

Table des matières

Probabilité discrète	3
Probabilité sur les espaces discrets :	3
Symétrie du hasard et probabilité uniforme :	3
Loi de probabilité sur un ensemble fini :	5
Probabilité sur un état d'espace dénombrable :	5
Probabilité conditionnelle et indépendance :	7
Théorème de Bayes :	7
Indépendance de deux évènements aléatoires :	8
Variables aléatoires discrètes :	9
Élément aléatoire et loi de probabilité :	9
Espérance d'une variable aléatoire :	10
Variance d'une variable aléatoire :	12
Lois de probabilité sur \mathbb{N} :	12
Fonction génératrice :	13
Loi binomiale :	13
Loi multinomiale :	14
Loi géométrique :	14
Loi binomiale négative :	14
Loi de Poisson :	15
Loi de probabilité à densité sur \mathbb{R} :	16
Théorie de l'intégration	16
1- Rappels sur l'intégrale de Riemann:	16
Définition :	16
L'intégrale de Lebesgue :	17
Les espaces $L^p\mathbb{R}$, $p \geq 1$:	24
3-2 L'espace $L^\infty\mathbb{R}$:	24
Quelques théorèmes pratiques du calcul intégral.	25
Loi de probabilité à densité sur \mathbb{R} :	28
Loi de probabilité uniforme :	28
Caractérisation d'une ddp sur \mathbb{R} :	29
Exemple de la densité gaussienne $\mathcal{N}(m, \sigma)$	30
Exemple de la densité Gamma $\Gamma(k, \theta)$	31
Espaces de probabilité continus construits sur \mathbb{R}	31

Probabilité discrète

Probabilité sur les espaces discrets :

Symétrie du hasard et probabilité uniforme :

Notion d'espace d'état :

L'état d'un système est un objet mathématique contenant toutes les informations qui le décrivent. En dynamique par exemple, un point est défini par le couple position-vitesse. La connaissance de l'état d'un système dans une théorie physique donnée permet de connaître complètement son comportement. Les variables d'état d'un système sont les variables qui le déterminent, comme les deux variables d'état position et vitesse pour le point en dynamique. Si elles sont numériques¹ et en nombre fini d , l'ensemble des états du système que nous noterons Ω est un sous ensemble de \mathbb{R}^d . Il est appelé espace d'états. Tout sous ensemble de cet espace, est appelé évènement.

Dans le cadre d'une théorie déterministe, l'état d'un système à un instant donné est un élément de l'espace d'états du système. Dans une modélisation aléatoire, l'état du système n'est pas défini : le comportement du système à un instant donné est décrit par une loi de probabilité, c'est-à-dire par des nombres décrivant la probabilité qu'a le système de se trouver dans un tel ou tel état. Le cas le plus simple à modéliser est celui où tous les états du système sont équiprobables. Comme nous le verrons ci-dessous, le calcul de la probabilité d'un évènement dans ce cas se ramène à une division du nombre d'états de l'évènement par le nombre de l'ensembles de tous les états. C'est le cas de la symétrie du hasard.

Exemples :

Le lancer d'un dé :

Considérons l'expérience aléatoire du lancer de dé à 6 faces. Le résultat de l'expérience est un entier de $\Omega = \{1,2,3,4,5,6\}$. Tout évènement provenant de l'expérience est un sous ensemble de E . Par exemple, l'évènement « le résultat est supérieur à 3 » est représenté par $ES3 = \{4,5,6\}$. L'évènement « le résultat est inférieur à 3 » est représenté par $EI3 = \{1,2\}$. L'espace d'état est l'ensemble Ω .

Un jeu de lancer de dé à deux joueurs :

Considérons maintenant un jeu à 2 joueurs, Le bleu et le rouge. Chaque joueur mise une somme unité et lance un dé à 6 faces de sa couleur. Le joueur ayant obtenu le résultat le plus élevé remporte la mise. Chacun reprend sa mise en cas d'égalité de résultat. L'espace d'état est donc $\Omega = \{(x, y) \in \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}\}$. Son cardinal est de 36. Un état (x, y) représente l'issue où le bleu a obtenu le résultat x et le rouge le résultat y . L'évènement « le bleu gagne » est représenté par l'ensemble des 15 états :

$$BG = \{(2,1), (3,1), (3,2), (4,1), (4,2), (4,3), (5,1), (5,2), (5,3), (5,4), (6,1), (6,2), (6,3), (6,4), (6,5)\}$$

L'évènement « le rouge gagne » est représenté par l'ensemble des 15 états :

$$RG = \{(1,2), (1,3), (2,3), (1,4), (2,4), (3,4), (1,5), (2,5), (3,5), (4,5), (1,6), (2,6), (3,6), (4,6), (5,6)\}$$

Et l'évènement chacun reprend sa mise est représenté par l'ensemble des 6 états :

$$EG = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$$

¹ Il arrive que les variables d'état soient qualitatives pour décrire par exemple l'état de marche ou de panne d'un système.

On peut considérer qu'il y a deux variables d'état, le résultat obtenu par le bleu et celui obtenu par le rouge. On remarque particulièrement que les états (1,2) et (2,1) se distinguent entre eux, ce qui n'aurait pas été le cas si la couleur des deux dés était la même.

La probabilité de gagner du bleu, aussi bien que celle du rouge est de $\frac{15}{36}$.

La probabilité que chacun reprenne sa mise est de $\frac{1}{6}$.

L'évènement « bleu gagne ou rouge gagne » est représenté par l'ensemble $BG \cup RG$. La probabilité de son avènement est de $\frac{5}{6}$, tandis que la probabilité de l'évènement « personne ne gagne » qui n'est autre que l'évènement « chacun reprend sa mise » est de $\frac{1}{6}$. Ce dernier évènement peut aussi être décrit par « Non (bleu gagne ou rouge gagne) ». Il est représenté par l'ensemble $(BG \cup RG)^c$, la complémentarité étant relative à l'espace des états $\Omega = BG \cup RG \cup EG$.

D'une manière générale, les locutions logiques sont traduites en langage mathématique via une expression ensembliste. Si ω modélise le résultat d'une expérience aléatoire dont Ω est son espace d'états, nous pouvons dresser le tableau suivant :

Logique	Notation ensembliste	Commentaire
Etat observable	ω	
Evènement toujours observable	Ω	Ensemble des états observables
Evènement jamais observable	\emptyset	Ensemble vide
Evènement aléatoire « A »	$A \subset \Omega$	
Opération « A ou B »	$A \cup B$	
Opération « A et B »	$A \cap B$	
Opérateur « Non A »	A^c	

Cadre Mathématique : Algèbre de Boole

Définition :

Soit Ω un ensemble. Un ensemble $\mathcal{A} \subset \mathcal{P}(\Omega)$ est une algèbre de Boole s'il vérifie :

- $\Omega \in \mathcal{A}$,
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
- $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

Exercice :

Montrer que si $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$

Probabilité uniforme :

Définition : (Loi de probabilité uniforme sur un ensemble fini Ω)

Soit Ω un espace d'états fini associée à une expérience aléatoire. On appelle loi de probabilité uniforme sur Ω , l'application :

$$\begin{aligned} P: \mathcal{P}(\Omega) &\rightarrow [0,1] \\ A &\mapsto \frac{\text{card}(A)}{\text{card}(\Omega)} \end{aligned}$$

Où $\text{card}(A)$ est le cardinal de A et $\text{card}(\Omega)$ est le cardinal de Ω .

Loi de probabilité sur un ensemble fini :

Définition :

Soit Ω un ensemble fini. On appelle loi de probabilité sur Ω une application P de $\mathcal{P}(\Omega)$ dans $[0,1]$ vérifiant :

- a. $P(\Omega) = 1$, (masse unité)
- b. Soient A et B deux éléments de Ω . Si $A \cap B = \emptyset$ alors $P(A \cup B) = P(A) + P(B)$ (additivité)

On peut facilement montrer que a. et b. sont bien vérifiés par la loi de probabilité uniforme sur un ensemble fini Ω .

Remarque :

$$P(\emptyset) = 0.$$

Théorème :

Soit Ω un ensemble fini et P une probabilité sur $\mathcal{P}(\Omega)$. On a :

- a. $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$,
- b. $\forall A \subset \Omega, P(A) = \sum_{\omega \in A} P(\{\omega\})$

Réciproquement, soit l'application $\begin{matrix} \Omega & \rightarrow & [0,1] \\ \omega & \mapsto & p_\omega \end{matrix}$ telle que $\sum_{\omega \in \Omega} p_\omega = 1$, alors il existe une et une seule probabilité P sur $\mathcal{P}(\Omega)$ telle que $\forall \omega \subset \Omega, P(\{\omega\}) = p_\omega$.

Commentaire :

Ce théorème montre comment se construisent les probabilités sur un ensemble d'états fini à partir de la donnée de la probabilité sur les états.

Exercice : Paradoxe du Grand Duc de Toscane

On dispose de trois dés similaires qu'on jette simultanément.

1. Donner toutes les combinaisons pour avoir une somme totale égale à 9,
2. Donner toutes les combinaisons pour avoir une somme totale égale à 10,
3. Donner la probabilité d'avoir une somme totale égale à 9 et celle d'avoir une somme totale égale à 10.

Exercice : Formule de Poincaré

Soit P une probabilité sur une algèbre de Boole \mathcal{A} de parties d'un ensemble.

1. Montrer que $\forall (A, B) \in \mathcal{A} \times \mathcal{A}, P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
2. Etendre cette formule par récurrence au cas de n ensembles A_1, \dots, A_n

Probabilité sur un état d'espace dénombrable :

Evolution du cadre mathématique :

Définition : Tribu

Soit Ω un ensemble. On appelle σ –algèbre de Boole ou tribu, un ensemble $\mathcal{T} \subset \mathcal{P}(\Omega)$ qui vérifie :

- a. $\Omega \in \mathcal{T}$,
- b. $A \in \mathcal{T} \Rightarrow A^c \in \mathcal{T}$,

et qui de plus possède une des propriétés équivalentes suivantes :

- c1. Soit $(A_n)_{n \in \mathbb{N}}$ une suite croissante de \mathcal{T} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$
- c2. Soit $(A_n)_{n \in \mathbb{N}}$ une suite quelconque de \mathcal{T} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{T}$

Définition : Probabilité sur une tribu

Soit \mathcal{T} une tribu d'un ensemble Ω . Une probabilité sur \mathcal{T} est une application de \mathcal{T} dans $[0,1]$ vérifiant :

- a. $P(\Omega) = 1$,
- b. Soient A et B deux éléments de \mathcal{T} . Si $A \cap B = \emptyset$ alors $P(A \cup B) = P(A) + P(B)$
- c. Soit $(A_n)_{n \in \mathbb{N}}$ une suite croissante de \mathcal{T} , alors $\lim_{n \rightarrow \infty} P(A_n) = P(\bigcup_{n \in \mathbb{N}} A_n)$ (continuité monotone croissante)

Remarque :

Les propriétés b. et c. entraînent la σ -additivité :

Soit $(A_n)_{n \in \mathbb{N}}$ une suite de \mathcal{T} d'ensembles disjoints deux à deux, alors $P(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$

Exercice : Montrer en passant à la complémentarité que la continuité monotone croissante est équivalente à la continuité monotone décroissante :

- d. Soit $(A_n)_{n \in \mathbb{N}}$ une suite décroissante de \mathcal{T} , alors $\lim_{n \rightarrow \infty} P(A_n) = P(\bigcap_{n \in \mathbb{N}} A_n)$

Définition :

On appelle espace de probabilité la donnée d'un espace d'état Ω , d'une tribu \mathcal{T} de Ω et d'une probabilité P sur \mathcal{A} .

