Cours Commun Scientifique de Probabilités & Statistiques

Laurent Tournier

Janvier 2016

UNIVERSITÉ PARIS 13

Eppur si muove !

Institut Galilée

Plan du cours

Espaces de probabilités

Définition

*Un **espace de probabilité** (Ω, P) est constitué de*

- *Ω, un ensemble*
- *P, une probabilité sur Ω.   (à définir tout à l'heure)*

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 :

Notation	Sens mathématique	Interprétation en probabilités
$A^c (= \Omega \setminus A)$	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, $\Omega = \{P,F\}$ ou $\{0,1\}$
- lancer d'un dé, $\Omega = \{1,2,3,4,5,6\}$
- lancer de deux pièces, $\Omega = \{P,F\}^2 = \{(P,P),(P,F),(F,P),(F,F)\}$
- choix de deux parts dans une galette coupée en 8 :

  $$\Omega = \{(i,j) \mid i,j \in \{1,\ldots,8\} \text{ et } i \neq j\}$$

  ou, si l'ordre n'a pas d'importance,

  $$\Omega = \{\{i,j\} \mid i,j \in \{1,\ldots,8\} \text{ et } i \neq j\}$$

- attente d'un bus qui passe toutes les T minutes, $\Omega = [0,T] \subset \mathbb{R}$
- placement d'une fève circulaire dans une galette, $\Omega = D(0,R) \subset \mathbb{R}^2$

Pour une même expérience, divers choix de Ω sont possibles. Souvent, on ne décrira pas Ω et on fera des hypothèses sur des événements (et des variables aléatoires) en sachant qu'*il existe* un espace de probabilité Ω convenable.

Espaces de probabilités

Pour $A \subset \Omega$, $P(A)$ est la « proportion de chance » que A se réalise.
Intuition : si on répète l'expérience, $P(A)$ est la proportion des fois où A se réalise (cf. Loi des grands nombres).

Définition

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

1. $P(\Omega) = 1$
2. *pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements **disjoints deux à deux**,*
$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n).$$

*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

Pour $A \subset \Omega$, $P(A)$ est la « proportion de chance » que A se réalise.
Intuition : si on répète l'expérience, $P(A)$ est la proportion des fois où A se réalise (cf. Loi des grands nombres).

Définition

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

1. $P(\Omega) = 1$
2. *pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'événements **disjoints deux à deux**,*
$$P\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} P(A_n).$$

*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. ».*

Pour simplifier, on suppose ici que l'on peut définir la probabilité **tous les événements**. En vérité, ce n'est pas possible dans certains cas, mais cela ne posera pas de problème pratique.

Espaces de probabilités

Propriétés

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

(ii) *Pour tout événement* A, $P(A^c) = 1 - P(A)$

(iii) *Si* $A \subset B$, *alors* $P(A) \leq P(B)$

(iv) *Pour tous événements* A *et* B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Espaces de probabilités

Propriétés

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

(ii) *Pour tout événement* A, $P(A^c) = 1 - P(A)$

(iii) *Si* $A \subset B$, *alors* $P(A) \leq P(B)$

(iv) *Pour tous événements* A *et* B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Preuve de (ii) : A et A^c sont disjoints ($A \cap A^c = \emptyset$), 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 (i) en prenant $A = \Omega$.

Espaces de probabilités

Propriétés

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

(ii) *Pour tout événement* A, $P(A^c) = 1 - P(A)$

(iii) *Si* $A \subset B$, *alors* $P(A) \leq P(B)$

(iv) *Pour tous événements* A *et* B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Preuve de (ii) : A et A^c sont disjoints ($A \cap A^c = \emptyset$), 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 (i) en prenant $A = \Omega$.
Preuve de (iii) : A et $B \setminus A$ sont disjoints, et $A \cup (B \setminus A) = B$ donc

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$$

et $P(B \setminus A) = P(B) - P(A)$.

Espaces de probabilités

Propriétés

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

(ii) *Pour tout événement* A, $P(A^c) = 1 - P(A)$

(iii) *Si* $A \subset B$, *alors* $P(A) \leq P(B)$

(iv) *Pour tous événements* A *et* B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Preuve de (ii) : A et A^c sont disjoints ($A \cap A^c = \emptyset$), 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 (i) en prenant $A = \Omega$.
Preuve de (iii) : A et $B \setminus A$ sont disjoints, et $A \cup (B \setminus A) = B$ donc

$$P(B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$$

et $P(B \setminus A) = P(B) - P(A)$.
Preuve de (iv) : $A \setminus (A \cap B)$, $A \cap B$ et $B \setminus (A \cap B)$ sont disjoints, d'union $A \cup B$, donc

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

Distribution uniforme de probabilité

On suppose que Ω est fini, avec $\operatorname{Card}\Omega = n$:

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}.$$

Si ces résultats jouent des rôles symétriques, il est naturel de considérer la probabilité uniforme sur Ω, telle que

$$P(\{\omega_1\}) = \cdots = P(\{\omega_n\}) = \frac{1}{n}.$$

Définition

*La **probabilité uniforme** sur Ω (ou distribution équiprobable) est la probabilité P définie par : pour tout $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\} \subset \Omega$,*

$$P(A) = \frac{k}{n} = \frac{\operatorname{Card} A}{\operatorname{Card}\Omega}.$$

Autrement dit,

$$P(\text{événement}) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Rappels de dénombrement :

Soit E un ensemble fini.

Une **permutation** de E est une façon d'ordonner les éléments de E.
Le nombre de permutations d'un ensemble à n éléments est

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

Un **arrangement** de k éléments de E est une suite de k éléments de E distincts 2 à 2. *L'ordre est important.*
Le nombre d'arrangements de k éléments parmi n éléments est

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

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.
Le nombre de combinaisons de k éléments parmi n éléments est

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

Exemple : paradoxe des anniversaires

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

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

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

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

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

et on cherche $P(A)$ où

$$A = \{2 \text{ étudiants sont nés le même jour}\} = \{(j_1, \ldots, j_n) \in \Omega \,|\, \exists k \neq l, j_k = j_l\}.$$

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

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

et on cherche $P(A)$ où

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

Alors

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

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

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

et on cherche $P(A)$ où

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

Alors

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

et

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

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

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

et on cherche $P(A)$ où

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

Alors

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

et

$$P(A) = 1 - P(A^c) = 1 - \frac{\mathrm{Card}(A^c)}{\mathrm{Card}(\Omega)} = 1 - \frac{A_N^n}{N^n} = 1 - \frac{N(N-1)\cdots(N-n+1)}{N^n}.$$

Exemple : Pour $n = 23$, $P(A) \simeq 0{,}5$. Pour $n = 57$, $P(A) \simeq 0{,}99$.

Quelles probabilités pour le bus et la galette ?

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

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

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

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

Pour la position d'une fève circulaire dans une galette, $\Omega = D(0,R)$

Si la fève est mise "complètement au hasard",
- la fève a autant de chance d'être dans A que dans B si A et B ont même *aire*.
- la fève a 2 fois plus de chances d'être dans A que dans B si l'aire est double.

$\rightsquigarrow$ la probabilité que la fève soit dans une partie A est proportionnelle à l'aire de A : (« loi uniforme sur $D(0,R)$ »)

$$P(A) = \frac{\text{aire}(A)}{\text{aire}(D(0,R))} = \frac{\text{aire}(A)}{\pi R^2}.$$

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.
$\rightsquigarrow$ C'est la probabilité de A si on dispose de l'information que B est réalisé.

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

$$P(A|B) = P(A)$$

$\rightsquigarrow$ Savoir que B est réalisé n'influence pas 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 %.
$\rightarrow$ Quelle est la probabilité que j'aie la fève ?

On note $F = \{\text{j'ai la fève}\}$ et $A_5 = \{\text{nous sommes } 5\}$, A_6 et A_7 de même.
Alors :

$$P(A_5) = 0{,}5 \quad P(A_5) = 0{,}3 \quad P(A_7) = 0{,}2$$

et

$$P(F|A_5) = \frac{1}{5}, \quad P(F|A_6) = \frac{1}{6}, \quad P(F|A_7) = \frac{1}{7},$$

d'où

$$\begin{aligned} P(F) &= P(F \cap A_5) + P(F \cap A_6) + P(F \cap A_7) \\ &= P(F|A_5)P(A_5) + P(F|A_6)P(A_6) + P(F|A_7)P(A_7) \\ &= 0{,}18. \end{aligned}$$

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 %.
$\rightarrow$ Quelle est la probabilité que j'aie la fève ?

On note $F = \{\text{j'ai la fève}\}$ et $A_5 = \{\text{nous sommes 5}\}$, A_6 et A_7 de même.
Alors :

$$P(A_5) = 0{,}5 \quad P(A_5) = 0{,}3 \quad P(A_7) = 0{,}2$$

et

$$P(F|A_5) = \frac{1}{5}, \quad P(F|A_6) = \frac{1}{6}, \quad P(F|A_7) = \frac{1}{7},$$

d'où

$$\begin{aligned} P(F) &= P(F \cap A_5) + P(F \cap A_6) + P(F \cap A_7) \\ &= P(F|A_5)P(A_5) + P(F|A_6)P(A_6) + P(F|A_7)P(A_7) \\ &= 0{,}18. \end{aligned}$$

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) :

$$\text{pour tous } i \neq j,\ A_i \cap A_j = \emptyset, \qquad \text{et} \qquad \Omega = \bigcup_n A_n.$$

Par exemple, pour tout événement B, le couple (B,B^c) est une partition de Ω.

Théorème (Théorème des probabilités totales)

$$P(A) = \sum_n P(A \cap A_n) = \sum_n P(A|A_n)P(A_n).$$

En particulier, pour tous A et B, $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$.

Théorème (Formule de Bayes)

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

En particulier, pour tous A et B,

$$P(B^c|A) = \frac{P(A|B^c)P(B^c)}{P(A|B)P(B) + P(A|B^c)P(B^c)}\ .$$

Événements indépendants : cas général

Rappel : 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 $A_{i_1}, \ldots, A_{i_k}$ on a

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

En particulier, des événements A, B et C sont indépendants si

$$P(A \cap B) = P(A)P(B), \qquad P(B \cap C) = P(B)P(C), \qquad P(A \cap C) = P(A)P(C)$$

$$\textbf{et} \qquad P(A \cap B \cap C) = P(A)P(B)P(C).$$

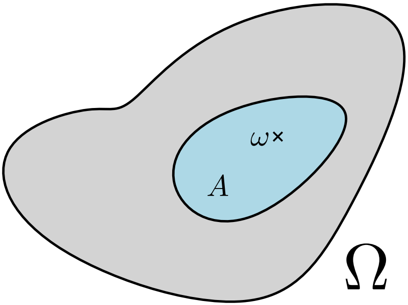
$\rightsquigarrow$ alors, par exemple, $A \cap B$ et C sont indépendants

Cours 2

–

Mercredi 20 janvier 2016

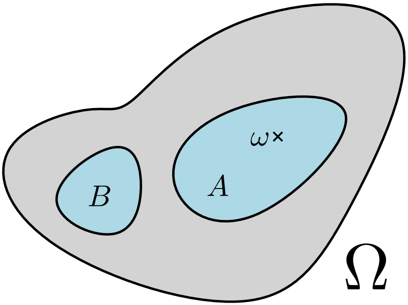
Espace de probabilités – Rappel



Espace de probabilités : (Ω, P)

- Ω, ensemble des résultats possibles d'une expérience aléatoire
- $\omega \in \Omega$, une *réalisation* de l'expérience
- $A \subset \Omega$, un *événement* relatif à l'expérience (peut être *réalisé* ou non)
- $P(A) \in [0,1]$, *probabilité* de l'événement A (d'où $P : \mathcal{P}(\Omega) \to [0,1]$)

Espace de probabilités – Rappel



Espace de probabilités : (Ω,P)

- Ω, ensemble des résultats possibles d'une expérience aléatoire
- $\omega \in \Omega$, une *réalisation* de l'expérience
- $A \subset \Omega$, un *événement* relatif à l'expérience (peut être *réalisé* ou non)
- $P(A) \in [0,1]$, *probabilité* de l'événement A (d'où $P : \mathcal{P}(\Omega) \to [0,1]$)

telle que $P(\Omega) = 1$ et, si A et B sont disjoints, $P(A \cup B) = P(A) + P(B)$.
Et, si on a une suite $(A_n)_n$ d'événements disjoints, $P(\bigcup_n A_n) = \sum_n P(A_n)$.

Événements indépendants

Rappel : 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 $A_{i_1}, \ldots, A_{i_k}$ on a

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

En particulier, des événements A, B et C sont indépendants si

$$P(A \cap B) = P(A)P(B), \qquad P(B \cap C) = P(B)P(C), \qquad P(A \cap C) = P(A)P(C)$$

$$\textbf{et} \qquad P(A \cap B \cap C) = P(A)P(B)P(C).$$

$\rightsquigarrow$ alors, par exemple, $A \cap B$ et C sont indépendants

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.

