

# Probabilités

Alexis Bienvenüe, Alexis.Bienvenue@univ-lyon1.fr

Vincent Lerouillois

- ▶ Probabilités : contrôle continu par QCM  
<https://clarolineconnect.univ-lyon1.fr/>

# Introduction

*Probabilités* : modélisation du hasard.

*Statistique* : caractérisation à partir des observations.

Modélisation et évaluation du risque.

# Plan

## Espaces probabilisés discrets

- Espace probabilisé

- Probabilité conditionnelle et indépendance

- Éléments aléatoires

## Théorie générale des probabilités

## Manipulation des vecteurs aléatoires

## Statistiques descriptives

# Espace probabilisé

Modélisation :

- ▶ **Expérience** : phénomène aléatoire étudié
- ▶ **Univers**  $\Omega$  : ensemble des **éventualités** (ou **états de la nature**). Quoi qu'il arrive à l'issue de l'expérience, le résultat de l'expérience doit pouvoir être associé à une et une seule éventualité.
- ▶ **Événement** : collection d'éventualités (qui peut être vide ou ne contenir qu'une éventualité), c'est-à-dire sous-ensemble de l'univers.

Quelques événements particuliers :

- ▶ L'**événement impossible**  $\emptyset$
- ▶ L'**événement certain**  $\Omega$
- ▶ Les **événements élémentaires**, contenant exactement une éventualité.

Correspondance entre les opérateurs logiques et ensemblistes : si  $A_1, A_2, \dots$  sont des événements d'un même univers  $\Omega$ ,

- ▶  $A_1 \cap A_2$  est réalisé si  $A_1$  et  $A_2$  sont réalisés ;
- ▶  $A_1 \cup A_2$  est réalisé si  $A_1$  ou  $A_2$  sont réalisés (ou les deux) ;
- ▶  $\bar{A}_1 = \Omega \setminus A_1$  (le complémentaire de  $A_1$ ) est l'événement contraire de  $A_1$  ;
- ▶ deux événements  $A_1$  et  $A_2$  sont dits incompatibles s'ils sont disjoints, c'est-à-dire si  $A_1 \cap A_2 = \emptyset$ .
- ▶ «  $A_1$  implique  $A_2$  » s'écrit  $A_1 \subset A_2$ , car dans ce cas  $\omega \in A_1 \Rightarrow \omega \in A_2$ .
- ▶ « Un nombre infini de  $A_k$  se réalisent » est représenté par

$$\bigcap_n \bigcup_{k \geq n} A_k = \limsup_n A_n ;$$

- ▶ « Les  $A_k$  se réalisent tous à partir d'un certain rang » est représenté par

$$\bigcup_n \bigcap_{k \geq n} A_k = \liminf_n A_n .$$

Jojo achète un billet de l'euro-million.

$\Omega = \{0, 1\}$ , avec la notation

- ▶  $1 = \ll \text{Jojo gagne le gros lot} \gg$
- ▶  $0 = \ll \text{Jojo ne gagne pas le gros lot} \gg$

Jojo lance sa pièce de 2€.

$\Omega = \{0, 1\}$ , avec la notation

- ▶  $1 = \ll \text{la pièce tombe sur pile} \gg$
- ▶  $0 = \ll \text{la pièce tombe sur face} \gg$

## Définition (Probabilité)

On appelle **probabilité** sur un univers discret (c'est-à-dire dénombrable ou fini)  $\Omega$  une application  $\mathbb{P}$  de  $\mathcal{P}(\Omega)$  dans  $[0, 1]$  telle que :

- ▶  $\mathbb{P}(\Omega) = 1$  ;
- ▶ pour toute suite  $(A_i)_{i \geq 1}$  d'événements deux à deux incompatibles,

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

Le couple  $(\Omega, \mathbb{P})$  sera alors appelé **espace probabilisé**.

## Lemme

Si  $\mathbb{P}$  est une probabilité sur un univers discret  $\Omega$ , et si  $A$ ,  $B$  et les  $A_n$  sont des événements, alors :

1.  $\mathbb{P}(\emptyset) = 0$ ;
2.  $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$ ;
3.  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ ;
4.  $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ ;
5. Si  $(A_n)_n$  est une suite croissante (i.e.  $A_n \subset A_{n+1}$ ), alors

$$\mathbb{P}\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n);$$

6. Si  $(A_n)_n$  est une suite décroissante (i.e.  $A_n \supset A_{n+1}$ ), alors

$$\mathbb{P}\left(\bigcap_{n=1}^{+\infty} A_n\right) = \lim_{n \rightarrow +\infty} \mathbb{P}(A_n);$$



## Lemme (suite)

7. On a :

$$\mathbb{P}\left(\liminf_n A_n\right) \leq \liminf_n \mathbb{P}(A_n), \quad \mathbb{P}\left(\limsup_n A_n\right) \geq \limsup_n \mathbb{P}(A_n);$$

8. Formule de *Poincaré* (appelée aussi principe d'inclusion-exclusion) :

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \dots \\ &+ (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n), \end{aligned}$$

De plus cette somme est alternée (inégalités de *Bonferroni*).

Un exemple important de probabilité :

## Définition (Probabilité uniforme)

Soit  $\Omega$  un ensemble **fini**. La **probabilité uniforme** (ou **équiprobabilité**) sur  $\Omega$  est l'unique probabilité  $\mathbb{P}$  qui prend la même valeur sur tous les événements élémentaires. Elle s'écrit

$$\mathbb{P}(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)} \quad \forall A \subset \Omega.$$

Comment modéliser les expériences suivantes ?

- ▶ On choisit un nombre au hasard entre 1 et 6.
- ▶ On choisit un entier naturel au hasard.

**Attention** : il n'existe pas de probabilité uniforme sur un ensemble dénombrable infini !

## Caractérisation des probabilités sur un univers discret

Si  $\Omega$  est dénombrable, toute probabilité  $\mathbb{P}$  sur  $\Omega$  est entièrement définie par la donnée des probabilités des événements élémentaires :  $p_\omega = \mathbb{P}(\{\omega\})$ , de la façon suivante :

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega \quad \forall A \subset \Omega.$$

De plus, si  $(p_\omega)_{\omega \in \Omega}$  est une famille de réels de  $[0, 1]$  telle que

$$\sum_{\omega \in \Omega} p_\omega = 1,$$

alors il existe une et une seule probabilité  $\mathbb{P}$  telle que  $\mathbb{P}(\{\omega\}) = p_\omega$  pour tout  $\omega$ .

**Exemple :** la probabilité de Poisson de paramètre  $\lambda > 0$  est la probabilité  $\mathbb{P}$  sur  $\mathbb{N}$  définie par :

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(\{n\}) = p_n = \frac{e^{-\lambda} \lambda^n}{n!}.$$

## Définition

Un événement  $A$  d'un espace probabilisé  $(\Omega, \mathbb{P})$  est dit **presque-impossible** ou **négligeable** si  $\mathbb{P}(A) = 0$ . Il est dit **presque-sûr** si  $\mathbb{P}(A) = 1$ , c'est-à-dire si son complémentaire est presque-impossible.

**Dans le cadre discret**, si on pose

$$\Omega_0 = \{\omega \in \Omega, \mathbb{P}(\{\omega\}) = 0\},$$

alors un événement  $A \subset \Omega$  est presque-impossible si et seulement si il est inclus dans  $\Omega_0$ , et  $A$  est presque-sûr si et seulement si  $\bar{\Omega}_0 \subset A$ .

# Probabilité conditionnelle et indépendance

Jojo lance deux dés, un rouge et un bleu. On modélise cette expérience par l'univers  $\Omega = \{1, \dots, 6\}^2$  muni de la probabilité uniforme  $\mathbb{P}$ , l'élément  $(i, j) \in \Omega$  représentant l'éventualité « le dé rouge a donné  $i$  et le dé bleu a donné  $j$  ».

	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- Utilisons par exemple le même univers

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- ▶ Utilisons par exemple le même univers
- ▶ La nouvelle probabilité a utiliser doit être nulle hors de  $B = \ll \text{la somme des deux dés est 4 ou 5} \gg$ .

	1	2	3	4	5	6
1	0	0	?	?	0	0
2	0	?	?	0	0	0
3	?	?	0	0	0	0
4	?	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- ▶ Utilisons par exemple le même univers
- ▶ La nouvelle probabilité à utiliser doit être nulle hors de  $B = \ll \text{la somme des deux dés est 4 ou 5} \gg$ .
- ▶ Dans  $B$ , les éventualités doivent rester équiprobables

	1	2	3	4	5	6
1	0	0	x	x	0	0
2	0	x	x	0	0	0
3	x	x	0	0	0	0
4	x	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0



On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- ▶ Utilisons par exemple le même univers
- ▶ La nouvelle probabilité à utiliser doit être nulle hors de  $B = \ll \text{la somme des deux dés est 4 ou 5} \gg$ .
- ▶ Dans  $B$ , les éventualités doivent rester équiprobables

	1	2	3	4	5	6
1	0	0	$\frac{1}{7}$	$\frac{1}{7}$	0	0
2	0	$\frac{1}{7}$	$\frac{1}{7}$	0	0	0
3	$\frac{1}{7}$	$\frac{1}{7}$	0	0	0	0
4	$\frac{1}{7}$	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0

Jojo lance deux dés, un rouge et un bleu. On modélise cette expérience par l'univers  $\Omega = \{2, \dots, 12\}$  muni de la probabilité  $\mathbb{P}$ , l'élément  $i \in \Omega$  représentant l'éventualité « la somme du résultat des deux dés est  $i$  ». La probabilité  $\mathbb{P}$  adaptée est définie de la manière suivante :

$i$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\{i\})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- Utilisons par exemple le même univers, muni de la probabilité  $\mathbb{P}'$ .

$i$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\{i\})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\mathbb{P}'(\{i\})$											

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- ▶ Utilisons par exemple le même univers, muni de la probabilité  $\mathbb{P}'$ .
- ▶ Cette probabilité ne doit pas donner de poids en-dehors de  $B = \ll \text{la somme du résultat des deux dés est 4 ou 5} \gg = \{4, 5\}$

$i$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\{i\})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\mathbb{P}'(\{i\})$	0	0	?	?	0	0	0	0	0	0	0

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- ▶ Utilisons par exemple le même univers, muni de la probabilité  $\mathbb{P}'$ .
- ▶ Cette probabilité ne doit pas donner de poids en-dehors de  $B = \ll \text{la somme du résultat des deux dés est 4 ou 5} \gg = \{4, 5\}$
- ▶ Le rapport entre les poids des événements  $\{4\}$  et  $\{5\}$  doit rester inchangé

$i$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\{i\})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\mathbb{P}'(\{i\})$	0	0	x	$\frac{4}{3}x$	0	0	0	0	0	0	0

On sait en plus que la somme des deux dés lancés par Jojo est 4 ou 5. Comment modéliser cette nouvelle expérience ?

- ▶ Utilisons par exemple le même univers, muni de la probabilité  $\mathbb{P}'$ .
- ▶ Cette probabilité ne doit pas donner de poids en-dehors de  $B = \ll \text{la somme du résultat des deux dés est 4 ou 5} \gg = \{4, 5\}$
- ▶ Le rapport entre les poids des événements  $\{4\}$  et  $\{5\}$  doit rester inchangé

$i$	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(\{i\})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
$\mathbb{P}'(\{i\})$	0	0	$\frac{3}{7}$	$\frac{4}{7}$	0	0	0	0	0	0	0

## Définition (Probabilité conditionnelle)

Soit  $(\Omega, \mathbb{P})$  un espace probabilisé discret,  $A$  et  $B$  deux événements tels que  $\mathbb{P}(B) \neq 0$ . On définit la **probabilité conditionnelle** de  $A$  sachant  $B$ , notée  $\mathbb{P}(A | B)$ , par :

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

On vérifie que

$$\begin{aligned} \mathbb{P}(\cdot | B) : \mathcal{P}(\Omega) &\longrightarrow [0, 1] \\ A &\longmapsto \mathbb{P}(A | B) \end{aligned}$$

est une nouvelle probabilité sur  $\Omega$  qui s'annule sur les événements incompatibles avec  $B$ .

Si  $(\Omega, \mathbb{P})$  modélise une certaine expérience, alors  $(\Omega, \mathbb{P}(\cdot | B))$  modélise cette expérience à laquelle on a rajouté l'hypothèse/la connaissance supplémentaire de la réalisation de  $B$ .

## Formule des probabilités totales

Soit  $I$  un ensemble d'indices fini ou dénombrable, et  $\{B_n\}_{n \in I}$  une partition de  $\Omega$  par des événements de probabilité non nulle. Alors

$$\forall A \subset \Omega, \quad \mathbb{P}(A) = \sum_{n \in I} \mathbb{P}(A | B_n) \mathbb{P}(B_n).$$

## Formule des probabilités composées

Soit  $(\Omega, \mathbb{P})$  un espace probabilisé et  $A_1, \dots, A_n$  des événements tels que  $\mathbb{P}(A_1 \cap \dots \cap A_n) \neq 0$ . Alors,

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 \cap A_2) \cdots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}).$$



# Indépendance

*On dit que deux événements  $A$  et  $B$  sont indépendants si la connaissance ou non de la réalisation de l'un n'affecte pas la probabilité que l'autre se réalise.*

## Définition (Indépendance de deux événements)

On dit que deux événements  $A$  et  $B$  d'un même espace probabilisé  $(\Omega, \mathbb{P})$  sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B).$$

## Propriétés

- ▶ Tout événement  $A$  est indépendant de  $\Omega$  et de  $\emptyset$ .
- ▶ Si  $A$  et  $B$  sont indépendants, alors  $A$  et  $\bar{B}$  le sont, de même que  $\bar{A}$  et  $B$ , ainsi que  $\bar{A}$  et  $\bar{B}$ .
- ▶ Si  $A$  et  $A'$  sont tous les deux indépendants de  $B$ , alors l'indépendance de  $A \cap A'$  et  $B$  est équivalente à l'indépendance de  $A \cup A'$  et  $B$ .

## Définition (Indépendance mutuelle d'événements)

Une famille  $(A_i)_{i \in I}$  est dite **indépendante dans son ensemble** – ou les  $(A_i)_{i \in I}$  sont dits **mutuellement indépendants** – si pour tout ensemble fini d'indices  $F \subset I$ , on a :

$$\mathbb{P} \left( \bigcap_{i \in F} A_i \right) = \prod_{i \in F} \mathbb{P}(A_i).$$

**Remarque :** Si les  $(A_i)_{i \in I}$  sont mutuellement indépendants, alors ils sont deux à deux indépendants (prendre  $F = \{i, j\}$ ), mais le contraire n'est pas vrai.

## Définition

Deux familles  $\mathcal{F}_1$  et  $\mathcal{F}_2$  d'événements sont dites **indépendantes** si

$$\forall A_1 \in \mathcal{F}_1, \forall A_2 \in \mathcal{F}_2, \text{ on a } \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

## Définition

Les familles d'événements  $\mathcal{F}_i$  d'une collection  $(\mathcal{F}_i)_{i \in I}$  sont dites **mutuellement indépendantes** si toute famille d'événements  $\mathcal{F}^* = (A_i)_{i \in I}$  telle que  $A_i \in \mathcal{F}_i$  pour tout  $i \in I$  est indépendante dans son ensemble.

# Espace probabilisé produit

Considérons  $n$  expériences modélisées par les espaces probabilisés

$$(\Omega_i, \mathbb{P}_i)_{i=1, \dots, n}.$$

Si ces expériences sont **indépendantes**, on peut les modéliser de manière conjointe par l'**espace probabilisé produit**

$$(\Omega, \mathbb{P}) = \left( \prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \mathbb{P}_i \right)$$

défini par

$$\mathbb{P}(\{(\omega_1, \dots, \omega_n)\}) = \prod_{i=1}^n \mathbb{P}_i(\{\omega_i\}) \quad \forall (\omega_1, \dots, \omega_n) \in \Omega.$$

- ▶ Cette dernière égalité définit bien une probabilité.
- ▶ Dans cet espace, les différentes expériences sont indépendantes.

# Éléments aléatoires

## Définition

Soit  $T$  un ensemble quelconque. Si  $(\Omega, \mathbb{P})$  est un espace probabilisé discret, on appelle **élément aléatoire** défini sur  $(\Omega, \mathbb{P})$  et à valeurs dans  $T$ , toute application  $X : \Omega \rightarrow T$ .

Pour  $B \subset T$ , l'événement  $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$  pourra être noté  $\{X \in B\}$  et sa probabilité  $\mathbb{P}(X^{-1}(B))$  pourra être notée  $\mathbb{P}[X \in B]$ .

- ▶ Si  $T = \mathbb{R}$ , on dit que  $X$  est une **variable aléatoire** réelle (v.a.r.).
- ▶ Si  $T = \mathbb{R}^k$ , on dit que  $X$  est un **vecteur aléatoire**, et dans ce cas les composantes  $X_i$  de  $X = (X_1, \dots, X_k)$  sont des v.a.r.
- ▶ De même, on peut définir une **fonction aléatoire** ou tout autre objet mathématique aléatoire (graphes, réseaux, formes, etc.).
- ▶ Si l'image de  $X$  dans  $T$  est fini ou dénombrable (c'est toujours le cas si  $\Omega$  est dénombrable), on dit que  $X$  est un **élément aléatoire discret**.

# Éléments aléatoires

## Définition

Soit  $T$  un ensemble quelconque. Si  $(\Omega, \mathbb{P})$  est un espace probabilisé discret, on appelle **élément aléatoire** défini sur  $(\Omega, \mathbb{P})$  et à valeurs dans  $T$ , toute application  $X : \Omega \rightarrow T$ .

Pour  $B \subset T$ , l'événement  $X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\}$  pourra être noté  $\{X \in B\}$  et sa probabilité  $\mathbb{P}(X^{-1}(B))$  pourra être notée  $\mathbb{P}[X \in B]$ .

## Exemples fondamentaux :

- ▶ Une fonction constante  $f \equiv a$  est un élément aléatoire, même si dans ce cas il n'y a plus d'aléatoire dans la valeur de  $f$ .
- ▶ Si  $A \subset \Omega$  est un événement, alors la fonction indicatrice  $\mathbf{1}_A : \Omega \rightarrow \{0, 1\}$  définie par

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

est une variable aléatoire réelle.

# Loi de probabilité d'un élément aléatoire

## Définition

Soit  $X : (\Omega, \mathbb{P}) \longrightarrow T$  un élément aléatoire discret. On définit l'application  $\mathbb{P}_X$  de  $\mathcal{P}(T)$  dans  $[0, 1]$  par :

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}[X \in B] \quad \forall B \subset T.$$

Alors  $\mathbb{P}_X$  est une probabilité sur  $T$  appelée **loi de probabilité** (ou **loi**, ou **distribution**) de  $X$ .

Donner la loi d'un élément aléatoire  $X$ , c'est :

- ▶ Donner l'ensemble des valeurs que peut prendre  $X$  :  $X(\Omega)$  ;
- ▶ pour chaque  $x \in X(\Omega)$ , donner la probabilité que  $X$  prenne effectivement cette valeur :  $\mathbb{P}_X(\{x\}) = \mathbb{P}[X = x]$ .

# Égalité en loi

## Définition

On dira que deux éléments aléatoires  $X$  et  $Y$  **ont même loi**, ou **sont identiquement distribués**, si leurs lois  $\mathbb{P}_X$  et  $\mathbb{P}_Y$  sont des probabilités identiques. Cette relation sera notée

$$X \sim Y.$$

Pour pouvoir comparer leurs lois, deux éléments aléatoires ne doivent pas forcément être définis sur le même espace probabilisé, mais doivent prendre leurs valeurs dans le même ensemble.

Par contre, pour pouvoir dire  $X = Y$ , il est nécessaire que  $X$  et  $Y$  soient définis sur le même espace de probabilité, et qu'ils prennent leurs valeurs dans le même espace. En effet, on compare ici deux fonctions de  $\Omega$  dans  $T$ .

Par ailleurs, si  $X = Y$ , alors il est évident que  $X \sim Y$ , mais le contraire n'est pas vrai.



# Indépendance d'éléments aléatoires

- ▶ Quelles sont les informations qu'un élément aléatoire peut donner ?
- ▶ Quand dit-on que deux éléments aléatoires sont indépendants ?

## Définition

Deux éléments aléatoires discrets  $X$  et  $Y$  à valeurs dans  $T$  et  $U$  définis sur le même espace probabilisé  $(\Omega, \mathbb{P})$  sont dits **indépendants** si, pour tous sous-ensembles  $A \subset T$  et  $B \subset U$ , les événements  $\{X \in A\}$  et  $\{Y \in B\}$  sont indépendants, c'est-à-dire si les familles  $X^{-1}(\mathcal{P}(T))$  et  $Y^{-1}(\mathcal{P}(U))$  sont indépendantes.

## Caractérisation dans le cas discret

Deux éléments aléatoires **discrets**  $X$  et  $Y$  à valeurs dans  $T$  et  $U$  sont indépendants si, pour tous  $x \in T$  et  $y \in U$ ,  $\{X = x\}$  et  $\{Y = y\}$  sont indépendants.

