Cours Commun Scientifique de Probabilités & Statistiques

Laurent Tournier

Janvier 2019

Plan du cours

1. Espaces de probabilité.
   - Définitions
   - Équiprobabilité
   - Probabilités conditionnelles

2. Variables aléatoires. Généralités

3. Couples de variables aléatoires

Espaces de probabilités

Définition

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

- *Ω, un ensemble*
- *P, une probabilité sur Ω.   (qui reste à définir)*

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 :

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

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

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

- 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.

Une expérience aléatoire d'actualité : tirer les rois

Selon une tradition française (du XV^{e} siècle), on « tire les rois » à l'Épiphanie (1er dimanche de janvier) : une **fève** est cachée dans une **galette**, qui est un gâteau feuilleté fourré à la frangipane (pâte aux amandes). On découpe cette galette, et la personne qui obtient la fève devient le "roi"/la "reine" de la journée.

La **fève** était à l'origine une fève (un haricot sec). On utilise maintenant plutôt un petit objet, généralement en porcelaine, qui peut se collectionner.

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

a) $P(\emptyset) = 0$

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

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

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

Propriétés

a) $P(\emptyset) = 0$

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

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

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

Preuve de b) : 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 a) en prenant $A = \Omega$.

Espaces de probabilités

Propriétés

a) $P(\emptyset) = 0$

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

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

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

Preuve de b) : 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 a) en prenant $A = \Omega$.
Preuve de c) : 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

a) $P(\emptyset) = 0$

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

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

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

Preuve de b) : 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 a) en prenant $A = \Omega$.
Preuve de c) : 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 d) : $A \setminus (A \cap B)$, $A \cap B$ et $B \setminus (A \cap B)$ sont disjoints, d'union $A \cup B$, donc

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

Propriétés

- e) *Pour toute suite $(A_n)_n$ finie ou infinie, $P(\bigcup_n A_n) \leq \sum_n P(A_n)$.*
  (on dit que P est sous-additive)

Preuve de e). On l'a vu pour 2 événements :
$P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$.
On en déduit le cas d'une suite finie par récurrence.

$(*)$ Pour une suite infinie $(A_n)_{n \geq 0}$, on peut poser $C_0 = A_0$ puis $C_n = A_n \setminus (A_1 \cup \cdots \cup A_{n-1})$, alors $C_0, C_1, \ldots$ sont disjoints, $C_0 \cup \cdots \cup C_n = A_0 \cup \cdots \cup A_n$ et $\bigcup_{n \geq 0} C_n = \bigcup_{n \geq 0} A_n$. Alors

$$P(\bigcup_{n \geq 0} A_n) = P(\bigcup_{n \geq 0} C_n) = \sum_{n \geq 0} P(C_n) = \lim_N \sum_{n=0}^{N} P(C_n) = \lim_N P(A_1 \cup \cdots \cup A_n)$$

et

$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + \cdots + P(A_n) \underset{n \to \infty}{\longrightarrow} \sum_{k \geq 0} P(A_k)$$

Distribution uniforme de probabilité

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

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

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 $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}\} \subset \Omega$,*

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

Autrement dit,

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

Rappels de dénombrement :

Calculer des probabilités dans ce cas se ramène donc à dénombrer (compter) les éléments d'un ensemble.

On commence par un résultat très simple :

Soit $n \in \mathbb{N}^*$. Soit $E_1, \ldots, E_n$ des ensembles finis.

Un n**-uplet** $(x_1, \ldots, x_n)$ est une suite de n éléments (l'ordre est important).
Le nombre de n-uplets $(x_1, \ldots, x_n)$ tels que $x_1 \in E_1$,..., $x_n \in E_n$, est

$$\operatorname{Card} E_1 \times \operatorname{Card} E_2 \times \cdots \times \operatorname{Card} E_n.$$

On retient que, s'il y a

- k_1 choix pour la valeur de x_1, puis
- k_2 choix pour la valeur de x_2 (quel que soit x_1),
- etc.,

alors il y a $k_1 k_2 \cdots k_n$ façons de choisir le n-uplet $(x_1, \ldots, x_n)$.

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

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

Un exemple très simple

Une galette comporte 8 parts, dont 2 contiennent une fève. On prend 3 parts au hasard (sans les remettre...). Quelle est la probabilité de ne pas avoir de fève ?

Un exemple très simple

Une galette comporte 8 parts, dont 2 contiennent une fève. On prend 3 parts au hasard (sans les remettre...). Quelle est la probabilité de ne pas avoir de fève ?

Il y a 6 parts sans fève, d'où

$$P(\text{ne pas avoir de fève}) = \frac{\text{nb de choix de 3 parts sans fève}}{\text{nb de choix de 3 parts}}$$

$$= \frac{C_6^3}{C_8^3} = \frac{A_6^3}{A_8^3} = \frac{6 \cdot 5 \cdot 4}{8 \cdot 7 \cdot 6} \simeq 36\,\%$$

Comme l'événement ne dépend pas de l'ordre, on peut choisir deux espaces Ω différents (au moins) : l'ensemble des combinaisons de 3 éléments parmi 8, ou l'ensemble des arrangements de 3 éléments parmi 8. Avec la probabilité uniforme sur Ω.

Un exemple très simple

Une galette comporte 8 parts, dont 2 contiennent une fève. On prend 3 parts au hasard (sans les remettre...). Quelle est la probabilité de ne pas avoir de fève ?

Il y a 6 parts sans fève, d'où

$$P(\text{ne pas avoir de fève}) = \frac{\text{nb de choix de 3 parts sans fève}}{\text{nb de choix de 3 parts}}$$

$$= \frac{C_6^3}{C_8^3} = \frac{A_6^3}{A_8^3} = \frac{6 \cdot 5 \cdot 4}{8 \cdot 7 \cdot 6} \simeq 36\,\%$$

Comme l'événement ne dépend pas de l'ordre, on peut choisir deux espaces Ω différents (au moins) : l'ensemble des combinaisons de 3 éléments parmi 8, ou l'ensemble des arrangements de 3 éléments parmi 8. Avec la probabilité uniforme sur Ω.

NB. Vous auriez peut-être envie d'écrire $\frac{6}{8} \cdot \frac{5}{7} \cdot \frac{4}{6}$, cela correspondrait à utiliser la notion de *probabilité conditionnelle*. La première part a 6 chances sur 8 de ne pas avoir de fèves ; *sachant cela,* la deuxième a 5 chances sur 7 de ne pas en avoir non plus ; etc.

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 l'ensemble des n-uplets dans $\{1, \ldots, N\}$:

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

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 l'ensemble des n-uplets dans $\{1, \ldots, N\}$:

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

avec P uniforme, 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 l'ensemble des n-uplets dans $\{1, \dots, N\}$:

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

avec P uniforme, et on cherche $P(A)$ où

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

Alors

$A^c = \{$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 l'ensemble des n-uplets dans $\{1, \ldots, N\}$:

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

avec P uniforme, et on cherche $P(A)$ où

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

Alors

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

est l'ensemble des arrangements de n éléments parmi N, donc

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

Exemple : 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 l'ensemble des n-uplets dans $\{1,\dots,N\}$:

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

avec P uniforme, 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

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

est l'ensemble des arrangements de n éléments parmi N, donc

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

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

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 l'ensemble des n-uplets dans $\{1,\dots,N\}$:

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

avec P uniforme, 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

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

est l'ensemble des arrangements de n éléments parmi N, donc

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

Exemple : Pour $n = 23$, $P(A) \simeq 0{,}5$. Pour $n = 57$, $P(A) \simeq 0{,}99$.
+ difficile : Si $n \geq 88$, $P(\text{3 étudiants ont leur anniversaire ensemble}) \geq 0{,}5$.

