MAP 311 - Aléatoire

Leçon 2

2016-2017

Espace de probabilité

On considère une expérience aléatoire.

- On note Ω l'ensemble des résultats possibles.
- On considère un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ qui contient les événements que l'on veut étudier.
  Dans le cas où Ω est dénombrable, on prend $\mathcal{A} = \mathcal{P}(\Omega)$.
  Mais dans le cas où Ω n'est pas dénombrable, $\mathcal{P}(\Omega)$ s'avère trop gros !
- On introduit une probabilité $\mathbb{P}$ qui quantifie la vraisemblance des événements de $\mathcal{A}$.

On appelle le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité [Kolmogorov (1933)].

Tribu

Définition

$\mathcal{A} \subset \mathcal{P}(\Omega)$ est une tribu si elle vérifie les propriétés suivantes :
- $\Omega \in \mathcal{A}$.
- Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.
- Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$, alors $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Par conséquent :
- $\emptyset \in \mathcal{A}$.
- Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'éléments de $\mathcal{A}$, alors $\cap_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Exemples :
- $\mathcal{A} = \{\emptyset, \Omega\}$ est une tribu.
- Si $A \subset \Omega$, alors $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$ est une tribu.
- $\mathcal{P}(\Omega)$ est une tribu.

Tribu borélienne

Si Ω est non dénombrable (par exemple, $\Omega = [0,1]$ ou $\Omega = \mathbb{R}$), alors $\mathcal{P}(\Omega)$ est trop riche pour que l'on puisse en "mesurer" tous les éléments de manière non triviale.

Définition

La tribu borélienne de $\mathbb{R}$ est la plus petite tribu qui contient les intervalles de la forme $]-\infty, a]$ pour $a \in \mathbb{Q}$.

On note que tous les intervalles, tous les ouverts et fermés de $\mathbb{R}$ sont dans la tribu borélienne. Mais $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.
Pour des raisons mathématiques, on va toujours munir $\mathbb{R}$ de cette tribu.
Ceci nous permettra de caractériser les probabilités sur $\mathbb{R}$.

Probabilité

Définition

Une probabilité sur $(\Omega, \mathcal{A})$ *est une application* $\mathbb{P}$ *de* $\mathcal{A}$ *dans* $[0,1]$ *telle que:*
(i) $\mathbb{P}(\Omega) = 1$.
(ii) Pour toute suite $(A_i)_{i\in I}$ *d'éléments de* $\mathcal{A}$ *deux-à-deux disjoints avec* I *au plus dénombrable, on a* $\mathbb{P}(\cup_{i\in I} A_i) = \sum_{i\in I} \mathbb{P}(A_i)$.

L'axiome (ii) s'appelle "axiome de σ-additivité".
Remarque 1 : dans le cadre de la théorie de la mesure, $\mathbb{P}$ est une mesure positive de masse 1 sur l'espace mesurable $(\Omega, \mathcal{A})$.
Remarque 2 : Si $\Omega = [0,1]$ et $\mathcal{A} = \mathcal{P}(\Omega)$, alors il n'existe pas de probabilité $\mathbb{P}$ telle que $\mathbb{P}([a,b]) = b - a$!
Remarque 3 : Si $\Omega = [0,1]$ et $\mathcal{A} = \mathcal{B}(\Omega)$, alors il existe une probabilité $\mathbb{P}$ telle que $\mathbb{P}([a,b]) = b - a$! C'est la mesure de Lebesgue.

Variable aléatoire réelle et sa loi

$\mathbb{R}$ sera toujours muni de sa tribu borélienne $\mathcal{B}(\mathbb{R})$.

Définition

Une application $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *est appelée variable aléatoire réelle si pour tout* $B \in \mathcal{B}(\mathbb{R})$, *on a*

$$X^{-1}(B) = \{X \in B\} \in \mathcal{A}$$

On réclame donc que les parties $\{X \in B\}$ de Ω soient des événements de $\mathcal{A}$.

On peut alors considérer les probabilités $\mathbb{P}(X \in B)$.

Remarque : Il est suffisant de vérifier que $\{X \leq x\} \in \mathcal{A}$ pour tout $x \in \mathbb{R}$.

