0,15	0,5
Total		0,2	0,5	0,3	1

Sur cet exemple, on voit bien que les lois marginales ne suffisent pas à déterminer la loi du couple.

Lois conjointes et lois marginales. Cas continu

Reprenons la fonction de répartition conjointe $F_{X,Y}$.

Définition 59 (densité de probabilité d'un couple).

Soit un couple de v.a. (X, Y) de fonction de répartition conjointe $F_{X,Y}$. Lorsque $F_{X,Y}$ est une fonction de classe \mathcal{C}^2 , on définit la **densité du couple** (X, Y), notée $f_{X,Y}$, par

$$f_{X,Y} = \frac{\partial^2 F_{X,Y}}{\partial x \partial y} \tag{3.7}$$

de sorte que

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv \tag{3.8}$$

On a bien sûr $\int_{\mathbb{R}^2} f_{X,Y}(u, v) du dv = 1$.

Proposition 60.

On a alors, pour les fonctions de répartition marginales

$$F_X(x) = \int_{-\infty}^x \int_{\mathbb{R}} f_{X,Y}(u, v) dv du \quad \text{et} \quad F_Y(y) = \int_{-\infty}^y \int_{\mathbb{R}} f_{X,Y}(u, v) dv du \tag{3.9}$$

La probabilité d'un événement se calcule alors par une intégrale (double). Soit A un événement, et soit D la partie de \mathbb{R}^2 correspondante ; c'est-à-dire $A = \{(u, v) \in D\}$. On a

$$\mathbf{P}(A) = \iint_D f_{X,Y} \tag{3.10}$$

On définit les fonctions de densité marginales.

Définition 61 (Densités marginales).

Soit un couple de variables aléatoires (X, Y) de fonction de répartition conjointe $F_{X,Y}$ de classe \mathcal{C}^2 . Les fonctions f_X et f_Y définies par

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, t) dt \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(t, y) dt \quad (3.11)$$

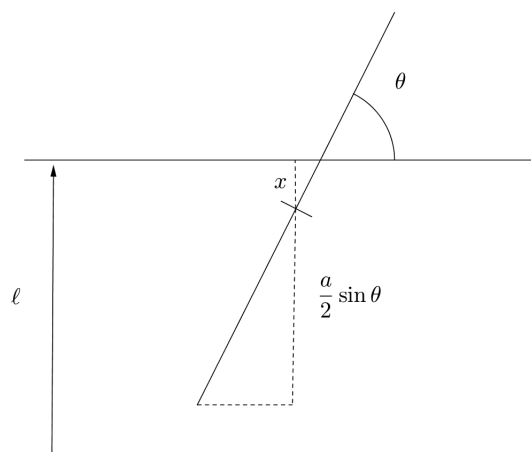
sont les **densités marginales** de probabilité de X et Y respectivement.

Cas particulier des variables aléatoires indépendantes.**Proposition 62.**

Deux variables aléatoires continues X et Y sont indépendantes si et seulement si pour presque tout couple (u, v) de réels, on a :

$$f_{X,Y}(u, v) = f_X(u) \cdot f_Y(v) \quad (3.12)$$

Exemple 3.1. L'aiguille de Buffon. Les lames d'un parquet sont disposées parallèlement et ont toutes une largeur égale à ℓ . On laisse tomber sur ce parquet une fine aiguille de longueur $a < \ell$ de façon aléatoire. On se propose de calculer la probabilité p pour que, après sa chute, l'aiguille coupe la ligne de séparation entre deux lames du parquet. Soit X la variable aléatoire représentant la distance du milieu de l'aiguille à la plus proche ligne de séparation des lames du parquet et Θ la variable aléatoire représentant l'angle aigu que fait la direction de l'aiguille et la ligne de séparation. Nous noterons x et θ les valeurs prises par X et Θ . On peut vérifier que x peut prendre toutes les valeurs comprises entre 0 et $\ell/2$ et que Θ peut prendre toutes les valeurs comprises entre 0 et $\pi/2$.



