

Introduction aux probabilités

Outils-Math

Bastien DOIGNIES

LIRIS

2023

Table of Contents

- 1 Probabilité : définition
- 2 Probabilités conditionnelles
- 3 Variable aléatoire et lois de probabilités
- 4 Espérance, Variance et Covariance
- 5 Théorèmes et propriétés diverse(s).
- 6 Probabilités multivariées

Probabilité : définition

Les probabilités sont la théorie mathématique qui étudie le caractère probable d'un évènement. On considère ainsi une expérience "aléatoire", c'est à dire une expérience dont on ne peut donner avec certitude le résultat (un lancé de dés, un tirage de boules, ...).

Les probabilités sont la théorie mathématique qui étudie le caractère probable d'un évènement. On considère ainsi une expérience "aléatoire", c'est à dire une expérience dont on ne peut donner avec certitude le résultat (un lancé de dés, un tirage de boules, ...).

Le résultat de l'expérience prend ainsi plusieurs valeurs. On note souvent Ω l'univers des possibles. Les parties de Ω sont appelées les évènements. L'idée sous-jacente est, pour une même expérience, de pouvoir assigner une probabilité à toutes les combinaisons de résultats possibles.

Les probabilités sont la théorie mathématique qui étudie le caractère probable d'un évènement. On considère ainsi une expérience "aléatoire", c'est à dire une expérience dont on ne peut donner avec certitude le résultat (un lancé de dés, un tirage de boules, ...).

Le résultat de l'expérience prend ainsi plusieurs valeurs. On note souvent Ω l'univers des possibles. Les parties de Ω sont appelées les évènements. L'idée sous-jacente est, pour une même expérience, de pouvoir assigner une probabilité à toutes les combinaisons de résultats possibles.

Exemple : je lance un dé tétraédrique. On a $\Omega = \{1, 2, 3, 4\}$ et les évènements sont : $\{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}, \{1, 2, 3\}, \{1, 3, 4\}, \{1, 2, 4\}, \{2, 3, 4\}, \Omega\}$

Probabilité

On appelle probabilité, ou loi de probabilité, toute application de $\mathcal{P}(\Omega)$ dans $[0, 1]$ vérifiant :

- $\mathbb{P}(\Omega) = 1$
- Si $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Probabilité

On appelle probabilité, ou loi de probabilité, toute application de $\mathcal{P}(\Omega)$ dans $[0, 1]$ vérifiant :

- $\mathbb{P}(\Omega) = 1$
- Si $A \cap B = \emptyset$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

La définition encapsule l'idée que :

- Le fait d'observer une valeur est toujours certain (de probabilité 1).
- Si deux évènements sont incompatibles (ils ne peuvent pas se produire en même temps), alors la probabilité d'observer l'un ou l'autre, est la somme des probabilités.

Exemple de probabilité pour le dé tétraédrique :

- $\mathbb{P}(\{1\}) = \frac{1}{4}$
- $\mathbb{P}(\{2\}) = \frac{1}{2}$
- $\mathbb{P}(\{3\}) = \frac{1}{8}$
- $\mathbb{P}(\{4\}) = \frac{1}{8}$

Et, par exemple :

- $\mathbb{P}(\{1, 2\}) = \mathbb{P}(\{1\} \cup \{2\}) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

Exemple de probabilité pour le dé tétraédrique :

- $\mathbb{P}(\{1\}) = \frac{1}{4}$
- $\mathbb{P}(\{2\}) = \frac{1}{2}$
- $\mathbb{P}(\{3\}) = \frac{1}{8}$
- $\mathbb{P}(\{4\}) = \frac{1}{8}$

Et, par exemple :

- $\mathbb{P}(\{1, 2\}) = \mathbb{P}(\{1\} \cup \{2\}) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$

De manière générale :

- $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup \bar{A}) = \mathbb{P}(A) + \mathbb{P}(\bar{A}) \implies \mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Probabilités conditionnelles

Probabilité conditionnelle d'évènements

On appelle probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A|B)$ la quantité :

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

Probabilité conditionnelle d'évènements

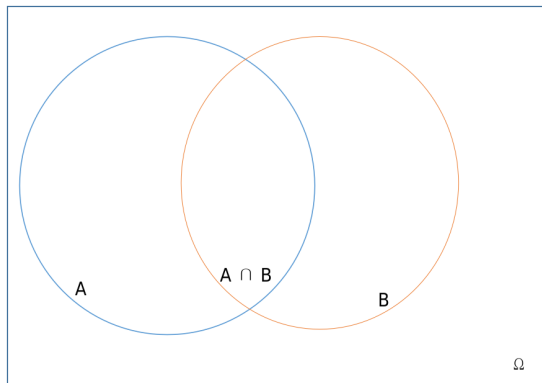
On appelle probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A|B)$ la quantité :

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

Probabilité conditionnelle d'évènements

On appelle probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A|B)$ la quantité :

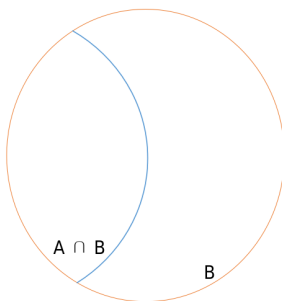
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(B) \mathbb{P}(A|B) = \mathbb{P}(A) \mathbb{P}(B|A)$$



Probabilité conditionnelle d'évènements

On appelle probabilité conditionnelle de A sachant B, notée $\mathbb{P}(A|B)$ la quantité :

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



Indépendance

On dit que deux évènements sont indépendants, si :

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

Ainsi, deux évènements sont indépendants si connaître l'un, n'influence pas la probabilité d'observer l'autre.