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 proportionnelle 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

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

Je vous dis que 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

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

Exemple : on tire deux pièces à pile-ou-face. $A = \{\text{la première est pile}\}$ et $B = \{\text{la deuxième est pile}\}$ sont indépendants, mais A, B et $C = \{\text{les deux sont du même côté}\}$ ne sont pas indépendants.

Par contre, A et C sont indépendants, et B et C aussi.

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

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

Exemple. Dans un jeu de pile-ou-face, si $A_i = \{\text{le } i^{\text{ème}} \text{ tirage est pile}\}$, $A_1, A_2, \dots$ 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

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

On note 1 pour Pile et 0 pour Face. Notons $A_i = \{\text{le tirage } i \text{ est pile}\}$
$\to$ chaque réalisation ω est une suite de 0 et de 1 de longueur n : $\Omega = \{0,1\}^n$.
$A_1, \ldots, A_n$ sont indépendants, donc par exemple (ici, $n = 4$)

$$P(\{(1,0,1,1)\}) = P(A_1 \cap A_2^c \cap A_3 \cap A_4) = p \times (1-p) \times p \times p = p^3(1-p)$$

et, si la suite $\omega = (\varepsilon_1, \ldots, \varepsilon_n)$ 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 suite $\omega \in B_k$, $P(\{\omega\}) = p^k(1-p)^{n-k}$. Par ailleurs, le nombre de telles suites 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

Par le même calcul, 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 a

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

Loi binomiale

Par le même calcul, 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 a

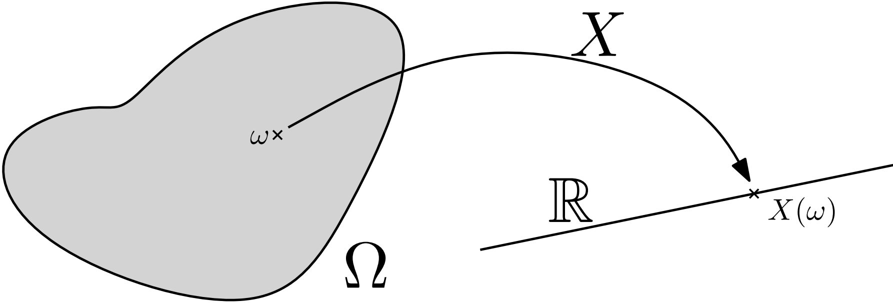
$$P(\text{exactement } k \text{ événements parmi } A_1,\ldots,A_n \text{ se réalisent}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

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

Définition

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



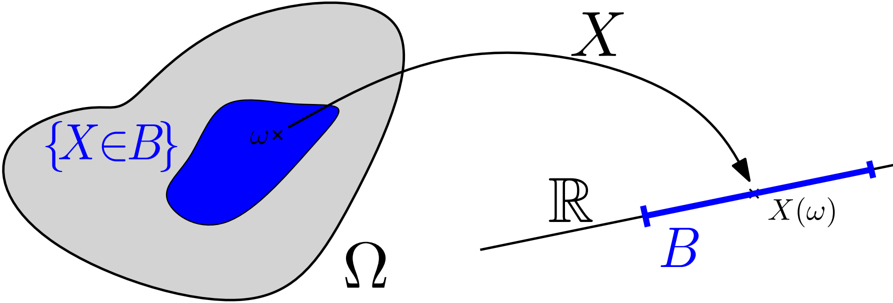
Variables aléatoires

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" au sens usuel (mais c'est un cas particulier de variable aléatoire).
  En général, la valeur de $X(\omega)$ *dépend* de la réalisation ω, et la distribution de ces valeurs 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

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

- 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 $\{1, \ldots, 6\}$

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 $\{0, 1, \ldots, n\}$ et

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

$\rightsquigarrow$ si $A_1, \ldots, A_n$ 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, $\Omega = \{1,\ldots,6\}^2 = \{(x_1,x_2) \mid x_1,x_2 \in \{1,\ldots,6\}\}$
  - Somme des résultats : $X((x_1,x_2)) = x_1 + x_2$
    (à valeurs dans $\{2,\ldots,12\}$)

Lois discrètes – Exemples

Retour sur la liste d'exemples :

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

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

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

$$\dots$$

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

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

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

$$\dots$$

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

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

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 $\{1,\ldots,8\}$*)*
    $\leadsto N_A$ 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 $\{1,2,\ldots\} = \mathbb{N}^*$*)*

  - 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 $\{1,\ldots,8\}$*)*
    $\rightsquigarrow$ 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 $n \in \mathbb{N}^*$,

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

    $\rightsquigarrow$ 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 $\{1,\ldots,8\}$)
    $\rightsquigarrow$ 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 $n \in \mathbb{N}^*$,

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

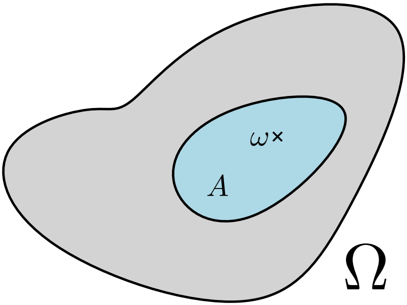
    $\rightsquigarrow$ N_B suit la **loi géométrique** de paramètre $p = \frac{1}{8}$.
  - Nombre de fèves obtenues en n jours : S_n
    (à valeurs dans $\{0,1,2,\ldots,n\}$)
    $\rightsquigarrow$ S_n suit la loi binomiale de paramètres $(n, \frac{1}{8})$.

Cours 2

–

Mercredi 30 janvier 2019

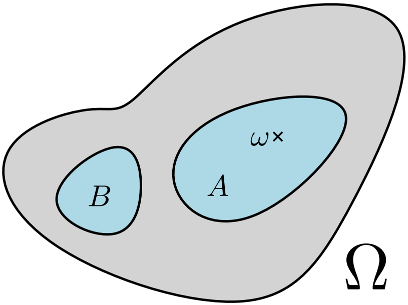
Espace de probabilités – Rappel



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



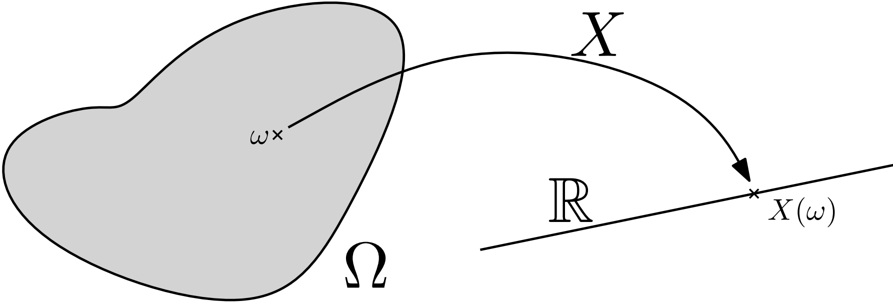
Espace de probabilités : (Ω, P)

- 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, $P(\bigcup_n A_n) = \sum_n P(A_n)$.

Définition

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



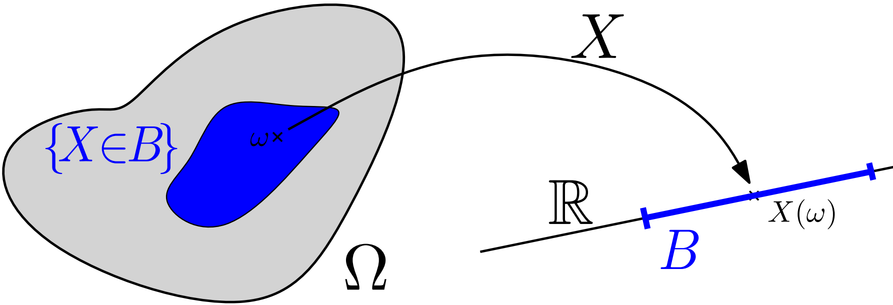
Variables aléatoires – Rappel

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. X est "à valeurs dans $X(\Omega)$"*
$\rightsquigarrow P_X$ 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.



