Cours de Probabilités

IMT Mines Alès - Département de Mathématiques

A. CAVAILLÉ

2020-2021

Table des matières

Première partie

Axiomatique

Chapitre 1

Espaces Probabilisés

1.1 Introduction

1.1.1 Une expérience et deux épreuves

Considérons l'expérience suivante : On lance une fois un dé cubique dont les faces sont numérotées de 1 à 6. On a 6 éventualités, ou résultats possibles. **On note Ω l'univers des possibles, ou l'ensemble des éventualités.**

- **Épreuve $n°1$** : on s'intéresse au chiffre apparu.
- **Épreuve $n°2$** : on s'intéresse à la parité du chiffre apparu.

On définit ainsi deux épreuves liées à la même expérience.
Chaque épreuve engendre une partition Ω' de Ω.
Ω' est l'ensemble des évènements élémentaires.

Dans cette expérience : $\Omega = \{1,2,3,4,5,6\}$.

- **Pour la première épreuve** :
  $\Omega = \{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\}$ avec la partition $\Omega'_1 = \big\{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}\big\}$
- **Pour la deuxième épreuve** :
  $\Omega = \{1,3,5\} \cup \{2,4,6\}$ avec la partition $\Omega'_2 = \big\{\{1,3,5\},\{2,4,6\}\big\} = \{I,P\}$

Un évènement composé d'une épreuve est une réunion d'évènements élémentaires de cette épreuve.
- Dans la première épreuve (canonique), il y a 6 évènements élémentaires qui forment la partition Ω'_1 de Ω constituée de tous les singletons éventualités de Ω : $\{i\}$ pour $i \in \{1,\ldots,6\}$.
- Dans la deuxième épreuve, il y a 2 évènements élémentaires qui forment une autre partition Ω'_2 de Ω, avec : $I = \{1,3,5\}$ (les impairs) et $P = \{2,4,6\}$ (les pairs).

Remarque : Soit $A = \{1,2,3\} = \{1\} \cup \{2\} \cup \{3\}$, alors A est un évènement (composé) dans la première épreuve, mais il n'est pas un évènement dans la deuxième épreuve.

Soit $\mathcal{B}$ l'ensemble des évènements liés à une épreuve :
- Dans la $1^{\text{ère}}$ épreuve, $\mathcal{B}_1 = \mathcal{P}(\Omega)$ avec $\operatorname{card}(\mathcal{B}_1) = 2^{\operatorname{card}(\Omega'_1)} = 2^6 = 64$. Il y a 64 évènements dans la première épreuve.
- Dans la $2^{\text{ème}}$ épreuve, $\mathcal{B}_2 = \{\emptyset, I, P, \Omega\}$ avec $\operatorname{card}(\mathcal{B}_2) = 2^{\operatorname{card}(\Omega'_2)} = 2^2 = 4$. Il y a 4 évènements dans la seconde épreuve.

Définition 1

On considère un ensemble, noté Ω, qui est l'ensemble des résultats possibles ω d'une épreuve aléatoire.
Des sous-ensembles de Ω sont des **évènements** liés à cette épreuve.
On désignera par $\mathcal{B} \subset \mathcal{P}(\Omega)$ l'ensemble de ces évènements.

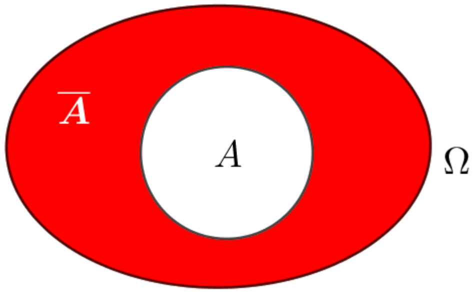
Ω et $\emptyset$ sont des évènements, Ω est appelé évènement sûr et $\emptyset$ est l'évènement impossible.

On veut pouvoir faire des opérations, ou des comparaisons entre évènements. On voudra donc que $\mathcal{B}$ soit **stable par certaines opérations**, qui vont suivre.

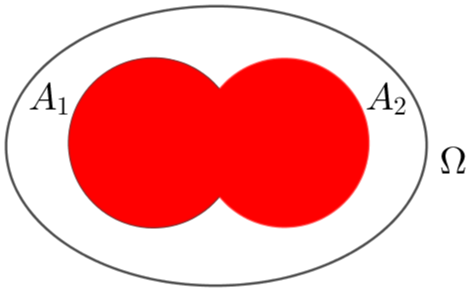
1.1.2 Opérations sur les évènements

$\mathcal{B} \subset \mathcal{P}(\Omega)$ est l'ensemble des évènements liés à l'épreuve considérée. Pour tout $(A, A_1, A_2) \in \mathcal{B}^3$:

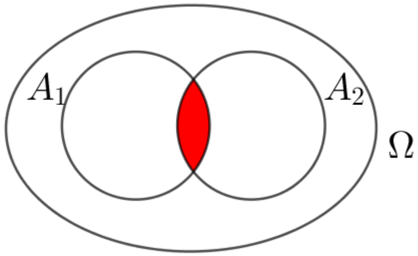
- $\overline{A}$ est l'évènement « contraire » de A (également noté A^C, complémentaire de A).



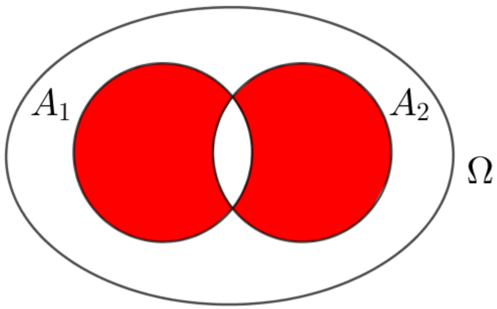
- $A_1 \cup A_2$ est l'évènement « A_1 ou A_2 ».