On émet l'hypothèse que ce sont deux variables aléatoires indépendantes, chacune de loi uniforme. X suit la loi continue uniforme de l'intervalle $[0, \ell/2]$ et Θ suit la loi continue uniforme de l'intervalle $[0, \pi/2]$. Autrement dit, les densités marginales de probabilités de X et Θ valent respectivement

$$\forall x \in [0, \ell/2], \quad f_X(x) = \frac{2}{\ell} \quad \text{et} \quad \forall \theta \in [0, \pi/2], \quad f_\Theta(\theta) = \frac{2}{\pi} \quad (3.13)$$

et 0 sinon. Comme les variables aléatoires X et Θ sont indépendantes, la densité de probabilité jointe est donnée par

$$\forall (x, \theta) \in [0, \ell/2] \times [0, \pi/2], \quad f_{X, \Theta}(x, \theta) = \frac{2}{\ell} \cdot \frac{2}{\pi} \quad (3.14)$$

Sur la figure, on remarque que l'aiguille touche effectivement la ligne de séparation des lames de parquet lorsque $x \geq \frac{a}{2} \sin \theta$. La probabilité de cet événement est donc

$$p = \frac{4}{\pi \cdot \ell} \int_0^{\pi/2} \int_0^{\frac{a}{2} \sin \theta} 1 \, dx \, d\theta = \frac{2a}{\pi \cdot \ell} \quad (3.15)$$

Avec la loi des grands nombres, on peut faire une approximation de l'expression ci-dessus avec la fréquence expérimentale des chutes après lesquelles l'aiguille coupe le ligne de séparation des lames du parquet, on obtient des valeurs approximatives de π . La convergence est très mauvaise.

3.1.2 Covariance et coefficient de corrélation

Les définitions introduites, dans le chapitre précédent, sur les moments d'une variable aléatoire peuvent être étendues à deux variables aléatoires X et Y (et plus) définies sur un même espace probabilisé $(\Omega, \mathcal{E}, \mathbf{P})$. Dans la pratique, on utilise seulement les moments du premier ordre et les moments centrés du second ordre.

Espérances

Les moments d'ordre un sont les espérances mathématiques des variables aléatoires X et Y , soit $\mathbf{E}(X)$ et $\mathbf{E}(Y)$.

Variances, covariances et coefficient de corrélation

Les moments centrés d'ordre deux sont:

- les variances des variables aléatoires X et Y , soit $\mathbf{Var}(X) = \sigma_X^2$ et $\mathbf{Var}(Y) = \sigma_Y^2$.
- le moment centré mixte d'ordre deux. Il joue un rôle particulièrement important: c'est la covariance des variables aléatoires X et Y .

Définition 63 (Covariance).

Pour un couple de variables aléatoires réelles, on définit la covariance du couple par

$$\text{cov}(X, Y) = \mathbf{E}((X - \mathbf{E}(X)) \cdot (Y - \mathbf{E}(Y))) \quad (3.16)$$

C'est "la moyenne du produit des écarts aux moyennes". On peut calculer la covariance par l'une des formules suivantes :

- si c'est un couple de variables discrètes

$$\text{cov}(X, Y) = \sum_{i,j} p_{i,j} (x_i - \mathbf{E}(X)) \cdot (y_j - \mathbf{E}(Y)) \quad (3.17)$$

- si c'est un couple de variables continues

$$\text{cov}(X, Y) = \int_{\mathbb{R}^2} f_{X,Y}(u, v) (u - \mathbf{E}(X)) \cdot (v - \mathbf{E}(Y)) \, du \, dv \quad (3.18)$$

Remarque 3.2. On note aussi $\sigma_{X,Y}$ la covariance de X et Y . Notons aussi que

$$\text{cov}(X, X) = \mathbf{Var}(X)$$

Pour toutes variables aléatoires X et Y , on a les propriétés :

Proposition 64.

- $\text{cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X) \cdot \mathbf{E}(Y)$,
- $\mathbf{Var}(X \pm Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) \pm 2 \text{cov}(X, Y)$,
- $\text{cov}(X, Y) = 0$ si X et Y sont indépendants,
- $|\text{cov}(X, Y)| \leq \sigma_X \times \sigma_Y$. C'est l'inégalité de CAUCHY-SCHWARZ.