Preuve :

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

Par récurrence, on peut obtenir :

Proposition

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

De là on pourrait déduire que, si $A_1, \ldots, A_n$ sont indépendants, alors des événements $B_1, \ldots, B_k$ qui dépendent de *paquets disjoints* d'événements parmi $A_1, \ldots, A_n$ sont indépendants.

$\rightsquigarrow$ Par ex., dans un jeu de pile-ou-face, si $A_i = \{\text{le } i^{\text{ème}} \text{ tirage est pile}\}$, $A_1, A_2, \ldots$ sont indépendants, et donc B_1, B_2, B_3 sont indépendants, où

$$B_1 = A_1 \cap A_2^c, \quad B_2 = A_5 \cup A_6, \quad B_3 = A_4.$$

Loi binomiale

On considère n tirages à Pile-ou-Face avec la même pièce biaisée, qui tombe sur Pile avec probabilité p (et donc sur Face avec probabilité $1-p$).

On note 1 pour Pile et 0 pour Face
$\to$ chaque réalisation ω est une suite de 0 et de 1 de longueur n : $\Omega = \{0,1\}^n$.
Les tirages sont indépendants, donc par exemple (ici, $n = 4$)

$$P(\{(1,0,1,1)\}) = p \times (1-p) \times p \times p = p^3(1-p)$$

et, si la séquence $\omega = (\varepsilon_1, \ldots, \varepsilon_n)$ contient k fois 1 (et donc $n-k$ fois 0),

$$P(\{\omega\}) = p^k(1-p)^{n-k}.$$

Soit $0 \leq k \leq n$. On définit l'événement

$$B_k = \{\text{Exactement } k \text{ pièces tombent sur Pile}\}.$$

On vient de voir que, pour toute séquence $\omega \in B_k$, $P(\{\omega\}) = p^k(1-p)^{n-k}$.
Par ailleurs, le nombre de telles séquences est $\operatorname{Card} B_k = \binom{n}{k}$. On en déduit

$$\boxed{P(B_k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

Loi binomiale

Plus généralement, si on a n événements **indépendants** $A_1, \ldots, A_n$ ayant tous **la même probabilité** $P(A_i) = p$, pour $k = 0, \ldots, n$, on obtient de même

$$P(\text{exactement } k \text{ événements parmi } A_1, \ldots, A_n \text{ se réalisent}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dans la situation précédente, $A_i = \{\text{le } i\text{-ième tirage est Pile}\}$.

Loi binomiale

Plus généralement, si on a n événements **indépendants** $A_1,\ldots,A_n$ ayant tous **la même probabilité** $P(A_i) = p$, pour $k = 0,\ldots,n$, on obtient de même

$$P(\text{exactement } k \text{ événements parmi } A_1,\ldots,A_n \text{ se réalisent}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dans la situation précédente, $A_i = \{\text{le } i\text{-ième tirage est Pile}\}$.

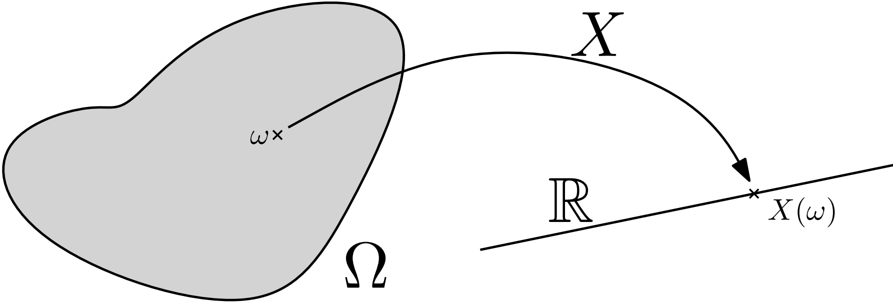
En notant X le nombre de fois où Pile est apparu parmi les n lancers, X est une **variable aléatoire** qui suit la **loi binomiale** $\mathcal{B}(n,p)$.

Plan du cours

Variables aléatoires

Définition

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



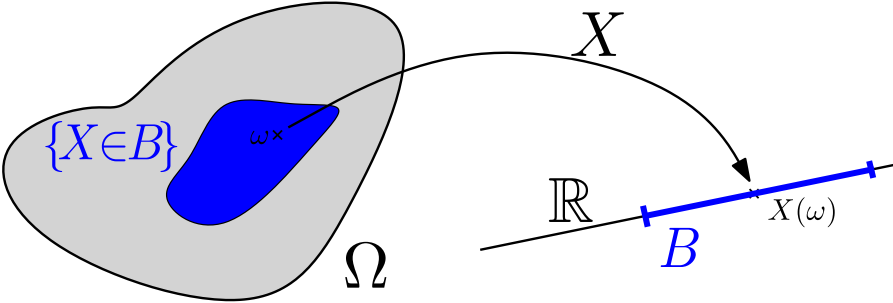
Variables aléatoires

Définition

*Une **variable aléatoire** est une application $X : \Omega \to \mathbb{R}$.*
*La **loi** de X est la probabilité P_X sur $\mathbb{R}$ définie par :*

$$\text{pour tout } B \subset \mathbb{R}, \qquad P_X(B) = P(\{\omega \in \Omega \,|\, X(\omega) \in B\}) = P(X \in B).$$

*$X(\Omega)$ (image de X) est le **support** de P_X.*
$\rightsquigarrow$ P_X peut aussi être vue comme une probabilité sur $X(\Omega)$.
On note parfois $X \sim P_X$ pour indiquer que X suit la loi 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** égale à c et n'est donc pas "aléatoire".
  En général, la valeur de $X(\omega)$ dépend de la réalisation ω, et sa distribution sur $\mathbb{R}$ est donnée par la loi de X.
- Notation : On a noté $\{X \in B\}$ l'événement formé des éventualités ω pour lesquelles $X(\omega) \in B$, et on abrège

$$P(X \in B) = P(\{X \in B\}) = P(\{\omega \in \Omega \,|\, X(\omega) \in B\}).$$

- Exemple le plus simple :

Définition

Si A est un événement, on introduit la variable aléatoire ***fonction indicatrice de*** *A, notée $\mathbf{1}_A$, qui indique si l'événement A est réalisé :*

$$\text{pour tout } \omega \in \Omega, \qquad \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Variables aléatoires – Exemples

- Lancer de deux dés, $\Omega = \{1, \ldots, 6\}^2 = \{(x_1, x_2) \mid x_1, x_2 \in \{1, \ldots, 6\}\}$
  - Valeurs des dés : $X_1((x_1, x_2)) = x_1$ et $X_2((x_1, x_2)) = x_2$
    (à valeurs dans $\{1, \ldots, 6\}$)
  - Somme des résultats : $X = X_1 + X_2$, c.-à-d. $X((x_1, x_2)) = x_1 + x_2$
    (à valeurs dans $\{2, \ldots, 12\}$)
- Placement d'une fève circulaire dans une galette, $\Omega = D(0, r) \subset \mathbb{R}^2$
  - Coordonnées du point : $X((x, y)) = x$, $Y((x, y)) = y$
    (à valeurs dans $[-r, r]$)
  - Distance au centre : $R = \sqrt{X^2 + Y^2}$
    (à valeurs dans $[0, r]$)
- On prend successivement les parts d'une galette (coupée en 8)
  - Nombre de parts à prendre jusqu'à avoir la fève : N_A
    (à valeurs dans $\{1, \ldots, 8\}$)
- Chaque jour, on prend une part d'une galette *différente* (coupée en 8)
  - Nombre de parts à prendre jusqu'à avoir la fève : N_B
    (à valeurs dans $\{1, 2, \ldots\} = \mathbb{N}^*$)
  - Nombre de fèves obtenues en n jours : S_n
    (à valeurs dans $\{0, 1, 2, \ldots, n\}$)

Lois discrètes

Définition

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

(C'est-à-dire que l'on peut trouver une suite qui *énumère tous les éléments* de $X(\Omega)$: par ex., si $X(\Omega)$ est un ensemble fini, $\mathbb{N}$, $\mathbb{Z}$ ou $\mathbb{Q}$, mais pas l'intervalle $[0,1]$ ni $\mathbb{R}$).

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

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

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

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

$\rightsquigarrow$ Pour caractériser une loi discrète, il suffit donc de se donner les **probabilités élémentaires** $p_X(x) = P(X = x)$ pour tout $x \in X(\Omega)$. On a

$$\text{pour tout } x \in X(\Omega),\ p_X(x) \geq 0, \quad \text{et} \quad \sum_{x \in X(\Omega)} p_X(x) = 1.$$

Lois discrètes – Exemples

Si $E \subset \mathbb{R}$ est fini, une variable aléatoire X suit la **loi uniforme sur** E si

$$\text{pour tout } x \in E, \quad P(X = x) = \frac{1}{\operatorname{Card} E}.$$

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

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

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

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

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

$$\text{pour } k = 0,\dots,n, \quad P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

$\rightsquigarrow$ si $A_1,\dots,A_n$ sont indépendants et $P(A_1) = \cdots = P(A_n) = p$, la loi de $S_n = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n} =$ « nombre d'événements réalisés » est $\mathcal{B}(n,p)$.

Lois discrètes – Exemples

Retour sur la liste d'exemples :

- Lancer de deux dés, $\Omega = \{1,\ldots,6\}^2 = \{(x_1,x_2)\,|\,x_1,x_2 \in \{1,\ldots,6\}\}$
    - Somme des résultats : $X((x_1,x_2)) = x_1 + x_2$
      (à valeurs dans $\{2,\ldots,12\}$*)*

Lois discrètes – Exemples

Retour sur la liste d'exemples :

- Lancer de deux dés, $\Omega = \{1,\ldots,6\}^2 = \{(x_1,x_2) \mid x_1,x_2 \in \{1,\ldots,6\}\}$
  - Somme des résultats : $X((x_1,x_2)) = x_1 + x_2$
    (à valeurs dans $\{2,\ldots,12\}$)

$$P(X=2) = P(\{(1,1)\}) = \frac{1}{36},$$

$$P(X=3) = P(\{(1,2),(2,1)\}) = \frac{2}{36},$$

$$\ldots$$

$$P(X=6) = P(\{(1,5),(2,4),\ldots,(5,1)\}) = \frac{5}{36}$$

$$P(X=7) = P(\{(1,6),(2,5),\ldots,(6,1)\}) = \frac{6}{36}$$

$$P(X=8) = P(\{(2,6),(3,5),\ldots,(6,2)\}) = \frac{5}{36}$$

$$\ldots$$

$$P(X=11) = P(\{(5,6),(6,5)\}) = \frac{2}{36}$$

$$P(X=12) = P(\{(6,6)\}) = \frac{1}{36}$$

Lois discrètes – Exemples

- On prend successivement les parts d'une galette (coupée en 8)
  - Nombre de parts à prendre jusqu'à avoir la fève : N_A
    (à valeurs dans $\{1, \dots, 8\}$*)*
- Chaque jour, on prend une part d'une galette *différente* (coupée en 8)
  - 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
    (à valeurs dans $\{0, 1, 2, \dots, n\}$*)*

Lois discrètes – Exemples

- On prend successivement les parts d'une galette (coupée en 8)
  - Nombre de parts à prendre jusqu'à avoir la fève : N_A
    (à valeurs dans $\{1,\dots,8\}$*)*
    $\rightsquigarrow$ N_A suit la loi uniforme sur $\{1,\dots,8\}$
- Chaque jour, on prend une part d'une galette *différente* (coupée en 8)
  - 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
    (à valeurs dans $\{0,1,2,\dots,n\}$*)*

Lois discrètes – Exemples

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

$$\begin{aligned} P(N_B = n) &= P(n-1 \text{ parts sans fève, puis une part avec fève}) \\ &= \left(1 - \frac{1}{8}\right)^{n-1} \frac{1}{8}. \qquad \text{(par indépendance)} \end{aligned}$$

    $\rightsquigarrow$ N_B suit la **loi géométrique** de paramètre $p = \frac{1}{8}$.
  - Nombre de fèves obtenues en n jours : S_n
    (à valeurs dans $\{0,1,2,\ldots,n\}$*)*

Lois discrètes – Exemples

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

$$\begin{aligned} P(N_B = n) &= P(n-1 \text{ parts sans fève, puis une part avec fève}) \\ &= \left(1 - \frac{1}{8}\right)^{n-1} \frac{1}{8}. \qquad \text{(par indépendance)} \end{aligned}$$

    $\rightsquigarrow$ N_B suit la **loi géométrique** de paramètre $p = \frac{1}{8}$.
  - Nombre de fèves obtenues en n jours : S_n
    (à valeurs dans $\{0,1,2,\dots,n\}$*)*
    $\rightsquigarrow$ S_n suit la loi binomiale de paramètres $(n, \frac{1}{8})$.