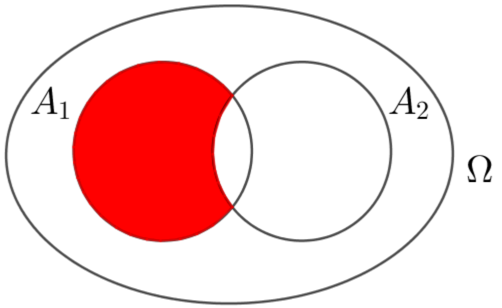
- $A_1 \cap A_2$ est l'évènement « A_1 et A_2 » (également noté $A_1 \cdot A_2$).



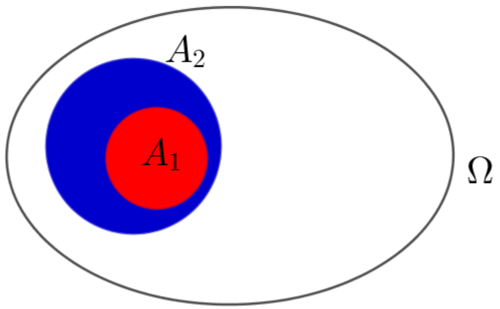
- $A_1 \Delta A_2$ est l'évènement « un et un seul des évènements A_1 ou A_2 ». Δ est le symbole de la différence symétrique.



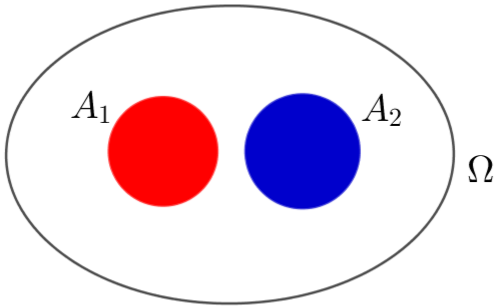
- $A_1 \setminus A_2 = A_1 \cap \overline{A_2}$ est l'évènement « l'évènement A_1 est réalisé mais A_2 ne l'est pas ».



- Si $A_1 \subset A_2$, on dit que « la réalisation de A_1 entraine la réalisation de A_2 ».



- Si $A_1 \cap A_2 = \emptyset$, A_1 et A_2 sont dits « incompatibles ». Ils ne peuvent pas se réaliser simultanément.



- $B = A_1 + A_2 \Leftrightarrow \begin{cases} A_1 \cap A_2 = \emptyset \\ B = A_1 \cup A_2 \end{cases}$
  Le symbole $+$ est la notation pour « union disjointe ».

1.2 Espaces probabilisables

1.2.1 Espace probabilisable fini

Définition 2

Soit Ω l'univers fini des possibles. Soit $\mathcal{B} \subset \mathcal{P}(\Omega)$ l'ensemble des évènements liés à une épreuve, avec $\mathcal{B} \neq \emptyset$.
On dit que $\mathcal{B}$ est une **algèbre de Boole** d'évènements (ou une **tribu** sur Ω) si :

- $\boldsymbol{\Omega \in \mathcal{B}}$
- $\mathcal{B}$ est stable par passage au **contraire** (ou complémentaire) : $\forall \boldsymbol{A \in \mathcal{B}, \overline{A} \in \mathcal{B}}$
- $\mathcal{B}$ est stable par **union finie** : $\forall \boldsymbol{(A_1, A_2) \in \mathcal{B}^2, A_1 \cup A_2 \in \mathcal{B}}$

$(\Omega, \mathcal{B})$ est appelé **espace probabilisable** fini.

Si $\mathcal{B} = \mathcal{P}(\Omega)$, on dit que la tribu est **totale** ou **canonique** sur Ω.

Retour sur l'expérience du jet d'un dé : On avait $\Omega = \{1, 2, 3, 4, 5, 6\}$ et $\mathcal{B}_2 = \{\emptyset, I, P, \Omega\}$.
Soit $\Omega'_2 = \{I, P\}$ l'ensemble des évènements élémentaires de cette épreuve, c'est une partition de Ω. On a $\mathcal{P}(\Omega'_2) = \big\{\emptyset, \{I\}, \{P\}, \Omega'_2\big\}$.

On considère l'application ϕ : $\begin{cases} \mathcal{P}(\Omega'_2) & \longrightarrow \mathcal{B}_2 \\ \emptyset & \longmapsto \quad \emptyset \\ \{I\} & \longmapsto \quad I \\ \{P\} & \longmapsto \quad P \\ \Omega'_2 & \longmapsto \quad \Omega \end{cases}$

Cette application est bijective par construction, donc : $\boxed{\mathrm{card}(\mathcal{B}_2) = \mathrm{card}(\mathcal{P}(\Omega'_2)) = 2^{\,\mathrm{card}(\Omega'_2)}}$

Ainsi, le nombre d'évènements de $\mathcal{B}_2$ est égal à $2^{\text{nombre d'évènements élémentaires}}$.
Cas général : L'exemple précédent se généralise avec le théorème suivant.

Théorème 3

Soit Ω un ensemble fini et $\mathcal{B}$ une algèbre de Boole de parties de Ω.
Il existe une partition de Ω en parties E_i $(1 \leqslant i \leqslant n)$ telle que $\mathcal{B}$ est constituée de l'ensemble vide et de toutes les unions finies d'ensembles $(E_i)_{1\leqslant i\leqslant n}$.

- Les $(E_i)_{1\leqslant i\leqslant n}$ sont les **évènements élémentaires** de l'algèbre de Boole $\mathcal{B}$.
- $\mathcal{B}$ est équipotent à $\mathcal{P}(\Omega')$ où Ω' est l'ensemble des $(E_i)_{1\leqslant i\leqslant n}$.

Toute algèbre de Boole $\mathcal{B}$ définie sur un univers fini Ω a donc un cardinal égal à une puissance de 2.

En effet : $\quad \mathrm{card}(\mathcal{B}) = \mathrm{card}(\mathcal{P}(\Omega')) = 2^{\,\mathrm{card}(\Omega')}$.