Preuve. La première propriété se démontre de la manière suivante en utilisant la linéarité de l'espérance

$$\begin{aligned} \text{cov}(X, Y) &= \mathbf{E}((X - \mathbf{E}(X)) \cdot (Y - \mathbf{E}(Y))) \\ &= \mathbf{E}(XY - \mathbf{E}(X)Y - \mathbf{E}(Y)X + \mathbf{E}(X) \cdot \mathbf{E}(Y)) \\ &= \mathbf{E}(XY) - \mathbf{E}(X) \cdot \mathbf{E}(Y) - \mathbf{E}(Y) \cdot \mathbf{E}(X) + \mathbf{E}(X) \cdot \mathbf{E}(Y) \\ \text{cov}(X, Y) &= \mathbf{E}(XY) - \mathbf{E}(X) \cdot \mathbf{E}(Y) \end{aligned}$$

La deuxième propriété de la covariance résulte des définitions de la variance et de la covariance. Pour la troisième, il suffit de prouver qu si X et Y sont indépendantes, alors

$$\mathbf{E}(XY) = \mathbf{E}(X) \cdot \mathbf{E}(Y)$$

Ce fait résulte des propositions 58 et 62 respectivement dans le cas discret et continu. En effet,

$$\begin{aligned} \mathbf{E}(XY) &= \sum_{i,j} p_{ij} x_i y_j &= \sum_{i,j} p_i \cdot p_j x_i y_j \\ &= \left(\sum_i p_i x_i \right) \cdot \left(\sum_j p_j y_j \right) &= \mathbf{E}(X) \cdot \mathbf{E}(Y) \end{aligned}$$

Le cas continu est similaire, le regroupement étant licite par le théorème de FUBINI. La quatrième propriété provient du fait qu'on travaille, en réalité, avec un produit scalaire dans un espace vectoriel. La covariance définit un produit scalaire sur l'espace vectoriel réel des variables aléatoires centrées d'ordre 2 (c'est-à-dire admettant une variance). ■

Remarque 3.3. La première égalité donne un deuxième moyen de calculer la covariance. Il faut calculer $\mathbf{E}(XY)$. Ce qui se fait avec l'une des deux formules

- si c'est un couple de variables discrètes

$$\mathbf{E}(XY) = \sum_{i,j} p_{i,j} x_i y_j \tag{3.19}$$

- si c'est un couple de variables continues

$$\mathbf{E}(XY) = \int_{\mathbb{R}^2} uv f_{X,Y}(u, v) du dv \tag{3.20}$$

Puisque la covariance est nulle si les variables sont indépendantes. La covariance est un indicateur du fait que les variables sont liées ou non.

Définition 65 (Coefficient de corrélation).

Pour caractériser seulement la liaison des variables X et Y on utilise le **coefficient de corrélation** défini par

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y} \quad (3.21)$$

C'est un nombre sans dimension, compris entre -1 et 1. C'est le "cosinus" de l'angle obtenu par les deux vecteurs formés par les variables aléatoires centrées $X - \mathbf{E}(X)$ et $Y - \mathbf{E}(Y)$ pour le produit scalaire défini ci-dessus. En utilisant les résultats connus des espaces préhilbertiens réels, on sait que, si $\rho_{XY} = \pm 1$, les "vecteurs" sont liés, l'un est multiple de l'autre. Il existe un lien linéaire entre les variables centrées $X - \mathbf{E}(X)$ et $Y - \mathbf{E}(Y)$, donc un lien affine entre les variables X et Y : $Y = aX + b$. Si $\rho_{XY} = 0$, les vecteurs sont orthogonaux. On dit que les variables sont **décorrélées**. On a donc la propriété suivante.

Proposition 66.

Si deux variables X et Y indépendantes, alors elles sont **décorrélées**

$$\rho_{XY} = 0$$

mais **attention**, la réciproque est fautive.

Preuve. On sait que deux variables indépendantes vérifient

$$\mathbf{E}(X \times Y) = \mathbf{E}(X) \times \mathbf{E}(Y) \quad \Rightarrow \quad \rho_{X,Y} = 0$$

Donnons un contre-exemple à une éventuelle réciproque. Considérons un système formé par deux variables aléatoires continues X et Y avec $X \sim \mathcal{N}(0; 1)$ et $Y = X^2$. On a bien

$$\sigma_{X,Y} = \mathbf{E}(XY) - \mathbf{E}(X) \cdot \mathbf{E}(Y) = 0 \quad \Rightarrow \quad \rho_{X,Y} = 0$$

Les variables aléatoires sont **décorrélées** bien qu'elles soient dépendantes. ■

Cas d'un vecteur gaussien

On dit qu'un couple ou vecteur (X, Y) est gaussien si pour tout couple $(\alpha, \beta) \in \mathbb{R}^2 \setminus (0, 0)$, $\alpha X + \beta Y$ suit une loi normale.

On démontre que pour un vecteur gaussien la réciproque de la propriété précédente est vraie. C'est-à-dire

Proposition 67.

Si (X, Y) est un vecteur gaussien tel que $\rho_{X,Y} = 0$, alors X et Y sont indépendantes.