Théorème :

Soit Ω un espace d'états dénombrable et P une probabilité sur $\mathcal{P}(\Omega)$. On a :

- a. $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$,
- b. $\forall A \subset \Omega, P(A) = \sum_{\omega \in A} P(\{\omega\})$

Réciproquement, soit l'application $\begin{matrix} \Omega & \rightarrow & [0,1] \\ \omega & \mapsto & p_\omega \end{matrix}$ telle que $\sum_{\omega \in \Omega} p_\omega = 1$, alors il existe une et une seule probabilité P sur $\mathcal{P}(\Omega)$ telle que $\forall \omega \subset \Omega, P(\{\omega\}) = p_\omega$.

Exemple de la poule à 3 joueurs :

On considère 3 joueurs A, B et C qui se livrent une série de duels. On suppose que les duels sont équilibrés (lancer d'une pièce à pile ou face avec une probabilité 0,5). Les joueurs A et B commencent et celui qui gagne affronte le joueur C. Et le jeu continue jusqu'à ce que l'un des joueurs gagne deux duels successifs. Il est alors déclaré vainqueur de la poule. Dans un premier temps, on suppose que le nombre de duels est limité à 4 lancers.

1. Quelle est la probabilité de chacun de gagner ?
2. Quelle est la probabilité que le jeu s'arrête sans déclaration de vainqueur ?

Réponse :

L'espace d'états est $\Omega_4 = \{AA, ACC, ACBB, ACBA, BB, BCC, BCBA, BCAB\}$ avec les probabilités $P(AA) = P(BB) = \frac{1}{4}$, tandis que $P(ACC) = P(BCC) = \frac{1}{8}$ et après les quatre duels, on obtient

$P(ACBB) = P(ACBA) = P(BCAA) = P(BCAB) = \frac{1}{16}$. Ceci fait que la probabilité de gain de A est égale à celle de B égale à $\frac{1}{4} + \frac{1}{16} = \frac{5}{16}$ et celle de C est égale à $\frac{2}{8}$. La probabilité que le jeu s'arrête sans déclaration de vainqueur est de $\frac{1}{8}$.

Supposons maintenant que nous voulons modéliser le jeu sans fixer d'étape d'arrêt. L'espace d'états devient alors infini. Les états à n duels ont une probabilité de $\frac{1}{2^n}$ et la probabilité que le jeu ne s'arrête jamais est donc nulle !!!

Exercice :

Dans le jeu de la poule à trois joueurs sans fixer une valeur maximale pour les duels :

1. Calculer la probabilité que le jeu s'arrête après n duels
2. Calculer la probabilité de gain de chacun des joueurs

Probabilité conditionnelle et indépendance :

Définition : Probabilité conditionnelle

Soit (Ω, \mathcal{T}, P) un espace de probabilité et $B \in \mathcal{T}$ un évènement aléatoire de probabilité non nulle. On appelle probabilité conditionnelle sachant B la nouvelle probabilité définie sur \mathcal{T} par

$$\begin{aligned} P(\cdot | B) : \mathcal{T} &\rightarrow [0,1] \\ A &\mapsto P(A|B) = \frac{P(A \cap B)}{P(B)} \end{aligned}$$

Exemple :

Considérons l'exemple des joueurs bleu et rouge. La probabilité que le bleu gagne sachant que le rouge a obtenu un 6 est calculée comme suit :

Le bleu gagne est l'évènement

$$BG = \{(2,1), (3,1), (3,2), (4,1), (4,2), (4,3), (5,1), (5,2), (5,3), (5,4), (6,1), (6,2), (6,3), (6,4), (6,5)\}$$

L'évènement le rouge a obtenu 6 est l'évènement :

$$R6 = \{(1,6), (2,6), (3,6), (4,6), (5,6), (6,6)\}$$

$$\text{On a : } \begin{cases} P(R6) = \frac{1}{6} \neq 0 \\ BG \cap R6 = \emptyset \Rightarrow P(BG \cap R6) = 0 \end{cases} \Rightarrow P(BG|R6) = 0$$

Exercice :

Calculer dans le cadre de l'exercice relatif au paradoxe du Grand Duc de Toscane, la probabilité d'avoir 9 et 10 sachant que deux dés ont eu des résultats semblables.

Théorème de Bayes :

Soit (Ω, \mathcal{T}, P) un espace de probabilité et $(B_n)_{n \geq 0}$ une partition finie ou dénombrable de Ω en évènements aléatoires, autrement dit :

- a. $\forall n \geq 0, B_n \in \mathcal{T}$,
- b. $n \neq p \Rightarrow B_n \cap B_p = \emptyset$,
- c. $\bigcup_n B_n = \Omega$

Alors on a :

$$\begin{cases} \forall A \in \mathcal{T} & P(A) = \sum_n P(A|B_n) P(B_n) \\ & P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_n P(A|B_n) P(B_n)} \end{cases}$$

Preuve :

On a par la formule de σ -additivité de P que $P(A) = \sum_n P(A \cap B_n) = \sum_n P(A|B_n) P(B_n)$. Par ailleurs, on a aussi $P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_n P(A|B_n)P(B_n)}$

■

Ce théorème important permet ainsi de formuler des hypothèses statistiques sur une propriété inconnue d'un système B_i à partir d'une observation inconnue A et d'une analyse à priori du système (connaissance des probabilités de B_n et de A sachant B_n).

Exemple :

Considérons encore le jeu des dés rouge et bleu. Soit B_p respectivement B_i l'événement le dé bleu est paire respectivement le dé bleu impaire. La famille (B_p, B_i) est une partition de l'espace des états :

$$B_p = \{(2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\}$$

Et

$$B_i = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6)\}$$

La probabilité de chacun des deux évènements est de $\frac{1}{2}$.

$$\text{Soit l'évènement } RG. \text{ On a } P(B_p|RG) = \frac{P(B_p \cap RG)}{P(RG)} = \frac{P\{(2,3), (2,4), (2,5), (2,6), (4,5), (4,6)\}}{\frac{5}{12}} = \frac{\frac{6}{12}}{\frac{5}{12}} = \frac{6}{5} = \frac{2}{5}$$

$$\text{Par ailleurs, on a : } \begin{cases} P(B_p) = P(B_i) = \frac{1}{2} \\ P(RG|B_p) = \frac{P(B_p \cap RG)}{P(B_p)} = \frac{\frac{6}{12}}{\frac{1}{2}} = \frac{1}{3} \\ P(RG|B_i) = \frac{P(B_i \cap RG)}{P(B_i)} = \frac{P\{(1,2), (1,3), (1,4), (1,5), (1,6), (3,4), (3,5), (3,6), (5,6)\}}{\frac{1}{2}} = \frac{\frac{6}{12}}{\frac{1}{2}} = \frac{1}{2} \end{cases} \Rightarrow$$

$$P(B_p|RG) = \frac{\frac{1}{3} \times \frac{1}{2}}{\frac{1}{3} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2}} = \frac{2}{5}$$

Indépendance de deux évènements aléatoires :

L'indépendance de deux évènements aléatoires est une notion fondamentale. Sa définition est une conséquence directe de la définition d'une propriété conditionnelle. Deux évènements aléatoires sont indépendants si la probabilité de l'un n'est pas changée par la connaissance de l'occurrence de l'autre.

Définition :

On dit que deux évènements aléatoires A et B d'une espace de probabilité (Ω, \mathcal{F}, P) sont indépendants si $P(A \cap B) = P(A)P(B)$.

Théorème :

A et B sont indépendants si et seulement si A et B^c sont indépendants.

Preuve :

On a $A = (A \cap B) \cup (A \cap B^c)$ avec $(A \cap B) \cap (A \cap B^c) = \emptyset \Rightarrow$

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A)(1 - P(B)) = P(A)P(B^c)$$

■

Définition : Indépendance de n évènements aléatoires

On dit que n évènements aléatoires $(A_i)_{1 \leq i \leq n}$ d'un espace de probabilité sont indépendants dans leur ensemble si quelle que soit la sous-famille $(A_i)_{i \in I}$ extraite de la famille initiale on a :

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

Remarque :

Les évènements peuvent être indépendants deux à deux sans être indépendants dans leur ensemble. Considérons par exemple l'espace de probabilité $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$ muni de la probabilité uniforme, alors les évènements $A = \{(1,0), (1,1)\}$, $B = \{(0,1), (1,1)\}$ et $C = \{(0,0), (1,1)\}$ sont indépendants 2 à 2, mais ne sont pas indépendants dans leur ensemble. En effet, on a :

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$P(A \cap B) = P(B \cap C) = P(A \cap C) = \frac{1}{4}$$

$$P(A \cap B \cap C) = \frac{1}{4}$$

Variables aléatoires discrètes :

Élément aléatoire et loi de probabilité :

Définitions et notations:

- On appelle élément aléatoire toute fonction X de Ω dans un ensemble de valeurs E .
- Soit $F \subset E$. L'évènement aléatoire $\{\omega \in \Omega / X(\omega) \in F\}$ est conventionnellement noté $(X \in F)$. De même la notation $(X \in F, Y \in G)$ est une notation conventionnelle de $\{\omega \in \Omega / X(\omega) \in F\} \cap \{\omega \in \Omega / Y(\omega) \in G\}$
- Soit X un élément aléatoire défini sur l'espace de probabilité fini ou dénombrable $(\Omega, \mathcal{P}(\Omega), P)$ et à valeurs dans E . L'application :

$$\begin{aligned} P_X: \mathcal{P}(E) &\rightarrow [0,1] \\ A &\mapsto P(X \in A) \end{aligned}$$

est une probabilité sur $(E, \mathcal{P}(E))$ appelée loi de probabilité de X et on notera $X \sim P_X$ pour indiquer que P_X est la loi de X .

