

Probabilités I

STEP, MINES ParisTech*

11 janvier 2020 (#3c3e161)

Table des matières

Introduction	2
Exemple de la modélisation du climat	3
Plan du cours	4
Historique	4
Probabilités des événements	5
Phénomènes aléatoires et événements	5
Exemples	5
Définition – L’espace fondamental	6
Exemples	6
Définition – Événement	6
Exemples	6
Correspondance entre opérations logiques et ensemblistes	7
Définition – Tribu	7
Exemples	7
Notion de densité de probabilité	8
Remarque	9
Probabilité	9
Définition – Probabilité	9
Définition – Propriété presque-sûre	9
Proposition – Propriétés élémentaires	9
Théorème de la continuité monotone	10
Remarque	10
Proposition	10
Probabilité conditionnelle	10
Exemple	11
Définition – Probabilité conditionnelle	11
Proposition	11
Formule des probabilités totales	11
Formule de Bayes	12

*Ce document est un des produits du projet  boisgera/CDIS, initié par la collaboration de (S)ébastien Boisgérault (CAOR), (T)homas Romary et (E)milie Chautru (GEOSCIENCES), (P)auline Bernard (CAS), avec la contribution de Gabriel Stoltz (Ecole des Ponts ParisTech, CERMICS). Il est mis à disposition selon les termes de la licence Creative Commons “attribution – pas d’utilisation commerciale – partage dans les mêmes conditions” 4.0 internationale.

Remarque	12
Exemple	12
Indépendance des événements	13
Définition – Indépendance de deux événements	13
Remarques	13
Proposition	14
Exemple	14
Remarque : réflexion sur le concept de probabilité	14
Probabilité sur \mathbb{R}	15
Fonction de répartition	15
Définition – fonction de répartition	15
Théorème	15
Théorème	16
Remarque	16
Remarque	16
Exemple	17
Densités de probabilités	19
Définition – densité de probabilité	19
Remarques	19
Exemples de lois à densités	20
Exercices	24
Propriétés élémentaires	24
Continuité monotone	26
Une définition alternative de la probabilité	26
Généralisation de l'inégalité de Boole	26
Indépendance et conditionnement	27
Problème de Monty Hall	27
Queues de distributions	27
Densité et fonction de répartition d'une loi Normale	29
Remarque	30
Solutions	30
Propriétés élémentaires	30
Continuité monotone	31
Une définition alternative de la probabilité	31
Généralisation de l'inégalité de Boole	31
Indépendance et conditionnement	32
Problème de Monty Hall	32
Queues de distributions	33
Densité et fonction de répartition d'une loi Normale	36
Références	38

Introduction

Le but de ce cours est de consolider et compléter les connaissances en théorie des probabilités acquises en CPGE mais surtout de permettre d'acquérir le

raisonnement probabiliste. En effet, les probabilités peuvent être vues comme un outil de modélisation de phénomènes qui ont la caractéristique d’être aléatoires. L’aléatoire peut intervenir de différentes manières dans ces phénomènes.

- Dans les cas d’école que sont les jeux de pile ou face ou de lancés de dés, la différence entre les résultats, si l’on réitère l’expérience, peut être liée à l’impulsion initiale communiquée au dé et à d’autres facteurs environnementaux comme le vent, la rugosité de la table, etc. Le hasard intervient du fait de la méconnaissance des conditions initiales, car la pièce ou le dé ont des trajectoires parfaitement définies par la mécanique classique.
- Dans beaucoup de cas de figure, on fait intervenir l’aléatoire dans la modélisation du fait d’une connaissance incomplète des phénomènes. On parle alors de modélisation de l’incertitude. C’est le cas par exemple en sciences du climat, dont nous discutons ci-dessous.
- Dans certains domaines, tels la physique quantique, l’aléatoire fait intrinsèquement partie de la théorie.

Elles sont aussi un préalable indispensable pour aborder l’analyse statistique des données et les méthodes d’apprentissage automatique.

En CPGE, les probabilités ont été vues dans le cadre de phénomènes aléatoires qui admettent un nombre au plus dénombrable de résultats possibles. Ce cadre restreint est supposé connu. On pourra se reporter au chapitre 3 du cours de Garnier (2018) ou aux deux premiers chapitres du cours de Jourdain (2016) pour une éventuelle mise à niveau.

Exemple de la modélisation du climat

Les prédictions de la météo et les projections du climat proviennent généralement de modèles numériques qui simulent les différents processus physiques à l’œuvre. Les incertitudes dans la construction et l’application de ces modèles sont variées et peuvent être réparties en quatre groupes : les conditions initiales (on ne connaît jamais parfaitement l’ensemble des variables climatiques en tout point du globe), les conditions aux limites (par exemple lorsqu’on travaille à l’échelle d’un continent ou d’un pays), les valeurs des paramètres intervenant dans les modèles (constantes issues d’observations diverses), les incertitudes structurelles enfin qui relèvent des choix de modélisation. Pour tenir compte de ces incertitudes, les climatologues effectuent des ensembles de simulations, où les différentes quantités incertaines sont échantillonnées selon des modèles probabilistes (voir en particulier Tebaldi and Knutti (2007)).

L’incertitude sur les conditions initiales est particulièrement influente aux faibles échelles de temps. La météo est un système chaotique, les prévisions sont extrêmement sensibles aux variations des conditions initiales utilisées pour initialiser les modèles. Ces dernières sont en revanche connues de manière imparfaite. Il est ainsi nécessaire de tenir compte de cette incertitude. Ceci est rendu possible par la modélisation probabiliste. Les conditions aux limites font intervenir des **variables continues** (température, pression, vitesse du vent, etc.), ce qui représente la principale nouveauté par rapport au programme de CPGE, et il est nécessaire de caractériser leurs relations de **dépendance**. Il convient par

ailleurs de tenir compte des observations (satellites, station de mesure, ...) en incluant cette information dans la modélisation probabiliste. On parle de **conditionnement** aux données. Il s'agit enfin d'en générer les valeurs via des algorithmes de **simulation stochastique**. La validité de l'approche est assurée par les **théorèmes limites**, qui garantissent la représentativité des ensembles générés.

Plan du cours

Le cours est organisé en 5 amphis et abordera consécutivement les notions évoquées ci-dessus, à savoir les probabilités définies sur \mathbb{R} , les variables et vecteurs aléatoires réels, l'indépendance et le conditionnement de variables aléatoires, l'étude des suites de variables aléatoires et, enfin, les méthodes de simulation stochastique.

Historique

Avant que l'étude des probabilités soit considérée comme une science, l'observation du hasard dans les événements naturels a amené les philosophes et les scientifiques à réfléchir sur la notion de liens entre événements, causes et conséquences, et lois de la nature. Les jeux de hasard, les situations météorologiques ou les trajectoires des astres ont fait partie des domaines étudiés. Les explications données sont alors liées au destin, à une colère céleste ou à une présence divine.

Il est communément admis que le début de la science des probabilités se situe au XVI^e siècle avec l'analyse de jeux de hasard par Jérôme Cardan et au XVII^e siècle avec les discussions entre Pierre de Fermat et Blaise Pascal au sujet de paradoxes issus de ces jeux, notamment posés par Antoine Gombaud, chevalier de Méré. Cette nouvelle théorie est nommée géométrie aléatoire par le chevalier de Méré en 1654, elle est appelée par la suite calcul conjectural, arithmétique politique et plus communément aujourd'hui théorie des probabilités. Cette théorie, dite des probabilités modernes, est alors étudiée par de nombreux penseurs jusqu'au XIX^e siècle : Kepler, Galilée, Leibniz, Huygens, Halley, Buffon, les frères Bernoulli, Moivre, Euler, D'Alembert, Condorcet, Laplace, Fourier. Elle est principalement basée sur les événements discrets et la combinatoire.

Des considérations analytiques ont forcé l'introduction de variables aléatoires continues dans la théorie. Cette idée prend tout son essor dans la théorie moderne des probabilités, dont les fondations ont été posées par Andreï Nikolaevich Kolmogorov. Kolmogorov combina la notion d'univers, introduite par Richard von Mises et la théorie de la mesure pour présenter son système d'axiomes pour la théorie des probabilités en 1933. Très vite, son approche devint la base incontestée des probabilités modernes.

Le XX^e siècle voit également le développement de l'application de la théorie des probabilités dans plusieurs sciences.

Avec la mécanique newtonienne, la théorie du champ électromagnétique ou la thermodynamique, la physique classique est la théorie utilisée jusqu'à la fin du XIX^e siècle. En 1925, Erwin Schrödinger étudie l'équation qui détermine

l'évolution d'une onde au cours du temps : l'équation de Schrödinger. Max Born utilise cette équation pour décrire une collision entre des particules telles que des électrons ou des atomes. Les observations de ces expériences l'amènent à supposer que le module de la fonction d'onde est la probabilité que la particule soit détectée en un point de l'espace. C'est le début d'une nouvelle approche de la physique quantique.

En 1900, Louis Bachelier fut un des premiers mathématiciens à modéliser les variations de prix boursiers grâce à des variables aléatoires. « le marché n'obéit qu'à une seule loi : la loi du hasard ». Bachelier utilise alors le calcul stochastique pour étudier les variations boursières au cours du temps. En 1970, Fischer Black et Myron Scholes reprennent les idées de Bachelier pour modéliser les rendements d'une action.

L'utilisation des probabilités en biologie a pris un essor dans les années 1970, notamment dans l'étude de l'évolution des espèces. La reproduction des individus est modélisée par un choix aléatoire des gènes transmis ainsi que des mutations apparaissant de manière aléatoire sur les individus. L'extinction des espèces ou des gènes est alors étudiée en fonction des effets stochastiques.

De nos jours, l'Ecole française de Probabilités est très active. La première Médaille Fields décernée à un probabiliste a été attribuée à Wendelin Werner en 2006. Les probabilités se développent de plus en plus, alimentées en particulier de manière essentielle par la physique, le développement des réseaux de télécommunications, la finance, la biologie, la médecine. . . Elles permettent de construire des modèles mathématiques, qui peuvent être validés par les données suivant la théorie statistique, et fournissent également des possibilités d'expérimentations fictives dans de multiples domaines d'applications.

Probabilités des événements

Phénomènes aléatoires et événements

L'objet de la théorie des probabilités est l'analyse mathématique de phénomènes dans lesquels le hasard intervient. Les phénomènes aléatoires résultent d'expériences dont le résultat ne peut être prédit à l'avance et qui **peut** varier si on répète l'expérience dans des conditions identiques.

Il est aisé de trouver des exemples de tels phénomènes.

Exemples

1. Jeu de Pile ou Face
2. Lancé de dés
3. Durée de vie d'une ampoule électrique
4. Température demain à 12h au sommet de la tour Eiffel
5. Evolution de la vitesse d'une molécule dans un gaz raréfié sur un intervalle de temps $[t_1, t_2]$

La théorie des probabilités vise à fournir un modèle mathématique pour décrire ces phénomènes. Elle repose sur trois ingrédients essentiels dont on donne ici les définitions.

Définition – L’espace fondamental

Noté habituellement Ω , l’*espace fondamental* (ou encore l’*espace d’état* ou *univers*) contient l’ensemble de tous les résultats possibles d’un phénomène aléatoire. Un résultat possible d’une expérience sera noté $\omega \in \Omega$.

Si on reprend les exemples précédents, on peut facilement définir les univers associés.