3.2 Loys conditionnelles

Dans ce paragraphe, pour alléger les notations, les fonctions $f_{X,Y}$ et $F_{X,Y}$ seront notées f et F , si il n'y a pas d'ambiguïté. Si deux variables X et Y sont dépendantes, on peut considérer que la probabilité pour que " $Y \leq y$ " est conditionnée par le fait que x se trouve dans l'intervalle $]a, b]$. Ceci s'exprime au moyen de la formule des probabilités conditionnelles

$$\mathbf{P}(Y \leq y \mid a < X \leq b) = \frac{\mathbf{P}(a < X \leq b, Y \leq y)}{\mathbf{P}(a < X \leq b)}$$

d'où

$$\mathbf{P}(Y \leq y \mid a < X \leq b) = \frac{\int_a^b \int_{-\infty}^y f(u, v) \, du \, dv}{\int_a^b \int_{\mathbb{R}} f(u, v) \, du \, dv}$$

Si l'on fait tendre a et b vers x , on obtient alors (quand la limite existe)

$$F(y \mid x) = \lim_{a, b \rightarrow x} \mathbf{P}(Y \leq y \mid a < X \leq b) = \frac{\int_{-\infty}^y f(x, t) \, dt}{\int_{\mathbb{R}} f(x, t) \, dt} = \frac{1}{f_X(x)} \int_{-\infty}^y f(x, t) \, dt \quad (3.22)$$

$F(y \mid x)$ définit la **fonction de répartition conditionnelle** de la variable aléatoire y pour $X = x$. Si cette fonction admet une dérivée par rapport à y ,

$$f(y \mid x) = \frac{\partial F(y \mid x)}{\partial y} \quad (3.23)$$

$f(y \mid x)$ définit la **densité de probabilité conditionnelle** de la variable aléatoire Y pour $X = x$. Les densités de probabilité conditionnelles ont la même propriété que les densités de probabilité unidimensionnelles, à savoir

$$\int_{\mathbb{R}} f(y \mid x) \, dy = 1 \quad (3.24)$$

En dérivant $F(y \mid x)$ par rapport à y , la formule (3.23) donne

$$f(y \mid x) = \frac{f(x, y)}{f_X(x)} \Rightarrow f(x, y) = f_X(x) \cdot f(y \mid x) \quad (3.25)$$

On obtient alors

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_{\mathbb{R}} f_X(x) \cdot f(y \mid x) \, dx \quad (3.26)$$

Cette formule permet d'exprimer la densité de probabilité marginale de Y en fonction de la densité de probabilité marginale de X et de la densité de probabilité conditionnelle de Y lorsque X est donné. On définit de même la répartition conditionnelle de la variable aléatoire X pour une valeur donnée de Y . Ainsi la densité de probabilité conditionnelle de X pour $Y = y$ est donnée par

$$f(x \mid y) = \frac{f(x, y)}{f_Y(y)} \quad (3.27)$$

Si les variables aléatoires X et Y sont indépendantes, on a $f(x, y) = f_X(x) \cdot f_Y(y)$. Comme le montrent les formules (3.25) et (3.27), la densité de probabilité conditionnelle coïncide avec la densité de probabilité marginale de la variable aléatoire.

En utilisant (3.25), (3.26) et (3.27), on peut aisément établir une relation entre les densités de probabilité marginale et conditionnelle des variables aléatoires

$$f(x \mid y) = \frac{f_X(x) \cdot f(y \mid x)}{\int_{\mathbb{R}} f_X(x) \cdot f(y \mid x) \, dx} \quad (3.28)$$

Annexe A

A.1 Rappels sur le dénombrements

- Nombre de permutations d'un ensemble à n éléments: $n!$
- Nombre de p -uplets d'un ensemble à n éléments: n^p
- Nombre de p -arrangements d'un ensemble à n éléments: $A_n^p := \frac{n!}{(n-p)!}$
- Nombre de parties d'un ensemble à n éléments: 2^n
- Nombre de parties à k éléments d'un ensemble à n éléments: $\binom{n}{k} := \frac{n!}{k!(n-k)!}$
- Rappelons la formule du binôme de Newton: $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$
- Et la formule de Stirling: $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$