# Indépendance d'éléments aléatoires

- ▶ Quelles sont les informations qu'un élément aléatoire peut donner ?
- ▶ Quand dit-on que deux éléments aléatoires sont indépendants ?

## Définition

Deux éléments aléatoires discrets  $X$  et  $Y$  à valeurs dans  $T$  et  $U$  définis sur le même espace probabilisé  $(\Omega, \mathbb{P})$  sont dits **indépendants** si, pour tous sous-ensembles  $A \subset T$  et  $B \subset U$ , les événements  $\{X \in A\}$  et  $\{Y \in B\}$  sont indépendants, c'est-à-dire si les familles  $X^{-1}(\mathcal{P}(T))$  et  $Y^{-1}(\mathcal{P}(U))$  sont indépendantes.

## Lien avec la probabilité produit

Deux éléments aléatoires **discrets**  $X$  et  $Y$  sont indépendants si et seulement si

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y .$$

# Espérance

## Définition

Soit  $X$  une variable aléatoire discrète, pouvant prendre les valeurs  $X(\Omega)$ . Si la série  $\sum_{x \in X(\Omega)} |x| \mathbb{P}[X = x]$  converge, on dit que  $X$  est **intégrable**, on pose

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

et on appelle ce nombre **espérance** mathématique de  $X$  ou **moyenne** de  $X$ . Si  $\sum_{x \in X(\Omega)} |x| \mathbb{P}[X = x] = +\infty$ , on dira que  $X$  n'est pas intégrable.

## Remarque

L'espérance d'une variable aléatoire *ne dépend que de sa loi*. On parlera donc aussi d'intégrabilité et d'espérance d'une *loi*.

## Représentation alternative

On a aussi :

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

Plus généralement, si la suite d'événements  $(A_i)_{i \geq 1}$  forme une partition de  $\Omega$  telle que  $X$  est constante et égale à  $x_i$  sur chaque  $A_i$ , alors :

$$\mathbb{E} X = \sum_{i \geq 1} x_i \mathbb{P}(A_i).$$

## Propriétés

Soient  $X$  et  $Y$  deux variables aléatoires intégrables et  $\alpha$  un réel. Alors

- ▶  $E(\alpha X) = \alpha E X$
- ▶  $E(X + Y) = E X + E Y$
- ▶ si  $X$  et  $Y$  sont indépendantes, alors  $E(XY) = E X \times E Y$

# Plan

Espaces probabilisés discrets

**Théorie générale des probabilités**

Tribus

Mesure positive, probabilité

Indépendance, espace probabilisé produit

Mesurabilité des fonctions, éléments aléatoires

Intégrale, espérance

Densité

Manipulation des vecteurs aléatoires

Statistiques descriptives

# Théorie générale des probabilités

- ▶ Les espaces probabilisés discrets sont-ils suffisants pour nos modèles ?
- ▶ Que se passe-t-il si on conserve les mêmes définitions dans le cas où  $\Omega$  n'est pas dénombrable ?

## Différence fondamentale entre le cas discret et le cas général

Dans la théorie générale des probabilités, on n'essayera plus de mesurer (par la probabilité  $\mathbb{P}$ ) tous les sous-ensembles de  $\Omega$ , mais seulement ceux qui appartiennent à une certaine famille  $\mathcal{A}$ .

La formule  $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$  ne sera valable que dans le cas où  $A$  est dénombrable.

Si l'on veut pouvoir manipuler  $\mathbb{P}$  sur  $\mathcal{A}$  de manière fluide,  $\mathcal{A}$  devrait être stable par les opérations ensemblistes qui interviennent dans la définition et les propriétés d'une probabilité discrète

# Tribus

## Définition (Tribu)

Un sous-ensemble  $\mathcal{A}$  de  $\mathcal{P}(\Omega)$  est une **tribu** (ou  **$\sigma$ -algèbre**) sur  $\Omega$  si

- ▶  $\Omega \in \mathcal{A}$
- ▶  $A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$
- ▶ Si  $(A_i)_i$  est une suite d'éléments de  $\mathcal{A}$ , alors  $\cup_{i \geq 1} A_i \in \mathcal{A}$

Les éléments de  $\mathcal{A}$  sont appelés **ensembles mesurables** ou **événements**. Le couple  $(\Omega, \mathcal{A})$  est alors appelé **espace mesurable**, ou **espace probabilisable**.

## Exemples

- ▶  $\mathcal{P}(\Omega)$  est une tribu sur  $\Omega$ .
- ▶  $\{\emptyset, \Omega\}$  est une tribu sur  $\Omega$ .
- ▶ Si  $A \subset \Omega$ ,  $\{\emptyset, A, \bar{A}, \Omega\}$  est une tribu sur  $\Omega$ .



## Propriété

Une intersection d'un nombre quelconque de tribus sur  $\Omega$  est une tribu sur  $\Omega$ .

Remarque : il n'en est pas de même pour les unions !

## Définition

Soit  $\mathcal{E}$  un sous ensemble de  $\mathcal{P}(\Omega)$ .

- ▶ La **tribu engendrée** par  $\mathcal{E}$ , notée  $\sigma(\mathcal{E})$ , est l'intersection de toutes les tribus contenant  $\mathcal{E}$ .

## Propriété

Soit  $\mathcal{E}$  un sous ensemble de  $\mathcal{P}(\Omega)$  et  $\mathcal{A}$  une tribu sur  $\Omega$ . Alors

$$\mathcal{E} \subset \mathcal{A} \implies \sigma(\mathcal{E}) \subset \mathcal{A} .$$

# Tribus boréliennes

## Définition

Si  $\Omega$  est un espace topologique (en particulier  $\mathbb{R}$  ou  $\mathbb{R}^d$ ), on appelle **tribu borélienne** sur  $\Omega$ , que l'on note  $\mathcal{B}(\Omega)$ , la tribu engendrée par les ouverts de  $\Omega$ . On appelle **borélien** un sous-ensemble de  $\Omega$  appartenant à la tribu borélienne.

## Définition alternative

La tribu borélienne de  $\mathbb{R}^d$  est la tribu engendrée par les pavés de  $\mathbb{R}^d$  qui s'écrivent sous la forme

$$\prod_{i=1}^d ]-\infty, b_i].$$

## Remarque

La tribu borélienne de  $\mathbb{R}^d$ , même si elle contient la plupart des sous-ensembles de  $\mathbb{R}^d$  auxquels on peut penser naturellement, est différente de  $\mathcal{P}(\mathbb{R}^d)$ .

# Mesure positive, probabilité

## Définition

Soit  $\Omega$  un ensemble quelconque, et  $\mathcal{A}$  une tribu sur  $\Omega$ . On appelle **mesure positive** sur  $(\Omega, \mathcal{A})$  une application  $\mu$  de  $\mathcal{A}$  dans  $\bar{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{+\infty\}$  telle que :

- ▶  $\mu(\emptyset) = 0$ .
- ▶ pour toute suite  $(A_n)_{n \geq 1}$  d'éléments deux à deux disjoints de  $\mathcal{A}$ ,

$$\mu \left( \bigcup_{n \geq 1} A_n \right) = \sum_{n \geq 1} \mu(A_n).$$

Cette dernière propriété est appelée  $\sigma$ -**additivité** de  $\mu$ .

Le triplet  $(\Omega, \mathcal{A}, \mu)$  est appelé **espace mesuré**.

On appelle **probabilité** sur  $(\Omega, \mathcal{A})$  une mesure positive  $\mathbb{P}$  de masse totale 1 (c'est-à-dire telle que  $\mathbb{P}(\Omega) = 1$ ). On dit alors que  $(\Omega, \mathcal{A}, \mathbb{P})$  est un **espace probabilisé**.

## Exemples de mesures positives

- ▶ Si  $\Omega$  est fini ou dénombrable et si  $\mathbb{P}$  est une probabilité sur  $\Omega$  suivant la définition donnée au premier chapitre, alors  $\mathbb{P}$  est une probabilité sur l'espace probabilisable  $(\Omega, \mathcal{P}(\Omega))$ .
- ▶ Soit  $(\Omega, \mathcal{A})$  un espace probabilisable, et  $x$  un point de  $\Omega$ . On appelle **mesure de Dirac** (ou **masse de Dirac**) au point  $x$  la probabilité notée  $\delta_x$  sur  $(\Omega, \mathcal{A})$  définie par :

$$\forall A \in \mathcal{A}, \quad \delta_x(A) = \mathbf{1}_A(x) = \begin{cases} 1 & \text{si } x \in A, \\ 0 & \text{sinon.} \end{cases}$$

- ▶ La **mesure de comptage**  $\mu$  sur  $(\Omega, \mathcal{P}(\Omega))$  est définie par

$$\forall A \subset \Omega, \quad \mu(A) = \begin{cases} \text{Card}(A) & \text{si } A \text{ est fini,} \\ +\infty & \text{sinon.} \end{cases}$$

## Propriétés

Soit  $\mu$  une mesure sur  $(\Omega, \mathcal{A})$  et  $(A_i)_{i \geq 1}$  une suite d'ensembles mesurables. Alors :

- ▶ Si  $A_1 \subset A_2$ ,  $\mu(A_1) \leq \mu(A_2)$  (croissance).
- ▶  $\mu(\cup_i A_i) \leq \sum_i \mu(A_i)$  (sous-additivité).
- ▶ Si  $(A_i)_i$  est croissante,  $\mu(\cup_i A_i) = \lim_i \mu(A_i)$ .
- ▶ Si  $(A_i)_i$  est décroissante et  $\mu(A_{i_0}) < +\infty$  pour un certain  $i_0$ , alors  $\mu(\cap_i A_i) = \lim_i \mu(A_i)$ .

De même, toutes les propriétés des probabilités montrées dans le cas discret sont encore valables pour la définition générale d'une probabilité.

## Transformations de mesures

Soit  $(\Omega, \mathcal{A}, \mu)$  un espace mesuré.

- ▶ Si  $A \in \mathcal{A}$ , alors l'application  $\mu_A : \mathcal{A} \rightarrow \bar{\mathbb{R}}_+$  définie par  $\mu_A(B) = \mu(A \cap B)$  est une mesure, appelée **mesure trace** de  $\mu$  sur  $A$ . On peut bien entendu aussi remplacer  $\mathcal{A}$  par la **tribu trace** de  $\mathcal{A}$  sur  $A$ , composée des  $A \cap B$  où  $B \in \mathcal{A}$ .
- ▶ Si  $\lambda > 0$ ,  $\lambda\mu$  est une mesure.
- ▶ Si  $\mu'$  est aussi une mesure sur  $(\Omega, \mathcal{A})$ , alors  $\mu + \mu'$  est une mesure.

### Définition

Si  $(\Omega, \mathcal{A}, \mathbb{P})$  est un espace probabilisé et  $C$  un ensemble mesurable tel que  $\mathbb{P}(C) > 0$ , alors la **probabilité conditionnelle** sachant  $C$  est la probabilité  $\mathbb{P}(\cdot | C)$  définie par

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

# Prolongement de mesures — définitions

## Définition

Une mesure  $\mu$  sur  $(\Omega, \mathcal{A})$  est dite  **$\sigma$ -finie** si il existe un recouvrement dénombrable de  $\Omega$  par des sous-ensembles de  $\mu$ -mesure finie.

## Définition

Un ensemble  $\mathcal{C}$  de parties de  $\Omega$  est appelé **semi-anneau d'ensembles** si :

- ▶ L'ensemble vide appartient à  $\mathcal{C}$  ;
- ▶ Pour tous  $A, B \in \mathcal{C}$ ,  $A \setminus B$  est une réunion disjointe finie d'éléments de  $\mathcal{C}$  ;
- ▶ Pour tous  $A, B \in \mathcal{C}$ ,  $A \cap B$  appartient à  $\mathcal{C}$ .

# Prolongement de mesures

## Théorème de prolongement de Carathéodory

Soit  $\mu$  une mesure positive définie sur un semi-anneau  $\mathcal{C}$  de parties de  $\Omega$  :  $\mu(\emptyset) = 0$ , et, pour toute suite  $(A_n)_{n \geq 1}$  d'éléments deux à deux disjoints de  $\mathcal{C}$  telle que  $\cup_n A_n \in \mathcal{C}$ ,

$$\mu \left( \bigcup_{n \geq 1} A_n \right) = \sum_{n \geq 1} \mu(A_n).$$

Alors la fonction  $\bar{\mu}$  définie sur  $\sigma(\mathcal{C})$  par

$$\bar{\mu}(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu(A_n); A_n \in \mathcal{C} \forall n \text{ et } A \subset \bigcup_{n=1}^{\infty} A_n \right\}$$

est une mesure qui prolonge  $\mu$ . Ce prolongement est unique si  $\mu$  est une mesure  $\sigma$ -finie.



## Mesures de Lebesgue

Sur  $\mathbb{R}$ , l'ensemble  $\mathcal{C}$  des intervalles du type  $]a, b]$  (avec  $a, b \in \bar{\mathbb{R}}$ ) forme un semi-anneau qui engendre la tribu borélienne  $\mathcal{B}(\mathbb{R})$ . Définissons  $\mu(]a, b]) = b - a$ . D'après le théorème de prolongement de Carathéodory,  $\mu$  se prolonge en une unique mesure  $\lambda$  sur  $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ . Cette mesure est appelée **mesure de Lebesgue** sur  $\mathbb{R}$ .

On définit de même la mesure de Lebesgue  $\lambda_d$  sur  $\mathbb{R}^d$  par ses valeurs sur les pavés :

$$\lambda_d \left( \prod_{i=1}^n ]a_i, b_i] \right) = \prod_{i=1}^n (b_i - a_i) \quad \forall a_i < b_i \in \mathbb{R}.$$

**Remarque :** Si  $A \subset \mathbb{R}^d$  est dénombrable, alors  $\lambda_d(A) = 0$ .

### Définition

Soit  $A$  un borélien de  $\mathbb{R}^d$  de mesure de Lebesgue finie et non-nulle. La **probabilité uniforme** sur  $A$  est la mesure  $\mathbb{P}$  définie par

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

# Indépendance

La probabilité conditionnelle, ainsi que l'indépendance et l'indépendance mutuelle d'un nombre fini d'événements ou de familles d'événements, sont définis de la même façon que dans le cas discret...

## Définition (Indépendance mutuelle)

On dit que les événements  $A_i$  d'une famille  $(A_i)_{i \in I} \subset \mathcal{A}$  sont **mutuellement indépendants** si, pour tout sous-ensemble fini  $J$  de  $I$ , les événements  $A_i$  pour  $i \in J$  sont mutuellement indépendants.

On dit que les familles  $\mathcal{F}_i \subset \mathcal{A}$  pour  $i \in I$  sont mutuellement indépendantes si, pour tout sous-ensemble fini  $J$  de  $I$ , les familles  $\mathcal{F}_i$  pour  $i \in J$  sont mutuellement indépendantes.

## Espace probabilisé produit de deux espaces

On considère deux expériences modélisées séparément par deux espaces probabilisés  $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$  et  $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ . Comment regrouper ces deux modélisations en une seule si les expériences sont indépendantes ?

De manière naturelle, on choisit  $\Omega = \Omega_1 \times \Omega_2$ .

### Définition (Tribu produit)

La **tribu produit**  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$  sur  $\Omega$  est la tribu engendrée par les pavés

$$A_1 \times A_2, \quad A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

### Définition (Probabilité produit)

La **probabilité produit** est l'unique probabilité  $\mathbb{P}$  qui vérifie

$$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1) \times \mathbb{P}_2(A_2) \quad \forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

# Espace probabilisé produit

## Définition

Soit  $I$  un ensemble quelconque, et, pour tout  $i \in I$ ,  $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$  un espace probabilisé. On appelle **espace probabilisé produit** des  $\Omega_i$  l'espace probabilisé constitué de l'univers  $\Omega = \prod_{i \in I} \Omega_i$ , de la tribu  $\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i$  et de la probabilité  $\mathbb{Q} = \bigotimes_{i \in I} \mathbb{P}_i$ , définis de la manière suivante :

- ▶ un **pavé** de  $\Omega = \prod_{i \in I} \Omega_i$  est un sous-ensemble de  $\Omega$  de la forme  $\prod_{i \in I} A_i$ , où, pour tout  $i \in I$ ,  $A_i \in \mathcal{A}_i$ , et où les  $A_i$  différents de  $\Omega_i$  sont en nombre fini.
- ▶  $\mathcal{A}$  est la tribu engendrée par les pavés de  $\Omega$ .
- ▶  $\mathbb{Q}$  est la probabilité sur  $(\Omega, \mathcal{A})$  définie de manière unique par sa valeur sur les pavés : si  $A = \prod_{i \in I} A_i$  est un pavé de  $\Omega$ , alors (dans le produit ci-dessous, il n'y a qu'un nombre fini de termes différents de 1, de sorte que ce produit peut être considéré comme fini) :

$$\mathbb{Q}(A) = \prod_{i \in I} \mathbb{P}_i(A_i).$$

**Remarque :** On peut définir de la même façon la *mesure* produit.

**Exemples :**

- ▶ La tribu de Borel sur  $\mathbb{R}^n$  est la tribu produit de tribus de Borel de  $\mathbb{R}$ , et la mesure de Lebesgue en dimension  $n$  est la mesure produit de  $n$  mesures de Lebesgue unidimensionnelles :

$$\mathcal{B}(\mathbb{R}^n) = \bigotimes_{i=1}^n \mathcal{B}(\mathbb{R}), \quad \lambda_n = \bigotimes_{i=1}^n \lambda.$$

- ▶ Modélisation d'une infinité de lancers d'un dé à 6 faces par  $(\Omega, \mathcal{A}, \mathbb{P})$ , avec

$$\Omega = \{1, \dots, 6\}^{\mathbb{N}^*}, \quad \mathcal{A} = \bigotimes_{i \in \mathbb{N}^*} \mathcal{P}(\{1, \dots, 6\}), \quad \mathbb{P} = \bigotimes_{i \in \mathbb{N}^*} \mathbb{P}_6,$$

où  $\mathbb{P}_6$  est la probabilité uniforme sur  $\{1, \dots, 6\}$ .

# Mesurabilité des fonctions, éléments aléatoires

Pour une variable aléatoire  $X$ , on souhaite pouvoir mesurer  $\{X \in B\} = X^{-1}(B)$ ...

# Mesurabilité des fonctions, éléments aléatoires

## Définition

Soient  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{A}')$  deux espaces probabilisables, et  $X$  une application de  $\Omega$  dans  $\Omega'$ . On dit que  $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$  est **mesurable** si

$$\forall B \in \mathcal{A}', \quad X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{A},$$

c'est-à-dire si

$$X^{-1}(\mathcal{A}') \subset \mathcal{A}.$$

Si  $(\Omega, \mathcal{A}, \mathbb{P})$  est un espace probabilisé et  $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$  est mesurable, alors on dit que  $X$  est un **élément aléatoire**.

**Remarque :** Dans le cas où  $\mathcal{A} = \mathcal{P}(\Omega)$ , toute application  $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$  est bien sûr mesurable, et ce quels que soient  $\Omega'$  et  $\mathcal{A}'$ .

# Mesurabilité des fonctions, éléments aléatoires

## Définition

Soient  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{A}')$  deux espaces probabilisables, et  $X$  une application de  $\Omega$  dans  $\Omega'$ . On dit que  $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$  est **mesurable** si

$$\forall B \in \mathcal{A}', \quad X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{A},$$

c'est-à-dire si

$$X^{-1}(\mathcal{A}') \subset \mathcal{A}.$$

Si  $(\Omega, \mathcal{A}, \mathbb{P})$  est un espace probabilisé et  $X : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$  est mesurable, alors on dit que  $X$  est un **élément aléatoire**.

## Exemples :

- ▶ Toute application constante est mesurable.
- ▶ Si  $A \subset \Omega$ ,  $\mathbf{1}_A$  est mesurable si et seulement si  $A \in \mathcal{A}$  (sauf cas inintéressants).



# Transport de tribus

## Lemme

Soient  $\Omega$  et  $\Omega'$  deux ensembles quelconques, et  $X$  une application de  $\Omega$  dans  $\Omega'$ .  
Alors :

1. si  $\mathcal{A}'$  est une tribu sur  $\Omega'$ , alors  $\mathcal{A} = X^{-1}(\mathcal{A}')$  est une tribu sur  $\Omega$ ;
2. si  $\mathcal{A}$  est une tribu sur  $\Omega$ , alors

$$\mathcal{A}' = \{A' \subset \Omega', X^{-1}(A') \in \mathcal{A}\}$$

est une tribu de  $\Omega'$ , appelée **tribu induite** de  $\mathcal{A}$  par  $X$ .

3. si  $\mathcal{C}' \subset \mathcal{P}(\Omega')$ , alors

$$X^{-1}(\sigma(\mathcal{C}')) = \sigma(X^{-1}(\mathcal{C}')).$$

## Critère de mesurabilité

Soient  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{A}')$  deux espaces probabilisables, et  $X$  une application de  $\Omega$  dans  $\Omega'$ . Soit  $\mathcal{C}'$  un sous-ensemble de  $\mathcal{P}(\Omega')$  tel que  $\sigma(\mathcal{C}') = \mathcal{A}'$ . Si

$$X^{-1}(\mathcal{C}') \subset \mathcal{A},$$

alors  $X$  est mesurable.

