

Probabilités et Statistiques

Dr Mory Ouattara

USP
SDM

17 octobre 2022

Tribus

Pour un ensemble X , on note $\mathcal{P}(X) = \{A : A \subset X\}$ l'ensemble de ses parties.

Définition

$\mathcal{A} \subset \mathcal{P}(X)$ est une tribu (ou une σ -algèbre) si :

- $X \in \mathcal{A}$
- Si $\mathcal{A}$ est stable par réunion dénombrable : pour tout $i \in \mathbb{N}$, $A_i \in \mathcal{A}$ alors $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$.
- Si $\mathcal{A}$ est stable par complémentaire : Si $A \in \mathcal{A}$ alors $A^c \in \mathcal{A}$

Tribu borélienne

Définition

Lorsque X est un espace topologique (c'est à dire muni d'une famille d'ouverts), la plus petite tribu contenant tous les ouverts est appelée la tribu borélienne. Elle est notée $\mathcal{B}(X)$.

Définition

Une application $f : (X; \mathcal{A}) \longrightarrow (Y; \mathcal{B})$ est dite mesurable si $\forall B \in \mathcal{B}, f^{-1}(B) \in \mathcal{A}$.

Mesures

On considère $(X; \mathcal{A})$ un espace mesurable.

Définition

Une mesure μ sur $(X; \mathcal{A})$ est une application de $\mathcal{A} \rightarrow [0; +\infty]$ telle que :

- $\mu(\emptyset) = 0$
- si $(A_n)_{n \geq 1}$ est une suite dénombrable d'ensembles de $\mathcal{A}$ deux à deux disjoints alors

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{+\infty} \mu(A_n)$$

Le triplet $(X; \mathcal{A}; \mu)$ est appelé un espace mesuré (espace mesurable + mesure).

Exemples de Mesures

- Mesure de Dirac sur $(X; \mathcal{P}(X))$: soit $a \in X$,

$$\begin{cases} 1 & \text{si } a \in X \\ 0 & \text{sinon} \end{cases}$$

- Mesure de Lebesgue Mesure de Lebesgue sur $(\mathbb{R}; B(\mathbb{R}))$: c'est la mesure qui généralise la notion de longueur des intervalles. Elle est invariante par translation :

$$\lambda([a, b]) = b - a, \quad \lambda(A + x) = \lambda(A)$$

- Une mesure μ est dite de probabilité sur $(X, \mathcal{A})$ si $\mu(X) = 1$. Traditionnellement, pour un espace de probabilité, on note $(\Omega, \mathcal{F}, \mathbb{P})$ au lieu de $(X, \mathcal{A}, \mu)$.

Espaces de probabilité

Un espace de probabilité $(\Omega, \mathbb{P})$ est constitué de

Définition

- Ω , un ensemble
- $\mathbb{P}$, une probabilité sur Ω .

Ω correspond à l'ensemble des résultats d'une expérience aléatoire. Un élément $\omega \in \Omega$ est appelé une réalisation, c'est un résultat possible d'une expérience aléatoire.

Un sous-ensemble $A \subset \Omega$ est appelé un événement. C'est un ensemble de réalisations (celles qui vérifient une certaine condition). Les opérations usuelles sur des événements A et B ont un sens logique :

Les opérations usuelles sur des événements A et B ont un sens logique

Notation	Sens mathématique	Interprétation en probabilités
$A^c (= \Omega - A)$	complémentaire de A	contraire de A , « non A »
$A \cup B$	réunion de A et B	« A ou B »
$A \cap B$	intersection de A et B	« A et B »
$A \cap B = \emptyset$	A et B sont disjoints	« A et B sont incompatibles »
$A \subset B$	A est inclus dans B	« A implique B »

Espaces de probabilités ; exemples

Ω correspond aux résultats de l'expérience :

- tirage à pile-ou-face, $\Omega = \{ , \}$
- lancer d'un dé, $\Omega = \{ , \}$
- lancer de deux pièces, $\Omega = \{ , \}$
- choix de deux parts dans une gâteau coupée en 8 : $\Omega = \{ , \}$
- attente d'un bus qui passe toutes les T minutes, $\Omega = \{ , \}$

Espaces de probabilités ; exemples

Ω correspond aux résultats de l'expérience :

- tirage à pile-ou-face, $\Omega = \{P, F\}$ ou $\{0, 1\}$
- lancer d'un dé, $\Omega = \{1, 2, 3, 4, 5, 6\}$
- lancer de deux pièces, $\Omega = \{P, F\}^2$
- choix de deux parts dans un gâteau coupé en 8 :
 $\Omega = \{(i, j), i, j \in \{1, \dots, 8\} \text{ et } i \neq j\}$
- attente d'un bus qui passe toutes les T minutes, $\Omega = [0, T]$

Espaces de probabilités ; exemples

Définition

Une probabilité sur Ω est une application $P : P(\Omega) \longrightarrow [0, 1]$, définie sur les événements, telle que :

- $\mathbb{P}(\Omega) = 1$
- pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'événements disjoints deux à deux,

$$\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)$$

Si un événement A vérifie $P(A) = 0$, on dit que A est négligeable ; et si $P(A) = 1$, on dit que A est presque sûr, ou que A a lieu presque sûrement, abrégé « p.s. ».

Espaces de probabilités

Propriétés

- ❶ $\mathbb{P}(\emptyset) = 0$
- ❷ Pour tout événement A , $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- ❸ Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$
- ❹ Pour tous événements A et B ,
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Espaces de probabilités

- Preuve de (2) : A et A^c sont disjoints ($A \cap A^c = \Omega$), et $\Omega = A \cup A^c$ donc
$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c),$$
d'où $P(A^c) = 1 - P(A)$. Et on obtient (1) en prenant $A = \Omega$.
- Preuve de (3) A et $B - A$ sont disjoints, et $A \cup (B - A) = B$ donc $P(B) = P(A \cup (B - A)) = P(A) + P(B - A) \geq P(A)$ et $P(B - A) = P(B) - P(A)$.
- Preuve de (4) : $A - (A \cap B)$, $A \cap B$ et $B - (A \cap B)$ sont disjoints, d'union $A \cup B$, donc
$$\begin{aligned} P(A \cup B) &= P(A - (A \cap B)) + P(A \cap B) + P(B - (A \cap B)) \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \end{aligned}$$

Distribution uniforme de probabilité

On suppose que Ω est fini, avec $\text{Card } \Omega = n$ $\Omega = \omega_1, \omega_2, \dots, \omega_n$. Si ces résultats jouent des rôles symétriques, il est naturel de considérer la probabilité uniforme sur Ω , telle que $P(\omega_1) = \dots = P(\omega_n) = \frac{1}{n}$

Définition

La probabilité uniforme sur Ω (ou distribution équiprobable) est la probabilité P définie par : pour tout $A = \omega_{i1}, \omega_{i2}, \dots, \omega_{ik} \subset \Omega$,

$$P(A) = \frac{k}{n} = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}}$$

Rappels de dénombrement

Soit E un ensemble fini.

Une permutation de E est une façon d'ordonner les éléments de E .

$$n! = 1 \times 2 \times 3 \times \dots \times (n-2) \times (n-1) \times n.$$

Un arrangement de k éléments de E est une suite de k éléments de E distincts 2 à 2. L'ordre est important.

$$A_n^k = n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!}$$

Une combinaison de k éléments de E est une façon de choisir k éléments de E , sans spécifier d'ordre : c'est un sous-ensemble de E à k éléments.

$$C_n^k = \frac{n(n-1) \dots (n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

Exemple : paradoxe des anniversaires

Dans un groupe de n étudiants, quelle est la probabilité que 2 (au moins) aient leur anniversaire le même jour ?

Exemple : paradoxe des anniversaires

Dans un groupe de n étudiants, quelle est la probabilité que 2 (au moins) aient leur anniversaire le même jour ?

On note $N = 365$, on suppose les dates équiprobables (et qu'il n'y a pas de jumeaux, ni d'années bissextiles).

On considère ainsi :

$$\Omega = \{(j_1, \dots, j_n) \mid j_1, \dots, j_n \in 1, \dots, N\} = \{1, \dots, N\}^n$$

et on cherche $P(A)$ où $A = \{ \text{2 étudiants sont nés le même jour} \} = \{(j_1, \dots, j_n) \in \Omega \mid \exists k \neq l, j_k = j_l\}$.

Alors

$$A^c = \{ \text{les étudiants sont nés des jours } \neq \} = \{(j_1, \dots, j_n) \in \Omega \mid \forall k \neq l, j_k \neq j_l\}.$$

et

$$P(A) = 1 - P(A^c) = 1 - \frac{\text{Card}(A^c)}{\text{Card}(\Omega)} = 1 - \frac{A_N^n}{N^n}$$

Exemple : Pour $n = 23$, $P(A) \sim 0.5$. Pour $n = 57$, $P(A) \sim 0.99$.

Quelle probabilité pour le bus

Pour l'attente du bus qui passe toutes les T minutes, $\Omega = [0, T]$

Le bus a autant de chances d'arriver dans $[t, t + \sigma]$ que dans $[t', t' + \sigma]$.

le bus a 2 fois plus de chances d'arriver dans $[t, t + 2\sigma]$ que dans $[t, t + \sigma]$.

la probabilité que le temps d'attente soit dans un intervalle I est proportionnel à sa longueur : (loi uniforme sur $[0, T]$)

$$P(I) = \frac{\text{Longueur}(I)}{T}$$

Probabilités conditionnelles

Définition

Soit B un événement tel que $P(B) > 0$. Pour $A \subset \Omega$, on définit

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

. $P(A | B)$ est appelée la probabilité conditionnelle de A sachant B . C'est la proportion de chance que A se réalise parmi les éventualités où B se réalise.

Définition

Deux événements A et B sont indépendants si $P(A \cap B) = P(A)P(B)$.

Si $P(B) \neq 0$, cela revient à $P(A | B) = P(A)$ Savoir que B est réalisé

influence la probabilité de A

Probabilités conditionnelles : exemple

On divise une galette selon le nombre d'invités, et chacun prend une part. Or le nombre d'invité n'est pas encore connu : Nous serons 5, 6 ou 7 avec probabilités 50 %, 30 % et 20 %.

Quelle est la probabilité que j'aie la fève ?

J'ai eu la fève. Quelle est la probabilité que nous étions 5 ?

Probabilités conditionnelles : exemple

On divise une galette selon le nombre d'invités, et chacun prend une part. Or le nombre d'invité n'est pas encore connu : Nous serons 5, 6 ou 7 avec probabilités 50 %, 30 % et 20 %.

Quelle est la probabilité que j'aie la fève ?

On note F = j'ai la fève et A_5 = nous sommes 5, A_6 et A_7 de même.