- d. Deux éléments aléatoires qui ne diffèrent que sur un ensemble d'états de probabilité nulle sont dits presque sûrement égaux et sont confondus en pratique.

Exemple :

Soit Ω l'ensemble des individus d'une population quelconque. La fonction

$$\begin{aligned} AGE: \Omega &\rightarrow \mathbb{R}_+ \\ \omega &\mapsto AGE(\omega) \end{aligned}$$

est un élément aléatoire.

Définition :

On dit que deux éléments aléatoires X_1 respectivement X_2 définis sur l'espace de probabilité $(\Omega, \mathcal{P}(\Omega), P)$ et à valeurs dans E_1 respectivement E_2 sont indépendants si pour tout $F_1 \subset E_1$ respectivement $F_2 \subset E_2$, $P(X_1 \in F_1, X_2 \in F_2) = P(X_1 \in F_1)P(X_2 \in F_2)$

Exemple :

Les éléments aléatoires AGE et Genre sont indépendants

Espérance d'une variable aléatoire :

Définition :

L'espérance d'une variable aléatoire réelle X définie sur un espace de probabilité $(\Omega, \mathcal{P}(\Omega), P)$ fini est la moyenne pondérée

$$E(X) = \sum_{\omega \in \Omega} P(\omega)X(\omega) = \sum_x P(X = x)x$$

Théorème :

Soit X une variable aléatoire réelle et f une fonction de la variable réelle. On définit la variable aléatoire $f(X)$ par $f(X)(\omega) = f(X(\omega))$. On a alors $E(f(X)) = \sum_{\omega \in \Omega} P(X = x)f(x)$

Remarques :

- Si $X = (X_1, X_2, \dots, X_d)$ est un vecteur aléatoire à valeur dans \mathbb{R}^d , la même définition permet d'obtenir $E(X)$ comme un vecteur de \mathbb{R}^d dont les composantes sont les espérances des variables aléatoires X_1, X_2, \dots, X_d . Autrement dit, $E(X) = (E(X_1), E(X_2), \dots, E(X_d))$.
- Si X est une variable aléatoire, respectivement vecteur aléatoire, définie sur un espace d'état Ω dénombrable, et si la série, respectivement les séries, généralisant la définition de l'espérance sont absolument convergentes, leurs sommes sont par définition l'espérance de X . On écrit dans ce cas $E(|X|) < +\infty$ ou $E(X) \in L^1(\Omega, \mathcal{P}(\Omega), P)$.

Théorème :

Soit X une variable aléatoire définie sur un espace de probabilité fini ou dénombrable $(\Omega, \mathcal{P}(\Omega), P)$. On a :

- a. Si $E(|X|) < +\infty$ et $X \geq 0$, alors $E(X) \geq 0$,
- b. Si X et Y sont dans $L^1(\Omega, \mathcal{P}(\Omega), P)$, et si $X \leq Y$ alors $E(X) \leq E(Y)$
- c. On a en particulier $|E(X)| \leq E(|X|)$

Théorème :

Soient X et Y deux variables aléatoires définies sur un espace de probabilité fini ou dénombrable $(\Omega, \mathcal{P}(\Omega), P)$. Si elles sont dans $L^1(\Omega, \mathcal{P}(\Omega), P)$, alors $\forall (a, b) \in \mathbb{R}^2$ $aX + bY \in L^1(\Omega, \mathcal{P}(\Omega), P)$ et $E(aX + bY) = aE(X) + bE(Y)$

Théorème : Inégalité de Markov

Soit X une variable aléatoire définie sur un espace de probabilité fini ou dénombrable $(\Omega, \mathcal{P}(\Omega), P)$ telle que $E(|X|) < +\infty$, alors $\forall \varepsilon > 0, P(X \geq \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$

Preuve :

On considère la variable aléatoire $1_{|X| \geq \varepsilon}$, égale à 1 si $|X(\omega)| \geq \varepsilon$, 0 sinon. On a $E(1_{|X| \geq \varepsilon}) = P(|X| \geq \varepsilon)$ et $\varepsilon 1_{|X| \geq \varepsilon} \leq X \Rightarrow E(\varepsilon 1_{|X| \geq \varepsilon}) = \varepsilon E(1_{|X| \geq \varepsilon}) = \varepsilon P(|X| \geq \varepsilon) \leq E(X)$

■

Théorème : Espérance du produit de variables aléatoires

Soient X et Y des variables aléatoires indépendantes telles que X , Y et XY admettent une espérance alors $E(XY) = E(X)E(Y)$.

Preuve :

On note E l'ensemble des valeurs de X et F l'ensemble des valeurs de Y . On a

$$\begin{aligned} E(XY) &= \sum_{\omega \in \Omega} X(\omega)Y(\omega)P(\omega) = \sum_{x \in E} \sum_{y \in F} P(X = x, Y = y)xy \\ &= \sum_{x \in E} \sum_{y \in F} P(X = x)P(Y = y)xy \\ &= E(X)E(Y) \end{aligned}$$

Exercice :

Quelle est la loi de probabilité et l'espérance de la variable aléatoire n égale au nombre de duels nécessaires pour achever le jeu de la poule à trois joueurs.

Exercice : Problème des parties du chevalier de Méré

Dans un jeu de pile ou face, chaque joueur mise une unité. Un joueur choisit pile et l'autre face. Le premier joueur qui obtient trois fois son choix gagne la partie et emporte les deux unités.

1. Soit N le nombre de lancers de pièce. Quelle est sa loi de probabilité ?
2. Le jeu doit s'arrêter au bout de trois lancers. Le joueur « pile » a obtenu deux fois son choix et le joueur « face » a obtenu une fois le sien. Comment répartir équitablement la mise entre les deux joueurs ?

Variance d'une variable aléatoire :

Définition :

Soit X une variable aléatoire réelle telle que $E(X^2) < +\infty$. On définit la variance de X par $Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$. On a $Var(X) \geq 0$ et si $Var(X) = 0$, alors X est presque sûrement constante : $X = E(X)$.

Définition :

Soit X une variable aléatoire réelle telle que $E(X^2) < +\infty$. On définit l'écart type de X noté σ_X par $\sigma_X = \sqrt{Var(X)}$.

Théorème :

Soient X et Y deux variables aléatoires du second ordre indépendantes. On a $Var(X + Y) = Var(X) + Var(Y)$.

Preuve :

On a $Var(X + Y) = E((X + Y - E(X) - E(Y))^2) = E((X - E(X))^2 + 2(X - E(X))(Y - E(Y)) + (Y - E(Y))^2) = E((X - E(X))^2) + E((Y - E(Y))^2) + 2E(XY) - 4E(X)E(Y) + 2E(X)E(Y) = Var(X) + Var(Y)$

■

Théorème : Théorème de Huygens pour la variance

Soit X une variable aléatoire du second ordre et $a \in \mathbb{R}$. On a :

$$E((X - a)^2) = Var(X) + (E(X) - a)^2$$

Preuve :

Appliquer le théorème précédent sur $X - a = (X - E(X)) + (E(X) - a)$

■

Théorème : Inégalité de Bienaymé Tchebychev

Soit X une variable aléatoire du second ordre. On a :

$$\forall \varepsilon > 0, P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Lois de probabilité sur \mathbb{N} :

Définition :

Soit X une variable aléatoire à valeur entière. La série de terme général $P(X = n)z^n$ a un rayon de convergence au moins égal à 1. Sa somme notée $G_X = E(z^X)$ est une fonction de la variable complexe z appelée fonction génératrice de X .

Propriété :

La fonction génératrice possède les propriétés suivantes :

- Deux variables aléatoires entières possédant la même fonction génératrice ont même loi,
- Si X et Y sont des variables aléatoires entières indépendantes, alors $G_{X+Y}(z) = G_X(z) + G_Y(z)$,
- Si le rayon de convergence définissant G_X est supérieur à 1, X est de second ordre et on a $E(X) = G'_X(1)$ et $Var(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$

Preuve :

- L'identité des fonctions génératrice entraîne l'égalité des séries entières donc des coefficients de la série entière, donc de la loi de probabilité
- $G_{X+Y}(z) = E(z^{X+Y}) = E(z^X z^Y) = E(z^X)E(z^Y)$
- On a $G'_X(z) = \sum_{n \geq 1} (P(X=n)n)z^{n-1} \Rightarrow G'_X(1) = \sum_{n \geq 1} (P(X=n)n) = E(X)$,

$$G''_X(z) = \sum_{n \geq 2} (P(X=n)n(n-1))z^{n-2}$$

$$\Rightarrow G''_X(1) = \sum_{n \geq 2} (P(X=n)n(n-1)) = E(X(X-1))$$

Fonction génératrice :

Loi binomiale :

Définition :

Soit $n \geq 1$ et $p \in]0,1[$. La loi de probabilité binomiale $\mathfrak{B}(n,p)$ sur les entiers de 0 à n est définie par

$$P(\{k\}) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Remarques :

- La définition repose sur l'égalité $(p+1-p)^n = \sum_{k=0}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = 1$.
- La loi de probabilité $\mathfrak{B}(1,p)$ est fondamentale et porte le nom de la loi de Bernoulli de paramètre p .

Fonction génératrice, espérance et variance de la loi binomiale :

$$G(z) = \sum_{k=0}^n \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} z^k = (pz + 1 - p)^n$$

$$G'(z) = np(pz + 1 - p)^{n-1} \Rightarrow E(X) = G'(1) = np$$

$$G''(z) = n(n-1)p^2(pz + 1 - p)^{n-2} \Rightarrow Var(X) = G''(1) + G'(1) - G'(1)^2$$

$$= n(n-1)p^2 + np - (np)^2 = -np^2 + np = np(1-p)$$

Exercice : Optimisation de stock

On veut calculer le nombre optimal d'équipements à acquérir en fonction du service à fournir. Dans le cas courant où l'usage de ces équipements est aléatoire, l'application d'un calcul simple de probabilité est indispensable.

- Soit X le nombre aléatoire d'équipements nécessaires à un instant donné (On suppose que la loi de probabilité de X est stationnaire) et soit k le nombre d'équipement du stock de coût unitaire 1DT. Exprimer le cout aléatoire $C(X, k)$ associé à un stock donné sachant que le coût occasionné par la non satisfaction d'un service est N fois supérieur au coût d'un équipement. Calculer k pour minimiser le coût moyen $E(C(X, k))$,

2. Détermination de la loi des besoins. On fait l'hypothèse que l'ensemble des clients du stock est un ensemble E fixe et connu de taille n et que chaque client potentiel a effectivement besoin d'un équipement avec la probabilité p indépendamment des autres clients². Sous cette hypothèse, la probabilité qu'un sous-ensemble donné F de E de cardinal x soit l'ensemble des clients ayant besoin de l'équipement est $p^x(1-p)^{n-x}$. On note X la variable aléatoire représentant le nombre d'équipement nécessaire à un instant donné. Montrer que la loi de X est une loi binomiale de paramètre n et p .
3. Application numérique : On cherche à résoudre numériquement la question 1 pour déterminer la taille optimale de stock k^* pour $N = 10, n = 10, p = 0,3$
4. Vérifier les calculs précédents par une simulation numérique.