Définition

La loi de X *est l'application* $P_X : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to [0, 1]$ *définie par* $P_X(B) = \mathbb{P}(X \in B)$.

Résultat : P_X est une probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Cas particulier : variables aléatoires discrètes réelles.
Soit $X : (\Omega, \mathcal{A}) \to (E, \mathcal{P}(E))$ avec $E \subset \mathbb{R}$ fini ou dénombrable.

- X est une v.a. dès que $\{X = x\} \in \mathcal{A}$ pour tout $x \in E$.
  En effet, si $B \in \mathcal{P}(E)$, alors $B = \cup_{x \in B}\{x\}$, B au plus dénombrable, et

$$\{X \in B\} = \cup_{x \in B}\{X = x\} \in \mathcal{A}$$

- La loi de X est caractérisée par $\{p_x, x \in E\}$ où $p_x = \mathbb{P}(X = x)$.
  En effet, si $B \in \mathcal{P}(E)$, alors $B = \cup_{x \in B}\{x\}$, et

$$P_X(B) = \mathbb{P}(X \in B) = \sum_{x \in B} \mathbb{P}(X = x) = \sum_{x \in B} p_x$$

Fonction de répartition

Définition

On appelle fonction de répartition de X la fonction suivante :

$$F_X(x) = P_X(]-\infty, x]) = \mathbb{P}(X \leq x), \quad \forall x \in \mathbb{R}.$$

Proposition

La fonction de répartition F_X caractérise la probabilité P_X et elle vérifie les trois conditions suivantes :
(i) elle est croissante,
(ii) elle est continue à droite,
(iii) $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to+\infty} F_X(x) = 1$.

Idée de preuve: Si $B = \cup_{i=1}^{n}]x_i, y_i]$, alors
$P_X(B) = \sum_{i=1}^{n}(F_X(y_i) - F_X(x_i))$, dès que $x_i < y_i < x_{i+1}$.

Proposition

Si deux variables aléatoires ont la même fonction de répartition, alors elles ont la même loi.

Cette proposition est un résultat d'unicité.

Proposition

Si F vérifie les conditions (i)-(iii) ci-dessus, c'est la fonction de répartition d'une (unique) probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Cette proposition est un résultat d'existence.
Résultat : Si F vérifie (i)-(iii) alors il existe une unique probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ telle que pour tout $x \in \mathbb{R}$,

$$F(x) = \mu(]-\infty, x]).$$

Attention : On ne peut pas en général définir cette probabilité sur la tribu $\mathcal{P}(\mathbb{R})$ des parties de $\mathbb{R}$!

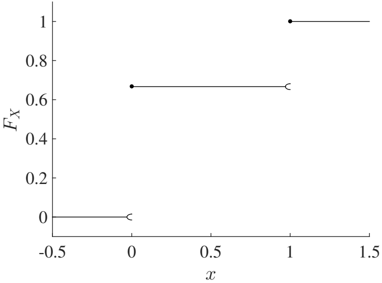
Fonctions indicatrices et variables de Bernoulli

- Une v.a. X à valeurs dans $\{0,1\}$ de loi de Bernoulli $p \in [0,1]$:

$$\mathbb{P}(X=0) = 1-p, \quad \mathbb{P}(X=1) = p,$$

a pour fonction de répartition

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & \text{si } x < 0, \\ 1-p & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$



- Soit $A \in \mathcal{A}$. Alors $\mathbf{1}_A$ est une v.a. à valeurs dans $\{0,1\}$:

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

C'est une v.a. de Bernoulli de paramètre $\mathbb{P}(A)$ car $\mathbb{P}(\mathbf{1}_A = 1) = \mathbb{P}(A)$.

Variables aléatoires discrètes réelles

Si X prend un nombre fini de valeurs, sa fonction de répartition est une fonction en escalier.