Alors : $P(A_5) = 0,5$ $P(A_6) = 0,3$ $P(A_7) = 0,2$

$P(F | A_5) = \frac{1}{5}$; $P(F | A_6) = \frac{1}{6}$; $P(F | A_7) = \frac{1}{7}$ d'où

$$P(F) = P(F \cap A_5) + P(F \cap A_6) + P(F \cap A_7) = P(F | A_5)P(A_5) + P(F | A_6)P(A_6) + P(F | A_7)P(A_7)$$

J'ai eu la fève. Quelle est la probabilité que nous étions 5 ?

$$P(A_5 | F) = \frac{P(A_5 \cap F)}{P(F)} = \frac{P(F|A_5)P(A_5)}{P(F)} = 0,56.$$

Probabilités conditionnelles

On suppose que $(A_n)_n$ est une partition de Ω (= un “découpage” de Ω) : *pour tous* $i \neq j$, $A_i \cap A_j = \emptyset$, et $\Omega = \bigcup_n A_n$.

Théorème (Théorème des probabilités totales)

$$P(A) = \sum_n P(A \cap A_n) = \sum_n P(A \mid A_n)P(A_n)$$

.

Théorème (Théorème de bayes)

$$P(A_i \mid A) = \frac{P(A \cap A_i)}{P(A)} = \frac{P(A \mid A_i)P(A_i)}{P(A)} = \frac{P(A \mid A_i)P(A_i)}{\sum_n P(A \mid A_n)P(A_n)}$$

.

Événements indépendants : cas général

Deux événements A et B sont indépendants si

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

.

Définition

Une famille $(A_i)_i$ d'événements est indépendante si pour toute sous-famille finie $A_{i_1}, \dots, A_{i_k}$ on a

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k})$$

Indépendance et complémentaire

Proposition

Si deux événements A et B sont indépendants, alors A^c et B^c le sont aussi, de même que A et B^c .

$$\begin{aligned} P(A^c \cap B^c) &= P((A \cup B)^c) = 1 - P(A \cup B) = \\ 1 - P(A) - P(B) + P(A \cap B) &= 1 - P(A) - P(B) + P(A)P(B) = \\ (1 - P(A))(1 - P(B)) &= P(A^c)P(B^c) \end{aligned}$$

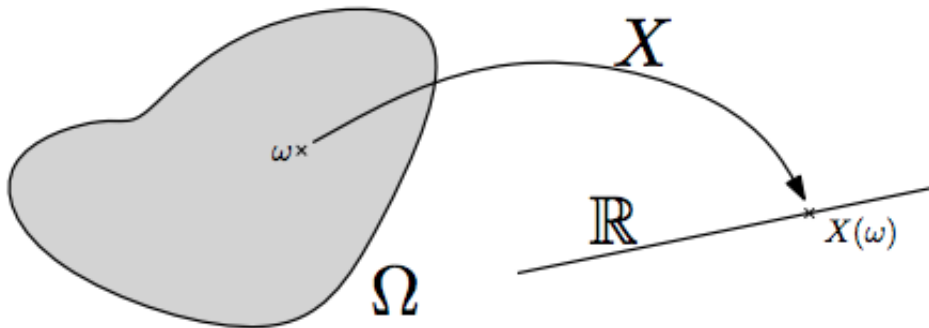
Par récurrence, on peut obtenir :

Proposition

Si $A_1, \dots, A_n$ sont indépendants, et $B_1, \dots, B_n$ sont tels que, pour tout i , $B_i = A_i$ ou $B_i = A_i^c$ alors $B_1, \dots, B_n$ sont indépendants.

Définition

Une variable aléatoire est une application $X : \Omega \longrightarrow \mathbb{R}$.



Loi

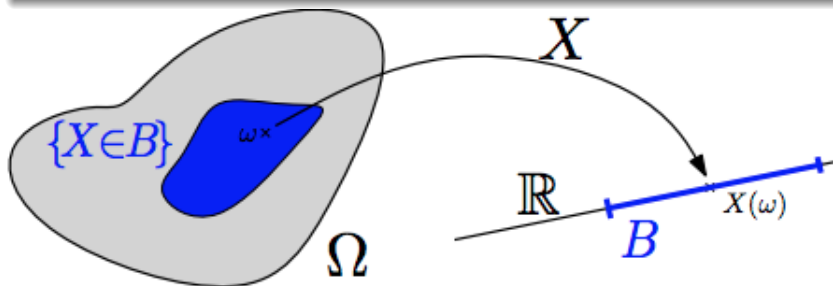
Définition

Une variable aléatoire est une application $X : \Omega \longrightarrow \mathbb{R}$.

La loi de X est la probabilité P_X sur $\mathbb{R}$ définie par : pour tout $B \subset \mathbb{R}$,

$$P_X(B) = P(\omega \in \Omega \mid X(\omega) \in B) = P(X \in B)$$

. $X(\Omega)$ (image de X) est le support de P_X .



Variables aléatoires - Remarques

- On précise parfois variable aléatoire réelle, ou à valeurs dans $\mathbb{R}$.
- S'il existe un réel c tel que $P(X = c) = 1$ alors X est constante.
- Notation : On a noté $X \in B$ l'événement formé des éventualités ω pour lesquelles $X(\omega) \in B$, et on abrège :

$$P(X \in B) = P(X \in B) = P(\omega \in \Omega \mid X(\omega) \in B).$$

Définition

Si A est un événement, on introduit la variable aléatoire fonction indicatrice de A , notée 1_A , qui indique si l'événement A est réalisé : pour tout $\omega \in \Omega$

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

Variables aléatoires - Exemples

- Lancer de deux dés, $\Omega = \{1, \dots, 6\}^2 = \{ \}$
Valeurs des dés : $X_1((x_1, x_2)) = x_1$ et $X_2((x_1, x_2)) = x_2$ (à valeurs dans $\{1, \dots, 6\}$)
Somme des résultats : $X =$ (à valeurs dans $2, \dots, 12$)

Variables aléatoires - Exemples

- Lancer de deux dés,

$$\Omega = \{1, \dots, 6\}^2 = \{(x_1, x_2) \mid x_1, x_2 \in \{1, \dots, 6\}\}$$

Valeurs des dés : $X_1((x_1, x_2)) = x_1$ et $X_2((x_1, x_2)) = x_2$ (à valeurs dans $\{1, \dots, 6\}$)

Somme des résultats : $X = X_1 + X_2$ c.-à-d. $X((x_1, x_2)) = x_1 + x_2$ (à valeurs dans $2, \dots, 12$)

Lois Discrètes

Définition

Une variable aléatoire X est dite discrète si l'ensemble $X(\Omega)$ des valeurs qu'elle prend est dénombrable.

Si X est discrète, alors pour tout $B \subset X(\Omega)$, on a $B = \{b_n \mid n = 1, 2, \dots, N\}$ avec des b_n distincts, et

$$X \in B = \cup_n \{X = b_n\}$$

or ces événements sont disjoints et forment une suite, d'où

$$P_X(B) = P(X \in B) = \sum_n P(X = b_n) = \sum_{x \in B} P(X = x)$$

Lois discrètes - Exemples

Uniforme

Si $E \subset \mathbb{R}$ est fini, une variable aléatoire X suit la loi uniforme sur E si pour tout $x \in E$, $P(X = x) = \frac{1}{\text{card}E}$

la loi du résultat d'un dé est la loi uniforme sur $\{1, \dots, 6\}$.

Bernoulli

Soit $p \in [0, 1]$. Une variable aléatoire X suit la loi de Bernoulli de paramètre p (notée $\mathcal{B}(p)$) si X est à valeurs dans $\{0, 1\}$ et $P(X = 1) = p$, $P(X = 0) = 1 - p$.

la loi de 1_A est $\mathcal{B}(P(A))$.

Lois discrètes - Exemples

Binomiale

Soit $n \in \mathbb{N}$ et $p \in [0, 1]$. Une variable aléatoire X suit la loi binomiale de paramètres n et p (notée $\mathcal{B}(n, p)$) si X est à valeurs dans $\{0, 1, \dots, n\}$ et pour $k = 0, \dots, n$

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

Lois discrètes - Exemples

- Expérience : Lancer de deux dés,
 $\Omega = \{1, \dots, 6\}^2 = \{(x_1, x_2) \mid x_1, x_2 \in \{1, \dots, 6\}\}$ associée à la
V.A $X((x_1, x_2)) = x_1 + x_2$. Quelle est sa loi ?

Lois discrètes - Exemples

- Expérience : Lancer de deux dés,
 $\Omega = \{1, \dots, 6\}^2 = \{(x_1, x_2) \mid x_1, x_2 \in \{1, \dots, 6\}\}$ associée à la
V.A $X((x_1, x_2)) = x_1 + x_2$. Quelle est sa loi ?
 $P(X = 2) = P((1, 1)) = \frac{1}{36}$
 $P(X = 3) = P((1, 2), (2, 1)) = \frac{2}{36}$
 $P(X = 6) = P((1, 5), (2, 4), \dots, (5, 1)) = \frac{5}{36}$
 $P(X = 7) = P((1, 6), (2, 5), \dots, (6, 1)) = \frac{6}{36}$
 $P(X = 8) = P((2, 6), (3, 5), \dots, (6, 2)) = \frac{5}{36}$
...
 $P(X = 12) = P((6, 6)) = \frac{1}{36}$

Lois discrètes - Exemples

- On prend successivement les parts d'une galette (coupée en 8)
 N_A : Nombre de parts à prendre jusqu'à avoir la fève : à valeurs dans $\{1, \dots, 8\}$
- Chaque jour, on prend une part d'une galette différente (coupée en 8)
 N_B : Nombre de parts à prendre jusqu'à avoir la fève : N_B , à valeurs dans $\{1, 2, \dots\} = \mathbb{N}^*$
- Nombre de fèves obtenues en n jours : S_n

Lois discrètes - Exemples

- On prend successivement les parts d'une galette (coupée en 8)
 N_A : Nombre de parts à prendre jusqu'à avoir la fève : à valeurs dans $\{1, \dots, 8\}$
 N_A suit la loi uniforme sur $\{1, \dots, 8\}$
- Chaque jour, on prend une part d'une galette différente (coupée en 8)
 N_B : Nombre de parts à prendre jusqu'à avoir la fève : N_B , à valeurs dans $\{1, 2, \dots\} = \mathbb{N}^*$
Pour tout $n \in \mathbb{N}^*$,
 $P(N_B = n) = (1 - \frac{1}{8})^{n-1} \frac{1}{8}$ (Indépendance).
- Nombre de fèves obtenues en n jours : S_n S_n suit la loi binomiale de paramètres $(n, \frac{1}{8})$.

Lois continues

Définition

Une variable aléatoire X est dite continue ou à densité s'il existe une fonction $f_X : \mathbb{R} \longrightarrow \mathbb{R}$ telle que, pour tout $B \subset \mathbb{R}$

$$P_X(B) = P(X \in B) = \int_B f_X(x) dx$$

La fonction f_X est appelée la densité de X . Une fonction f est la densité d'une variable aléatoire si, et seulement si

- pour tout $x \in \mathbb{R}$, $f(x) \geq 0$
- $\int_{\mathbb{R}} f_X(x) dx = 1$

Si X a pour densité f_X , pour tous $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx$$

Densités classiques

Uniforme

Soit $a < b$. La loi uniforme sur $[a; b]$ est la loi de densité

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

Une variable aléatoire X de loi $U([a; b])$ est donc à valeurs dans $[a; b]$.