Indépendance

On dit que deux évènements sont indépendants, si :

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

Ainsi, deux évènements sont indépendants si connaître l'un, n'influence pas la probabilité d'observer l'autre.

Indépendance de plusieurs évènements

On dit que des évènements $A_1, \dots, A_N$ sont indépendants si, pour tout $n \leq N$, et pour tout entiers distincts $i_1, \dots, i_n \leq N$, on a :

$$\mathbb{P}(\cap_{i=1}^n A_i) = \prod_{i=1}^n \mathbb{P}(A_i)$$

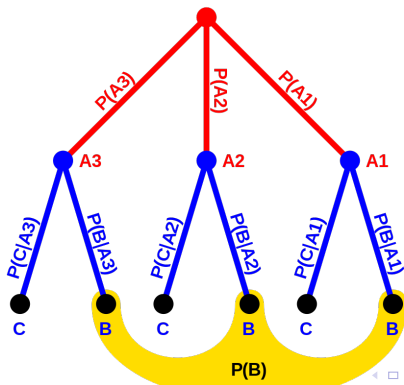
ATTENTION, IL FAUT VÉRIFIER TOUTES LES COMBINAISONS

Probabilités totales

Si $(A_i)_i$ est une partition d'un événement A , alors :

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

Par exemple : $\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|\bar{A})\mathbb{P}(\bar{A})$



Formules de Bayes

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

De manière équivalente :

Formules de Bayes

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

Daniel Kahneman et son prix Nobel (1/3)

Daniel Kahneman est un chercheur en psychologie qui gagne en 2002 le prix Nobel d'économie pour avoir "intégrer des principes psychologique dans les sciences économique". Ses travaux sont liés à la formule de Bayes. L'histoire est racontée par Grant Sanderson (3B1B) dans sa vidéo sur la formule de Bayes.

Daniel Kahneman et son prix Nobel (1/3)

Daniel Kahneman est un chercheur en psychologie qui gagne en 2002 le prix Nobel d'économie pour avoir "intégrer des principes psychologique dans les sciences économique". Ses travaux sont liés à la formule de Bayes. L'histoire est racontée par Grant Sanderson (3B1B) dans sa vidéo sur la formule de Bayes.

Voci la description d'une personne : Steve est très timide et introverti. Il est d'une aide très préciseuse mais ne s'intéresse ni au monde ni aux gens. Il a besoin d'ordre et a une passion pour le détail.

Daniel Kahneman et son prix Nobel (1/3)

Daniel Kahneman est un chercheur en psychologie qui gagne en 2002 le prix Nobel d'économie pour avoir "intégrer des principes psychologique dans les sciences économique". Ses travaux sont liés à la formule de Bayes. L'histoire est racontée par Grant Sanderson (3B1B) dans sa vidéo sur la formule de Bayes.

Voci la description d'une personne : Steve est très timide et introverti. Il est d'une aide très préciseuse mais ne s'intéresse ni au monde ni aux gens. Il a besoin d'ordre et a une passion pour le détail.

Question: Quel est le plus probable : Steve est-il bibliothécaire ou fermier ?

Daniel Kahneman et son prix Nobel (1/3)

Daniel Kahneman est un chercheur en psychologie qui gagne en 2002 le prix Nobel d'économie pour avoir "intégrer des principes psychologique dans les sciences économique". Ses travaux sont liés à la formule de Bayes. L'histoire est racontée par Grant Sanderson (3B1B) dans sa vidéo sur la formule de Bayes.

Voci la description d'une personne : Steve est très timide et introverti. Il est d'une aide très préciseuse mais ne s'intéresse ni au monde ni aux gens. Il a besoin d'ordre et a une passion pour le détail.