A.2 Lois discrètes classiques

Nom	Notation	Ω	Loi	Espérance	Variance
Binomiale	$\mathcal{B}(n; p)$	$\llbracket 0, n \rrbracket$	$\binom{n}{k} p^k q^{n-k}$	np	npq
Bernoulli	$\mathcal{B}(1; p)$	$\{0; 1\}$	$\begin{cases} \mathbf{P}(X = 0) = q \\ \mathbf{P}(X = 1) = p \end{cases}$	p	pq
Poisson	$\mathcal{P}(\lambda)$	$\llbracket 0, n \rrbracket$	$e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ
Géométrique	$\mathcal{G}(p)$	\mathbb{N}^*	pq^{k-1}	p^{-1}	$q \cdot p^{-2}$
Hypergéométrique	$\mathcal{H}(N, n, p)$	$\llbracket 0, N \rrbracket$	$\frac{\binom{Np}{k} \cdot \binom{Nq}{n-k}}{\binom{N}{n}}$	np	$npq \frac{N-n}{N-1}$
Uniforme	$\mathcal{U}_{\llbracket 1, n \rrbracket}$	$\llbracket 1, n \rrbracket$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Pascal	$\mathcal{P}(r, p)$	$\llbracket r, +\infty \llbracket$	$\binom{k-1}{r-1} p^r q^{k-r}$	$r \cdot p^{-1}$	$rq \cdot p^{-2}$
Binomiale négative	$\mathcal{J}(r, p)$	\mathbb{N}	$\binom{k+r-1}{k} p^r q^k$	$r \cdot p^{-1}$	$rq \cdot p^{-2}$

A.3 Lois continues classiques

Nom et Notation	Ω	Densité	Espérance	Variance
Uniforme $\mathcal{U}_{[a,b]}$	$[a, b]$	$\frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Exponentielle $\mathcal{E}(\lambda)$	\mathbb{R}^+	$\lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_*^+}(x)$	λ^{-1}	λ^{-2}
Normale $\mathcal{N}(\mu, \sigma)$	\mathbb{R}	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2
Log-Normale $\mathcal{LN}(\mu, \sigma)$	\mathbb{R}_*^+	$\frac{1}{\sigma\sqrt{2\pi x}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \mathbf{1}_{\mathbb{R}_*^+}(x)$	$e^{\mu+\sigma^2/2}$	$(e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$
Gamma $\Gamma(\alpha, \beta)$	\mathbb{R}_*^+	$x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} \mathbf{1}_{\mathbb{R}_*^+}(x)$	$\alpha\beta^{-1}$	$\alpha\beta^{-2}$
Cauchy $\mathcal{C}(a)$	\mathbb{R}	$\frac{a}{\pi^2(a^2 + x^2)}$	\times	\times
Pareto	$[a, +\infty[$	$\frac{\alpha}{a} \left(\frac{a}{x}\right)^{\alpha+1} \mathbf{1}_{[a, +\infty[}(x)$	$\frac{\alpha a}{\alpha - 1}$ si $\alpha > 1$	$\frac{\alpha a^2}{(\alpha - 1)^2(\alpha - 2)}$ si $\alpha > 2$
Rayleigh	\mathbb{R}^+	$x \exp\left(-\frac{x^2}{2}\right)$	$\sqrt{\frac{\pi}{2}}$	$2 - \frac{\pi}{2}$
χ_2	\mathbb{R}^+	$\frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{\frac{k}{2}-1} e^{-x/2}$	k	$2k$

A.4 Tables de probabilité et statistiques

A.4.1 La loi de POISSON

Probabilité et fonction de répartition de la loi de POISSON $\mathcal{P}(\lambda)$.

Soit X une variable aléatoire réelle suivant la loi $\mathcal{P}(a)$. On pose

$$p_a(k) = \mathbf{P}(X = k) \quad \text{et} \quad F_a(k) = \mathbf{P}(X \leq k)$$