Exponentielle

Soit $\lambda > 0$. La loi exponentielle de paramètre λ est la loi de densité

$$f(x) = \lambda e^{-\lambda x} 1_{\mathbb{R}_+}(x)$$

Une variable X de loi $\varepsilon(\lambda)$ est donc à valeurs dans $\mathbb{R}_+$.

Densités classiques

Exponentielle

Soit $\lambda > 0$. La loi exponentielle de paramètre λ est la loi de densité

$$f(x) = \lambda e^{-\lambda x} 1_{\mathbb{R}_+}(x)$$

Une variable X de loi $\varepsilon(\lambda)$ est donc à valeurs dans $\mathbb{R}_+$.

La loi exponentielle est une loi « sans mémoire ». En effet, pour tous $s, t \geq 0$,

$$P(X \geq s + t \mid X > s) = \frac{P(\{X \geq s + t\}) \cap \{X \geq s\}}{P(\{X \geq s\})}$$

$$P(X \geq s + t \mid X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda(s)}} = P(\{X \geq t\})$$

Utilisée pour modéliser les durées de vie de machine sans vieillissement.

Exemple : soit X une variable aléatoire de loi $U([0; 1])$. On définit
 $Y = \min(X; \frac{1}{2})$
On a $Y(\omega) = \min(X(\omega); \frac{1}{2})$ pour tout $\omega \in \Omega$

De nombreuses variables ne sont ni discrètes, ni à densité. Exemple : soit X une variable aléatoire de loi $U([0; 1])$. On définit

$$Y = \min(X; \frac{1}{2})$$

On a $Y(\omega) = \min(X(\omega); \frac{1}{2})$ pour tout $\omega \in \Omega$

- Y est à valeurs dans $[0; \frac{1}{2}]$ car X est à valeurs dans $[0; 1]$.
- On a
$$P(Y = \frac{1}{2}) = P(X \geq \frac{1}{2}) = \int_{\frac{1}{2}}^{\infty} 1_{[0;1]}(x)dx = \int_{\frac{1}{2}}^1 dx = \frac{1}{2} > 0.$$
n'a pas de densité
- Pour tout $0 \leq x < \frac{1}{2}$ $P(Y = x) = P(X = x) = 0$
 Y n'est pas discrète

Fonction de répartition

Définition

Soit X une variable aléatoire. La fonction de répartition de X est la fonction $F(X) : \mathbb{R} \longrightarrow \mathbb{R}$ définie par : $\forall a \in \mathbb{R}$;

$$F_X(a) = P(X \leq a)$$

Proposition

- La fonction de répartition F_X est une fonction croissante,
- $\lim_{a \rightarrow -\infty} F_X(a) = 0$ et $\lim_{a \rightarrow +\infty} F_X(a) = 1$
- Si X et Y sont deux variables aléatoires telles que $F_X(t) = F_Y(t)$ pour tout $t \in \mathbb{R}$, alors X et Y ont même loi.

$\implies$ la fonction de répartition caractérise la loi d'une variable aléatoire

Fonction de répartition-cas discret

Proposition

Si X est une variable aléatoire discrète, F_X est une fonction constante par morceaux, dont les sauts se situent aux points de $X(\Omega)$, et le saut en $x \in X(\Omega)$ a pour hauteur $P(X = x)$.

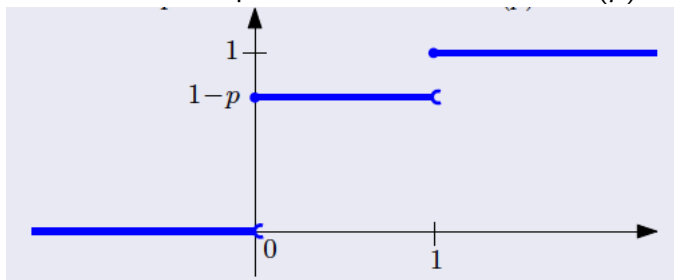
Fonction de répartition d'une v.a. de loi $\mathcal{B}(p)$

Fonction de répartition-cas discret

Proposition

Si X est une variable aléatoire discrète, F_X est une fonction constante par morceaux, dont les sauts se situent aux points de $X(\Omega)$, et le saut en $x \in X(\Omega)$ a pour hauteur $P(X = x)$.

Fonction de répartition d'une v.a. de loi $\mathcal{B}(p)$



Fonction de répartition-cas discret

Proposition

Si X est une variable aléatoire discrète, F_X est une fonction constante par morceaux, dont les sauts se situent aux points de $X(\Omega)$, et le saut en $x \in X(\Omega)$ a pour hauteur $P(X = x)$.

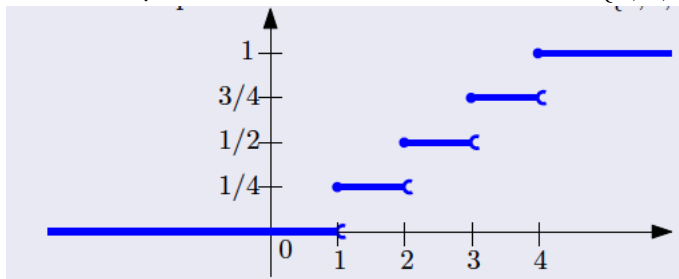
Fonction de répartition d'une v.a. de loi unif. sur $\{1, 2, 3, 4\}$

Fonction de répartition-cas discret

Proposition

Si X est une variable aléatoire discrète, F_X est une fonction constante par morceaux, dont les sauts se situent aux points de $X(\Omega)$, et le saut en $x \in X(\Omega)$ a pour hauteur $P(X = x)$.

Fonction de répartition d'une v.a. de loi unif. sur $\{1, 2, 3, 4\}$

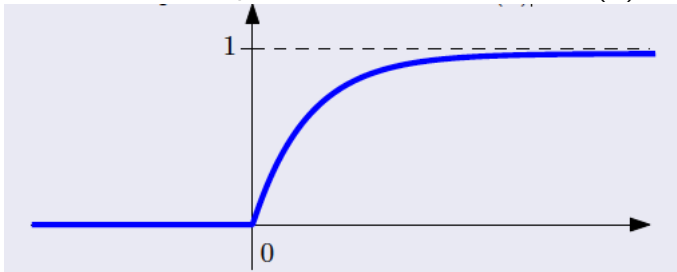


Fonction de répartition-cas à densité

Proposition

Si X est une variable aléatoire de densité f_X , on a pour tout $x \in \mathbb{R}$;
 $F_X(x) = \int_{-\infty}^x f_X(t)dt$ et on a la dérivée $(F_X)'(x) = f_X(x)$ (pour tout x où f_X est continue).

Fonction de répartition d'une v.a. de loi $\mathcal{E}(\lambda)$



Fonction de répartition-cas à densité

Proposition

Si X est une variable aléatoire de densité f_X , on a pour tout $x \in \mathbb{R}$;
 $F_X(x) = \int_{-\infty}^x f_X(t)dt$ et on a la dérivée $(F_X)'(x) = f_X(x)$ (pour tout x où f_X est continue).

Fonction de répartition d'une v.a. de loi $\mathcal{E}(\lambda)$

Fonction de répartition-cas à densité

Proposition

Si X est une variable aléatoire de densité f_X , on a pour tout $x \in \mathbb{R}$;
 $F_X(x) = \int_{-\infty}^x f_X(t)dt$ et on a la dérivée $(F_X)'(x) = f_X(x)$ (pour tout x où f_X est continue).

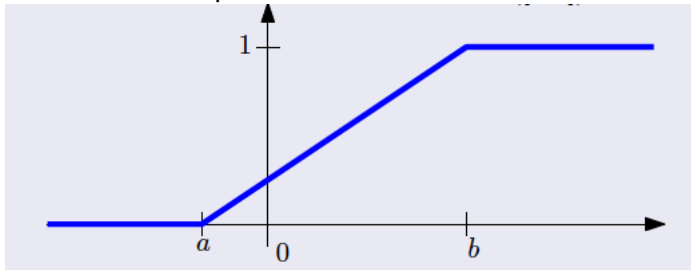
Fonction de répartition d'une v.a. de loi Uniforme

Fonction de répartition-cas à densité

Proposition

Si X est une variable aléatoire de densité f_X , on a pour tout $x \in \mathbb{R}$;
 $F_X(x) = \int_{-\infty}^x f_X(t)dt$ et on a la dérivée $(F_X)'(x) = f_X(x)$ (pour tout x où f_X est continue).

Fonction de répartition d'une v.a. de loi Uniforme



Fonction de répartition-cas à densité

Proposition

Si X est une variable aléatoire de densité f_X , on a pour tout $x \in \mathbb{R}$; $F_X(x) = \int_{-\infty}^x f_X(t)dt$ et on a la dérivée $(F_X)'(x) = f_X(x)$ (pour tout x où f_X est continue).

Inversement, si X est une v.a. telle que F_X est

- continue sur $\mathbb{R}$
- dérivable sauf peut-être en un nombre fini de points, alors X a pour densité $f_X = F'_X$.

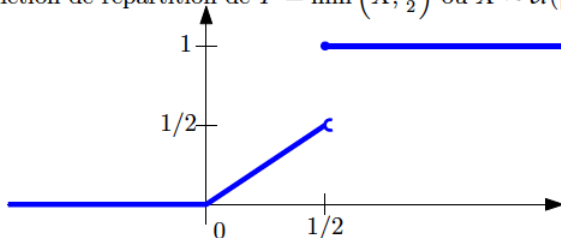
Fonction de répartition-cas à densité

Suite du premier (contre-)exemple : soit X une v.a. de loi $\mathcal{U}([0,1])$. On définit

$$Y = \min\left(X, \frac{1}{2}\right).$$

- Y est à valeurs dans $[0, \frac{1}{2}]$, d'où $F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y > \frac{1}{2} \end{cases}$
- pour $0 \leq y \leq \frac{1}{2}$, $F_Y(y) = P(Y \leq y) = P(X \leq y) = y$
(car si $Y \leq \frac{1}{2}$ alors $Y = X$)

Fonction de répartition de $Y = \min\left(X, \frac{1}{2}\right)$ où $X \sim \mathcal{U}([0,1])$



Calcul de la loi de $Y = \phi(X)$

Soit X une variable aléatoire, de loi connue, et $\phi : X(\Omega) \longrightarrow \mathbb{R}$ une fonction. On cherche la loi de la variable aléatoire $Y = \phi(X)$.

- Si Y est discrète (par exemple, si X est discrète),
 - Déterminer les valeurs possibles de Y ,
 - Calculer chacune de leurs probabilités en se ramenant à X .
- Si X a une densité f_X , et ϕ est monotone (croissante ou décroissante),
 - Déterminer les valeurs possibles de Y ,
 - Calculer la **fonction de répartition** de Y ,
 - Si F_Y est continue sur $\mathbb{R}$, et dérivable (sauf en quelques points), dériver pour obtenir f_Y .