1.2.2 Espace probabilisable infini

Définition 4

Soit Ω l'univers infini des possibles. Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$ l'ensemble des évènements liés à une épreuve, avec $\mathcal{A} \neq \emptyset$.
On dit que $\mathcal{A}$ est une **σ-algèbre de Boole** d'évènements (ou une **tribu** sur Ω) si :

- $\boldsymbol{\Omega \in \mathcal{A}}$
- $\mathcal{A}$ est stable par passage au **contraire** : $\forall \boldsymbol{A \in \mathcal{A}, \overline{A} \in \mathcal{A}}$.
- $\mathcal{A}$ est stable par **union dénombrable** : $\Big(\forall \boldsymbol{i} \in \mathbb{N}^*,\ \boldsymbol{A_i \in \mathcal{A}}\Big) \Longrightarrow \Big(\bigcup\limits_{\boldsymbol{i}\in\mathbb{N}^*} \boldsymbol{A_i \in \mathcal{A}}\Big)$.

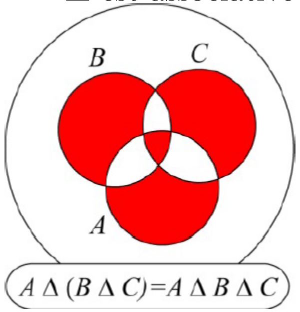
$(\Omega, \mathcal{A})$ est appelé **espace probabilisable** infini.

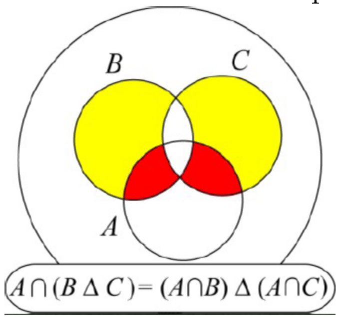
Proposition 5

$\big(\mathcal{P}(\Omega), \Delta, \cap\big)$ a une structure d'anneau commutatif.

En particulier :

- $\emptyset$ est le neutre pour Δ
- Δ est associative
- Ω est le neutre pour $\cap$.
- $\cap$ est distributive par rapport à Δ.





1.2.3 Premières propriétés

Propriété 6

Pour toute algèbre de Boole $\mathcal{B}$ sur un univers Ω et pour tout $n \in \mathbb{N}^*$, on a :

- $\emptyset \in \mathcal{B}$
- **Stabilité par intersection finie** : $\forall (A_1, \dots, A_n) \in \mathcal{B}^n, \quad \bigcap_{i=1}^{n} A_i \in \mathcal{B}$
- **Stabilité par union finie** : $\forall (A_1, \dots, A_n) \in \mathcal{B}^n, \quad \bigcup_{i=1}^{n} A_i \in \mathcal{B}$
- **Stabilité par différence** : $\forall (A_1, A_2) \in \mathcal{B}^2, \quad A_1 \setminus A_2 \in \mathcal{B}$
- **Stabilité par différence symétrique** Δ : $\forall (A_1, A_2) \in \mathcal{B}^2, \quad A_1 \Delta A_2 \in \mathcal{B}$.

Toute σ-algèbre de Boole $\mathcal{A}$ est en particulier une algèbre de Boole et vérifie donc les propriétés ci-dessus.
De plus, on a la **stabilité par l'union dénombrable** (c'est dans la définition) et la **stabilité pour l'intersection dénombrable** : $\left(\forall i \in \mathbb{N}^*, A_i \in \mathcal{A}\right) \implies \left(\bigcap_{i \in \mathbb{N}^*} A_i \in \mathcal{A}\right)$.

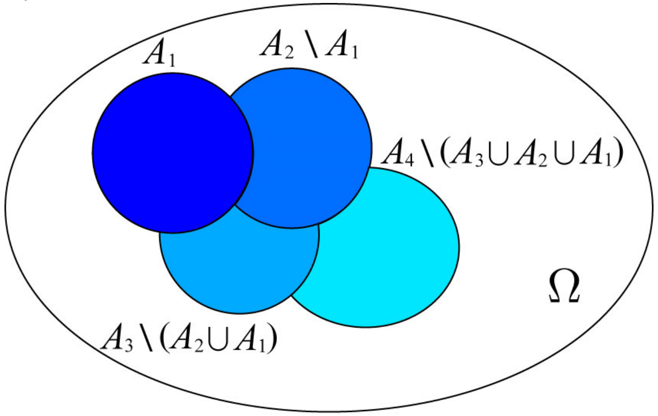
Proposition 7

Pour toute algèbre de Boole $\mathcal{B}$, $\big(\mathcal{B}, \Delta, \cap\big)$ est un sous anneau idempotent ($\forall A \in \mathcal{B}, A \cap A = A$, donc $A^2 = A$) de $\big(\mathcal{P}(\Omega), \Delta, \cap\big)$.

1.2.4 Lemme de décomposition

Lemme 1. *Pour tout $n \geqslant 2$ et pour toute famille $(A_i)_{i\in\{1,\dots,n\}}$ de parties de Ω :*

$$\begin{aligned}\bigcup_{i=1}^{n} A_i &= \sum_{i=1}^{n}\left(A_i \setminus \bigcup_{j=1}^{i-1} A_j\right)\\ &= A_1 + (A_2 \setminus A_1) + (A_3 \setminus (A_2 \cup A_1)) + \cdots\\ &\overset{\textit{autre écriture}}{=} \sum_{i=1}^{n}\left[A_i \cap \left(\bigcap_{j=1}^{i-1} \overline{A_j}\right)\right]\end{aligned}$$



Démonstration. On raisonne par récurrence.

- **Initialisation** : si $n = 2$, $A_1 \cup A_2 = A_1 + (A_2 \cap \overline{A_1}) = A_1 + (A_2 \setminus A_1)$.
- **Hérédité** : on suppose qu'il existe $n \geqslant 2$ tel que la formule est vraie au rang n.

$$\begin{aligned}\bigcup_{i=1}^{n+1} A_i &= \left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1} &&= \bigcup_{i=1}^{n} A_i + \left(A_{n+1} \setminus \bigcup_{i=1}^{n} A_i\right)\\ &= \sum_{i=1}^{n}\left[A_i \setminus \bigcup_{j=1}^{i-1} A_j\right] + \left(A_{n+1} \setminus \bigcup_{i=1}^{n} A_i\right) &&= \sum_{i=1}^{n+1}\left[A_i \setminus \bigcup_{j=1}^{i-1} A_j\right]\end{aligned}$$

  Donc la relation est vraie au rang $n+1$.
- **Conclusion** : par le théorème de récurrence, on a donc le résultat voulu. $\square$

1.2.5 Tribu engendrée

Théorème 8

L'intersection de tribus de Ω est encore une tribu de Ω.