Exemples

1. Jeu de Pile ou Face, $\Omega = \{\text{pile, face}\}$
2. Lancé de dés, $\Omega = \{1, 2, 3, 4, 5, 6\}$
3. Durée de vie d’une ampoule électrique, $\Omega = [0, +\infty[$
4. Température demain à 12h au sommet de la tour Eiffel (en degrés Kelvin), $\Omega = [0, +\infty[$
5. Evolution de la vitesse d’une molécule dans un gaz raréfié sur un intervalle de temps $[t_1, t_2]$, Ω : ensemble des application continues sur $[t_1, t_2]$ à valeurs dans \mathbb{R}^3

Cette liste d’exemples montre que l’espace Ω peut varier énormément dans sa structure, d’une expérience à l’autre. Cela permet de réaliser la richesse de la théorie qu’il faut mettre en place, pour créer un modèle qui englobe tous ces cas. Nous verrons également ultérieurement que le modèle abstrait que nous allons construire permettra de s’affranchir du fait que Ω décrit précisément tous les résultats possibles de l’expérience.

Définition – Evénement

Un *événement* est une propriété qui est vérifiée ou non une fois l’expérience réalisée. On identifie un événement A à un sous-ensemble ou *partie* de Ω , i.e. $A = \{\omega \in \Omega : A \text{ est vérifiée pour } \omega\}$.

Exemples

1. Jeu de Pile ou Face : $A = \{\text{pile}\}$.
2. Lancé de dés : $A = \{1, 3, 5\}$.
3. Durée de vie d’une ampoule électrique : $A = [t_1, t_2] \subset \mathbb{R}_+$.
4. Température demain à 12h au sommet de la tour Eiffel (en degrés Kelvin) : $A = [T_1, T_2] \cup [T_3, T_4] \subset \mathbb{R}_+$.
5. Evolution de la vitesse d’une molécule dans un gaz raréfié sur un intervalle de temps $[t_1, t_2] \subset \mathbb{R}_+$: $A = \{f \in C([t_1, t_2], \mathbb{R}^3) : \|f - g\|_\infty \leq a\}$, où $g \in C([t_1, t_2], \mathbb{R}^3)$ et $a \in \mathbb{R}_+$.

Les événements étant des ensembles, les opérations ensemblistes classiques admettent une interprétation probabiliste.

Correspondance entre opérations logiques et ensemblistes

Terminologie probabiliste	Terminologie ensembliste	Notation
événement certain	ensemble entier	Ω
événement impossible	ensemble vide	\emptyset
événement contraire	complémentaire	A^c
événement atomique	singleton	$\{\omega\}$
implication	inclusion	\subset
et	intersection	\cap
ou	réunion	\cup
événements incompatibles	ensembles disjoints	$A_1 \cap A_2 = \emptyset$

On doit maintenant répondre à la question de savoir quels sont les événements dont on va vouloir évaluer la probabilité d'occurrence. On va ainsi regrouper les événements en un ensemble \mathcal{A} qui constitue une collection de sous-ensembles de Ω . On va souhaiter en particulier pouvoir combiner des événements au sein de \mathcal{A} par les opérations ensemblistes courantes.

Ceci conduit à la notion de *tribu de parties* de Ω .

Définition – Tribu

Une *tribu* (ou *σ -algèbre*) \mathcal{A} est une collection de sous-ensembles de Ω tels que :

1. $\Omega \in \mathcal{A}$,
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
3. $\forall n \in \mathbb{N}, A_n \in \mathcal{A} \Rightarrow \bigcup_n A_n \in \mathcal{A}$.

Le couple (Ω, \mathcal{A}) est appelé *espace probabilisable*.

Exemples

1. $\mathcal{A} = \{\emptyset, \Omega\}$ est la tribu *grossière* ou *triviale* : c'est la plus petite tribu de Ω .
2. Dans le cas où Ω est au plus dénombrable, on choisit systématiquement l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω dont on vérifie aisément qu'il s'agit d'une tribu ; c'est le cadre étudié en CPGE. On verra plus loin que cette tribu est trop grande dans le cas où Ω est infini non dénombrable.
3. Si $\Omega = \mathbb{R}$, on peut le munir de la tribu formée des ensembles mesurables de \mathbb{R} , dite *tribu de Lebesgue*.
4. On verra ultérieurement qu'il est aisé de définir une tribu sur tout espace topologique.

Notion de densité de probabilité

Une des nouveautés majeures de ce cours par rapport au programme de CPGE est le cas où l'espace fondamental n'est plus fini ni dénombrable. On va voir ici que les outils développés dans le cours de calcul intégral vont nous permettre de définir une probabilité sur \mathbb{R} muni de la tribu des ensembles mesurables.

Soit $\Omega = \mathbb{R}$ et $f : \Omega \rightarrow \mathbb{R}_+$ une fonction intégrable telle que

$$\int_{\Omega} f(x) dx = 1.$$

Soit \mathcal{A} la collection des ensembles mesurables sur Ω ; les propriétés élémentaires des ensembles mesurables (cf. "Calcul Intégral II") établissent que \mathcal{A} est une tribu, sur laquelle on peut définir

$$\mathbb{P}(A) = \int_{\Omega} 1_A f(x) dx = \int_A f(x) dx.$$

On vérifie aisément que \mathbb{P} vérifie les 3 propriétés suivantes :

1. $\forall A \in \mathcal{A}, \mathbb{P}(A) \in [0, 1]$.
2. $\mathbb{P}(\Omega) = \int_{\Omega} f(x) dx = 1$.
3. Si A_n désigne une suite (dénombrable) d'événements **disjoints** de \mathcal{A} , on a, en appliquant le théorème de convergence dominée à la suite de fonctions $f(x)1_{\{\bigcup_{n=0}^m A_n\}}(x) = \sum_{n=0}^m f(x)1_{A_n}(x)$ ($m \in \mathbb{N}^*$), majorée trivialement par f intégrable :

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \int_{\Omega} \lim_{m \rightarrow +\infty} 1_{\{\bigcup_{n=0}^m A_n\}} f(x) dx \\ &= \lim_{m \rightarrow +\infty} \int_{\Omega} \sum_{n=0}^m 1_{A_n} f(x) dx \\ &= \lim_{m \rightarrow +\infty} \sum_{n=0}^m \int_{\Omega} 1_{A_n} f(x) dx \\ &= \lim_{m \rightarrow +\infty} \sum_{n=0}^m \mathbb{P}(A_n) \\ &= \sum_{n=0}^{+\infty} \mathbb{P}(A_n). \end{aligned}$$

Ces trois propriétés correspondent aux axiomes de Kolmogorov qui définissent une probabilité sur un espace probabilisable général. La fonction f est appelée *densité de probabilité*. On verra plus loin que l'on ne peut pas caractériser toutes les probabilités sur \mathbb{R} via cette notion. Celle-ci constitue néanmoins un exemple fondamental que l'on approfondira dans la suite du cours, notamment dans le cadre de l'étude des variables aléatoires.

Remarque

On pourra faire l'analogie entre la densité de probabilité et la loi de probabilité sur un univers discret, dans le sens où elle va "pondérer" les valeurs réelles, en remarquant cependant que :

- $f(x)$ n'est pas nécessairement inférieure à 1,
- $\mathbb{P}(\{x\}) = \int_{\{x\}} f(x) dx = 0$ et plus généralement, $\mathbb{P}(A) = 0$ si A est négligeable.

Probabilité

Définition – Probabilité

Une *probabilité* sur l'espace (Ω, \mathcal{A}) est une application $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$, telle que :

1. $\mathbb{P}(\Omega) = 1$,
2. pour toute suite (dénombrable) $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} **deux à deux disjoints**, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé *espace probabilisé*. La modélisation probabiliste consiste ainsi à décrire une expérience aléatoire par la donnée d'un espace probabilisé.

La définition suivante est fondamentale en théorie des probabilités. Elle introduit une notion de "vrai ou faux" qui dépend de la probabilité choisie sur l'espace fondamental.

Définition – Propriété presque-sûre

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On dit qu'un événement $A \in \mathcal{A}$ se réalise *\mathbb{P} -presque sûrement* (en abrégé \mathbb{P} -p.s.) si $\mathbb{P}(A) = 1$.

Proposition – Propriétés élémentaires

1. $\forall A \in \mathcal{A}, \mathbb{P}(A) \in [0, 1]$ et $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\forall A, B \in \mathcal{A}, A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\forall A, B \in \mathcal{A}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
4. Inégalité de Boole : $\forall n \in \mathbb{N}^*, \forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}$,

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i).$$

5. Formule de Poincaré : $\forall n \in \mathbb{N}^*, \forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}$

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^{n+1} \mathbb{P}\left(\bigcap_{i=1}^n A_i\right).$$

Démonstration Exercice. ■

Théorème de la continuité monotone

Dans le cas d'une suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} croissante, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Démonstration Exercice. ■

Remarque

Dans le cas d'une suite décroissante, on a

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}^*} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Le second point de la définition de la probabilité donne la probabilité de la réunion $\cup_n A_n$ en fonction des $\mathbb{P}(A_n)$ lorsque les événements sont deux à deux disjoints. Si ce n'est pas le cas, on a tout de même la majoration suivante, très utile dans la pratique :

Proposition

Soit \mathbb{P} une probabilité et soit $(A_n)_{n \in \mathbb{N}^*}$ une famille dénombrable d'événements. On a alors

$$\mathbb{P}(\cup_n A_n) \leq \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n)$$

Démonstration Exercice. ■

Probabilité conditionnelle

La construction d'un modèle probabiliste repose sur l'information connue **a priori** sur l'expérience aléatoire. Ce modèle permet de quantifier les probabilités de réalisation de certains résultats de l'expérience. Il est fondamental de remarquer que si l'information change, les probabilités de réalisation changent.

Exemple

On cherche pour un lancer de deux dés, la probabilité de l'événement "la somme est supérieure ou égale à 10". Elle vaut $1/6$ sans information supplémentaire, $1/2$ si l'on sait que le résultat d'un des dés est 6, 0 si l'on sait a priori que le résultat d'un des dés est 2. Pour obtenir ces résultats, on a calculé dans chaque cas le rapport du nombre de résultats favorables sur le nombre de cas possibles. Il est ainsi indispensable de bien définir l'espace de probabilité lié à l'expérience munie de l'information a priori. On remarque également que l'information a priori a changé la valeur de la probabilité de l'événement.

L'outil qui va nous permettre d'introduire de l'information est la probabilité conditionnelle dont nous donnons ici la définition.

Définition – Probabilité conditionnelle

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $A, B \in \mathcal{A}$ tels que $\mathbb{P}(B) > 0$. La *probabilité conditionnelle* de A sachant B est le nombre

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

Cela définit une probabilité comme le montre la proposition suivante.