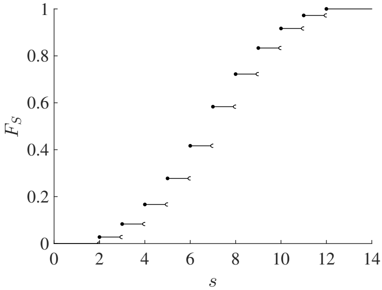
Si X est à valeurs dans $\{x_i, i \in I\}$, de loi $\{p_i, i \in I\}$, alors

$$F_X(x) = P_X\left(]-\infty, x]\right) = P_X\left(\bigcup_{i\,,\,x_i \leq x} \{x_i\}\right) = \sum_{i\,,\,x_i \leq x} P_X\left(\{x_i\}\right) = \sum_{i\,,\,x_i \leq x} p_i$$

La fonction de répartition saute de p_i en x_i.

Exemple : $S =$ somme des lancers de deux dés.

$$p_S(s) = \frac{\min\left[(13-s)(s-1)\right]}{36},$$
$s \in \{2, \ldots, 12\}$



Variables aléatoires réelles générales

Soit X une variable aléatoire réelle de fonction de répartition F_X.
Comme F_X est croissante, elle admet une limite à gauche en chaque point, notée $F_X(x-)$, et $F_X(x-) = \mathbb{P}(X < x) = P_X(]-\infty, x[)$.

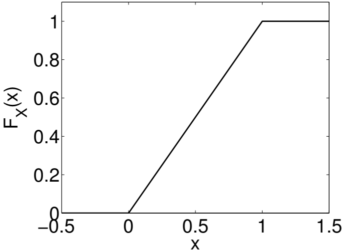
$$P_X(]x, y]) = \mathbb{P}(x < X \leq y) = F_X(y) - F_X(x),$$
$$P_X([x, y]) = \mathbb{P}(x \leq X \leq y) = F_X(y) - F_X(x-).$$

En particulier, $P_X(\{x\}) = F_X(x) - F_X(x-)$ est le saut de la fonction F_X au point x. On a donc

Proposition

$$P_X(\{x\}) = \mathbb{P}(X = x) = 0 \iff F_X \textit{ est continue en } x.$$

Exemple : $F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ x & \text{si } x \in [0,1] \\ 1 & \text{si } x \geq 1 \end{cases}$



F_X détermine une probabilité P_X appelée probabilité uniforme sur $[0,1]$ ou encore mesure de Lebesgue sur $[0,1]$, qui correspond à la "longueur" des intervalles : $P_X([x,y]) = F_X(y) - F_X(x) = y - x$.
On dit que X suit la loi uniforme sur $[0,1]$. Pour tout $0 \leq x \leq y \leq 1$:

$$\mathbb{P}\big(X \in [x,y]\big) = y - x$$

On écrit $X \sim \mathcal{U}(0,1)$.

Variables aléatoires réelles à densité

Définition

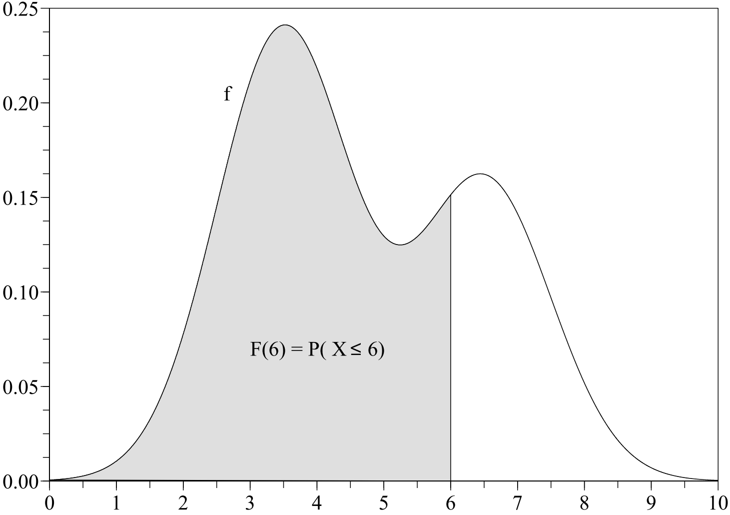
1) Une fonction réelle $f : \mathbb{R} \to \mathbb{R}$ *est appelée une densité de probabilité, ou simplement une "densité", si elle est positive, intégrable, et vérifie*

$$\int_{-\infty}^{+\infty} f(x)dx = 1.$$

2) Soit X *une variable aléatoire réelle. On dit que* X *a une loi de densité* f *et que* P_X *a pour densité* f*, si pour tout réel* x*,*

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Remarque : Si X est à densité, alors $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$.