La méthode s'étend aux fonctions non monotones, mais il faut alors être plus vigilant, ou se ramener à des intervalles où ϕ est monotone.

Exemples de calculs de loi

Soit X une variable aléatoire de loi uniforme sur $\{-1, 0, 1\}$. On pose $Y = |X|$. Alors Y est à valeurs dans $\{0, 1\}$, et

$$P(Y = 1) = P(|X| = 1) = P(X = 1 \text{ ou } X = -1) = 2/3$$

et ainsi $P(Y = 0) = 1/3$ et $P(Y = 1) = 2/3$ donc Y suit la loi $B(2/3)$.

Exemples de calculs de loi

Soit X une variable aléatoire de loi $\mathcal{E}(\lambda)$. On pose

$$Y = \frac{1}{1+X}$$

On a

$$Y = \phi(X)$$

où : $\phi(x) \longrightarrow \frac{1}{1+x}$ Comme $X \sim \varepsilon(\lambda)$, on a $X > 0$ p.s.

ϕ est strictement décroissante sur $]0; +1[$, $\phi(0) = 1$ et $\lim_{x \rightarrow +\infty} \phi(x) = 0$
donc $\phi(]0; +1[) =]0; 1[$.

Exemples de calculs de loi

Ainsi, Y est à valeurs dans $]0; 1[$.

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y \geq 1 \end{cases}$$

et pour $0 \leq y \leq 1$,

$$\begin{aligned} P(Y \leq y) &= P\left(\frac{1}{1+X} \leq y\right) = P\left(X \geq \frac{1}{y} - 1\right) \\ &= 1 - P\left(X \leq \frac{1}{y} - 1\right) = 1 - F_X\left(\frac{1}{y} - 1\right) = e^{-\lambda(\frac{1}{y}-1)} \end{aligned}$$

F_Y est continue sur $\mathbb{R}$ (on vérifie $F_Y(0^+) = 0$ et $F_Y(1^-) = 1$, et dérivable sauf peut-être en 0 et 1. Donc Y a pour densité la dérivée

$$f_Y(y) = (F_Y)'(y) = \begin{cases} \frac{\lambda}{y^2} e^{-\lambda(\frac{1}{y}-1)} & \text{si } y \in]0, 1[\\ 0 & \text{si } y \geq 1 \end{cases}$$

Espérance

Dans un jeu de hasard A, on peut

- Gagner 100 fcfa, avec probabilité 0,1
- Perdre 1 fcfa, avec probabilité 0,9.

Dans un autre jeu de hasard B, on peut

- Gagner 100 fcfa, avec probabilité 0,2
- Perdre 1 fcfa, avec probabilité 0,8.

À quel jeu devrait-on jouer ?

Dans un jeu de hasard A, on peut

- Gagner 100 fcfa, avec probabilité 0,1
- Perdre 1 fcfa, avec probabilité 0,9.

Dans un autre jeu de hasard B, on peut

- Gagner 10 fcfa, avec probabilité 0,5
- Perdre 1 fcfa, avec probabilité 0,5.

À quel jeu devrait-on jouer ?

Espérance

Définition

L'espérance d'une variable aléatoire X , notée $E[X]$, est la moyenne de ses valeurs, pondérées par leurs probabilités. Si X est discrète,

$$E[X] = \sum_{x \in X(\Omega)} xP(X = x)$$

Si X est continue, de densité f_X ,

$$E[X] = \int_{\mathbb{R}} xf_X(x)dx$$

Attention. L'espérance n'est pas toujours définie. Il faut pour cela que la série ou l'intégrale ci-dessus converge absolument.

Espérance

Intérêt, interprétation :

- $E[X]$ donne une indication de l'ordre de grandeur typique de X .
- $E[X]$ est souvent plus simple à calculer (et à interpréter) que la loi de X .
- $E[X]$ correspond au “prix équitable” à faire payer pour jouer à un jeu de hasard où le gain est X (dans l'idée que l'on joue un grand nombre de fois). $\rightsquigarrow$ prix d'assurances, d'actifs financiers

Espérance - Exemples discrets

Si X suit la loi de Bernoulli $\mathbb{B}(p)$

X est à valeurs dans $\{0; 1\}$ et $P(X = 1) = p$, $P(X = 0) = 1 - p$, d'où

$$E[X] = 1 * p + 0 * (1 - p) = p$$

Si X suit la loi uniforme sur $\{1, 2, \dots, n\}$

X est à valeurs dans $\{1, 2, \dots, n\}$ et $P(X = 1) = \dots = P(X = n) = \frac{1}{n}$, d'où

$$E[X] = \frac{n+1}{2}$$

Espérance - Exemples discrets

Si X suit la loi géométrique $\mathcal{G}(p)$

X est à valeurs dans $N^* = \{1, 2, \dots\}$ et $P(X = k) = (1 - p)^{k-1}p$,
d'où

$$E[X] = \sum_{k=1}^{\infty} k(1 - p)^{k-1}p = \frac{1}{p}$$

car

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{d}{dx} \left(\sum_{k=1}^{\infty} x^k \right) = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^2}$$

$$\forall x \in]-1, 1[$$

Espérance – Exemples à densité

Si X suit la loi uniforme sur $[a; b]$, où $a < b$

Espérance – Exemples à densité

Si X suit la loi uniforme sur $[a; b]$, où $a < b$

X a pour densité $f(x) = \frac{1}{b-a}1_{[a;b]}(x)$, d'où

$$E[X] = \int_{\mathbb{R}} x \frac{1}{b-a} 1_{[a;b]}(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{b-a}{2}$$

Espérance – Exemples à densité

Si X suit la loi exponentielle $\mathcal{E}(\lambda)$, où $\lambda > 0$,

Espérance – Exemples à densité

Si X suit la loi exponentielle $\mathcal{E}(\lambda)$, où $\lambda > 0$,

X a pour densité $f(x) = \lambda e^{-\lambda x} 1_{]0;+\infty[}(x)$,
d'où

$$E[X] = \int_{\mathbb{R}} x \lambda e^{-\lambda x} 1_{]0;+\infty[}(x) dx = \int_0^{\infty} x \lambda e^{-\lambda x}(x) dx = \frac{1}{\lambda}$$

Espérance – Propriétés

Propriétés

- 1 Si X est constante, égale à $c \in \mathbb{R}$ (pour tout $\omega \in \Omega$), alors $E[X] = E[c] = c$.
- 2 Pour tout événement $A \subset \Omega$ $E[1_A] = P(A)$.
- 3 L'espérance est linéaire : pour toutes variables aléatoires X et Y , et tout réel a ,

$$E[aX] = aE[X] \text{ et } E[X + Y] = E[X] + E[Y]$$

- 4 L'espérance est croissante : si $X \leq Y$ p.s., alors $E[X] \leq E[Y]$.

Espérance – Propriétés

Si $A_1, \dots, A_n$ sont des événements indépendants et

$$P(A_1) = \dots = P(A_n) = p$$

on a vu que

$$S_n = 1_{A_1} + \dots + 1_{A_n}$$

suit la loi binomiale $\mathcal{B}(n; p)$: pour $k = 0, \dots, n$,

$$P(S_n = k) = C_n^k p^k (1 - p)^{n-k}$$

Par linéarité,

$$E[S_n] = E[1_{A_1}] + \dots + E[1_{A_n}] = P(A_1) + \dots + P(A_n) = np$$

Espérance de $\phi(X)$

Soit X une variable aléatoire, et $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction.
Si X est discrète, alors

$$E[\phi(X)] = \sum_{x \in X(\Omega)} \phi(x)P(X = x)$$

Si X est continue, alors

$$E[\phi(X)] = \int_{\mathbb{R}} \phi(x)f_X(x)dx$$

(À condition que la série et l'intégrale soient bien définies)

Espérance de $\phi(X)$

Si X suit la loi uniforme sur $\{1, 2, 3\}$,

$$E\left[\frac{X}{1+X}\right] = \frac{23}{36}$$

Si X suit la loi uniforme sur $[0; 1]$,

$$E\left[\frac{X}{1+X}\right] = \ln 2$$

Variance

- L'espérance $E[X]$ représente-t-elle bien les valeurs typiques de X ?
- Comment les valeurs de X sont-elles dispersées autour de $E[X]$?

Variance

Définition

Soit X une variable aléatoire. La variance de X est l'espérance des carrés des écarts de X à sa moyenne :

$$\text{Var}(X) = E(X - E[X])^2 \geq 0$$

L'écart type de X est $\sigma(X) = \sqrt{\text{Var}(X)}$

NB. À la différence de la variance, l'écart type $\sigma(X)$ est homogène à X : si par exemple X est une distance, alors $\sigma(X)$ est une distance aussi. Ceci justifie l'intérêt de l'écart type.

Variance

Définition

Pour toutes variables aléatoires X et Y et toute constante a ,

- $Var(X) = E[X^2] - E[X]^2$
- $Var(aX) = a^2 Var(X)$
- $Var(X + a) = Var(X)$
- $Var(X + Y) = Var(X) + 2Cov(X, Y) + Var(Y)$, où la covariance est définie par :

$$Cov(X; Y) = E((X - E[X])(Y - E[Y])) = E[XY] - E[X]E[Y]$$

Standardisation d'une variable

Pour toute variable aléatoire X possédant une variance, la variable aléatoire

$$Z = \frac{X - E(X)}{\sigma(X)}$$

est centrée ($E(Z) = 0$) et réduites ($\text{var}(Z) = 1$)

Moments

Plus généralement, pour $r > 0$, on définit (s'il existe) le moment d'ordre r :

$$\mu_r(X) = E[X^r]$$

et le moment centré d'ordre r :

$$\mu_r(X) = E((X - E[X])^r)$$

Variance exemples

Si X suit la loi de Bernoulli $B(p)$

$$E(X^2) = 1^2 * p + 0^2 * (1 - p) = p; \text{ donc } \text{Var}(X) = p - p^2 = p(1 - p)$$

Variance exemples

Si X suit la loi géométrique $G(p)$

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

Indication : dériver deux fois

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

pour obtenir

$$\sum_{k=2}^{\infty} k(k-1)x^k = \frac{2}{1-x}^3$$

et en déduire le calcul de $E[X(X-1)]$ puis
 $E[X^2] = E[X(X-1) + X] = \dots$

Variance exemples

Si X suit la loi uniforme sur $[a;b]$,

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) = \frac{a^2 + ab + b^2}{3}$$

d'où

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Inégalités

Proposition (Inégalité de Markov)

Soit X une variable aléatoire. Pour tout $a > 0$,

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

Plus généralement, pour tout $a > 0$ et $r > 0$,

$$P(|X| \geq a) \leq \frac{E(|X|^r)}{a^r}$$

Preuve

Inégalités

Preuve(Inégalité de Markov)

$$E(|X|^r) = \int_{\mathbb{R}} |X|^r f_X(x) dx \quad (1)$$

$$= \int_{]-a,a[} |X|^r f_X(x) dx + \int_{\mathbb{R}-]-a,a[} |X|^r f_X(x) dx \quad (2)$$

$$\geq 0 + \int_{\mathbb{R}-]-a,a[} a^r f_X(x) dx = a^r P(X \in \mathbb{R}-]-a,a[) \quad (3)$$

$$= a^r P(|X| > a) \quad (4)$$

Inégalités

Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire. Pour tout $a > 0$,

$$P(|X - E[X]| \geq a) < \frac{\text{Var}(X)}{a^2}$$

avec probabilité 75 %, $|X - E[X]| \leq 2\sigma(X)$.