Un exemple non discret

- Placement d'une fève circulaire dans une galette, $\Omega = D(0,r) \subset \mathbb{R}^2$
  - Distance au centre : $R((x,y)) = \sqrt{x^2 + y^2}$
    (à valeurs dans $[0,r]$)

Rappel : On munit Ω de la loi uniforme, $P(A) = \dfrac{\text{aire}(A)}{\text{aire}(D(0,r))} = \dfrac{\text{aire}(A)}{\pi r^2}$

Un exemple non discret

- Placement d'une fève circulaire dans une galette, $\Omega = D(0,r) \subset \mathbb{R}^2$
  - Distance au centre : $R((x,y)) = \sqrt{x^2+y^2}$
    (à valeurs dans $[0,r]$)

Rappel : On munit Ω de la loi uniforme, $P(A) = \dfrac{\text{aire}(A)}{\text{aire}(D(0,r))} = \dfrac{\text{aire}(A)}{\pi r^2}$

• Pour $0 \leq x \leq r$,

$$P(R = x) = \frac{\text{aire(cercle de rayon } x)}{\pi r^2} = 0.$$

$\Rightarrow$ approche précédente inadaptée.

Un exemple non discret

- Placement d'une fève circulaire dans une galette, $\Omega = D(0,r) \subset \mathbb{R}^2$
  - Distance au centre : $R((x,y)) = \sqrt{x^2+y^2}$
    (à valeurs dans $[0,r]$)

Rappel : On munit Ω de la loi uniforme, $P(A) = \dfrac{\text{aire}(A)}{\text{aire}(D(0,r))} = \dfrac{\text{aire}(A)}{\pi r^2}$

• Pour $0 \leq x \leq r$,

$$P(R = x) = \frac{\text{aire(cercle de rayon } x)}{\pi r^2} = 0.$$

$\Rightarrow$ approche précédente inadaptée.

• Pour $0 \leq a \leq b \leq r$,

$$P(a \leq R \leq b) = \frac{\text{aire(couronne)}}{\pi r^2} = \frac{\pi b^2 - \pi a^2}{\pi r^2} = \frac{b^2 - a^2}{r^2} = \int_a^b \frac{2t}{r^2} dt.$$

$\rightsquigarrow$ la fonction $f(t) = \dfrac{2t}{r^2}\mathbf{1}_{[0,r]}(t)$ représente la **densité de probabilité** de R.

Lois continues

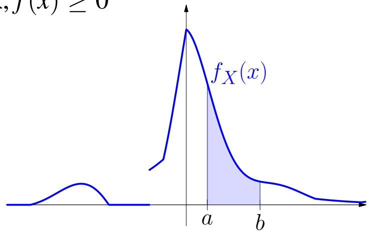
Définition

Une variable aléatoire X est dite ***continue*** *ou* ***à densité*** *s'il existe une fonction $f_X : \mathbb{R} \to \mathbb{R}$ 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

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



Si X a pour densité f_X, pour tous $a \leq b$, $P(a \leq X \leq b) = \displaystyle\int_a^b f_X(x)dx$

Définition

*Une variable aléatoire X est dite **continue** ou **à densité** s'il existe une fonction $f_X : \mathbb{R} \to \mathbb{R}$ 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

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

Remarques

- Si X a une densité alors, pour tout $x \in \mathbb{R}$, $P(X = x) = \int_x^x f_X(t)dt = 0$.
- Si $f_X(x) = 0$ pour tout $x \in B$, alors $P(X \in B) = \int_B f_X(x)dx = 0$.
  $\Rightarrow X$ est à valeurs dans

$$\mathrm{Supp}(f_X) = \{x \in \mathbb{R} \,|\, f_X(x) > 0\}.$$

Densités classiques

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

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

Une variable aléatoire X de loi $\mathcal{U}([a,b])$ est donc à valeurs dans $[a,b]$.

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

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

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

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

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

$\rightsquigarrow$ Utilisée pour modéliser les durées de vie de machine sans vieillissement

Attention

De nombreuses variables ne sont ni discrètes, ni à densité.

Exemple : soit X une variable aléatoire de loi $\mathcal{U}([0,1])$. On définit

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

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

Attention

De nombreuses variables ne sont ni discrètes, ni à densité.

Exemple : soit X une variable aléatoire de loi $\mathcal{U}([0,1])$. 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 $[0,1]$.

Attention

De nombreuses variables ne sont ni discrètes, ni à densité.

Exemple : soit X une variable aléatoire de loi $\mathcal{U}([0,1])$. 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 $P(Y = \dfrac{1}{2}) = P(X \geq \dfrac{1}{2}) = \displaystyle\int_{1/2}^{\infty} \mathbf{1}_{[0,1]}(x)dx = \int_{\frac{1}{2}}^{1} 1dx = \frac{1}{2} > 0$

$\Rightarrow$ Y n'a pas de densité

Attention

De nombreuses variables ne sont ni discrètes, ni à densité.

Exemple : soit X une variable aléatoire de loi $\mathcal{U}([0,1])$. 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 $P(Y = \frac{1}{2}) = P(X \geq \frac{1}{2}) = \int_{1/2}^{\infty} \mathbf{1}_{[0,1]}(x)dx = \int_{\frac{1}{2}}^{1} 1dx = \frac{1}{2} > 0$
$\Rightarrow$ Y n'a pas de densité
• Pour tout $0 \leq x < \frac{1}{2}$, $P(Y = x) = P(X = x) = 0$
$\Rightarrow$ Y n'est pas discrète

Fonction de répartition

But : avoir une façon de représenter et étudier n'importe quelle loi.

Définition

Soit X une variable aléatoire. La ***fonction de répartition de** X* *est la fonction $F_X : \mathbb{R} \to \mathbb{R}$ définie par*

$$\text{pour tout } x \in \mathbb{R}, \qquad F_X(x) = P(X \leq x).$$

Cours 3

–

Mercredi 3 février 2016

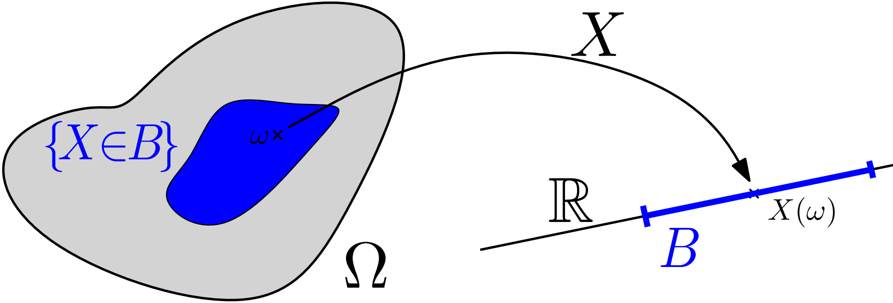
Rappel – Variables aléatoires

Définition

*Une **variable aléatoire** est une application* $X : \Omega \to \mathbb{R}$.
*La **loi** de* X *est la probabilité* P_X *sur* $\mathbb{R}$ *définie par :*

$$\text{pour tout } B \subset \mathbb{R}, \qquad P_X(B) = P(\{\omega \in \Omega \,|\, X(\omega) \in B\}) = P(X \in B).$$

$X(\Omega)$ *(image de* X*) est le **support** de* P_X.
$\leadsto$ P_X *peut aussi être vue comme une probabilité sur* $X(\Omega)$.
On note parfois $X \sim P_X$ *pour indiquer que* X *suit la loi* P_X.



Rappel – Lois discrètes et à densité

Une variable aléatoire X est dite **discrète** si l'ensemble $X(\Omega)$ des valeurs qu'elle prend est *dénombrable*. Alors, pour tout ensemble $B \subset \mathbb{R}$,

$$P_X(B) = P(X \in B) = \sum_{x \in B} P(X = x) = \sum_{x \in B \cap X(\Omega)} P(X = x)$$

$\rightsquigarrow$ pour se donner la loi d'une v.a. discrète, il suffit de se donner les probabilités élémentaires $p_X(x) = P(X = x)$ pour tout $x \in X(\Omega)$, telles que

$$\text{pour tout } x \in X(\Omega),\ p_X(x) \geq 0, \quad \text{et} \quad \sum_{x \in X(\Omega)} p_X(x) = 1.$$

Une variable aléatoire X a pour **densité** la fonction $f_X : \mathbb{R} \to \mathbb{R}$ si, pour tout ensemble $B \subset \mathbb{R}$,

$$P_X(B) = P(X \in B) = \int_B f_X(x)dx = \int_{B \cap \mathrm{Supp}(f_X)} f_X(x)dx.$$

$\rightsquigarrow$ pour se donner la loi d'une v.a. à densité, il suffit de se donner f_X, telle que

$$\text{pour tout } x \in \mathbb{R},\ f_X(x) \geq 0, \quad \text{et} \quad \int_{\mathbb{R}} f_X(x)dx = 1.$$

Attention

De nombreuses variables ne sont ni discrètes, ni à densité.

Exemple : soit X une variable aléatoire de loi $\mathcal{U}([0,1])$. On définit

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

On a $P\left(Y = \frac{1}{2}\right) = \frac{1}{2} > 0$ donc Y n'a pas de densité.
On a $P(Y = x) = 0$ pour tout $x \neq \frac{1}{2}$, donc $\sum_{x\in\mathbb{R}} P(Y = x) \neq 1$: Y n'est pas discrète.

Fonction de répartition

But : avoir une façon unifiée de représenter et étudier n'importe quelle loi.

Définition

Soit X une variable aléatoire. La ***fonction de répartition de*** X *est la fonction $F_X : \mathbb{R} \to \mathbb{R}$ définie par*

$$\text{pour tout } x \in \mathbb{R}, \qquad F_X(x) = P(X \leq x).$$

Fonction de répartition

But : avoir une façon unifiée de représenter et étudier n'importe quelle loi.

Définition

*Soit X une variable aléatoire. La **fonction de répartition de** X est la fonction $F_X : \mathbb{R} \to \mathbb{R}$ définie par*

$$\text{pour tout } x \in \mathbb{R}, \qquad F_X(x) = P(X \leq x).$$

Proposition

a) *La fonction de répartition F_X est une fonction croissante,*

$$\lim_{x \to -\infty} F_X(x) = 0 \quad \text{et} \quad \lim_{x \to +\infty} F_X(x) = 1.$$

b) *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.*

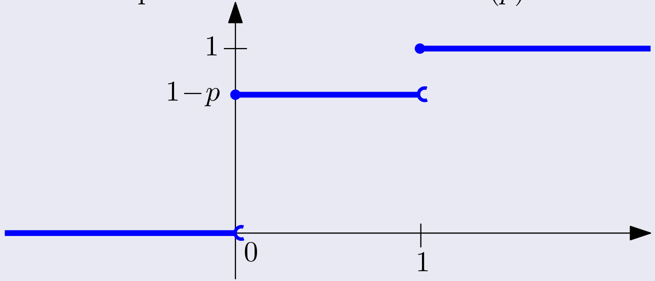
$\rightsquigarrow$ vu b), la fonction de répartition caractérise la loi d'une variable aléatoire.

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 $\mathcal{B}(p)$

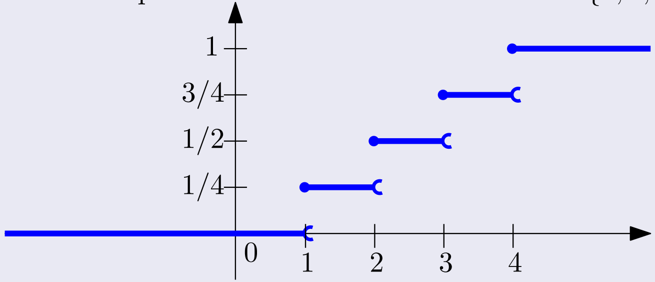


Fonction de répartition – Cas discret

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 unif. sur $\{1, 2, 3, 4\}$



Fonction de répartition – Cas à densité

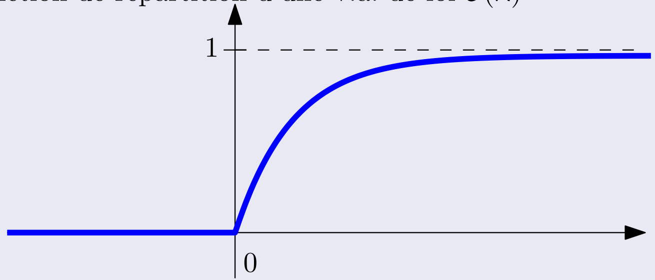
Proposition

Si X est une variable aléatoire de densité f_X, on a

$$\text{pour tout } x \in \mathbb{R}, \qquad F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

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

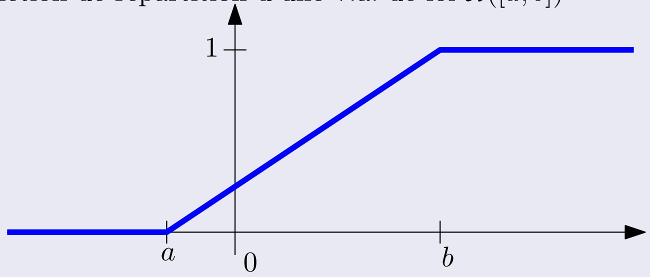
Proposition

Si X est une variable aléatoire de densité f_X, on a

$$\text{pour tout } x \in \mathbb{R}, \qquad F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

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 $\mathcal{U}([a,b])$



Fonction de répartition – Cas à densité

Proposition

Si X est une variable aléatoire de densité f_X, on a

$$\text{pour tout } x \in \mathbb{R}, \qquad F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

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 – Autre exemple

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 $Y = X$)

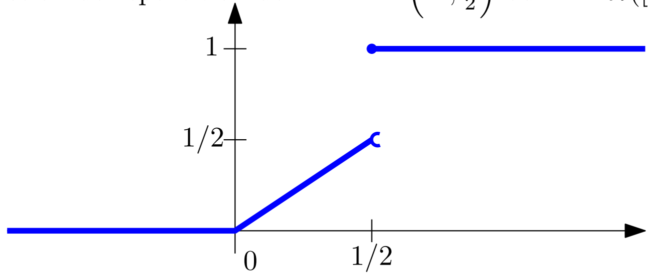
Fonction de répartition – Autre exemple

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 $Y = \varphi(X)$

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

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

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

• Si Y est discrète (par exemple, si X est discrète),
$\to$ Déterminer les valeurs possibles de Y,
$\to$ Calculer chacune de leurs probabilités en se ramenant à X.