### Exemples :

- ▶ Si  $(\Omega, \mathcal{B}(\Omega))$  et  $(\Omega', \mathcal{B}(\Omega'))$  sont deux espaces topologiques munis de leurs tribus boréliennes, alors toute application continue  $X$  de  $\Omega$  dans  $\Omega'$  est mesurable.
- ▶ Toute fonction croissante de  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  est mesurable.

# Manipulations de fonctions mesurables

## Propriétés

- ▶ Soient  $f$  et  $g$  deux applications mesurables :

$$(\Omega, \mathcal{A}) \xrightarrow{f} (\Omega', \mathcal{A}') \xrightarrow{g} (\Omega'', \mathcal{A}'').$$

Alors  $g \circ f$  est mesurable.

- ▶ Soient  $f : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{B})$  et  $g : (\Omega, \mathcal{A}) \rightarrow (E', \mathcal{B}')$  deux applications mesurables. Alors l'application

$$v = (f, g) : (\Omega, \mathcal{A}) \longrightarrow (E \times E', \mathcal{B} \otimes \mathcal{B}')$$

est mesurable.

- ▶ Soient  $f$  et  $g$  deux applications mesurables à valeurs réelles, et  $\alpha$  un réel. Alors les applications  $\alpha f$ ,  $f + g$ ,  $\sup(f, g)$ ,  $\inf(f, g)$  et  $f \times g$  sont mesurables. Si  $(f_n)_{n \geq 1}$  est une suite de fonctions mesurables, alors les fonctions  $\lim_n f_n$  (quand elle existe),  $\sup_n f_n$ ,  $\inf_n f_n$ ,  $\limsup_n f_n$  et  $\liminf_n f_n$  sont mesurables.

## Propriétés

- ▶ Soit  $(f_n)_n$  une suite de fonctions mesurables de  $(\Omega, \mathcal{A})$  dans un espace métrique  $(E, d)$  muni de sa tribu borélienne. Si  $f_n$  converge simplement vers  $f$ , alors  $f$  est mesurable.

**Preuve :** Soit  $U$  un ouvert de  $E$ , et

$$U_r = \{x \in U, d(x, E \setminus U) > 1/r\}.$$

Alors

$$f^{-1}(U) = \bigcup_r \liminf_n f_n^{-1}(U_r).$$

# Tribu engendrée par une application

## Définition

Soit  $\Omega$  un ensemble quelconque,  $(\Omega', \mathcal{A}')$  un espace probabilisable, et  $X$  une application de  $\Omega$  dans  $\Omega'$ . On appelle **tribu engendrée par  $X$** , que l'on note  $\sigma(X)$ , la plus petite tribu sur  $\Omega$  qui rend  $X$  mesurable. Elle s'écrit  $\sigma(X) = X^{-1}(\mathcal{A}')$ .

**Remarque :** une application  $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  est mesurable si et seulement si  $\sigma(X) \subset \mathcal{A}$ .

### Exemples :

- ▶ Si  $X$  est constante,  $\sigma(X) = \{\emptyset, \Omega\}$ .
- ▶ Si  $A$  est un événement et  $X = \mathbf{1}_A$ , alors (en prenant une tribu raisonnable sur l'espace d'arrivée)  $\sigma(X) = \{\emptyset, A, \bar{A}, \Omega\}$ .

# Mesure image

## Définition

Soit  $X : (\Omega, \mathcal{A}, \mu) \rightarrow (E, \mathcal{B})$  une application mesurable. On appelle **mesure image de  $\mu$  par  $X$**  la mesure notée  $\mu_X$  définie par

$$\forall B \in \mathcal{B}, \quad \mu_X(B) = \mu(X^{-1}(B)).$$

Dans le cas où  $\mu$  est une probabilité,  $\mu_X$  est une probabilité appelée **loi (de probabilité) de  $X$** , ou **distribution de  $X$** .

# Indépendance d'éléments aléatoires

## Définition

Soit  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace probabilisé, et  $X : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E, \mathcal{B})$  et  $Y : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (E', \mathcal{B}')$  deux éléments aléatoires. On dit que  $X$  et  $Y$  sont **indépendants** si  $\sigma(X)$  et  $\sigma(Y)$  sont deux familles indépendantes d'événements, c'est-à-dire si, pour tous  $A \in \mathcal{B}$ ,  $B \in \mathcal{B}'$ ,  $\{X \in A\}$  et  $\{Y \in B\}$  sont indépendants.

## Définition

Soient  $X_1, \dots, X_n$   $n$  éléments aléatoires définis sur un même espace probabilisé, à valeurs dans des espaces probabilisables  $(E_i, \mathcal{B}_i)$ . On dit que les  $X_i$  sont **mutuellement indépendants** si les familles  $\sigma(X_i)$  sont mutuellement indépendantes, c'est-à-dire si

$$\mathbb{P} \left( \bigcap_{i=1}^n \{X_i \in A_i\} \right) = \prod_{i=1}^n \mathbb{P} [X_i \in A_i] \quad \forall A_1 \in \mathcal{B}_1, \dots, A_n \in \mathcal{B}_n.$$

## Définition

Soit  $I$  un ensemble quelconque et  $(X_i)_{i \in I}$  une famille d'éléments aléatoires définis sur un même espace probabilisé. On dit que les  $X_i$  sont **mutuellement indépendants** si, pour tout sous-ensemble fini  $J$  de  $I$ , les éléments aléatoires  $(X_i)_{i \in J}$  sont mutuellement indépendants.

**Remarque :** Un élément aléatoire constant est indépendant de tout autre élément aléatoire.

**Propriété :** Par définition, deux éléments aléatoires  $X$  et  $Y$  sont indépendants si et seulement si

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y .$$

De même, les éléments aléatoires d'une famille  $X = (X_i)_{i \in I}$  sont mutuellement indépendants si et seulement si

$$\mathbb{P}_X = \bigotimes_{i \in I} \mathbb{P}_{X_i} .$$



# Critère d'indépendance

## Définition

Un ensemble  $\mathcal{C}$  de sous-ensembles de  $\Omega$  est appelé  $\pi$ -**système** si il est stable par intersection :

$$\forall A, B \in \mathcal{C}, A \cap B \in \mathcal{C}.$$

## Théorème

Si  $\mathcal{C}_1$  et  $\mathcal{C}_2$  sont deux  $\pi$ -systèmes indépendants dans l'espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ , alors les tribus  $\sigma(\mathcal{C}_1)$  et  $\sigma(\mathcal{C}_2)$  sont indépendantes.

## Rassemblement d'éléments aléatoires indépendants

Soit  $(X_i)_{i \in I}$  une famille quelconque d'éléments mutuellement indépendants, à valeurs dans des espaces probabilisables  $(E_i, \mathcal{B}_i)$ , et  $(J_k)_{k \in K}$  une famille de sous-ensembles de  $I$  deux à deux disjoints (par exemple une partition de  $I$ ). Alors les éléments aléatoires

$$(Y_k)_{k \in K} = ((X_i)_{i \in J_k})_{k \in K}$$

à valeurs dans les espaces probabilisables produits

$$\left( \Omega_k = \prod_{i \in J_k} E_i, \mathcal{C}_k = \bigotimes_{i \in J_k} \mathcal{B}_i \right), \quad k \in K,$$

sont mutuellement indépendants.

## Transformations d'éléments aléatoires indépendants

Soit  $(X_i)_{i \in I}$  une famille quelconque d'éléments mutuellement indépendants, et  $(f_i)_{i \in I}$  une famille d'applications mesurables. Alors les  $(f_i(X_i))_{i \in I}$  sont mutuellement indépendants.

# Intégration

**Objectif** : définir dans le cadre général une espérance pour les fonctions mesurables à valeurs dans  $\mathbb{R}^d$ , sous la forme

$$\mathbb{E} X = \int_{\Omega} X d\mathbb{P} .$$

Ingrédients :

- ▶ prolongement des définitions du cas discret ;
- ▶ linéarité.

On considère un espace mesuré  $(\Omega, \mathcal{A}, \mu)$ . On placera toujours la tribu borélienne sur  $\mathbb{R}^d$ .

## Définition (Intégrale de fonctions indicatrices)

Si  $A \in \mathcal{A}$ , la fonction  $\mathbf{1}_A$  est mesurable, et on définit son intégrale par rapport à la mesure  $\mu$  par

$$\int \mathbf{1}_A d\mu = \int_{\Omega} \mathbf{1}_A(\omega) d\mu(\omega) = \mu(A).$$

## Définition (Intégrale de fonctions indicatrices)

Si  $A \in \mathcal{A}$ , la fonction  $\mathbf{1}_A$  est mesurable, et on définit son intégrale par rapport à la mesure  $\mu$  par

$$\int \mathbf{1}_A d\mu = \int_{\Omega} \mathbf{1}_A(\omega) d\mu(\omega) = \mu(A).$$

## Définition (Intégrale de fonctions étagées)

On appelle **fonction étagée positive** une fonction  $f : \Omega \rightarrow \mathbb{R}$  qui peut s'écrire sous la forme

$$f = \sum_{i=1}^n a_i \mathbf{1}_{A_i},$$

où les  $a_i$  sont des réels positifs et les  $A_i \in \mathcal{A}$  sont deux à deux disjoints.

On définit l'intégrale de ces fonctions par rapport à la mesure  $\mu$  par

$$\int f d\mu = \sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n a_i \int \mathbf{1}_{A_i} d\mu.$$

**Remarque :** Dans le cas d'un espace de départ *probabilisé*, les fonctions étagées sont les variables aléatoires qui peuvent prendre un nombre fini de valeurs : dans ce cas,

$$X = \sum_{x \in X(\Omega)} x \mathbf{1}_{\{X=x\}}.$$

$$\text{et } \mathbb{E} X = \int X d\mathbb{P} = \sum_{x \in X(\Omega)} x \mathbb{P}[X = x].$$

**Propriétés :** (pour les fonction étagées positives)

- ▶ croissance,
- ▶ linéarité.

## Définition (Intégrale de fonctions mesurables positives)

Soit  $f : \Omega \rightarrow \mathbb{R}$  une fonction mesurable positive. On définit son intégrale par rapport à  $\mu$  par

$$\int f d\mu = \sup \left\{ \int g d\mu, g \text{ étagée positive telle que } g \leq f \right\}.$$

**Propriétés :** (pour les fonction mesurables positives)

1. Si  $0 \leq f \leq g$ , alors  $0 \leq \int f d\mu \leq \int g d\mu$ .
2. Si  $c \geq 0$  est une constante,  $\int cf d\mu = c \int f d\mu$ .
3.  $\int (f + g) d\mu = \int f d\mu + \int g d\mu$ .
4.  $\int f d\mu = 0$  est équivalent à  $f = 0$   $\mu$ -presque-partout.

## Théorème de convergence monotone

Soit  $(f_n)_n$  une suite croissante de fonctions mesurables positives convergeant simplement vers  $f$ . Alors  $f$  est mesurable et

$$\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu.$$

**Corollaire :** Si  $(f_n)_n$  est une suite de fonctions mesurables positives et  $f = \sum_{n \geq 1} f_n$ . Alors

$$\int f d\mu = \sum_{n \geq 1} \int f_n d\mu.$$

**Corollaire (Lemme de Fatou) :** Si  $(f_n)_n$  est une suite de fonctions mesurables positives,

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$



## Approximation par des fonctions étagées

Soit  $f$  une fonction mesurable positive. Il existe une suite croissante  $(f_n)_n$  de fonctions étagées positives qui converge simplement vers  $f$ .

**preuve :** Choisir par exemple

$$f_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq f < \frac{k+1}{2^n}\}} + n \mathbf{1}_{\{f \geq n\}}.$$

# Intégrale de fonctions quelconques

## Définition

Soit  $f$  une fonction mesurable,  $f^+ = f \vee 0$  et  $f^- = -(f \wedge 0)$ . Si  $\int |f| d\mu < +\infty$ , on dit que  $f$  est  $\mu$ -**intégrable**, et on définit son intégrale par rapport à  $\mu$  par

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu.$$

**Remarque** : l'intégrale d'une fonction positive est toujours définie dans  $\bar{\mathbb{R}}_+$  (éventuellement infinie), mais l'intégrale d'une fonction quelconque n'est pas définie si sa valeur absolue n'est pas intégrable.

## Définition

Si  $A \in \mathcal{A}$  et  $f$  est une fonction mesurable, on appelle intégrale de  $f$  sur  $A$  l'intégrale de  $\mathbf{1}_A f$ , et on note :

$$\int_A f d\mu = \int \mathbf{1}_A f d\mu.$$

## Propriétés :

- ▶ Si  $f$  et  $g$  sont intégrables et si  $\alpha, \beta \in \mathbb{R}$ , alors

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

- ▶ Si  $f \leq g$ , alors  $\int f d\mu \leq \int g d\mu$ .

## Schéma de la preuve :

- ▶ Cas des fonctions étagées.
- ▶ Cas des fonctions positives, par convergence monotone.
- ▶ Cas général, par décomposition en parties positive et négative.

## Théorème de convergence dominée de Lebesgue

Soit  $f_n$  une suite de fonctions mesurables convergeant simplement vers  $f$ , et  $g$  une fonction intégrable, telles que  $|f_n| \leq g$  pour tout  $n$ . Alors  $f$  est intégrable, et

$$\lim_{n \rightarrow +\infty} \int f_n d\mu = \int f d\mu.$$

## Inégalité de Jensen

Soit  $\phi$  une fonction convexe de  $\mathbb{R}$  dans  $\mathbb{R}$ , et  $X$  une fonction mesurable telle que  $X$  et  $\phi(X)$  sont intégrables par rapport à une mesure de probabilité  $\mathbb{P}$ . Alors

$$\phi(\mathbb{E} X) \leq \mathbb{E}(\phi(X)).$$

## Théorème de Fubini

Soit  $f$  une fonction mesurable à valeurs réelles définie sur un espace probabilisé produit  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$ . Les trois propriétés suivantes sont équivalentes :

- ▶  $g_1(\omega_1) = \int_{\Omega_2} |f(\omega_1, \omega_2)| d\mu_2(\omega_2)$  est  $\mu_1$ -intégrable ;
- ▶  $g_2(\omega_2) = \int_{\Omega_1} |f(\omega_1, \omega_2)| d\mu_1(\omega_1)$  est  $\mu_2$ -intégrable ;
- ▶  $f$  est  $\mu_1 \otimes \mu_2$ -intégrable.

De plus, si ces conditions sont vérifiées (ou si  $f$  est à valeurs positives), alors

$$\int_{\Omega_1 \times \Omega_2} f d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left( \int_{\Omega_2} f d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left( \int_{\Omega_1} f d\mu_1 \right) d\mu_2.$$

## Densité

Dans ce paragraphe, toutes les mesures sont supposées  $\sigma$ -finies, c'est-à-dire que pour toute mesure  $\mu$  sur  $(\Omega, \mathcal{A})$  que l'on considèrera, il existe une partition dénombrable  $(A_i)_i$  de  $\Omega$  telle que  $\mu(A_i) < +\infty$  pour tout  $i$ .

### Définition

Soit  $f$  une fonction sur  $(\Omega, \mathcal{A})$  à valeurs dans  $\mathbb{R}$ , mesurable, positive et intégrable. La fonction  $\nu : \mathcal{A} \rightarrow \bar{\mathbb{R}}_+$  définie par

$$\nu(A) = \int_A f d\mu$$

est une mesure appelée **mesure de densité  $f$  par rapport à  $\mu$** .

# Densité

## Définition

Soit  $f$  une fonction sur  $(\Omega, \mathcal{A})$  à valeurs dans  $\mathbb{R}$ , mesurable, positive et intégrable. La fonction  $\nu : \mathcal{A} \rightarrow \overline{\mathbb{R}}_+$  définie par

$$\nu(A) = \int_A f d\mu$$

est une mesure appelée **mesure de densité  $f$  par rapport à  $\mu$** .

## Propriétés

Si  $\nu$  est la mesure de densité  $f$  par rapport à  $\mu$ , alors :

- ▶ si  $\mu(A) = 0$ , alors  $\nu(A) = 0$  ;
- ▶ si  $g$  est positive mesurable,  $\int g d\nu = \int gf d\mu$  ;
- ▶ une fonction  $g$  mesurable est  $\nu$ -intégrable si et seulement si  $gf$  est  $\mu$ -intégrable, et, dans ce cas,  $\int g d\nu = \int gf d\mu$ .

## Exemples :

- ▶ La mesure discrète

$$\nu = \sum_{i=1}^{\infty} \alpha_i \delta_{a_i} .$$

- ▶ Fabrication de mesures de probabilité non-discrètes et non-uniformes sur  $\mathbb{R}^d$ .  
On dira que le **vecteur aléatoire**  $X$  est **continu, de densité**  $f$ , si sa loi  $\mathbb{P}_X$  est absolument continue par rapport à la mesure de Lebesgue et a pour densité  $f$ , c'est-à-dire si

$$\mathbb{P}[X \in A] = \int_A f(x) d\lambda_d(x) .$$

**Remarque :** Une fonction  $f$  Riemann-intégrable sur un pavé  $A \subset \mathbb{R}^d$  est Lebesgue-intégrable, et son intégrale de Lebesgue est égale à son intégrale de Riemann.



## Définition

Soient  $\mu$  et  $\nu$  deux mesures définies sur le même espace mesurable  $(\Omega, \mathcal{A})$ .

- ▶ Si, pour tout  $A \in \mathcal{A}$ ,  $\mu(A) = 0 \Rightarrow \nu(A) = 0$ , on dit que  $\nu$  est **absolument continue par rapport à  $\mu$** , et on note  $\nu \ll \mu$ .
- ▶ Si  $\nu \ll \mu$  et  $\mu \ll \nu$ , on dit que  $\mu$  et  $\nu$  sont **équivalentes**.
- ▶ Si il existe un ensemble mesurable  $A$  tel que  $\mu(A) = 0$  et  $\nu(\bar{A}) = 0$ , alors on dit que  $\mu$  et  $\nu$  sont **étrangères**.

## Exemples :

- ▶ La mesure  $\nu$  de densité  $f$  par rapport à  $\mu$  est absolument continue par rapport à  $\mu$ .
- ▶ Toute mesure est absolument continue par rapport à la mesure de comptage.
- ▶ Une mesure de probabilité discrète est étrangère à la mesure de Lebesgue.

## Théorème

Soient  $\mu$  et  $\nu$  deux mesures  $\sigma$ -finies sur  $(\Omega, \mathcal{A})$ . Il existe une unique décomposition  $\nu = \nu_c + \nu_\perp$ , où  $\nu_c$  est absolument continue par rapport à  $\mu$  et  $\nu_\perp$  est étrangère à  $\mu$ .

## Théorème de Radon-Nikodym

Si  $\mu$  et  $\nu$  sont deux mesures  $\sigma$ -finies sur  $(\Omega, \mathcal{A})$  telles que  $\nu \ll \mu$ , alors il existe une fonction mesurable positive  $f$  telle que  $\nu$  soit la mesure de densité  $f$  par rapport à  $\mu$ . On dit que  $f$  est une densité de  $\nu$  par rapport à  $\mu$ , et on note

$$f = \frac{d\nu}{d\mu}.$$

## Théorème de transport

Soit  $f$  une application mesurable de  $(\Omega, \mathcal{A}, \mu)$  dans  $(E, \mathcal{E})$  et  $\phi$  une application mesurable de  $(E, \mathcal{E})$  dans  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ .

Alors  $\phi$  est  $\mu_f$ -intégrable si et seulement si  $\phi \circ f$  est  $\mu$ -intégrable, et, dans ce cas (ou si  $\phi$  est à valeurs positives),

$$\int_E \phi d\mu_f = \int_{\Omega} \phi \circ f d\mu.$$

### Application aux mesures de probabilité à densité sur $\mathbb{R}^n$ :

Dans la situation suivante :

$$(\Omega, \mathcal{A}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \xrightarrow{\phi} (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

où  $\mathbb{P}_X$  est la loi de densité  $f_X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}^n$ , le théorème de transfert implique

$$\mathbb{E} \phi(X) = \int_{\Omega} \phi(X(\omega)) d\mathbb{P}(\omega) = \int_{\mathbb{R}^n} \phi(x) d\mathbb{P}_X(x) = \int_{\mathbb{R}^n} \phi(x) f_X(x) d\lambda_n(x).$$

## Définition

Un vecteur aléatoire  $X : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  est dit **continu** de densité  $f : \mathbb{R}^n \mapsto \mathbb{R}_+$  si la loi de  $X$  est une probabilité absolument continue par rapport à la mesure de Lebesgue  $\lambda_d$ , de densité  $f$  :

$$\mathbb{P}[X \in A] = \mathbb{P}_X(A) = \int_A f(x) d\lambda_d(x) \quad \forall A \in \mathcal{B}(\mathbb{R}^n).$$

# Espérance et indépendance

## Propriété

Soient  $X$  et  $Y$  deux éléments aléatoires *indépendants*, définis sur le même espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$  et à valeurs dans  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$ . Soient  $\phi : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et  $\psi : (F, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  deux applications mesurables, telles que  $E\phi(X)$  et  $E\psi(Y)$  soient bien définis.