Lois discrètes – Rappel

Définition

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

Si X est discrète, alors pour tout $B \subset X(\Omega)$, on a

$$\{X \in B\} = \bigcup_{x \in B} \{X = x\},$$

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

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

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

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

Lois discrètes – Exemples

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

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

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 $\{0,1,\ldots,n\}$ et

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

$\rightsquigarrow$ si $A_1,\ldots,A_n$ 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 $\mathcal{B}(n,p)$.

Soit $p \in]0,1]$. Une variable aléatoire X suit la **loi géométrique de paramètre** p (notée $\mathcal{G}(p)$) si X est à valeurs dans $\mathbb{N}^* = \{1,2,\ldots\}$ et

$$\text{pour } n \geq 1, \quad P(X=n) = (1-p)^{n-1} p.$$

$\rightsquigarrow$ si $A_1, A_2, \ldots$ sont indépendants et $P(A_n) = p$ pour tout n, alors la loi de $X = \inf\{n \geq 1 \mid \mathbf{1}_{A_n} = 1\} =$ « nombre de tentatives jusqu'à un succès » est $\mathcal{G}(p)$.

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 $0 \leq x \leq r$,

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

$\Rightarrow$ l'approche précédente est 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 $0 \leq x \leq r$,

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

$\Rightarrow$ l'approche précédente est 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

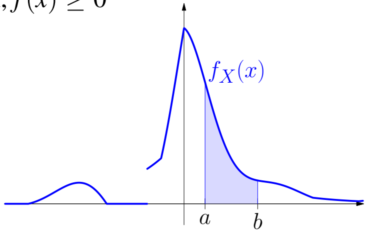
Définition

*Une variable aléatoire X est dite **continue** ou **à densité** s'il existe une fonction (intégrable) $f_X : \mathbb{R} \to [0, +\infty[$ 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 $a \leq b$, $P(a \leq X \leq b) = \displaystyle\int_a^b f_X(x)dx$

Lois continues

Définition

*Une variable aléatoire X est dite **continue** ou **à densité** s'il existe une fonction (intégrable) $f_X : \mathbb{R} \to [0, +\infty[$ 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 $x \in \mathbb{R}$, $P(X = x) = \int_{\{x\}} f_X(t)dt = 0.$
- Si $f_X(x) = 0$ pour tout $x \in B$, alors $P(X \in B) = \displaystyle\int_B f_X(x)dx = 0.$

  $\Rightarrow$ les valeurs prises par X sont dans le *support* de f_X :

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

Interprétation intuitive de la densité

$f_X(x)$ représente la probabilité que X est dans un (petit) voisinage de x, *rapportée à la longueur de ce voisinage* (d'où le terme "densité") :

Supposons que X a pour densité f_X, continue au point $x \in \mathbb{R}$. Alors

$$\frac{P(X \in [x - \frac{\delta}{2}, x + \frac{\delta}{2}])}{\delta} \underset{\delta \to 0^+}{\longrightarrow} f_X(x).$$

Interprétation intuitive de la densité

$f_X(x)$ représente la probabilité que X est dans un (petit) voisinage de x, *rapportée à la longueur de ce voisinage* (d'où le terme “densité”) :

Supposons que X a pour densité f_X, continue au point $x \in \mathbb{R}$. Alors

$$\frac{P(X \in [x - \frac{\delta}{2}, x + \frac{\delta}{2}])}{\delta} \underset{\delta \to 0^+}{\longrightarrow} f_X(x).$$

Soit $\varepsilon > 0$. Pour un certain $\delta > 0$, on a $|f_X(t) - f_X(x)| < \varepsilon$ dès que $|t - x| < \delta$, d'où

$$\left| P\left(X \in \left[x - \frac{\delta}{2}, x + \frac{\delta}{2}\right]\right) - \delta f_X(x) \right| = \left| \int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} f_X(t)dt - \delta f_X(x) \right|$$

$$= \left| \int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} \big(f_X(t) - f_X(x)\big)dt \right| \leq \int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} |f_X(t) - f_X(x)|dt \leq \delta\varepsilon.$$

Densités classiques

Soit $a < b$. La **loi uniforme sur** $[a,b]$ (notée $\mathcal{U}([a,b])$) est la loi de densité

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

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 (notée $\mathcal{E}(\lambda)$) a pour 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 $s,t \ge 0$,

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

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

Quelques exemples de calculs

Si X suit la loi uniforme sur $[a,b]$, et si $[c,d] \subset [a,b]$, on retrouve

$$P(X \in [c,d]) = \int_{[c,d]} \frac{1}{b-a} dt = \frac{d-c}{b-a} = \frac{\text{longueur}([c,d])}{\text{longueur}([a,b])}.$$

Supposons que X suit la loi exponentielle de paramètre 2 : densité $f : x \mapsto 2e^{-2x}\mathbf{1}_{[0,+\infty[}(x)$.
Par exemple $P(X > 5) = \int_{]5,+\infty[} 2e^{-2x}\mathbf{1}_{[0,+\infty[}(x)dx = \int_5^\infty 2e^{-2x}dx = e^{-10}$.

Supposons que X suit la loi uniforme sur $[0,1]$. Posons $Y = \lfloor 5X \rfloor$ (partie entière). Quelle est la loi de Y ?

- Y est à valeurs dans $\{0,1,2,3,4\}$: elle est donc discrète
- pour $k = 0,1,2,3,4$, $P(Y = k) = P(k \leq 5X < k+1)$
$$= P(\frac{k}{5} \leq X < \frac{k+1}{5}) = \int_{\frac{k}{5}}^{\frac{k+1}{5}} \mathbf{1}_{[0,1]} dx = \frac{1}{5}$$

Donc Y suit la loi uniforme sur $\{0,1,2,3,4\}$.

Attention

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

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

$$Y = \min(X, \frac{1}{2}) = \begin{cases} X & \text{si } X < \frac{1}{2} \\ \frac{1}{2} & \text{sinon.} \end{cases}$$

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

Attention

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

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

$$Y = \min(X, \frac{1}{2}) = \begin{cases} X & \text{si } X < \frac{1}{2} \\ \frac{1}{2} & \text{sinon.} \end{cases}$$

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

- Y est à valeurs dans $[0, \frac{1}{2}]$ car X est à valeurs dans $[0,1]$.

Attention

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

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

$$Y = \min(X,\frac{1}{2}) = \begin{cases} X & \text{si } X < \frac{1}{2} \\ \frac{1}{2} & \text{sinon.} \end{cases}$$

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

- Y est à valeurs dans $[0,\frac{1}{2}]$ car X est à valeurs dans $[0,1]$.
- On a $P\left(Y = \dfrac{1}{2}\right) = P\left(X \geq \dfrac{1}{2}\right) = \displaystyle\int_{1/2}^{\infty} \mathbf{1}_{[0,1]}(x)dx = \int_{\frac{1}{2}}^{1} 1dx = \frac{1}{2} > 0$

$\Rightarrow$ Y n'a pas de densité

Attention

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

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

$$Y = \min(X, \frac{1}{2}) = \begin{cases} X & \text{si } X < \frac{1}{2} \\ \frac{1}{2} & \text{sinon.} \end{cases}$$

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

• Y est à valeurs dans $[0, \frac{1}{2}]$ car X est à valeurs dans $[0,1]$.