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

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

• Si Y est discrète (par exemple, si X est discrète),
$\to$ Déterminer les valeurs possibles de Y,
$\to$ Calculer chacune de leurs probabilités en se ramenant à X.

• Si X a une densité f_X, et φ est monotone (croissante ou décroissante),
$\to$ Déterminer les valeurs possibles de Y,
$\to$ Calculer la **fonction de répartition** de Y,
$\to$ Si F_Y est continue sur $\mathbb{R}$, et dérivable (sauf en quelques points), dériver pour obtenir f_Y.

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

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

• Si Y est discrète (par exemple, si X est discrète),
$\to$ Déterminer les valeurs possibles de Y,
$\to$ Calculer chacune de leurs probabilités en se ramenant à X.

• Si X a une densité f_X, et φ est monotone (croissante ou décroissante),
$\to$ Déterminer les valeurs possibles de Y,
$\to$ Calculer la **fonction de répartition** de Y,
$\to$ 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 vigilent, 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

$$\begin{aligned} P(Y=1) &= P(|X|=1) = P(X=1 \text{ ou } X=-1) = P(X=1)+P(X=-1) \\ &= \frac{1}{3}+\frac{1}{3}=\frac{2}{3}, \end{aligned}$$

et ainsi $P(Y=0) = 1 - P(Y=1) = \frac{1}{3}$, donc Y suit la loi $\mathcal{B}(2/3)$.

Exemples de calculs de loi

Soit X une variable aléatoire de loi $\mathcal{E}(\lambda)$. On pose $Y = \frac{1}{1+X}$.
On a $Y = \varphi(X)$ où $\varphi : x \mapsto \frac{1}{1+x}$. Comme $X \sim \mathcal{E}(\lambda)$, on a $X > 0$ p.s..
φ est strictement décroissante sur $]0, +\infty[$, $\varphi(0) = 1$ et $\lim_{x\to+\infty} \varphi(x) = 0$ donc $\varphi(]0, +\infty[) =]0,1[$. Ainsi, Y est à valeurs dans $]0,1[$.

Alors, $F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ 1 & \text{si } y \geq 1 \end{cases}$ et, pour $0 \leq y \leq 1$,

$$P(Y \leq y) = P\Big(\frac{1}{1+X} \leq y\Big) = P\Big(1 + X \geq \frac{1}{y}\Big) = P\Big(X \geq \frac{1}{y} - 1\Big)$$
$$= 1 - P\Big(X < \frac{1}{y} - 1\Big) = 1 - F_X\Big(\frac{1}{y} - 1\Big) = e^{-\lambda(\frac{1}{y}-1)}.$$

NB. $P(X < x) = P(X \leq x) - P(X = x) = P(X \leq x)$ car X a une densité.
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

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

(avec valeurs quelconques en 0 et 1)

Espérance – Motivation

Dans un jeu de hasard A, on peut

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

Dans un autre jeu de hasard B, on peut

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

À quel jeu devrait-on jouer ?

Espérance – Motivation

Dans un jeu de hasard A, on peut

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

Dans un autre jeu de hasard B, on peut

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

À quel jeu devrait-on jouer ? *Évidemment B*. Il suffit de comparer les probabilités de gain.

Espérance – Motivation

Dans un jeu de hasard A, on peut

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

Dans un autre jeu de hasard B, on peut

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

À quel jeu devrait-on jouer ?

Espérance – Motivation

Dans un jeu de hasard A, on peut

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

Dans un autre jeu de hasard B, on peut

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

À quel jeu devrait-on jouer ? *Moins clair...* Ici il faut prendre en compte les montants, pas seulement les probabilités.
Si on joue un grand nombre de fois, la quantité importante est le gain moyen, ou **espérance** de gain $\Rightarrow$ on choisit A
Si on ne joue qu'un petit nombre de fois, cela reste une bonne indication, mais la décision dépend du risque que l'on est prêt à prendre.

(voir "Paradoxe de Saint-Petersbourg" sur Wikipedia)

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, $E[X] = \sum_{x \in X(\Omega)} x P(X = x)$.
Si X est continue, de densité f_X, $E[X] = \int_{\mathbb{R}} x f_X(x) dx$.
***Attention.** L'espérance n'est pas toujours définie. Il faut pour cela que la série ou l'intégrale ci-dessus converge absolument.*

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)
$\rightarrow$ prix d'assurances, d'actifs financiers,...

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, $E[X] = \int_{\mathbb{R}} x f_X(x) dx$.
***Attention.** L'espérance n'est pas toujours définie. Il faut pour cela que la série ou l'intégrale ci-dessus converge absolument.*

Remarque : Dans ce cours, on se contentera des cas discret et à densité.
Si X n'est ni discrète ni à densité, on pourrait utiliser F_X et définir

$$E[X] = \int_0^{\infty} (1 - F_X(x))dx - \int_{-\infty}^{0} F_X(x)dx$$

en vérifiant que dans les cas discret et à densité, cela redonne la définition.
Une meilleure approche est en fait de définir une intégrale généralisée pour pouvoir avoir $E[X] = \int_\Omega X(\omega)dP(\omega) = \int_{\mathbb{R}} x\, dP_X(x)$

Espérance – Exemples discrets

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

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

$$E[X] = 1\cdot p + 0\cdot(1-p) = p.$$

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

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

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

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

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

$$E[X] = \sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}p = p\frac{1}{(1-(1-p))^2} = \frac{p}{p^2} = \frac{1}{p}$$

car $\displaystyle\sum_{k=1}^{\infty} kx^{k-1} = \frac{d}{dx}\Big(\sum_{k=0}^{\infty} x^k\Big) = \frac{d}{dx}\Big(\frac{1}{1-x}\Big) = \frac{1}{(1-x)^2}$ pour $-1<x<1$.

Espérance – Exemples à densité

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

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

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

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

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

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x \cdot \lambda e^{-\lambda x}\mathbf{1}_{]0,+\infty[}(x)dx = \int_0^\infty x\lambda e^{-\lambda x}dx \\ &= [-xe^{-\lambda x}]_{x=0}^{\infty} + \int_0^\infty e^{-\lambda x}dx = \frac{1}{\lambda}. \end{aligned}$$

Espérance – Propriétés

Propriétés

(i) *Si* X *est constante, égale à* $c \in \mathbb{R}$ *(pour tout* $\omega \in \Omega$*,* $X(\omega) = c$*), alors* $E[X] = E[c] = c$*.*

(ii) *Pour tout événement* $A \subset \Omega$*,* $E[\mathbf{1}_A] = P(A)$*.*

(iii) *L'espérance est linéaire : pour toutes variables aléatoires* X *et* Y*, et tout réel* a*,*

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

(iv) *L'espérance est croissante : si* $X \leq Y$ *p.s., alors* $E[X] \leq E[Y]$*.*

Espérance – Propriétés

Propriétés

(i) *Si* X *est constante, égale à* $c \in \mathbb{R}$ *(pour tout* $\omega \in \Omega$*,* $X(\omega) = c$*), alors* $E[X] = E[c] = c$*.*

(ii) *Pour tout événement* $A \subset \Omega$*,* $E[\mathbf{1}_A] = P(A)$*.*

(iii) *L'espérance est linéaire : pour toutes variables aléatoires* X *et* Y*, et tout réel* a*,*

$$E[aX] = aE[X] \qquad \textit{et} \qquad E[X+Y] = E[X] + E[Y].$$

(iv) *L'espérance est croissante : si* $X \leq Y$ *p.s., alors* $E[X] \leq E[Y]$*.*

Si $A_1, \ldots, A_n$ sont des événements indépendants et $P(A_1) = \cdots = P(A_n) = p$, on a vu que $S_n = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}$ suit la loi binomiale $\mathcal{B}(n,p)$:

$$\text{pour } k = 0, \ldots, n, \quad P(S_n = k) = C_n^k p^k (1-p)^{n-k}.$$

Par linéarité, $E[S_n] = E[\mathbf{1}_{A_1}] + \cdots + E[\mathbf{1}_{A_n}] = P(A_1) + \cdots + P(A_n) = np$.

Espérance de $\varphi(X)$

Proposition

Soit X une variable aléatoire, et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction.

- *Si X est discrète, alors*

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

- *Si X est continue, alors*

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

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

Si X suit la loi uniforme sur $\{1,2,3\}$, $E\left[\frac{X}{1+X}\right] = \frac{1}{1+1}\frac{1}{3} + \frac{2}{1+2}\frac{1}{3} + \frac{3}{1+3}\frac{1}{3} = \frac{23}{36}$
Si X suit la loi uniforme sur $[0,1]$,
$E\left[\frac{1}{1+X}\right] = \int_{\mathbb{R}} \frac{1}{1+x} \mathbf{1}_{[0,1]}(x) dx = \int_0^1 \frac{1}{1+x} dx = [\ln(1+x)]_{x=0}^1 = \ln 2$

Variance

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

Définition

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

$$\mathrm{Var}(X) = E\left[\left(X - E[X]\right)^2\right] \geq 0.$$

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

Attention. *La variance n'est pas toujours définie. Il faut que l'espérance $E[X]$ soit définie* ***et*** *que l'espérance ci-dessus converge.*
$\rightarrow$ *Ceci revient à demander à ce que $E[X^2]$ converge.*

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 – Propriétés

Propriétés

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

1. $\mathrm{Var}(X) = E[X^2] - E[X]^2$
2. $\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$
3. $\mathrm{Var}(X+a) = \mathrm{Var}(X)$
4. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)$, *où la **covariance** est définie par*

$$\mathrm{Cov}(X,Y) = E\Big[\big(X - E[X]\big)\big(Y - E[Y]\big)\Big] = E[XY] - E[X]E[Y].$$

Pour toute variable aléatoire X possédant une variance, la variable aléatoire $Y = \dfrac{X - E[X]}{\sigma(X)}$ est centrée ($E[Y] = 0$) et réduite ($\mathrm{Var}(Y) = 1$).

Moments

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

$$m_r(X) = E[X^r],$$

et le **moment centré d'ordre** r :

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

Variance – Exemples discrets

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

$E[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$, donc $\mathrm{Var}(X) = p - p^2 = p(1-p)$

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

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

Indication : dériver deux fois $\displaystyle\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ pour obtenir

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

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

Variance – Exemples à densité

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

$$\mathrm{Var}(X) = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

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

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = [x^2 e^{-\lambda x}]_{x=0}^\infty + \int_0^\infty 2x e^{-\lambda x} dx = \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2}$$

d'où

$$\mathrm{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

Cours 4 & 5

–

Mercredi 10 février 2016

Rappel – Espérance, variance

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, $E[X] = \int_{\mathbb{R}} x f_X(x) dx$.
***Attention.** L'espérance n'est pas toujours définie. Il faut pour cela que la série ou l'intégrale ci-dessus converge absolument.*

Intérêt, interprétation :
• $E[X]$ est la moyenne des valeurs de X observées en "répétant l'expérience" un grand nombre de fois (loi des grands nombres)
• $E[X]$ 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)
$\rightarrow$ prix d'assurances, d'actifs financiers,...

Rappel – Espérance, variance

Définition

*L'**espérance** d'une variable aléatoire X, notée $E[X]$, est la moyenne de ses valeurs, pondérées par leurs probabilités. Pour $\varphi : X(\Omega) \to \mathbb{R}$,*

Si X est discrète, $E[\varphi(X)] = \sum_{x \in X(\Omega)} \varphi(x) P(X = x)$.

Si X est continue, de densité f_X, $E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f_X(x) dx$.