Définition 9

On appelle **tribu engendrée** par une partie $\mathcal{A}_0$ de $\mathcal{P}(\Omega)$ la plus petite tribu contenant $\mathcal{A}_0$. C'est l'intersection de toutes les tribus contenant $\mathcal{A}_0$.

Elle contient les éléments de $\mathcal{A}_0$, leurs complémentaires, les réunions dénombrables de ceux-ci et de ceux-là, ...

Exemple fondamental : tribu des boréliens de $\mathbb{R}$. $\mathcal{A} = \mathcal{B}_{\mathbb{R}}$ est la **tribu des boréliens de** $\mathbb{R}$, c'est la σ-algèbre de Boole engendrée par les sous-ensembles de $\mathbb{R}$, demi-droites ouvertes inachévées : $B_a =]-\infty, a[$ où $a \in \mathbb{R}$.

- Par stabilité du complémentaire, $B_a \in \mathcal{B}_{\mathbb{R}}$ donc $\overline{B_a} = [a, +\infty[\in \mathcal{B}_{\mathbb{R}}$.
- Par stabilité de l'union dénombrable, $]a, +\infty[= \displaystyle\bigcup_{n \in \mathbb{N}^*} \left[a + \frac{1}{n}, +\infty\right[$ donc $]a, +\infty[\in \mathcal{B}_{\mathbb{R}}$.
- Alors $\overline{]a, +\infty[} =]-\infty, a] \in \mathcal{B}_{\mathbb{R}}$, donc aussi $]-\infty, a] \cap [a, +\infty[= \{a\} \in \mathcal{B}_{\mathbb{R}}$.

La tribu des boréliens contient donc tous les **singletons** de réels, tous les **intervalles** et leurs **réunions dénombrables**. $\mathbb{Q}$ l'ensemble des rationnels est dénombrable donc $\mathbb{Q} \in \mathcal{B}_{\mathbb{R}}$. L'ensemble des irrationnels est le complémentaire de $\mathbb{Q}$ dans $\mathbb{R}$ donc il appartient à $\mathcal{B}_{\mathbb{R}}$.

Cependant, $\mathcal{B}_{\mathbb{R}} \neq \mathcal{P}(\mathbb{R})$ car il existe des sous-ensembles de $\mathbb{R}$ (compliqués) qui ne sont pas des boréliens.

$(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ est un espace probabilisable infini indénombrable.

1.3 Espaces probabilisés

1.3.1 Espaces probabilisés finis ou infinis

Définition 10 (Axiomes de Kolmogorov (1933), cas fini)

Une **probabilité** sur $(\Omega, \mathcal{B})$ où Ω est **fini** et $\mathcal{B}$ une tribu, est une application $\mathbb{P} : \left\{ \begin{array}{lcl} \mathcal{B} & \longrightarrow & [0, +\infty[\\ A & \longmapsto & \mathbb{P}(A) \end{array} \right.$ telle que :

1. $\mathbb{P}(\Omega) = 1$
2. **Additivité forte**
   $\forall (A_1, A_2) \in \mathcal{B}^2$, tels que $A_1 \cap A_2 = \emptyset$, on a : $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$. Autrement dit, $\mathbb{P}(A_1 + A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$.

On dit que $(\Omega, \mathcal{B}, \mathbb{P})$ est un **espace probabilisé fini**.

Définition 11 (Axiomes de Kolmogorov (1933) première version, cas infini)

Une **probabilité** sur $(\Omega, \mathcal{A})$, où Ω est **infini** et $\mathcal{A}$ une tribu, est une application $\mathbb{P} : \left\{ \begin{array}{lcl} \mathcal{A} & \longrightarrow & [0, +\infty[\\ A & \longmapsto & \mathbb{P}(A) \end{array} \right.$ telle que :

1. $\mathbb{P}(\Omega) = 1$
2. **Additivité forte**
   $\forall (A_1, A_2) \in \mathcal{A}^2$, tels que $A_1 \cap A_2 = \emptyset$, on a : $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$. Autrement dit, $\mathbb{P}(A_1 + A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$.
3. **Continuité monotone**
   La limite de la probabilité de toute suite décroissante (pour l'inclusion) vers l'évènement impossible vaut 0. C'est-à-dire :
   si $\forall n \in \mathbb{N}^*,\ A_{n+1} \subset A_n$ et $\displaystyle\bigcap_{n \in \mathbb{N}^*} A_n = \emptyset$, alors on a : $\displaystyle\lim_{n \to +\infty} \mathbb{P}(A_n) = 0$

On dit que $(\Omega, \mathcal{A}, \mathbb{P})$ est un **espace probabilisé infini**.

Définition 12 (Axiomes de Kolmogorov (1933) seconde version, cas infini)

Une **probabilité** sur $(\Omega, \mathcal{A})$, où Ω est **infini** et $\mathcal{A}$ une tribu, est une application $\mathbb{P} : \begin{cases} \mathcal{A} & \longrightarrow & [0, +\infty[\\ A & \longmapsto & \mathbb{P}(A) \end{cases}$ telle que :

1. $\mathbb{P}(\Omega) = 1$
2. **σ-additivité**
   Pour toute famille $(A_i)_{i \in \mathbb{N}^*}$ d'évènements de $\mathcal{A}$ incompatibles deux à deux (c'est-à-dire $A_i \cap A_j = \emptyset$ si $i \neq j$), on a :
   $\mathbb{P}\left(\bigcup_{i \in \mathbb{N}^*} A_i\right) = \sum_{i \in \mathbb{N}^*} \mathbb{P}(A_i)$

On démontre que les axiomatiques des définitions 11 et 12 sont équivalentes (admis). Dans la littérature, les deux versions co-existent. Dans la pratique, nous utiliserons l'une ou l'autre à notre convenance.