Indépendance de variables aléatoires

Définition

Des variables aléatoires $\{X_1; \dots; X_n\}$ sont indépendantes si, pour tous

$$B_1, \dots, B_n \subset \mathbb{R}$$

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n)$$

Par exemple, deux variables aléatoires X et Y sont indépendantes si les événements qui ne dépendent que de X sont indépendants des événements qui ne dépendent que de Y : pour $B, C \subset \mathbb{R}$,

$$P(X \in B; Y \in C) = P(X \in B)P(Y \in C)$$

Indépendance-Retour sur un exemple

On considère deux tirages de dés : espace de probabilité $\Omega = \{1, \dots, \}^2$, avec la probabilité P uniforme. On note X_1, X_2 les résultats des dés : pour tout tirage $(k, l) \in \Omega$

$X_1(k, l) = k$ et $X_2(k, l) = l$

Alors, pour $A, B \subset \{1, \dots, \}$

$$P(X_1 \in A; X_2 \in B) = P(\{(k; l) \in \Omega \mid k \in A; l \in B\}) =$$

$$P(A \times B) = \frac{\text{card}(A \times B)}{\text{card}\Omega} = \frac{\text{card}A \times \text{Card}B}{\text{card}\Omega}$$

car X_1 et X_2 suivent la loi uniforme/

Indépendance- Propriétés

- Si $X_1, \dots, X_n$ sont indépendantes, alors les variables aléatoires $f_1(X_1), \dots, f_n(X_n)$ sont indépendantes, quelles que soient les fonctions $f_1, \dots, f_n$.
- Si des événements $A_1, \dots, A_n$ sont indépendants alors leurs fonctions indicatrices $1_{A_1}, \dots, 1_{A_n}$ sont des variables aléatoires indépendantes; et réciproquement.

Indépendance et espérance

Proposition

Si $X_1, \dots, X_n$ sont des variables aléatoires indépendantes, alors

- 1 si leurs espérances sont bien définies

$$E(X_1 \dots X_n) = E(X_1) \dots E(X_n)$$

- 2 si leurs variances sont bien définies, alors on a $\text{Cov}(X_i; X_j) = 0$

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

(le 1. est évident si $X_i = 1_{A_i}$, et le cas général s'en déduit par approximation)

Par le 1) on déduit, si $X_1, \dots, X_n$ sont indépendantes

$$E[f_1(X_1) \dots f_n(X_n)] = E[f_1(X_1)]E[f_n(X_n)]$$

Théorème Loi des grands nombres

Théorème

Soit $X_n, n \geq 1$ une suite de variables aléatoires indépendantes, et de même loi, d'espérance m et de variance σ^2 . On définit la variable aléatoires $\bar{X}_n$, appelée moyenne empirique, par :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

On a : Pour tout $\epsilon > 0$, $P(m - \epsilon \leq \bar{X}_n \leq m + \epsilon) \rightarrow 1$

Si $(A_n)n_1$ est une suite d'événements indépendants et qui ont même probabilité p (par exemple, dans une suite de tirages à Pile-ou-Face, $A_n = \{ \text{le } n\text{-ième tirage est Pile} \}$, et $p = \frac{1}{2}$), alors en posant

$$X_i = 1_{A_i}, \text{ on a } \bar{X}_n = \frac{1_{A_1} + \dots + 1_{A_n}}{n}$$

Loi normale $\mathcal{N}(m; \sigma^2)$

La loi normale centrée ($m = 0$) réduite ($\sigma = 1$), notée $\mathcal{N}(0; 1)$, est la loi de densité

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

Si $m \in \mathbb{R}$ et $\sigma \in]0; +\infty[$ la loi normale de moyenne m et de variance σ^2 , notée $\mathcal{N}(m, \sigma^2)$, est la loi de la variable aléatoire $X = m + \sigma Z$, où Z suit la loi $\mathcal{N}(0, 1)$.

Si X suit une loi normale, on dit que X est une v.a. gaussienne

Si $Z \sim \mathcal{N}(0; 1)$, sa fonction de répartition est

$$\phi(x) = P(Z \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$$

ϕ ne peut pas s'exprimer à l'aide des fonctions usuelles, donc on utilise une table (imprimée, ou dans un logiciel de calcul numérique).

Loi normale $\mathcal{N}(m; \sigma^2)$

Proposition

Toute combinaison linéaire de variables aléatoires gaussiennes indépendantes est une variable aléatoire gaussienne.

Plus précisément, si $X_1, \dots, X_n$ sont indépendantes et $X_i \sim \mathcal{N}(m_i; \sigma_i^2)$ alors, pour tous $a_1, \dots, a_n \in \mathbb{R}$

$$X = a_1 X_1 + \dots + a_n X_n \sim \mathcal{N}(M, \Sigma^2)$$

$$M = E(X) = \sum_{i=1}^n a_i m_i$$

et

$$\Sigma^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$$

Les lois normales interviendront en statistique (pour étudier la marge

Loi du couple

Proposition

Soit (X, Y) deux variables aléatoires. La loi du couple (X, Y) est la probabilité $P(X, Y)$ sur $\mathbb{R}^2$ qui vérifie : pour tous $A, B \subset \mathbb{R}$;

$$P(X, Y)(A \times B) = P(X \in A; Y \in B) :$$

Les lois de X et Y se déduisent de $P(X, Y)$: pour $A \subset \mathbb{R}$,

$$P_X(A) = P(X \in A) = P(X \in A; Y \in \mathbb{R}) = P(X, Y)(A \times \mathbb{R})$$

Inversement, les lois de X et de Y sont les lois marginales de $P(X, Y)$.

Si X et Y sont indépendantes, la loi du couple est fournie par les lois de X et de Y :

$$P(X, Y)(A \times B) = P_X(A)P_Y(B)$$

Loi du couple

Si X et Y sont indépendantes, la loi du couple est fournie par les lois de X et de Y :

$$P_{(X, Y)}(A \times B) = P_X(A)P_Y(B)$$

La loi du couple contient davantage d'information que P_X et P_Y : elle indique aussi la façon dont les variables dépendent l'une de l'autre (connaître X peut renseigner sur Y).

On choisit au hasard (uniformément) un étudiant entré à l'université en 2017. On note

- $S \in \{H; F\}$ son sexe
- $D \in \{bio - sant, droit, lettres, sciences, sport, sciencesco\}$ la discipline où il est inscrit. Ce sont deux variables aléatoires.

Décrire la loi de (S, D) revient à se donner les proportions d'étudiants dans chaque cas :

	bio-santé	droit	lettres	sciences	sport	sciences éco
H	6 %	7 %	15 %	6 %	4 %	5 %
F	14 %	10 %	24 %	3 %	2 %	4 %

les variables D et S sont elles indépendantes ?

On choisit au hasard (uniformément) un étudiant entré à l'université en 2017. On note

	bio-santé	droit	lettres	sciences	sport	sciences éco	Total
H	6 %	7 %	15 %	6 %	4 %	5 %	43 %
F	14 %	10 %	24 %	3 %	2 %	4 %	57 %
Total	20 %	17 %	39 %	9 %	6 %	9 %	100 %

Le total de droite est la loi de S . Le total du bas est la loi de D .

D et S ne sont pas indépendantes : $P(D = \text{bio}; S = H) = 0.06$ et $P(D = \text{bio})P(S = H) = 0.2 \cdot 0.43 = 0.086$

Cas de deux variables discrètes

Si X et Y sont discrètes alors la loi de (X, Y) est donnée par les probabilités élémentaires :

$p_{(X,Y)}(x, y) = P(X = x, Y = y)$ pour tous $x \in X(\Omega)$; $y \in Y(\Omega)$:
Elles vérifient $p_{(X,Y)}(x, y) \in [0; 1]$ pour tous x, y , et

$$\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{(X,Y)}(x, y) = 1$$

Inversement, les lois marginales se déduisent des $(p_{(X,Y)}(x, y))$: pour tout $x \in X(\Omega)$

$$p_X(x) = P(X = x) = \sum_{y \in Y(\Omega)} P(X = x, Y = y) = \sum_{y \in Y(\Omega)} p_{(X,Y)}(x, y)$$

X et Y sont indépendantes si, et seulement si

$p_{(X,Y)}(x, y) = p_X(x)p_Y(y)$ pour tous x, y .

Cas de deux variables à densité

On dit que le couple (X, Y) a une densité s'il y a une fonction $f_{(X,Y)} : \mathbb{R} \rightarrow \mathbb{P}(X = x, Y = y)$ telle que $D \subset \mathbb{R}$

$$P_{(X,Y)} = \int_{\mathbb{D}} \int_{\mathbb{D}} f_{(X,Y)}(x,y) dx dy$$

$f_{(X,Y)}$ est appelée la densité du couple (X, Y) . Alors $f_{(X,Y)}(x,y) \geq 0$ pour tous $x, y \in \mathbb{R}$, et

$$P_{(X,Y)} = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{(X,Y)}(x,y) dx dy = 1$$

Presque sûrement, $(X, Y) \in \text{Supp}(f_{(X,Y)})$ où le support de la fonction $f_{(X,Y)}$ est défini par

$$\text{Supp}(f_{(X,Y)}) = \{(x,y) \in \mathbb{R}^2 \mid f_{(X,Y)}(x,y) > 0\}$$

Cas de deux variables à densité

On dit que le couple (X, Y) a une densité s'il y a une fonction $f_{(X,Y)} : \mathbb{R} \rightarrow \mathbb{P}(X = x, Y = y)$ telle que $D \subset \mathbb{R}$

$$P_{(X,Y)} = \int_{\mathbb{D}} \int_{\mathbb{D}} f_{(X,Y)}(x, y) dx dy$$

$f_{(X,Y)}$ est appelée la densité du couple (X, Y) . Alors $f_{(X,Y)}(x, y) \geq 0$ pour tous $x, y \in \mathbb{R}$, et

$$P_{(X,Y)} = \int_{\mathbb{R}} \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx dy = 1$$

Presque sûrement, $(X, Y) \in \text{Supp}(f_{(X,Y)})$ où le support de la fonction $f_{(X,Y)}$ est défini par

$$\text{Supp}(f_{(X,Y)}) = \{(x, y) \in \mathbb{R}^2 \mid f_{(X,Y)}(x, y) > 0\}$$

Cas où $P(X, Y)$ a une densité

On déduit les lois marginales de la loi du couple et, dans le cas indépendant, on déduit la loi du couple des lois marginales :

Proposition

- ❶ Si (X, Y) a pour densité $f(X, Y)$, alors X et Y ont des densités f_X et f_Y données par

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy \text{ et } f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dx$$

- ❷ Si X et Y ont des densités f_X et f_Y et sont indépendantes, alors (X, Y) a pour densité

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$$

Cas où $P(X, Y)$ a une densité

On lance 2 dés à 4 faces, dont on note X et Y les résultats, entre 1 et 4.