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

Intérêt, interprétation :

- $E[X]$ est la moyenne des valeurs de X observées en "répétant l'expérience" un grand nombre de fois (loi des grands nombres)
- $E[X]$ 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)

$\to$ prix d'assurances, d'actifs financiers,...

Rappel – Espérance, variance

Définition

*L'**espérance** d'une variable aléatoire X, notée $E[X]$, est la moyenne de ses valeurs, pondérées par leurs probabilités. Pour $\varphi : X(\Omega) \to \mathbb{R}$,*

Si X est discrète, $E[\varphi(X)] = \sum_{x \in X(\Omega)} \varphi(x) P(X = x)$.

Si X est continue, de densité f_X, $E[\varphi(X)] = \int_{\mathbb{R}} \varphi(x) f_X(x) dx$.

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

Définition

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

$$\mathrm{Var}(X) = E\left[\left(X - E[X]\right)^2\right] \geq 0.$$

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

On a aussi $\mathrm{Var}(X) = E[X^2] - E[X]^2$.

Inégalités

Proposition (Inégalité de Markov)

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

$$P(|X| \geq a) \leq \frac{E\big[|X|\big]}{a}.$$

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

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

Preuve (si X a pour densité f_X)

$$\begin{aligned}
E[|X|^r] = \int_{\mathbb{R}} |x|^r f_X(x)dx &= \int_{]-a,a[} |x|^r f_X(x)dx + \int_{\mathbb{R}\setminus]-a,a[} |x|^r f_X(x)dx \\
&\geq 0 + \int_{\mathbb{R}\setminus]-a,a[} a^r f_X(x)dx \\
&= a^r P\big(X \in \mathbb{R}\setminus]-a,a[\big) = a^r P(|X| \geq a).
\end{aligned}$$

Inégalités

Proposition (Inégalité de Bienaymé-Tchebychev)

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

$$P\Big(\big|X - E[X]\big| \geq a\Big) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

Preuve : Appliquer l'inégalité de Markov à $r = 2$ et $X - E[X]$.

Autre écriture

Pour tout $A > 0$,

$$P\Big(E[X] - A\sigma(X) \leq X \leq E[X] + A\sigma(X)\Big) \geq 1 - \frac{1}{A^2}.$$

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

Parenthèse : loi de Poisson $\mathcal{P}(\lambda)$

Soit $\lambda > 0$. Une variable aléatoire X suit la **loi de Poisson de paramètre** λ (notée $\mathcal{P}(\lambda)$) si X est à valeurs dans $\mathbb{N} = \{0,1,2,\ldots\}$ et

$$\text{pour tout } k \in \mathbb{N}, \qquad P(X = k) = e^{-k}\lambda^k k!.$$

On a

$$E[X] = \lambda \qquad \text{et} \qquad \mathrm{Var}(X) = \lambda\,.$$

C'est la loi limite de la loi binomiale $\mathcal{B}(n,p)$, avec $np \sim \lambda$:

Proposition

Si, pour tout n, S_n suit la loi $\mathcal{B}(n,p_n)$, et $np_n \underset{n\to\infty}{\longrightarrow} \lambda$, alors

$$\textit{pour tout } k \in \mathbb{N}, \qquad P(S_n = k) \underset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}.$$

Dans la pratique, on peut approcher la loi $\mathcal{B}(n,p)$ par la loi $\mathcal{P}(np)$ lorsque $n \geq 50$ et $p \leq 0{,}1$ (erreur inférieure à 5 % dans les calculs de probabilités).

Soit $\lambda > 0$. Une variable aléatoire X suit la **loi de Poisson de paramètre** λ (notée $\mathcal{P}(\lambda)$) si X est à valeurs dans $\mathbb{N} = \{0,1,2,\dots\}$ et

$$\text{pour tout } k \in \mathbb{N}, \qquad P(X = k) = e^{-k}\lambda^k k!.$$

Proposition

Si, pour tout n, S_n suit la loi $\mathcal{B}(n,p_n)$, et $np_n \underset{n\to\infty}{\longrightarrow} \lambda$, alors

$$\textit{pour tout } k \in \mathbb{N}, \qquad P(S_n = k) \underset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}.$$

Dans la pratique, on peut approcher la loi $\mathcal{B}(n,p)$ par la loi $\mathcal{P}(np)$ lorsque $n \geq 50$ et $p \leq 0{,}1$ (erreur inférieure à 5 % dans les calculs de probabilités).
Ex. Une usine produit 500 pièces par jour, dont 1 % sont défaillantes. Le nombre N de pièces défaillantes suit la loi $\mathcal{B}(n,p)$ avec $n = 500$, $p = 0{,}01$.
Le nombre moyen d'erreurs est $\lambda = E[N] = np = 5$.
Alors, N suit approx. la loi $\mathcal{P}(5)$, donc $P(N \leq 7) \simeq e^{-5}\sum_{k=0}^{7}\frac{5^k}{k!} \simeq 0{,}866$.
En vérité, $P(N \leq 7) = \sum_{k=0}^{7}\binom{500}{k}0{,}01^k \cdot 0{,}99^{500-k} \simeq 0{,}868$.

Indépendance de variables aléatoires

Définition

*Des variables aléatoires $X_1,\dots,X_n$ sont **indépendantes** si, pour tous $B_1,\dots,B_n \subset \mathbb{R}$,*

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

où les virgules se lisent « et » :

$$P(X_1 \in B_1,\dots,X_n \in B_n) = P(\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\})$$

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 $B,C \subset \mathbb{R}$,

$$P(X \in B, Y \in C) = P(X \in B)P(Y \in C).$$

$\rightsquigarrow$ Connaître X ne renseigne pas sur Y. Notion intuitive d'« indépendance ».
Exemple : tirages de dés,...

Indépendance – Retour sur un exemple

On considère deux tirages de dés : espace de probabilité $\Omega = \{1, \dots, 6\}^2$, avec la probabilité P uniforme.
On note X_1, X_2 les résultats des dés : pour tout tirage $(k,l) \in \Omega$,

$$X_1((k,l)) = k \quad \text{et} \quad X_2((k,l)) = l.$$

Alors, pour $A,B \subset \{1, \dots, 6\}$,

$$\begin{aligned} P(X_1 \in A, X_2 \in B) &= P(\{(k,l) \in \Omega \,|\, k \in A, l \in B\}) \\ &= P(A \times B) = \frac{\mathrm{Card}(A \times B)}{\mathrm{Card}(\Omega)} \\ &= \frac{\mathrm{Card}\, A}{6} \frac{\mathrm{Card}\, B}{6} = P(X_1 \in A) P(X_2 \in B), \end{aligned}$$

car X_1 et X_2 suivent la loi uniforme sur $\{1, \dots, 6\}$. Donc X_1 et X_2 sont indépendantes.
$\rightsquigarrow$ On a ainsi déjà utilisé des v.a. indépendantes sans le dire.
(cf. aussi le paradoxe des anniversaires)

Indépendance – Propriétés

Proposition

1. *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 $f_1,\dots,f_n$.*
2. ***« Indépendance par paquets ».** Si $X_1,\dots,X_n$ sont indépendantes alors, par exemple, les variables aléatoires $f_{1,2}(X_1,X_2), f_3(X_3), f_{4,5,6}(X_4,X_5,X_6),\dots$ sont indépendantes : les fonctions de « paquets disjoints » de variables sont indépendantes.*
3. *Si des événements $A_1,\dots,A_n$ sont indépendants alors leurs fonctions indicatrices $\mathbf{1}_{A_1},\dots,\mathbf{1}_{A_n}$ sont des variables aléatoires indépendantes ; et réciproquement.*

Par 2), si X,Y,Z,T sont indépendantes,

$$X\sqrt{|Z|},\ Y^2 \text{ et } \tfrac{1}{T} \text{ sont indépendantes ;}$$

et de même,

$$X+Y^2 \text{ et } Z(1-\tfrac{T}{Z}) \text{ sont indépendantes.}$$

Indépendance et espérance

Proposition

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

1. *si leurs espérances sont bien définies,*

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n]$$

2. *si leurs variances sont bien définies, alors on a* $\mathrm{Cov}(X_i, X_j) = 0$ *pour tous* $i \neq j$*, d'où*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

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

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

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

Indépendance et espérance

Proposition

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

1. *si leurs espérances sont bien définies,*

$$E[X_1 \cdots X_n] = E[X_1] \cdots E[X_n]$$

2. *si leurs variances sont bien définies, alors on a $\mathrm{Cov}(X_i, X_j) = 0$ pour tous $i \neq j$, d'où*

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n).$$

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

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

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

Application : Variance de la loi binomiale. Si $A_1, \dots, A_n$ sont indépendants et $P(A_1) = \cdots = P(A_n)$, alors $S_n = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}$ suit la loi $\mathcal{B}(n,p)$. D'où, comme $\mathbf{1}_{A_1}, \dots, \mathbf{1}_{A_n}$ sont indépendantes,

$$\mathrm{Var}(S_n) = \mathrm{Var}(\mathbf{1}_{A_1}) + \cdots + \mathrm{Var}(\mathbf{1}_{A_n}) = n\,\mathrm{Var}(\mathbf{1}_{A_1}) = np(1-p).$$

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 $\overline{X}_n$, appelée **moyenne empirique**, par*

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

On a :

$$\text{pour tout } \varepsilon > 0, \qquad P\Big(m - \varepsilon \leq \overline{X}_n \leq m + \varepsilon\Big) \underset{n\to\infty}{\longrightarrow} 1.$$

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* $\overline{X}_n$, *appelée* ***moyenne empirique****, par*

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

On a :

$$\textit{pour tout } \varepsilon > 0, \qquad P\Big(m - \varepsilon \leq \overline{X}_n \leq m + \varepsilon\Big) \underset{n\to\infty}{\longrightarrow} 1.$$

NB. Si $(A_n)_{n\geq 1}$ 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 $X_i = \mathbf{1}_{A_i}$, on a

$$\overline{X}_n = \frac{\mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_n}}{n} = \frac{\text{nombre d'événements réalisés parmi } A_1, \ldots, A_n}{n}$$

donc $\overline{X}_n$ est la **fréquence de réalisation** des événements $A_1, \ldots, A_n$.

Simulation stochastique et méthode de Monte-Carlo

Principe de la simulation aléatoire (ou “stochastique”)

La “fonction” `rand()` de Scilab (Matlab) renvoie une suite d'observations de variables aléatoires indépendantes et de loi uniforme sur $[0,1]$.

$\rightsquigarrow$ Il existe une suite $U_1, U_2, \ldots$ de variables aléatoires indépendantes, et de loi uniforme sur $[0,1]$, et un élément $\omega \in \Omega$ tel que la fonction `rand()` renvoie d'abord $U_1(\omega)$, puis $U_2(\omega), \ldots$

• On peut dire que la valeur de ω correspond à la **graine** du générateur aléatoire : c'est une valeur (un entier) qui détermine la suite des tirages. On peut la choisir au hasard, par exemple à l'aide de l'horloge.
Dans Scilab, `rand("seed",n)` donne à la graine la valeur `n`.

• *En réalité,* $U_1, U_2, \ldots$ *ne sont pas vraiment indépendantes et de loi uniforme sur* $[0,1]$, *mais se comportent “presque” comme si elles l'étaient. On parle de nombres* ***pseudo-aléatoires***.

• Si on souhaite des variables qui suivent d'autres lois, on peut les construire à partir de $U_1, U_2, \ldots$ ou, pour les lois usuelles, utiliser une autre fonction du logiciel (pour Scilab, `grand`)

• Par la **loi des grands nombres**, on peut calculer des valeurs approchées de probabilités ou d'espérances : **méthode de Monte-Carlo**.

Parenthèse : 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

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

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)
- ou une approximation : $P(Z > x) = 1 - \Phi(x) \underset{x \to \infty}{\sim} \dfrac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}$

(avec une erreur relative inférieure à 0,2 si $x > 1{,}9$)
Si $X \sim \mathcal{N}(m,\sigma^2)$, on pose $Z = \frac{X-m}{\sigma}$ pour se ramener à $\mathcal{N}(0,1)$.

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* $a_1,\dots,a_n \in \mathbb{R}$*,*

$$X = a_1X_1 + \cdots + a_nX_n \sim \mathcal{N}(M,\Sigma^2),$$

où

$$M = E[X] = \sum_{i=1}^{n} a_i m_i \qquad \textit{et} \qquad \Sigma^2 = \mathrm{Var}(X) = \sum_{i=1}^{n} a_i^2\sigma_i^2.$$

Les lois normales interviendront en statistique (pour étudier la marge d'erreur dans la loi des grands nombres).

Plan du cours

Loi du couple

Définition

*Soit X,Y deux variables aléatoires. La **loi du couple** (X,Y) est la probabilité $P_{(X,Y)}$ sur $\mathbb{R}^2$ qui vérifie :*

$$\text{pour tous } A,B \subset \mathbb{R}, \qquad P_{(X,Y)}(A \times B) = P(X \in A,\ Y \in B).$$

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 :

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

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).