Proposition

1. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $B \in \mathcal{A}$ tel que $\mathbb{P}(B) > 0$. Alors l'application de \mathcal{A} dans $[0, 1]$ qui à A associe $\mathbb{P}(A|B)$ définit une nouvelle probabilité sur Ω , appelée probabilité conditionnelle sachant B .
2. Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, nous avons

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

Démonstration Il est clair que $0 \leq \mathbb{P}(A|B) \leq 1$. Par ailleurs, les deux propriétés de la définition de la probabilité pour $\mathbb{P}(\cdot|B)$ proviennent des mêmes propriétés pour \mathbb{P} et des remarques suivantes : $\Omega \cap B = B$, et $(\bigcup_{n \in \mathbb{N}^*} A_n) \cap B = \bigcup_{n \in \mathbb{N}^*} (A_n \cap B)$. De plus, si A et C sont disjoints, il en est de même de $A \cap B$ et $C \cap B$. L'assertion 2 est évidente, d'après la définition de la Probabilité conditionnelle. ■

Formule des probabilités totales

Soit $(B_n)_{n \in \mathbb{N}^*}$ une partition finie ou dénombrable d'événements de Ω (i.e. telle que $\bigcup_{n \in \mathbb{N}^*} B_n = \Omega$ et les B_n sont deux-à-deux disjoints), telle que $\mathbb{P}(B_n) > 0$ pour tout $n \in \mathbb{N}^*$. Pour tout $A \in \mathcal{A}$, on a alors

$$P(A) = \sum_{n \in \mathbb{N}^*} P(A \cap B_n) = \sum_{n \in \mathbb{N}^*} P(A|B_n) P(B_n).$$

Démonstration Nous avons $A = \bigcup_{n \in \mathbb{N}^*} (A \cap B_n)$. Par hypothèse, les ensembles $(A \cap B_n)$ sont deux-à-deux disjoints et de plus $\mathbb{P}(A \cap B_n) = \mathbb{P}(A|B_n) \mathbb{P}(B_n)$. Le résultat découle du deuxième point de la définition de la probabilité. ■

Formule de Bayes

Selon les mêmes hypothèses que ci-dessus et si $\mathbb{P}(A) > 0$, on a

$$\forall i \in \{1, \dots, n\}, \mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\sum_{n \in \mathbb{N}^*} \mathbb{P}(A|B_n) \mathbb{P}(B_n)}.$$

Démonstration Le dénominateur vaut $\mathbb{P}(A)$ d'après la Formule des probabilités totales. La définition de la probabilité conditionnelle implique :

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\mathbb{P}(A)}.$$

■

Remarque

La formule de Bayes, simple conséquence des axiomes et de la définition de la probabilité conditionnelle, tient une place à part dans le calcul des probabilités en raison de son importance pratique considérable et des controverses auxquelles son application pratique a donné lieu : elle est à la base de toute une branche de la statistique appelée statistique bayésienne.

Exemple

Un individu est tiré au hasard dans une population où l'on trouve une proportion 10^{-4} de séropositifs. On lui fait passer un test de détection de la séropositivité. Par ailleurs, des expérimentations antérieures ont permis de savoir que les probabilités d'avoir un résultat positif lors de l'application du test si l'individu est séropositif, ou s'il ne l'est pas, sont respectivement égales à 0,99 (c'est la sensibilité du test) et à 0,001 ($0,999 = 1 - 0,001$ est la spécificité du test). Sachant que le test donne un résultat positif, quelle est la probabilité pour que l'individu soit effectivement séropositif ?

On considère les événements A "l'individu est séropositif", et B "le test de détection donne un résultat positif". Les données ci-dessus nous indiquent $\mathbb{P}(A) = 10^{-4}$ d'où $\mathbb{P}(A^c) = 0,9999$, $\mathbb{P}(B|A) = 0,99$ et $\mathbb{P}(B|A^c) = 0,001$. Nous trouvons

alors

$$\begin{aligned}\mathbb{P}(A|B) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} \\ &= \frac{0,99 \times 10^{-4}}{0,99 \times 10^{-4} + 0,001 \times 0,9999} \\ &\approx 0,09.\end{aligned}$$

On remarque que contrairement à l'intuition, cette probabilité est petite.

Indépendance des événements

La notion d'indépendance est absolument fondamentale en probabilités et nous verrons par la suite toutes ses implications dans la modélisation de l'aléatoire.

Intuitivement, deux événements A et B sont indépendants si le fait de savoir que A est réalisé ne donne aucune information sur la réalisation de B et réciproquement.

Si B est un événement de probabilité strictement positive, A sera dit indépendant de B si

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

On remarque que cette formule se symétrise et la notion d'indépendance se définit finalement comme suit.

Définition – Indépendance de deux événements

Deux événements A et B sont *indépendants* si et seulement si

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

Remarques

- La probabilité de voir A réalisé ne dépend pas de la réalisation de B , et réciproquement.
- Cette notion est liée au choix de la probabilité \mathbb{P} et n'est pas une notion ensembliste. Cela n'a en particulier rien à voir avec le fait que A et B soient disjoints ou non.
- Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, alors

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

Nous laissons en exercice (très simple à vérifier) la démonstration de la proposition suivante, dont le résultat est tout-à-fait intuitif.

Proposition

Si les événements A et B sont indépendants, alors il en est de même des couples (A^c, B) , (A, B^c) et (A^c, B^c) .

Exemple

1. On lance 3 fois un dé. Si A_i est un événement qui ne dépend que du $i^{\text{ème}}$ lancer, alors A_1 , A_2 , A_3 sont indépendants.
2. On tire une carte au hasard dans un jeu de 52 cartes. Soit $A = \{\text{la carte est une dame}\}$ et $B = \{\text{la carte est un coeur}\}$. Il est facile de voir que $\mathbb{P}(A) = 4/52$ et $\mathbb{P}(B) = 13/52$ et $\mathbb{P}(A \cap B) = \mathbb{P}(\{\text{la carte est la dame de coeur}\}) = 1/52 = \mathbb{P}(A)\mathbb{P}(B)$. Ainsi, les événements A et B sont indépendants pour la probabilité uniforme \mathbb{P} .
3. On suppose maintenant que le jeu de cartes soit trafiqué. Soit $\tilde{\mathbb{P}}$ la nouvelle probabilité correspondant au tirage de cartes. On suppose également que

$$\tilde{\mathbb{P}}(\{\text{As de trèfle}\}) = \frac{1}{2}, \quad \tilde{\mathbb{P}}(\{\text{autre carte}\}) = \frac{1}{2} \frac{1}{51} = \frac{1}{102},$$

alors

$$\tilde{\mathbb{P}}(A \cap B) = \frac{1}{102} \neq \tilde{\mathbb{P}}(A)\tilde{\mathbb{P}}(B) = \frac{2}{51} \frac{13}{102}.$$

Les événements A et B ne sont pas indépendants sous la probabilité $\tilde{\mathbb{P}}$.

Remarque : réflexion sur le concept de probabilité

La théorie mathématiques des probabilités ne dit pas quelle loi de probabilité choisir sur un espace (Ω, \mathcal{A}) parmi l'infinité de lois possibles. Ce problème qui concerne ceux qui veulent appliquer le calcul des probabilités, renvoie à la nature "physique" du concept de probabilité qui formalise et quantifie le sentiment d'incertitude vis-à-vis d'un événement. Ce problème d'ordre conceptuel oppose deux écoles de pensée, la conception objectiviste et la conception subjectiviste.

Pour les tenants du premier point de vue, la probabilité d'un événement peut être déterminée de manière unique. Dans la vision dite classique, héritée des jeux de hasard, Ω est fini et on donne à chaque événement élémentaire la même probabilité. Le calcul des probabilités se résume alors à un problème de dénombrement et la probabilité d'un événement est le rapport du nombre de cas favorables sur le nombre de cas possibles. Dans le cas infini, la vision fréquentiste repose sur la loi des grands nombres : si on répète un grand nombre de fois l'expérience, la proportion de fois où un événement sera réalisé va converger vers la probabilité de cet événement. Dans ce cadre, il est impossible de donner une valeur et même un sens à un événement non répétable comme "pleuvra-t-il demain ?". En outre, la répétition à l'infini d'une même expérience étant physiquement irréalisable, la loi des grands nombres étant un résultat qui suppose défini le concept de probabilité, la vision fréquentiste est logiquement intenable.

Dans la conception subjectiviste, la probabilité objective d'un événement n'existe pas et n'est donc pas une grandeur mesurable analogue à la masse d'un corps, par

exemple. C'est simplement une mesure d'incertitude qui reflète un degré de croyance pouvant varier avec les circonstances et l'observateur, donc subjective, la seule exigence étant qu'elle satisfasse aux axiomes du calcul des probabilités. Des méthodes ont alors été proposées pour passer d'un simple pré-ordre sur les événements, à une probabilité. Puisque la répétition n'est plus nécessaire, on peut probabiliser des événements non répétables et étendre ainsi le domaine d'application du calcul des probabilités, notamment pour orienter des prises de décisions. On notera que la formule de Bayes permet d'intégrer facilement de l'information a priori, dans la mesure où celle-ci est probabilisée.

On arrête ici ces quelques remarques sans prendre parti dans une querelle qui dure encore. L'un ou l'autre point de vue sera adopté selon les ouvrages rencontrés. Dans tous les cas, les outils mathématiques développés dans ce cours seront adaptés. On rappelle tout de même que la modélisation probabiliste a prouvé son efficacité dans de nombreuses applications mais que, comme tout modèle, ce n'est qu'une représentation simplificatrice de la réalité et que ses hypothèses doivent être mises à l'épreuve des faits. A ce titre, on citera Georges Matheron qui dans son essai sur la pratique des probabilités *Estimer et Choisir* (Matheron (1978)) écrit fort justement : "Il n'y a pas de probabilités en soi. Il n'y a que des modèles probabilistes".

Probabilité sur \mathbb{R}

Fonction de répartition

Nous avons vu précédemment la définition générale d'une probabilité \mathbb{P} sur un espace quelconque Ω muni d'une tribu \mathcal{A} . Un problème fondamental est de construire et de caractériser ces probabilités. La résolution de ce problème lorsque Ω est fini ou dénombrable est connu. Le cas général fait l'objet de la théorie de la mesure et sera développé ultérieurement.

Nous allons ici nous contenter de résoudre, sans démonstrations complètes, le cas où $\Omega = \mathbb{R}$ et où la tribu \mathcal{A} est la tribu formée des ensembles mesurables.

Définition – fonction de répartition

La *fonction de répartition* de la probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{A})$ est la fonction

$$F(x) = \mathbb{P}([-\infty, x]), \quad x \in \mathbb{R}.$$

Théorème

La fonction de répartition F caractérise la probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{A})$.

Ce résultat sera démontré ultérieurement.

Théorème

Une fonction F est la fonction de répartition d'une unique probabilité \mathbb{P} sur $(\mathbb{R}, \mathcal{A})$ si et seulement si elle vérifie les trois conditions suivantes :

1. elle est croissante,
2. elle est continue à droite,
3. $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1$.

Démonstration La première assertion est immédiate d'après la définition. Pour la seconde, on remarque que si x_n décroît vers x , alors $] - \infty, x_n]$ décroît vers $] - \infty, x]$ et donc $F(x_n)$ décroît vers $F(x)$ par le théorème de la continuité monotone. La troisième assertion se montre de manière analogue en remarquant que $] - \infty, x]$ décroît vers \emptyset (resp. croît vers \mathbb{R}) lorsque x décroît vers $-\infty$ (resp. croît vers $+\infty$).