• On a $P\left(Y = \dfrac{1}{2}\right) = P\left(X \geq \dfrac{1}{2}\right) = \displaystyle\int_{1/2}^{\infty} \mathbf{1}_{[0,1]}(x)dx = \int_{\frac{1}{2}}^{1} 1dx = \frac{1}{2} > 0$

$\Rightarrow$ Y n'a pas de densité

• Pour tout $0 \leq x < \frac{1}{2}$, $P(Y = x) = P(X = x) = 0$

$\Rightarrow$ Y n'est pas discrète (si elle l'était, $\displaystyle\sum_{y \in Y(\Omega)} P(Y = y) = 1$, mais cette somme vaut $\frac{1}{2}$ ici)

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*

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

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.*

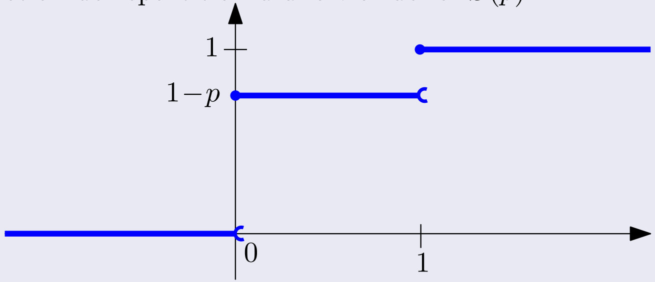
$\rightsquigarrow$ Autrement dit, la fonction de répartition caractérise la loi d'une variable aléatoire.

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 $\mathcal{B}(p)$

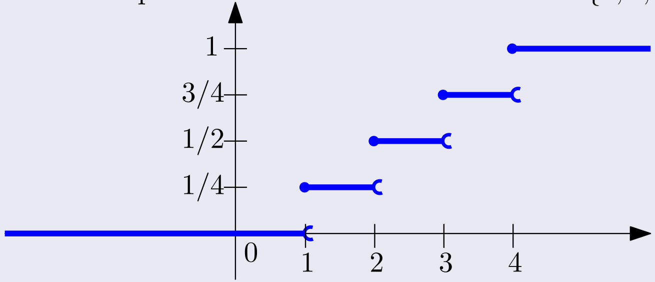


Fonction de répartition – Cas discret

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

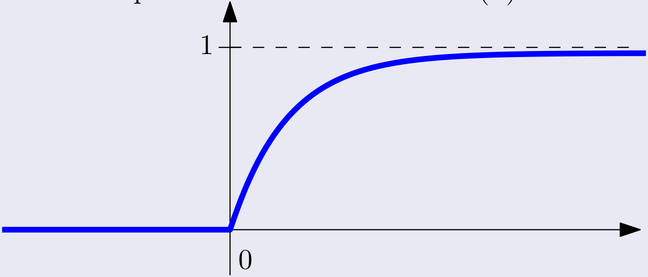
Proposition

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

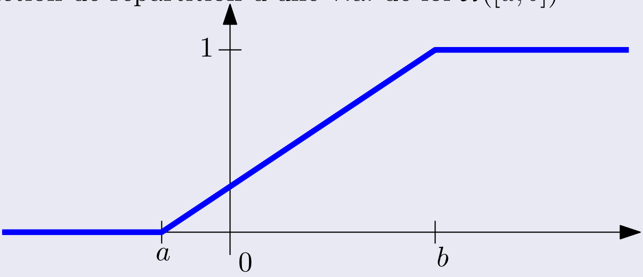
Proposition

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 $f_X = F'_X$. (Avec une valeur quelconque aux points où il n'y a pas de dérivée)

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 $Y = X$)

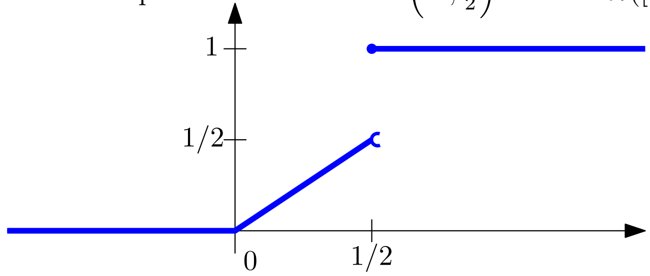
Fonction de répartition – Autre exemple

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 $X \sim \mathcal{U}([0,1])$



Application : 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 $Y = \varphi(X)$.

$\to$ Déterminer les valeurs possibles de Y, puis

Application : 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 $Y = \varphi(X)$.

$\to$ Déterminer les valeurs possibles de Y, puis

• Si Y est discrète (l'ensemble des valeurs possibles est dénombrable),
$\to$ Calculer chacune de leurs probabilités en se ramenant à X.

Application : 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 $Y = \varphi(X)$.

$\to$ Déterminer les valeurs possibles de Y, puis

• Si Y est discrète (l'ensemble des valeurs possibles est dénombrable),
$\to$ Calculer chacune de leurs probabilités en se ramenant à X.

• Si X a une densité f_X, et φ est **monotone** (croissante ou décroissante),
$\to$ 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.

Application : 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 $Y = \varphi(X)$.

$\to$ Déterminer les valeurs possibles de Y, puis

• Si Y est discrète (l'ensemble des valeurs possibles est dénombrable),
$\to$ Calculer chacune de leurs probabilités en se ramenant à X.

• Si X a une densité f_X, et φ est **monotone** (croissante ou décroissante),
$\to$ 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 $Y = \frac{1}{1+X}$.

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 $X > 0$.
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 $]0{,}1[$.

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 $X > 0$.
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, pour $y \in \mathbb{R}$, $F_Y(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ 1 & \text{si } y \geq 1 \end{cases}$ et, si $0 < y \leq 1$,

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

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

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 $X > 0$.
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, pour $y \in \mathbb{R}$, $F_Y(y) = \begin{cases} 0 & \text{si } y \leq 0 \\ 1 & \text{si } y \geq 1 \end{cases}$ et, si $0 < y \leq 1$,

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

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)

Bilan

- Pour modéliser une expérience aléatoire, on a défini un espace de probabilité (Ω, P). Les grandeurs (réelles) mesurées sur l'expérience correspondent à des variables aléatoires $X : \Omega \to \mathbb{R}$.
- Différentes modélisations (choix de Ω, P et donc X) sont possibles pour une même expérience.
- En revanche, si X représente une certaine grandeur qui dépend de l'expérience, la **loi** de X (c'est-à-dire toutes les probabilités $P(X \in A)$) ne dépend pas du choix de la modélisation.

Bilan

- Pour modéliser une expérience aléatoire, on a défini un espace de probabilité (Ω, P). Les grandeurs (réelles) mesurées sur l'expérience correspondent à des variables aléatoires $X : \Omega \to \mathbb{R}$.
- Différentes modélisations (choix de Ω, P et donc X) sont possibles pour une même expérience.
- En revanche, si X représente une certaine grandeur qui dépend de l'expérience, la **loi** de X (c'est-à-dire toutes les probabilités $P(X \in A)$) ne dépend pas du choix de la modélisation.

La loi de X est une probabilité sur $\mathbb{R}$; deux cas sont très fréquents :

- le cas discret : X ne prend des valeurs que dans un ensemble dénombrable $\{x_1, x_2, \ldots\}$. La loi de X équivaut alors à connaître $P(X = x_i)$ pour tout i (ce sont les *probabilités élémentaires*)
- le cas continu/à densité : $P(X = x) = 0$ pour tout $x \in \mathbb{R}$ mais il y a une densité de probabilité, c'est-à-dire que pour tous $a < b$,

$$P(a < X < b) = \int_a^b f(x)dx$$

pour une certaine fonction f. La loi de X équivaut alors à connaître f.