1.3.2 Propriétés des probabilités

Propriété 13

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

1. $\forall A \in \mathcal{A}, \quad \mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$

2. $\mathbb{P}(\emptyset) = 0$

3. $\forall (A, B) \in \mathcal{A}^2$:
- $A \subset B \Longrightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$. D'où $\forall A \in \mathcal{A},\ 0 \leqslant \mathbb{P}(A) \leqslant 1$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $\mathbb{P}(A \,\Delta\, B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\,\mathbb{P}(A \cap B)$

4. $\forall (A_i)_{i\in\mathbb{N}^*} \subset \mathcal{A}, \quad \mathbb{P}\left(\bigcup_{i\in\mathbb{N}^*} A_i\right) \leqslant \sum_{i\in\mathbb{N}^*} \mathbb{P}(A_i)$

5. $\left(\forall n \in \mathbb{N}^*, A_{n+1} \subset A_n \text{ et } \bigcap_{n\in\mathbb{N}^*} A_n = A\right) \Longrightarrow \left(\lim_{n\to+\infty} \mathbb{P}(A_n) = \mathbb{P}(A)\right)$

Démonstration.

1. A et $\overline{A}$ forment une réunion disjointe : $A + \overline{A} = \Omega$ donc $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = 1$.

2. $\emptyset = \overline{\Omega}$ donc $\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 0$.

3.
- Soit $A \subset B$. Alors on a la réunion disjointe $B = A + (B \cap \overline{A})$ donc $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A})$. Comme $\mathbb{P}(B \cap \overline{A}) \geqslant 0$, on a $\mathbb{P}(A) \leqslant \mathbb{P}(B)$. En prenant $B = \Omega$, on obtient l'encadrement voulu.
- D'une part $A \cup B = A + (B \cap \overline{A})$ donc $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A})$. D'autre part $B = B \cap \Omega = B \cap (A + \overline{A}) = B \cap A + B \cap \overline{A}$ (probabilités totales) donc $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap \overline{A})$, d'où le résultat.
- On a $A \,\Delta\, B = (A \cap \overline{B}) + (B \cap \overline{A})$. Or (probabilités totales) $A = A \cap B + A \cap \overline{B}$ et $B = B \cap A + B \cap \overline{A}$, d'où le résultat.

4. -Cas fini uniquement- D'après le lemme de décomposition, on a une union disjointe $\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n}\left(A_i \setminus \bigcup_{j=1}^{i-1} A_j\right)$ donc $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}\left(A_i \setminus \bigcup_{j=1}^{i-1} A_j\right)$. Pour chaque i : $A_i \setminus \bigcup_{j=1}^{i-1} A_j \subset A_i$ donc $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} \mathbb{P}(A_i)$ d'où le résultat.
5. On a $\forall n \in \mathbb{N}^*, A_{n+1} \subset A_n$ et $A_n = A + (A_n \setminus A)$ donc $\mathbb{P}(A_n) = \mathbb{P}(A) + \mathbb{P}(A_n \setminus A)$. Or la suite $(A_n \setminus A)_{n\in\mathbb{N}^*}$ est décroissante vers l'évènement vide $\emptyset$ donc d'après l'axiome 3, $\lim_{n\to+\infty} \mathbb{P}(A_n \setminus A) = 0$ d'où le résultat. □

1.3.3 Espaces probabilisés complets

Définition 14 (Evènement $\mathbb{P}$-négligeable)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Un évènement $A \in \mathcal{A}$ tel que $\mathbb{P}(A) = 0$ est dit **$\mathbb{P}$-négligeable**.

Définition 15 (Espace probabilisé complet)

Si tout sous-ensemble d'un évènement $\mathbb{P}$-négligeable d'une tribu $\mathcal{A}$ appartient à cette tribu, on dit que l'espace probabilisé est **complet**.

Dans un espace probabilisé complet, tout sous-ensemble d'un évènement $\mathbb{P}$-négligeable est aussi $\mathbb{P}$-négligeable.
En effet, si $A \subset N$ avec $\mathbb{P}(N) = 0$, alors $0 \leqslant \mathbb{P}(A) \leqslant \mathbb{P}(N) = 0$ donc $\mathbb{P}(A) = 0$ et A est aussi $\mathbb{P}$-négligeable.

Théorème 16

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé non complet. On considère la classe des sous-ensembles $\mathbb{P}$-négligeables de $\mathcal{A}$, notée $\mathcal{N}$ et on construit $\widetilde{\mathcal{A}} = \{A \cup N, A \in \mathcal{A}, N \in \mathcal{N}\}$. Alors $\widetilde{\mathcal{A}}$ est une σ-algèbre de Boole.
On définit $\widetilde{\mathbb{P}}(A \cup N) = \mathbb{P}(A)$. Alors $\widetilde{\mathbb{P}}$ est une probabilité sur $\widetilde{\mathcal{A}}$.

Ainsi, $\left(\Omega, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}}\right)$ est un espace probabilisé **complet**.

Rappel : Toute tribu $\mathcal{A}$ est incluse dans $\mathcal{P}(\Omega)$, mais toute partie de Ω n'est pas nécessairement un évènement.

Définition 17 (Probabilité extérieure et intérieure)

Pour tout $\Omega_0 \notin \mathcal{A}$, on définit :
- la **probabilité extérieure** $\mathbb{P}^*(\Omega_0) = \inf\{\mathbb{P}(A), A \in \mathcal{A}, \Omega_0 \subset A\}$
- la **probabilité intérieure** $\mathbb{P}_*(\Omega_0) = \sup\{\mathbb{P}(A), A \in \mathcal{A}, A \subset \Omega_0\}$.