Alors

$$E(\phi(X)\psi(Y)) = E\phi(X) \cdot E\psi(Y).$$

## Cas particulier :

Si  $X$  et  $Y$  sont deux variables aléatoires intégrables *indépendantes*, alors

$$E(XY) = EX \cdot EY.$$

## Exemple des changements de probabilité

Si  $Y$  est une variable aléatoire positive telle que  $E Y = 1$ , on peut considérer la probabilité  $\mathbb{P}'$  de densité  $Y$  par rapport à  $\mathbb{P}$  :

$$\mathbb{P}'(A) = \int_A Y d\mathbb{P} .$$

### Exemples :

- ▶ Si  $X$  et  $Y$  sont indépendantes,  $Y$  positive et d'espérance 1, et  $\mathbb{P}'$  est la probabilité de densité  $Y$  par rapport à  $\mathbb{P}$ , la loi de  $X$  sous  $\mathbb{P}$  est la même que celle de  $X$  sous  $\mathbb{P}'$ .
- ▶ Si  $X$  est continue de densité  $f$ , peut-on trouver  $Y$  telle que la loi de  $X$  sous  $\mathbb{P}'$  soit de densité  $g$  ? On essaiera de prendre  $Y$  sous la forme  $\psi(X)$ .

# Plan

Espaces probabilisés discrets

Théorie générale des probabilités

**Manipulation des vecteurs aléatoires**

- Densité

- Fonction de répartition, fonction de survie

- Moments

- Convolution

- Lois de probabilité usuelles

- Caractérisations fonctionnelles des lois de probabilité

Statistiques descriptives

# Densité

## Remarque :

On peut déterminer la loi d'une des composantes  $X_i$  d'un vecteur aléatoire  $X = (X_1, \dots, X_d)$  à partir de celle de  $X$  :

$$\begin{aligned}\mathbb{P}_{X_i}(A) &= \mathbb{P}[X_i \in A] = \mathbb{P}[X \in \mathbb{R} \times \dots \times \mathbb{R} \times A \times \mathbb{R} \times \dots \times \mathbb{R}] \\ &= \mathbb{P}_X(\mathbb{R}^{i-1} \times A \times \mathbb{R}^{d-i}).\end{aligned}$$

De la même façon, on peut déterminer la loi d'un sous-vecteur  $(X_i)_{i \in I}$  à partir de celle de  $X$ .

Par contre, la connaissance de toutes les lois *marginales* (les lois des  $X_i$ ) ne permet pas de déterminer la loi de  $X$  !



**Mise en œuvre dans le cas continu :**

Si  $X$  est un vecteur aléatoire continu de densité  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , alors  $X_i$  est une variable aléatoire continue de densité  $f_i : \mathbb{R} \rightarrow \mathbb{R}$  et

$$f_i(x) = \int_{\mathbb{R}^{d-1}} f(t_1, \dots, t_{i-1}, x, t_{i+1}, \dots, t_d) d\lambda_{d-1}(t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_d).$$

**Exemples :**

- ▶ Si  $(X, Y)$  est un vecteur aléatoire de densité

$$f(x, y) = 2\mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,x]}(y),$$

quelle est la loi de  $X$  ?

- ▶ Si  $X$  est une variable aléatoire de loi uniforme sur  $[0, 1]$ , le vecteur aléatoire  $(X, X)$  est-il continu ?

**Exemple 1**

$$f(x, y) = 2\mathbf{1}_{[0,1]}(x)\mathbf{1}_{[0,x]}(y).$$

Pour  $x \notin [0, 1]$ ,

$$f_X(x) = \int f(x, y) dy = 0.$$

Pour  $x \in [0, 1]$ ,

$$f_X(x) = \int f(x, y) dy = \int 2\mathbf{1}_{[0,x]}(y) dy = 2 \int_0^x 1 dy = 2x.$$

Donc finalement,

$$f_X(x) = 2x\mathbf{1}_{[0,1]}(x).$$

## Exemple 2

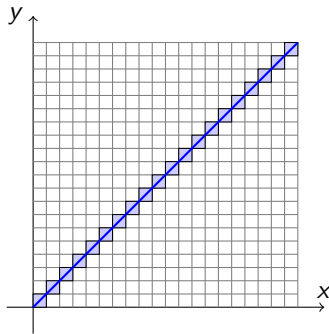
$$X \sim \mathcal{U}([0, 1])$$

En notant

$$\Delta = \{(x, y) \in \mathbb{R}^2, 0 \leq x = y \leq 1\},$$

$$\mathbb{P}_{(X, X)}(\Delta) = \mathbb{P}[(X, X) \in \Delta] = 1,$$

mais  $\lambda_2(\Delta) = 0$ , de sorte que  $\mathbb{P}_{(X, X)}$  n'est **pas** absolument continue par rapport à  $\lambda_2$ , c'est-à-dire que  $(X, X)$  n'est **pas** un vecteur aléatoire continu.



$$\lambda_2(\Delta) \leq n \frac{1}{n^2} = \frac{1}{n}$$

## Théorème de changement de variables

Soit  $X$  un vecteur aléatoire continu à valeurs dans  $\mathbb{R}^d$ , de densité  $f_X$ ,  $D = \{x \in \mathbb{R}^d, f_X(x) > 0\}$ , et  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ . On suppose que

- ▶  $\phi$  est bijective de  $D$  dans  $\phi(D)$ ;
- ▶  $\phi^{-1}$  est continument différentiable sur  $\phi(D)$ .

Alors le vecteur aléatoire  $Y = \phi(X)$  est continu, de densité  $f_Y$  définie par

$$f_Y(y) = \mathbf{1}_{\phi(D)}(y) f_X(\phi^{-1}(y)) |\det \text{Jac } \phi^{-1}(y)|.$$

### Exemples :

- ▶ Soit  $U$  une variable aléatoire de loi  $\mathcal{U}([0, 1])$ . Quelle est la loi de  $2U + 1$  ?
- ▶ Soit  $U$  une variable aléatoire de loi  $\mathcal{U}([0, 1])$ . Quelle est la loi de  $U^2$  ?  
Combien vaut  $E(U^2)$  ?
- ▶ Soit  $(U, V)$  une variable aléatoire de loi uniforme sur le disque unité. On note  $(R, \Theta)$  les coordonnées polaires correspondantes à  $(U, V)$ . Quelle est leur loi ?

### Exemple 1

$$U \sim \mathcal{U}([0, 1]), \quad Y = 2U + 1.$$

#### À la main

Soit  $I = [a, b]$ .

$$\begin{aligned} \mathbb{P}[Y \in I] &= \mathbb{P}[2U + 1 \in I] = \mathbb{P}[U \in (I - 1)/2] \\ &= \lambda((I - 1)/2 \cap [0, 1]) = \lambda\left(\left[\frac{a-1}{2}, \frac{b-1}{2}\right] \cap [0, 1]\right) \\ &= \frac{1}{2}\lambda([a-1, b-1] \cap [0, 2]) = \frac{1}{2}\lambda([a, b] \cap [1, 3]) \end{aligned}$$

Donc  $Y \sim \mathcal{U}([1, 3])$ .

### Exemple 1

$$U \sim \mathcal{U}([0, 1]), \quad Y = 2U + 1.$$

#### Par changement de variable

Ici,  $f_U(x) = \mathbf{1}_{[0,1]}(x)$ .

$Y = \phi(U)$ , avec  $\phi(u) = 2u + 1$  et donc  $\phi^{-1}(y) = (y - 1)/2$ .

En dimension 1,  $\text{Jac } \phi^{-1} = \phi^{-1'}$ , et ici  $\phi^{-1'}(y) = 1/2$ .

Donc finalement, avec  $D = \mathbb{R}$  et  $\phi(D) = \mathbb{R}$  :

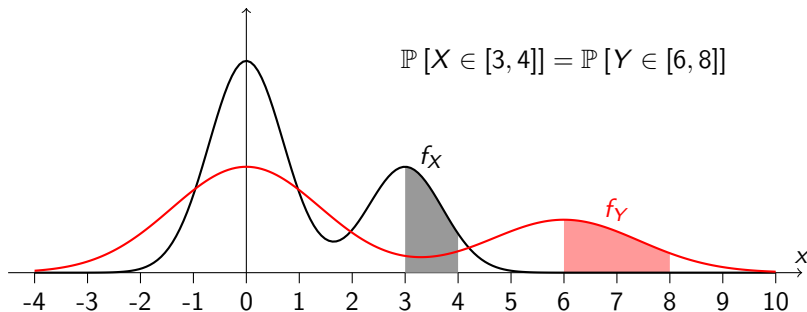
$$\begin{aligned} f_Y(y) &= \mathbf{1}_{\phi(D)}(y) f_U(\phi^{-1}(y)) |\det \text{Jac } \phi^{-1}(y)| \\ &= f_U((y - 1)/2) \times \frac{1}{2} = \frac{1}{2} \mathbf{1}_{[0,1]}((y - 1)/2) \\ &= \mathbf{1}_{[1,3]}(y) / \lambda([1, 3]). \end{aligned}$$

On retrouve bien  $Y \sim \mathcal{U}([1, 3])$ .

**Exemple 1 bis**

Si  $X$  est continue de densité  $f_X$ , et  $Y = 2X$ , alors  $Y$  est continue de densité

$$f_Y(y) = \frac{1}{2} f_X\left(\frac{y}{2}\right).$$

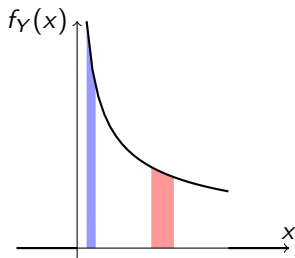
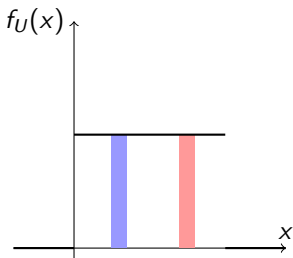


## Exemple 2

$$U \sim \mathcal{U}([0, 1]), \quad Y = U^2.$$

On peut choisir  $D = \mathbb{R}_+^*$ ,  $\phi(u) = u^2$ , de sorte que  $\phi^{-1}(y) = \sqrt{y}$  et  $\phi^{-1}'(y) = 1/(2\sqrt{y})$ .

$$\begin{aligned} f_Y(y) &= \mathbf{1}_{\phi(D)}(y) f_U(\phi^{-1}(y)) |\det \text{Jac } \phi^{-1}(y)| \\ &= \mathbf{1}_{\mathbb{R}_+^*}(y) \mathbf{1}_{[0,1]}(\sqrt{y}) \frac{1}{2\sqrt{y}} \\ &= \mathbf{1}_{]0,1]}(y) \frac{1}{2\sqrt{y}}. \end{aligned}$$





**Exemple 3**

$D$  disque unité de  $\mathbb{R}^2$ ,  $(U, V) \sim \mathcal{U}(D)$ .

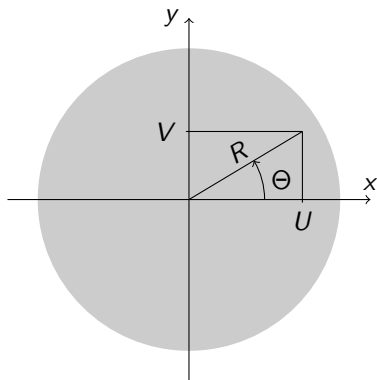
$$(R, \Theta) = \phi(U, V)$$

$$\phi : \mathbb{R}^2 \setminus \{(0, 0)\} \longrightarrow \mathbb{R}_+^2 \times [0, 2\pi[$$

$$\phi^{-1}(r, \theta) = (r \cos(\theta), r \sin(\theta))$$

$$\text{Jac } \phi^{-1}(r, \theta) = \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix}$$

$$|\det \text{Jac } \phi^{-1}(r, \theta)| = r$$



$$\begin{aligned} f_{(R, \Theta)}(r, \theta) &= \mathbf{1}_{\phi(D)}(r, \theta) f_{(U, V)}(\phi^{-1}(r, \theta)) |\det \text{Jac } \phi^{-1}(r, \theta)| \\ &= \frac{1}{\pi} \mathbf{1}_{]0, 1[}(r) \mathbf{1}_{[0, 2\pi[}(\theta) \times r \\ &= 2r \mathbf{1}_{]0, 1[}(r) \times \frac{1}{2\pi} \mathbf{1}_{[0, 2\pi[}(\theta) \end{aligned}$$

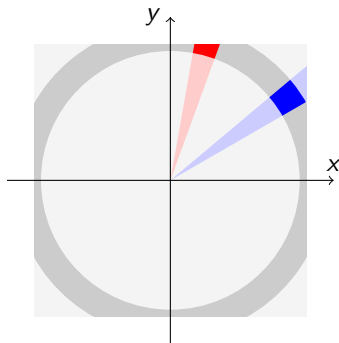
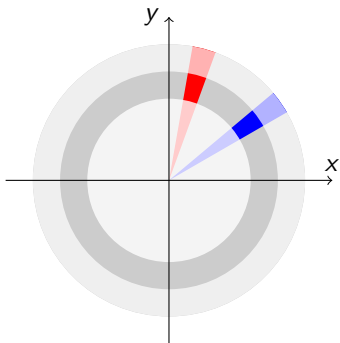
### Exemple 3 (suite)

$$f_{(R,\Theta)}(r, \theta) = 2r \mathbf{1}_{]0,1]}(r) \times \frac{1}{2\pi} \mathbf{1}_{[0,2\pi[}(\theta)$$

$$f_R(r) = \int f_{(R,\Theta)}(r, \theta) d\theta = 2r \mathbf{1}_{]0,1]}(r)$$

$$f_\Theta(\theta) = \int f_{(R,\Theta)}(r, \theta) dr = \frac{1}{2\pi} \mathbf{1}_{[0,2\pi[}(\theta)$$

$$\Theta \sim \mathcal{U}([0, 2\pi[)$$



# Densité et indépendance

## Propriété

1. Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes, continues, de densités respectives  $f_X$  et  $f_Y$ , alors  $(X, Y)$  est un vecteur aléatoire continu de densité  $f$  définie par

$$f(x, y) = f_X(x)f_Y(y).$$

2. Si  $(X, Y)$  est un vecteur aléatoire continu dont la densité  $f$  peut s'écrire sous la forme

$$f(x, y) = g(x)h(y),$$

alors  $X$  et  $Y$  sont deux variables aléatoires indépendantes de densités respectives  $g^*$  et  $h^*$ , avec

$$g^*(x) = \frac{g(x)}{\int g(u) du}, \quad h^*(x) = \frac{h(x)}{\int h(u) du}.$$

### Preuve du point 1

Soient  $X$  et  $Y$  deux v.a. indépendantes, de densités  $f_X$  et  $f_Y$ . Alors, si  $A$  et  $B$  sont deux boréliens de  $\mathbb{R}$ ,

$$\begin{aligned}\mathbb{P}_{(X,Y)}(A \times B) &= \mathbb{P}[(X, Y) \in A \times B] = \mathbb{P}[X \in A \text{ et } Y \in B] \\ &= \mathbb{P}[X \in A] \times \mathbb{P}[Y \in B] \\ \text{(Fubini)} &= \int \mathbf{1}_A(x) f_X(x) d\lambda(x) \times \int \mathbf{1}_B(y) f_Y(y) d\lambda(y) \\ &= \int \mathbf{1}_A(x) f_X(x) \mathbf{1}_B(y) f_Y(y) d\lambda_2(x, y) \\ &= \int \mathbf{1}_{A \times B}(x, y) f_X(x) f_Y(y) d\lambda_2(x, y)\end{aligned}$$

Donc  $\mathbb{P}_{(X,Y)}$  coïncide sur les pavés avec la mesure  $\mu$  de densité  $f(x, y) = f_X(x) f_Y(y)$  par rapport à la mesure de Lebesgue  $\lambda_2$ .

Donc  $\mathbb{P}_{(X,Y)} = \mu$ , c'est-à-dire que  $(X, Y)$  est continu de densité  $f$ .

## Preuve du point 2

Soit  $(X, Y)$  un vecteur aléatoire de densité  $f(x, y) = g(x)h(y)$ .

Toujours grâce au théorème de Fubini,

$$1 = \mathbb{P}_{(X, Y)}(\mathbb{R}^2) = \int f(x, y) d\lambda_2(x, y) = \int g(x) d\lambda(x) \times \int h(y) d\lambda(y)$$

On peut donc réécrire  $f$  sous la forme

$$f(x, y) = \underbrace{\frac{g(x)}{\int g(u) d\lambda(u)}}_{g^*(x)} \times \underbrace{\frac{h(y)}{\int h(v) d\lambda(v)}}_{h^*(y)},$$

où  $\int g^* d\lambda = \int h^* d\lambda = 1$ .

Ainsi,  $X$  est continue de densité

$$\int f(x, y) d\lambda(y) = g^*(x) \int h^*(y) d\lambda(y) = g^*(x),$$

et  $Y$  est continue de densité  $h^*$ .

## Preuve du point 2 (suite)

De plus, si  $A$  et  $B$  sont deux boréliens de  $\mathbb{R}$ ,

$$\begin{aligned}\mathbb{P}[X \in A \text{ et } Y \in B] &= \mathbb{P}_{(X,Y)}(A \times B) \\ &= \int_{A \times B} g^*(x)h^*(y) d\lambda_2(x,y) \\ (\text{Fubini}) &= \int_A g^*(x) d\lambda(x) \times \int_B h^*(y) d\lambda(y) \\ &= \mathbb{P}[X \in A] \times \mathbb{P}[Y \in B]\end{aligned}$$

# Fonction de répartition, fonction de survie

## Définition

On appelle **fonction de répartition** de la variable aléatoire  $X$  la fonction  $F_X$  définie sur  $\mathbb{R}$ , et à valeurs dans  $[0, 1]$ , par :

$$F_X(x) = \mathbb{P}[X \leq x] \quad \forall x \in \mathbb{R}.$$

On appelle **fonction de survie** de  $X$  la fonction  $S_X$  définie sur  $\mathbb{R}$ , et à valeurs dans  $[0, 1]$ , par :

$$S_X(x) = \mathbb{P}[X > x] \quad \forall x \in \mathbb{R}.$$

Si la variable aléatoire  $X$  modélise l'âge qu'aura Jojo lors de son décès,  $S_X(x)$  est la probabilité qu'il soit encore en vie à l'âge  $x$ .

# Fonction de répartition, fonction de survie

## Définition

On appelle **fonction de répartition** du vecteur aléatoire  $X = (X_1, \dots, X_d)$  à valeurs dans  $\mathbb{R}^d$  la fonction  $F_X$  définie sur  $\mathbb{R}^d$ , et à valeurs dans  $[0, 1]$ , par :

$$F_X(x) = \mathbb{P}[X_1 \leq x_1, \dots, X_d \leq x_d] \quad \forall x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

On appelle **fonction de survie** de  $X$  la fonction  $S_X$  définie sur  $\mathbb{R}^d$ , et à valeurs dans  $[0, 1]$ , par :

$$S_X(x) = \mathbb{P}[X_1 > x_1, \dots, X_d > x_d] \quad \forall x = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

On remarquera qu'en dimension 1,  $S = 1 - F$ , mais que ceci n'est plus vrai en dimension supérieure.



**Remarque :**

$$F_X(x) = \mathbb{P}_X(\{u \in \mathbb{R}^d, u_1 \leq x_1, \dots, u_d \leq x_d\}).$$

- ▶ La fonction de répartition ne dépend que de la *loi* de  $X$ . On parlera donc aussi de fonction de répartition *de la loi*  $\mathbb{P}_X$ .
- ▶ D'après le théorème d'identification de mesures, deux vecteurs aléatoires qui ont même fonction de répartition ont même loi. On dit que la fonction de répartition caractérise la loi (il en est de même pour la fonction de survie).

**Exemples :**

- ▶ loi exponentielle de paramètre  $\lambda$ ,
- ▶ lois discrètes.
- ▶ loi uniforme sur  $[0, 1]^2$ .

**Cas continu :** Si  $X$  est continue de densité  $f$ , alors  $F_X$  est dérivable aux points de continuité de  $f$ , avec  $F'_X = f$  en ces points.