Espérance d'une variable aléatoire – 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 d'une variable aléatoire – 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 d'une variable aléatoire – 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 d'une variable aléatoire – 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 ⇒ 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, $E[X] = \sum_{x \in X(\Omega)} xP(X = x)$.
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 :
• $E[X]$ 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,...
• $E[X]$ est la limite, quand $n \rightarrow \infty$, de la moyenne $\frac{1}{n}(X_1 + \cdots + X_n)$ de n réalisations de X obtenues en répétant l'expérience... On y reviendra.

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 $\displaystyle\sum_{k=1}^{\infty} kx^{k-1} = \frac{d}{dx}\left(\sum_{k=0}^{\infty} x^k\right) = \frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$ 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

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

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 $X \leq Y$, 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*,*

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

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

Si $A_1, \dots, A_n$ 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 $\mathcal{B}(n,p)$:

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

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

Retour sur les anniversaires

On choisit n personnes au hasard. Combien en moyenne y a-t-il de jours dans l'année où au moins 2 ont leur anniversaire ?

Retour sur les anniversaires

On choisit n personnes au hasard. Combien en moyenne y a-t-il de jours dans l'année où au moins 2 ont leur anniversaire ?

On note $N = 365$ et, pour $j = 1, \ldots, N$, on définit l'événement

$$A_j = \{\text{il y a} \geq 2 \text{ anniversaires le jour } j\}.$$

Alors $P(A_j) = 1 - P(A_j^c) = 1 - \left(\frac{N-1}{N}\right)^n - n\frac{1}{N}\left(\frac{N-1}{N}\right)^{n-1}$. Donc le nombre de jours avec ≥ 2 anniversaires est

$$X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_N}$$

et

$$\begin{aligned} E[X] &= E[\mathbf{1}_{A_1}] + \cdots + E[\mathbf{1}_{A_N}] = P(A_1) + \cdots + P(A_N) \\ &= NP(A_1) = N - N\left(\frac{N-1}{N}\right)^n - n\left(\frac{N-1}{N}\right)^{n-1}. \end{aligned}$$

Application numérique : pour $n = 100$, $E[X] \simeq 11{,}4$.

Retour sur les anniversaires

On choisit n personnes au hasard. Combien en moyenne y a-t-il de jours dans l'année où au moins 2 ont leur anniversaire ?

On note $N = 365$ et, pour $j = 1, \ldots, N$, on définit l'événement

$$A_j = \{\text{il y a} \geq 2 \text{ anniversaires le jour } j\}.$$

Alors $P(A_j) = 1 - P(A_j^c) = 1 - \left(\frac{N-1}{N}\right)^n - n\frac{1}{N}\left(\frac{N-1}{N}\right)^{n-1}$. Donc le nombre de jours avec ≥ 2 anniversaires est

$$X = \mathbf{1}_{A_1} + \cdots + \mathbf{1}_{A_N}$$

et

$$\begin{aligned} E[X] &= E[\mathbf{1}_{A_1}] + \cdots + E[\mathbf{1}_{A_N}] = P(A_1) + \cdots + P(A_N) \\ &= NP(A_1) = N - N\left(\frac{N-1}{N}\right)^n - n\left(\frac{N-1}{N}\right)^{n-1}. \end{aligned}$$

Application numérique : pour $n = 100$, $E[X] \simeq 11{,}4$.

NB. $A_1, \ldots, A_N$ ne sont pas indépendants ! Et la loi de X n'est pas binomiale.

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 $[0,1]$,

$E\left[\frac{1}{1+X}\right] = \int_{\mathbb{R}} \frac{1}{1+x} \mathbf{1}_{[0,1]}(x) dx = \int_0^1 \frac{1}{1+x} dx = [\ln(1+x)]_{x=0}^1 = \ln 2$

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

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

On a

$$E[X] = \lambda$$

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

Proposition

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

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

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).

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

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

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, $P(N \leq 7) = \sum_{k=0}^{7}\binom{500}{k}0{,}01^k \cdot 0{,}99^{500-k} \simeq 0{,}868$.

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 $E[X]$?

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** l'espérance ci-dessus aussi.*
$\rightarrow$ Ceci revient à demander à ce que $E[X^2]$ soit définie.

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 ($\mathrm{Var}(Y) = 1$).

Cours 3

–

Mercredi 13 février 2019

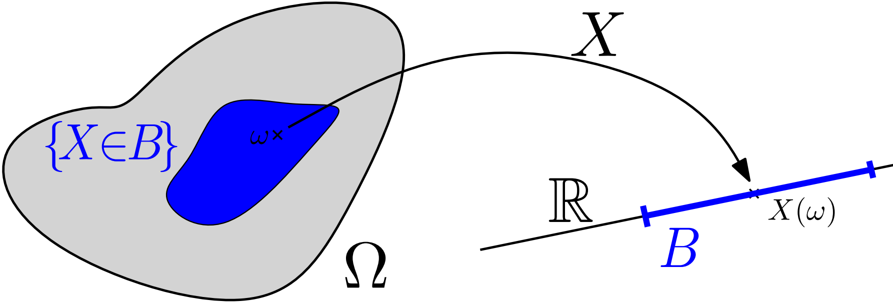
Rappel – Variables aléatoires

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*.*
$\rightsquigarrow$ 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*.*



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, $E[X] = \displaystyle\sum_{x \in X(\Omega)} xP(X = x)$.
Si X est continue, de densité f_X, $E[X] = \displaystyle\int_{\mathbb{R}} x f_X(x)dx$.
***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,...

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 $E[X]$?

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** l'espérance ci-dessus aussi.*
$\rightarrow$ Ceci revient à demander à ce que $E[X^2]$ soit définie.

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*

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

$\mathrm{Var}(X + Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$ en général ! Par exemple,
$\mathrm{Var}(X + X) = \mathrm{Var}(2X) = 4\,\mathrm{Var}(X)$ mais $\mathrm{Var}(X) + \mathrm{Var}(X) = 2\,\mathrm{Var}(X)$...

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 ($\mathrm{Var}(Y) = 1$).

Variance – Exemples discrets

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

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

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

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

Bilan

- Étant donnée une variable aléatoire X, son **espérance** $E[X]$ (si elle existe) est la moyenne de ses valeurs, pondérées par leurs probabilités d'apparition.
- $E[X]$ donne une idée de l'ordre de grandeur "typique" des réalisations de X. C'est en particulier utile si on ne connaît pas la loi de X (on peut en effet souvent calculer $E[X]$ sans connaître la loi de X).
- Afin de mesurer la dispersion des valeurs prises par X autour de $E[X]$, on peut calculer l'**écart-type** $\sigma(X)$ de X.

L'espérance et la variance fournissent également des informations sur certaines probabilités via les inégalités de Markov et Tchebychev...

Inégalités

Proposition (Inégalité de Markov)

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

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

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

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

Preuve

Inégalités

Proposition (Inégalité de Markov)

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

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

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

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

Preuve

On définit une variable aléatoire Y par

$$Y = \begin{cases} a & \text{si } |X| \geq a, \\ 0 & \text{sinon.} \end{cases}$$

Alors on a toujours $|X| \geq Y$.

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

On définit une variable aléatoire Y par

$$Y = \begin{cases} a & \text{si } |X| \geq a, \\ 0 & \text{sinon.} \end{cases}$$

Alors on a toujours $|X| \geq Y$. Donc $E\big[|X|\big] \geq E[Y]$.