Propriété 18

1. $\forall \Omega_0 \in \mathcal{P}(\Omega) : \mathbb{P}_*(\Omega_0) \leqslant \mathbb{P}^*(\Omega_0)$ **2.** Si $\Omega_0 \in \mathcal{A}$, alors $\mathbb{P}_*(\Omega_0) = \mathbb{P}^*(\Omega_0) = \mathbb{P}(\Omega_0)$. **3.** $\forall \Omega_0 \in \mathcal{P}(\Omega) : \mathbb{P}_*(\Omega_0) = 1 - \mathbb{P}^*(\overline{\Omega_0})$

Démonstration. Démontrons la propriété **3.** Soit $A \in \mathcal{A}$: $A \subset \Omega_0 \Leftrightarrow \overline{\Omega_0} \subset \overline{A}$, et $\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$. Ainsi on a :

$$\begin{aligned}\mathbb{P}_*(\Omega_0) &= \sup\{\mathbb{P}(A), A \in \mathcal{A}, A \subset \Omega_0\} = \sup\{1 - \mathbb{P}(\overline{A}), A \in \mathcal{A}, A \subset \Omega_0\} = 1 - \inf\{\mathbb{P}(\overline{A}), A \in \mathcal{A}, A \subset \Omega_0\} \\ &= 1 - \inf\{\mathbb{P}(\overline{A}), \overline{A} \in \mathcal{A}, \overline{\Omega_0} \subset \overline{A}\} = 1 - \mathbb{P}^*(\overline{\Omega_0})\end{aligned}$$

□

Théorème 19

On pose $\widetilde{\mathcal{A}} = \{\Omega_0 \subset \Omega,\ \ \mathbb{P}_*(\Omega_0) = \mathbb{P}^*(\Omega_0)\}$ et on définit pour $\Omega_0 \in \widetilde{\mathcal{A}}$: $\widetilde{\mathbb{P}}(\Omega_0) = \mathbb{P}_*(\Omega_0) = \mathbb{P}^*(\Omega_0)$.
Alors l'espace $\left(\Omega,\ \widetilde{\mathcal{A}},\ \widetilde{\mathbb{P}}\right)$ est un espace probabilisé complet.

1.4 Équiprobabilité des évènements élémentaires

Considérons $(\Omega, \mathcal{B}, \mathbb{P})$ un espace probabilisé **fini**. On note $\Omega' = \{E_i, 1 \leqslant i \leqslant n\}$ une partition de Ω en évènements élémentaires (donc $\text{card}(\mathcal{B}) = \text{card}(\mathcal{P}(\Omega')) = 2^n$).

Définition 20

On dit qu'il y a **équiprobabilité** des évènements élémentaires si : $\forall (i,j) \in \{1,\dots,n\}^2, \;\; \mathbb{P}(E_i) = \mathbb{P}(E_j)$.

Proposition 21

- S'il y a équiprobabilité des évènements élémentaires, alors $\forall\, i \in \{1,\dots,n\}, \;\; \mathbb{P}(E_i) = \dfrac{1}{n}$ où $n = \text{card}(\Omega')$.
- Si A est un évènement composé : $A = \displaystyle\sum_{i=1}^{m} E_i$, alors $\quad \mathbb{P}(A) = \dfrac{m}{n}\cdot$

  Où m est le nombre de cas favorables et n le nombre de cas possibles.

Démonstration.

- $\mathbb{P}(\Omega) = \mathbb{P}\left(\displaystyle\sum_{i=1}^{n} E_i\right) = \displaystyle\sum_{i=1}^{n} \mathbb{P}(E_i) = n\,\mathbb{P}(E_1) = 1$ donc $\mathbb{P}(E_1) = \dfrac{1}{n}\cdot$
- $\mathbb{P}(A) = \mathbb{P}\left(\displaystyle\sum_{i=1}^{m} E_i\right) = \displaystyle\sum_{i=1}^{m} \mathbb{P}(E_i) = \dfrac{m}{n}\cdot$ □

Cas particulier : si $\mathcal{B} = \mathcal{P}(\Omega)$, alors $\text{card}(\Omega) = \text{card}(\Omega') = n$ et $\mathbb{P}(\{\omega_i\}) = \dfrac{1}{n} = \dfrac{1}{\text{card}(\Omega)}\cdot$

$$\boxed{\text{Alors pour un évènement } A \text{ de cardinal } m : \quad \mathbb{P}(A) = \frac{m}{n} = \frac{\text{card}(A)}{\text{card}(\Omega)}\cdot}$$

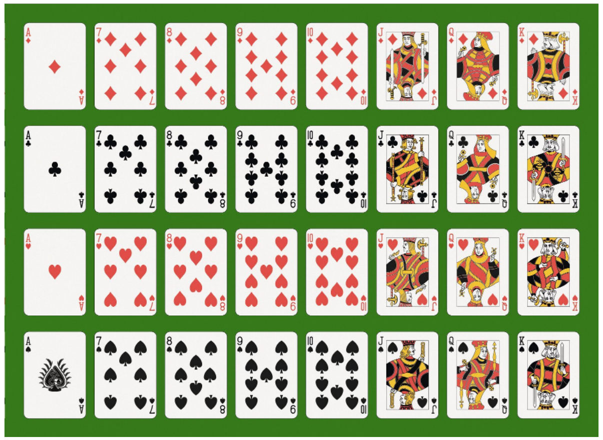
Chapitre 2

Probabilités conditionnelles

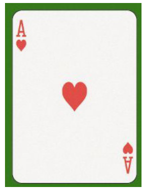
2.1 Axiome de la probabilité conditionnelle

Exemple d'un tirage : on tire une carte d'un jeu de 32.

Jeu de 32 cartes

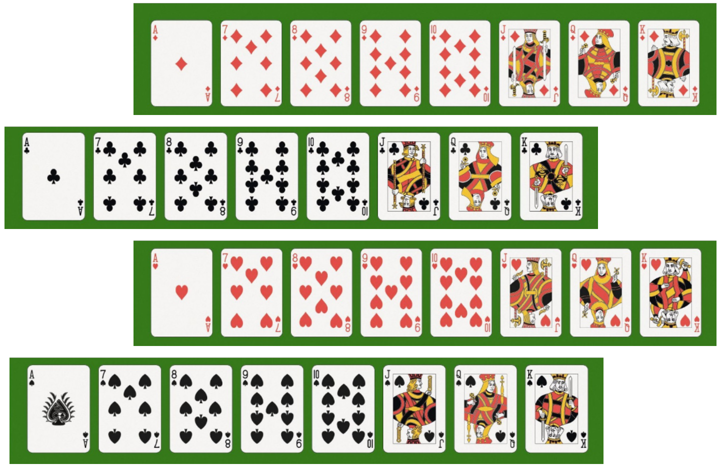


Par exemple, l'as de coeur



Probabilité d'obtenir l'as de coeur :

$$\mathbb{P}(\text{« As de coeur »}) = \frac{1}{32}.$$



Probabilité d'obtenir l'as de coeur, sachant qu'on a tiré une carte rouge :

$\mathbb{P}(\text{« Carte rouge »}) = \dfrac{1}{2}$,

$\mathbb{P}(\text{« As de coeur » sachant « Carte rouge »}) = \dfrac{1}{16}$.