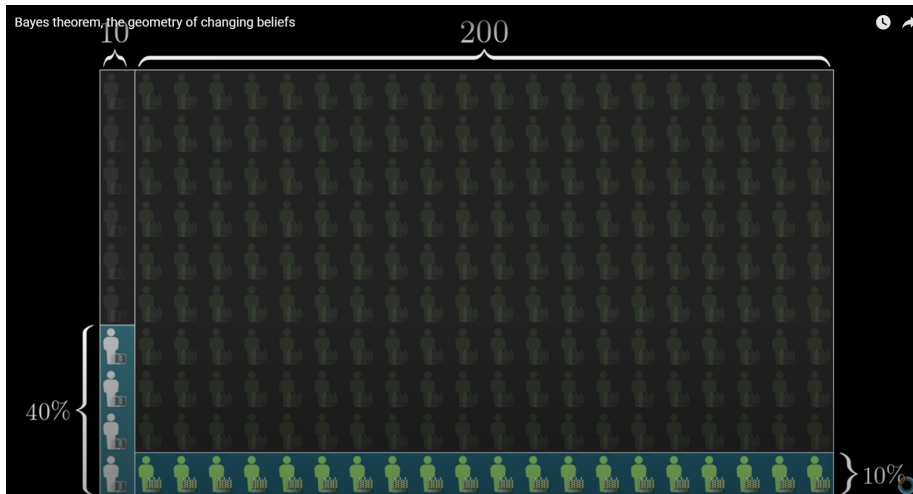
Question: Quel est le plus probable : Steve est-il bibliothécaire ou fermier ?

La plupart des interrogés répondent 'bibliothécaire'. Mais cette réponse est "irrationnelle"...

En effet, presque personne ne prend en compte le nombre de personne étant bibliothécaire ou fermier indépendamment de la description. Dans les faits, il y a 20 fois plus de fermiers que de bibliothécaires.

Si l'on suppose que 100% des bibliothécaires correspondent à la description, mais que seulement 10% des fermiers correspondent à la description : il est toujours deux fois plus probable que Steve soit un fermier.

Daniel Kahneman et son prix Nobel (3/3)



La formule de Bayes géométriquement expliquée

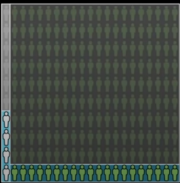
Bayes theorem, the geometry of changing beliefs

Bayes' theorem

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)} = \frac{P(H)P(E|H)}{P(H)P(E|H) + P(\neg H)P(E|\neg H)}$$

“Prior” $\rightarrow P(H) = 1/21$

“Likelihood”

$P(E|H) = 0.4$  $P(E|\neg H) = 0.1$

H = Hypothèse, E = Evidence (les données)

Daniel Kahneman et son prix Nobel (Trivia)

Petit trivia. On donne maintenant la description de Lynda : 'Lynda a 31 ans, est célibataire, franche et intelligente. Elle a majoré en philosophie. En tant qu'étudiante, elle était très concernée par les problèmes de discrimination et de justice sociale et participait aux manifestations contre le nucléaire'. Quel est le plus probable ?

Daniel Kahneman et son prix Nobel (Trivia)

Petit trivia. On donne maintenant la description de Lynda : 'Lynda a 31 ans, est célibataire, franche et intelligente. Elle a majoré en philosophie. En tant qu'étudiante, elle était très concernée par les problèmes de discrimination et de justice sociale et participait aux manifestations contre le nucléaire'. Quel est le plus probable ?

- Lynda est-elle une banquière ?
- Lynda est-elle une banquière active dans les mouvements féministes ?

Daniel Kahneman et son prix Nobel (Trivia)

Petit trivia. On donne maintenant la description de Lynda : 'Lynda a 31 ans, est célibataire, franche et intelligente. Elle a majoré en philosophie. En tant qu'étudiante, elle était très concernée par les problèmes de discrimination et de justice sociale et participait aux manifestations contre le nucléaire'. Quel est le plus probable ?

- Lynda est-elle une banquière ?
- Lynda est-elle une banquière active dans les mouvements féministes ?

85% des gens interrogés ont répondu la deuxième option...

Daniel Kahneman et son prix Nobel (Trivia)

Petit trivia. On donne maintenant la description de Lynda : 'Lynda a 31 ans, est célibataire, franche et intelligente. Elle a majoré en philosophie. En tant qu'étudiante, elle était très concernée par les problèmes de discrimination et de justice sociale et participait aux manifestations contre le nucléaire'. Quel est le plus probable ?

- Lynda est-elle une banquière ?
- Lynda est-elle une banquière active dans les mouvements féministes ?

85% des gens interrogés ont répondu la deuxième option...

Globalement, on n'a pas d'intuition en probabilités...

Variable aléatoire et lois de probabilités

Variable aléatoire

On appelle variable aléatoire toute fonction de Ω dans un ensemble.

Une variable aléatoire a pour rôle d'assigner à chaque évènement une valeur, que l'on considère ici comme un nombre. Exemples :

- Compter le nombre de "pile" après 10 lancers de pièces.
- La durée de vie d'une pièce mécanique...

Variable aléatoire

On appelle variable aléatoire toute fonction de Ω dans un ensemble.

Une variable aléatoire a pour rôle d'assigner à chaque évènement une valeur, que l'on considère ici comme un nombre. Exemples :

- Compter le nombre de "pile" après 10 lancers de pièces.
- La durée de vie d'une pièce mécanique...

Lorsque l'ensemble est dénombrable, on parle de loi discrète. Dans le cas contraire, on parle de loi continue.