k	$a = 1$		$a = 2$		$a = 3$		$a = 4$		$a = 5$	
	$p_a(k)$	$F_a(k)$	$p_a(k)$	$F_a(k)$	$p_a(k)$	$F_a(k)$	$p_a(k)$	$F_a(k)$	$p_a(k)$	$F_a(k)$
0	0,36788	0,36788	0,13534	0,13534	0,04979	0,04979	0,01832	0,01832	0,00674	0,00674
1	0,36788	0,73576	0,27067	0,40601	0,14936	0,19915	0,07326	0,09158	0,03369	0,04043
2	0,18394	0,91970	0,27067	0,67668	0,22404	0,42319	0,14653	0,23810	0,08422	0,12465
3	0,06131	0,98101	0,18045	0,85712	0,22404	0,64723	0,19537	0,43347	0,14037	0,26503
4	0,01533	0,99634	0,09022	0,94735	0,16803	0,81526	0,19537	0,62884	0,17547	0,44049
5	0,00307	0,99941	0,03609	0,98344	0,10082	0,91608	0,15629	0,78513	0,17547	0,61596
6	0,00051	0,99992	0,01203	0,99547	0,05041	0,96679	0,10420	0,88933	0,14622	0,76218
7	0,00007	0,99999	0,00344	0,99890	0,02160	0,98810	0,05954	0,94887	0,10444	0,86663
8	0,00001	1,00000	0,00086	0,99976	0,00810	0,99620	0,02977	0,97864	0,06528	0,93191
9			0,00019	0,99995	0,00270	0,99890	0,01323	0,99187	0,03627	0,96817
10			0,00004	0,99999	0,00081	0,99971	0,00529	0,99716	0,01813	0,98630
11			0,00001	1,00000	0,00022	0,99993	0,00192	0,99908	0,00824	0,99455
12					0,00006	0,99998	0,00064	0,99973	0,00343	0,99798
13					0,00002	1,00000	0,00020	0,99992	0,0132	0,99930
14							0,00006	0,99998	0,00047	0,99977
15							0,00002	1,00000	0,00016	0,99993
16									0,00005	0,99998
17									0,00001	0,99999
18									0,00001	1,00000

Quantiles de la loi normale centrée réduite $\mathcal{N}(0; 1)$.

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,5	0,0000	0,0251	0,0502	0,0753	0,1004	0,1257	0,1510	0,1764	0,2019	0,2275
0,6	0,2533	0,2793	0,3055	0,3319	0,3585	0,3853	0,4125	0,4399	0,4677	0,4959
0,7	0,5244	0,5534	0,5828	0,6128	0,6433	0,6745	0,7063	0,7388	0,7722	0,8064
0,8	0,8416	0,8779	0,9154	0,9542	0,9945	1,0364	1,0803	1,1264	1,1750	1,2265
0,9	1,2816	1,3408	1,4051	1,4758	1,5548	1,6449	1,7507	1,8808	2,0537	2,3263

α	0,990	0,991	0,992	0,993	0,994	0,995	0,996	0,997	0,998	0,999
$\Phi^{-1}(\alpha)$	2,3263	2,3656	2,4089	2,4573	2,5121	2,5758	2,6521	2,7478	2,8782	3,0902

α	0,9990	0,9991	0,9992	0,9993	0,9994	0,9995	0,9996	0,9997	0,9998	0,9999
$\Phi^{-1}(\alpha)$	3,0902	3,1214	3,1559	3,1947	3,2389	3,2905	3,3528	3,4316	3,5401	3,7190

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,5758	2,3263	2,1701	2,0537	1,9600	1,8808	1,8119	1,7507	1,6954
0,1	1,6449	1,5982	1,5548	1,5141	1,4758	1,4395	1,4051	1,3722	1,3408	1,3106
0,2	1,2816	1,2536	1,2265	1,2004	1,1750	1,1503	1,1264	1,1031	1,0803	1,0581
0,3	1,0364	1,0152	0,9945	0,9741	0,9542	0,9346	0,9154	0,8965	0,8779	0,8596
0,4	0,8416	0,8239	0,8064	0,7892	0,7722	0,7554	0,7388	0,7225	0,7063	0,6903
0,5	0,6745	0,6588	0,6433	0,6280	0,6128	0,5978	0,5828	0,5681	0,5534	0,5388
0,6	0,5244	0,5101	0,4959	0,4817	0,4677	0,4538	0,4399	0,4261	0,4125	0,3989
0,7	0,3853	0,3719	0,3585	0,3451	0,3319	0,3186	0,3055	0,2924	0,2793	0,2663
0,8	0,2533	0,2404	0,2275	0,2147	0,2019	0,1891	0,1764	0,1637	0,1510	0,1383
0,9	0,1257	0,1130	0,1004	0,0878	0,0753	0,0627	0,0502	0,0376	0,0251	0,0125

α	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$z_{1-\alpha/2}$	3,2905	3,8906	4,4172	4,8916	5,3267	5,7307	6,1094

Annexe B

Compléments en probabilité

Cette partie donne quelques définitions et exemples d'outils importants en probabilité.