### Exemple 1

Soit  $X$  une variable aléatoire de loi exponentielle de paramètre  $\lambda$ .  $X$  est continue de densité

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

La fonction de survie  $S$  de  $X$  peut donc s'écrire

$$\begin{aligned} S(x) &= \mathbb{P}[X > x] = \int_x^{+\infty} f(t) dt \\ &= \int_x^{+\infty} \mathbf{1}_{\mathbb{R}_+}(t) \lambda \exp(-\lambda t) dt \end{aligned}$$

Si  $x \geq 0$ ,

$$S(x) = \int_x^{+\infty} \lambda \exp(-\lambda t) dt = [\exp(-\lambda t)]_x^{+\infty} = e^{-\lambda x},$$

et si  $x < 0$ ,

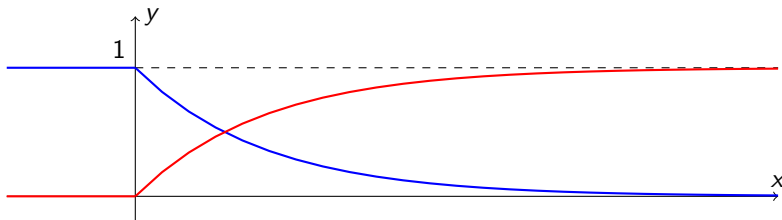
$$S(x) = \int_0^{+\infty} \lambda \exp(-\lambda x) dx = 1$$

## Exemple 1 (suite)

$$X \sim \mathcal{E}(\lambda)$$

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

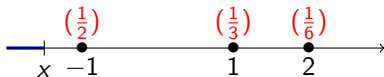
$$F_X(x) = 1 - S_X(x) = \mathbf{1}_{\mathbb{R}_+}(x)(1 - e^{-\lambda x})$$



## Exemple 2

Soit  $X$  une variable aléatoire discrète, dont la loi est déterminée par

$$X \in \{-1, 1, 2\}, \quad \mathbb{P}[X = -1] = \frac{1}{2}, \quad \mathbb{P}[X = 1] = \frac{1}{3}, \quad \mathbb{P}[X = 2] = \frac{1}{6}.$$

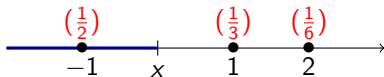


► Si  $x < -1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = 0$ .

## Exemple 2

Soit  $X$  une variable aléatoire discrète, dont la loi est déterminée par

$$X \in \{-1, 1, 2\}, \quad \mathbb{P}[X = -1] = \frac{1}{2}, \quad \mathbb{P}[X = 1] = \frac{1}{3}, \quad \mathbb{P}[X = 2] = \frac{1}{6}.$$

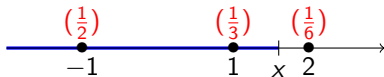


- ▶ Si  $x < -1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = 0$ .
- ▶ Si  $-1 \leq x < 1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2}$ .

## Exemple 2

Soit  $X$  une variable aléatoire discrète, dont la loi est déterminée par

$$X \in \{-1, 1, 2\}, \quad \mathbb{P}[X = -1] = \frac{1}{2}, \quad \mathbb{P}[X = 1] = \frac{1}{3}, \quad \mathbb{P}[X = 2] = \frac{1}{6}.$$

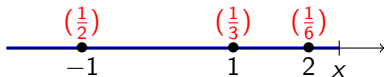


- ▶ Si  $x < -1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = 0$ .
- ▶ Si  $-1 \leq x < 1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2}$ .
- ▶ Si  $1 \leq x < 2$ ,  $F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ .

## Exemple 2

Soit  $X$  une variable aléatoire discrète, dont la loi est déterminée par

$$X \in \{-1, 1, 2\}, \quad \mathbb{P}[X = -1] = \frac{1}{2}, \quad \mathbb{P}[X = 1] = \frac{1}{3}, \quad \mathbb{P}[X = 2] = \frac{1}{6}.$$

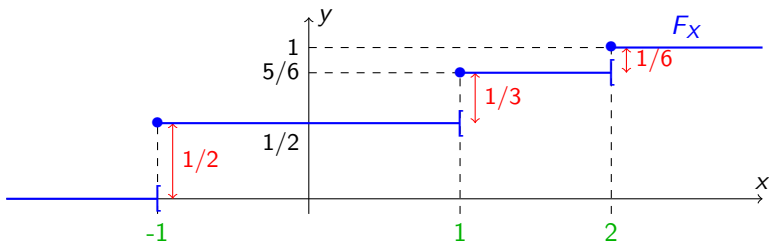


- ▶ Si  $x < -1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = 0$ .
- ▶ Si  $-1 \leq x < 1$ ,  $F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2}$ .
- ▶ Si  $1 \leq x < 2$ ,  $F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$ .
- ▶ Si  $x \geq 2$ ,  $F_X(x) = \mathbb{P}[X \leq x] = \frac{1}{2} + \frac{1}{3} + \frac{1}{6} = 1$ .

### Exemple 2 (suite)

Soit  $X$  une variable aléatoire discrète, dont la loi est déterminée par

$$X \in \{-1, 1, 2\}, \quad \mathbb{P}[X = -1] = \frac{1}{2}, \quad \mathbb{P}[X = 1] = \frac{1}{3}, \quad \mathbb{P}[X = 2] = \frac{1}{6}.$$



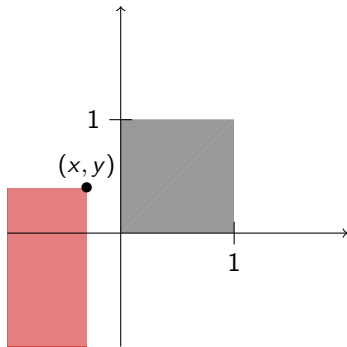


**Exemple 3**

$(X, Y) \sim \mathcal{U}([0, 1]^2)$ , de fonction de répartition  $F$ .

$$F(x, y) = \mathbb{P}_{(X, Y)}(\underbrace{]-\infty, x] \times ]-\infty, y]}_{Q_{x,y}} \cap [0, 1]^2).$$

► Si  $x < 0$  ou  $y < 0$ ,  $F(x, y) = 0$ .

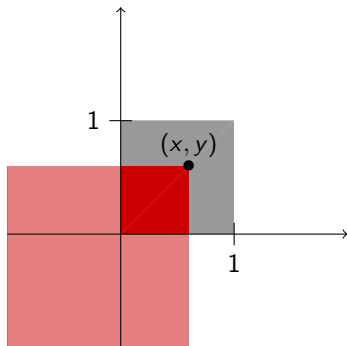


**Exemple 3**

$(X, Y) \sim \mathcal{U}([0, 1]^2)$ , de fonction de répartition  $F$ .

$$F(x, y) = \mathbb{P}_{(X, Y)}(\underbrace{]-\infty, x] \times ]-\infty, y]}_{Q_{x,y}} \cap [0, 1]^2).$$

- ▶ Si  $x < 0$  ou  $y < 0$ ,  $F(x, y) = 0$ .
- ▶ Si  $x, y \in [0, 1]$ ,  $F(x, y) = xy$ .

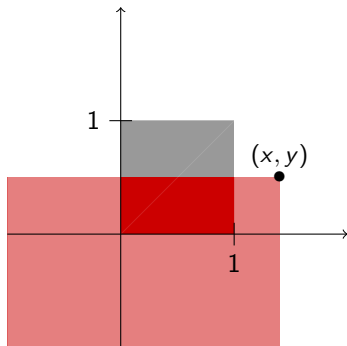


**Exemple 3**

$(X, Y) \sim \mathcal{U}([0, 1]^2)$ , de fonction de répartition  $F$ .

$$F(x, y) = \mathbb{P}_{(X, Y)}(\underbrace{]-\infty, x] \times ]-\infty, y]}_{Q_{x,y}} \cap [0, 1]^2).$$

- ▶ Si  $x < 0$  ou  $y < 0$ ,  $F(x, y) = 0$ .
- ▶ Si  $x, y \in [0, 1]$ ,  $F(x, y) = xy$ .
- ▶ Si  $x > 1$  et  $y \in [0, 1]$ ,  $F(x, y) = y$ .

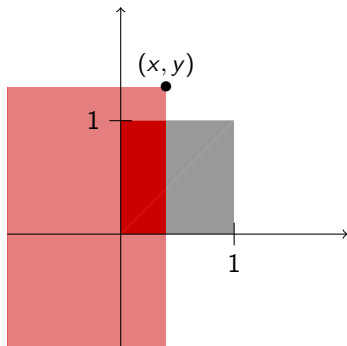


**Exemple 3**

$(X, Y) \sim \mathcal{U}([0, 1]^2)$ , de fonction de répartition  $F$ .

$$F(x, y) = \mathbb{P}_{(X, Y)}(\underbrace{]-\infty, x] \times ]-\infty, y]}_{Q_{x,y}} \cap [0, 1]^2).$$

- ▶ Si  $x < 0$  ou  $y < 0$ ,  $F(x, y) = 0$ .
- ▶ Si  $x, y \in [0, 1]$ ,  $F(x, y) = xy$ .
- ▶ Si  $x > 1$  et  $y \in [0, 1]$ ,  $F(x, y) = y$ .
- ▶ Si  $x \in [0, 1]$  et  $y > 1$ ,  $F(x, y) = x$ .

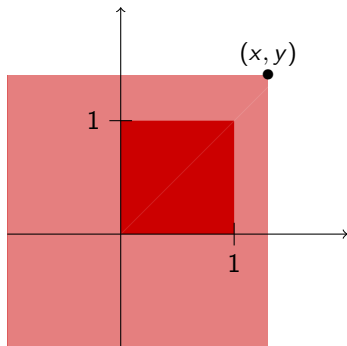


**Exemple 3**

$(X, Y) \sim \mathcal{U}([0, 1]^2)$ , de fonction de répartition  $F$ .

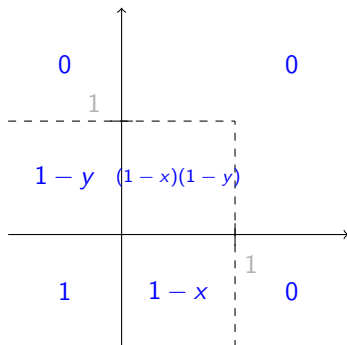
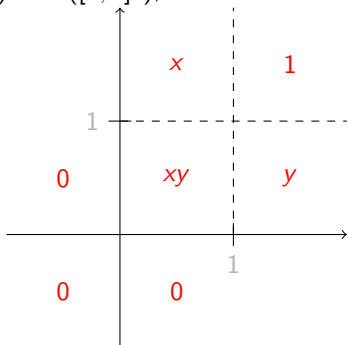
$$F(x, y) = \mathbb{P}_{(X, Y)}(\underbrace{]--\infty, x] \times ]-\infty, y]}_{Q_{x,y}} \cap [0, 1]^2).$$

- ▶ Si  $x < 0$  ou  $y < 0$ ,  $F(x, y) = 0$ .
- ▶ Si  $x, y \in [0, 1]$ ,  $F(x, y) = xy$ .
- ▶ Si  $x > 1$  et  $y \in [0, 1]$ ,  $F(x, y) = y$ .
- ▶ Si  $x \in [0, 1]$  et  $y > 1$ ,  $F(x, y) = x$ .
- ▶ Si  $x > 1$  et  $y > 1$ ,  $F(x, y) = 1$ .



**Exemple3 (suite)**

$(X, Y) \sim \mathcal{U}([0, 1]^2)$ , de fonction de répartition  $F$  et de fonction de survie  $S$ .



**Remarque :**  $F$  et  $S$  sont continues sur  $\mathbb{R}^2$ .

## Propriétés

Soit  $F_X$  la fonction de répartition d'une variable aléatoire  $X$ .

1.  $F_X$  est croissante.
2.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .
3.  $F_X$  est continue à droite.
4.  $F_X$  a une limite à gauche en tout point.
5. En tout point  $x_0$ ,  $F_X(x_0) - \lim_{x_0^-} F_X = \mathbb{P}(X = x_0)$ .

## Propriétés

Soit  $F_X$  la fonction de répartition d'un vecteur aléatoire  $X = (X_1, \dots, X_d)$ .

1.  $F_X$  est croissante, càd-làg en chacune des coordonnées.
2.  $\lim_{x_i \rightarrow -\infty} F_X(x) = 0$  pour tout  $i$ , et  $\lim_{x_1, \dots, x_d \rightarrow +\infty} F_X(x) = 1$ .
3. Si  $J \subset \{1, \dots, d\}$ , le sous-vecteur  $Y = (X_i)_{i \in J}$  a pour fonction de répartition la limite de  $F_X$  quand les  $x_i$ ,  $i \notin J$  tendent vers  $+\infty$ .

**Preuve** ( $d = 1$ )

1.  $F_X$  est croissante.

Si  $x \leq x'$ ,  $] - \infty, x] \subset ] - \infty, x']$  donc

$$\mathbb{P}_X(] - \infty, x]) \leq \mathbb{P}_X(] - \infty, x'])$$

C'est-à-dire

$$F_X(x) \leq F_X(x')$$



**Preuve** ( $d = 1$ )

2.  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

$$\bigcap_n ]-\infty, -n] = \emptyset$$

donc

$$\lim_{n \rightarrow \infty} \mathbb{P}_X(]-\infty, -n]) = 0$$

soit

$$\lim_{n \rightarrow -\infty} F_X(n) = 0.$$

Comme  $F_X$  est croissante, on en déduit que

$$\lim_{x \rightarrow -\infty} F_X(x) = 0.$$

On démontre de même que  $\lim_{x \rightarrow +\infty} F_X(x) = 1$  à partir de  $\cup_n ]-\infty, n] = \mathbb{R}$ .

**Preuve** ( $d = 1$ )

3.  $F_X$  est continue à droite.

$$\bigcap_n ]-\infty, x_0 + 1/n] = ]-\infty, x_0]$$

donc

$$\lim_{n \rightarrow \infty} F_X(x_0 + 1/n) = F_X(x_0).$$

Comme  $F_X$  est croissante, on en déduit que

$$\lim_{x \rightarrow x_0^+} F_X(x) = F_X(x_0).$$

**Preuve** ( $d = 1$ )

4.  $F_X$  a une limite à gauche en tout point.
5. En tout point  $x_0$ ,  $F_X(x_0) - \lim_{x \rightarrow x_0^-} F_X = \mathbb{P}(X = x_0)$ .

$$\bigcup_n ]-\infty, x_0 - 1/n] = ]-\infty, x_0[$$

et la croissance de  $F_X$  impliquent

$$\lim_{x \rightarrow x_0^-} F_X(x) = \mathbb{P}[X < x_0] = \mathbb{P}[X \leq x_0] - \mathbb{P}[X = x_0].$$

**Remarque :**  $F_X$  est continue en  $x_0$  si et seulement si  $\mathbb{P}[X = x_0] = 0$ .

**Preuve** ( $d = 2$ )

1.  $F_X$  est croissante, càd-làg en chacune des coordonnées.

$$F_X(x, y) = \mathbb{P}_{(X, Y)}(Q_{x, y})$$

Si  $x \leq x'$  et  $y \leq y'$ ,  $Q_{x, y} \subset Q_{x', y}$  et  $Q_{x, y} \subset Q_{x, y'}$

$$\bigcap_n Q_{x_0+1/n, y_0} = Q_{x_0, y_0}$$

$$\bigcup_n Q_{x_0-1/n, y_0} = \{(x, y), x < x_0, y \leq y_0\}$$

**Preuve** ( $d = 2$ )

2.  $\lim_{x_i \rightarrow -\infty} F_X(x) = 0$  pour tout  $i$ , et  $\lim_{x_1, \dots, x_d \rightarrow +\infty} F_X(x) = 1$ .

$$\bigcap_n Q_{-n, y_0} = \emptyset, \quad \bigcap_n Q_{x_0, -n} = \emptyset$$

$$\bigcup_{m, n} Q_{m, n} = \mathbb{R}^2$$

**Preuve** ( $d = 2$ )

3. Si  $J \subset \{1, \dots, d\}$ , le sous-vecteur  $Y = (X_i)_{i \in J}$  a pour fonction de répartition la limite de  $F_X$  quand les  $x_i$ ,  $i \notin J$  tendent vers  $+\infty$ .

$$\bigcup_n Q_{n,y_0} = \{(x_1, x_2), x_2 \leq y_0\}$$

donc

$$\lim_{x \rightarrow +\infty} F_{(X_1, X_2)}(x, y_0) = \mathbb{P}[X_2 \leq y_0] = F_{X_2}(y_0)$$

et de même

$$\lim_{y \rightarrow +\infty} F_{(X_1, X_2)}(x_0, y) = \mathbb{P}[X_1 \leq x_0] = F_{X_1}(x_0)$$

## Propriétés caractéristiques (dimension 1)

Une fonction  $F : \mathbb{R} \rightarrow \mathbb{R}$  càd-làg croissante qui a pour limite 0 en  $-\infty$  et 1 en  $+\infty$  est la fonction de répartition d'une loi de probabilité sur  $\mathbb{R}$ .

## Propriétés caractéristiques (dimension $d$ )

Soit  $F : \mathbb{R}^d \rightarrow \mathbb{R}$  une fonction telle que :

- ▶  $F$  est croissante et càd-làg en toutes les coordonnées,
- ▶  $\lim_{x_i \rightarrow -\infty} F(x) = 0 \forall i$  et  $\lim_{x_1, \dots, x_d \rightarrow +\infty} F(x) = 1$ ,
- ▶ pour tout  $\varepsilon \in \mathbb{R}_+^d$ , pour tout  $x \in \mathbb{R}^d$ , et avec la notation  $\varepsilon_J = (\mathbf{1}_J(i)\varepsilon_i)_i$ ,

$$F(x + \varepsilon) - \sum_{|J|=d-1} F(x + \varepsilon_J) + \sum_{|J|=d-2} F(x + \varepsilon_J) - \dots + (-1)^d F(x) \geq 0.$$

Alors  $F$  est la fonction de répartition d'une mesure de probabilité sur  $\mathbb{R}^d$ .

## Idée de la preuve en dimension 1

On définit une fonction  $\mu$  sur les intervalles de la forme  $]a, b]$  par

$$\mu(]a, b]) = F(b) - F(a)$$

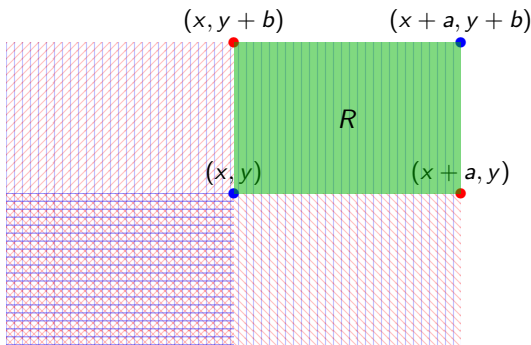
puis on utilise le théorème de prolongement de Carathéodory (ces intervalles forment un semi-anneau qui engendre  $\mathcal{B}(\mathbb{R})$ ) pour la prolonger en une mesure  $\mu$  sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  qui vérifie donc  $\mu(]-\infty, b]) = F(b)$ .



## Note sur la dernière propriété en dimension $d = 2$

$$F(x + \varepsilon) - \sum_{|J|=d-1} F(x + \varepsilon_J) + \sum_{|J|=d-2} F(x + \varepsilon_J) - \cdots + (-1)^d F(x) \geq 0.$$

$$\begin{aligned} & F(x + a, y + b) - F(x + a, y) - F(x, y + b) + F(x, y) \\ &= P_{(X, Y)}(Q_{x+a, y+b}) - P_{(X, Y)}(Q_{x+a, y}) - P_{(X, Y)}(Q_{x, y+b}) + P_{(X, Y)}(Q_{x, y}) \\ &= P_{(X, Y)}(R) \end{aligned}$$



# Fonctions de répartition et indépendance

## Propriété

Deux vecteurs aléatoires  $X$  et  $Y$  à valeurs dans  $\mathbb{R}^n$  et  $\mathbb{R}^m$  sont indépendants si et seulement si

$$F_{(X,Y)}(x,y) = F_X(x)F_Y(y) \quad \forall x \in \mathbb{R}^n, y \in \mathbb{R}^m.$$

## Preuve :

- ▶ Si  $X$  et  $Y$  sont indépendantes,

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

- ▶ Si  $F_{(X,Y)}(x,y) = F_X(x)F_Y(y)$ , on peut montrer par le critère d'indépendance que  $X$  et  $Y$  sont indépendantes. En effet,

$$\mathcal{C}_X = \{\{X \leq x\}, x \in \mathbb{R}\} = X^{-1}(\{]-\infty, x], x \in \mathbb{R}\})$$

est un  $\pi$ -système qui engendre  $\sigma_X = X^{-1}(\mathcal{B}(\mathbb{R}))$ .

# Fonctions dérivées

## Définition

La fonction quantile d'une variable aléatoire  $X$  est la fonction  $Q_X = F_X^{\leftarrow}$  définie par

$$Q_X(p) = F_X^{\leftarrow}(p) = \inf\{x, F_X(x) > p\} \quad \forall p \in ]0, 1[.$$

## Exemples :

- ▶ Soit  $X$  le montant aléatoire d'un sinistre. Quelle est la valeur  $x$  telle que le montant d'un sinistre est inférieur à  $x$  avec probabilité 0.95 ? C'est  $Q_X(0.95)$ .
- ▶ Soit  $X$  le revenu annuel d'un ménage fiscal choisi au hasard. « 10 % des ménages ont un revenu inférieur à 9 400 € » peut s'écrire  $Q_X(0.1) = 9\,400$ .