$\rightsquigarrow X$ et Y sont indépendantes, de loi uniforme sur $\{1, \dots, 4\}$.

On définit $Z = |X - Y|$

$\rightsquigarrow Z$ est à valeurs dans $\{0, 1, 2, 3\}$ et la loi de (X, Z) est donnée par :

$Z \backslash X$	1	2	3	4
0	1/16	1/16	1/16	1/16
1	1/16	1/8	1/8	1/16
2	1/16	1/16	1/16	1/16
3	1/16	0	0	1/16

Donner les loi de X et Y .

Cas où $P(X, Y)$ a une densité

On lance 2 dés à 4 faces, dont on note X et Y les résultats, entre 1 et 4.

$\rightsquigarrow$ X et Y sont indépendantes, de loi uniforme sur $\{1, \dots, 4\}$.

On définit $Z = |X - Y|$

$\rightsquigarrow$ Z est à valeurs dans $\{0, 1, 2, 3\}$ et la loi de (X, Z) est donnée par :

$Z \backslash X$	1	2	3	4	Total (loi de Z)
0	1/16	1/16	1/16	1/16	1/4
1	1/16	1/8	1/8	1/16	3/8
2	1/16	1/16	1/16	1/16	1/4
3	1/16	0	0	1/16	1/8
Total (loi de X)	1/4	1/4	1/4	1/4	1

Cas où $P(X, Y)$ a une densité : exemple

On place aléatoirement un point dans le disque de centre 0 et de rayon r .
On note X et Y ses coordonnées.

↪ Alors (X, Y) a pour densité

$$f_{(X,Y)}(x,y) = \frac{1}{\pi r^2} \mathbf{1}_{D(0,r)}(x,y) = \begin{cases} \frac{1}{\pi r^2} & \text{si } \sqrt{x^2 + y^2} \leq r \\ 0 & \text{sinon.} \end{cases}.$$

Déterminer la loi de X : la variable aléatoire X a pour densité.

Cas où $P(X, Y)$ a une densité : exemple

On place aléatoirement un point dans le disque de centre 0 et de rayon r . On note X et Y ses coordonnées.

$\leadsto$ Alors (X, Y) a pour densité

$$f_{(X,Y)}(x,y) = \frac{1}{\pi r^2} \mathbf{1}_{D(0,r)}(x,y) = \begin{cases} \frac{1}{\pi r^2} & \text{si } \sqrt{x^2 + y^2} \leq r \\ 0 & \text{sinon.} \end{cases}$$

D'où la loi de X : la variable aléatoire X a pour densité

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y) dy = \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2} dy = 2 \frac{\sqrt{r^2-x^2}}{\pi r^2} \quad \text{si } -r < x < r \\ = 0 \quad \text{sinon.}$$

Déterminer la loi de X : la variable aléatoire X a pour densité.

Calculs d'espérances

Soit $\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction.

- Si X et Y sont discrètes, alors

$$E(\phi(X, Y)) = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} \phi(x, y) P(X = x, Y = y)$$

- Si (X, Y) a pour densité $f_{(X, Y)}$, alors

$$E(\phi(X, Y)) = \int_{\mathbb{R}} \int_{\mathbb{R}} \phi(x, y) f_{(X, Y)}(x, y) dx dy$$

(À condition que les séries et les intégrales soient bien définies)
Que se passe t'il lorsque X et Y sont indépendants

Calculs d'espérances : exemple

Soit X, Y deux variables aléatoires indépendantes, de loi $\mathcal{E}(\lambda)$. On cherche

$$E\left[\frac{1}{X+Y}\right].$$

Calculs d'espérances : exemple

Soit X, Y deux variables aléatoires indépendantes, de loi $\mathcal{E}(\lambda)$. On cherche

$$E\left[\frac{1}{X+Y}\right].$$

$$\begin{aligned} E\left[\frac{1}{X+Y}\right] &= \int_0^\infty \int_0^\infty \frac{1}{x+y} f_X(x) f_Y(y) dx dy \\ &= \int_0^\infty \int_0^\infty \frac{1}{x+y} \lambda e^{-\lambda x} \lambda e^{-\lambda y} dx dy \\ &= \int_0^\infty \left(\int_0^\infty \frac{1}{x+y} \lambda^2 e^{-\lambda(x+y)} dx \right) dy \\ &= \int_0^\infty \left(\int_y^\infty \frac{1}{z} \lambda^2 e^{-\lambda z} dz \right) dy \quad \text{en posant } x \mapsto z = x + y \\ &= \int_0^\infty \left(\int_0^\infty \mathbf{1}_{(y \leq z)} \frac{1}{z} \lambda^2 e^{-\lambda z} dz \right) dy \\ &= \int_0^\infty \left(\int_0^\infty \mathbf{1}_{(y \leq z)} \frac{1}{z} \lambda^2 e^{-\lambda z} dy \right) dz \\ &= \int_0^\infty \left(\int_0^z \frac{1}{z} \lambda^2 e^{-\lambda z} dy \right) dz = \int_0^\infty \frac{1}{z} \lambda^2 e^{-\lambda z} dz = \lambda^2 \int_0^\infty e^{-\lambda z} dz = \lambda \end{aligned}$$

Somme de variables aléatoires indépendantes.

On peut souvent calculer la loi de fonctions de X et Y . Par exemple

Proposition

On suppose X et Y indépendantes, de densités f_X et f_Y et on considère

$$Z = X + Y$$

Z a pour densité.

$$f_{X+Y}(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x) dx = f_X * f_Y$$

f_{X+Y} est le produit de convolution de f_X et f_Y .

Statistiques

Les statistiques descriptives visent à décrire un ensemble, en général important, de données, c'est-à-dire à en résumer certaines particularités :

- sous la forme de représentations graphiques
- à l'aide de grandeurs numériques (moyenne,...).

L'interprétation des résultats est ensuite propre à chaque champ d'application.

On ne suppose pas les données aléatoires) $\rightarrow$ pas de probabilités.
(simplement des méthodes pour présenter un ensemble de données)
En revanche, la partie suivante (Estimation) supposera les données issues de la répétition d'une expérience aléatoire, et fera appel aux statistiques pour obtenir des renseignements sur l'expérience.

Vocabulaire : Notion de variable statistique

En statistiques, les données dont on dispose associent à chaque individu d'un certain ensemble (la population), une ou plusieurs variables qui quantifient ou qualifient certains caractères des individus. Ces données sont aussi appelées une série statistique. Pour une population de taille n (où n est l'effectif total), les données sont de la forme : $(x_1, y_1, \dots), (x_2, y_2, \dots), \dots (x_n, y_n, \dots)$ où $(x_n, y_n, \dots)$ sont les observations des variables associées au i -ième individu.

Une variable peut avoir divers types

Variable qualitative-Représentations graphiques

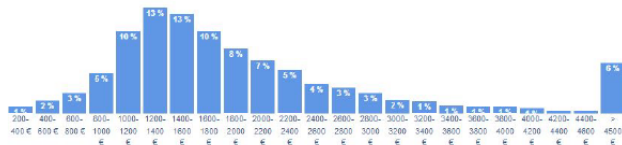
Pour une variable qualitative (ou quantitative discrète),

- l'effectif d'une modalité (ou d'une valeur) est le nombre de fois où elle est présente dans la population. Représentation : diagramme en bâtons.
- la fréquence d'une modalité (ou d'une valeur) est le quotient de l'effectif de cette valeur par l'effectif total. Représentation : diagramme circulaire (ou diagramme en bâtons, éventuellement empilés).

Variable quantitative-Représentations graphiques

Pour représenter graphiquement des observations de variables quantitatives continues (ou discrètes avec de nombreuses valeurs différentes), on choisit en général un nombre fini d'intervalles et on classe les individus selon l'intervalle qui contient leur valeur
on approche la variable continue par une variable discrète. La représentation des effectifs (ou fréquences) est un histogramme

Répartition des salaires nets mensuels des salariés



Quelle est la différence entre un histogramme et digramme en bâton

Variable quantitative-Représentations graphiques

On dispose d'observations $x_1, \dots, x_n$ d'une variable quantitative.
 Pour $x \in \mathbb{R}$, la **fréquence cumulée jusqu'à** x est la proportion d'observations inférieures à x :

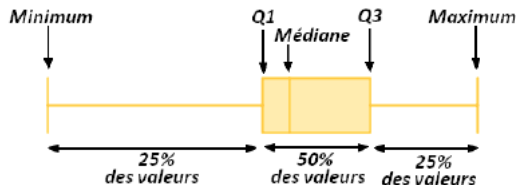
$$F(x) = \frac{1}{n} \text{Card}\{i \in \{1, \dots, n\} \mid x_i \leq x\}.$$

Pour $\alpha \in [0, 1]$, un **quantile d'ordre** α est un réel x_α tel que $F(x_\alpha) = \alpha$, c'est-à-dire qu'une proportion α des données est inférieure ou égale à x_α .
 Cas particuliers :

- Une *médiane* est un quantile d'ordre $1/2$: la moitié des observations lui sont inférieures.
- Un *premier quartile* est un quantile d'ordre $1/4$; un *troisième quartile* est un quantile d'ordre $3/4$.
- Le *minimum* est le plus grand quantile d'ordre 0, et le *maximum* est le plus petit d'ordre 1.

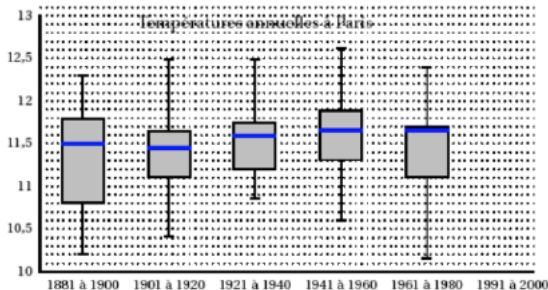
On peut représenter ces valeurs par une *boîte de dispersion* (ou « boîte à moustaches »)

Boxplot



les boîtes de dispersion peuvent avoir d'autres interprétations :
vérifier comment la boîte est définie

Boxplot



les boîtes de dispersion peuvent avoir d'autres interprétations :
vérifier comment la boîte est définie

Représentation numérique

Comment synthétiser des observations d'une variable quantitative graphiquement, via un histogramme, etc.
numériquement, en calculant des quantités qui mesurent l'ordre de grandeur des valeurs observées (ex. : médiane), quantifient leur dispersion (ex. : quantiles), leur asymétrie, etc.

Représentation numérique

Proposition

La moyenne est l'unique réel c qui minimise

$$\sum_{i=1}^n (x_i - c)^2$$

Les médianes sont les réels c qui minimisent

$$\sum_{i=1}^n |x_i - c|$$

La moyenne est sensible aux valeurs aberrantes, à la différence de la médiane.