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

On définit une variable aléatoire Y par

$$Y = \begin{cases} a & \text{si } |X| \geq a, \\ 0 & \text{sinon.} \end{cases}$$

Alors on a toujours $|X| \geq Y$. Donc $E\big[|X|\big] \geq E[Y]$. D'où l'inégalité, car

$$E[Y] = aP(|X| \geq a) + 0P(|X| < a) = aP(|X| \geq a).$$

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

On définit une variable aléatoire Y par

$$Y = \begin{cases} a & \text{si } |X| \geq a, \\ 0 & \text{sinon.} \end{cases}$$

Alors on a toujours $|X|^r \geq Y^r$. Donc $E\big[|X|^r\big] \geq E[Y^r]$. D'où l'inégalité, car

$$E[Y^r] = a^r P(|X| \geq a) + 0P(|X| < a) = a^r P(|X| \geq a).$$

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 à la v.a. $X - E[X]$.

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 $\geq 75\,\%$, $|X - E[X]| \leq 2\sigma(X)$.
$\rightarrow$ avec probabilité $\geq 99\,\%$, $|X - E[X]| \leq 10\sigma(X)$.

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 à la v.a. $X - E[X]$.

Autre écriture

Pour tout $A > 0$,

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

$\to$ avec probabilité $\geq 75\,\%$, $|X - E[X]| \leq 2\sigma(X)$.
$\to$ avec probabilité $\geq 99\,\%$, $|X - E[X]| \leq 10\sigma(X)$.

Ces inégalités sont intéressantes si on ne connaît pas la loi de X (ou si elle est compliquée), mais que l'on connaît $E[X]$ et $\sigma(X)$. Cela arrive notamment quand X est définie à partir de *plusieurs* variables aléatoires.

Désormais, on va s'intéresser à **plusieurs** variables aléatoires, et à la façon dont elles sont liées.

Indépendance de variables aléatoires

Définition

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

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

où les virgules se lisent « et » :

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

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

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

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 choix de (Ω,P) dans le paradoxe des anniversaires)

Indépendance – Propriétés (admises)

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 les fonctions de « paquets disjoints » de variables sont indépendantes : 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.*
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 $X_1, \dots, X_n$ 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,

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

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,

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

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

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

Inégalité de Tchebychev pour la loi $\mathcal{B}(n,p)$

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

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

Appliquons l'inégalité de Tchebychev à S_n : pour tout $\delta > 0$,

$$P(|S_n - np| \geq \delta) \leq \frac{np(1-p)}{\delta^2}.$$

Inégalité de Tchebychev pour la loi $\mathcal{B}(n,p)$

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

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

Appliquons l'inégalité de Tchebychev à S_n : pour tout $\delta > 0$,

$$P(|S_n - np| \geq \delta) \leq \frac{np(1-p)}{\delta^2}.$$

Pour $\delta = n\varepsilon$,

$$P(|S_n - np| \geq n\varepsilon) \leq \frac{np(1-p)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0.$$

Inégalité de Tchebychev pour la loi $\mathcal{B}(n,p)$

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

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

Appliquons l'inégalité de Tchebychev à S_n : pour tout $\delta>0$,

$$P(|S_n-np|\geq\delta)\leq\frac{np(1-p)}{\delta^2}.$$

Pour $\delta=n\varepsilon$,

$$P\left(\left|\frac{S_n}{n}-p\right|\geq\varepsilon\right)=P(|S_n-np|\geq n\varepsilon)\leq\frac{np(1-p)}{n^2\varepsilon^2}=\frac{p(1-p)}{n\varepsilon^2}\underset{n\to\infty}{\longrightarrow}0.$$

$\rightsquigarrow$ La proportion de "succès" $\frac{S_n}{n}=\frac{\mathbf{1}_{A_1}+\cdots+\mathbf{1}_{A_n}}{n}$ est proche de p avec grande probabilité, si le nombre n est grand. C'est un cas particulier de la **loi des grands nombres.**

Exemple : si $n=1000$ et $p=\frac{1}{2}$, avec probab. $\geq 75\%$, $S_n\in[0.468,0.531]$.

(On verra plus tard comment raffiner cet intervalle)

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 :

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

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 $A_1, \ldots, A_n$.

Application : Simulation stochastique

Principe de la simulation aléatoire (ou “stochastique”)

La “fonction” `rand()` de 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, toutes de loi uniforme sur $[0,1]$, et un élément $\omega \in \Omega$, tels que la fonction `rand()` renvoie d’abord $U_1(\omega)$, puis $U_2(\omega)$, puis...

• 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. Dans Matlab, `rng(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, il faut les construire à partir de $U_1, U_2, \ldots$

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

Calcul espérance/variance : collectionneur d'images

Dans chaque paquet de céréales, on trouve une image. Il existe en tout $N = 50$ images différentes. Combien de paquets faut-il ouvrir pour en avoir (au moins) une de chaque type ?

Calcul espérance/variance : collectionneur d'images

Dans chaque paquet de céréales, on trouve une image. Il existe en tout $N = 50$ images différentes. Combien de paquets faut-il ouvrir pour en avoir (au moins) une de chaque type ?

On note $U_1, U_2, \ldots$ le numéro de l'image du paquet 1, du paquet 2,… Alors $U_1, U_2, \ldots$ sont des variables aléatoires indépendantes, et de loi uniforme dans $\{1, 2, \ldots, N\}$. Notons X le nombre de paquets à ouvrir.

Calcul espérance/variance : collectionneur d'images

Dans chaque paquet de céréales, on trouve une image. Il existe en tout $N = 50$ images différentes. Combien de paquets faut-il ouvrir pour en avoir (au moins) une de chaque type ?

On note $U_1, U_2, \ldots$ le numéro de l'image du paquet 1, du paquet 2,… Alors $U_1, U_2, \ldots$ sont des variables aléatoires indépendantes, et de loi uniforme dans $\{1, 2, \ldots, N\}$. Notons X le nombre de paquets à ouvrir.
On peut décomposer $X = K_1 + K_2 + \cdots + K_N$, où K_i est le nombre de nouveaux paquets à ouvrir pour avoir i images différentes, quand on en a déjà $i - 1$ différentes (avec $K_1 = 1$). Quelle est la loi de K_i ?

Calcul espérance/variance : collectionneur d'images

Dans chaque paquet de céréales, on trouve une image. Il existe en tout $N = 50$ images différentes. Combien de paquets faut-il ouvrir pour en avoir (au moins) une de chaque type ?

On note $U_1, U_2, \dots$ le numéro de l'image du paquet 1, du paquet 2,...
Alors $U_1, U_2, \dots$ sont des variables aléatoires indépendantes, et de loi uniforme dans $\{1, 2, \dots, N\}$. Notons X le nombre de paquets à ouvrir.
On peut décomposer $X = K_1 + K_2 + \cdots + K_N$, où K_i est le nombre de nouveaux paquets à ouvrir pour avoir i images différentes, quand on en a déjà $i-1$ différentes (avec $K_1 = 1$). Quelle est la loi de K_i ?
K_2 suit la loi géométrique de paramètre $\frac{N-1}{N}$.
K_3 suit la loi géométrique de paramètre $\frac{N-2}{N}$.
...
K_N suit la loi géométrique de paramètre $\frac{1}{N}$. Donc

$$E[X] = E[K_1] + E[K_2] + \cdots + E[K_N] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1}$$
$$= N\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}\right) \simeq 225 \qquad (\text{si } N = 50)$$

Calcul espérance/variance : collectionneur d'images

On note $U_1, U_2, \ldots$ le numéro de l'image du paquet 1, du paquet 2,…
Alors $U_1, U_2, \ldots$ sont des variables aléatoires indépendantes, et de loi uniforme dans $\{1, 2, \ldots, N\}$. Notons X le nombre de paquets à ouvrir.
On peut décomposer $X = K_1 + K_2 + \cdots + K_N$, où K_i est le nombre de nouveaux paquets à ouvrir pour avoir i images différentes, quand on en a déjà $i-1$ différentes (avec $K_1 = 1$). Quelle est la loi de K_i ?
K_2 suit la loi géométrique de paramètre $\frac{N-1}{N}$.
K_3 suit la loi géométrique de paramètre $\frac{N-2}{N}$.
…
K_N suit la loi géométrique de paramètre $\frac{1}{N}$. Donc

$$E[X] = E[K_1] + E[K_2] + \cdots + E[K_N] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{1}$$

$$= N\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}\right) \simeq 225 \qquad (\text{si } N = 50)$$

et $K_1, \ldots, K_n$ sont indépendants, donc

$$\mathrm{Var}(X) = \mathrm{Var}(K_1) + \cdots + \mathrm{Var}(K_N) = 0 + \frac{\frac{1}{N}}{(1 - \frac{1}{N})^2} + \cdots + \frac{\frac{N-1}{N}}{(1 - \frac{N-1}{N})^2} \simeq 62^2$$

$\rightsquigarrow$ (Par inég. de Tchebychev) Avec probabilité $\geq 75\%$, $101 \leq X \leq 349$.