Exemple

On choisit au hasard (uniformément) un étudiant entré à l'université en 2012. On note

- $S \in \{H,F\}$ son sexe
- $D \in \{$bio-santé, droit, lettres, sciences, sport, sciences éco $\}$ 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 %

Exemple

On choisit au hasard (uniformément) un étudiant entré à l'université en 2012. On note

- $S \in \{H,F\}$ son sexe
- $D \in$ {bio-santé, droit, lettres, sciences, sport, sciences éco } 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	**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 %

$\rightarrow$ 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 : $P(D = \text{bio}, S = \text{H}) = 0{,}06$ et
$P(D = \text{bio})P(S = \text{H}) = 0{,}2 \cdot 0{,}43 = 0{,}086 \neq 0{,}06$

Cas de deux variables discrètes

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

$$p_{(X,Y)}(x,y) = P(X = x,\ Y = y) \quad \text{pour tous } x \in X(\Omega),\ y \in Y(\Omega).$$

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 $x \in X(\Omega)$,

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

pour tout $y \in Y(\Omega)$,

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

NB. 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 où $P_{(X,Y)}$ a une densité

On dit que le couple (X,Y) a une densité s'il y a une fonction $f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$ telle que

$$\text{pour tout } D \subset \mathbb{R}^2, \qquad P_{(X,Y)}(D) = \iint_D f_{(X,Y)}(x,y)dx\,dy.$$

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

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

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

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

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

1. *Si (X,Y) a pour densité $f_{(X,Y)}$, alors X et Y ont des densités f_X et f_Y données par*

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

2. *Si X et Y ont des densités f_X et f_Y et sont indépendantes, alors (X,Y) a pour densité*

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

   Réciproquement, si $f_{(X,Y)}(x,y) = f(x)g(y)$ pour deux fonctions f et g, alors X et Y sont indépendantes, et les densités de X et Y sont proportionnelles à f et g.

Cours 6

–

Mercredi 17 février 2016

Loi d'un couple de variables aléatoires

Définition

Soit X,Y *deux variables aléatoires. La* ***loi du couple*** (X,Y) *est la probabilité* $P_{(X,Y)}$ *sur* $\mathbb{R}^2$ *qui vérifie :*

$$\text{pour tous } A,B \subset \mathbb{R}, \qquad P_{(X,Y)}(A \times B) = P(X \in A,\, Y \in B).$$

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* $P_{(X,Y)}$*.*

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).

Loi d'un couple – cas discret et à densité

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

$$P(X = x,\, Y = y) \quad \text{pour tous } x \in X(\Omega),\, y \in Y(\Omega).$$

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

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

On dit que le couple (X,Y) a une **densité** s'il existe $f_{(X,Y)} : \mathbb{R}^2 \to \mathbb{R}$ telle que

$$P(X \in A,\, Y \in B) = \int_A \int_B f_{(X,Y)}(x,y)\,dx\,dy \qquad \text{pour tous } A,B \subset \mathbb{R}^2.$$

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

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

Exemple discret

On lance 2 dés à 4 faces, dont on note X et Y les résultats, entre 1 et 4.
$\leadsto$ X et Y sont indépendantes, de loi uniforme sur $\{1,\dots,4\}$.
On définit $Z = |X - Y|$
$\leadsto$ Z est à valeurs dans $\{0,1,2,3\}$ et la loi de (X,Z) est donnée par :

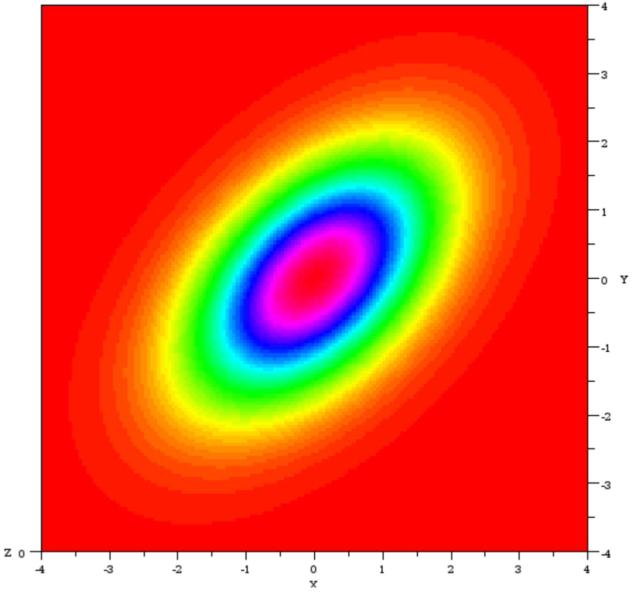
$Z \backslash X$	1	2	3	4
0	$1/16$	$1/16$	$1/16$	$1/16$
1	$1/16$	$1/8$	$1/8$	$1/16$
2	$1/16$	$1/16$	$1/16$	$1/16$
3	$1/16$	0	0	$1/16$

Exemple discret

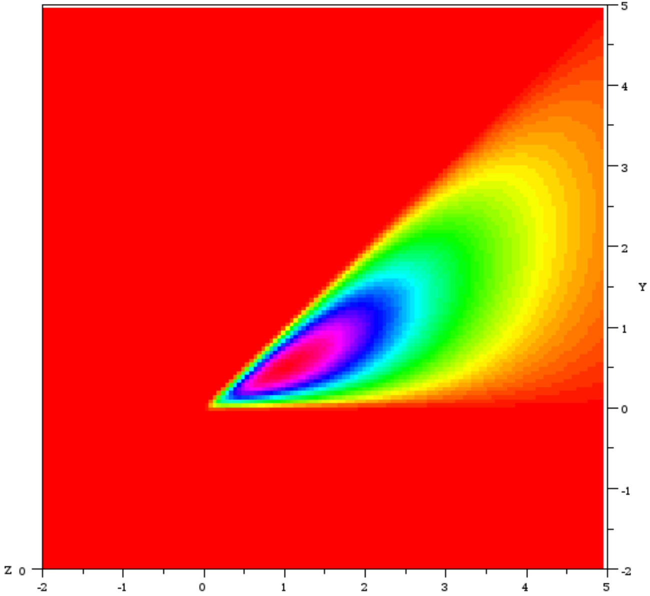
On lance 2 dés à 4 faces, dont on note X et Y les résultats, entre 1 et 4.
$\rightsquigarrow$ X et Y sont indépendantes, de loi uniforme sur $\{1,\ldots,4\}$.
On définit $Z = |X - Y|$
$\rightsquigarrow$ Z est à valeurs dans $\{0,1,2,3\}$ et la loi de (X,Z) est donnée par :

Z \ 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

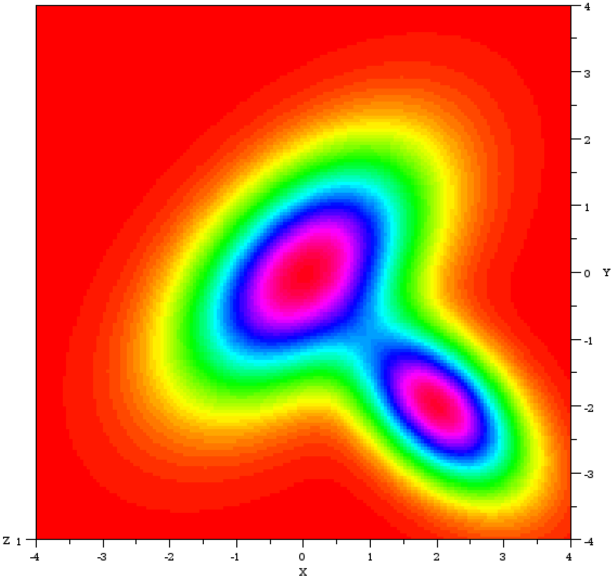
Densité sur $\mathbb{R}^2$, représentation par couleur (rouge = 0)



Densité sur $\mathbb{R}^2$, représentation par couleur (rouge = 0)



Densité sur $\mathbb{R}^2$, représentation par couleur (rouge = 0)



Exemple à densité

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

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

Couple de variables aléatoires indépendantes

Si X et Y sont indépendantes, la loi du couple (X,Y) se déduit des lois de X et de Y, car

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

Cas discret

Si X et Y sont discrètes, X et Y sont indépendantes si et seulement si

$$P(X = x,\, Y = y) = P(X = x)P(Y = y)$$

pour toutes les valeurs x,y possibles.

Cas à densité

Si X et Y ont pour densité f_X et f_Y, X et Y sont indépendantes si et seulement si (X,Y) a pour densité la fonction $f_{(X,Y)}$ telle que

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

pour tous $x,y \in \mathbb{R}$.

Calculs d'espérances

Avec la loi du couple (X,Y), on calcule l'espérance de fonctions de X et Y :

Proposition

Soit $\varphi : \mathbb{R}^2 \to \mathbb{R}$ une fonction.

- *Si X et Y sont discrètes, alors*

$$E[\varphi(X,Y)] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} \varphi(x,y) P(X = x,\ Y = y).$$

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

$$E[\varphi(X,Y)] = \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x,y) f_{(X,Y)}(x,y)\, dx\, dy.$$

(À condition que les séries et les intégrales soient bien définies)

Rappel : si X,Y sont indépendantes,

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

Exemple de calcul

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

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

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

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

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

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

Cours 7

–

Mercredi 9 mars 2016

Plan du cours

Les **statistiques descriptives** visent à **décrire** un ensemble, en général important, de données, c'est-à-dire à en résumer certaines particularités,
$\rightarrow$ sous la forme de représentations graphiques
$\rightarrow$ à l'aide de grandeurs numériques (moyenne,...).

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

NB. 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'une 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 $(x_i,y_i,\dots)$ sont les **observations** des variables associées au i-ième individu.

Une variable peut avoir divers types :

$\rightarrow$ Variable **quantitative**, à valeurs numériques :

- $\rightarrow$ **discrète** s'il y a un nombre fini de valeurs possibles
- $\rightarrow$ **continue** sinon.

$\rightarrow$ Variable **qualitative**, à valeurs non numériques (catégories). Les valeurs possibles sont les **modalités** de la variable.

NB. Une variable quantitative discrète peut se voir comme une variable qualitative.

Variable qualitative – Représentations graphiques

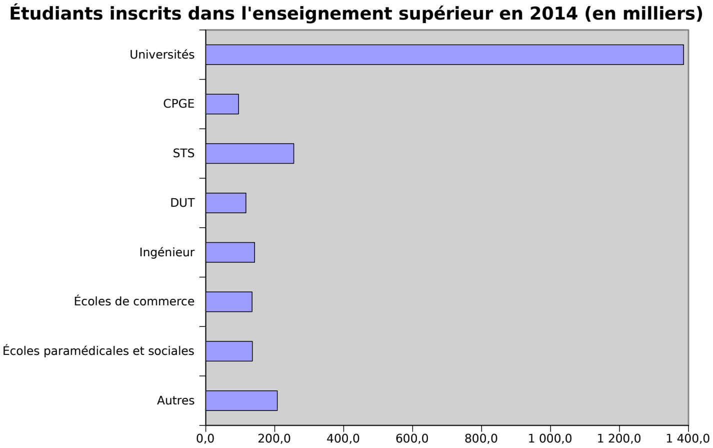
Pour une variable qualitative (ou quantitative discrète),

- 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 qualitative

Population : Étudiants inscrits dans l'enseignement supérieur en 2014
Variable : type d'établissement d'inscription
Données : (longue) liste des établissements où est inscrit chaque étudiant
Effectifs : en milliers,

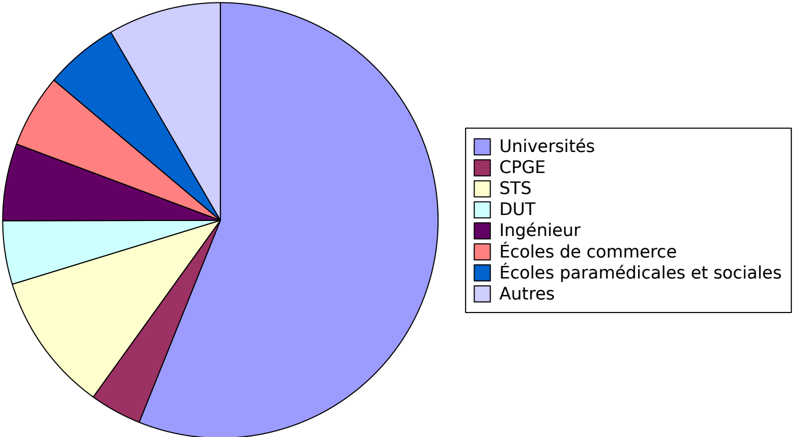
Universités	CPGE	STS	DUT	Ingénieur	Commerce	Paramédic.	Autres
1385,8	95,0	255,2	116,4	141,6	134,3	135,1	207,4