B.1 Lemme de BOREL-CANTELLI

Soit $(A_n)_{n \in \mathbb{N}}$ une suite d'événements, on pose $\limsup_{n \in \mathbb{N}} A_n = \bigcap_{n \geq 0} \left(\bigcup_{k \geq n} A_k \right)$.

Théorème 68 (De Borel-Cantelli).

- Si $\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) < +\infty$ alors $\mathbf{P}(\limsup_{n \in \mathbb{N}} A_n) = 0$.
- Si les événements sont indépendants alors $\sum_{n \in \mathbb{N}} \mathbf{P}(A_n) < +\infty$ implique

$$\mathbf{P}(\limsup_{n \in \mathbb{N}} A_n) = 1$$

Exemple B.1. Le singe dactylographe. Supposons qu'un singe, placé devant une machine à écrire, tape au hasard sur toutes touches. Supposons que ce singe appuie sur chaque touche avec une fréquence égale. Alors, tôt ou tard, il tapera n'importe quel texte choisi à l'avance.

“ Parce que le hasard, il a bon dos. Prenons un des singes, là. Objectivement, quelles sont ses chances de sortir un livre de 200 pages au hasard ? Mettons que ça fasse une combinaison de 400 000 signes. 100 caractères sur un clavier, en général. Donc ça fera 100 puissance 400 000 possibilités. Un “ 1 ” avec 800 000 zéros derrière. En gros, si un milliard d'ordinateurs essayaient chacun un milliard de combinaisons par seconde depuis le big bang, ils auraient à peine eu le temps de faire suffisamment d'essais pour tomber sur le titre. Tous ces efforts pour 200 pages. Le livre a intérêt à être bon. La haine si tu tombes sur du DAN BROWN.” Boulet, Notes.

B.2 Fonctions génératrices

Définition 69 (Fonction génératrice).

La fonction génératrice G_X d'une variable aléatoire réelle discrète X est donnée par

$$G_X(t) = \mathbf{E}(t^X) = \sum_{k=0}^{+\infty} t^k p_k$$

G_X est bien définie pour tout $t \in [-1; 1]$ et elle caractérise la loi de X . Si X et Y sont indépendantes alors

$$G_{X+Y} = G_X G_Y$$

B.3 Espérances conditionnelles et martingales

On se place dans le cas général d'un espace de probabilité $(\Omega, \mathcal{E}, \mathbf{P})$. Soit $\mathcal{F} \subset \mathcal{E}$ une sous-tribu, ainsi qu'une variable aléatoire intégrable X . Alors il existe une variable aléatoire Z , \mathcal{F} -mesurable et LEBESGUE-intégrable, telle que, pour toute variable aléatoire U bornée et \mathcal{F} -mesurable,

$$\mathbf{E}(XU) = \mathbf{E}(ZU)$$

On note alors $Z = \mathbf{E}(X|\mathcal{F})$. De plus, si Y est une seconde v.a engendrant la tribu $\sigma(Y)$, on note

$$\mathbf{E}(X|Y) = \mathbf{E}(X|\sigma(Y))$$

Définition 70 (Martingale).

Soit $(\Omega, \mathcal{E}, \mathbf{P})$ un espace probabilisé. Considérons $(X_n)_n$ une suite de v.a.r et $(\mathcal{E}_n)_n$ une suite croissante de tribus inclus dans \mathcal{E} (on parle de filtration). On dit que $(X_n)_n$ est une martingale si

- $\mathbf{E}(X_n) < +\infty$ et X_n est \mathcal{E}_n -mesurable pour tout n .
- On a pour tout n , $\mathbf{E}(X_{n+1}|\mathcal{E}_n) = X_n$.

B.4 Chaînes de MARKOV

Une chaîne de MARKOV est une suite X_0, X_1, \dots de variables aléatoires à valeurs dans un ensemble E (dit ensemble des états) où l'avenir ne dépend que du présent. Autrement dit, pour tous états $x_0, \dots, x_{n+1} \in E$,

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

On rajoute souvent l'hypothèse supplémentaire que la chaîne est indépendante du temps, c'est-à-dire que les probabilités conditionnelles ne dépendent pas de n ,

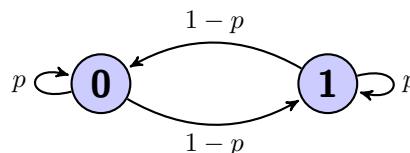
$$\mathbf{P}(X_{n+1} = j | X_n = i) = p_{ij}$$

On parle de chaînes de MARKOV homogènes. La matrice $\mathbf{P} = (p_{ij})_{i,j}$ s'appelle la matrice de transition.