Loi multinomiale :

Définition :

Soient $(p_1, p_2, \dots, p_m) \in]0,1[^m$ tel que $\sum_{i=1}^m p_i = 1$ et soit $n \in \mathbb{N}$. La loi de probabilité P définie sur l'ensemble fini $\{(n_1, n_2, \dots, n_m) / \sum_{i=1}^m n_i = n\}$ par $P(n_1, n_2, \dots, n_m) = \frac{n!}{n_1! n_2! \dots n_m!} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}$ s'appelle loi multinomiale de paramètre $(n, p_1, p_2, \dots, p_m)$

Loi géométrique :

Définition :

Soit $p \in]0,1[$. On appelle loi de probabilité géométrique $\mathcal{G}(p)$ de paramètre p , la loi de probabilité sur les entiers strictement positifs par $P(\{n\}) = (1-p)^{n-1}p$

Fonction génératrice :

$$G(z) = \sum_{n=1}^{\infty} (1-p)^{n-1} p z^n = \frac{pz}{1 - (1-p)z}$$

De rayon de convergence strictement supérieur à 1.

On a :

$$E(X) = \frac{1}{p}$$

Et

$$Var(X) = \frac{1-p}{p^2}$$

Loi binomiale négative :

Définition :

Soit $p \in]0,1[$. On appelle loi binomiale négative $\mathcal{B}_n(n, p)$ de paramètre n, p la loi de probabilité :

$$P(X = k) = \frac{(k+n-1)!}{k! (n-1)!} (1-p)^k p^n$$

Fonction génératrice :

La fonction génératrice de la loi binomiale négative est

² Cette hypothèse est réaliste s'il s'agit d'un stock de pièces détachées pour une machine travaillant en continu. Elle n'est pas réaliste s'il s'agit d'une entreprise de travaux agricoles où l'aléa météorologique est le même pour tous les clients.

$$G(z) = \left(\frac{p}{1 - (1-p)z} \right)^n$$

$$E(X) = \frac{n(1-p)}{p}$$

$$Var(X) = \frac{n(1-p)}{p^2}$$

Loi de Poisson :

Définition :

On appelle loi de poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ la loi de probabilité sur \mathbb{N} définie par :

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Fonction génératrice :

$$G(z) = e^{\lambda(z-1)}$$

$$E(X) = Var(X) = \lambda$$

Loi de probabilité à densité sur \mathbb{R} :

Théorie de l'intégration

La notion d'intégrale est étroitement liée au calcul de l'air de la surface plane délimitée par les droites d'équations $x = a$ et $x = b$, l'axe des abscisses et la courbe représentative d'une fonction f positive et continue sur l'intervalle $I = [a, b] \subset \mathbb{R}$.

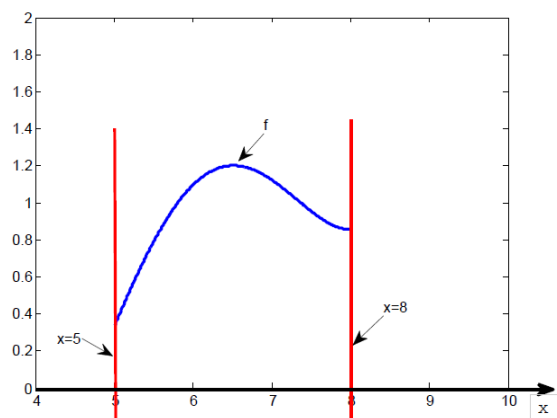


Figure 1 : Aire correspondant à l'intégrale d'une fonction f sur un intervalle I borné

Plusieurs théories ont été alors élaborées pour élargir le spectre des espaces fonctionnels sur lesquels on peut développer cette notion. Les plus connues sont celles de [Riemann](#)³ dont l'étude commence au cours de l'enseignement secondaire, et celle de [Lebesgue](#)⁴ qui fait l'essentiel de ce chapitre et qui nécessite des connaissances mathématiques plus compliquées.

1- Rappels sur l'intégrale de Riemann:

Riemann a développé sa théorie en se basant sur l'ensemble des [fonctions en escalier](#)⁵ définies sur un intervalle $[a, b]$ noté $Esc([a, b])$. Il a d'abord considéré les quantités :

$$s(f) = \sup_{\{u \in Esc[a, b]\}} \left\{ \int_a^b u(x) dx, u \leq f \right\}$$

et,

$$S(f) = \inf_{\{v \in Esc[a, b]\}} \left\{ \int_a^b v(x) dx, f \leq v \right\}$$

et définit l'intégrale comme suit :

Définition :

La fonction f est dite Riemann-intégrable sur $[a, b]$ si et seulement si $s(f) = S(f)$. Dans ce cas on note $\int_a^b f(x) dx = s(f) = S(f)$

Exercice :

³ http://fr.wikipedia.org/wiki/Bernhard_Riemann

⁴ http://fr.wikipedia.org/wiki/Henri-Léon_Lebesgue

⁵ http://fr.wikipedia.org/wiki/Fonction_étagée

Soit la fonction irrégulière de [Dirichlet](#)⁶ définie par :

$$D: [0,1] \rightarrow \mathbb{R} \\ x \mapsto \begin{cases} 1 & \text{si } x \in \mathbb{Q} \\ 0 & \text{sinon} \end{cases}$$

1- Calculer $s(D)$ puis $S(D)$.

2- Conclure.

3- Soit $n \in \mathbb{N}^*$. On considère la suite $(x_i)_{1 \leq i \leq n}$ de rationnels de l'intervalle $[0,1]$ distincts

$$f_n: [0,1] \rightarrow \mathbb{R} \\ \text{deux à deux et la fonction } x \mapsto \begin{cases} 1 & \text{si } x \in \{x_1, x_2, \dots, x_n\} \\ 0 & \text{sinon} \end{cases}$$

a- Montrer que la suite des fonctions $(f_n)_{n \in \mathbb{N}}$ est une [suites⁷ de Cauchy](#)⁸ et qu'elle converge vers la fonction D .

b- Dédurre que l'ensemble des fonctions intégrables au sens de Riemann, muni de la semi norme $\|f\|_1 = \int_a^b |f(x)| dx$ n'est pas un [espace complet](#)⁹.

L'intégrale de Lebesgue :

Introduction :

La technique de Lebesgue pour déterminer la valeur d'une intégrale est fondamentalement différente de celle de Riemann. Elle permet de donner des valeurs d'intégrales pour des fonctions irrégulières comme celle de Dirichlet ou sur des ensembles non classiques pour Riemann comme $\mathbb{R} \setminus \mathbb{Q}$ par exemple!!!

Dans cette introduction, et juste pour l'illustration de la différence entre les deux techniques, on présente sur un cas simple qui est celui de l'intégrale d'une fonction f continue sur un intervalle fermé borné $I = [a, b] \subset \mathbb{R}$.

Pratique numérique pour le calcul de l'intégrale selon Riemann :

Soit la famille de subdivisions de I , $(\Delta_n = \{a = x_0, x_1, \dots, x_n = b\})_{n \geq 1}$ avec $x_0 < x_1 < \dots < x_n$ et telle que $\lim_{n \rightarrow +\infty} (\|\Delta_n\| = \max_{0 \leq i \leq n-1} (x_{i+1} - x_i)) = 0$. Lorsque f est Riemann-intégrable, la valeur de l'intégrale est égale à :

$$s_R = \lim_{n \rightarrow +\infty} \left(\sum_{i=0}^{n-1} f(\tilde{x}_i) (x_{i+1} - x_i) \right)$$

Où $\tilde{x}_i \in [x_i, x_{i+1}]$.

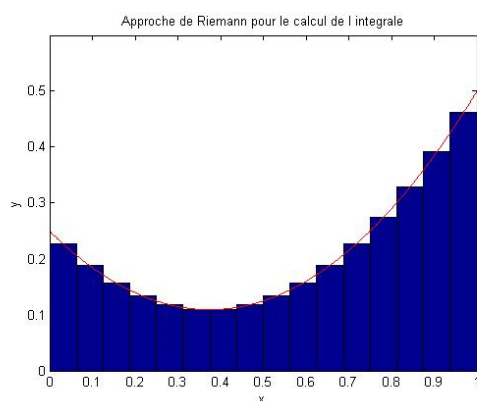
Les vidéos 1 et 2 illustrent l'évaluation numérique de l'intégrale par l'approche de Riemann pour $n = 16$ et $n = 32$.

⁶ http://fr.wikipedia.org/wiki/Johann_Peter_Gustav_Lejeune_Dirichlet

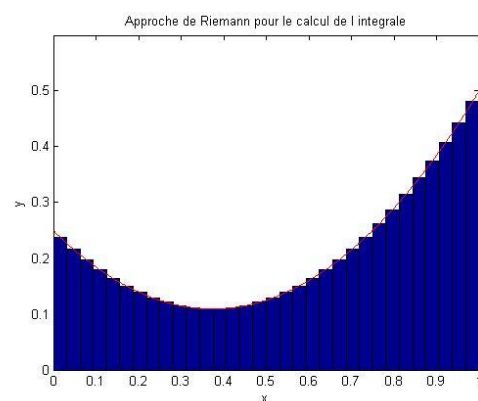
⁷ http://fr.wikipedia.org/wiki/Suite_de_Cauchy#D.C3.A9finition

⁸ http://fr.wikipedia.org/wiki/Augustin_Louis_Cauchy

⁹ http://fr.wikipedia.org/wiki/Espace_complet



Vidéo 1 : Evaluation numérique de l'intégrale par Riemann pour $n = 16$



Vidéo 2 : Evaluation numérique de l'intégrale par Riemann pour $n = 32$

Pratique numérique pour le calcul de l'intégrale selon Lebesgue :

Soient $y_0 = \min_{x \in [a,b]} f(x)$ et $y_n = \max_{x \in [a,b]} f(x)$ ¹⁰ et soit $(\Xi_n = \{y_0, y_1, \dots, y_n\})_{n \geq 1}$ une famille de subdivisions de l'intervalle $J = \left[\min_{x \in [a,b]} f(x), \max_{x \in [a,b]} f(x) \right]$ vérifiant $\lim_{n \rightarrow +\infty} (\|\Xi_n\| = \max_{0 \leq i \leq n-1} (y_{i+1} - y_i)) = 0$. Lorsque f est Lebesgue-intégrable, la valeur de son intégrale est égale à :

$$s_L = \lim_{n \rightarrow +\infty} \left(\sum_{i=0}^{n-1} y_i \mu(f^{-1}([y_i, y_{i+1}[)) + y_n \mu(f^{-1}(\{y_n\})) \right) \quad (*)$$

où pour $A \subset J$, $f^{-1}(A) \subset I$ est l'image réciproque de l'ensemble A par f et μ est la fonction qui mesure la « longueur » de $f^{-1}(A)$.