La réciproque est plus difficile à obtenir et nécessite des éléments de théorie de la mesure qui nous font pour l'instant défaut. Nous renvoyons donc sa démonstration à un chapitre ultérieur. ■

Remarque

Comme F est croissante, elle admet une limite à gauche en chaque point notée $F(x^-)$. En remarquant que $] - \infty, y[= \lim_{n \rightarrow +\infty}] - \infty, y_n[$ si y_n tend vers y par valeurs décroissantes, on obtient pour $x < y$:

- $\mathbb{P}(]x, y]) = F(y) - F(x)$
- $\mathbb{P}(]x, y[) = F(y-) - F(x)$
- $\mathbb{P}([x, y]) = F(y) - F(x^-)$
- $\mathbb{P}([x, y[) = F(y-) - F(x^-)$

En particulier, $\mathbb{P}(\{x\}) = F(x) - F(x^-)$ est le **saut** de la fonction F au point x . On a donc $\mathbb{P}(\{x\}) = 0$ pour tout x si et seulement si F est continue en tout point.

Remarque

Le théorème ci-dessus explique pourquoi, d'un point de vue strictement mathématique, il est nécessaire d'introduire les tribus en probabilités, malgré la complexité que cela engendre.

Plus concrètement, considérons l'exemple suivant : soit $\Omega = [0, 1]$ et \mathbb{P} telle que $\mathbb{P}(]a, b]) = b - a$ pour $0 \leq a \leq b \leq 1$ (il s'agit de la loi uniforme sur $[0, 1]$). C'est une probabilité naturelle qui assigne à tout intervalle sa longueur comme probabilité. Supposons maintenant que l'on souhaite étendre de manière unique \mathbb{P} aux $2^{[0, 1]}$ éléments de $\mathcal{P}([0, 1])$ de manière à ce que $\mathbb{P}(\Omega) = 1$ et $\mathbb{P}(\cup_{n \in \mathbb{N}^*} A_n) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n)$ pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ tels que $A_n \cap A_m = \emptyset$ pour $n \neq m$. On peut prouver qu'un tel \mathbb{P} n'existe pas. $\mathcal{P}([0, 1])$ est trop "grand" pour définir un tel \mathbb{P} . Il contient en particulier des ensembles non mesurables.

Si l'on voulait travailler avec la tribu $\mathcal{A} = \mathcal{P}(\mathbb{R})$, il n'existerait que très peu de probabilités sur \mathbb{R} , à savoir les probabilités discrètes que l'on décrit rapidement ci-dessous.

Exemple

1. Les masses de Dirac (ou **mesures** de Dirac).

Soit $a \in \mathbb{R}$, on appelle mesure de Dirac en a , la probabilité \mathbb{P} sur \mathbb{R} qui vérifie pour $A \in \mathcal{P}(\mathbb{R})$

$$\mathbb{P}(A) = \begin{cases} 1 & \text{si } a \in A, \\ 0 & \text{sinon.} \end{cases}$$

Sa fonction de répartition est $F(x) = 1_{[a, +\infty[}(x)$.

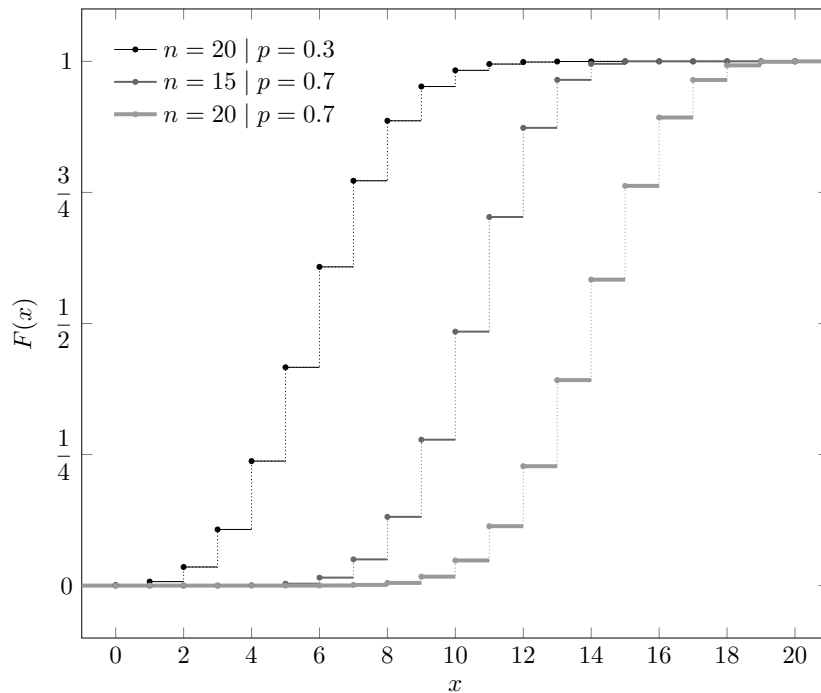
2. Les probabilités portées par \mathbb{N} .

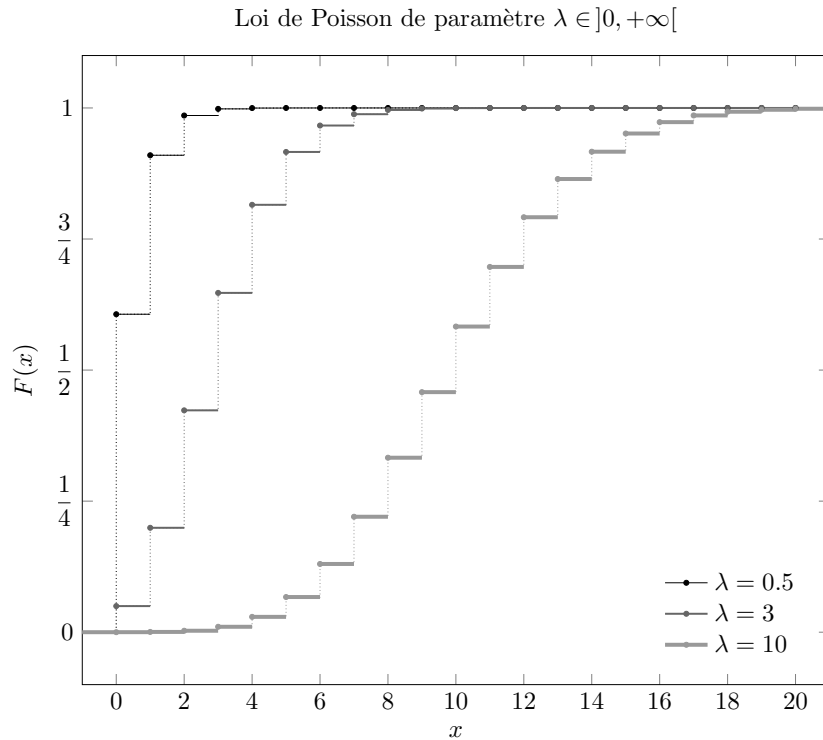
Comme \mathbb{N} est une partie de \mathbb{R} , toute probabilité sur \mathbb{N} peut être considérée comme une probabilité sur \mathbb{R} qui ne “charge” que \mathbb{N} . Plus précisément, si Q est une probabilité sur \mathbb{N} , on définit son “extension” \mathbb{P} à \mathbb{R} en posant $\mathbb{P}(A) = Q(A \cap \mathbb{N})$. Si $q_n = Q(\{n\})$ pour $n \in \mathbb{N}$, la fonction de répartition F de \mathbb{P} est

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ \sum_{i=0}^{\lfloor x \rfloor} q_i & \text{sinon,} \end{cases}$$

où $\lfloor \cdot \rfloor$ désigne la partie entière. A titre d'exemple, on représente ci-dessous les fonctions de répartitions de la loi binomiale et de la loi de Poisson.

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





3. Les probabilités discrètes.

Plus généralement, si E est une partie finie ou dénombrable de \mathbb{R} , toute probabilité Q sur E peut être considérée comme une probabilité \mathbb{P} sur \mathbb{R} , via la formule $\mathbb{P}(A) = Q(A \cap E)$. Si pour tout $i \in E$, on pose $q_i = Q(\{i\})$, la fonction de répartition F de \mathbb{P} est alors

$$F(x) = \sum_{\substack{i \in E \\ i \leq x}} q_i,$$

avec la convention qu'une somme "vide" vaut 0. On retrouve bien l'exemple 2 si $E = \mathbb{N}$. On voit que F est **purement discontinue** au sens où elle est complètement caractérisée par ses sauts $\Delta F(x) = F(x) - F(x^-)$:

$$F(x) = \sum_{y \leq x} \Delta F(y).$$

Il existe bien d'autres probabilités, non discrètes, sur \mathbb{R} . Le paragraphe suivant est consacré à un exemple très important, celui des probabilités à densité.

Densités de probabilités

Définition – densité de probabilité

Une fonction réelle f sur \mathbb{R} est une *densité de probabilité* (ou plus simplement une *densité*) si elle est positive, intégrable et vérifie

$$\int_{\mathbb{R}} f(x) dx = 1.$$

Si f est une densité, la fonction

$$F(x) = \int_{-\infty}^x f(y) dy$$

est la fonction de répartition d'une probabilité \mathbb{P} sur \mathbb{R} . On dit que f est la densité de \mathbb{P} ou que \mathbb{P} admet la densité f . Dans ce cas, F est continue, de sorte que $\mathbb{P}(\{x\}) = 0$ pour tout x , et elle est même dérivable et de dérivée f en tout point où f est continue. À l'inverse, si la fonction de répartition d'une probabilité \mathbb{P} est dérivable, ou seulement continue partout et dérivable par morceaux, alors \mathbb{P} admet une densité.

Il existe bien sûr des fonctions de répartition qui n'ont pas de densité : c'est le cas des probabilités discrètes données en exemple ci-dessus. Il existe des cas "mixtes" : soient d'une part f une fonction positive, intégrable et d'intégrale strictement positive et d'autre part une partie finie ou dénombrable E de \mathbb{R} et des poids $p_i > 0$ indexés par $i \in E$, tels que :

$$\int_{\mathbb{R}} f(x) dx + \sum_{i \in E} p_i = 1.$$

Alors la fonction

$$F(x) = \int_{-\infty}^x f(x) dx + \sum_{\substack{i \in E \\ i \leq x}} p_i$$

est une fonction de répartition, et la probabilité associée \mathbb{P} n'admet pas de densité et n'est pas non plus discrète.

