

Probabilités I

STEP, MINES ParisTech*

26 septembre 2019 (#b161997)

Table des matières

Introduction	2
Exemple de la modélisation du climat	3
Plan du cours	3
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	10
Théorème de la continuité monotone	10
Remarque	10
Probabilité conditionnelle	10
Définition – Probabilité conditionnelle	11
Proposition	11

*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.

Exemple : approche bayésienne (subjective)	11
Formule des probabilités totales	11
Formule de Bayes	12
Exemple d'application TODO	12
Indépendance des événements	12
Définition – Indépendance de deux événements	12
Remarques	12
Proposition	12
Probabilité sur \mathbb{R}	13
Fonction de répartition	13
Définition – fonction de répartition	13
Théorème	13
Théorème	13
Remarque	14
Remarque	14
Exemple	14
Densités de probabilités	15
Définition – densité de probabilité	15
Remarques	16
Exemples de lois à densités	16
Exercices	18

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 de Garnier (2018) ou aux deux premiers chapitres 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 particulièrement 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 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 ultérieurement 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 par la suite 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 de ses ensembles mesurables.

Soit $\Omega = \mathbb{R}$ et $f : \Omega \rightarrow \mathbb{R}^+$ une fonction absolument 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 :

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

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 inférieur 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 (dénumbrable) $(A_n)_{n \in \mathbb{N}^*}$ d'éléments de \mathcal{A} **deux à deux disjoints**, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} \mathbb{P}(A_n). \quad (1)$$

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.

Une propriété est vraie *\mathbb{P} -presque-sûrement* (en abrégé \mathbb{P} -p.s.), si l'ensemble des $\omega \in \Omega$ pour lesquels elle est vraie est de probabilité égale à 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 : $\forall n \in \mathbb{N}^*, \forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}, \mathbb{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$.
5. Formule de Poincaré : $\forall n \in \mathbb{N}^*, \forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}$

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

Demonstration 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

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

Demonstration Exercice

■

Remarque

Dans le cas d'une suite décroissante, on a

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

ou version plus complète avec l'équivalence de la σ -additivité??? -
Dans les compléments

Lemme de Borel-Cantelli ??? ou alors à la fin - *Dans les suites de va*

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. L'outil qui va nous permettre d'introduire cette 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

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (2)$$

Cela définit une probabilité.

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. ■

Exemple : approche bayésienne (subjective)

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

$$P(A) = \sum_{n \in \mathbb{N}} P(A \cap B_n) = \sum_{n \in \mathbb{N}} P(A|B_n) P(B_n). \quad (3)$$

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

$$\forall i \in \{1, \dots, n\}, \mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\sum_{n \in \mathbb{N}} \mathbb{P}(A|B_n) \mathbb{P}(A)}. \quad (4)$$

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

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_i) \mathbb{P}(B_i)}{\mathbb{P}(A)}. \quad (5)$$

■

Exemple d'application TODO

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.

Définition – Indépendance de deux événements

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B). \quad (6)$$

Remarques

- La probabilité de voir A réalisé ne dépend pas de la réalisation de B , et réciproquement.
- Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, alors

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}(B) \Leftrightarrow \mathbb{P}(A) = \mathbb{P}(A|B) \Leftrightarrow \mathbb{P}(B) = \mathbb{P}(B|A). \quad (7)$$

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) .

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} est la fonction

$$F(x) = \mathbb{P}(]-\infty, x]), \quad x \in \mathbb{R}. \quad (8)$$

Théorème

La fonction de répartition F caractérise la probabilité.

Nous démontrerons ce résultat ultérieurement.

Théorème

Une fonction F est la fonction de répartition d'une unique probabilité \mathbb{P} sur \mathbb{R} muni de la tribu des ensembles mesurables si et seulement si elle vérifie les trois conditions suivantes :

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

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 plus tard. ■

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 $x < y$:

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

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} au $2^{[0, 1]}$ éléments de $\mathcal{P}(\mathbb{R})$ 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. $\mathcal{P}(\mathbb{R})$ 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 $A \in \mathcal{B}_{\mathbb{R}}$

$$\mathbb{P}(A) = \begin{cases} 1 & \text{si } a \in A, \\ 0 & \text{sinon.} \end{cases} \quad (9)$$

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

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ \sum_{i=0}^{\lfloor x \rfloor} q_i & \text{sinon,} \end{cases} \quad (10)$$

où $\lfloor \cdot \rfloor$ désigne la partie entière.

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 $\Delta F(x) = F(x) - F(x^-)$:

$$F(x) = \sum_{\substack{y \in E \\ y \leq x}} q_i.$$

Notons aussi que l’ensemble E , bien qu’au plus dénombrable, peut tout à fait être partout dense dans \mathbb{R} , par exemple $E = \mathbb{Q}$: si alors $q_i > 0$ pour tout $i \in \mathbb{Q}$, la fonction F est discontinue en tout rationnel.

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

Densités de probabilités

Définition – densité de probabilité

Une fonction réelle f sur \mathbb{R} est une *densité de probabilité* (ou plus simplement *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'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, nous

avons “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). Nous avons donc une probabilité nulle de tomber exactement en un point x fixé à l'avance. On vérifie aisément que sa fonction de répartition est

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

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

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

et de fonction de répartition

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

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. Nous pouvons 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.

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

Exercices

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>.

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>.