Définition 22

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Pour $A \in \mathcal{A}$ tel que $\mathbb{P}(A) \neq 0$, on appelle probabilité conditionnelle sachant A la probabilité définie sur Ω par :

$$\forall B \in \mathcal{A}, \quad \mathbb{P}_A(B) \text{ notée aussi } \mathbb{P}(B/A), \text{ avec } \mathbb{P}_A(B) = \mathbb{P}(B/A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

(ces deux notations se lisent : probabilité de B sachant A.)

Si on reprend l'exemple initial, et les évènements :

A : « obtenir une carte rouge ». On a $\mathbb{P}(A) = \dfrac{16}{32} = \dfrac{1}{2}$. B : « obtenir l'as de coeur ». On a $\mathbb{P}(B) = \dfrac{1}{32}$.

Alors $\mathbb{P}(A) \neq 0$, et $\mathbb{P}_A(B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(A)} = \dfrac{1}{16}$.

Remarques :

- L'application $\mathbb{P}_A$ définit un nouvel espace probabilisé : $(\Omega, \mathcal{A}, \mathbb{P}_A)$.
- On peut aussi considérer une nouvelle tribu $\mathcal{A}_A$, la tribu-trace sur A constituée des évènements $B \cap A$ où $B \in \mathcal{A}$.

On a ainsi un autre espace probabilisé : $(A, \mathcal{A}_A, \mathbb{P}_A)$.

2.2 Indépendance de deux évènements

Définition 23

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.
Soient A et B deux évènements tels que $\mathbb{P}(A) \times \mathbb{P}(\overline{A}) \times \mathbb{P}(B) \times \mathbb{P}(\overline{B}) \neq 0$.
On dit que A et B sont $\mathbb{P}$-indépendants (ou indépendants par rapport à $\mathbb{P}$) lorsque $\mathbb{P}(A/B) = \mathbb{P}(A/\overline{B})$.

Savoir que B est vrai, ou que $\overline{B}$ est vrai, ne modifie pas la probabilité que A soit vrai.

Propriété 24 (propriété caractéristique)

A et B sont deux évènements $\mathbb{P}$-indépendants si et seulement si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Démonstration.
Montrons que si les hypothèses de la Définition 23 sont vraies, alors Définition 23 $\Leftrightarrow$ Définition 24.
$\forall (A, B) \in \mathcal{A}^2, \quad B = (A \cap B) + (\overline{A} \cap B), \quad \mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B)$

$$\begin{aligned} \text{et} \quad \mathbb{P}(B) &= \mathbb{P}(A)\,\mathbb{P}(B/A) + \mathbb{P}(\overline{A})\,\mathbb{P}(B/\overline{A}) = \big(1 - \mathbb{P}(\overline{A})\big)\,\mathbb{P}(B/A) + \mathbb{P}(\overline{A})\,\mathbb{P}(B/\overline{A}) \\ &= \mathbb{P}(B/A) + \mathbb{P}(\overline{A})\,\big[\mathbb{P}(B/\overline{A}) - \mathbb{P}(B/A)\big] \end{aligned}$$

$\square$

Donc $\Big(\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)\Big) \Leftrightarrow \Big(\mathbb{P}(B) = \mathbb{P}(B/A)\Big) \Leftrightarrow \Big(\mathbb{P}(B/A) = \mathbb{P}(B/\overline{A})\Big)$ et de même en échangeant les rôles de A et B.

A et B jouent des rôles symétriques dans leur $\mathbb{P}$-indépendance.

Proposition 25

Si A et B sont des évènements $\mathbb{P}$-indépendants, alors $\overline{A}$ et B, A et $\overline{B}$, $\overline{A}$ et $\overline{B}$ sont aussi $\mathbb{P}$-indépendants.

Démonstration. Il suffit de démontrer que si A et B sont $\mathbb{P}$-indépendants, alors A et $\overline{B}$ sont $\mathbb{P}$-indépendants :
$A = (B \cap A) + (\overline{B} \cap A)$, donc $\mathbb{P}(A) = \mathbb{P}(B \cap A) + \mathbb{P}(\overline{B} \cap A) = \mathbb{P}(A)\,\mathbb{P}(B) + \mathbb{P}(\overline{B} \cap A)$
donc $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(A) - \mathbb{P}(A)\,\mathbb{P}(B) = \mathbb{P}(A)\Big(1 - \mathbb{P}(B)\Big) = \mathbb{P}(A)\,\mathbb{P}(\overline{B})$ □

2.3 Indépendance de n évènements (avec $n \geqslant 2$)

Définition 26

Soient A, B et C 3 évènements de l'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$.
Ils sont dits **$\mathbb{P}$-mutuellement indépendants** si et seulement si :

- ils sont 2 à 2 $\mathbb{P}$-indépendants
- $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$

Pour $n = 3$, l'indépendance mutuelle entraîne l'indépendance deux à deux mais la réciproque est fausse.

Généralisation :
n avec $n \geqslant 2$ évènements $\Big(A_i\Big)_{1 \leqslant i \leqslant n}$ sont dit P-**mutuellement indépendants** si et seulement si pour toute famille finie $K \subset \{1, \ldots, n\}$, avec $\text{card}(K) \geqslant 2$, on a : $\mathbb{P}\left(\bigcap_{i \in K} A_i\right) = \prod_{i \in K} \mathbb{P}(A_i)$
Pour $n \geqslant 3$, l'indépendance mutuelle entraîne l'indépendance deux à deux mais la réciproque est fausse.
L'indépendance mutuelle est assurée en remplaçant éventuellement tout évènement A_i par son contraire $\overline{A_i}$.