Remarques

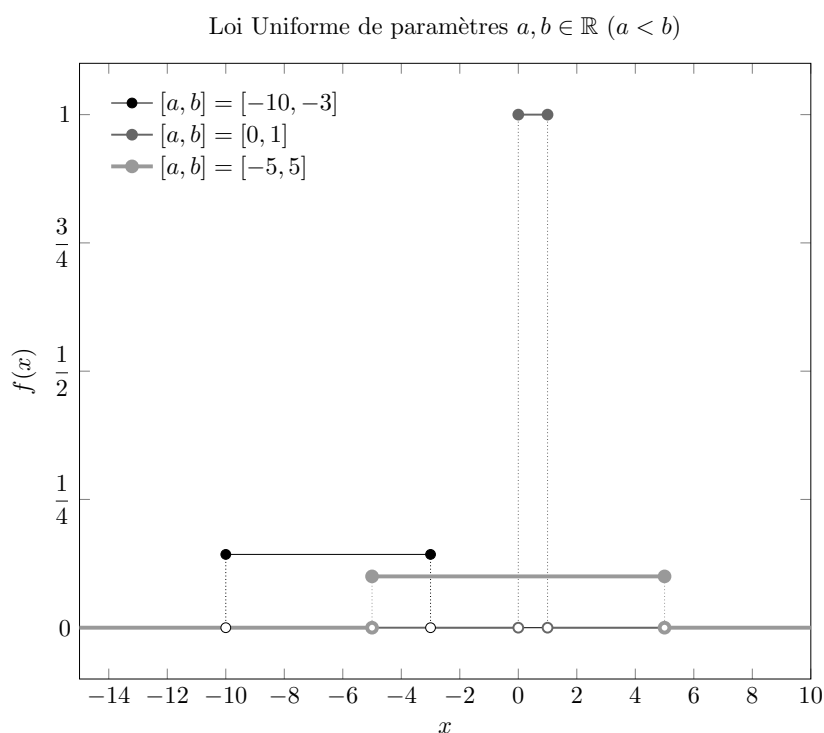
1. La fonction de répartition est entièrement déterminée par la probabilité \mathbb{P} . Il n'en est pas de même de la densité lorsqu'elle existe : si en effet on a $F(x) = \int_{-\infty}^x f(y) dy$ et si on pose $g(x) = f(x)$ si $x \notin E$ et $g(x) = f(x) + 1$ (par exemple) si $x \in E$, où E est un ensemble négligeable, alors g est encore une densité de \mathbb{P} .
2. Une interprétation intuitive de la densité f de \mathbb{P} : si dx est un petit accroissement de la variable x , on a (si du moins f est continue en x) :

$$f(x) \sim \frac{\mathbb{P}([x, x + dx])}{dx}.$$

Exemples de lois à densités

1. La *loi uniforme* sur $[a, b]$, avec $a, b \in \mathbb{R}$ tels que $a < b$, notée $\mathcal{U}_{[a,b]}$, est la probabilité \mathbb{P} qui admet la densité

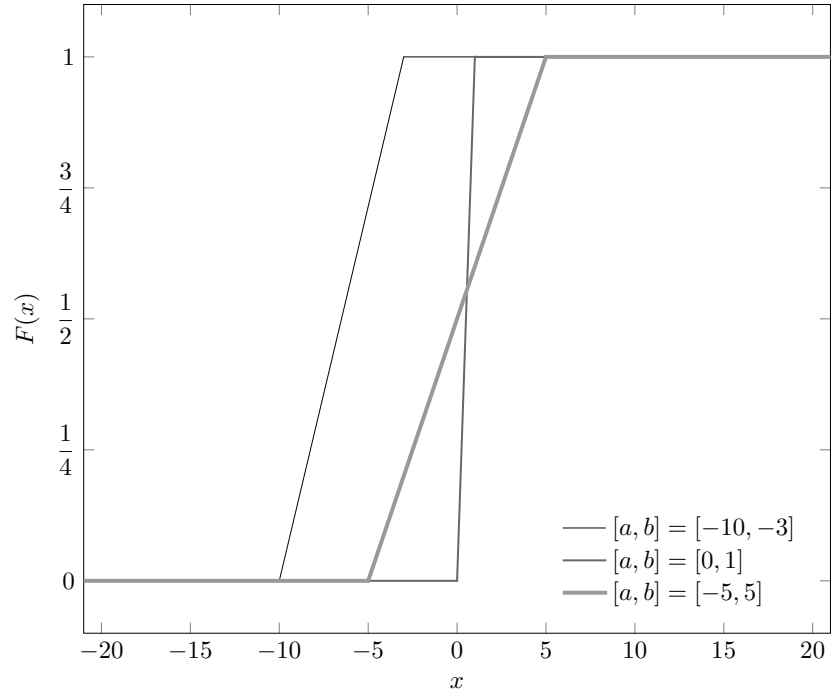
$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b, \\ 0 & \text{sinon.} \end{cases}$$



En vertu de la remarque 1. ci-dessus, on aurait pu choisir $f(a) = f(b) = 0$. Au vu de l'interprétation 2. ci-dessus, le fait que f soit constante sur $[a, b]$ exprime que si on choisit un point selon la probabilité uniforme, on a "autant de chances" de tomber au voisinage de chaque point de l'intervalle $[a, b]$. Cela explique le nom "uniforme". On remarque aussi que $\mathbb{P}(\{x\}) = 0$ pour tout x (comme pour toutes les probabilités à densité). On a donc une probabilité nulle de tomber exactement en un point x fixé à l'avance. On vérifie aisément que sa fonction de répartition est

$$F(x) = \begin{cases} 0 & \text{si } x < a, \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b, \\ 1 & \text{si } x > b. \end{cases}$$

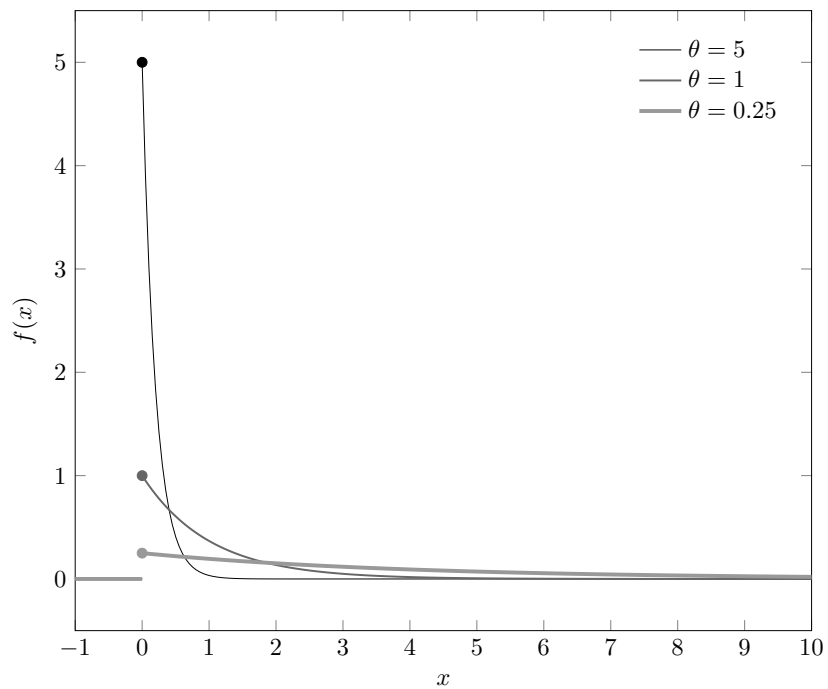
Loi Uniforme de paramètres $a, b \in \mathbb{R}$ ($a < b$)



2. La *loi exponentielle* de paramètre θ , notée \mathcal{E}_θ , est la loi de densité

$$f(x) = \begin{cases} 0 & \text{si } x < 0, \\ \theta e^{-\theta x} & \text{sinon.} \end{cases}$$

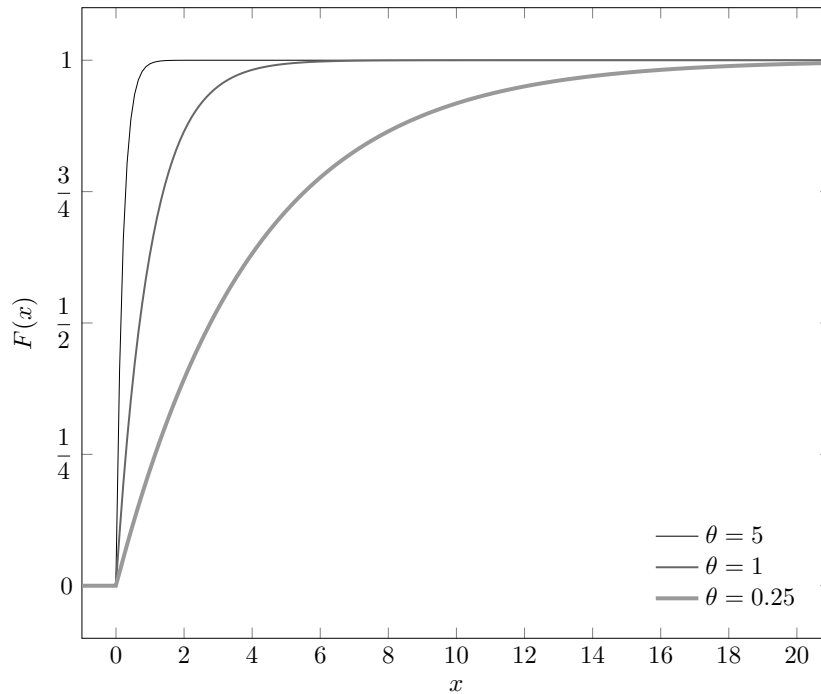
Loi Exponentielle de paramètre $\theta \in]0, +\infty[$



et de fonction de répartition

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - e^{-\theta x} & \text{sinon.} \end{cases}$$

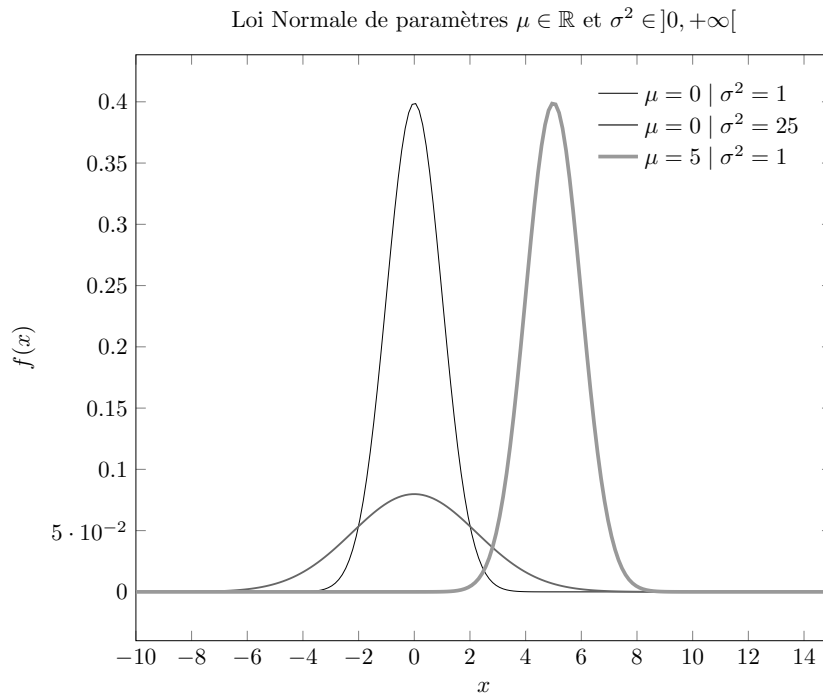
Loi Exponentielle de paramètre $\theta \in]0, +\infty[$



Dans la pratique, on utilise fréquemment la loi exponentielle pour modéliser une durée de vie ou le temps d'attente avant l'arrivée d'un événement spécifique. Par exemple la durée de vie d'une bactérie, d'un composant électronique, la durée d'une conversation téléphonique ou le temps qui nous sépare du prochain tremblement de terre peuvent être considérées comme des variables aléatoires de loi exponentielle.

3. La *loi normale* (ou de *Gauss* ou encore *gaussienne*) de paramètres μ et σ^2 , notée $\mathcal{N}(\mu, \sigma^2)$, est la loi de densité

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Lorsque $\mu = 0$ et $\sigma^2 = 1$, on l'appelle la loi normale *centrée réduite* pour des raisons que nous expliciterons au prochain chapitre.