Graphe de la densité f.

Proposition

1) Soit X de loi de densité f. La fonction F_X est dérivable en tout point x où f est continue, et

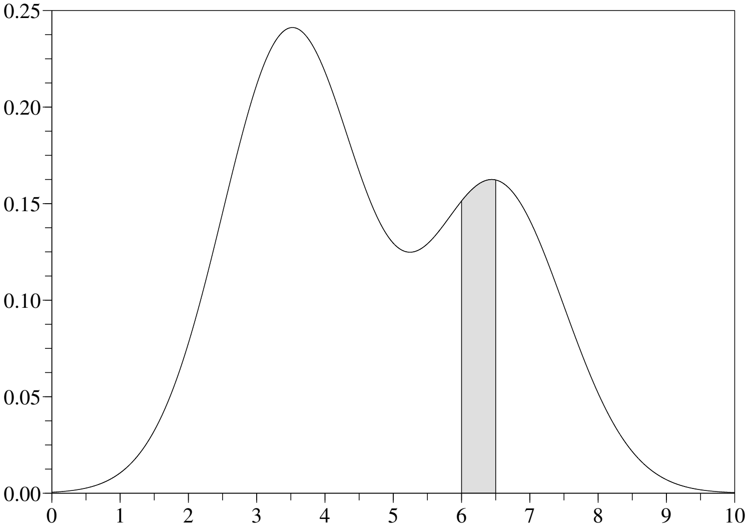
$$F'_X(x) = f(x).$$

2) Si F_X est dérivable en tout point, alors X a une loi de densité $f = F'_X$.

Preuve : Dérivée d'une intégrale fonction de sa borne supérieure.

Interprétation : si Δx est un "petit" accroissement de la quantité x, on a (du moins si f est continue en x) :

$$f(x) \simeq \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} = \frac{P_X([x, x+\Delta x])}{\Delta x}$$



$$f(6) \simeq \frac{F(6{,}5)-F(6)}{0{,}5}$$

Variable aléatoire uniforme

Variable aléatoire uniforme sur $[a, b]$, $a < b$: $X : \Omega \to [a, b]$, et

$$F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } x \in [a, b[\\ 1 & \text{si } x \geq b \end{cases}$$

$$f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

Autant de chances de tomber au voisinage de chaque point de $[a, b]$.
On note $X \sim \mathcal{U}(a, b)$.
Exercice : montrer que $X \sim \mathcal{U}(a, b)$ ssi $Y = \frac{X-a}{b-a} \sim \mathcal{U}(0, 1)$.

Simulation d'une variable aléatoire

But : générer les valeurs d'une variable aléatoire X de fonction de répartition F donnée.
Problème : il n'y a rien d'aléatoire dans un ordinateur !
Cas le plus simple : simuler les valeurs d'une variable aléatoire U de loi uniforme sur $[0, 1]$.
$\hookrightarrow$ On utilise un générateur de nombres pseudo-aléatoires.

Simulation d'une variable aléatoire

Exemple : (un générateur congruentiel linéaire à 32 bits) :

$$x_{n+1} = a x_n \mod b,$$

avec $a = 7^5$, $b = 2^{31} - 1$, et un certain x_0.
$\hookrightarrow$ donne une suite d'entiers dans $\{0, 1, \ldots, 2^{31} - 2\}$.
$\hookrightarrow$ $u_n = x_n/(2^{31} - 1)$ donne une suite de nombres quasi-réels (sur 32 bits) entre 0 et 1.
Note : la suite est périodique, de période $2^{31} - 2$.
C'est le générateur mcg16807 de matlab (utilisé dans les premières versions).
Aujourd'hui : matlab et python utilisent mt19937ar, basé sur l'algorithme de Mersenne Twister (récurrence linéaire matricielle). Sa période est $2^{19937} - 1$.