Probabilité d'une variable aléatoire

On définit la probabilité d'une variable aléatoire par :

$$\mathbb{P}(X \in B) = \mathbb{P}(\{x \in \Omega | X(x) \in B\})$$

Probabilité d'une variable aléatoire

On définit la probabilité d'une variable aléatoire par :

$$\mathbb{P}(X \in B) = \mathbb{P}(\{x \in \Omega | X(x) \in B\})$$

L'idée de cette définition est que la probabilité qu'une variable aléatoire ait une valeur c'est la probabilité d'observer n'importe quel évènement qui lui donnera la valeur en question.

Probabilité d'une variable aléatoire

On définit la probabilité d'une variable aléatoire par :

$$\mathbb{P}(X \in B) = \mathbb{P}(\{x \in \Omega | X(x) \in B\})$$

L'idée de cette définition est que la probabilité qu'une variable aléatoire ait une valeur c'est la probabilité d'observer n'importe quel événement qui lui donnera la valeur en question.

Notations

On note :

- $\mathbb{P}(X = a)$ pour $\mathbb{P}(X \in \{a\})$
- $\mathbb{P}(X \leq a)$ pour $\mathbb{P}(X \in] - \infty, a])$...

Fonction de répartition (Cumulative distribution function - CDF)

Définition - fonction de répartition

Pour une variable aléatoire, X , on appelle F_X la fonction de répartition de la loi de X , la fonction définie par :

$$\begin{aligned} F_X &: \mathbb{R} \rightarrow [0, 1] \\ x &\mapsto \mathbb{P}(X \leq x) \end{aligned}$$

Fonction de répartition (Cumulative distribution function - CDF)

Définition - fonction de répartition

Pour une variable aléatoire, X , on appelle F_X la fonction de répartition de la loi de X , la fonction définie par :

$$\begin{aligned} F_X &: \mathbb{R} \rightarrow [0, 1] \\ x &\mapsto \mathbb{P}(X \leq x) \end{aligned}$$

Propriétés

Pour toute fonction de répartition :

- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
- F_X est croissante
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$
- $\lim_{x \rightarrow +\infty} F_X(x) = 1$

Cette fonction est caractéristique de la loi !!!

Densité - (Mass function, probability density function - pdf)

Densité - Cas discret

Pour une variable aléatoire X discrète, on appelle f_X 'densité de la loi de X ' la fonction :

$$\begin{aligned} f_X &: \mathbb{R} \rightarrow \mathbb{R}^+ \\ x &\mapsto \mathbb{P}(X = x) \end{aligned}$$

Densité - Cas continu

Pour une variable aléatoire X continue, on appelle f_X la densité de la loi de X la fonction, lorsqu'elle existe :

$$\begin{aligned} f_X &: \mathbb{R} \rightarrow \mathbb{R}^+ \\ x &\mapsto F'(x) \end{aligned}$$

Probabilité et densité

Dans le cas continu densité $\neq$ probabilité. Pour tout x réel, on a $f_X(x) \geq 0$, mais $\mathbb{P}(X = x) = 0$!!!

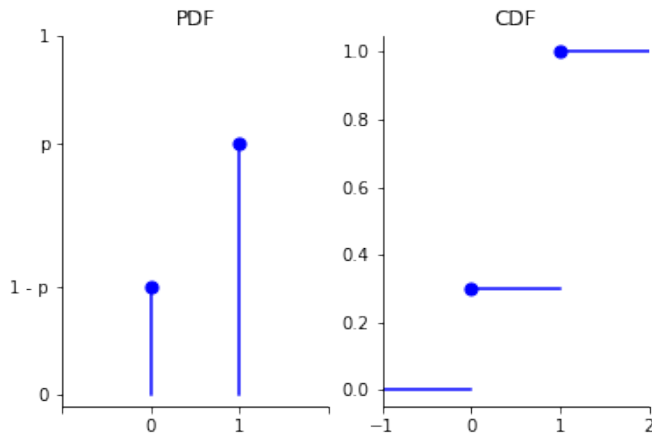
Probabilité et densité

Dans le cas continu densité $\neq$ probabilité. Pour tout x réel, on a $f_X(x) \geq 0$, mais $\mathbb{P}(X = x) = 0$!!!

Les valeurs de f_X donnent tout de même une bonne indication d'où sont les valeurs probables : $\mathbb{P}(x - \frac{\epsilon}{2} \leq X \leq x + \frac{\epsilon}{2}) \approx f_X(x)\epsilon$