Variable qualitative

Population : Étudiants inscrits dans l'enseignement supérieur en 2014
Variable : type d'établissement d'inscription
Données : (longue) liste des établissements où est inscrit chaque étudiant
Effectifs et fréquences : en milliers,

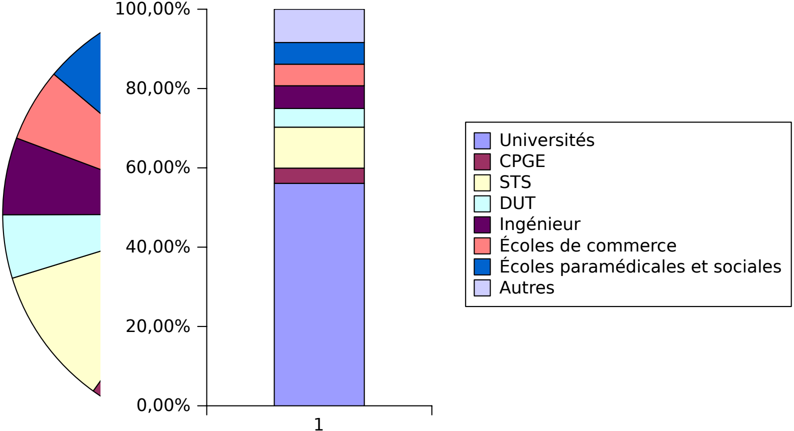
Universités	CPGE	STS	DUT	Ingénieur	Commerce	Paramédic.	Autres
1385,8	95,0	255,2	116,4	141,6	134,3	135,1	207,4
56,1 %	3,8 %	10,3 %	4,7 %	5,7 %	5,4 %	5,5 %	8,4 %



Variable qualitative

Population : Étudiants inscrits dans l'enseignement supérieur en 2014
Variable : type d'établissement d'inscription
Données : (longue) liste des établissements où est inscrit chaque étudiant
Effectifs et fréquences : en milliers,

Universités	CPGE	STS	DUT	Ingénieur	Commerce	Paramédic.	Autres
1385,8	95,0	255,2	116,4	141,6	134,3	135,1	207,4
56,1 %	3,8 %	10,3 %	4,7 %	5,7 %	5,4 %	5,5 %	8,4 %



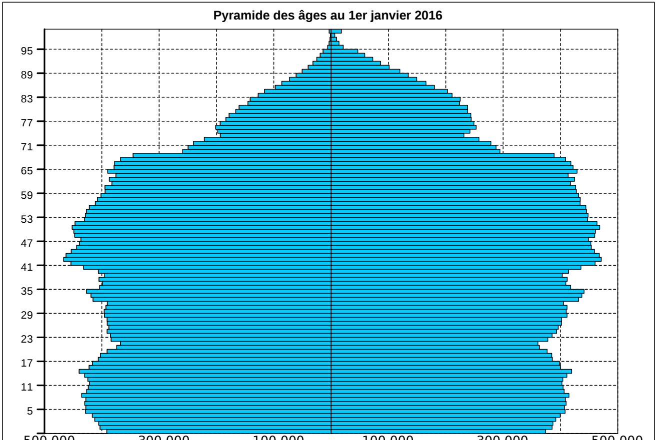
Variable quantitative discrète

Population : personnes de nationalité française, en 2016

Variables : âge *(quantitative discrète)* et sexe *(qualitative)*

Effectifs :

Année de naissance	Âge révolu	Nombre d'hommes	Nombre de femmes	Total
2015	0	391 371	374 179	765 550
2014	1	403 204	385 442	788 646
2013	2	405 502	386 831	792 333
2012	3	412 383	391 853	804 236
2011	4	416 626	399 632	816 258
...				



Variable quantitative discrète

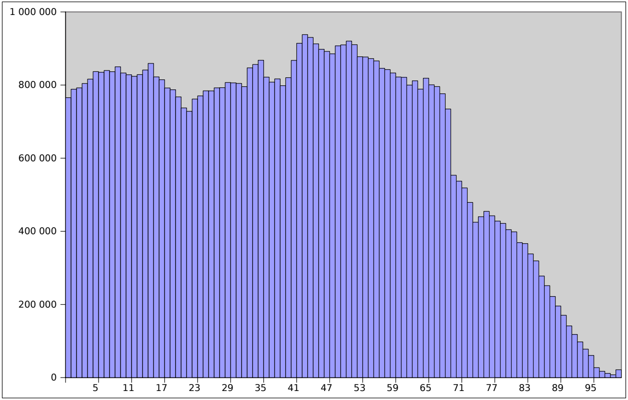
Population : personnes de nationalité française, en 2016

Variables : âge *(quantitative discrète)* et sexe *(qualitative)*

Effectifs :

Année de naissance	Âge révolu	Nombre d'hommes	Nombre de femmes	Total
2015	0	391 371	374 179	765 550
2014	1	403 204	385 442	788 646
2013	2	405 502	386 831	792 333
2012	3	412 383	391 853	804 236
2011	4	416 626	399 632	816 258
...				

Nombre de français, par âge



Variable quantitative discrète

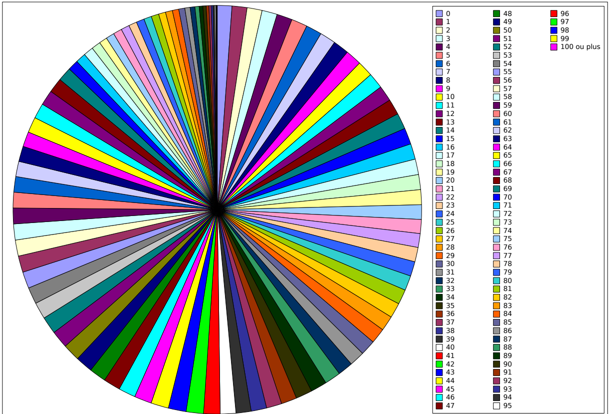
Population : personnes de nationalité française, en 2016

Variables : âge *(quantitative discrète)* et sexe *(qualitative)*

Effectifs :

Année de naissance	Âge révolu	Nombre d'hommes	Nombre de femmes	Total
2015	0	391 371	374 179	765 550
2014	1	403 204	385 442	788 646
2013	2	405 502	386 831	792 333
2012	3	412 383	391 853	804 236
2011	4	416 626	399 632	816 258
...				

Diagramme circulaire ? Illisible : trop de catégories ⇝ regrouper.



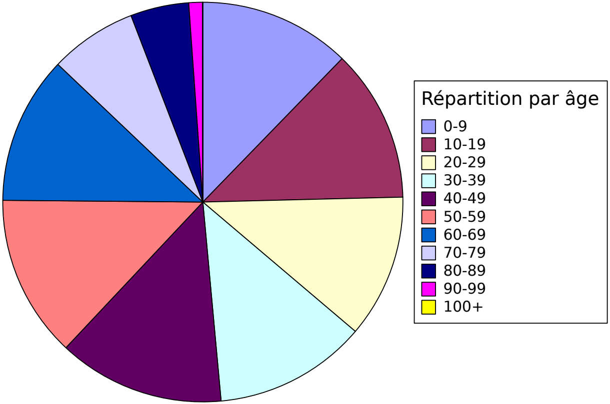
Variable quantitative discrète

Population : personnes de nationalité française, en 2016

Variables : âge *(quantitative discrète)* et sexe *(qualitative)*

Effectifs :

Année de naissance	Âge révolu	Nombre d'hommes	Nombre de femmes	Total
2015	0	391 371	374 179	765 550
2014	1	403 204	385 442	788 646
2013	2	405 502	386 831	792 333
2012	3	412 383	391 853	804 236
2011	4	416 626	399 632	816 258
...				

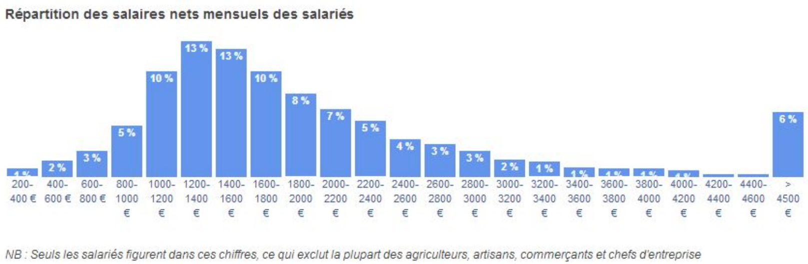


Variable quantitative – Représentation graphique

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 classifie les individus selon l'intervalle qui contient leur valeur

$\rightarrow$ on approche la variable continue par une variable *discrète*.

La représentation des effectifs (ou fréquences) est un **histogramme**.



Attention, le choix des intervalles est essentiel et ne suit pas une règle générale.

Répartition des observations

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

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

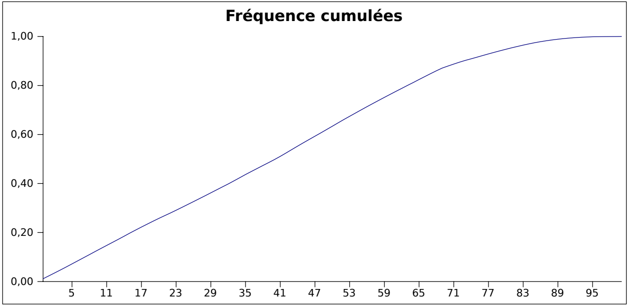
Pour $\alpha \in [0,1]$, un **quantile d'ordre** α est un réel x_α tel que $F(x_\alpha) = \alpha$, c'est-à-dire qu'une proportion α des données est inférieure ou égale à x_α.
Cas particuliers :

- 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 »).

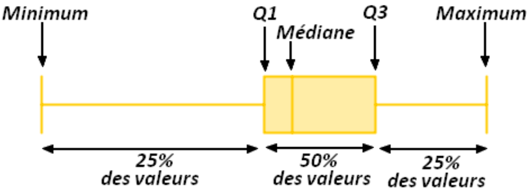
Exemple

Fréquences cumulées des âges de la population française



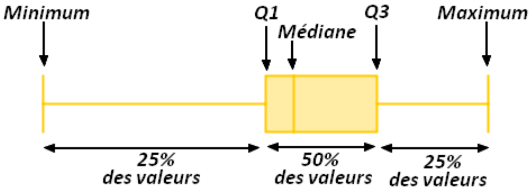
Premier quartile : 19 ans
Médiane : 40 ans
Troisième quartile : 59 ans

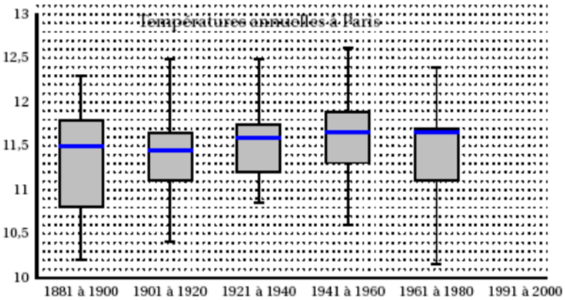
Boîte de dispersion (ou boîte à moustaches)



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

Boîte de dispersion (ou boîte à moustaches)





Attention, 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 ?
$\rightarrow$ *graphiquement*, via un histogramme, etc.
$\rightarrow$ *numériquement*, en calculant des quantités qui mesurent l'ordre de grandeur des valeurs observées (ex. : médiane), quantifient leur dispersion (ex. : quantiles), leur asymétrie, etc.

Moyenne et médiane

Si $x_1, \ldots, x_n$ sont les observations d'une variable quantitative, la **moyenne** de cette série statistique est

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

Pour une variable quantitative discrète, de valeurs $v_1, \ldots, v_r$, ayant pour effectifs $n_1, \ldots, n_r$ et fréquences $f_1, \ldots, f_r$, on a aussi

$$\bar{x} = \frac{n_1 v_1 + \cdots + n_r v_r}{n} = f_1 v_1 + \cdots + f_r v_r.$$

Moyenne et médiane « approchent » au mieux les données :

Proposition

La moyenne est l'unique réel c qui minimise $\displaystyle\sum_{i=1}^{n} (x_i - c)^2$.

Les médianes sont les réels c qui minimisent $\displaystyle\sum_{i=1}^{n} |x_i - c|$.

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

Variance

La **variance** de la série statistique est

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \quad (= \overline{x^2} - \overline{x}^2),$$

et son **écart type** est $\sigma_x = \sqrt{s_x^2}$.
C'est une mesure de la dispersion des valeurs autour de la moyenne : si σ_x est petit par rapport à $\overline{x}$, les données sont concentrées autour de $\overline{x}$.

Couple de variables, corrélation

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

La *covariance* des deux variables est

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

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

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.