Couple de variables

On observe **deux variables** pour chaque individu, si bien que l'on dispose de données $(x_1, y_1), \dots, (x_n, y_n)$.

La *covariance* des deux variables est

$$s_{x,y} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \bar{x} \bar{y} \quad (= \overline{xy} - \bar{x} \bar{y}),$$

(donc $s_x^2 = s_{x,x}$) et leur *corrélation* est $\rho_{x,y} = \frac{s_{x,y}}{\sigma_x \sigma_y}$.

Corrélation positive (et proche de 1) $\rightarrow$ les variables ont tendance à être simultanément grandes, ou simultanément petites.

Corrélation négative (et proche de -1) $\rightarrow$ variations en sens opposés.

Principe de l'estimation statistique

Définition

Soit X une variable aléatoire. Un échantillon de taille n de X est une famille $X_1, \dots, X_n$ de variables aléatoires indépendantes, de même loi que X .

On souhaite étudier la loi de X .

Exemple : X est la taille en centimètres d'un individu choisi uniformément dans la population adulte Ivoirienne. Son espérance est donc la taille moyenne d'un ivoirien adulte, que l'on peut vouloir estimer.

Que faire ?

Principe de l'estimation statistique

Définition

Soit X une variable aléatoire. Un échantillon de taille n de X est une famille $X_1, \dots, X_n$ de variables aléatoires indépendantes, de même loi que X .

On ne dispose pour cela que d'une réalisation d'un échantillon de taille n : une réalisation $X_1, \dots, X_n \in \mathbb{R}^n$ de n variables aléatoires indépendantes $(X_1, \dots, X_n)$ qui ont la même loi que X . À défaut de pouvoir mesurer toute la population, ce qui serait long, coûteux et compliqué, on se contente de mesurer la taille de n personnes choisies au hasard parmi les Ivoiriens adultes.

Objectif de l'estimation statistique : déduire certaines propriétés de la loi de X (espérance, variance, paramètres...) à partir d'un échantillon $X_1, \dots, X_n$

Statistiques simples

Les quantités les plus classiques pour décrire un échantillon sont

- la **moyenne empirique** : $\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$
- la **variance empirique** : $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X}_n)^2$
- la **variance empirique modifiée** : $\Sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Statistiques simples

Proposition

Si X a pour espérance m et pour écart type σ , alors :



$$E(\bar{X}_n) = m \text{ et } \text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$$



$$E(\bar{S}_n^2) = \frac{n-1}{n}\sigma^2 \text{ et } E(\Sigma_n^2) = \sigma^2$$

Estimateur

On suppose que X suit une loi P_θ qui dépend d'un paramètre $\theta \in \Theta$, où $\Theta \subseteq \mathbb{R}$ est l'ensemble des valeurs a priori possibles du paramètre. On ignore la valeur de θ , et on souhaite l'estimer.

Définition

- Un estimateur de θ est une variable aléatoire $T_n = f(X_1, \dots, X_n)$ qui dépend d'un échantillon $X_1, \dots, X_n$ de X . On utilise souvent la notation $\hat{\theta}$ pour un estimateur de θ .
- Une estimation de θ est la valeur réelle $t_n = f(x_1, \dots, x_n)$ prise par une réalisation particulière de l'échantillon.

Estimateur

On suppose que X suit une loi P_θ qui dépend d'un paramètre $\theta \in \Theta$, où $\Theta \subseteq \mathbb{R}$ est l'ensemble des valeurs a priori possibles du paramètre. On ignore la valeur de θ , et on souhaite l'estimer.

Définition : qualités d'un estimateur T_n de θ

Le biais de T_n est la différence $E(T_n) - \theta$

- On dit que T_n est sans biais si $E(T_n) = \theta$, quel que soit $\theta \in \Theta$.
- On dit que T_n est asymptotiquement sans biais si $E(T_n) \rightarrow \theta$, quel que soit $\theta \in \Theta$.
- On dit que T_n converge si quel que soit $\theta \in \Theta$. pour tout $\alpha > 0$

$$P(|T_n - \theta| > \alpha) \rightarrow 0$$

Estimateur

Proposition

On suppose que X a pour espérance m et variance σ^2 .

- La moyenne empirique est un estimateur sans biais et convergent de m .
- La variance empirique est un estimateur asymptotiquement sans biais et convergent de σ^2 , et la variance empirique modifiée est un estimateur sans biais et convergent de σ^2 .

Exemple : Si on sait que $X_1, \dots, X_n$ suivent la loi $P(\lambda)$, et que l'on cherche la valeur de λ , on peut utiliser le fait que λ est l'espérance des X_i , donc la moyenne empirique $\bar{X}_n$ est un estimateur sans biais, convergent, de λ .

En fait, λ est aussi la variance des X_i donc on pourrait aussi utiliser la variance empirique (modifiée ou non) pour estimer λ .

Estimateurs-Risque quadratique

Comment mesurer la qualité et comparer deux estimateurs ?

Proposition

Le risque quadratique d'un estimateur T_n de θ est

$$R_{T_n}(\theta) = E((T_n - \theta)^2)$$

On dit que l'estimateur S_n est meilleur que T_n si, quel que soit θ .

$$R_{S_n}(\theta) \leq R_{T_n}(\theta)$$

Par l'inégalité de Markov, un estimateur dont le risque quadratique tend vers 0 (quel que soit θ) est convergent. NB. Si T_n est sans biais, alors $R_{T_n}(\theta) = \text{Var}(T_n)$.

Construction d'estimateurs

Problème

Comment trouver un “bon” estimateur d'un paramètre θ ?

Deux méthodes classiques :

- 1 méthode des moments
- 2 méthode du maximum de vraisemblance

Méthode des moment : Karl Pearson en 1894

Le principe est d'utiliser la loi des grands nombres pour estimer les moments, et d'utiliser ensuite ces estimateurs des moments pour estimer θ . Par la loi des grands nombres, on a :

Proposition

Soit X une variable aléatoire d'espérance m et de variance σ^2 .

- 1 La moyenne empirique est un estimateur sans biais et convergent de m .
- 2 La variance empirique est un estimateur asymptotiquement sans biais et convergent de σ^2 , et la variance empirique modifiée est un estimateur sans biais et convergent de σ^2 .
- 3 Pour tout $r > 0$, le moment empirique d'ordre r ,

$$\widehat{m}_r = \frac{1}{n} \sum_{i=1}^n (X_i)^r$$

Construction d'estimateurs

On en déduit la méthode des moments :

- 1 calculer les moments $m_1 = E(X)$, $m_2 = E(X^2)$, etc., jusqu'à pouvoir exprimer θ à l'aide de ceux-ci ;
- 2 remplacer dans cette expression les moments par les moments empiriques : m_1 remplacé par $\bar{X}_n$ etc.

Ceci fournit un estimateur convergent de θ . L'expression peut aussi faire intervenir σ^2 , que l'on remplace par S_n^2 ou Σ_n^2

Exemple

EMM pour le modèle exponentiel $\{\mathcal{E}(\theta), \theta > 0\}$. La densité de F_θ est

$$f(x, \theta) = \theta^{-1} e^{-x/\theta} \mathbb{1}_{\{x > 0\}}.$$

$$\mu_1(\theta) = E_\theta(X_1) = \frac{1}{\theta} \int_0^\infty x e^{-\frac{x}{\theta}} dx = \theta,$$

Exemple

et la solution $\tilde{\theta}_n^{(1)} = \bar{X}$ de l'équation

$$\theta = \bar{X}, \quad (\text{ou } \mu_1(\theta) = m_1 = \bar{X})$$

est un estimateur par la méthode des moments du paramètre θ . On remarque aussi qu

$$\mu_2(\theta) = \frac{1}{\theta} \int_0^\infty x^2 e^{-\frac{x}{\theta}} dx = 2\theta^2.$$

Exemple

Par conséquent, la solution $\hat{\theta}_n^{(2)} = \sqrt{\frac{\bar{X}^2 + s^2}{2}}$ de l'équation $2\theta^2 = m_2$, où

$$m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \bar{X}^2 + s^2$$

est un autre estimateur par la méthode des moments de θ . De plus, comme $\bar{X}$ (p.s.) et $s^2 \rightarrow \sigma^2 = E_\theta[(X_1 - \mu_1)^2] = \theta^2$ (p.s.), d'après le Premier théorème (Proposition 1.10),

$$\sqrt{\frac{\bar{X}^2 + s^2}{2}} \rightarrow \sqrt{\frac{\mu_2(\theta)}{2}} = \theta \quad (\text{p.s.}).$$

Méthode du maximum de vraisemblance

Principe : estimer θ par la valeur qui maximise la densité de $(X_1, \dots, X_n)$.

La vraisemblance de l'échantillon $(X_1, \dots, X_n)$ est la fonction L , où :

- si X est discrète, de loi P_θ pour tous $X_1, \dots, X_n \in X(\Omega)$

$$L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n P_\theta(x_i)$$

- si X est continue, de densité f_θ , pour tous $X_1, \dots, X_n \in X(\Omega)$

$$L(X_1, \dots, X_n, \theta) = \prod_{i=1}^n f_\theta(x_i)$$

Méthode du maximum de vraisemblance

Principe : estimer θ par la valeur qui maximise la densité de $(X_1, \dots, X_n)$.

La vraisemblance de l'échantillon $(X_1, \dots, X_n)$ est la fonction L , où :
Un estimateur du maximum de vraisemblance pour θ est un estimateur $h(X_1, \dots, X_n)$ tel que, pour tous $x_1, \dots, x_n \in X(\Omega)$,

$$L(x_1, \dots, x_n; h(x_1, \dots, x_n)) = \max_{\theta \in \Omega} L(x_1, \dots, x_n; \theta)$$

Méthode du maximum de vraisemblance

En pratique,

- on calcule la vraisemblance $L(x_1, \dots, x_n, \theta)$
- pour $x_1, \dots, x_n$ constants, on cherche θ qui maximise la fonction

$$\phi(\theta) = L(x_1, \dots, x_n, \theta)$$

- en général, on dérive ϕ et on résout $\phi'(\theta) = 0$ (et on étudie les variations de ϕ' pour voir que c'est un maximum).
- souvent, il est plus simple de chercher le maximum de

$$\psi(\theta) = \ln L(x_1, \dots, x_n, \theta)$$

(c'est équivalent car le logarithme est strictement croissant) et donc calculer de ϕ , etc.

- L'estimateur du maximum de vraisemblance s'obtient en prenant $x_1, \dots, x_n$ égaux aux variables aléatoires $x_1, \dots, X_n$.

Intervalles de confiance

Un estimateur fournit une valeur (l'estimation), sensée être proche du paramètre θ estimé. Quelle est la marge d'erreur "typique" d'un estimateur ?

IC

Un intervalle de confiance de niveau $1 - \alpha$ est un intervalle I , qui dépend de $X_1, \dots, X_n$, contenant la valeur θ avec probabilité $1 - \alpha$.

NB. L'intervalle dépend du niveau de confiance, qu'il faut en principe choisir selon l'enjeu du résultat. En pratique, on prendra souvent un niveau de 95% (c'est-à-dire vérifié 19 fois sur 20).