La distribution normale fut introduite par De Moivre en 1733. Celui-ci l'utilisa pour approximer une variable aléatoire binomiale quand le paramètre n de celle-ci était grand. Ce résultat fut ensuite progressivement généralisé par Laplace et d'autres confrères pour devenir le théorème connu sous le nom de théorème de la limite centrale, qui sera démontré au Chapitre 4. Ce théorème est l'un des plus importants de la théorie des probabilités et prouve que de très nombreux phénomènes aléatoires suivent approximativement une loi normale. On peut citer à titre d'exemple la taille d'un individu choisi au hasard, les composantes de la vitesse d'une molécule de gaz ou l'erreur de mesure d'une quantité physique.

Il est difficile de faire des calculs avec la loi normale car sa densité n'admet pas de primitive explicite. Aussi des tables numériques ont-elles été construites pour permettre aux utilisateurs d'obtenir très rapidement des valeurs numériques. Elles sont disponibles dans la plupart des logiciels et permettent entre autre de représenter sa fonction de répartition.

Nous aurons l'occasion de voir par la suite un grand nombre d'autres exemples de probabilités à densité.

Exercices

Propriétés élémentaires

A partir de la définition de la probabilité, démontrer les propriétés suivantes :

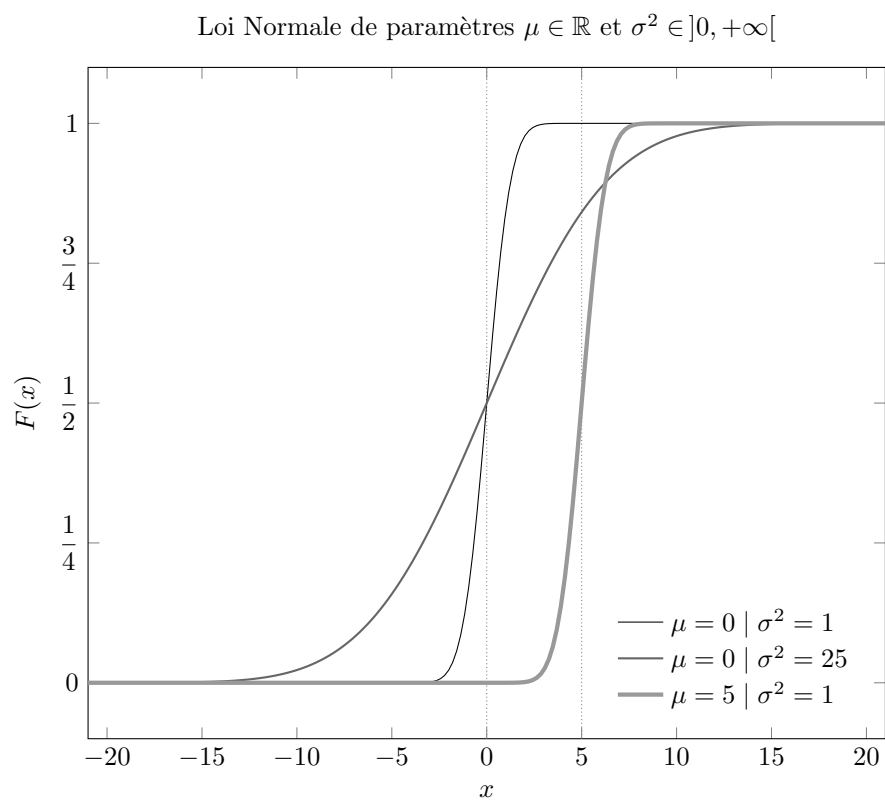


FIGURE 1 – fonction de répartition de la loi normale

1. $\forall A \in \mathcal{A}, \mathbb{P}(A) \in [0, 1]$ et $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
2. $\forall A, B \in \mathcal{A}, A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\forall A, B \in \mathcal{A}, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
4. Inégalité de Boole : $\forall n \in \mathbb{N}^*, \forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}, \mathbb{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$.
5. Formule de Poincaré : $\forall n \in \mathbb{N}^*, \forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}$

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^n \mathbb{P}\left(\bigcap_{i=1}^n A_i\right).$$

(?)

Continuité monotone

Soit une suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} croissante (au sens de l'inclusion).
Montrer que

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

(?)

Une définition alternative de la probabilité

Soit (Ω, \mathcal{A}) un espace probabilisable. Supposons que $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ vérifie :

1. $\mathbb{P}(\Omega) = 1$,
2. Pour $A, B \in \mathcal{A}$, tels que $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (additivité),
3. Pour toute suite $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} croissante

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Montrer que \mathbb{P} vérifie la propriété de σ -additivité. (?)

Généralisation de l'inégalité de Boole

Soit \mathbb{P} une probabilité et soit $(A_n)_{n \in \mathbb{N}^*}$ une famille dénombrable d'événements.
On a alors

$$\mathbb{P}(\bigcup_n A_n) \leq \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n)$$

(?)

Indépendance et conditionnement

Question 1 Dans quelles circonstances particulières deux événements A et B tels que $A \cup B$ soit presque sûr sont-ils indépendants ? (?)

Question 2 Soit A, B et C trois événements indépendants **2 à 2**, avec $\mathbb{P}(C) > 0$. Donner une condition nécessaire et suffisante sur A, B et C pour que A et B soient indépendants relativement à la probabilité conditionnelle $\mathbb{P}(\cdot | C)$, à la place de \mathbb{P} . (?)

Question 3 Etablir que deux événements A et B sont indépendants ssi

$$\mathbb{P}(A \cap B)\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A \cap B^c)\mathbb{P}(A^c \cap B)$$

(?)

Problème de Monty Hall

Dans un jeu télévisé, vous êtes confrontés au problème suivant : devant vous se trouvent 3 portes fermées. Derrière l'une d'entre elle se trouve une voiture, derrière chacune des deux autres, une chèvre. Vous désignez au hasard l'une d'entre elle. Le présentateur, qui sait où se trouve la voiture, ouvre alors une porte, qui n'est ni celle que vous avez choisie, ni celle qui cache la voiture. Il vous offre alors la possibilité de réviser votre choix. Que choisissez-vous ?

Calculer la probabilité de remporter la voiture selon les deux stratégies (changer ou non son choix de porte). (?)

Queues de distributions

Soit une densité f sur \mathbb{R} étendue à $[-\infty, +\infty]$ telle que la fonction $h : x \in [-\infty, +\infty] \mapsto x f(x)$ est absolument intégrable. Cette hypothèse sera appelée (\mathcal{H}) . On note F la fonction de répartition correspondante sur la droite réelle achevée. Nous allons montrer que¹

$$\left| \begin{array}{l} 1 - F(x) \underset{x \rightarrow +\infty}{=} o\left(\frac{1}{x}\right), \quad (1) \\ F(x) \underset{x \rightarrow -\infty}{=} o\left(\frac{1}{x}\right). \quad (2) \end{array} \right.$$

Etude en $+\infty$

1. Soient $a \in [-\infty, +\infty]$ et f et g deux fonctions définies sur un même voisinage I de a . On dit que f est *négligeable* par rapport à g (ou que g est *prépondérante* par rapport à f) au voisinage de a et on note $f(x) \underset{x \rightarrow a}{=} o(g(x))$ s'il existe une fonction ϵ définie sur I telle que $\epsilon(x) \underset{x \rightarrow a}{\longrightarrow} 0$ et $f = \epsilon \times g$.

1. Vérifier que pour tout $u > 0$,

$$\int_{-\infty}^{+\infty} h(x) dx \geq \int_{-\infty}^u h(x) dx + u (1 - F(u)).$$

2. En déduire (1).

(?)

Etude en $-\infty$

3. En vous inspirant du raisonnement précédent, montrer (2).

(?)

Interprétation

4. Proposez une interprétation de ce résultat.

(?)

Etude de la réciproque Considérons maintenant la fonction de répartition

$$F : x \in \mathbb{R} \mapsto \begin{cases} 1 - \frac{e}{x \ln(x)} & \text{si } x \geq e, \\ 0 & \text{si } x < e, \end{cases}$$

que l'on étend à la droite réelle achevée en posant $F(-\infty) = 0$ et $F(+\infty) = 1$.

5. Vérifier que cette fonction possède bien une densité f et l'explicitier.
6. A-t-on les développements asymptotiques (1) et (2) ?
7. La fonction h correspondante est-elle absolument intégrable ?
8. Qu'en concluez-vous ?

(?)

Liens entre les développements asymptotiques Soient les fonctions de répartition

$$F : x \in \mathbb{R} \mapsto \begin{cases} 1 - \frac{1}{\ln(x)} & \text{si } x \geq e, \\ 0 & \text{sinon,} \end{cases}$$

et $G : x \in \mathbb{R} \mapsto 1 - F(-x)$, toutes deux étendues à la droite réelle achevée en posant $\lim_{x \rightarrow -\infty} F(x) = \lim_{x \rightarrow -\infty} G(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = \lim_{x \rightarrow +\infty} G(x) = 1$.

9. Que peut-on dire sur les éventuels liens entre (1) et (2) à l'aide de ces deux fonctions ?

(?)

(?)

Densité et fonction de répartition d'une loi Normale

On considère la densité d'une loi Normale centrée réduite :

$$f : x \in [-\infty, +\infty] \mapsto \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{x^2}{2} \right\},$$

où $f(-\infty) := \lim_{x \rightarrow -\infty} f(x)$ et $f(+\infty) := \lim_{x \rightarrow +\infty} f(x)$.

Propriétés générales

1. Faire l'étude de f (domaine, parité, limites aux bornes, dérivabilité, variations, convexité/concavité).
2. Donner la définition de la fonction de répartition F associée à f . En donner une expression faisant apparaître la *fonction d'erreur* (qui est une fonction spéciale)

$$\operatorname{erf} : x \in \mathbb{R} \mapsto \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du.$$

(?)

Encadrement de $1 - F(x)$ pour tout $x > 0$ Nous allons maintenant démontrer la propriété suivante : $\forall x > 0$

$$\frac{x}{1+x^2} f(x) \leq 1 - F(x) \leq \frac{f(x)}{x}.$$

3. En observant que $u \geq x > 0$ implique $\frac{u}{x} \geq 1$, montrer l'inégalité de droite.
4. En observant que $u \geq x > 0$ implique $u^{-2} \leq x^{-2}$, montrer que $\forall x > 0$,

$$\left(1 + \frac{1}{x^2}\right) (1 - F(x)) \geq \frac{f(x)}{x}.$$

En déduire l'inégalité de gauche.

(?)

Equivalent de $1 - F(x)$ quand $x \rightarrow +\infty$.

5. Dédurre de l'encadrement précédent un équivalent de $1 - F(x)$ lorsque $x \rightarrow +\infty$.

(?)

Remarque

Pour $x \in \mathbb{R}$, le rapport $\frac{1 - F(x)}{f(x)}$ est appelé *ratio de Mills*. Il est beaucoup utilisé en statistique, en particulier pour l'analyse des modèles de régression en présence de biais de sélection.

6. Comparer ce résultat à celui de l'exercice sur les développements limités pour les fonctions de répartition. En particulier, $h : x \in [-\infty, +\infty] \mapsto x f(x)$ est-elle absolument intégrable ?

Loi Normale générale. Considérons maintenant la densité plus générale

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\},$$

où $\mu \in \mathbb{R}$ et $\sigma^2 > 0$. On note G sa fonction de répartition associée.

7. En réécrivant g en fonction de f , déduire des questions précédentes un équivalent de $1 - G(x)$ lorsque $x \rightarrow +\infty$.