Exemple B.2. • Un canal de transmission transmet des bits avec erreur selon le modèle suivant: il transmet fidèlement un bit avec probabilité p et de façon erronée avec probabilité $1 - p$ avec $p \in [0, 1]$. On considère n canaux en série, et que chaque canal fonctionne indépendamment des autres. On note X_k le bit reçu en sortie du k -ième canal et X_0 le bit à l'entrée du premier canal. On désire calculer la probabilité qu'au bout des n canaux, le signal reste inchangé. Cette situation se modélise par une chaîne de MARKOV à 2 états:

$$\mathbf{P}(X_{k+1} = 1 | X_k = 0) = 1 - p \quad \text{et} \quad \mathbf{P}(X_{k+1} = 1 | X_k = 1) = p$$

et $\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \in \mathcal{M}_2(\mathbb{R})$. Il est commode de modéliser cela par un graphe



Le calcul des puissances de la matrice \mathbf{P} donne la probabilité qu'un bit soit fidèlement transmis au bout de n canaux

$$\mathbf{P}(X_n = X_0) = \frac{1 + (2p - 1)^n}{2} \xrightarrow{n \rightarrow +\infty} 1/2$$

À la limite, on ne transmet rien...

- L'algorithme PageRank utilisé par GOOGLE donnant un indice de popularité d'une page Web.

- L'exemple proposé par d'Andrei Andreevich MARKOV, en 1913. On considère une suite de 20 000 caractères pris dans *Eugène Onegin* d'Alexandre POUCHKINE, et on regarde la succession entre les voyelles et les consonnes. En russe, ce dernier avait obtenu la matrice de transition suivante

$$P = \begin{pmatrix} 12,8 & 87,2 \\ 66,3 & 33,7 \end{pmatrix}$$

Par exemple, la probabilité qu'une voyelle soit suivie d'une consonne est de 87,2%.

B.5 Un exemple d'inégalité de concentration

Théorème 71 (Inégalité de Hoeffding - 1963).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires indépendantes à valeurs dans $[a, b]$, alors pour tout $\varepsilon > 0$ avec $\bar{X}_n = \sum_{i=1}^n X_i$

$$\mathbf{P}(\bar{X}_n \geq \mathbf{E}(\bar{X}_n) + \varepsilon) \leq \exp\left(-2 \frac{n\varepsilon^2}{(b-a)^2}\right) \tag{B.1}$$

B.6 Entropie (de SHANNON)

Définition 72 (Entropie).

Soient X et Y deux variables aléatoires discrètes finies. On définit

- l'entropie de X par $H(X) = - \sum_{i=1}^n p_i \log_2 p_i$.
- l'entropie conjointe des variables X et Y

$$H(X, Y) = - \sum_{i,j} \mathbf{P}(X = x_i, Y = y_j) \log_2 \mathbf{P}(X = x_i, Y = y_j)$$

L'entropie dépend que de la loi de X. On vérifie que si X_c et X_{eq} suivent respectivement une loi certaine et une loi équiprobable alors $0 = H(X_c) \leq H(X) \leq H(X_{eq})$.

Interprétation en théorie de l'information. Supposons qu'un receveur ignore ce que l'émetteur va lui envoyer. Toutefois, le receveur connaît la probabilité de chaque message possible. Par exemple, le receveur connaît chaque p_i , la probabilité d'un mot x_i pris dans un dictionnaire $\chi = \{x_1, \dots, x_n\}$. Alors $H(p) = - \sum_{i=1}^n p_i \log_2 p_i$ mesure en un certain sens l'incertitude lors de l'envoi du message. La relation précédente confirme cette intuition, l'entropie est nulle pour un message certain alors qu'elle est maximale pour le cas d'équiprobabilité. On vérifie aussi que

$$H(X, Y) \leq H(X) + H(Y)$$

Si X et Y sont indépendants, on a égalité. Citons aussi une phrase de John VON NEUMANN à Claude SHANNON en 1949 justifiant le terme d'entropie: *“La théorie est excellente mais elle a besoin d'un bon nom pour “information perdue”. Pourquoi ne l'appelles-tu pas entropie? Premièrement, un développement mathématique ressemblant fort au tien existe déjà dans la mécanique statistique de BOLTZMANN, et deuxièmement, personne ne comprend vraiment bien l'entropie, donc dans une discussion tu te trouverais dans une position avantageuse.”*