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

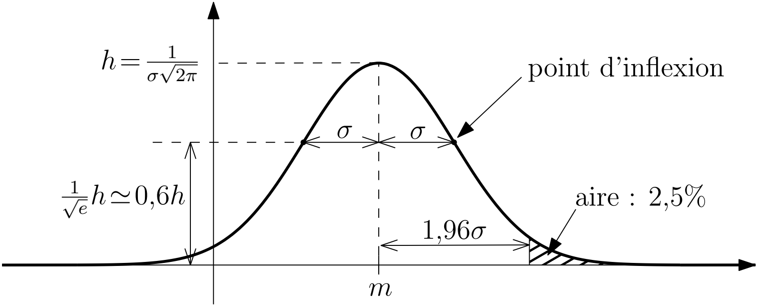
(avec une erreur relative inférieure à 0,2 si $x > 1{,}9$)
Si $X \sim \mathcal{N}(m,\sigma^2)$, on pose $Z = \frac{X-m}{\sigma}$ pour se ramener à $\mathcal{N}(0,1)$.

Courbe en cloche : densité de la loi $\mathcal{N}(m,\sigma^2)$

La densité de la loi $\mathcal{N}(m,\sigma^2)$ est

$$f : x \mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Cette fonction est appelée une *gaussienne* ou "*courbe en cloche*".



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

Proposition

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

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

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

où

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

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

Autre 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 :

$Z \setminus X$	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	$1/16$

Autre 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

Cours 4

–

Lundi 25 février 2019

Rappel – Bilan

On a vu jusque-là comment étudier *une seule* variable aléatoire X :

- X est une fonction $\Omega \to \mathbb{R}$
- son ensemble de valeurs possibles (ou *support*) est son image $X(\Omega)$
- sa loi est la donnée de $P_X(A) = P(X \in A)$, pour tout $A \subset \mathbb{R}$
  - cas discret : équivaut à $P(X = x)$ pour tous les $x \in X(\Omega)$
  - cas à densité f_X : équivaut à $f_X(x)$ pour tous les $x \in \mathbb{R}$
- on peut aussi se donner sa loi par sa *fonction de répartition*

$$F_X : x \mapsto F_X(x) = P(X \leq x).$$

- On peut calculer des espérances $E[\phi(X)]$ pour $\phi : \mathbb{R} \to \mathbb{R}$:
  - cas discret : $E[\phi(X)] = \sum_x \phi(x) P(X = x)$
  - cas à densité f_X : $E[\phi(X)] = \int_{-\infty}^{+\infty} \phi(x) f_X(x) dx.$

  (si la série/intégrale converge)

Rappel – Bilan

On a aussi vu comment étudier *plusieurs* variables aléatoires X,Y **indépendantes** :

- X et Y sont, chacunes, des fonctions $\Omega \to \mathbb{R}$
- (X,Y) peut prendre toute valeur dans $X(\Omega) \times Y(\Omega)$ (pas d'influence)
- On sait calculer des probabilités du type

$$P(X \in A \text{ et } Y \in B) = P(X \in A)P(Y \in B)$$

- On sait calculer des espérances du type

$$E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)]$$

  (si elles convergent), et en particulier $\mathrm{Cov}(X,Y) = 0$.

Et pour calculer, disons $P(XY > 0)$ ou $E[(X+Y)^2]$, on se ramène aux calculs ci-dessus. Cela a permis aussi de démontrer la **loi des grands nombres.**

Rappel – Bilan

On a aussi vu comment étudier *plusieurs* variables aléatoires X, Y **indépendantes** :

- X et Y sont, chacunes, des fonctions $\Omega \to \mathbb{R}$
- (X, Y) peut prendre toute valeur dans $X(\Omega) \times Y(\Omega)$ (pas d'influence)
- On sait calculer des probabilités du type

$$P(X \in A \text{ et } Y \in B) = P(X \in A)P(Y \in B)$$

- On sait calculer des espérances du type

$$E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)]$$

  (si elles convergent), et en particulier $\mathrm{Cov}(X, Y) = 0$.

Et pour calculer, disons $P(XY > 0)$ ou $E[(X+Y)^2]$, on se ramène aux calculs ci-dessus. Cela a permis aussi de démontrer la **loi des grands nombres.**
Mais en général, les variables dans une expérience ne sont pas indépendantes (ni de simples fonctions $Y = f(X)$)... Pour étudier ces corrélations, on a besoin d'étudier le **couple** (X, Y) et en particulier sa loi.

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

Y \ X	1	2	3	4	5	6	Total
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

Y \ X	1	2	3	4	5	6	Total
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

$\boxed{X,Y \text{ indépendantes}}$

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X=i \text{ et } Y=j)$, pour i,j entre 1 et 6.

Y \ X	1	2	3	4	5	6	Total
1	$1/6$	0	0	0	0	0	$1/6$
2	0	$1/6$	0	0	0	0	$1/6$
3	0	0	$1/6$	0	0	0	$1/6$
4	0	0	0	$1/6$	0	0	$1/6$
5	0	0	0	0	$1/6$	0	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

Y \ X	1	2	3	4	5	6	Total
1	$1/6$	0	0	0	0	0	$1/6$
2	0	$1/6$	0	0	0	0	$1/6$
3	0	0	$1/6$	0	0	0	$1/6$
4	0	0	0	$1/6$	0	0	$1/6$
5	0	0	0	0	$1/6$	0	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

$\boxed{Y = X}$

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

$Y \backslash X$	1	2	3	4	5	6	Total
1	0	0	0	0	0	$1/6$	$1/6$
2	0	0	0	0	$1/6$	0	$1/6$
3	0	0	0	$1/6$	0	0	$1/6$
4	0	0	$1/6$	0	0	0	$1/6$
5	0	$1/6$	0	0	0	0	$1/6$
6	$1/6$	0	0	0	0	0	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X, Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

$Y \backslash X$	1	2	3	4	5	6	Total
1	0	0	0	0	0	$1/6$	$1/6$
2	0	0	0	0	$1/6$	0	$1/6$
3	0	0	0	$1/6$	0	0	$1/6$
4	0	0	$1/6$	0	0	0	$1/6$
5	0	$1/6$	0	0	0	0	$1/6$
6	$1/6$	0	0	0	0	0	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

$\boxed{Y = 7 - X}$

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

Y \ X	1	2	3	4	5	6	Total
1	$1/12$	$1/12$	0	0	0	0	$1/6$
2	$1/12$	$1/12$	0	0	0	0	$1/6$
3	0	0	$1/6$	0	0	0	$1/6$
4	0	0	0	$1/6$	0	0	$1/6$
5	0	0	0	0	$1/6$	0	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X = i \text{ et } Y = j)$, pour i,j entre 1 et 6.