(?)

Solutions

Propriétés élémentaires

1. $\mathbb{P}(A) \in [0, 1]$ est donné par la définition. Par ailleurs, on a $\Omega = A \cup A^c$ et $A \cap A^c = \emptyset$. En considérant la suite $A_1 = A$, $A_2 = A^c$ et $A_n = \emptyset$ pour $n > 2$, le point 2. de la définition nous donne $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c)$.
2. Puisque $A \subset B$, on a l'union disjointe $B = (B \cap A) \cup (B \cap A^c) = A \cup (B \cap A^c)$. Comme $\mathbb{P}(B \cap A^c) \geq 0$ par définition, on en déduit le résultat.
3. On a $(A \cup B) = B \cup (A \cap B^c)$, avec $B \cap (A \cap B^c) = \emptyset$ ainsi que $A = (A \cap B^c) \cup (A \cap B)$, où $(A \cap B^c) \cap (A \cap B) = \emptyset$. On en déduit $\mathbb{P}(A \cup B) = \mathbb{P}(B) + \mathbb{P}(A \cap B^c)$ et $\mathbb{P}(A) = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B)$, d'où le résultat.
4. Par récurrence. On a déjà vu que $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$. Notant $B = \bigcup_{i=1}^n A_i$, alors $\mathbb{P}(\bigcup_{i=1}^{n+1} A_i) \leq \mathbb{P}(B) + \mathbb{P}(A_{n+1})$, et l'hypothèse de récurrence nous indique $\mathbb{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$, d'où le résultat.
5. Par récurrence. On a déjà vu que $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Notant $B = \bigcup_{i=1}^n A_i$, alors

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}(B \cup A_{n+1}) \\
&= \mathbb{P}(B) + \mathbb{P}(A_{n+1}) - \mathbb{P}(B \cap A_{n+1}) \\
&= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^n \mathbb{P}\left(\bigcap_{i=1}^n A_i\right) \\
&\quad + \mathbb{P}(A_{n+1}) \\
&\quad - \sum_{i=1}^n \mathbb{P}(A_i \cap A_{n+1}) + \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^{n+1} \mathbb{P}\left(\bigcap_{i=1}^n A_i \cap A_{n+1}\right) \\
&= \sum_{i=1}^{n+1} \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n+1} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^{n+1} \mathbb{P}\left(\bigcap_{i=1}^{n+1} A_i\right).
\end{aligned}$$

Continuité monotone

On définit une suite $(B_n)_{n \in \mathbb{N}^*}$ telle que $B_1 = A_1$ et $B_n = A_n \setminus B_{n-1}$, pour $n \geq 2$. Les B_n ainsi définis sont deux à deux disjoints et $\bigcup B_n = \bigcup A_n$, donc :

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_n \mathbb{P}(B_n) = \lim_{n \rightarrow \infty} \sum_{p=1}^n \mathbb{P}(B_p) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Une définition alternative de la probabilité

Soit A_n une suite d'éléments de \mathcal{A} deux-à-deux disjoints. On définit $B_n = \bigcup_{p \leq n} A_p$ et $B = \bigcup_n A_n$. Comme \mathbb{P} est additive, on a $\mathbb{P}(B_n) = \sum_{p \leq n} \mathbb{P}(A_p)$ qui croît vers $\sum_n \mathbb{P}(A_n)$ et aussi vers $\mathbb{P}(B)$ d'après le point 3.

En considérant les résultats de ces deux exercices, on obtient une définition alternative de la probabilité en substituant la continuité monotone et l'additivité à la propriété de σ -additivité.

Généralisation de l'inégalité de Boole

On a déjà vu le cas fini avec l'inégalité de Boole. On pose $B_n = \bigcup_{i=1}^n A_i$, qui croît vers l'ensemble $C = \bigcup_n A_n$. D'après l'inégalité de Boole, on a

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

Mais $\mathbb{P}(B_n) \rightarrow_{n \rightarrow \infty} \mathbb{P}(C)$ d'après le théorème de continuité monotone, tandis que $\sum_{i=1}^n \mathbb{P}(A_i) \rightarrow_{n \rightarrow \infty} \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n)$. En passant à la limite, on obtient donc le résultat.

Indépendance et conditionnement

Question 1 On a $1 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$, donc $1 = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B)$ si et seulement si A et B sont indépendants. Cette égalité se réécrit $(1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) = 0$, qui est vérifiée ssi A ou B est presque sûr.

Question 2 Par définition, on a A et B indépendants relativement à $\mathbb{P}(\cdot|C)$ ssi

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

car A et B sont chacun indépendants de C , d'où

$$\frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} = \mathbb{P}(A)\mathbb{P}(B)$$

par définition ; et comme A , B et C sont indépendants 2 à 2,

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

équivalent à A , B et C sont mutuellement indépendants.

Question 3 On a $\mathbb{P}(A^c \cap B^c) = \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A) - \mathbb{P}(B) + \mathbb{P}(A \cap B)$, donc

$$\mathbb{P}(A \cap B)\mathbb{P}(A^c \cap B^c) = \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B)(\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(A \cap B))$$

Par ailleurs, on a

$$\begin{cases} \mathbb{P}(A \cap B^c) &= \mathbb{P}(A \setminus (A \cap B)) = \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ \mathbb{P}(A^c \cap B) &= \mathbb{P}(B \setminus (A \cap B)) = \mathbb{P}(B) - \mathbb{P}(A \cap B) \end{cases}$$

Donc

$$\mathbb{P}(A \cap B^c)\mathbb{P}(A^c \cap B) = \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A \cap B)(\mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B))$$

On a donc bien l'égalité ssi $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B)$

Problème de Monty Hall

Considérons les événements A = 'la porte choisie en premier est la bonne' et B = 'la dernière porte est la bonne'.

On a évidemment $\mathbb{P}(A) = \frac{1}{3}$, puisque le choix initial se fait uniformément parmi les 3 portes.

Pour B , on doit tenir compte de l'information donnée par le présentateur. On a par la formule des probabilités totales

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}$$

En effet $\mathbb{P}(B|A^c) = 1$ car si la porte choisie initialement n'est pas la bonne, c'est nécessairement la dernière.

Il convient donc de changer son choix compte tenu de la nouvelle information. Pour se convaincre du résultat, on peut refaire le calcul avec disons 100 portes et le présentateur qui ouvre 98 autres porte après le choix du candidat.

Queues de distributions

Etude en $+\infty$

1. Soit $u > 0$. Comme h est absolument intégrable, elle est intégrable et on peut écrire

$$\begin{aligned} \int_{-\infty}^{+\infty} h(x) dx &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^u x f(x) dx + \int_u^{+\infty} x f(x) dx \\ &\geq \int_{-\infty}^u x f(x) dx + u \int_u^{+\infty} f(x) dx \\ &= \int_{-\infty}^u h(x) dx + u (1 - F(u)). \end{aligned}$$

2. Soit toujours $u > 0$. Toute fonction de répartition étant à valeurs dans $[0, 1]$, on a $u (1 - F(u)) \geq 0$. En utilisant la question précédente on obtient donc l'encadrement

$$\begin{aligned} 0 \leq u (1 - F(u)) &\leq \int_{-\infty}^{+\infty} h(x) dx - \int_{-\infty}^u h(x) dx \\ &= \int_u^{+\infty} h(x) dx = \int_{-\infty}^{+\infty} h(x) 1_{[u, +\infty]}(x) dx. \end{aligned}$$

Or pour tout $x \in [-\infty, +\infty]$ on a $|h(x) 1_{[u, +\infty]}(x)| \leq |h(x)|$, qui est intégrable. On peut donc appliquer le théorème de convergence dominée, qui nous donne :

$$\begin{aligned} \lim_{u \rightarrow +\infty} u (1 - F(u)) &= \int_{-\infty}^{+\infty} h(x) \lim_{u \rightarrow +\infty} 1_{[u, +\infty]}(x) dx \\ &= \int_{-\infty}^{+\infty} h(x) 1_{\emptyset}(x) dx = 0. \end{aligned}$$

On en conclut que $1 - F(x) = o\left(\frac{1}{x}\right)$ lorsque $x \rightarrow +\infty$.

Etude en $-\infty$

3. Prenons maintenant $u < 0$. Comme h est (absolument) intégrable, on peut écrire

$$\begin{aligned}\int_{-\infty}^{+\infty} h(x) dx &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^u x f(x) dx + \int_u^{+\infty} x f(x) dx \\ &\leq u \int_{-\infty}^u f(x) dx + \int_u^{+\infty} x f(x) dx \\ &= u F(u) + \int_u^{+\infty} h(x) dx.\end{aligned}$$

Or, comme F prend ses valeurs dans $[0, 1]$, $u F(u) \leq 0$. On obtient donc

$$\begin{aligned}0 &\geq u F(u) \geq \int_{-\infty}^{+\infty} h(x) dx - \int_u^{+\infty} h(x) dx \\ &= \int_{-\infty}^u h(x) dx = \int_{-\infty}^{+\infty} h(x) 1_{[-\infty, u]}(x) dx.\end{aligned}$$

Comme précédemment, cette dernière fonction sous l'intégrale est bornée par $-|h|$ (à gauche) et $|h|$ (à droite), toutes deux intégrables, donc d'après le théorème de convergence dominée :

$$\begin{aligned}\lim_{u \rightarrow -\infty} u F(u) &= \int_{-\infty}^{+\infty} h(x) \lim_{u \rightarrow -\infty} 1_{[-\infty, u]}(x) dx \\ &= \int_{-\infty}^{+\infty} h(x) 1_{\emptyset}(x) dx = 0.\end{aligned}$$

On a donc bien $F(x) = o\left(\frac{1}{x}\right)$ lorsque $x \rightarrow -\infty$.

Interprétation

4. L'hypothèse que h est (absolument) intégrable nous donne une indication sur la vitesse de convergence de $1 - F(x)$ et $F(x)$ lorsque x tend respectivement vers $+\infty$ et $-\infty$: elles convergent vers 0 au moins aussi vite que $\frac{1}{x}$. On dit que les queues de la distribution sont “relativement” fines.

Pour mieux comprendre l'impact de ce résultat, prenons un exemple pratique. Considérons un phénomène aléatoire comme la concentration en polluant dans l'air, modélisé par une loi de probabilité sur \mathbb{R} de densité f nulle sur \mathbb{R}_+^* . Etre capable de déterminer la probabilité que la concentration dépasse un seuil critique x est alors très important pour les organismes de contrôle de la qualité de l'air. Or cet événement correspond mathématiquement à l'événement $[x, +\infty]$. Dans ce cas, $1 - F(x)$ nous donne cette probabilité selon le modèle considéré. Le résultat nous indique que si l'on choisit un modèle tel que h est intégrable, alors cette probabilité décroît plus rapidement que $\frac{1}{x}$ lorsque le seuil x augmente. En d'autres termes, les événements extrêmes (quand x est grand) restent “assez” rares. L'impact du choix de modèle n'est donc pas négligeable : si on prend une

distribution à queues trop fines alors que les pics de pollution ne sont en réalité pas si rares que ça, $1 - F(x)$ risque de sous-estimer le risque réel de dépassement du seuil !