Simulation d'une variable aléatoire

Pour simuler X, on peut très souvent se ramener à la simulation de U.
Exemple : Soit X variable aléatoire de Bernoulli de paramètre p.
(i) on tire $U \sim \mathcal{U}(0,1)$.
(ii) si $U \in [0, 1-p]$, on pose $X = 0$, sinon on pose $X = 1$.
Alors $X = \mathbf{1}_{U>1-p}$ suit une loi de Bernoulli de paramètre $\mathbb{P}(U > 1-p) = p$.

Méthode générale : inversion de la fonction de répartition de X.

Proposition

Supposons que F soit continue et strictement croissante. Alors $F^{-1} :]0,1[\to \mathbb{R}$ existe et $X = F^{-1}(U)$ est une variable aléatoire de loi P_X, où P_X a pour fonction de répartition F.

Preuve : $\mathbb{P}(X \leq x) = \mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$.
Conclusion : Pour simuler X, on simule U et on calcule son image par F^{-1}.

Variable aléatoire exponentielle

Définition

X suit la loi exponentielle de paramètre $\lambda > 0$ si $X : \Omega \to \mathbb{R}^+$ et P_X est la loi de densité

$$f(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0,+\infty[}(x)$$

On note $X \sim \mathcal{E}(\lambda)$.

Modélisation de durée de vie ou de temps d'attente entre événements : durée de vie d'une bactérie, durée d'une conversation téléphonique, temps entre deux tremblements de terre.

Simulation d'une variable aléatoire de loi $\mathcal{E}(\lambda)$:
Montrer que si $U \sim \mathcal{U}(0,1)$ et $X = -\frac{1}{\lambda} \log U$, alors $X \sim \mathcal{E}(\lambda)$.

Propriété de non-vieillissement (absence de mémoire) : Soit X une v.a. de loi $\mathcal{E}(\lambda)$, alors pour tout $t, s \in \mathbb{R}^+$

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s) = e^{-\lambda s}$$

Exemple : durée de vie d'un atome radioactif.
Propriété caractéristique : Si une v.a. positive X vérifie

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

pour tout $t, s \in \mathbb{R}^+$, alors $X = 0$ ou X suit une loi exponentielle.
Preuve : $t \mapsto \rho(t) = \mathbb{P}(X > t)$ (décroissante) vérifie $\rho(t+s) = \rho(t)\rho(s)$.
Version rapide en supposant la dérivabilité : on a (en dérivant en s, avec $s = 0$), $\rho'(t) = -\lambda\rho(t)$ avec $\lambda = -\rho'(0) \geq 0$. Ainsi $\rho(t) = e^{-\lambda t}$ (car $\rho(0) = 1$), et $f(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0,+\infty[}(t)$ (loi exponentielle).

Propriété de non-vieillissement (absence de mémoire) : Soit X une v.a. de loi $\mathcal{E}(\lambda)$, alors pour tout $t, s \in \mathbb{R}^+$

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s) = e^{-\lambda s}$$

Exemple : durée de vie d'un atome radioactif.
Propriété caractéristique : Si une v.a. positive X vérifie

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

pour tout $t, s \in \mathbb{R}^+$, alors $X = 0$ ou X suit une loi exponentielle.
Preuve : $t \mapsto \rho(t) = \mathbb{P}(X > t)$ (décroissante) vérifie $\rho(t + s) = \rho(t)\rho(s)$.
Version rigoureuse sans supposer la dérivabilité : on note $\rho_1 = \rho(1)$. On montre par récurrence pour tout $n \in \mathbb{N}$: $\rho(n) = \rho_1^n$. Puis on montre pour tout $q = n/p \in \mathbb{Q}^+$: $\rho(n) = \rho(n/p)^p$ et donc $\rho(q) = \rho_1^{n/p} = \rho_1^q$. Puis on montre pour tout $t \in \mathbb{R}^+$: $\rho((n+1)[t/n]) \leq \rho(t) \leq \rho(n[t/n])$ par décroissance de ρ, et on applique le théorème des gendarmes pour obtenir $\rho(t) = \rho_1^t$.