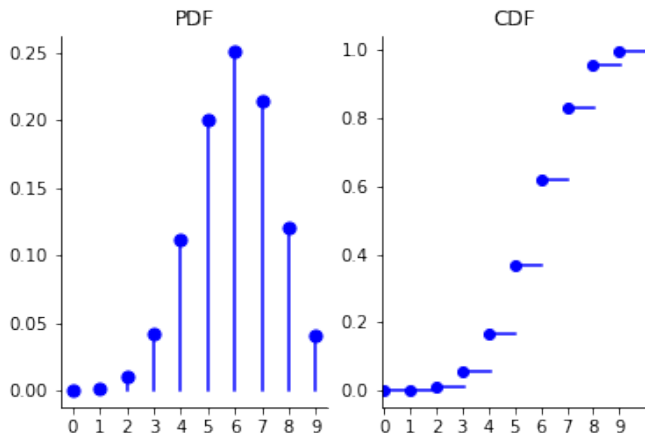
Lois discrètes classiques - Loi de Bernoulli

On note $X \sim \mathcal{B}(p)$. $\mathbb{P}(X = 0) = 1 - p$, $\mathbb{P}(X = 1) = p$



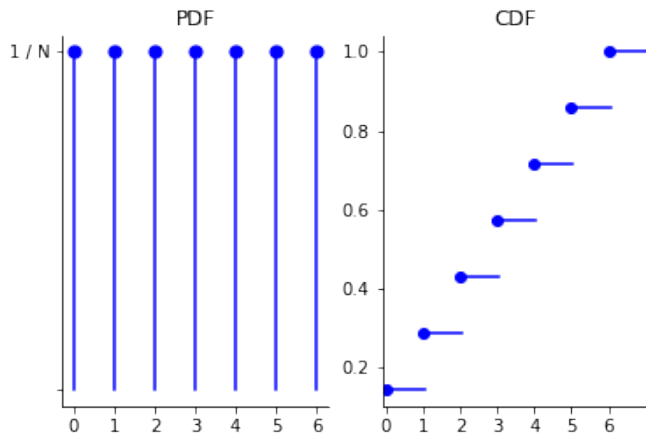
Lois discrètes classiques - Loi Binomiale

On note $X \sim \mathcal{B}(N, p)$. On montre que : $\mathbb{P}(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$



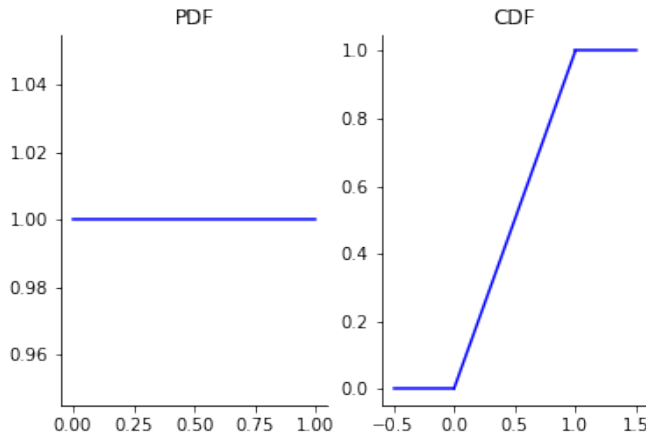
Lois discrètes classiques - Loi uniforme

On note $X \sim \mathcal{U}(S)$. $\mathbb{P}(X = i) = \frac{1}{|S|}$, pour $i \in S$



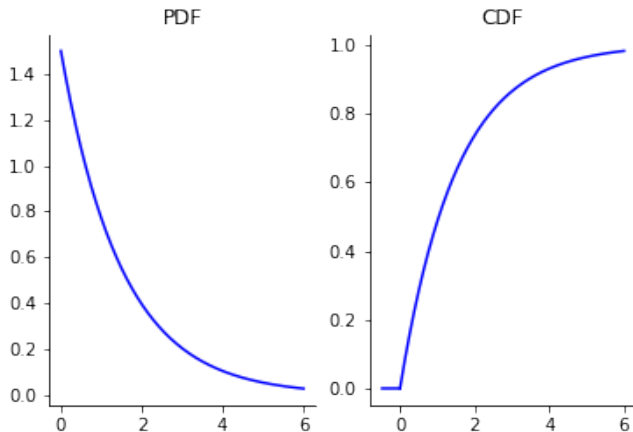
Lois continues classiques - Loi uniforme

On note $X \sim \mathcal{U}([a, b])$. $f_X(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$



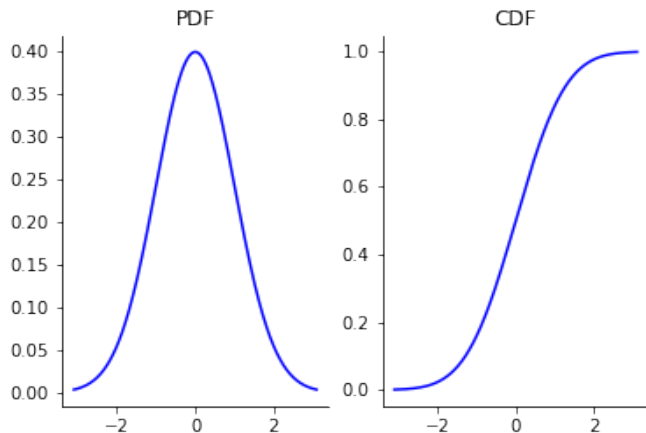
Lois continues classiques - Loi exponentielle

