

Probabilités I

STEP, Mines Paristech*

4 septembre 2019 (#b33c81a)

Table des matières

Introduction	2
Historique	3
Plan du cours	4
Probabilités des évènements	4
Phénomènes aléatoires et évènements	4
Exemples	4
L'espace fondamental	4
Exemples	5
Evènement	5
Exemples	5
Correspondance entre opérations logiques et ensemblistes	5
Tribu	6
Exemples	6
Notion de densité de probabilité	6
Remarque	7
Probabilité	7
Probabilité	7
Définition	8
Propriétés élémentaires	8
Théorème de la continuité monotone	8
Remarque	9
Lemme de Borel-Cantelli ??? ou alors à la fin	9
Probabilité conditionnelle	9
Probabilité conditionnelle	9
Proposition	9
Exemple : approche bayésienne (subjective)	10
Formule des probabilités totales	10
Formule de Bayes	10
Exemple d'application TODO	10

*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.

Indépendance des évènements	10
Indépendance de deux évènements	10
Remarque	11
Proposition	11
Probabilité sur \mathbb{R}	11
Fonction de répartition	11
Définition - fonction de répartition	11
Exemple	11
Proposition	11
Remarque	12
Théorème	12
Remarque	12
Exemple	13
Densités de probabilités	13
Définition	13
Remarques	14
TODO : ajouts d'exemples, exercices	14
Variables aléatoires absolument continues -> plutôt en II ?	14
Exercices	15

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. . .
- 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 classe préparatoire, 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 aux deux premiers chapitres de Jourdain (2016) pour une éventuelle mise à niveau.

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.

Plan du cours

Le cours est organisé en 5 amphis : Description des notions importantes du poly : variables aléatoires, conditionnement et indépendance, théorèmes limites, simulation Ouverture vers les stats/ML

TODO à reprendre/développer

Probabilités des évènements

Phénomènes aléatoires et évènements

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 du prix d'un actif financier 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.

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, $\Omega = \{p, f\}$
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 du prix d'un actif financier sur un intervalle de temps $[t_1, t_2]$, $\Omega = C([t_1, t_2], \mathbb{R}^+)$, ensemble des application continues de $[t_1, t_2]$ dans \mathbb{R}^+

Evènement

Un *évènement* est une propriété qui est vérifiée ou non une fois l'expérience réalisée. On identifie un évènement 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, $A = \{p\}$
2. Lancé de dés, $A = \{1, 3, 5\}$
3. Durée de vie d'une ampoule électrique, $A = [t_1, t_2]$
4. Température demain à 12h au sommet de la tour Eiffel (en degrés Kelvin), $A = [T_1, T_2] \cup [T_3, T_4]$
5. Evolution du prix d'un actif financier sur un intervalle de temps $[t_1, t_2]$, $A = \{f \in C([t_1, t_2], \mathbb{R}^+), \|f - g\|_\infty \leq a\}$, où $g \in C([t_1, t_2], \mathbb{R}^+)$, $a \in \mathbb{R}^+$

Les évènements étant des ensembles, les opérations ensemblistes classiques admettent une interprétation probabiliste :

Correspondance entre opérations logiques et ensemblistes

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

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

Tribu

Une *tribu* \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. 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

La nouveauté majeure de ce cours par rapport au programme des classes préparatoires 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 de \mathbb{R} via la notion de *densité de probabilité*.

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'évènements **disjoints** de \mathcal{A} , on a en utilisant le théorème de convergence monotone à la suite croissante de fonctions $1_{\{\bigcup_{n=1}^m A_n\}} = \sum_{n=1}^m 1_{A_n}$:

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

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 reste 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

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) d'éléments de \mathcal{A} **deux à deux disjoints**, on a

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n). \quad (1)$$

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

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.