Remarque :

La cohérence du cours veut que s_L prenne l'expression donnée ci-dessus en (*). La formule ci-dessous est autant valable :

$$s_L = \lim_{n \rightarrow +\infty} \left(\sum_{i=0}^{n-1} \tilde{y}_i \mu(f^{-1}([y_i, y_{i+1}[)) + y_n \mu(f^{-1}(\{y_n\})) \right)$$

où $\tilde{y}_i \in (y_i, y_{i+1})$

¹⁰ http://fr.wikipedia.org/wiki/Théorème_des_bornes

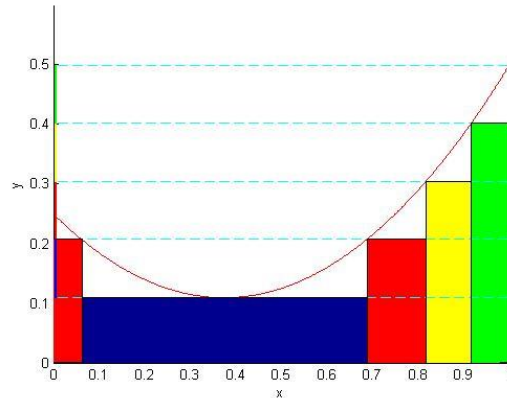
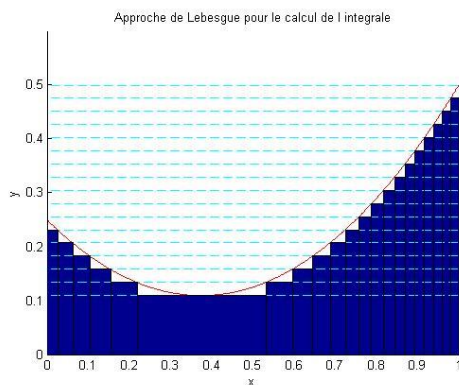
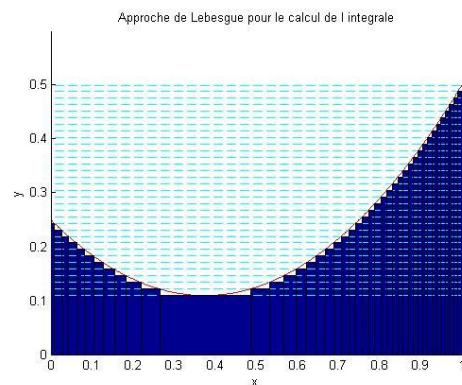


Figure 2 : Description de la méthode d'intégration de Lebesgue

Les vidéos 3 et 4 illustrent l'évaluation numérique de l'intégrale par l'approche de Lebesgue pour $n = 16$ et $n = 32$.



Vidéo 3 : Evaluation numérique de l'intégrale par Lebesgue pour $n = 16$



Vidéo 4 : Evaluation numérique de l'intégrale par Lebesgue pour $n = 32$

Commentaire :

[Les méthodes d'intégration numériques](http://fr.wikipedia.org/wiki/Calcul_numérique_d%27une_intégrale)¹¹ connues sont développées, pour des raisons évidentes de facilité de mise en œuvre, pour l'approche de Riemann. L'apport de Lebesgue se situe surtout dans la facilitation des calculs théoriques comme on va le voir ci-dessous.

L'Intégrale de Lebesgue :

Comme on vient de le voir par l'illustration faite à l'introduction ci-dessus, le calcul de l'intégrale selon Lebesgue passe par la prise en compte d'une famille de subdivisions de l'image de la fonction et par la mesure des ensembles images réciproques des éléments des différentes subdivisions. Ce calcul ne se restreint pas aux fonctions continues. Les outils développés par Lebesgue permettent même de calculer les intégrales d'une fonction irrégulière comme celle de Dirichlet. Ces outils font l'objet de ce qui suit. Par ailleurs, la série 1 permet d'approfondir des notions qui dépassent l'objet de ce cours.

¹¹ http://fr.wikipedia.org/wiki/Calcul_numérique_d%27une_intégrale

Tribus, Boréliens

Définition d'une tribu:

Soit E un ensemble quelconque et soit $\mathcal{P}(E)$ [l'ensemble des parties](#)¹² de E . On appelle Tribu sur E tout ensemble $\mathcal{T} \subset \mathcal{P}(E)$ de parties de E vérifiant les conditions:

- E et \emptyset appartiennent à \mathcal{T}
- Pour tout $\mathfrak{M} \in \mathcal{T}$ on a $E \setminus \mathfrak{M} \in \mathcal{T}$
- Si $(\mathfrak{M}_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{T} alors $\bigcup_{n \in \mathbb{N}} \mathfrak{M}_n \in \mathcal{T}$.

Exemples :

- $\{E, \emptyset\}$ est la plus petite tribu de \mathcal{T} ,
- $\mathcal{P}(E)$ est la plus grande tribu de \mathcal{T} ,
- Soit $a \in E$, alors $\{\emptyset, a, E \setminus \{a\}, E\}$, est la tribu dite engendrée par a ,
- Soit $n \in \mathbb{N}^*$. La tribu borélienne de \mathbb{R}^n définie par [E. Borel](#)¹³ est la tribu engendrée par tous les ouverts de \mathbb{R}^n . On la note $\mathfrak{B}(\mathbb{R}^n)$. Un ensemble borélien est un élément de $\mathfrak{B}(\mathbb{R}^n)$.

Espaces mesurables et espaces mesurés

Définition d'un espace mesurable:

On appelle espace mesurable tout couple (E, \mathcal{T}) où E est un ensemble quelconque et \mathcal{T} est une tribu sur E . Les éléments de \mathcal{T} sont appelés ensembles mesurables.

Exemples :

- Soit $a \in \mathbb{R}$, alors $\{a\} \in \mathcal{C}$. En effet, $\{a\} = \mathbb{R} \setminus (]-\infty, a[\cup]a, +\infty[)$,
- Toute union [dénombrable](#)¹⁴ de singletons de \mathbb{R} est un borélien.

Définition d'une mesure:

On appelle mesure toute fonction μ définie sur la tribu \mathcal{T} à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$ et telle que:

$$\mu(\emptyset) = 0$$

- $\mu(\bigcup_{n \in \mathbb{N}} \mathfrak{M}_n) = \sum_{n \in \mathbb{N}} \mu(\mathfrak{M}_n)$ pour toute famille dénombrable $(\mathfrak{M}_n)_{n \in \mathbb{N}}$ d'éléments deux à deux disjoints de \mathcal{T} .

Exemples :

- Soient E un ensemble et \mathcal{T} la plus petite tribu sur E . Alors la fonction μ définie
par
$$\begin{aligned} \mu : \mathcal{T} &\rightarrow \mathbb{R}_+ \cup \{+\infty\} \\ \mathfrak{M} &\mapsto \mu(\mathfrak{M}) = \begin{cases} 0 & \text{si } \mathfrak{M} = \emptyset \\ 1 & \text{si } \mathfrak{M} = E \end{cases} \end{aligned}$$
 est une mesure sur \mathcal{T} .

¹² http://fr.wikipedia.org/wiki/Ensemble_des_parties_d'un_ensemble

¹³ http://fr.wikipedia.org/wiki/%C3%89mile_Borel

¹⁴ http://fr.wikipedia.org/wiki/Ensemble_d%C3%A9nombrable

- Soit $n \in \mathbb{N}$. La mesure de Lebesgue sur \mathbb{R}^n est une fonction qui mesure le volume de toute partie de \mathbb{R}^n . A un intervalle $]a, b[\subset \mathbb{R}$ elle associe sa longueur $b - a$, à un rectangle $]a_1, b_1[\times]a_2, b_2[\subset \mathbb{R}^2$ elle lui associe son aire $(b_2 - a_2) \times (b_1 - a_1)$, et d'une manière générale à un pavé $\prod_{1 \leq i \leq n}]a_i, b_i[\subset \mathbb{R}^n$ elle lui associe son volume $\prod_{i=1}^n (b_i - a_i)$. Un ensemble Lebesgue-mesurable, ou mesurable tout simplement, est une partie de \mathbb{R}^n dont on peut calculer la mesure par l'approche de Lebesgue, c'est-à-dire dont on peut mesurer le volume. La tribu de Lebesgue \mathcal{L} est l'ensemble des parties Lebesgue-mesurables. Ainsi, tous les boréliens sont contenus dans la tribu de Lebesgue.

Remarque:

La mesure de tout ensemble mesurable peut être approchée soit par des ouverts, soit par des fermés, en considérant la notion de limite inf et sup et on a:

$$\forall \mathfrak{M} \in \mathfrak{L}, \mu(\mathfrak{M}) = \inf_{\{\mathcal{O} \in \text{ouverts de } \mathbb{R}\}} \{\mu(\mathcal{O}), \mathfrak{M} \subset \mathcal{O}\} = \sup_{\{F \in \text{compacts de } \mathbb{R}\}} \{\mu(F), F \subset \mathfrak{M}\}$$

Exemples :

- Pour $\mathfrak{M} = \{a\}$, on a :

$$\mu(\mathfrak{M}) = \inf \left\{ \mu \left(\left] a - \frac{1}{n}, a + \frac{1}{n} \right] \right), n \in \mathbb{N} \right\} = 0$$

- Pour $\mathfrak{M} = \{a_1, a_2, \dots, a_n\}$, on a $\mu(\mathfrak{M}) = \sum_{i=1}^n \mu(\{a_i\}) = 0$.

- La seconde propriété qui définit une mesure fait aussi que tout [ensemble dénombrable](#) est mesurable et est de mesure nulle. Ainsi $\mu(\mathbb{N}) = \mu(\mathbb{Z}) = \mu(\mathbb{Q}) = 0$.

Définition d'un ensemble négligeable:

Un ensemble E mesurable est dit négligeable, si et seulement si, sa mesure est nulle.

Définition d'une propriété vraie presque partout:

On dit qu'une propriété P sur un ensemble E est vraie presque partout (pp), si et seulement si P est vérifiée par tous les éléments $E \setminus M$ où M est un sous-ensemble négligeable de E .

Exemples :

- La fonction de Dirichlet D est nulle presque partout sur $[0,1]$.

- La fonction de Heaviside H , définie sur \mathbb{R} par :

$$H: \mathbb{R} \rightarrow \mathbb{R} \\ t \mapsto \begin{cases} 0 & \text{si } t \in]-\infty, 0[\\ 1 & \text{si } t \in [0, +\infty[\end{cases}$$

est dérivable pp .