## Définition (Taux de hasard)

On appelle taux de hasard d'une variable aléatoire  $X$  la fonction  $h : \mathbb{R} \rightarrow \mathbb{R}$  définie (quand elle existe) par

$$h(x) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{P}[X \leq x + \delta \mid X > x].$$

**Exemple** : taux de défaillance d'un équipement, taux de mortalité en temps continu, équipement sans usure.

# Table de mortalité

<https://www.ined.fr/fr/tout-savoir-population/chiffres/france/mortalite-cause-deces/table-mortalite/>

TABLEAU 68 - TABLE DE MORTALITÉ DES ANNÉES 2016 – 2018

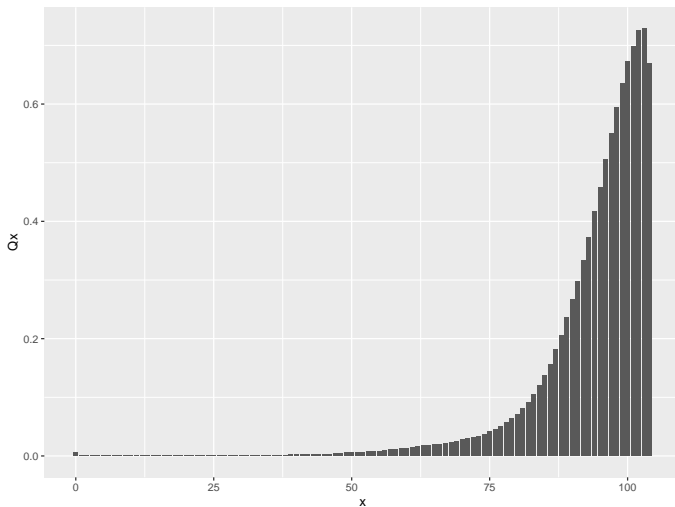
Âge x	Survivants S(x) à l'âge x	Ensemble	
		Quotient de mortalité Q(x, x+1) pour 100 000 survivants à l'âge x (âge en années révolues)	Espérance de vie E(x) à l'âge x
0	100 000		82,35
...	...	...	...
18	99 439	29	64,80
19	99 410	34	63,82
20	99 377	37	62,84
21	99 340	40	61,86
22	99 300	40	60,89
23	99 261	42	59,91
24	99 219	42	58,94
25	99 177	45	57,96

Si  $X$  modélise l'âge au décès d'un Français (en années),

$$S_X(x) = \frac{S(x)}{100\,000}, \quad \mathbb{P}[X \leq x+1 | X > x] = \frac{Q(x, x+1)}{100\,000}$$

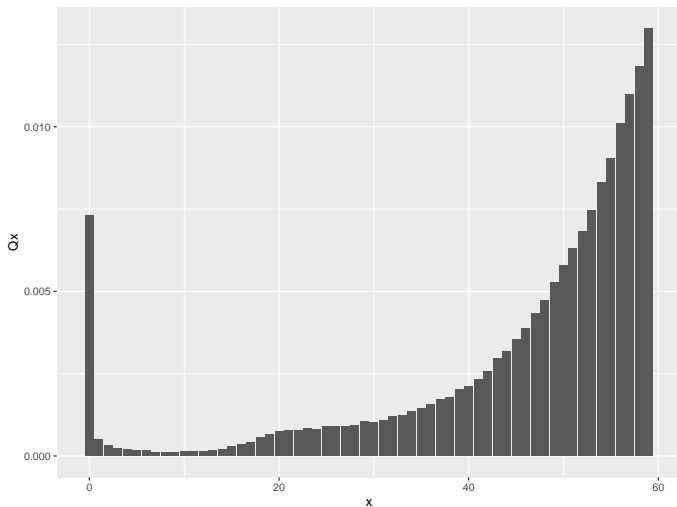
## Probabilités de décès dans l'année

$$\mathbb{P}[X \leq x + 1 | X > x] = \frac{Q(x, x + 1)}{100\,000}$$



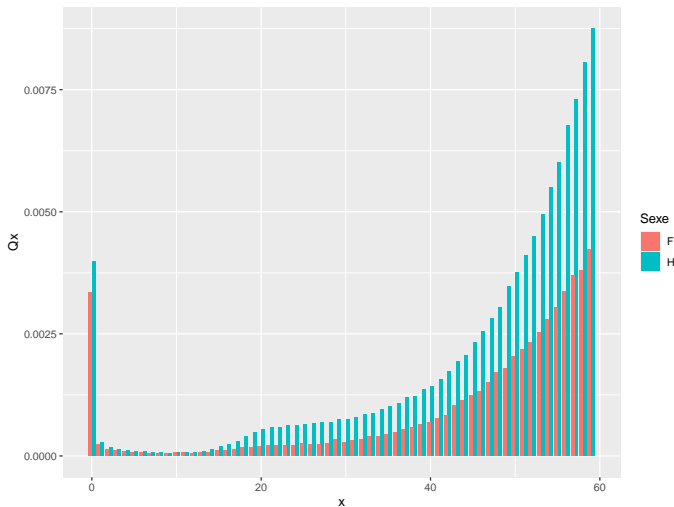
## Probabilités de décès dans l'année

$$\mathbb{P}[X \leq x + 1 | X > x] = \frac{Q(x, x + 1)}{100\,000}$$



# Probabilités de décès dans l'année

$$\mathbb{P}[X \leq x + 1 | X > x] = \frac{Q(x, x + 1)}{100\,000}$$





## Taux de mortalité instantané

$$\begin{aligned} & \mathbb{P}[X \leq x + 1 \text{ min} \mid X > x] \\ &= 1 - \mathbb{P}[X > x + 1 \text{ min} \mid X > x] \\ &= 1 - \mathbb{P}[X > x + 1 \text{ s} \mid X > x] \times \mathbb{P}[X > x + 2 \text{ s} \mid X > x + 1 \text{ s}] \times \cdots \\ &\quad \cdots \times \mathbb{P}[X > x + 60 \text{ s} \mid X > x + 59 \text{ s}] \\ &\simeq 1 - \mathbb{P}[X > x + 1 \text{ s} \mid X > x]^{60} \\ &\simeq 1 - (1 - \mathbb{P}[X \leq x + 1 \text{ s} \mid X > x])^{60} \\ &\simeq 60 \times \mathbb{P}[X \leq x + 1 \text{ s} \mid X > x] \\ &\simeq \frac{1}{1/60} \times \mathbb{P}\left[X \leq x + \frac{1}{60} \text{ min} \mid X > x\right] \end{aligned}$$

D'où la définition du taux de mortalité *instantané* :

$$h(x) = \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbb{P}[X \leq x + \delta \mid X > x] .$$

## Lien avec la densité

Dans le cas d'une v.a.  $X$  continue de densité  $f$  et de fonction de répartition  $F$ ,

$$\begin{aligned}\frac{1}{\delta} \mathbb{P}[X \leq x + \delta \mid X > x] &= \frac{1}{\delta} \mathbb{P}[x < X \leq x + \delta] / \mathbb{P}[X > x] \\ &= \frac{1}{1 - F(x)} \frac{1}{\delta} \int_x^{x+\delta} f(t) dt\end{aligned}$$

Donc

$$h(x) = \frac{f(x)}{1 - F(x)}$$

## Équipement sans usure

Dans le cas « sans usure »,  $h$  est une fonction constante :  $h(x) = \lambda$  pour  $x \geq 0$ .

$$\frac{f(x)}{1 - F(x)} = \lambda$$

$$\frac{(1 - F)'(x)}{1 - F(x)} = -\lambda$$

$$1 - F(x) = \exp(-\lambda x + c)$$

$X \geq 0$ , donc  $F(0) = 0$ , donc  $c = 0$  et ainsi

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

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

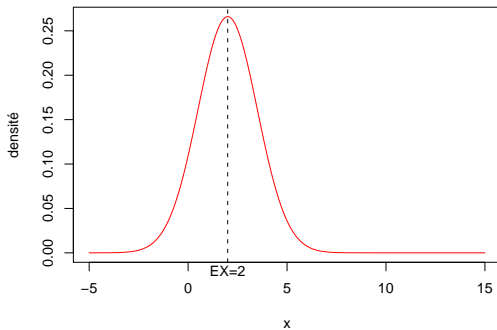
# Moments

## Définition (Moments)

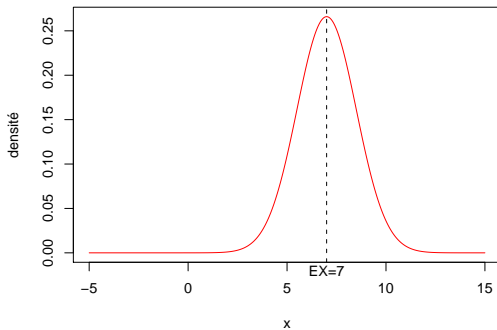
Si  $X$  est une variable aléatoire et  $n \in \mathbb{N}^*$ , on appelle **moment d'ordre  $n$**  de  $X$  l'espérance  $m_n = E(X^n)$ , quand elle existe.

On appelle **moment centré d'ordre  $n$**  de la variable aléatoire  $X$  l'espérance  $\mu_n = E[(X - E X)^n]$ , quand elle existe.

- L'**espérance** est le moment d'ordre 1 d'une variable aléatoire intégrable. Elle permet de mesurer la *localisation* de la variable.



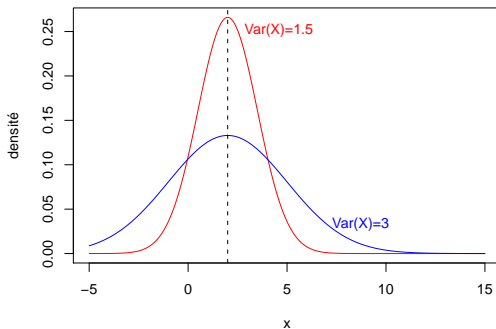
- L'**espérance** est le moment d'ordre 1 d'une variable aléatoire intégrable. Elle permet de mesurer la *localisation* de la variable.



- ▶ L'**espérance** est le moment d'ordre 1 d'une variable aléatoire intégrable. Elle permet de mesurer la *localisation* de la variable.
- ▶ La **variance** est le moment centré d'ordre 2 d'une variable aléatoire de carré intégrable :

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

Elle permet de mesurer la *dispersion* de la variable.



On définit également l'**écart-type** de  $X$  par  $\sigma_X = \sqrt{\text{Var}(X)}$ .

- ▶ L'**espérance** est le moment d'ordre 1 d'une variable aléatoire intégrable. Elle permet de mesurer la *localisation* de la variable.
- ▶ La **variance** est le moment centré d'ordre 2 d'une variable aléatoire de carré intégrable :

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

Elle permet de mesurer la *dispersion* de la variable. On définit également l'**écart-type** de  $X$  par  $\sigma_X = \sqrt{\text{Var}(X)}$ .

**Remarque** : la variance d'une variable aléatoire est positive ou nulle, et est nulle si et seulement si cette variable est presque-sûrement *constante*.

**Exemples** : loi de Poisson et loi exponentielle.



# Normalisation

On dit qu'une variable aléatoire  $X$  est **centrée** si son espérance est nulle. On dit qu'une variable aléatoire est **réduite** si sa variance vaut 1.

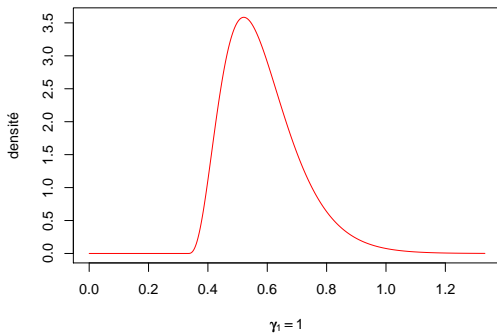
Si  $X$  est une variable aléatoire de carré intégrable non constante, on appellera variable  $X$  centrée réduite la variable

$$\frac{X - EX}{\sqrt{\text{Var } X}}.$$

- Le **coefficient de dissymétrie (skewness)** d'une variable aléatoire  $X$  est le moment d'ordre 3 centré réduit de  $X$ , défini par

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X - E X)^3]}{(\text{Var } X)^{3/2}}.$$

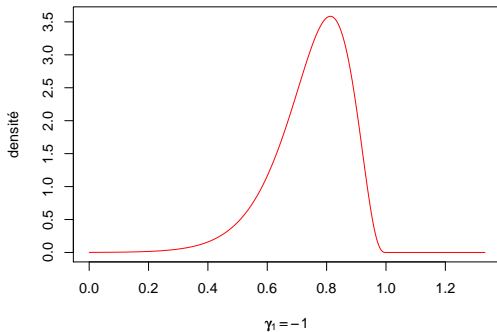
Il mesure l'*asymétrie* de la loi de  $X$ .



- Le **coefficient de dissymétrie (skewness)** d'une variable aléatoire  $X$  est le moment d'ordre 3 centré réduit de  $X$ , défini par

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X - E X)^3]}{(\text{Var } X)^{3/2}}.$$

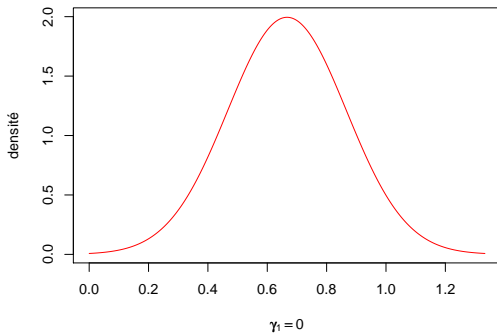
Il mesure l'*asymétrie* de la loi de  $X$ .



- Le **coefficient de dissymétrie (skewness)** d'une variable aléatoire  $X$  est le moment d'ordre 3 centré réduit de  $X$ , défini par

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X - E X)^3]}{(\text{Var } X)^{3/2}}.$$

Il mesure l'*asymétrie* de la loi de  $X$ .



- Le **coefficient d'applatissage (kurtosis)** d'une variable aléatoire  $X$  est le moment d'ordre 4 centré réduit de  $X$ , défini par

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{\mathbb{E}[(X - \mathbb{E} X)^4]}{(\text{Var } X)^2} \in [1, +\infty[.$$

Une grande valeur de  $\beta_2$  est caractéristique de densités avec un pic prononcé et de grandes valeurs loin de zéro, tandis qu'une petite valeur de  $\beta_2$  correspond à une densité avec un mode arrondi et qui tend rapidement vers zéro en l'infini.

**Remarque :** on peut aussi rencontrer la définition  $\mu_4/\sigma^4 - 3$ ; 3 étant la valeur du coefficient d'applatissage de la loi normale.

# Covariance

## Définition

On appelle **covariance** de deux variables aléatoires  $X$  et  $Y$  de carré intégrable la valeur

$$\text{Cov}(X, Y) = E[(X - E X)(Y - E Y)] = E(XY) - E X \times E Y .$$

On appelle **coefficient de corrélation linéaire** de  $X$  et  $Y$  la valeur

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} .$$

## Propriétés de la variance et de la covariance

Soient  $X$  et  $Y$  deux variables aléatoires de carré intégrable.

1. La covariance est bilinéaire et symétrique.
2. Si  $X$  et  $Y$  sont indépendantes, alors  $\text{Cov}(X, Y) = 0$ .
3. Si  $f : \mathbb{R} \rightarrow \mathbb{R}$  est croissante,  $\text{Cov}(X, f(X)) \geq 0$ .
4.  $\text{Var } X = \text{Cov}(X, X)$ .
5. Si  $\alpha, \beta \in \mathbb{R}$ ,  $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$ .
6.  $\text{Var}(X + Y) = \text{Var } X + \text{Var } Y + 2 \text{Cov}(X, Y)$ .
7.  $|\rho_{X,Y}| \leq 1$ .

**Remarque :** si  $\text{Cov}(X, Y) = 0$ , alors  $X$  et  $Y$  ne sont pas forcément indépendantes !

# Inégalités de Bienaymé-Tchebychev et Markov

## Inégalité de Markov

Soient  $X$  une variable aléatoire et  $r > 0$  tels que  $X^r$  est intégrable. Alors, pour tout  $\delta > 0$ ,

$$\mathbb{P}[|X| > \delta] \leq \frac{\mathbb{E}(|X|^r)}{\delta^r}.$$

## Inégalité de Bienaymé-Tchebychev

Soit  $X$  une variable aléatoire de carré intégrable, et  $\delta \in \mathbb{R}_+^*$ . Alors

$$\mathbb{P}[|X - \mathbb{E}X| > \delta] \leq \frac{\text{Var} X}{\delta^2}.$$



## Exemple : risques corrélés ou indépendants

Soit  $X$  le montant d'un risque auquel on est exposé, et  $\alpha \in ]0, 1[$ . Déterminons, en fonction de  $m = \mathbb{E} X$  et  $\sigma^2 = \text{Var} X$ , une valeur  $c^*$  telle que

$$\mathbb{P}[X > c^*] \leq \alpha.$$

On considère  $N$  risques individuels  $X_i$ , et on note  $X = \sum X_i$  le risque agrégé. Comparons  $c^*$  dans un cas d'indépendance et dans un cas de dépendance forte, lorsque  $N$  est grand.

# Cas vectoriel

## Définition

Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ .

- ▶ On dit que  $X$  est intégrable si toutes ses coordonnées sont intégrables, et, dans ce cas, on note

$$E X = (E X_1, \dots, E X_n).$$

- ▶ Si toutes les coordonnées de  $X$  sont de carré intégrable, on appelle **matrice de covariance** (ou matrice de variance-covariance) de  $X$ , la matrice  $n \times n$  notée  $\text{Var } X$  et dont les coefficients sont les covariances des coordonnées de  $X$  :

$$(\text{Var } X)_{ij} = \text{Cov}(X_i, X_j).$$

## Propriétés

Soit  $X$  un vecteur aléatoire.



$$\text{Var } X = E[(X - E X)^t (X - E X)].$$

- ▶ Si  $A$  est une matrice et  $B$  un vecteur déterministes tels que les dimensions sont compatibles dans la formule suivante,

$$\text{Var}(AX + B) = A(\text{Var } X) {}^t A.$$

- ▶ Pour tout  $\delta > 0$ ,

$$\mathbb{P} [\|X - E X\| > \delta] \leq \frac{\text{Trace Var } X}{\delta^2}.$$

# Convolution

## Rappel : loi d'un vecteur, que peut-on déduire de quoi ?

On considère un vecteur aléatoire  $X = (X_1, \dots, X_d)$ .

- ▶ Si la loi de  $X$  est connue, alors la loi de chacune des coordonnées est déterminée.
- ▶ Si la loi de chacune des variables  $X_i$  est connue, **et que les  $X_i$  sont mutuellement indépendantes**, alors la loi de  $X$  est déterminée.
- ▶ Si la loi de  $X$  est connue, et si  $f : \mathbb{R}^d \rightarrow \mathbb{R}^n$  est mesurable, alors la loi de  $f(X)$  est déterminée.

Exemple :  $X_1 + X_2$

# Convolution

## Rappel : loi d'un vecteur, que peut-on déduire de quoi ?

On considère un vecteur aléatoire  $X = (X_1, \dots, X_d)$ .

- ▶ Si la loi de chacune des variables  $X_i$  est connue, sans hypothèse supplémentaire sur la manière dont les  $X_i$  sont liées les unes aux autres, alors la loi du vecteur  $X$  n'est pas déterminée, et les lois de variables du type  $(X_1 + X_2)$  ne sont pas déterminées.

Exemple : dans les deux situations suivantes,  $X_1$  et  $X_2$  suivent tous les deux la loi uniforme sur  $[0, 1]$ . Par contre, la loi de  $X_1 + X_2$  dans la situation 1 est différente de celle de  $X_1 + X_2$  dans la situation 2.

1. Soit  $X_1$  de loi  $\mathcal{U}([0, 1])$ , et  $X_2 = 1 - X_1$ . La loi de  $(X_1, X_2)$  n'est pas continue, et  $X_1 + X_2 = 1$ .
2. Soient  $X_1$  et  $X_2$  deux variables indépendantes de loi  $\mathcal{U}([0, 1])$ . La loi de  $(X_1, X_2)$  est continue, et, par exemple,  $\mathbb{P}[X_1 + X_2 < 1] = 1/2$ .

## Définition

On appelle **produit de convolution** des probabilités  $\mathbb{P}_1$  et  $\mathbb{P}_2$ , notée  $\mathbb{P}_1 \star \mathbb{P}_2$ , l'image par  $(x, y) \mapsto x + y$  de  $\mathbb{P}_1 \otimes \mathbb{P}_2$ . Ainsi, si  $X_1$  et  $X_2$  sont indépendantes de lois respectives  $\mathbb{P}_1$  et  $\mathbb{P}_2$ , alors  $X_1 + X_2$  a pour loi  $\mathbb{P}_1 \star \mathbb{P}_2$ .

**Cas discret** : si  $X$  et  $Y$  sont discrètes et indépendantes, alors  $X + Y$  est discrète, et sa loi est définie par  $(X + Y)(\Omega) = X(\Omega) + Y(\Omega)$  et

$$\mathbb{P}[X + Y = k] = \sum_{y \in Y(\Omega)} \mathbb{P}[Y = y] \mathbb{P}[X = k - y].$$

**Cas continu** : si  $X$  et  $Y$  sont continues et indépendantes, de densités respectives  $f_X$  et  $f_Y$ , alors  $X + Y$  est continue, de densité

$$f_{X+Y}(z) = \int f_Y(y) f_X(z - y) dy.$$

On dira aussi que  $f_{X+Y}$  est le produit de convolution de  $f_X$  et  $f_Y$ , et on notera  $f_{X+Y} = f_X \star f_Y$ .