On note $X \sim \mathcal{E}(\lambda)$. $f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}^+}(x)$



Lois continues classiques - Loi normale

On note $X \sim \mathcal{N}(\mu, \sigma^2)$. $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



Espérance, Variance et Covariance

Espérance

Pour une variable aléatoire X continue (resp. discrète), on appelle espérance de X , la quantité, lorsqu'elle existe :

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf_X(x)dx \text{ (resp. } \sum_{x \geq 0} x\mathbb{P}(X = x))$$

Espérance

Pour une variable aléatoire X continue (resp. discrète), on appelle espérance de X , la quantité, lorsqu'elle existe :

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf_X(x)dx \text{ (resp. } \sum_{x \geq 0} x\mathbb{P}(X = x))$$

Cela est contre-intuitif, mais certaines variables aléatoires n'admettent pas d'espérance (loi de Cauchy).

Espérance

Pour une variable aléatoire X continue (resp. discrète), on appelle espérance de X , la quantité, lorsqu'elle existe :

$$\mathbb{E}[X] = \int_{\mathbb{R}} xf_X(x)dx \text{ (resp. } \sum_{x \geq 0} x\mathbb{P}(X = x))$$

Cela est contre-intuitif, mais certaines variables aléatoires n'admettent pas d'espérance (loi de Cauchy).

Théorème: formule de transfert

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f_X(x)dx \text{ (resp. } \sum_{i \geq 0} g(x)\mathbb{P}(X = x))$$

Soit X une v.a. suivant une loi uniforme sur $[a, b]$. On calcule $\mathbb{E}[X]$:

Soit X une v.a. suivant une loi uniforme sur $[a, b]$. On calcule $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \int_{\mathbb{R}} \frac{x}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{b+a}{2}$$

Soit X une v.a. suivant une loi uniforme sur $[a, b]$. On calcule $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \int_{\mathbb{R}} \frac{x}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{b+a}{2}$$

De même, en utilisant la formule de transfert on obtient :

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} \frac{x^2}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left. \frac{x^3}{3} \right|_a^b = \frac{a^2+ab+b^2}{3}$$

Soit X une v.a. suivant une loi uniforme sur $[a, b]$. On calcule $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \int_{\mathbb{R}} \frac{x}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{b+a}{2}$$

De même, en utilisant la formule de transfert on obtient :

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} \frac{x^2}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left. \frac{x^3}{3} \right|_a^b = \frac{a^2+ab+b^2}{3}$$

D'une manière plus générale, si X est une variable admettant une espérance et une densité, pour tous réels a et b :

Espérance - Exemple

Soit X une v.a. suivant une loi uniforme sur $[a, b]$. On calcule $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \int_{\mathbb{R}} \frac{x}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left. \frac{x^2}{2} \right|_a^b = \frac{b+a}{2}$$

De même, en utilisant la formule de transfert on obtient :

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} \frac{x^2}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left. \frac{x^3}{3} \right|_a^b = \frac{a^2+ab+b^2}{3}$$

D'une manière plus générale, si X est une variable admettant une espérance et une densité, pour tous réels a et b :

$$\mathbb{E}[aX + b] = \int_{\mathbb{R}} (ax+b) f_X(x) dx = a \int_{\mathbb{R}} x f_X(x) dx + b \int_{\mathbb{R}} f_X(x) dx = a\mathbb{E}[X] + b$$

L'espérance est donc linéaire !

Propriétés

- L'espérance est linéaire : $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$
- Si $X \leq 0$, $\mathbb{E}[X] \leq 0$
- Si $X \leq Y$, $\mathbb{E}[X] \leq \mathbb{E}[Y]$
- Si $\mathbb{E}[|X|] = 0$ alors, $X = 0$
- Soit A un ensemble. On a: $\mathbb{E}[\mathbf{1}_A(X)] = \mathbb{P}(X \in A)$

Propriétés de l'espérance (1 / 2)

Propriétés

- L'espérance est linéaire : $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$
- Si $X \leq 0$, $\mathbb{E}[X] \leq 0$
- Si $X \leq Y$, $\mathbb{E}[X] \leq \mathbb{E}[Y]$
- Si $\mathbb{E}[|X|] = 0$ alors, $X = 0$
- Soit A un ensemble. On a : $\mathbb{E}[\mathbf{1}_A(X)] = \mathbb{P}(X \in A)$

Propriétés

- Inégalité de Jensen : si f est une fonction convexe, $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$
- Inégalité de Markov : si X est une variable aléatoire positive et a un réel strictement positif, alors $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

Variance

Pour une v.a. X , on appelle variance de X , la quantité, lorsqu'elle existe :

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Variance

Pour une v.a. X , on appelle variance de X , la quantité, lorsqu'elle existe :

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Exemple pour la loi uniforme :

$$\mathbb{V}[X] = \int_{\mathbb{R}} (x - \frac{a+b}{2})^2 \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Variance

Pour une v.a. X , on appelle variance de X , la quantité, lorsqu'elle existe :

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Exemple pour la loi uniforme :

$$\mathbb{V}[X] = \int_{\mathbb{R}} (x - \frac{a+b}{2})^2 \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) dx = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Propriétés

- $\mathbb{V}[X] \geq 0$
- $\mathbb{V}[X] = 0 \Leftrightarrow X = c$
- $\mathbb{V}[aX + bY + c] =$
 $a^2\mathbb{V}[X] + b^2\mathbb{V}[Y] + 2ab \underbrace{\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]}_{\text{Cov}(X, Y)}$

Théorèmes et propriétés diverse(s).