Définition d'un espace mesuré:

On appelle espace mesuré la donnée d'une espace mesurable (E, \mathcal{T}) menu d'une mesure μ définie \mathcal{T} . On note (E, \mathcal{T}, μ) un espace mesuré.

Fonctions mesurables et intégrale de Lebesgue:

Définition d'une fonction mesurable:

Une fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ est dite mesurable, si pour tout ensemble mesurable $E \subset \mathbb{R}$, l'ensemble $f^{-1}(E)$ est mesurable. On note l'ensemble des fonctions mesurables définies sur un intervalle $I \subset \mathbb{R}$, $\mathcal{L}^0(I)$.

Exercice:

Montrer que toute fonction continue est mesurable.

Montrer que la relation \overline{pp} définie sur $\mathcal{L}^0(\mathbb{R})$ par $\forall (f, g) \in \mathcal{L}^0(\mathbb{R}) \times \mathcal{L}^0(\mathbb{R}), f \overline{pp} g \Leftrightarrow f - g = 0$ est une relation d'équivalence.

On note l'espace quotient¹⁵ $\mathcal{L}^0(\mathbb{R})/\overline{pp}$ par $L^0(\mathbb{R})$.

Définition d'une fonction étagée:

Une fonction $f: \mathbb{R} \rightarrow \mathbb{R}$ est dite étagée si et seulement s'il existe un nombre **fini** d'ensembles mesurables $(\mathfrak{M}_i)_{1 \leq i \leq n}$ et de réels $(\alpha_i)_{1 \leq i \leq n}$ tels que $f = \sum_{i=1}^n \alpha_i \chi_i$ où χ_i est la fonction caractéristique sur \mathfrak{M}_i .

Exercice:

Soit $f: \mathbb{R} \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ une fonction mesurable.

- 1- Montrer que si f est positive, alors il existe une suite de fonctions étagées positives et croissantes qui converge simplement vers f .
- 2- En déduire que si f est à valeurs quelconques alors est limite simple d'une suite de fonctions étagées.

Réponse :

- 1- Soit $(e_n)_{n \in \mathbb{N}}$ la suite de fonctions étagées définies par :

$$e_n(x) = \begin{cases} 2^{-n}k & \text{si } 2^{-n}k \leq f(x) < 2^{-n}(k+1) \text{ pour } k = 0, \dots, 2^{2n} - 1 \\ 2^n & \text{si } 2^n \leq f(x) \end{cases}$$

Si $f(x) < +\infty$ alors il existe $N \in \mathbb{N}^* / \forall n \geq N, 0 \leq f(x) - e_n(x) \leq \frac{1}{2^n}$.

Si $f(x) = +\infty$ alors $e_n(x) = 2^n \forall n \in \mathbb{N}$, et $\lim_{n \rightarrow +\infty} e_n(x) = f(x)$

- 2- Soient f^+ et f^- telles que :

$$f^+(x) = \begin{cases} f(x) & \text{si } 0 \leq f(x) \\ 0 & \text{si } f(x) < 0 \end{cases}$$

et

$$f^-(x) = \begin{cases} -f(x) & \text{si } f(x) \leq 0 \\ 0 & \text{si } 0 < f(x) \end{cases}$$

$f^+(x)$ et $f^-(x)$ sont positives et $\forall x \in \mathbb{R}, f(x) = f^+(x) - f^-(x)$

La déduction de la première question est alors évidente

Exercice:

¹⁵ http://fr.wikipedia.org/wiki/Relation_d'équivalence#Ensemble_quotient

Montrer que la fonction de Dirichlet est étagée.

Exercice:

Toute fonction étagée est mesurable et il en est de même pour la somme, le produit et le module de deux fonctions mesurables.

Définition de l'intégrale d'une fonction étagée:

Soit f une fonction étagée à valeurs positives. L'intégrale de f est définie comme le nombre positif, éventuellement infini $\int_{\mathbb{R}} f(x) d\mu(x) = \sum_{i=1}^n \alpha_i \mu(\mathfrak{M}_i)$, avec la convention que $0 \times \infty = 0$. Si ce nombre est fini, la fonction f est dite intégrable au sens de Lebesgue ou encore Lebesgue-intégrable ou intégrable en absence d'ambiguïté.

Si E est un sous ensemble mesurable de \mathbb{R} , alors l'intégrale de f sur E est définie par $\int_E f(x) d\mu(x) = \sum_{i=1}^n \alpha_i \mu(\mathfrak{M}_i \cap E)$.

Application:

Montrer que La fonction de Dirichlet est Lebesgue-intégrable et son intégrale au sens de Lebesgue vaut zéro.

Définition de l'intégrale de Lebesgue:

Soit $f: \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable à valeurs positives. Alors on définit l'intégrale de f au sens de Lebesgue sur $E \subset \mathbb{R}$ par :

$$\int_E f(x) d\mu(x) = \sup_{\{g/g \text{ fonction étagée}\}} \left\{ \int_E g(x) d\mu(x) ; 0 \leq g \leq f \right\}$$

Remarques :

- Pour $f: \mathbb{R} \rightarrow \mathbb{R}$ est à valeurs réelles quelconques. Soit la décomposition $f = f^+ - f^-$ avec $f^+(x) = \begin{cases} f(x) & \text{si } f(x) \geq 0 \\ 0 & \text{si } f(x) < 0 \end{cases}$ et $f^-(x) = \begin{cases} -f(x) & \text{si } f(x) \leq 0 \\ 0 & \text{si } f(x) > 0 \end{cases}$. On définit alors l'intégrale de f au sens de Lebesgue sur $E \subset \mathbb{R}$ par $\int_E f(x) d\mu(x) = \int_E f^+(x) d\mu(x) - \int_E f^-(x) d\mu(x)$.
- Pour $f: \mathbb{R} \rightarrow \mathbb{C}$ on a $f = \text{Re}(f) + i \text{Im}(f)$ et on définit l'intégrale de f au sens de Lebesgue sur $E \subset \mathbb{R}$ par $\int_E f(x) d\mu(x) = \int_E \text{Re}(f)(x) d\mu(x) + i \int_E \text{Im}(f)(x) d\mu(x)$

Propriétés du calcul intégral.

Soient f et g deux fonctions mesurables définies sur \mathbb{R} à valeurs réelles ou complexes.

Linéarité:

Soient f et g deux fonctions intégrables sur \mathbb{R} au sens de Lebesgue, alors $\forall (\alpha, \beta) \in \mathbb{C}^2$ on a :

$$\int_E (\alpha f(x) + \beta g(x)) d\mu(x) = \alpha \int_E f(x) d\mu(x) + \beta \int_E g(x) d\mu(x)$$

Comparaison:

- Soient f et g deux fonctions intégrables sur \mathbb{R} au sens de Lebesgue avec $f \leq g$ alors $\int_E f(x) d\mu(x) \leq \int_E g(x) d\mu(x)$
- Soit f intégrable sur \mathbb{R} , alors $\left| \int_E f(x) d\mu(x) \right| \leq \int_E |f(x)| d\mu(x)$
- Soit g intégrable sur \mathbb{R} , et f telle que $|f| \leq g$, alors f est intégrable et $\int_E |f(x)| d\mu(x) \leq \int_E g(x) d\mu(x)$

Intégrales nulles:

Soit $f: \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction mesurable alors $\int_E f(x) d\mu(x) = 0 \Leftrightarrow f = 0$ pp.

Comparaison entre l'intégrale de Riemann et l'intégrale de Lebesgue :

- Toute fonction Riemann-intégrable est Lebesgue-intégrable et les deux intégrales coïncident. La réciproque est fausse. En effet, la fonction $f = \chi_{\mathbb{Q} \cap [0,1]} - \chi_{(\mathbb{R} \setminus \mathbb{Q}) \cap [0,1]}$ définie sur $[0,1]$ qui est Lebesgue intégrable n'est pas Riemann intégrable.
- Pour Lebesgue, une fonction f mesurable est intégrable si et seulement si $|f|$ est intégrable. Ceci n'est pas vrai pour Riemann. En effet, la fonction $f = \chi_{\mathbb{Q} \cap [0,1]} - \chi_{(\mathbb{R} \setminus \mathbb{Q}) \cap [0,1]}$ définie sur $[0,1]$ n'est pas Riemann intégrable mais $|f|$ est Riemann intégrable, par ailleurs, la fonction $f(x) = \frac{\sin(x)}{x}$ est Riemann-intégrable mais $|f|$ ne l'est pas.

Les espaces $L^p(\mathbb{R})$, $p \geq 1$:

Les espace $L^p(\mathbb{R})$, $1 \leq p < +\infty$:

Soit l'entier naturel $1 \leq p < +\infty$. On considère $\mathcal{L}^p(\mathbb{R}) = \left\{ f \in \mathcal{L}^0(\mathbb{R}) / \left(\int_{\mathbb{R}} |f(x)|^p d\mu(x) \right)^{\frac{1}{p}} < +\infty \right\}$ et l'espace quotient $L^p(\mathbb{R}) = \mathcal{L}^p(\mathbb{R}) / \overline{pp}$. L'application $\| \cdot \|_p : f \rightarrow$

$\left(\int_{\mathbb{R}} |f(x)|^p d\mu(x) \right)^{\frac{1}{p}}$ est une semi norme sur $\mathcal{L}^p(\mathbb{R})$ et est une norme sur $L^p(\mathbb{R})$. $(L^p(\mathbb{R}), \| \cdot \|_p)$ est un espace vectoriel complet. En particulier, $L^2(\mathbb{R})$ est un espace de Hilbert¹⁶ avec $\int_{\mathbb{R}} f(x) \overline{g(x)} d\mu(x)$ comme produit scalaire associé à $\| \cdot \|_2$.

3-2 L'espace $L^\infty(\mathbb{R})$:

Définition :

Soit $f \in \mathcal{L}^0(\mathbb{R})$. f est dite essentiellement bornée s'il existe une constante réelle $c \geq 0$ telle que l'ensemble $\{x / f(x) > c\}$ soit négligeable.

¹⁶ http://fr.wikipedia.org/wiki/Espace_de_Hilbert

¹⁷ http://fr.wikipedia.org/wiki/David_Hilbert

Définition :

Soit $L^\infty(\mathbb{R})$ l'espace vectoriel des fonctions mesurables essentiellement bornées

et $L^\infty(\mathbb{R}) = \mathcal{L}^\infty(\mathbb{R}) / \overline{pp}$. L'application $\| \cdot \|_\infty : f \rightarrow \inf_{\mathbb{R}} \{c \geq$

$0 / \{x / |f(x)| > c\} \text{ soit négligeable} \}$ est une norme sur $L^\infty(\mathbb{R})$ et $(L^\infty(\mathbb{R}), \| \cdot \|_\infty)$ est un espace complet.