# Lois de probabilité usuelles

Un petit catalogue des lois de probabilité que nous rencontrerons très régulièrement...

# Loi de Bernoulli

**Application :** loi du pile ou face généralisé, loi d'une indicatrice.

**Expression :** pour  $p \in [0, 1]$ ,  $X \sim \mathcal{B}(p)$  signifie  $X(\Omega) = \{0, 1\}$  et

$$\mathbb{P}[X = 1] = p, \quad \mathbb{P}[X = 0] = 1 - p.$$

**Caractéristiques :**

$$E X = p, \quad \text{Var } X = p(1 - p).$$



## Loi binomiale

**Application :** si on réalise  $n$  expériences indépendantes qui ont chacune la même probabilité  $p$  de réussir, c'est la loi du nombre de réussites.

**Expression :** pour  $n \in \mathbb{N}$ ,  $p \in [0, 1]$ ,  $X \sim \mathcal{B}(n, p)$  signifie  $X(\Omega) = \{0, \dots, n\}$  et

$$\mathbb{P}[X = k] = C_n^k p^k (1 - p)^{n-k} \quad \forall k \in \{0, \dots, n\}.$$

**Caractéristiques :**

$$E X = np, \quad \text{Var } X = np(1 - p).$$

**Propriétés :**

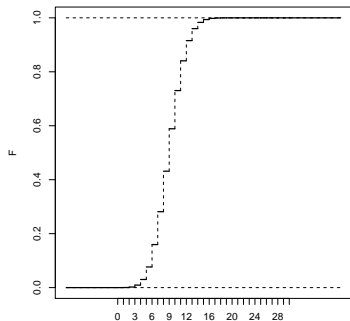
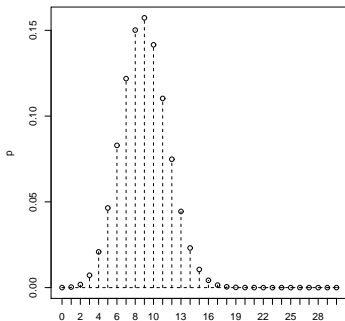
- ▶ Si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes de loi de Bernoulli de paramètre  $p$ , alors  $X = X_1 + \dots + X_n$  suit le loi  $\mathcal{B}(n, p)$ .
- ▶ Si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes de lois respectives  $\mathcal{B}(n_i, p)$ ,  $i = 1, \dots, n$ , alors  $X = X_1 + \dots + X_n$  suit le loi  $\mathcal{B}(n, p)$ , où  $n = \sum_i n_i$ .

## Loi binomiale

**Application** : si on réalise  $n$  expériences indépendantes qui ont chacune la même probabilité  $p$  de réussir, c'est la loi du nombre de réussites.

**Expression** : pour  $n \in \mathbb{N}$ ,  $p \in [0, 1]$ ,  $X \sim \mathcal{B}(n, p)$  signifie  $X(\Omega) = \{0, \dots, n\}$  et

$$\mathbb{P}[X = k] = C_n^k p^k (1-p)^{n-k} \quad \forall k \in \{0, \dots, n\}.$$



$\mathcal{B}(30, 0.3)$

## Loi géométrique

**Application :** si on réalise une suite (infinie) d'expériences indépendantes qui ont chacune la même probabilité  $p$  de réussir, c'est la loi du nombre d'expériences réalisées jusqu'à la première réussite.

**Expression :** pour  $p \in ]0, 1[$ ,  $X \sim \mathcal{G}(p)$  signifie  $X(\Omega) = \mathbb{N}^*$  et

$$\mathbb{P}[X = k] = (1 - p)^{k-1} p \quad \forall k \in \mathbb{N}^* .$$

**Caractéristiques :**

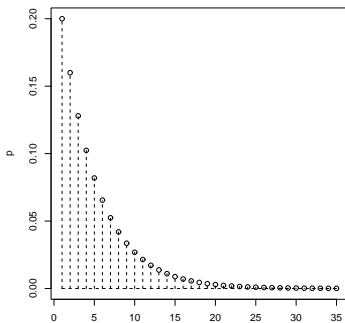
$$E X = 1/p, \text{Var } X = (1 - p)/p^2.$$

## Loi géométrique

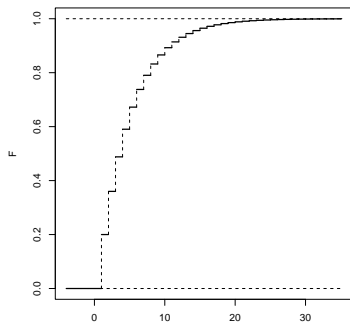
**Application** : si on réalise une suite (infinie) d'expériences indépendantes qui ont chacune la même probabilité  $p$  de réussir, c'est la loi du nombre d'expériences réalisées jusqu'à la première réussite.

**Expression** : pour  $p \in ]0, 1[$ ,  $X \sim \mathcal{G}(p)$  signifie  $X(\Omega) = \mathbb{N}^*$  et

$$\mathbb{P}[X = k] = (1 - p)^{k-1} p \quad \forall k \in \mathbb{N}^* .$$



$\mathcal{G}(0.2)$



## Loi de Poisson

**Application** : modélisation du nombre d'occurrences d'un événement dans une période de temps donnée (nombre de sinistres déclarés en un mois), ou une population dans une zone donnée (nombre de possesseurs d'un téléphone portable de la marque JojoMobiles dans la zone de couverture de l'émetteur de Saint-Vaast-la-Hougue).

**Expression** : pour  $\lambda > 0$ ,  $X \sim \mathcal{P}(\lambda)$  signifie  $X(\Omega) = \mathbb{N}$  et

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \quad \forall k \in \mathbb{N}.$$

**Caractéristiques** :

$E X = \lambda$ ,  $\text{Var } X = \lambda$ .

**Propriétés** :

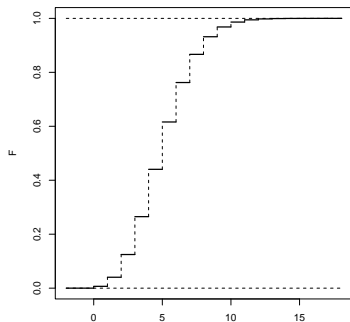
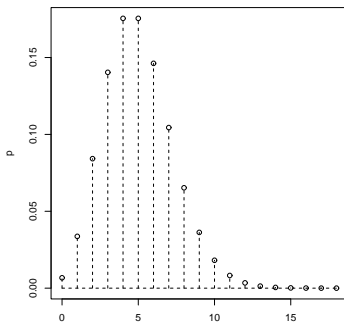
Stabilité par convolution : la somme de deux variables aléatoires indépendantes de loi de Poisson de paramètres respectifs  $\lambda_1$  et  $\lambda_2$  suit une loi de Poisson de paramètre  $\lambda_1 + \lambda_2$ .

## Loi de Poisson

**Application** : modélisation du nombre d'occurrences d'un événement dans une période de temps donnée, ou une population dans une zone donnée.

**Expression** : pour  $\lambda > 0$ ,  $X \sim \mathcal{P}(\lambda)$  signifie  $X(\Omega) = \mathbb{N}$  et

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda} \quad \forall k \in \mathbb{N}.$$



$\mathcal{P}(5)$

## Loi uniforme

**Application** : elle est portée par  $B$  mais ne privilégie aucune valeur dans l'ensemble  $B$ .

**Expression** : pour tout borélien  $B$  de  $\mathbb{R}^d$  de mesure de Lebesgue non nulle, la loi de  $X \sim \mathcal{U}(B)$  est définie par  $\mathbb{P}_X(A) = \lambda(A \cap B)/\lambda(B)$ , et a pour densité

$$f_X(x) = \frac{1}{\lambda(B)} \mathbf{1}_B(x).$$

**Caractéristiques** : Si  $B = [a, b]$ ,  
 $\mathbb{E}X = (a + b)/2$ ,  $\text{Var} X = (b - a)^2/12$ .

**Propriétés** :

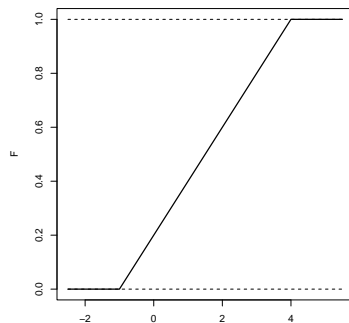
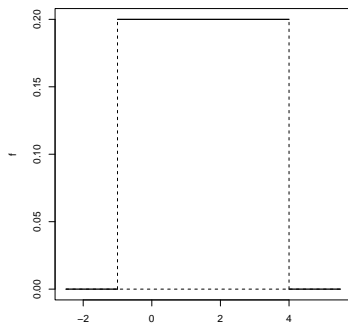
- ▶ La loi uniforme sur  $[a, b[$ , par exemple, est *la même loi* que la loi uniforme sur  $[a, b]$ .
- ▶ Si  $X \sim \mathcal{U}(B)$ , et  $\alpha \neq 0$ , alors  $\alpha X + \beta \sim \mathcal{U}(\alpha B + \beta)$ .

## Loi uniforme

**Application** : elle est portée par  $B$  mais ne privilégie aucune valeur dans l'ensemble  $B$ .

**Expression** : pour tout borélien  $B$  de  $\mathbb{R}^d$  de mesure de Lebesgue non nulle, la loi de  $X \sim \mathcal{U}(B)$  est définie par  $\mathbb{P}_X(A) = \lambda(A \cap B)/\lambda(B)$ , et a pour densité

$$f_X(x) = \frac{1}{\lambda(B)} \mathbf{1}_B(x).$$



$\mathcal{U}([-1, 4])$



## Loi exponentielle

**Application** : elle modélise la durée de vie d'un système sans usure. C'est aussi la loi de l'intervalle entre deux événements poissonniens

**Expression** : pour  $\lambda > 0$ , la densité de  $X \sim \mathcal{E}(\lambda)$  est

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

**Caractéristiques** :

$$E X = 1/\lambda, \text{Var } X = 1/\lambda^2.$$

**Propriétés** :

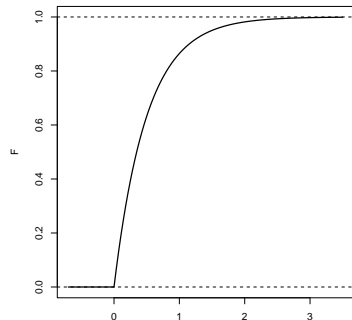
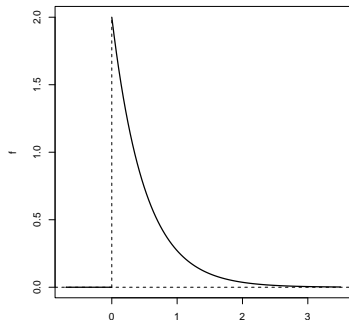
- ▶ Si  $X \sim \mathcal{E}(\lambda)$  et  $\alpha > 0$ , alors  $\alpha X \sim \mathcal{E}(\lambda/\alpha)$ .
- ▶ Si  $X_1$  et  $X_2$  sont deux variables aléatoires indépendantes de lois exponentielles de paramètres respectifs  $\lambda_1$  et  $\lambda_2$ , alors  $\min(X_1, X_2)$  suit la loi exponentielle de paramètre  $\lambda_1 + \lambda_2$ .

## Loi exponentielle

**Application** : elle modélise la durée de vie d'un système sans usure. C'est aussi la loi de l'intervalle entre deux événements poissonniens

**Expression** : pour  $\lambda > 0$ , la densité de  $X \sim \mathcal{E}(\lambda)$  est

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



$\mathcal{E}(2)$

## Loi gamma

**Expression :** pour  $\alpha > 0$  et  $\lambda > 0$ , la densité de  $X \sim \Gamma(\alpha, \lambda)$  est

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+^*}(x),$$

où  $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$ , de sorte que, pour  $n$  entier,  $\Gamma(n) = (n-1)!$ .

**Caractéristiques :**

$$E X = \alpha/\lambda, \text{ Var } X = \alpha/\lambda^2.$$

**Propriété :**

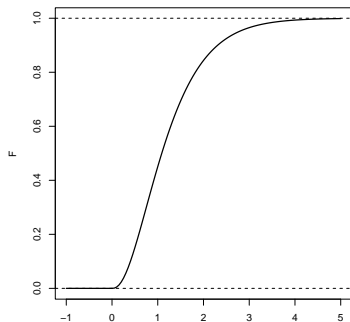
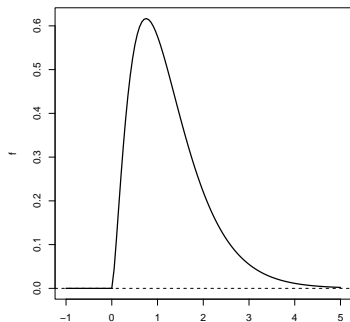
Si  $X_1, \dots, X_n$  sont  $n$  variables aléatoires indépendantes de loi  $\mathcal{E}(\lambda)$ , alors  $X = X_1 + \dots + X_n$  suit la loi  $\Gamma(n, \lambda)$ .

## Loi gamma

**Expression** : pour  $\alpha > 0$  et  $\lambda > 0$ , la densité de  $X \sim \Gamma(\alpha, \lambda)$  est

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+^*}(x),$$

où  $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$ , de sorte que, pour  $n$  entier,  $\Gamma(n) = (n-1)!$ .



$\Gamma(2.5, 2)$

## Loi normale

**Application** : elle peut modéliser une erreur de mesure, de positionnement, un bruit, etc.

**Expression** : pour  $m \in \mathbb{R}$  et  $\sigma \in \mathbb{R}_+^*$ , la densité de  $X \sim \mathcal{N}(m, \sigma^2)$  est

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

On définit aussi la loi  $\mathcal{N}(m, 0) = \delta_m$ .

**Caractéristiques** :

$$\mathbb{E} X = m, \text{Var } X = \sigma^2.$$

**Propriétés** :

- ▶ Si  $X$  suit la loi  $\mathcal{N}(m, \sigma^2)$ , et  $a \in \mathbb{R}^*$ ,  $b \in \mathbb{R}$ , alors  $aX + b$  suit la loi normale  $\mathcal{N}(am + b, a^2\sigma^2)$ .
- ▶ Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes de lois normales  $\mathcal{N}(m, \sigma^2)$  et  $\mathcal{N}(m', \sigma'^2)$ , alors leur somme  $X + Y$  suit la loi normale  $\mathcal{N}(m + m', \sigma^2 + \sigma'^2)$ .

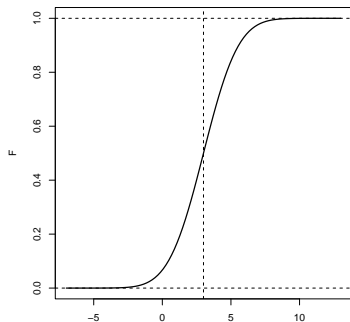
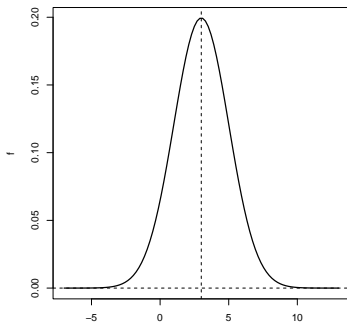
## Loi normale

**Application** : elle peut modéliser une erreur de mesure, de positionnement, un bruit, etc.

**Expression** : pour  $m \in \mathbb{R}$  et  $\sigma \in \mathbb{R}_+^*$ , la densité de  $X \sim \mathcal{N}(m, \sigma^2)$  est

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

On définit aussi la loi  $\mathcal{N}(m, 0) = \delta_m$ .



$\mathcal{N}(3, 4)$

## Loi de Pareto

**Application** : modélisation des excès au-delà de valeurs extrêmes.

**Expression** : pour  $x_m \in \mathbb{R}$  et  $\alpha > 0$ , la densité de  $X \sim \text{Pareto}(x_m, \alpha)$  est

$$f_X(x) = \mathbf{1}_{[x_m, +\infty[}(x) \frac{\alpha x_m^\alpha}{x^{\alpha+1}},$$

et sa fonction de survie vérifie  $S_X(x) = (x/x_m)^{-\alpha}$  pour  $x \geq x_m$ .

**Caractéristiques** :

$$E X = \alpha x_m / (\alpha - 1), \quad \text{Var } X = \alpha x_m^2 / [(\alpha - 1)^2 (\alpha - 2)].$$

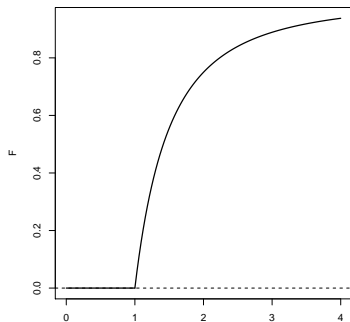
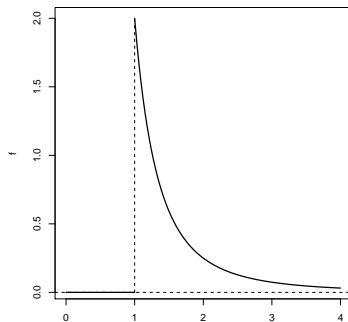
# Loi de Pareto

**Application** : modélisation des excès au-delà de valeurs extrêmes.

**Expression** : pour  $x_m \in \mathbb{R}$  et  $\alpha > 0$ , la densité de  $X \sim \text{Pareto}(x_m, \alpha)$  est

$$f_X(x) = \mathbf{1}_{[x_m, +\infty[}(x) \frac{\alpha x_m^\alpha}{x^{\alpha+1}},$$

et sa fonction de survie vérifie  $S_X(x) = (x/x_m)^{-\alpha}$  pour  $x \geq x_m$ .



Pareto(1, 2)



# Caractérisations fonctionnelles des lois de probabilité

La loi de probabilité de la variable aléatoire  $X$  est caractérisée par les valeurs

$$\mathbb{P}[X \in A] = \mathbb{E}(\mathbf{1}_A(X)) = \int_A f(x) dx$$

pour les boréliens  $A$ . On a vu qu'on peut se restreindre aux boréliens de la forme  $A = ]a, b]$ , ou à ceux de la forme  $A = ]-\infty, b]$ .

Dans cette partie, nous remplacerons les fonctions indicatrices par certaines classes spécifiques de fonctions.

# Grande classe de fonctions

## Propriété

Soient  $X$  et  $Y$  deux variables aléatoires. Si  $E f(X) = E f(Y)$  pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  continue bornée, alors  $X$  et  $Y$  ont même loi.

## Remarques :

- ▶ On peut aussi utiliser l'ensemble des fonctions  $\mathcal{C}^\infty$  support borné.
- ▶ Cette propriété est aussi vérifiée pour les vecteurs aléatoires.

**Corollaire :** Si il existe une fonction intégrable positive  $f$  telle que  $E g(X) = \int g(x)f(x) dx$  pour toute fonction continue bornée  $g$ , alors la variable aléatoire  $X$  est continue de densité  $f$ .

**Exemple :** Si  $X$  a pour densité  $f$ , quelle est la densité de  $|X|$  ?

# Fonction caractéristique

## Définition

Soit  $f$  une fonction mesurable à valeurs complexes. On dit que  $f$  est intégrable par rapport à la mesure  $\mu$  si  $|f|$  est  $\mu$ -intégrable (c'est-à-dire si  $\Re(f)$  et  $\Im(f)$  sont  $\mu$ -intégrables). Dans ce cas, on définit l'intégrale de  $f$  par

$$\int f d\mu = \int \Re(f) d\mu + i \int \Im(f) d\mu.$$

## Définition (dimension 1)

Soit  $\mu$  une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . On appelle **transformée de Fourier** de  $\mu$ , ou **fonction caractéristique** de  $\mu$ , la fonction  $\phi_\mu : \mathbb{R} \rightarrow \mathbb{C}$  définie par

$$\phi_\mu(t) = \int_{\mathbb{R}} e^{itx} d\mu(x).$$

Si  $X$  est une variable aléatoire, on appelle **transformée de Fourier** de  $X$ , ou **fonction caractéristique** de  $X$ , la fonction caractéristique de la loi de  $X$ .

## Définition (dimension $d$ )

Soit  $\mu$  une mesure de probabilité sur  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . La fonction caractéristique de  $\mu$  est la fonction  $\phi_\mu : \mathbb{R}^d \rightarrow \mathbb{C}$  définie par

$$\phi_\mu(t) = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} d\mu(x).$$

## Propriétés

1. Si  $\mu$  et  $\nu$  ont même fonction caractéristique, alors elles sont égales. Si  $X$  et  $Y$  ont même fonction caractéristique, alors elles ont même loi.
2.  $\phi_X(0) = 1$ .
3.  $|\phi_X(t)| \leq 1$  pour tout  $t$ .
4.  $\phi_X(-t) = \overline{\phi_X(t)}$ .
5.  $\phi_{aX+b}(t) = e^{itb} \phi_X(at)$ ,  $\phi_{AX+B}(t) = e^{i\langle t, B \rangle} \phi_X({}^t A t)$ .
6.  $\phi_X(t) = \phi_{\langle t, X \rangle}(1)$ .
7.  $\phi_X$  est uniformément continue.