Etude de la réciproque

5. La fonction F considérée est continue et dérivable par morceaux sur $[-\infty, +\infty]$; le seul point où F n'est pas dérivable est e (le taux d'accroissement n'a pas les mêmes limites à gauche et à droite). Elle possède donc une densité, qui s'obtient en dérivant chaque morceau : pour tout $x \in \mathbb{R}$,

$$f(x) = \begin{cases} e \frac{\ln(x) + 1}{x^2 \ln(x)^2} & \text{si } x \geq e, \\ 0 & \text{sinon.} \end{cases}$$

On peut l'étendre à la droite réelle achevée en posant $f(\pm\infty) = \lim_{x \rightarrow \pm\infty} f(x) = 0$.

6. On remarque que pour tout $x \geq e$, $x(1 - F(x)) = \frac{e}{\ln(x)} \xrightarrow{x \rightarrow +\infty} 0$, donc le développement asymptotique (1) est bien vérifié. En outre, comme $x F(x) = 0$ pour tout $x < e$, le développement asymptotique (2) est aussi bien respecté.
7. Pour tout $x \in \mathbb{R}$,

$$h(x) = |h(x)| = \begin{cases} e \frac{\ln(x) + 1}{x \ln(x)^2} & \text{si } x \geq e, \\ 0 & \text{sinon,} \end{cases}$$

et on pose $h(\pm\infty) = \lim_{x \rightarrow \pm\infty} h(x) = 0$. Par conséquent, h est clairement (absolument) intégrable sur $[-\infty, e]$, où elle est égale presque partout (partout sauf en e) à la fonction nulle. Pour savoir si h est (absolument) intégrable sur toute la droite réelle achevée, il nous faut donc regarder si elle l'est sur $[e, +\infty]$. Pour cela, nous allons utiliser le théorème de Hake, qui nous dit de vérifier que h est intégrable sur tout segment $[a, b] \subset [e, +\infty]$, puis que $\lim_{t \rightarrow +\infty} \int_e^t h(x) dx$ existe et est finie. Remarquons d'abord que h est continue, donc elle est intégrable sur tout segment strictement inclus dans $[e, +\infty]$; le premier point est vérifié. On remarque en outre que pour tout $x \geq e$,

$$h(x) \geq \frac{1}{x \ln(x)}.$$

Par conséquent, pour tout $t > e$ on a :

$$\int_e^t h(x) dx \geq \int_e^t \frac{1}{x \ln(x)} dx = [\ln(\ln(x))]_e^t = \ln(\ln(t)).$$

Or cette dernière quantité tend vers $+\infty$ quand $t \rightarrow +\infty$. Donc $\int_e^t h(x) dx$ n'a pas de limite finie quand $t \rightarrow +\infty$. On en conclut que h n'est pas (absolument) intégrable.

8. La fonction de répartition F considérée respecte bien les développements asymptotiques (1) et (2), mais la fonction h associée n'est pas (absolument) intégrable. Nous avons donc montré que la réciproque de la propriété est fausse.

Liens entre les développements asymptotiques

9. On remarque que F satisfait (2) mais pas (1), tandis que G satisfait (1) mais pas (2). On en conclut que (1) n'implique pas (2) et inversement.

Densité et fonction de répartition d'une loi Normale

Propriétés générales

- Faisons l'étude de f .
 - **Domaine.** $f : [-\infty, +\infty] \rightarrow \mathbb{R}^+$.
 - **Parité.** On remarque que f est paire : $\forall x \in \mathbb{R}, f(x) = f(-x)$.
 - **Limites aux bornes.** $\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow +\infty} f(x) = 0$. Ainsi, la parité est aussi vraie aux bornes.
 - **Dérivabilité.** f est continue sur $[-\infty, +\infty]$ et infiniment dérivable sur \mathbb{R} en tant que composée de fonctions infiniment dérivables. Pour tout $x \in \mathbb{R}$ sa dérivée première s'écrit $f'(x) = -x f(x)$ et a dérivée seconde $f''(x) = (x^2 - 1) f(x)$. Ces deux dernières fonctions peuvent être étendues à la droite réelle achevée en posant $f'(\pm\infty) := \lim_{x \rightarrow \pm\infty} f'(x) = 0$ et $f''(\pm\infty) := \lim_{x \rightarrow \pm\infty} f''(x) = 0$.
 - **Variations.** f' est strictement positive (resp. négative) ssi $-\infty < x < 0$ (resp. $+\infty > x > 0$). Elle est nulle en 0, $+\infty$ et $-\infty$. Ainsi, f est strictement croissante sur $]-\infty, 0[$, vaut $(2\pi)^{-1/2}$ en 0, puis est strictement décroissante sur $]0, +\infty[$.
 - **Convexité/Concavité.** f'' est strictement positive (resp. négative) sur $\mathbb{R} \setminus]-1, 1[$ (resp. $] -1, 1[$). f est donc convexe sur $\mathbb{R} \setminus]-1, 1[$ et concave sur $] -1, 1[$.
- Par définition, pour tout $x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du.$$

On pose $F(+\infty) = 1$ et $F(-\infty) = 0$. Lorsque $x = 0$, comme f est paire et que son intégrale sur $[-\infty, +\infty]$ vaut 1 (c'est une densité), son intégrale sur $[-\infty, 0]$ (i.e. $F(0)$) vaut $\frac{1}{2}$. Lorsque $x > 0$, en décomposant l'intégrale puis en utilisant un petit changement de variable on obtient

$$F(x) = \frac{1}{2} + \int_0^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du = \frac{1}{2} + \int_0^{\frac{x}{\sqrt{2}}} \frac{1}{\sqrt{\pi}} e^{-v^2} dv = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right).$$

Lorsque $x < 0$, on peut procéder de la même manière avec $1 - F(x)$ et obtenir la même égalité que ci-dessus.

Encadrement de $1 - F(x)$ pour tout $x > 0$

3. Soit $x > 0$. Par définition,

$$1 - F(x) = \int_{-\infty}^{+\infty} f(u) du - \int_{-\infty}^x f(u) du = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du.$$

Or pour tout $u \geq x$ on a $u/x \geq 1$, donc

$$1 - F(x) \leq \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \frac{u}{x} \exp\left\{-\frac{u^2}{2}\right\} du = \frac{-1}{x} \int_x^{+\infty} f'(u) du.$$

Or f' a pour primitive f donc d'après le théorème fondamental du calcul (extension) on a

$$1 - F(x) \leq -\frac{f(+\infty) - f(x)}{x} = \frac{f(x)}{x}.$$

4. Soit $x > 0$. Par définition,

$$\left(1 + \frac{1}{x^2}\right) (1 - F(x)) = \int_x^{+\infty} \left(1 + \frac{1}{x^2}\right) f(u) du.$$

Or pour tout $u \geq x$ on a $\frac{1}{u^2} \leq \frac{1}{x^2}$, d'où

$$\left(1 + \frac{1}{x^2}\right) (1 - F(x)) \geq \int_x^{+\infty} \left(1 + \frac{1}{u^2}\right) f(u) du.$$

Remarquons maintenant que d'après la question 1, $\left(\frac{f(x)}{x}\right)' = \frac{-x^2 f(x) - f(x)}{x^2} = -\left(1 + \frac{1}{x^2}\right) f(x)$. En d'autres termes, $A : x \in \mathbb{R} \mapsto -\frac{f(x)}{x}$ est une primitive de $x \in \mathbb{R} \mapsto \left(1 + \frac{1}{x^2}\right) f(x)$. On peut étendre ses valeurs à la droite réelle achevée en posant $A(\pm\infty) = \lim_{x \rightarrow \pm\infty} A(x) = 0$. On obtient alors par le théorème fondamental du calcul (extension)

$$\left(1 + \frac{1}{x^2}\right) (1 - F(x)) \geq -(A(+\infty) - A(x)) = A(x) = \frac{f(x)}{x}.$$

On en déduit directement l'inégalité de gauche :

$$1 - F(x) \geq f(x) \frac{x^2}{(x^2 + 1)x} = f(x) \frac{x}{x^2 + 1}.$$

Equivalent de $1 - F(x)$ quand $x \rightarrow +\infty$.

5. Soit $x > 0$. L'encadrement que nous avons démontré peut être réécrit

$$\frac{x^2}{1+x^2} \leq \frac{1-F(x)}{\frac{f(x)}{x}} \leq 1,$$

et comme $\frac{x^2}{1+x^2} \rightarrow 1$ lorsque $x \rightarrow +\infty$, on obtient que $1-F(x) \sim \frac{f(x)}{x}$ quand $x \rightarrow +\infty$.

6. Pour tout $x \in [-\infty, +\infty]$ on a $|h(x)| = |x f(x)| = |x| f(x)$ puisque $f \geq 0$. Lorsque $x \in [0, +\infty]$, $|h(x)| = h(x)$, or d'après la question 1, h admet pour primitive $-f$, qui vaut 0 en $+\infty$ et $-\infty$ puis $(2\pi)^{-1/2}$ en 0. Par conséquent, (d'après l'extension théorème fondamental du calcul) l'intégrale de $|h|$ sur $[0, +\infty]$ existe, et vaut $(2\pi)^{-1/2}$. Or $|h|$ est paire, son intégrale sur $[-\infty, 0]$ existe donc aussi et vaut $(2\pi)^{-1/2}$. On en conclut que h est bien absolument intégrable sur $[-\infty, +\infty]$. En remarquant qu'elle est impaire, on obtient immédiatement que son intégrale vaut 0.

On est donc bien dans le cadre de l'exercice sur les queues de distributions. L'équivalent que nous avons trouvé implique bien que $1-F(x) = o\left(\frac{1}{x}\right)$ lorsque $x \rightarrow +\infty$. Il nous donne juste plus de précisions quant au comportement de $1-F(x)$ lorsque $x \rightarrow +\infty$ dans le cas spécifique de la loi Normale centrée réduite : on connaît sa vitesse de convergence.

Loi Normale générale.

7. Soit $x \in \mathbb{R}$. On remarque que $g(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$. Cela implique en particulier que

$$G(x) = \int_{-\infty}^x \frac{1}{\sigma} f\left(\frac{u-\mu}{\sigma}\right) du = \int_{-\infty}^{\frac{x-\mu}{\sigma}} f(u) du = F\left(\frac{x-\mu}{\sigma}\right).$$

On obtient alors l'encadrement suivant, pour tout $x > \mu$:

$$\frac{x-\mu}{1+\frac{(x-\mu)^2}{\sigma^2}} g(x) \leq 1-G(x) \leq \frac{\sigma^2 g(x)}{x-\mu}.$$

On en déduit que $1-G(x) \sim \frac{\sigma^2 g(x)}{x-\mu} \sim \sigma^2 \frac{g(x)}{x}$ lorsque $x \rightarrow +\infty$.

Références

Garnier, Josselin. 2018. *Aléatoire*. http://josselin-garnier.org/wp-content/uploads/2018/03/MAP311_2018.pdf.

Jourdain, Benjamin. 2016. *Probabilités et Statistique*. Ellipses. <http://cermics.enpc.fr/~jourdain/probastat/poly.pdf>.

Matheron, Georges. 1978. *Estimer et choisir: essai sur la pratique des probabilités*. Fontainebleau, France: Ecole nationale supérieure des mines de Paris. http://cg.ensmp.fr/bibliotheque/public/MATHERON_Ouvrage_00208.pdf.

Tebaldi, Claudia, and Reto Knutti. 2007. “The Use of the Multi-Model Ensemble in Probabilistic Climate Projections.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (1857): 2053–75. <https://doi.org/10.1098/rsta.2007.2076>.