Théorème Central Limite

Soit X une variable aléatoire de moyenne m et de variance σ^2 . Soit $(X_1, \dots, X_n)$ un échantillon de X . On cherche un intervalle de confiance pour la moyenne, à partir des estimateurs sans biais $\bar{X}_n$ et Σ_n^2 . Résultat crucial : quelle que soit la loi de X , $\bar{X}_n$ a le même type de loi :

TCL

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes, de même loi de moyenne m et d'écart type σ . Soit Z_n la variable aléatoire définie par

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \rightarrow Z \sim \mathcal{N}(0, 1)$$

On a une convergence en loi.

Dans la pratique on pourra souvent appliquer ce résultat dès que $n > 30$.

Intervalle de confiance pour la moyenne m

Soit $Z \sim \mathcal{N}(0, 1)$ et a et α définis par

$$P(-a \leq Z \leq a) = 1 - \alpha$$

Si $n \geq 30$, le théorème de la limite centrale permet d'écrire

$$P\left(-a \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \leq a\right) \simeq 1 - \alpha$$

On obtient, vu que Σ_n^2 est un estimateur convergent de σ^2

$$P\left(-a \leq \frac{\sqrt{n}(\bar{X}_n - m)}{\Sigma_n} \leq a\right) \simeq 1 - \alpha$$

ce qui peut se réécrire

$$P\left(\bar{X}_n - a \frac{\Sigma_n}{\sqrt{n}} \leq m \leq \bar{X}_n + a \frac{\Sigma_n}{\sqrt{n}}\right) \simeq 1 - \alpha$$

Intervalle de confiance pour une proportion p

Ici X est une variable aléatoire de Bernoulli de paramètre p , pour n grand :

$$E(X) = p \text{ et } \text{Var}(X) = p(1 - p)$$

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1 - p)}} \sim \mathcal{N}(0, 1)$$

et l'intervalle de confiance pour p de niveau $1 - \alpha$ précédent est donc

$$IC = \left[\bar{X}_n - a \frac{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}{\sqrt{n}}, \bar{X}_n + a \frac{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}{\sqrt{n}} \right]$$

Intervalle de confiance pour une proportion p

- Attention : pour que ces approximations soient justifiées, les valeurs de $n\pi_1$, $n\pi_2$, $n(1 - \pi_1)$, $n(1 - \pi_2)$ doivent être toutes les quatre supérieures ou égales à 5.
- Les intervalles de confiance donnés ci-dessus permettent aussi de déterminer la grandeur n de l'échantillon nécessaire pour avoir une précision donnée pour l'estimation d'une proportion.

Tests

- On souhaite, à partir de l'échantillon observé $x_1, \dots, x_n$, savoir si l'on peut raisonnablement conclure qu'une certaine hypothèse sur la loi de X est fausse (en vue de prendre une décision).
- L'hypothèse est appelée hypothèse nulle, et notée H_0 . C'est une hypothèse que l'on veut avoir *peu de chance* de rejeter si elle est vraie (dans ce cas, on parle *d'erreur de 1re espèce*).
- Un test est une condition sur l'échantillon $x_1, \dots, x_n$ pour décider si on rejette H_0 , ou si on considère les données compatibles avec H_0 . Le seuil de risque du test est la probabilité d'erreur de première espèce :

$$\alpha = P(\text{rejet de } H_0) \text{ quand } H_0 \text{ est vraie}$$

α doit être petit, en général on souhaite $\alpha = 5\%$ (varie selon l'enjeu du test).

Tests

- On se donne également une hypothèse alternative H_1 (incompatible avec H_0), que l'on veut détecter. Souvent H_1 sera le contraire de H_0 . La puissance du test est $1 - \beta$, où β est la probabilité d'erreur de 2e espèce :

$$\beta = P(\text{acceptation de } H_0) \text{ quand } H_1 \text{ est vraie}$$

- On souhaite avoir β petit (donc une puissance $1 - \beta$ proche de 1) ; cependant en général on ne contrôle pas cette erreur, mais on choisit un test qui la minimise, et on veut s'assurer qu'elle converge vers 0 quand $n \rightarrow \infty$.

Test sous toutes ses formes par la pratique

Chez un individu adulte, le logarithme du dosage en d-dimères, variable que nous noterons X_i , est modélisé par une loi normale d'espérance μ et de variance σ^2 en d'autres termes la fonction de densité associée est : $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. La variable X_i est un indicateur de risque cardio-vasculaire : on considère que chez les individus sains, μ vaut -1, alors que chez les individus à risque, μ vaut 0. Dans les deux cas, la valeur de σ est la même.

Le Dr Soum ne souhaite pas alarmer inutilement ses patients. Il pose l'hypothèse $H_0 : \mu = -1$ contre l'alternative $H_1 : \mu = 0$.

Test d'adéquation de m à une valeur théorique

1-) Donner la règle de décision pour son test (C'est à dire définir la région critique au niveau α), au seuil de $\alpha = 5\%$. Application

Numérique : prendre $n=1$, $\sigma = 0.09$

2-) Calculer le risque de deuxième espèce et la puissance du test de la question précédente. Application Numérique : prendre $n=1$, $\sigma = 0.09$

3-) Un patient numéro i présente une valeur de X noté X_i égale à -0.46 . Calculer la p-valeur du test du Dr. Soum. Application

Numérique : prendre $n=1$, $\sigma = 0.09$

4-) Le Dr. O a pour point de vue qu'il vaut mieux alarmer à tort un patient plutôt que de ne pas l'avertir d'un risque réel. Il propose le test d'hypothèse suivant : $H'_0 : \mu = 0$ contre l'alternative

$H'_1 : \mu = -1$. Donner la règle de décision pour son test, au seuil de 5% . Application Numérique : prendre $n=1$, $\sigma = 0.09$

5-) Selon le seuil, pour quelles valeurs de la statistique X_n les deux médecins seront-ils d'accord ? Application Numérique : prendre $n=1$, 

Test d'adéquation de m à une valeur théorique

Échantillon $X_1, \dots, X_n$ de même loi que X , $E[X] = m$ et $\text{Var}(X) = \sigma^2$.

On veut savoir si l'observation $X_1, \dots, X_n$ est compatible avec l'hypothèse

$$\mathcal{H}_0 : "m = 7,5"$$

(par exemple)

Si $\mathcal{H}_0$ est vraie, le TCL montre que $P\left(-a \leq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - 7,5) \leq a\right) \simeq 1 - \alpha$,
ou encore en approchant σ par $\sqrt{\Sigma_n^2}$,

$$P\left(-a \leq \frac{\sqrt{n}}{\sqrt{\Sigma_n^2}}(\bar{X}_n - 7,5) \leq a\right) \simeq 1 - \alpha.$$

Test d'adéquation de m à une valeur théorique

D'où le test, de seuil de risque α : (test bilatère)

$$\text{Si } \frac{\sqrt{n}}{\sqrt{\Sigma_n^2}} |\bar{X}_n - 7,5| > a, \text{ alors on rejette } \mathcal{H}_0.$$

NB. Si $\mathcal{H}_0$ est fausse, $\bar{X}_n - 7,5 \simeq m - 7,5 \neq 0$ donc le test a tendance à être rejeté, pour n grand. La puissance du test (avec $\mathcal{H}_1 : "m \neq 7,5"$) tend vers 1.

Test d'adéquation de m à une valeur théorique

Si on veut seulement exclure les cas où m est trop grand par rapport à la valeur théorique, on utilise l'hypothèse alternative

$$\mathcal{H}_1 : "m > 7,5"$$

et on adapte le test pour rejeter surtout lorsque $\mathcal{H}_1$ est vraie (et donc améliorer la puissance) :

Si $\mathcal{H}_0$ est vraie, le TCL montre que $P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - 7,5) \leq A\right) \simeq 1 - \alpha$, avec A tel que $P(Z \leq A) \simeq 1 - \alpha$ pour $Z \sim \mathcal{N}(0,1)$ (par ex. $P(Z \leq 1,65) \simeq 0,95$) ou encore, en approchant σ par $\sqrt{\Sigma_n^2}$,

$$P\left(\frac{\sqrt{n}}{\sqrt{\Sigma_n^2}}(\bar{X}_n - 7,5) \leq A\right) \simeq 1 - \alpha.$$

D'où le test, de seuil de risque α : **(test unilatère)**

$$\text{Si } \frac{\sqrt{n}}{\sqrt{\Sigma_n^2}}(\bar{X}_n - 7,5) > A, \text{ alors on rejette } \mathcal{H}_0.$$

NB. On a $A \leq a$ donc si $\mathcal{H}_1$ est vraie, le test est plus souvent rejeté que le précédent, ce qui montre que sa puissance est plus grande.

Produit de convolution

Produit de convolution

calculer $f * g$ avec

$$f(x) = \begin{cases} 1 & \text{si } x \in [-1, 3[\\ 0 & \text{si } x \geq 3 \end{cases}$$

et

$$g(x) = \begin{cases} \frac{x}{2} & \text{si } x \in [0, 2] \\ 0 & \text{si } x \geq 2 \end{cases}$$

Nombre de blessés

Le modèle suivant peut être utilisé pour représenter le nombre de blessés dans les accidents de la circulation au cours d'un week-end. Le nombre d'accidents suit une loi de Poisson de paramètre λ . Le nombre de blessés par accident, suit une loi de Poisson de paramètre μ . Le nombre total de blessés est donc :

$S = X_1 + X_2 + \dots + X_N$ S est la somme d'un nombre aléatoire de variables de Poisson, indépendantes et de même loi.

- ❶ Donner une expression pour $P(S = s)$
- ❷ Calculer $P(S = 0)$.
- ❸ Calculer $E(S)$ et $V(S)$

Soit X une variable aléatoire suivant une loi de densité :

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x}$ pour $x > 0$. Soit Y une autre variable aléatoire. On suppose que la loi conditionnelle de Y sachant X est une loi normale de paramètres $m = 0$ et $\sigma^2 = \frac{1}{2X}$

- 1 Calculer la loi du couple (Y, X)
- 2 Quelle est la loi conditionnelle de X sachant Y ?
- 3 En déduire $E(X/Y)$

Une Compagnie pétrolière BEA a le choix entre trois décisions :
 vendre la concession (gain direct de 125 000) conserver la concession pendant 1 an et vendre

- Baisse du prix futur du pétrole la vente n'est plus que de 110K (prob. associée 0.9)
- Hausse du prix du pétrole la vente est de 440K (prob. complémentaire de 0.1)

exploiter la concession avec un coût de 200 000

- Le puits est sec (prob. associé 0.5), il n'y aura pas de recette
- Le puits est peu productif (prob. associée 0.4), il y aura 400k de recettes
- Le puits est très productif (prob. associée 0.1), il y aura 1500k de recettes

La compagnie pétrolière BEA décide de lancer une analyse sismique préalable qui déterminera la structure géologique du sol dans lequel forer. Il existe 3 sortes de structures : - absence de structure (N) ; - structure ouverte (O) ; - la structure est fermée (C).