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, $\forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}, \mathbb{P}(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathbb{P}(A_i)$
5. Formule de Poincaré $\forall (A_i)_{1 \leq i \leq n} \in \mathcal{A}$

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

Démonstration Exercice ■

Théorème de la continuité monotone

Dans le cas d'une suite $(A_n) \in \mathcal{A}$ croissante, on a

$$\mathbb{P}\left(\bigcup_n A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$$

Démonstration Exercice ■

Remarque

Dans le cas d'une suite décroissante, on a

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$$

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

Lemme de Borel-Cantelli ??? ou alors à la fin

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:

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. En effet, on a :

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

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A).$$

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 $P(\cdot|B)$ proviennent des mêmes propriétés pour \mathbb{P} et des remarques suivantes : $\Omega \cap B = B$, et $(\bigcup_n A_n) \cap B = \bigcup_n (A_n \cap B)$. 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) une partition finie ou dénombrable d'événements de Ω (i.e. telle que $\bigcup_n B_n = \Omega$ et les B_n sont deux-à-deux disjoints), telle que $\mathbb{P}(B_n) > 0$ pour tout n . Pour tout $A \in \mathcal{A}$, on a alors

$$P(A) = \sum_n P(A \bigcup_n B_n) = \sum_n P(A|B_n)P(B_n). \quad (3)$$

Démonstration Nous avons $A = \bigcup_n (A \cap B_n)$. 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

$$\forall i, \mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_n \mathbb{P}(A|B_n)\mathbb{P}(A)} \quad (4)$$

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.

Indépendance de deux événements

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

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

Remarque

- 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 évènements 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 $\Omega = \mathbb{R}$ (\mathbb{R}^d ?) et où la tribu \mathcal{A} est la tribu borélienne (**ou de Lebesgue cf CI II**) $\mathcal{B}_{\mathbb{R}}$ engendrée par les ouverts, ou par les fermés, ou par les intervalles de la forme $] -\infty, a]$ pour $a \in \mathbb{Q}$ (cf CI II).

Définition - fonction de répartition

La *fonction de répartition* de la probabilité \mathbb{P} sur \mathbb{R} est la fonction

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

Exemple

à développer \mathbb{P} mesure de Dirac $\rightarrow F$ fonction de Heavyside

Proposition

La fonction de répartition F caractérise la probabilité \mathbb{P} sur \mathbb{R} , et 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 Se reporter à Jacod and Protter (2003)¹ pour la caractérisation.

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 $+\infty$). ■

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.

La proposition admet une réciproque que nous admettrons. On se reportera à Jacod and Protter (2003) pour une démonstration.

Théorème

Si F est une fonction réelle sur \mathbb{R} qui vérifie les trois conditions de la proposition, c’est la fonction de répartition d’une (unique) probabilité \mathbb{P} sur \mathbb{R} munie de la tribu borélienne $\mathcal{B}_{\mathbb{R}}$. On ne peut pas, en général, définir \mathbb{P} sur la tribu $\mathcal{P}(\mathbb{R})$ de toutes les parties de \mathbb{R} .

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. Sinon, cela reviendrait à prendre (sans le dire) la tribu $\mathcal{A} = \mathcal{P}(\mathbb{R})$ et il n’existerait que très peu de probabilités sur \mathbb{R} , à savoir les probabilités discrètes que l’on décrit rapidement ci-dessous :

1. on évoquera ça dans calcul intégral V (applis théorie mesure): la fonction de répartition est de variation bornée (car croissante) donc elle à une dérivée faible (ou au sens des distributions) qui est une mesure (positive).

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

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

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 $\triangle F(x) = F(x) - F(x-)$:

$$F(x) = \sum_{y \in E; y \leq x} q_y.$$

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

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. A 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 densités : c'est le cas des probabilités discrètes données en exemple ci-dessus. Il existe des cas "mixtes" : soit d'une part f une fonction positive intégrable et d'autre part une partie finie ou dénombrable E de \mathbb{R} et des indices $p_i > 0$ indicés par $i \in E$, tels que :

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

Alors la fonction

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

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

Remarques

- 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 $g(x) = f(x) + 1$ si $x \in E$, où E est un ensemble négligeable, alors g est encore une densité de \mathbb{P} .
- 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) :

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

TODO : ajouts d'exemples, exercices

Variables aléatoires absolument continues -> plutôt en II ?

- Fonction de répartition
- Variables aléatoires à densité
- Exemples
- Lien avec la définition axiomatique ? (vérification de la σ -additivité dans le cas d'un univers réel, cf note SB)

aura-t-on le temps ? Sinon se limiter à une ouverture vers les v.a. à valeurs réelles

Exercices

exos sur les probas discrètes

Jacod, J., and P. Protter. 2003. *L'essentiel En Théorie Des Probabilités*. Cassini. <https://hal.archives-ouvertes.fr/hal-00104956>.

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