Variable aléatoire normale (variable gaussienne)

Définition

On appelle variable aléatoire normale centrée réduite une variable aléatoire réelle X de loi de densité

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

Remarque : On a bien $I = \int_{-\infty}^{+\infty} f(x)dx = 1$.

En effet, en considérant I^2 et en passant aux coordonnées polaires :

$$\begin{aligned} I^2 = \iint_{\mathbb{R}^2} f(x)f(y)dxdy &= \frac{1}{2\pi} \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2+y^2}{2}\right)dxdy \\ &= \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\infty} \exp\left(-\frac{r^2}{2}\right) rdr = 1 \end{aligned}$$

f n'a pas de primitive évidente : table ou approximation numériques (cf fonction "erf").

Variable aléatoire normale (variable gaussienne)

Définition

Pour $m \in \mathbb{R}$ et $\sigma > 0$, une variable normale de loi $\mathcal{N}(m, \sigma^2)$ a pour densité

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(y-m)^2}{2\sigma^2}\Big)$$

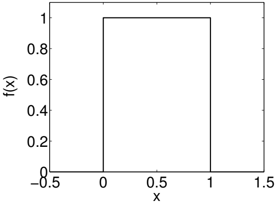
On a $Y \sim \mathcal{N}(m, \sigma^2)$ ssi $\frac{Y-m}{\sigma} \sim \mathcal{N}(0,1)$.

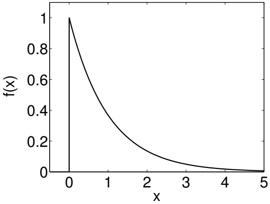
La distribution normale se rencontre "partout" :

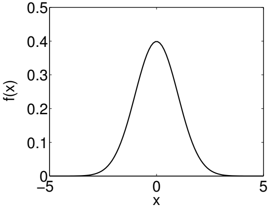
- taille d'un individu choisi au hasard,
- composantes de la vitesse d'une molécule de gaz,
- erreur de mesure d'une quantité physique.

Approximation de somme de variables indépendantes et de même loi : théorème de la limite centrale.

Résumé : lois à densité usuelles



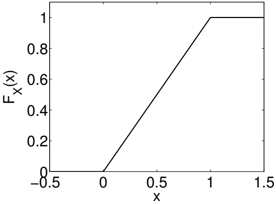


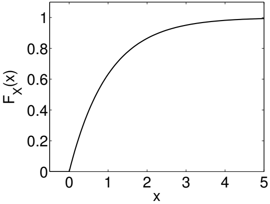


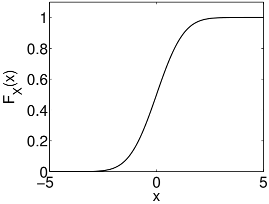
$$f(x) = \mathbf{1}_{[0,1]}(x)$$
$\mathcal{U}(0,1)$

$$f(x) = e^{-x}\mathbf{1}_{[0,\infty[}(x)$$
$\mathcal{E}(1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{x^2}{2}\Big)$$
$\mathcal{N}(0,1)$







Espérance d'une variable aléatoire réelle

Cas dénombrable : Si Ω est au plus dénombrable, alors X est à valeurs dans $E \subset \mathbb{R}$ au plus dénombrable, et

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

Cas général : On a une somme infinie non dénombrable.
Comment définir $\int_\Omega X(\omega)\mathbb{P}(d\omega)$? Ω peut être très compliqué !
Espérance : intégrale de Lebesgue pour la mesure abstraite $\mathbb{P}$.
On connaît bien l'espace d'arrivée de $X : \Omega \to \mathbb{R}$.
Cas simple : Si Y ne prend qu'un nombre fini de valeurs $y_1, \ldots, y_p$ (variable aléatoire étagée). Alors

$$\mathbb{E}(Y) = \sum_{i=1}^{p} y_i \mathbb{P}(Y = y_i)$$

Idée : On va approcher une variable aléatoire par une suite de variables aléatoires étagées.