Quelques théorèmes pratiques du calcul intégral.

Théorème de convergence dominée de Lebesgue

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables positives sur \mathbb{R} convergeant presque partout vers f . S'il existe une fonction g mesurable, intégrable et telle que $\forall n \in \mathbb{N}, |f_n(x)| \leq g(x)$ pp alors :

- f est intégrable, et

- Pour tout ensemble mesurable $E \subset \mathbb{R}$, on a $\lim_{n \rightarrow +\infty} \int_E f_n(x) dx = \int_E \lim_{n \rightarrow +\infty} f_n(x) dx = \int_E f(x) dx$

Exercice :

Calculer $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} \frac{ne^{-x^2} \cos(x)}{1+x^2 x^2} dx$

Intégrales dépendantes d'un paramètre :

Cadre de travail :

Soit $[a, b] \subset \mathbb{R}$ un intervalle borné, et $f: [a, b] \times \mathbb{R} \rightarrow \mathbb{R}$ mesurable. On pose pour $t \in [a, b]$, $I(t) = \int_{\mathbb{R}} f(t, x) dx < +\infty$. Soient $t_0 \in [a, b]$ et $V(t_0)$ un voisinage de t_0 .

Théorème : Continuité d'une intégrale à paramètre

Si $f_x: [a, b] \rightarrow \mathbb{R}$ est continue en t_0 pour presque tout $x \in \mathbb{R}$ et qu'il existe une fonction $g: \mathbb{R} \rightarrow \mathbb{R}$ intégrable et telle que pour tout $t \in V(t_0)$, $f(t, x) \leq g(x)$ pp, alors I est continue en t_0 .

Preuve :

Soit $(t_n)_{n \in \mathbb{N}}$ une suite dans $V(t_0)$ qui converge vers t_0 . Posons $f_n: \mathbb{R} \rightarrow \mathbb{R}$ par $f_n(x) = f(t_n, x)$. On a $\lim_{n \rightarrow +\infty} f_n(x) = f(t_0, x)$ pour presque tout $x \in \mathbb{R}$. De plus, $|f_n(x)| \leq g(x)$ pp \Rightarrow $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}} f_n(x) dx = \int_{\mathbb{R}} \lim_{n \rightarrow +\infty} f_n(x) dx \Rightarrow \lim_{n \rightarrow +\infty} I(t_n) = \int_{\mathbb{R}} f(t_0, x) dx = I(t_0)$.

Théorème : Dérivabilité d'une intégrale à paramètre

Si $f_x: [a, b] \rightarrow \mathbb{R}$ est dérivable dans un voisinage $V(t_0)$ de t_0 pour presque tout $x \in \mathbb{R}$ et qu'il existe une fonction $g: \mathbb{R} \rightarrow \mathbb{R}$ intégrable et telle que pour tout $t \in V(t_0)$, $\left| \frac{\partial f(t, x)}{\partial t} \right| \leq g(x)$ pp, alors I est dérivable en t_0 et $\frac{dI}{dt}(t_0) = \int_{\mathbb{R}} \frac{\partial f(t, x)}{\partial t} \Big|_{t_0} dx$

Théorème de [Fubini](#)¹⁸:

Soit $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable.

1- Supposons que f est intégrable sur \mathbb{R}^2 , alors :

- la fonction $f_x: y \rightarrow f_x(y) = f(x, y) \in L^1(\mathbb{R})$ pour presque tout $x \in \mathbb{R}$ et $k(x) = \int_{\mathbb{R}} f(x, y) dy \in L^1(\mathbb{R})$, et
- $\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx.$

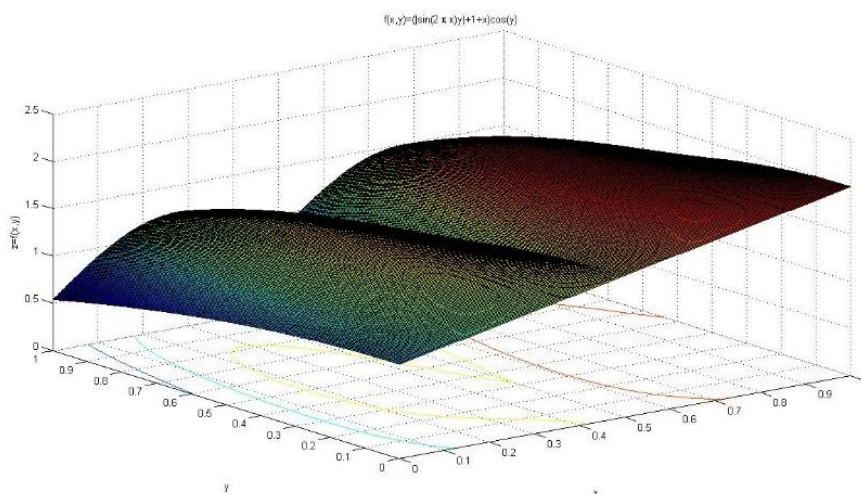
2- Réciproquement, si la fonction $f_x: y \rightarrow f_x(y) = f(x, y) \in L^1(\mathbb{R})$ pour presque tout $x \in \mathbb{R}$ et s'il existe $k^* \in L^1(\mathbb{R})$ telle que $k(x) = \int_{\mathbb{R}} |f(x, y)| dy \leq k^*(x)$ alors

- $f \in L^1(\mathbb{R}^2)$ et
- $\int_{\mathbb{R}^2} f(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx.$

Remarques :

- La preuve de ce théorème important fait l'objet de la série théorème de [Fubini](#).
- Le théorème de [Fubini](#) stipule que si $f \in L^1(\mathbb{R}^2)$ alors la valeur de l'intégrale en considérant la topologie de \mathbb{R}^2 est la même que sa valeur après deux intégrations successives en considérant la topologie de \mathbb{R} . Les figures et les vidéos ci-dessous montrent la différence entre les différents calculs pour l'exemple :

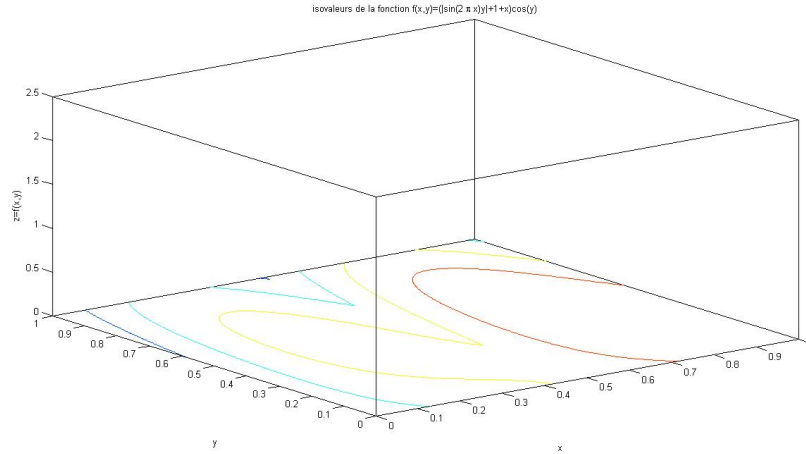
$$\begin{array}{ccc} f: [0,1] \times [0,1] & \rightarrow & \mathbb{R} \\ (x, y) & \mapsto & (|\sin(2\pi x)y| + 1 + x)\cos(y) \end{array}$$



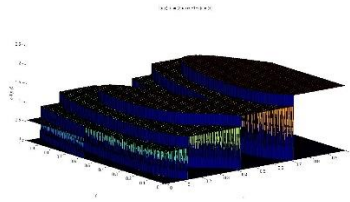
Les iso valeurs de la fonction pour $z_i = \min_{[0,1] \times [0,1]} f(x, y) + ih$, $i = 1, 2, 3, 4$ et 5 avec

$h = \frac{\left(\max_{[0,1] \times [0,1]} f(x, y) - \min_{[0,1] \times [0,1]} f(x, y) \right)}{5}$ sont présentées sur la figure ci-dessous :

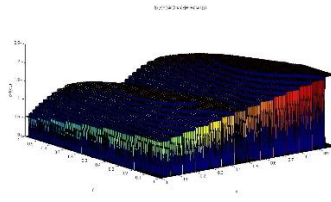
¹⁸ http://fr.wikipedia.org/wiki/Guido_Fubini



Et l'approximation de la valeur de l'intégrale pour deux différentes subdivisions est illustrée dans les vidéos 5 et 6 suivantes :



Vidéo 5 : Approximation de l'intégrale
 $\int_{[0,1] \times [0,1]} (|\sin(2\pi x)y| + 1 + x) \cos(y) d\mu(x, y)$
 en subdivisant uniformément
 $\left[\min_{[0,1] \times [0,1]} f(x, y), \max_{[0,1] \times [0,1]} f(x, y) \right]$ en cinq intervalles



Vidéo 6 : Approximation de l'intégrale
 $\int_{[0,1] \times [0,1]} (|\sin(2\pi x)y| + 1 + x) \cos(y) d\mu(x, y)$
 en subdivisant uniformément
 $\left[\min_{[0,1] \times [0,1]} f(x, y), \max_{[0,1] \times [0,1]} f(x, y) \right]$ en vingt intervalles

Par ailleurs, les fonctions :

$$k_x: [0,1] \rightarrow \mathbb{R}$$

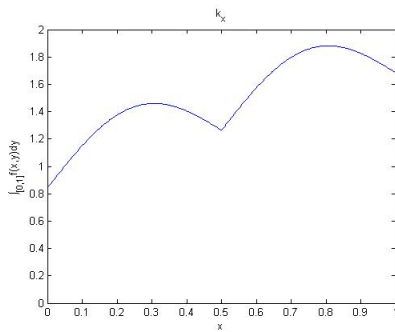
$$x \mapsto \int_{[0,1]} (|\sin(2\pi x)y| + 1 + x) \cos(y) d\mu(y)$$

et

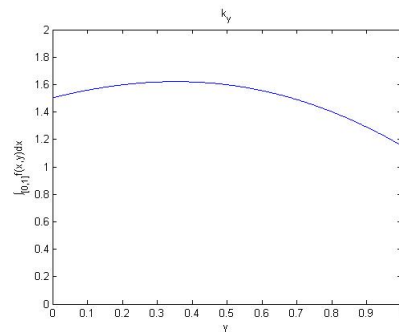
$$k_y: [0,1] \rightarrow \mathbb{R}$$

$$y \mapsto \int_{[0,1]} (|\sin(2\pi x)y| + 1 + x) \cos(y) d\mu(x)$$

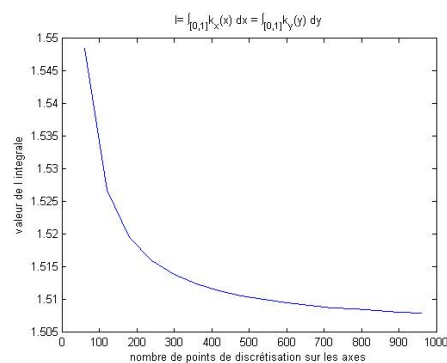
représentées dans les figures ci-dessous ont des valeurs d'intégrales identiques $I = \int_{[0,1]} k_x(x) d\mu(x) = \int_{[0,1]} k_y(y) d\mu(y)$, et ce quel que soit la discrétisation des intervalles $[0,1]$ suivant l'axe des abscisses et des ordonnées.