$Y \backslash X$	1	2	3	4	5	6	Total
1	$1/12$	$1/12$	0	0	0	0	$1/6$
2	$1/12$	$1/12$	0	0	0	0	$1/6$
3	0	0	$1/6$	0	0	0	$1/6$
4	0	0	0	$1/6$	0	0	$1/6$
5	0	0	0	0	$1/6$	0	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

X dé, $Z \in \{1,2\}$ pièce indépendante ; $Y = \begin{cases} X & \text{si } X \geq 3, \\ Z & \text{si } X \in \{1,2\} \end{cases}$

Loi du couple : des exemples discrets

Pour décrire la loi de deux variables aléatoires X,Y à valeurs dans $\{1,2,3,4,5,6\}$, il suffit de connaître $P(X=i \text{ et } Y=j)$, pour i,j entre 1 et 6.

Y \ X	1	2	3	4	5	6	Total
1	$1/12$	$1/12$	0	0	0	0	$1/6$
2	$1/12$	$1/12$	0	0	0	0	$1/6$
3	0	0	$1/6$	0	0	0	$1/6$
4	0	0	0	$1/6$	0	0	$1/6$
5	0	0	0	0	$1/6$	0	$1/6$
6	0	0	0	0	0	$1/6$	$1/6$
Total	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

$$\boxed{X \text{ dé, } Z \in \{1,2\} \text{ pièce indépendante ; } Y = \begin{cases} X & \text{si } X \geq 3, \\ Z & \text{si } X \in \{1,2\} \end{cases}}$$

Etc. : si on connaît la loi de X et celle de Y, on ne connaît pas la loi de (X,Y), celle-ci donne le lien entre les valeurs de X et de Y à un **même** tirage.

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 :*

$$\text{pour tout } C \subset \mathbb{R}^2, \qquad P_{(X,Y)}(C) = P((X,Y) \in C)$$

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

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 de (X,Y) est fournie par celles de X et 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 à un même tirage (connaître X peut renseigner sur Y).

Cas de deux variables discrètes

Si X et Y sont discrètes alors la loi de (X,Y) est donnée par le **tableau des 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

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

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 ssi $p_{(X,Y)}(x,y) = p_X(x)p_Y(y)$ 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

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

NB. En pratique, le calcul d'intégrale double se ramène à deux intégrales simples $\int(\int f_{(X,Y)}(x,y)dx)dy$, où les bornes peuvent dépendre du point y

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

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

Interprétation de la densité

On rappelle que, si X a pour densité f_X alors

$$f_X(x) = \lim_{\delta \to 0} \frac{P(X \in [x-\delta, x+\delta])}{\text{longueur}([x-\delta, x+\delta])} = \lim_{\delta \to 0} \frac{P(X \in [x-\delta, x+\delta])}{2\delta},$$

si f_X est continue en x.

De façon similaire,

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

$$f_{X,Y}(x,y) = \lim_{\delta \to 0} \frac{P\big((X,Y) \in D((x,y),\delta)\big)}{\text{aire}(D((x,y),\delta))} = \lim_{\delta \to 0} \frac{P\big((X,Y) \in D((x,y),\delta)\big)}{\pi\delta^2},$$

si $f_{(X,Y)}$ est continue en (x,y).

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.

Exemple à densité

Par définition, la **loi uniforme sur le disque** $D(0,r)$ est la loi d'un couple (X,Y) de densité

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

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

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

Et Y a même loi que X.
NB : X,Y ne sont pas indépendantes car on sait que $(X,Y) \in D(0,r)$.
(on peut aussi voir que $f_{(X,Y)}(x,y) \neq f_X(x)f_Y(y)$)

Autre exemple

Soit (U,V) un couple de variables aléatoire de densité

$$f(u,v) = Ce^{-2(u+v)}\mathbf{1}_D(u,v),$$

où $D = \{(u,v) \in \mathbb{R}^2 \,|\, 0 \leq u \leq v\}$, et C est un réel à déterminer.

Autre exemple

Soit (U,V) un couple de variables aléatoire de densité

$$f(u,v) = Ce^{-2(u+v)}\mathbf{1}_D(u,v),$$

où $D = \{(u,v) \in \mathbb{R}^2 \mid 0 \leq u \leq v\}$, et C est un réel à déterminer.
Alors U a pour densité $f_U(u) = \int_{\mathbb{R}} f(u,v)dv$ donc $f_U(u) = 0$ si $u < 0$ et, si $u > 0$,

$$f_U(u) = C\int_u^{\infty} e^{-2u-2v}dv = \frac{C}{2}e^{-4u},$$

donc U suit la loi $\mathcal{E}(4)$ et $C = 8$.

Autre exemple

Soit (U,V) un couple de variables aléatoire de densité

$$f(u,v) = Ce^{-2(u+v)}\mathbf{1}_D(u,v),$$

où $D = \{(u,v) \in \mathbb{R}^2 \,|\, 0 \leq u \leq v\}$, et C est un réel à déterminer.

Alors U a pour densité $f_U(u) = \displaystyle\int_{\mathbb{R}} f(u,v)dv$ donc $f_U(u) = 0$ si $u < 0$ et, si $u > 0$,

$$f_U(u) = C\int_u^{\infty} e^{-2u-2v}dv = \frac{C}{2}e^{-4u},$$

donc U suit la loi $\mathcal{E}(4)$ et $C = 8$.

Et V a pour densité $f_V(v) = \displaystyle\int_{\mathbb{R}} f(u,v)du$ donc $f_V(v) = 0$ si $v < 0$ et, si $v > 0$,

$$f_V(v) = 8\int_0^{v} e^{-2u-2v}du = 4(e^{-2v} - e^{-4v}).$$

Autre exemple

Soit (U,V) un couple de variables aléatoire de densité

$$f(u,v) = Ce^{-2(u+v)}\mathbf{1}_D(u,v),$$

où $D = \{(u,v) \in \mathbb{R}^2 \,|\, 0 \leq u \leq v\}$, et C est un réel à déterminer.
Alors U a pour densité $f_U(u) = \displaystyle\int_{\mathbb{R}} f(u,v)dv$ donc $f_U(u) = 0$ si $u < 0$ et, si $u > 0$,

$$f_U(u) = C\int_u^{\infty} e^{-2u-2v}dv = \frac{C}{2}e^{-4u},$$

donc U suit la loi $\mathcal{E}(4)$ et $C = 8$.
Et V a pour densité $f_V(v) = \displaystyle\int_{\mathbb{R}} f(u,v)du$ donc $f_V(v) = 0$ si $v < 0$ et, si $v > 0$,

$$f_V(v) = 8\int_0^{v} e^{-2u-2v}du = 4(e^{-2v} - e^{-4v}).$$

NB. U et V ne sont pas indépendantes. $((U,V) \in D$, ou $f(u,v) \neq f_U(u)f_V(v))$

Loi d'un couple : Bilan

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

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

Calculs d'espérances

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

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

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

Autre exemple de calcul (moins astucieux)

Pour le couple (U,V) précédent, calculer $E[e^{U+V}]$.

$$\begin{aligned}
E\left[e^{U+V}\right] &= \iint_{\mathbb{R}^2} e^{u+v} f(u,v)\,du\,dv \\
&= \iint_{D} e^{u+v} 8e^{-2(u+v)}\,du\,dv \\
&= \int_0^{\infty} \left(\int_u^{\infty} 8e^{-u}e^{-v} dv \right) du \\
&= 8 \int_0^{\infty} \left(e^{-u} \int_u^{\infty} e^{-v} dv \right) du \\
&= 8 \int_0^{\infty} e^{-u}e^{-u} du = 8\frac{1}{2} = 4.
\end{aligned}$$