Construction de l'espérance

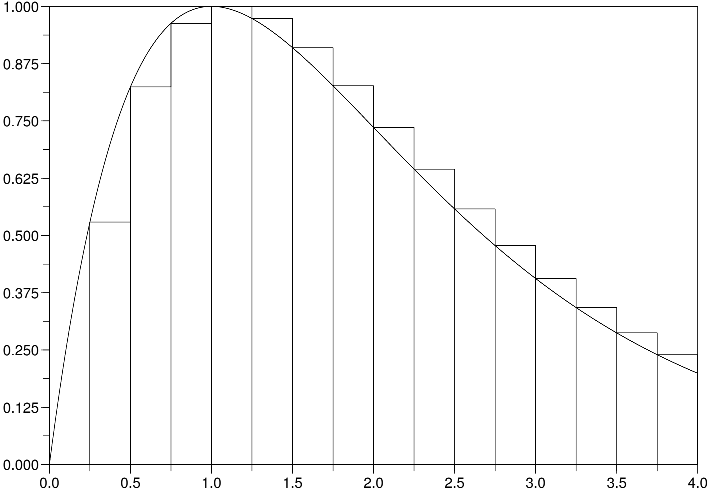
Soit X une variable aléatoire réelle positive.
On définit la suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires étagées croissant vers X :

$$X_n(\omega) = \begin{cases} k2^{-n} & \text{si } k2^{-n} \leq X(\omega) < (k+1)2^{-n} \text{ pour } k = 0, \dots, n2^n - 1, \\ n & \text{sinon} \end{cases}$$

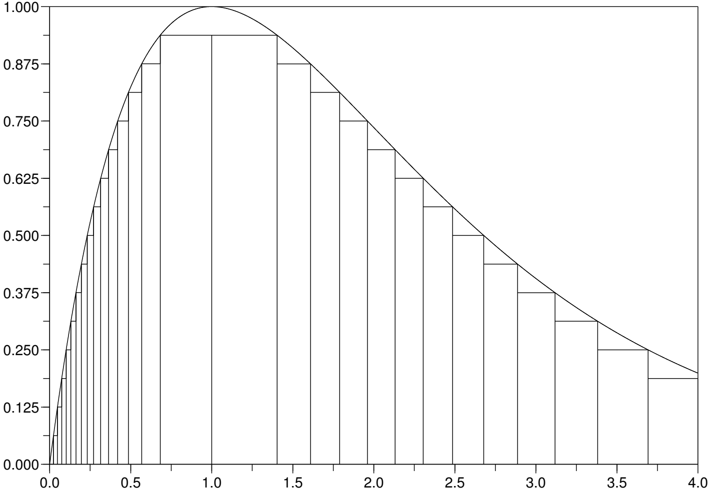
X_n est à valeurs dans un sous-ensemble fini de $[0, n]$. Donc $\mathbb{E}(X_n)$ est bien définie.
Pour tout ω, $X_n(\omega) \leq X_{n+1}(\omega)$, donc $\mathbb{E}(X_n) \leq \mathbb{E}(X_{n+1})$.
On pose

$$\mathbb{E}(X) = \lim_{n\to+\infty} \mathbb{E}(X_n)$$

Cette limite ne dépend pas de la suite de variables aléatoires étagées croissant vers X. Elle peut valoir $+\infty$.



Découpage “à la Riemann”, en 16 intervalles de même longueur de l’abscisse.



Découpage “à la Lebesgue”, en 16 intervalles de même longueur de l’ordonnée

Soit X une variable aléatoire réelle de signe quelconque.
Soient les variables aléatoires positives $X^+ = \sup(X, 0)$, $X^- = \sup(-X, 0)$. Alors $X = X^+ - X^-$ et $|X| = X^+ + X^-$.
X est intégrable si $\mathbb{E}(|X|) < +\infty$. Alors,

$$\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$$

$\mathbb{E}(X)$ est l'intégrale de Lebesgue de X pour la mesure $\mathbb{P}$, on la note aussi $\int_\Omega X(\omega)\mathbb{P}(d\omega)$.