Courbe de k_x



Courbe de k_y



Valeur de $I = \int_{[0,1]} k_x(x) d\mu(x) = \int_{[0,1]} k_y(y) d\mu(y)$
en fonction de la discrétisation des axes des abscisses et des ordonnées



Henri Lebesgue

Beauvais, 28 juin 1875 à – Paris, 26 juillet 1941

Loi de probabilité à densité sur \mathbb{R} :

Loi de probabilité uniforme :

Définition :

Soit $a < b$, on appelle loi de probabilité uniforme sur $[a, b]$ et on note $\mathcal{U}_{[a,b]}$ la probabilité sur $[a, b]$ vérifiant $\forall \alpha, \beta \in [a, b] / a \leq \alpha \leq \beta \leq b, P([\alpha, \beta]) = \frac{\beta - \alpha}{b - a}$.

Définition : Espérance de $f(X)$ si $X \sim \mathcal{U}_{[a,b]}$

Soit f une fonction intégrable sur $[a, b]$ et X une var (variable aléatoire réelle) de loi uniforme sur $[a, b]$. On définit l'espérance de la variable aléatoire $f(X)$ par $E(f(X)) = \frac{1}{b-a} \int_a^b f(x) dx$

Théorème :

Soit X une variable aléatoire de loi uniforme sur $[a, b]$. L'application

$$\begin{aligned} L^1([a, b]) &\rightarrow \mathbb{R} \\ f &\mapsto E(f(X)) = \frac{1}{b-a} \int_a^b f(x) dx \end{aligned}$$

Est une forme linéaire positive sur $L^1([a, b])$. On a $|E(f(X))| \leq E(|f(X)|) \leq (b-a) \sup_{[a,b]} |f(X)|$.

Définition : Variance et écart type de $f(X)$ si $X \sim \mathcal{U}_{[a,b]}$

Si $f \in L^2([a, b])$, on dit que $Y = f(X)$ est une variable aléatoire d'ordre 2. On définit la variance et l'écart type de la variable par $Var(Y) = E(Y - E(Y))^2 = E(Y^2) - E(Y)^2$ et $\sigma_Y = \sqrt{Var(Y)}$

Théorème :

Si $X \sim \mathcal{U}_{[a,b]}$, alors $Var(X) = \frac{(b-a)^2}{12}$ et $\sigma_X = \frac{b-a}{2\sqrt{3}}$

Exercice :

Calcul et simulation de l'espérance et de la variance de $\cos(X)$ ou $X \sim \mathcal{U}_{[0,\pi]}$.

Caractérisation d'une ddp sur \mathbb{R} :

Définition : Densité de probabilité sur \mathbb{R}

Une fonction h définie sur \mathbb{R} positive, intégrable et d'intégrale 1 est appelée densité de probabilité (ddp). Si h est une ddp, il existe une unique loi de probabilité P sur $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ vérifiant $\forall a < b, P([a, b]) = \int_a^b h(x) dx$. Cette loi est dite de densité h .

Définition : Fonction de répartition d'une loi de probabilité sur \mathbb{R}

Soit $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$ un espace de probabilité sur \mathbb{R} . La fonction de répartition de P (cdf : cumulative distribution function) est la fonction croissante de \mathbb{R} dans $[0,1]$ définie par $F(x) = P([-\infty, x])$.

Théorème :

Soit F la fonction de répartition d'une loi de probabilité de densité h . Si h est continue en x , alors F est dérivable en x et de dérivée $h(x)$.

Exemple : Loi de probabilité uniforme

Soit $P = \mathcal{U}_{[a,\pi b]}$. Sa fonction de répartition est : $F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$

Corollaire : Théorème de la fonction de répartition

Si la ddp h a pour support l'intervalle $I = [a, b]$, respectivement $[a, +\infty[$, $]-\infty, b]$, $]-\infty, +\infty[$, alors sa fonction de répartition F est une bijection monotone croissante de l'intérieur I sur $]0,1[$ de dérivée h et si U est une var de densité uniforme sur $[0,1]$, $X = F^{-1}(U)$ est une var de densité h .

Preuve :

La fonction de répartition de la loi X est $F_X(x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$

■

Ce théorème est important pour simuler une loi de probabilité dont l'inverse de la fonction de répartition est connu et programmé à partir d'un simulateur de la loi uniforme.

Théorème : Espérance d'une var dont la loi de probabilité admet une densité

Soit X une var de ddp h et f une fonction de la variable réelle telle que $fh \in L^1(\mathbb{R})$, alors $E(f(x)) = \int f(x)h(x)dx$

Preuve :

On a vu que si F est la fonction de répartition de X et G la fonction inverse, alors $U = F(X)$ a une loi de probabilité uniforme sur $[0,1]$. Ceci implique que

$$E(f(X)) = E(f(G(U))) = \int_0^1 f(G(U))du = \int f(x)h(x)dx$$

Exercice : Calcul de la fonction de répartition et simulation de la ddp triangle

Soit $h(x) = x\chi_{[0,1]}(x) + (2-x)\chi_{[1,2]}(x)$

1. Montrer que h est une ddp
2. Soit X une var de ddp h . Calculer sa fonction de répartition F_X
3. Calculer F_X^{-1} pour programmer un simulateur de la loi de ddp h
4. Comparer l'histogramme d'une simulation (nbin=10) avec l'histogramme d'un échantillon simulé de taille 1000
5. Calculer l'espérance et la variance de X et comparer avec les résultats empiriques issus de la simulation précédente

Définition : Fonction caractéristique

Soit X une var de loi de probabilité P_X . La fonction caractéristique de X ou de sa loi de probabilité P_X est notée φ_X et est égale à $\varphi_X(t) = E(e^{itX})$

Théorème :

Soit X une var du second ordre et φ_X sa fonction caractéristique. On a : $\varphi_X(0) = 1$, $\varphi_X'(0) = iE(X)$, $\varphi_X''(0) = -E(X^2)$, $Var(X) = -\varphi_X''(0) + \varphi_X'(0)^2$

Exercice :

Calculer la fonction caractéristique de la loi uniforme sur $[0,1]$ et retrouver son espérance et sa variance.

Exemple de la densité gaussienne $\mathcal{N}(m, \sigma)$

La loi de densité gaussienne, ou loi normale, est la loi la plus importante en probabilité. Elle gouverne de nombreux phénomènes en raison de propriétés remarquables. Par exemple, la plupart des représentations du bruit en électronique et en théorie du signal utilisent des lois gaussiennes. D'autre part, la famille des lois gaussiennes est stable par les transformations simples mises en œuvre par les systèmes de communications (transformation linéaire).

Définition :

Soit $(m, \sigma) \in \mathbb{R} \times \mathbb{R}^+$. La fonction $x \rightarrow \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$ est une ddp dont la loi de probabilité est la loi gaussienne $\mathcal{N}(m, \sigma)$, dite encore, loi normale.

Théorème :

Soit $X \sim \mathcal{N}(m, \sigma)$. On a :

- $a \neq 0, Y = aX + b \Rightarrow Y \sim \mathcal{N}(am + b, |a|\sigma)$. En particulier, $Y = \frac{X-m}{\sigma} \Rightarrow Y \sim \mathcal{N}(0,1)$,
- $E(X) = m$ et $Var(X) = \sigma^2$
- $\varphi_X(t) = e^{itm - \frac{\sigma^2 t^2}{2}}$

Exemple de la densité Gamma $\Gamma(k, \theta)$

Le but de cette fonction est de représenter par une intégrale des fonctions à croissance très rapide associées aux produits du type $(x+1) \dots (x+n)$

Définition :

On appelle fonction Γ la fonction définie pour $k > 0$, $\Gamma(k) = \int_0^{+\infty} t^{k-1} e^{-t} dt$

Définition : Loi Gamma $\Gamma(k, \theta)$

Soit $(k, \theta) \in \mathbb{R}^{+*} \times \mathbb{R}^{+*}$. La fonction $x \rightarrow \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$ est une ddp de loi de probabilité la loi $\Gamma(k, \theta)$. Le paramètre θ est appelé paramètre d'échelle et le paramètre k de forme. La loi $\Gamma(1, \theta)$ est aussi appelée loi exponentielle de paramètre $\frac{1}{\theta}$

Théorème :

Soit $X \sim \Gamma(k, \theta)$. On a :

- $\varphi_X(t) = \frac{1}{1 - \theta i t k}$,
- $E(X) = k\theta$,
- $Var(X) = k\theta^2$,
- $\forall a > 0, aX \sim \Gamma(k, a\theta)$

Espaces de probabilité continus construits sur \mathbb{R}

Soit $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ la droite réelle munie de la tribu borélienne et P une probabilité sur \mathbb{R} de ddp h , c'est-à-dire définie par $P(B) = \int_B h(x) dx$, alors le triplet $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$ est appelé espace de probabilité continu sur \mathbb{R} .

Théorème :

Soit X une var de loi de probabilité P de densité h , et soit $Y = f(X) \in L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$. On a $E(Y) = \int f(x) h(x) dx$

Théorème : Caractérisation d'une ddp par la méthode de la fonction test

Soit X une var et h une ddp. Si pour toute fonction continue à support borné ϕ , on a $E(\phi(X)) = \int \phi(x) h(x) dx$, alors h est la ddp de X

Théorème :

Soit $X \in L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$ une variable aléatoire. On a :

- a. Si $E(|X|) < +\infty$ et $X \geq 0$, alors $E(X) \geq 0$,
- b. Si X et Y sont dans $L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$, et si $X \leq Y$ alors $E(X) \leq E(Y)$
- c. On a en particulier $|E(X)| \leq E(|X|)$

Théorème :

Soient X et Y deux var. Si elles sont dans $L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$, alors $\forall (a, b) \in \mathbb{R}^2$ $aX + bY \in L^1(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P)$ et $E(aX + bY) = aE(X) + bE(Y)$

Théorème : Inégalité de Markov

Soit X une variable aléatoire telle que $E(|X|) < +\infty$, alors $\forall \varepsilon > 0, P(X \geq \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$