Théorème central limite

Soient $X_1, \dots, X_N$ des observations indépendantes, issues d'une même loi admettant une variance. La moyenne de ces observations suit, lorsque N est grand une loi normale. Autrement dit :

$$\left(\frac{1}{N} \sum_{i=1}^N (X_i - \mathbb{E}[X_i]) \right) / \left(\sqrt{\sum_{i=1}^N \mathbb{V}[X_i]} \right) \rightarrow \mathcal{N}(0, 1)$$

Théorème central limite

Soient $X_1, \dots, X_N$ des observations indépendantes, issues d'une même loi admettant une variance. La moyenne de ces observations suit, lorsque N est grand une loi normale. Autrement dit :

$$\left(\frac{1}{N} \sum_{i=1}^N (X_i - \mathbb{E}[X_i]) \right) / \left(\sqrt{\frac{1}{N} \sum_{i=1}^N \mathbb{V}[X_i]} \right) \rightarrow \mathcal{N}(0, 1)$$

Sous certaines conditions, le théorème est vrai pour des variables indépendantes de lois différentes (Lyapunov), ou même lorsqu'elles sont dépendantes, mais pas trop quand même... (Billingsley 1995).

Loi des grands nombres

Soit $X_1, \dots, X_N$ des observations indépendantes, issues d'une même loi admettant une espérance. La moyenne empirique converge vers l'espérance, c'est à dire :

$$\frac{1}{N} \sum_{i=1}^N X_i \rightarrow \mathbb{E}[X_1]$$

Loi des grands nombres

Soit $X_1, \dots, X_N$ des observations indépendantes, issues d'une même loi admettant une espérance. La moyenne empirique converge vers l'espérance, c'est à dire :

$$\frac{1}{N} \sum_{i=1}^N X_i \rightarrow \mathbb{E}[X_1]$$

Sous certaines conditions, le théorème est vrai pour des observations non indépendantes (théorème ergodique) ou des observations de lois différentes (théorème de Kolmogorov).

Famille exponentielle

Soit θ un vecteur de paramètres. Une loi est dite appartenant à la famille exponentielle, si la densité s'écrit comme :

$$f_X(x; \theta) = a(x)b(\theta)e^{\eta(\theta) \cdot T(x)}$$

Avec a , b , η et T des fonctions connues.

Famille exponentielle

Soit θ un vecteur de paramètres. Une loi est dite appartenant à la famille exponentielle, si la densité s'écrit comme :

$$f_X(x; \theta) = a(x)b(\theta)e^{\eta(\theta) \cdot T(x)}$$

Avec a , b , η et T des fonctions connues.

- Cette famille regroupe (entre autre) : lois normales, lois exponentielles, chi2, ...
- Cette famille de loi est très importante en statistique. Cela vient du fait qu'une somme de variables indépendantes correspond à un produit des densités.

Vraisemblance et maximum de vraisemblance

Supposons que l'on dispose d'une famille de loi paramétrée par un certain θ : $f_X(\cdot, \theta)$. On observe un certain nombre de données indépendantes $X_1, \dots, X_N$ issues d'une loi $f_X(\cdot, \theta^*)$ avec θ^* inconnu.

Vraisemblance et maximum de vraisemblance

Supposons que l'on dispose d'une famille de loi paramétrée par un certain θ : $f_X(\cdot, \theta)$. On observe un certain nombre de données indépendantes $X_1, \dots, X_N$ issues d'une loi $f_X(\cdot, \theta^*)$ avec θ^* inconnu.

Vraisemblance

On appelle vraisemblance, la fonction :

$$\theta \mapsto \prod_{i=1}^N f_X(x_i, \theta)$$

Vraisemblance et maximum de vraisemblance

Supposons que l'on dispose d'une famille de loi paramétrée par un certain θ : $f_X(\cdot, \theta)$. On observe un certain nombre de données indépendantes $X_1, \dots, X_N$ issues d'une loi $f_X(\cdot, \theta^*)$ avec θ^* inconnu.

Vraisemblance

On appelle vraisemblance, la fonction :

$$\theta \mapsto \prod_{i=1}^N f_X(x_i, \theta)$$

- Cette fonction peut se lire comme étant 'la probabilité d'observer ces données sachant θ '.
- Maximiser cette fonction sur θ , revient à chercher 'le paramètre le plus probable sachant les données'.
- Les lois de la famille exponentielle se comportent bien avec cette fonction (produit $\implies$ somme !).