L^1 : espace vectoriel des variables aléatoires intégrables.
Par passage la limite, toutes les propriétés de l'espérance vues dans le cas discret restent vraies.
Si $X, Y \in L^1$, pour $a, b \in \mathbb{R}$,

- Si $X(\omega) = a$ pour tout ω, alors $\mathbb{E}(X) = a$.
- Linéarité :
$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$
- Positivité :
$$|\mathbb{E}(X)| \leq \mathbb{E}(|X|).$$
Si $X \geq 0$, alors $\mathbb{E}(X) \geq 0$.
Si $X \leq Y$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$.

Remarque : Soient $A \in \mathcal{A}$ et $X = \mathbf{1}_A$. Alors $\mathbb{E}(X) = \mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A)$.

Variance

L^2 : ensemble des variables aléatoires de carré intégrable.
Si $X \in L^2$, alors $X \in L^1$ (i.e. $\mathbb{E}(|X|) < +\infty$) car $|X| \leq (1 + X^2)/2$.

Définition

Si $X \in L^2$, alors

$$\mathrm{Var}(X) = \mathbb{E}\big((X - \mathbb{E}(X))^2\big)$$

- Formule de Huygens : $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.
- $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.
- $\sigma_X = \sqrt{\mathrm{Var}(X)} =$ écart-type. Si $\sigma_X > 0$, alors la variable aléatoire $(X - \mathbb{E}(X))/\sigma_X$ est d'espérance 0 et d'écart-type 1 : elle est centrée et réduite.

Résultat :

$$\mathbb{E}(X) = \underset{a \in \mathbb{R}}{\mathrm{argmin}}\ \mathbb{E}\big[(X - a)^2\big]$$

$\mathbb{E}(X)$ est la meilleure approximation de X par une constante déterministe, au sens des moindres carrés.

Propriété de transport

Proposition

Soient $X : \Omega \to \mathbb{R}$ *et* $g : \mathbb{R} \to \mathbb{R}$ *(mesurable) avec* $g \geq 0$ *ou* $g(X)$ *intégrable. Alors*

$$\mathbb{E}(g(X)) = \int_\Omega g(X(\omega))\mathbb{P}(d\omega) = \int_\mathbb{R} g(x) P_X(dx)$$

Le calcul d'une espérance se ramène au calcul d'une intégrale réelle.

Calculs d'espérance : cas avec densité

Proposition

Soient X de densité f, et g telle que $\int_{-\infty}^{+\infty} |g(x)|f(x)dx < +\infty$, alors

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x)f(x)dx$$

En particulier,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx$$

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - \Big(\int_{-\infty}^{+\infty} xf(x)dx\Big)^2$$

Argument "heuristique": Supposons f et g continues. Soit

$$X_n = \frac{1}{n}[nX] = \frac{i}{n} \ \text{ si } \ \frac{i}{n} \leq X < \frac{i+1}{n}, \quad i \in \mathbb{Z}$$

Alors

$$\mathbb{E}(g(X_n)) = \sum_{i \in \mathbb{Z}} g\left(\frac{i}{n}\right) \mathbb{P}\left(\frac{i}{n} \leq X < \frac{i+1}{n}\right) \simeq \sum_{i \in \mathbb{Z}} g\left(\frac{i}{n}\right) f\left(\frac{i}{n}\right) \frac{1}{n}$$

car nous avons vu que pour n grand,

$$f\left(\frac{i}{n}\right) \simeq \frac{\mathbb{P}\left(\frac{i}{n} \leq X < \frac{i+1}{n}\right)}{\frac{1}{n}}$$

Nous utilisons la convergence des sommes de Riemann pour conclure.

Espérances et variances des lois usuelles

Montrer que

- Si $X \sim \mathcal{U}(a,b)$, alors

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

- Si $X \sim \mathcal{E}(\lambda)$, alors

$$\mathbb{E}(X) = \frac{1}{\lambda} \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

- Si $X \sim \mathcal{N}(0,1)$, alors

$$\mathbb{E}(X) = 0 \qquad \mathrm{Var}(X) = 1$$

- Si $X \sim \mathcal{N}(m,\sigma^2)$, alors

$$\mathbb{E}(X) = m \qquad \mathrm{Var}(X) = \sigma^2$$