## Exemples :

- ▶ Si  $X$  suit la loi exponentielle  $\mathcal{E}(1)$ , sa fonction caractéristique est  $f(t) = 1/(1 - it)$ .
- ▶ Si  $X$  suit la loi de Poisson  $\mathcal{P}(\lambda)$ , sa fonction caractéristique est  $f(t) = \exp(\lambda(e^{it} - 1))$ .

## Formule d'inversion de Fourier

Soit  $\phi_X$  la fonction caractéristique d'un vecteur aléatoire  $X$ . Si  $\phi_X$  est intégrable par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$ , alors  $X$  est continue, de densité

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t, x \rangle} \phi_X(t) dt.$$

## Transformée de Fourier et indépendance

Si  $\mu_1$  et  $\mu_2$  sont deux mesures de probabilité, alors

$$\phi_{\mu_1 \otimes \mu_2}(t_1, t_2) = \phi_{\mu_1}(t_1) \phi_{\mu_2}(t_2) \quad \forall (t_1, t_2).$$

De la même manière, si  $X = (X_1, X_2)$  est un vecteur aléatoire, alors  $X_1$  et  $X_2$  sont indépendantes si et seulement si

$$\phi_X(t_1, t_2) = \phi_{X_1}(t_1) \phi_{X_2}(t_2) \quad \forall (t_1, t_2).$$

**Remarque :**  $\phi_{X_1}(t_1) = \phi_X(t_1, 0)$ .

## Transformée de Fourier et convolution

Si  $\mu$  et  $\mu'$  sont deux mesures de probabilité, alors

$$\phi_{\mu \star \mu'} = \phi_{\mu} \phi_{\mu'}.$$

**Exemple :** fonction caractéristique de la loi binomiale.

## Régularité de la fonction caractéristique

Soit  $X$  une variable aléatoire réelle et  $\phi_X$  sa fonction caractéristique.

1. Si  $E(|X|^n) < \infty$ , alors  $\phi_X$  est  $\mathcal{C}^n$  et, pour tout  $k \in \{1, \dots, n\}$ ,

$$\phi_X^{(k)}(t) = i^k E(X^k e^{itx}) \quad \forall t \in \mathbb{R},$$

de sorte qu'en particulier,

$$\phi_X^{(k)}(0) = i^k E(X^k).$$

2. Si  $n$  est pair et si  $\phi_X$  est  $n$  fois dérivable en zéro, alors  $X$  admet un moment d'ordre  $n$  fini.



# Transformée de Fourier Discrète

## Cas discret

Si  $X$  est une variable aléatoire à valeurs dans  $\{0, \dots, M-1\}$ , la loi de  $X$  est déterminée par les  $M$  valeurs de sa transformée de Fourier aux points  $t_j = 2\pi j/M$ , pour  $j \in \{0, \dots, M-1\}$  :

$$\hat{p}(t_j) = \phi_X(t_j) = \sum_{k=0}^{M-1} \mathbb{P}[X = k] \exp(i \frac{2\pi j}{M} k).$$

En effet,

$$\mathbb{P}[X = k] = \frac{1}{M} \sum_{j=0}^{M-1} \hat{p}(t_j) \exp(-i \frac{2\pi j}{M} k) \quad \forall k \in \{0, \dots, M-1\}.$$

**Remarque** : la complexité algorithmique du calcul de  $\hat{p}$  est au plus de l'ordre de  $M^2$ , mais, dans le cas où  $M$  est une puissance de 2, on peut la ramener à l'ordre  $M \ln M$  (Cooley-Tuckey 1965 / Gauss 1805).

### Exemple de calcul d'un produit de convolution par FFT

On considère la somme  $S = \sum_{i=1}^N X_i$  de  $N$  variables aléatoires discrètes  $X_i$  indépendantes et toutes de même loi. Comment déterminer la loi de  $S$  ?

1. Calculer la transformée de Fourier discrète  $\hat{p}$  de la loi des  $X_i$ , avec une valeur de  $M$  puissance de 2 et telle que  $S \leq M$ .
2. Calculer  $\hat{p}_S = \hat{p}^N$ .
3. Calculer la loi de  $S$ .

# Transformée de Laplace

## Définition (Transformée de Laplace)

Si  $X$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^d$ , on appelle **transformée de Laplace** de  $X$  (ou transformée de Laplace de la loi de  $X$ ) la fonction  $L_X : \mathbb{R}^d \rightarrow \mathbb{R}$  définie (quand l'espérance existe) par

$$L_X(t) = \mathbb{E}(e^{-\langle t, X \rangle}).$$

## Propriétés

- ▶  $L_X(0) = 1$
- ▶  $L_X$  est convexe.
- ▶ Si  $X^n$  est une variable aléatoire intégrable,  $L_X^{(n)}(0) = (-1)^n \mathbb{E}(X^n)$ .

**Exemple** : moments de la loi exponentielle.

## Théorème

Si  $L_X$  et  $L_Y$  coïncident sur un ouvert non vide contenant 0, alors  $X$  et  $Y$  ont même loi.

## Propriétés

- ▶ Si  $X$  et  $Y$  sont indépendantes, alors

$$L_{(X,Y)}(t, u) = L_X(t)L_Y(u).$$

- ▶ Si  $X$  et  $Y$  sont indépendantes, alors

$$L_{X+Y}(t) = L_X(t)L_Y(t).$$

- ▶ Si  $A$  est une matrice et  $B$  un vecteur constants, alors

$$L_{AX+B}(t) = e^{-\langle B, t \rangle} L_X({}^t A t).$$

# Fonction génératrice

## Définition

Si  $X$  est une variable aléatoire à valeurs entières, on appelle **fonction génératrice** de  $X$  la fonction  $G_X : \mathbb{R} \rightarrow \mathbb{R}$  définie (quand l'espérance existe) par

$$G_X(t) = E(t^X).$$

Si  $X \in \mathbb{R}^d$  est un vecteur aléatoire, sa fonction génératrice est  $G_X : \mathbb{R}^d \rightarrow \mathbb{R}$  définie par

$$G_X(t_1, \dots, t_d) = E(t_1^{X_1} \cdots t_d^{X_d}).$$

## Propriétés

- ▶ Si  $X$  et  $Y$  sont indépendantes, alors

$$G_{(X,Y)}(t, u) = G_X(t)G_Y(u).$$

- ▶ Si  $X$  et  $Y$  sont indépendantes, alors

$$G_{X+Y}(t) = G_X(t)G_Y(t).$$

- ▶ Si  $X^n$  est une variable aléatoire intégrable, alors

$$G_X^{(n)}(1) = E[X(X-1)(X-2)\cdots(X-n+1)].$$

# Plan

Espaces probabilisés discrets

Théorie générale des probabilités

Manipulation des vecteurs aléatoires

**Statistiques descriptives**

Représentations graphiques

Indicateurs statistiques

# Statistiques descriptives

Objectif : décrire des données de la manière la plus efficace possible.

Illustrations avec le logiciel R (<http://r-project.org>).



# Représentations graphiques

## Diagramme en bâtons

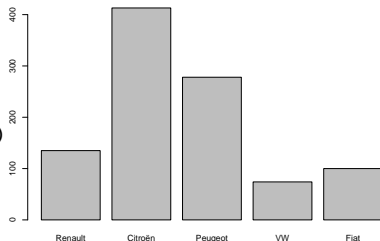
Jojo compte le nombre de voitures de chaque marque qui passent le péage de Jojoville-Sud durant une après-midi.

# Représentations graphiques

## Diagramme en bâtons

Jojo compte le nombre de voitures de chaque marque qui passent le péage de Jojoville-Sud durant une après-midi.

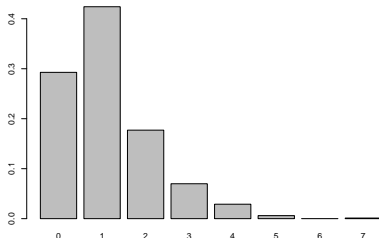
```
nv<-c( 135,413,278,74,100 )  
marques<-c('Renault','Citroën',  
            'Peugeot','VW','Fiat')  
barplot(nv,names=marques)
```



Répartition des 830 assurés à JojoTranquile suivant le nombre de sinistres déclarés les cinq dernières années.

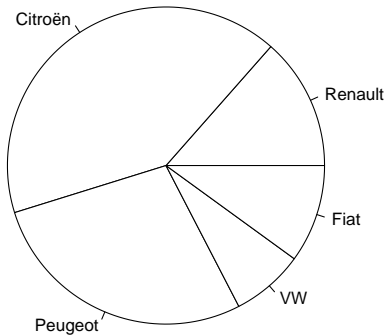
Répartition des 830 assurés à JojoTranquile suivant le nombre de sinistres déclarés les cinq dernières années.

```
eff<-c(243,352,147,58,  
       24,5,0,1)  
barplot(eff/sum(eff),  
        names=0:7)
```



## Camembert

```
pie(nv, labels=marques, col="white")
```



## Histogramme

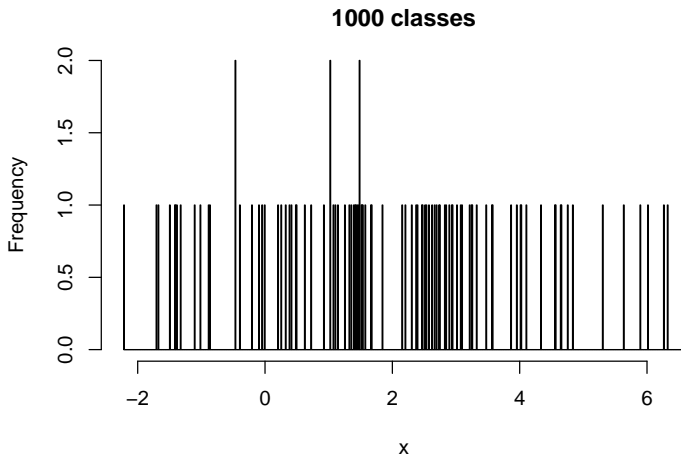
On partitionne un intervalle  $]a_0, a_k]$  contenant toutes les données en  $k$  intervalles  $]a_{i-1}, a_i]$  appelés **classes**. On appelle **effectif** de la classe  $i$  le nombre  $n_i$  d'observations appartenant à cette classe. La **fréquence** de la classe  $i$  est  $n_i/n$ . L'**histogramme** est constitué des rectangles dont les bases sont les classes et dont les **surfaces** sont proportionnelles aux fréquences de ces classes. Le **polygone des fréquences** est la ligne brisée reliant les milieux des sommets des rectangles.

- ▶ Il est recommandé d'avoir au moins 5 classes. La règle empirique de Sturges préconise un nombre de classes égal à  $k \simeq 1 + \log_2 n$ , ce qui donne par exemple  $k = 5$  pour  $n \leq 22$ ,  $k = 6$  pour  $23 \leq n \leq 45$ , etc.
- ▶ En ce qui concerne la largeur des classes, le choix le plus fréquent est de prendre soit des classes de même largeur, soit des classes de même effectif.

## Histogramme à largeur constante

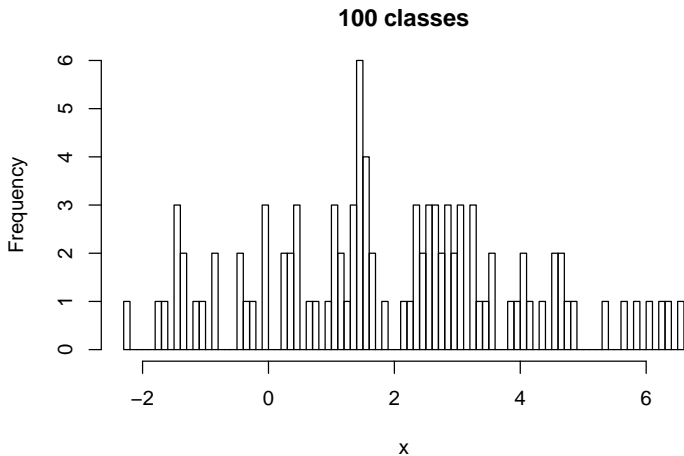
Jojo mesure le temps d'accès à un point donné du réseau internet depuis son ordinateur à de multiples reprises. Il obtient des temps compris entre 183,3 ms et 413,9 ms.

## Choix du nombre de classes

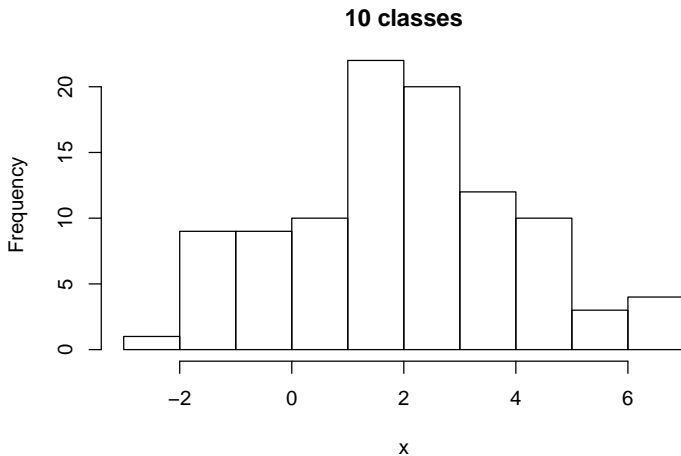




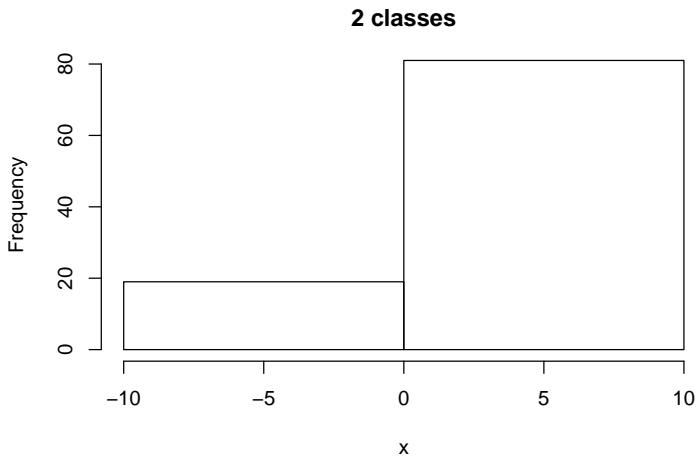
## Choix du nombre de classes



## Choix du nombre de classes



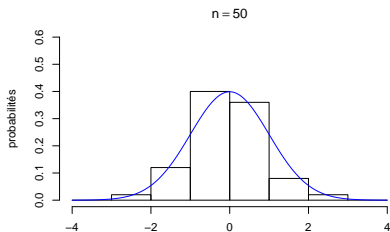
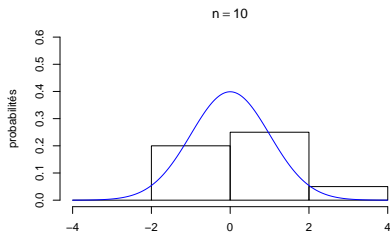
## Choix du nombre de classes



## Histogramme d'un échantillon

Soient  $X_1, X_2, \dots$  une suite de variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ .  
Pour un  $\omega \in \Omega$  fixé, tracé de l'histogramme des

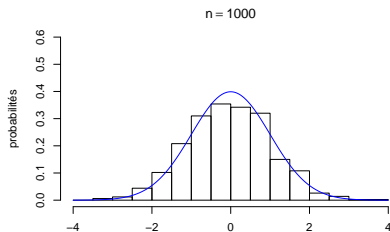
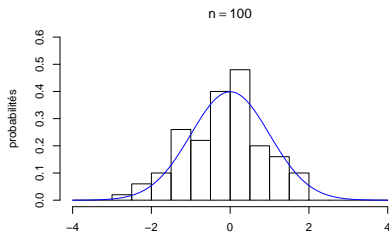
$$X_1(\omega), \dots, X_n(\omega).$$



## Histogramme d'un échantillon

Soient  $X_1, X_2, \dots$  une suite de variables aléatoires indépendantes de loi  $\mathcal{N}(0, 1)$ .  
Pour un  $\omega \in \Omega$  fixé, tracé de l'histogramme des

$$X_1(\omega), \dots, X_n(\omega).$$



## Fonction de répartition empirique

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(x_i) \\ &= \begin{cases} 0 & \text{si } x < x_1^*, \\ i/n & \text{si } x_i^* \leq x < x_{i+1}^*, \\ 1 & \text{si } x_i \geq x_n^*, \end{cases} \end{aligned}$$

## Théorème de Glivenko-Cantelli

Soit  $\{X_i\}_i$  une suite de variables aléatoires indépendantes et de même loi de fonction de répartition  $F$ . Pour tout  $n$ , on note  $F_n$  la fonction de répartition empirique (aléatoire) des observations  $X_1, \dots, X_n$ . Alors la suite  $(F_n)$  converge presque-sûrement uniformément vers  $F$  :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0.$$

## Grphe de probabilités – loi exponentielle

Pour la loi exponentielle  $\mathcal{E}(\lambda)$ ,

$$F(x) = 1 - e^{-\lambda x} \quad \forall x > 0,$$

c'est-à-dire

$$\ln(1 - F(x)) = -\lambda x.$$

Les points de coordonnées

$$(x, \ln(1 - F(x)))$$

sont donc alignés sur la droite d'équation

$$y = -\lambda x.$$

Pour la fonction de répartition empirique, en choisissant les abscisses  $x_i^*$ , de sorte que  $F_n(x_i^*) = i/n$ , on considère les points de coordonnées

$$(x_i^*, \ln(1 - i/n)).$$

Ici,

$$\ln(1 - F(x)) = -\lambda x + \alpha$$

$$F(x) = 1 - \exp(-\lambda(x - \alpha/\lambda)) \quad x \geq \alpha/\lambda$$

En posant  $x_0 = \alpha/\lambda$ , on en déduit

$$X - x_0 \sim \mathcal{E}(\lambda).$$



## Graphe de probabilités – loi normale

Pour la loi normale, si  $X \sim \mathcal{N}(m, \sigma^2)$ , alors

$$\frac{X - m}{\sigma} \sim \mathcal{N}(0, 1),$$

de sorte que

$$\begin{aligned} F(x) &= \mathbb{P}[X \leq x] = \mathbb{P}[(X - m)/\sigma \leq (x - m)/\sigma] \\ &= \Phi((x - m)/\sigma) \end{aligned}$$

Les points de coordonnées

$$(x, \Phi^{-1}(F(x)))$$

sont donc alignés sur la droite d'équation

$$y = (x - m)/\sigma.$$

Pour la fonction de répartition empirique, en choisissant les abscisses  $x_i^*$ , de sorte que  $F_n(x_i^*) = i/n$ , on considère les points de coordonnées

$$(x_i^*, \Phi^{-1}(i/n)).$$

# Indicateurs statistiques

## Localisation

- ▶ Moyenne empirique

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \text{mean}(\mathbf{x})$$

- ▶ Minimum, maximum, moyenne des deux

$$x_1^* = \min(\mathbf{x}), \quad x_n^* = \max(\mathbf{x})$$

$$\frac{x_1^* + x_n^*}{2} = \text{mean}(\text{range}(\mathbf{x}))$$

- ▶ Médiane empirique

$$\begin{aligned} \tilde{x}_n &= \begin{cases} x_{(n+1)/2}^* & \text{si } n \text{ est impair} \\ (x_{n/2}^* + x_{n/2+1}^*)/2 & \text{si } n \text{ est pair} \end{cases} \\ &= \text{median}(\mathbf{x}) \end{aligned}$$

## Dispersion

- ▶ Variance et écart-type empiriques

$$s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \text{var}(\mathbf{x}) * (n-1) / n$$

- ▶ Coefficient de variation empirique

$$cv_n = s_n / \bar{x}_n$$

- ▶ Étendue

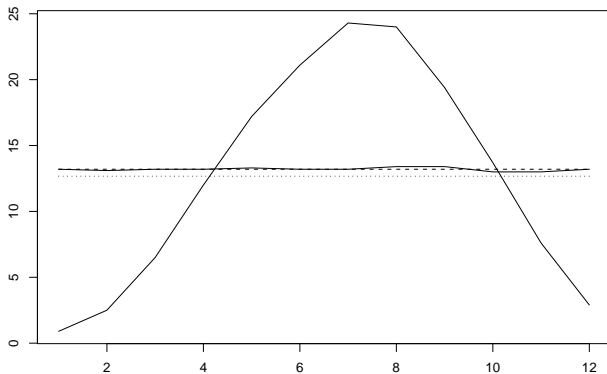
$$x_n^* - x_1^*$$

- ▶ Quantiles empiriques

$$\tilde{q}_{n,p} = \begin{cases} \frac{1}{2}(x_{np}^* + x_{np+1}^*) & \text{si } np \text{ est entier,} \\ x_{[np]+1}^* & \text{sinon.} \end{cases}$$

$$\simeq \text{quantile}(\mathbf{x}, p)$$

## Température à Quito et Tbilissi



	$\bar{x}_n$	$s_n^2$	$s_n$	$CV_n$
t° Tbilissi	12,7	66,9	8,18	64,5 %
t° Quito	13,2	0,015	0,122	0,93 %

## Répartition des températures moyennes mensuelles