Régression linéaire

Si les variables sont fortement corrélées (corrélation proche de ± 1), on peut penser qu'une variable est la cause de l'autre, et on peut chercher à approcher y par une fonction de x. Le cas le plus simple est le cas affine ($y_i \simeq ax_i + b$), qui peut s'aborder par *régression linéaire* :

Proposition (Droite des moindres carrés)

La droite $y = ax + b$ qui minimise la moyenne des carrés des erreurs

$$E(a,b) = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (ax_i + b) \right)^2$$

est donnée par

$$a = \frac{s_{x,y}}{s_x^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} \quad \textit{et} \quad b = \overline{y} - a\overline{x},$$

et le minimum est $\min_{a,b} E(a,b) = s_y^2 (1 - \rho_{x,y}^2)$.

NB. Si on cherche une relation de la forme $y \simeq Cx^\alpha$ ou $y \simeq Ce^{\alpha x}$, prendre le log ramène à une régression linéaire entre $\ln(x)$ et $\ln(y)$, ou entre x et $\ln(y)$.

Cours 8

–

Mercredi 16 mars 2016

Plan du cours

1. Espaces de probabilité.
2. Variables aléatoires. Généralités
3. Couples de variables aléatoires
4. Statistiques descriptives
5. Estimation
    - Principe, statistiques classiques
    - Construction d'estimateurs

Principe de l'estimation statistique

Définition

Soit X une variable aléatoire. Un ***échantillon de taille*** n ***de*** X *est une famille $X_1, \ldots, X_n$ 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 française. Son espérance est donc la taille moyenne d'un Français adulte, que l'on peut vouloir estimer.

- On ne dispose pour cela que d'une réalisation d'un **échantillon de taille** n : une réalisation $(x_1, \ldots, x_n) \in \mathbb{R}^n$ de n variables aléatoires indépendantes $(X_1, \ldots, X_n)$ 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 Français adultes.

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

Principe de l'estimation statistique

- On souhaite étudier la loi de X.

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

- On ne dispose pour cela que d'une réalisation d'un **échantillon de taille** n : une réalisation $(x_1, \ldots, x_n) \in \mathbb{R}^n$ de n variables aléatoires indépendantes $(X_1, \ldots, X_n)$ 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 Français adultes.

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

*Autre exemple : **Sondage**. X est la réponse (0 ou 1) à une question posée à un individu choisi uniformément au hasard dans la population française. Sa loi est $\mathcal{B}(p)$, où p est la proportion de Français répondant "1".*
On cherche p à partir des réponses $x_1, \ldots, x_n$ obtenues en interrogeant n personnes choisies au hasard.

Statistiques simples

Les quantités les plus classiques pour décrire un échantillon sont

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

Proposition

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

$$E\left[\overline{X}_n\right] = m, \qquad \mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma^2}{n},$$

$$E\left[S_n^2\right] = \frac{n-1}{n}\sigma^2, \qquad E\left[\Sigma_n^2\right] = \sigma^2.$$

Estimateurs

On suppose que X suit une loi P_θ qui dépend d'un paramètre $\theta \in \Theta$, où $\Theta \subset \mathbb{R}$ 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 $T_n = f(X_1, \ldots, X_n)$ qui dépend d'un échantillon $X_1, \ldots, X_n$ de X.*
On utilise souvent la notation $\widehat{\theta}$ pour un estimateur de θ.
Une ***estimation*** *de θ est la valeur réelle $t_n = f(x_1, \ldots, x_n)$ prise par une réalisation particulière de l'échantillon.*

Définitions : 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 $E[T_n] \underset{n\to\infty}{\longrightarrow} \theta$, quel que soit $\theta \in \Theta$.
On dit que T_n est **convergent** si, quel que soit $\theta \in \Theta$,

$$\text{pour tout } \alpha > 0, \qquad P\big(|T_n - \theta| > \alpha\big) \underset{n\to\infty}{\longrightarrow} 0.$$

Exemples d'estimateurs (espérance et variance)

Proposition

On suppose que X a pour espérance m et 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.*

Exemple : Si on sait que $X_1, \ldots, X_n$ suivent la loi $\mathcal{P}(\lambda)$, 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 $\overline{X}_n$ 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.

$\rightsquigarrow$ Souvent, on disposera de *plusieurs* estimateurs de la même quantité

Estimateurs – Risque quadratique

Comment mesurer la qualité et comparer deux estimateurs ?

Définition

*Le **risque quadratique** d'un estimateur T_n de θ est*

$$R_{T_n}(\theta) = E\left[(T_n - \theta)^2\right].$$

*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 $R_{T_n}(\theta) = \mathrm{Var}(T_n)$.

Construction d'estimateurs

Problème

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

Deux méthodes classiques :

$\rightarrow$ méthode des moments
$\rightarrow$ méthode du maximum de vraisemblance.

Méthode des moments

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,*

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

   est un estimateur sans biais et convergent de $m_r = E[X^r]$. (Si m_r est bien défini)

Méthode des moments

On en déduit la méthode des moments :

- calculer les moments $m_1 = E[X]$, $m_2 = E[X^2]$, etc., jusqu'à pouvoir exprimer θ à l'aide de ceux-ci ;
- remplacer dans cette expression les moments par les moments empiriques : m_1 remplacé par $\overline{X}_n$, m_2 remplacé par $\widehat{m}_2$, 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.

Méthode du maximum de vraisemblance

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

Définition

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

- *si X est discrète, de loi P_θ, pour tous $x_1, \ldots, x_n \in X(\Omega)$,*

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} P_\theta(x_i)$$

- *si X est continue, de densité f_θ, pour tous $x_1, \ldots, x_n \in X(\Omega)$,*

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i).$$

*Un **estimateur du maximum de vraisemblance** pour θ est un estimateur $h(X_1, \ldots, X_n)$ tel que, pour tous $x_1, \ldots, x_n \in X(\Omega)$,*

$$L\big(x_1, x_2, \ldots, x_n; h(x_1, \ldots, x_n)\big) = \max_{\theta \in \Theta} L(x_1, \ldots, x_n; \theta).$$

Méthode du maximum de vraisemblance

En pratique,

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

$$\varphi(\theta) = L(x_1, \ldots, x_n; \theta).$$

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

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

  (c'est équivalent car le logarithme est strictement croissant) et donc calculer de $\psi'(\theta)$, etc.
- L'estimateur du maximum de vraisemblance s'obtient en prenant $x_1, \ldots, x_n$ égaux aux variables aléatoires $X_1, \ldots, X_n$.

Sous des hypothèses assez générales, on montre que ceci définit un bon estimateur convergent.

Cours 9

–

Mercredi 23 mars 2016

Cadre statistique

On dispose d'**observations** $x_1, \ldots, x_n \in \mathbb{R}$.

On suppose que ces valeurs sont issues de la répétition d'une expérience aléatoire : il y a des **variables aléatoires** $X_1, \ldots, X_n$ indépendantes et de même loi, et $\omega \in \Omega$, telles que

$$x_1 = X_1(\omega), \ldots, x_n = X_n(\omega).$$

Ici, ω représente le tirage aléatoire observé.

On note X une variable aléatoire suivant aussi cette loi. On dit que $(X_1, \ldots, X_n)$ est un **échantillon** de X.

Énoncé du problème

On suppose que la loi de X (et donc de $X_1, \ldots, X_n$) est une loi P_θ qui dépend d'un **paramètre** θ, où $\theta \in \mathbb{R}^d$. On ignore la valeur de θ, et on souhaite en donner une **estimation** grâce à $x_1, \ldots, x_n$, par une fonction $f(x_1, \ldots, x_n)$.

Pour mesurer la qualité de l'estimation, on étudie la loi de l'**estimateur**, qui est la variable aléatoire $\widehat{\theta} = f(X_1, \ldots, X_n)$: est-ce que $\widehat{\theta}$ est souvent proche de θ ? ("proche" en probabilité, en espérance, en moyenne quadratique,...)

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 ?

Définition

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

Cela sera aussi utile pour prendre une décision : si un sondage donne 52 % de "Oui", peut-on conclure que plus de la moitié de la population pense "Oui", ou est-ce une illusion liée au hasard utilisé pour faire le sondage ?

Connaître la marge d'erreur du résultat permettra de conclure, ou non, selon si l'intervalle est plutôt $[50{,}5\,;\, 53{,}5]$ ou $[49{,}5\,;\, 54{,}5]$.

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 $(X_1,X_2,\cdots,X_n)$ un échantillon de X. On cherche un intervalle de confiance pour la moyenne, à partir des estimateurs sans biais $\overline{X}_n$ et Σ_n^2.
Résultat crucial : quelle que soit la loi de X, $\overline{X}_n$ a le même type de loi :

Théorème (Théorème central limite)

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

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - m)}{\sigma}.$$

Autrement dit,

$$\overline{X}_n = m + \frac{\sigma}{\sqrt{n}} Z_n.$$

Lorsque $n \to +\infty$, Z_n converge en loi vers une variable $Z \sim \mathcal{N}(0,1)$:

pour tout intervalle $I \subset \mathbb{R}$, $\quad P(Z_n \in I) \underset{n\to\infty}{\longrightarrow} P(Z \in I) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_I e^{-\frac{x^2}{2}} dx.$

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

Intervalle de confiance pour la moyenne m

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

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

Par exemple, on sait que (voir la table)

$$P(|Z| \leq 1{,}96) = 95\% \quad \text{et} \quad P(|Z| \leq 2{,}58) = 99\%.$$

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

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

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

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

ce qui se réécrit

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

Si $\overline{x}_n$ est la moyenne observée et σ_n^2 la variance corrigée observée, en posant $\pi_1 = \overline{x}_n - a\frac{\sigma_n}{\sqrt{n}}$ et $\pi_2 = \overline{x}_n + a\frac{\sigma_n}{\sqrt{n}}$, on en déduit que $[\pi_1, \pi_2]$ est un intervalle de confiance de m de niveau $1 - \alpha$.

Intervalle de confiance pour une proportion p.

Ici X est une variable aléatoire de Bernoulli de paramètre p, d'où

$$E[X] = p \qquad \text{et} \qquad \mathrm{Var}(X) = p(1-p).$$

On a, pour n grand,

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

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

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

Comme, pour tout p tel que $0 < p < 1$, on a $p(1-p) \leq \frac{1}{4}$, on déduit

$$I \subset J = \left[\overline{X}_n - \frac{a}{2\sqrt{n}}, \overline{X}_n + \frac{a}{2\sqrt{n}}\right].$$

On utilise souvent plutôt J (plus simple) comme intervalle de confiance de niveau $1-\alpha$.

Pour $\alpha = 5\,\%$, on a même $a = 1{,}96 \leq 2$ donc $J \subset \left[\overline{X}_n - \frac{1}{\sqrt{n}}, \overline{X}_n + \frac{1}{\sqrt{n}}\right]$.

Remarques.

1. Attention : pour que ces approximations soient justifiées, les valeurs de $n\pi_1$, $n\pi_2$, $n(1-\pi_1)$ et $n(1-\pi_2)$ doivent être toutes les quatre supérieures ou égales à 5.
2. 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 $x_1, \ldots, x_n$, 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 $\mathcal{H}_0$. 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 $X_1, \ldots, X_n$ pour décider si on rejette $\mathcal{H}_0$, ou si on considère les données compatibles avec $\mathcal{H}_0$.
Le **seuil de risque** du test est la probabilité d'erreur de première espèce :

$$\alpha = P(\text{rejet de } \mathcal{H}_0) \text{ quand } \mathcal{H}_0 \text{ est vraie.}$$

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

On se donne également une **hypothèse alternative** $\mathcal{H}_1$ (incompatible avec $\mathcal{H}_0$), que l'on veut détecter. Souvent $\mathcal{H}_1$ sera le contraire de $\mathcal{H}_0$.
La **puissance** du test est $1 - \beta$, où β est la probabilité d'erreur de 2e espèce :

$$\beta = P(\text{acceptation de } \mathcal{H}_0) \text{ quand } \mathcal{H}_1 \text{ est vraie.}$$

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 $n \to \infty$.

Test d'adéquation de m à une valeur théorique

Échantillon $X_1, \ldots, X_n$ de même loi que X, $E[X] = m$ et $\mathrm{Var}(X) = \sigma^2$.
On veut savoir si l'observation $X_1, \ldots, X_n$ est compatible avec l'hypothèse

$$\mathcal{H}_0 \ : \ \text{“}m = 7{,}5\text{”}$$

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

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

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

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

NB. Si $\mathcal{H}_0$ est fausse, $\overline{X}_n - 7{,}5 \simeq m - 7{,}5 \neq 0$ 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.

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

$$\mathcal{H}_1 \ : \ \text{“}m > 7{,}5\text{”}$$

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 $P\big(\frac{\sqrt{n}}{\sigma}(\overline{X}_n - 7{,}5) \le A\big) \simeq 1 - \alpha$, avec A tel que $P(Z \le A) \simeq 1 - \alpha$ pour $Z \sim \mathcal{N}(0{,}1)$ (par ex. $P(Z \le 1{,}65) \simeq 0{,}95$) ou encore, en approchant σ par $\sqrt{\Sigma_n^2}$,

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

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

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

NB. On a $A \le a$ 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.