Probabilités multivariées

Parfois, il arrive que l'on observe plusieurs variables aléatoires en même temps; c'était le cas du lancé de dés. Ces dernières peuvent interagir entre-elles et les étudier séparément n'est parfois pas possible. Il peut alors être utiles de pouvoir considérer la loi de probabilités qu'elles définissent toutes ensemble. On parle alors de loi jointe.

Parfois, il arrive que l'on observe plusieurs variables aléatoires en même temps; c'était le cas du lancé de dés. Ces dernières peuvent interagir entre-elles et les étudier séparément n'est parfois pas possible. Il peut alors être utiles de pouvoir considérer la loi de probabilités qu'elles définissent toutes ensemble. On parle alors de loi jointe.

Fonction de répartition multivariée

Soit $\mathbf{X} = (X_1, \dots, X_n)$ un vecteur aléatoire de $\mathbb{R}^n$. On appelle fonction de répartition, la fonction :

$$\begin{aligned} F_{\mathbf{X}} : \mathbb{R}^n &\rightarrow [0, 1] \\ x &\mapsto \mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) \end{aligned}$$

Cette fonction possède exactement les mêmes propriétés que sa consœur 1D décrite plus haut.

Densité jointe

De manière analogue à la densité dans le cas univarié, on définit la densité jointe comme la dérivée de la fonction de répartition :

$$\begin{aligned} f_{\mathbf{X}} : \mathbb{R}^n &\rightarrow [0, 1] \\ x &\mapsto \frac{\partial^n F_{\mathbf{X}}}{\partial x_1 \dots \partial x_n} \quad (\text{discret: } \mathbb{P}(X_1 = x_1, \dots, X_N = x_n)) \end{aligned}$$

Densité jointe

De manière analogue à la densité dans le cas univarié, on définit la densité jointe comme la dérivée de la fonction de répartition :

$$\begin{aligned} f_{\mathbf{X}} : \mathbb{R}^n &\rightarrow [0, 1] \\ x &\mapsto \frac{\partial^n F_{\mathbf{X}}}{\partial x_1 \dots \partial x_n} \quad (\text{discret: } \mathbb{P}(X_1 = x_1, \dots, X_N = x_n)) \end{aligned}$$

Lois marginales

Soit $i \in [0, n]$. On appelle loi marginale de la variable X_i la fonction :

$$\begin{aligned} f_{X_i} : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f_{\mathbf{X}} dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n \end{aligned}$$

Dans le cas multivariée, l'espérance n'est plus un nombre, mais un vecteur défini par :

Vecteur espérance

$$\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$$

La formule de transfert s'écrit :

$$\mathbb{E}[g(X_1, \dots, X_N)] = \int \cdots \int g(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Dans le cas multivariée, l'espérance n'est plus un nombre, mais un vecteur défini par :

Vecteur espérance

$$\mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$$

La formule de transfert s'écrit :

$$\mathbb{E}[g(X_1, \dots, X_n)] = \int \cdots \int g(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Propriétés

- Pour une matrice A quelconque et un vecteur b ,
 $\mathbb{E}[AX + b] = A\mathbb{E}[X] + b$
- $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Dans le cas multivariée, la variance est une matrice appelée matrice de covariance:

Matrice de covariance

$$\mathbb{V}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

Dans le cas multivariée, la variance est une matrice appelée matrice de covariance:

Matrice de covariance

$$\mathbb{V}[\mathbf{X}] = \mathbb{E}[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^T]$$

Propriétés

- La matrice de covariance est semi-définie positive
- $\mathbb{V}[A\mathbf{X}] = A\mathbb{V}[\mathbf{X}]A^T$

Probabilité conditionnelle

Pour deux variables aléatoires, on définit la probabilité conditionnelle de x sachant y :

$$\mathbb{P}_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) = \frac{P(\{X=x\} \cap \{Y=y\})}{\mathbb{P}(X=x)}$$

Probabilité conditionnelle

Pour deux variables aléatoires, on définit la probabilité conditionnelle de x sachant y :

$$\mathbb{P}_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) = \frac{P(\{X=x\} \cap \{Y=y\})}{\mathbb{P}(X=x)}$$

Densité conditionnelle

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

Indépendance

On dit que deux variables aléatoires sont indépendantes si :

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y), F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Indépendance

On dit que deux variables aléatoires sont indépendantes si :

$$\mathbb{P}(X \leq x, Y \leq y) = \mathbb{P}(X \leq x)\mathbb{P}(Y \leq y), F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Espérance du produit

Si X et Y sont deux variables aléatoires indépendantes, on a :

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Espérance conditionnelle

L'espérance conditionnelle de X sachant Y est une variable aléatoire notée $\mathbb{E}[X|Y]$.

$$\mathbb{E}[X|Y=y] = \frac{1}{f_Y(y)} \int_{\mathbb{R}} x f_{X,Y}(x,y) dx$$