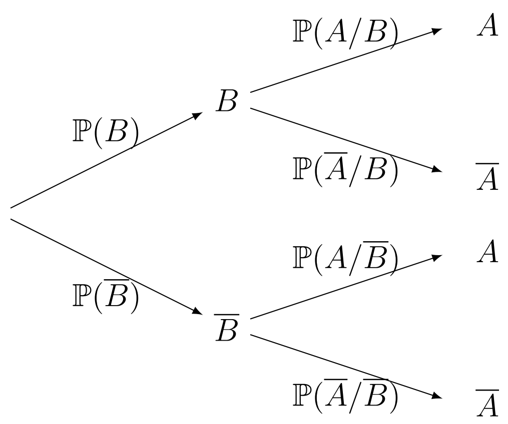
Exemple d'un contrôle de qualité :

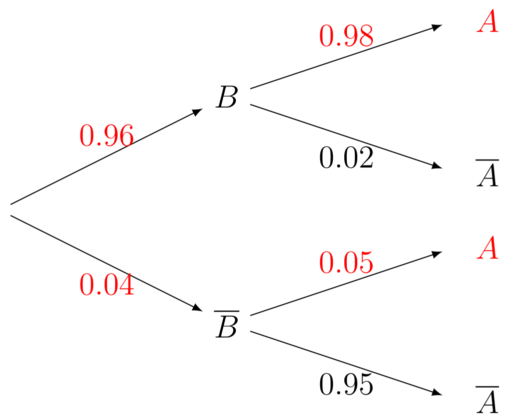
Une fabrique produit des articles, avec une probabilité globale de 4% qu'ils soient défectueux.
Un procédé de contrôle rapide mais imparfait, conduit à mettre au rebut les articles corrects avec une probabilité de 2% et à accepter des articles défectueux avec une probabilité de 5%. Quelle est la probabilité qu'un article pris au hasard soit accepté ?

Pour un article, on distingue les évènements :
- B il est bon ou correct
- $\overline{B}$ il est défectueux
- A il est accepté au contrôle
- $\overline{A}$ il est refusé

Données : on sait que
- $\mathbb{P}(\overline{B}) = 0.04$
- $\mathbb{P}(\overline{A}/B) = 0.02$
- $\mathbb{P}(A/\overline{B}) = 0.05$

Arbre des causes : on représente tous les cas possibles





On a : $\Omega = B + \overline{B}$
donc $A = A \cap (B + \overline{B}) = (A \cap B) + (A \cap \overline{B})$.
D'où

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \\ &= \mathbb{P}(B)\,\mathbb{P}(A/B) + \mathbb{P}(\overline{B})\,\mathbb{P}(A/\overline{B}) \\ &= 0.96 \times 0.98 + 0.04 \times 0.05 = 0.9428 \end{aligned}$$

Ainsi : $\mathbb{P}(A) = 0.9428$

2.4 Formule de la probabilité complète

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé fini.

Définition 27

Soit $n \geqslant 2$. n évènements $\left(B_i\right)_{1\leqslant i\leqslant n}$ de probabilités non nulles forment un **système complet d'évènements** si et seulement s'ils constituent une **partition** de Ω.

Ω est ainsi la réunion disjointe de ces évènements : $\Omega = B_1 + \cdots + B_n$.
Soit un système complet d'évènements $\left(B_i\right)_{1\leqslant i\leqslant n}$ tel que $\forall i \in \{1, \ldots, n\}, \mathbb{P}(B_i) \neq \emptyset$.
Soit A un évènement de $\mathcal{A}$:

$$\begin{aligned} A &= A \cap \Omega = (A \cap B_1) + \cdots + (A \cap B_n) \\ \mathbb{P}(A) &= \mathbb{P}(A \cap B_1) + \cdots + \mathbb{P}(A \cap B_n) \\ &= \mathbb{P}(B_1)\mathbb{P}(A/B_1) + \cdots + \mathbb{P}(B_n)\mathbb{P}(A/B_n) \end{aligned}$$

Théorème 28 (Formule de la probabilité complète)

On suppose $\forall i,\ \mathbb{P}(B_i) \neq 0$. Si $\left(B_i\right)_{1\leqslant i\leqslant n}$ est un système complet d'évènements, pour tout évènement A on a :

$$\mathbb{P}(A) = \sum_{1\leqslant i\leqslant n} \mathbb{P}(A/B_i)\,\mathbb{P}(B_i)$$

2.5 Formule de Bayes

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé fini.

Théorème 29 (Formule de Bayes, cas simple)

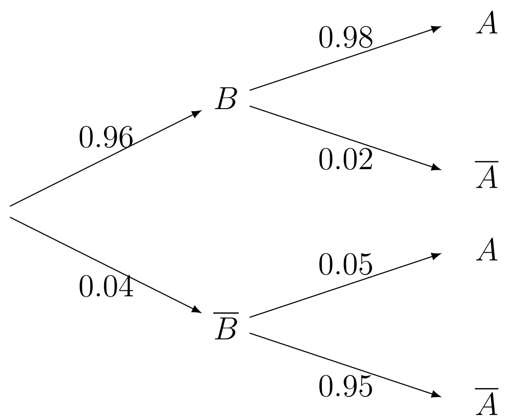
- Cas de deux évènements A et B tels que $\mathbb{P}(A) \cdot \mathbb{P}(B) \neq 0$, on a : $\quad \mathbb{P}(B/A) = \dfrac{\mathbb{P}(A/B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)}$.

On a aussi : $\mathbb{P}(A/B)\,\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B/A)\,\mathbb{P}(A)$.

C'est un **changement de point de vue** : on passe de probabilités « sachant B », à des probabilités « sachant A ».

Exemples :
1) Exemple du contrôle de qualité (suite) :

Pour un article, on a distingué les évènements :
- B il est bon ou correct
- $\overline{B}$ il est défectueux
- A il est accepté au contrôle
- $\overline{A}$ il est refusé



a) *Risque de première espèce :*
Quelle est la probabilité pour qu'un article accepté par ce contrôle rapide soit en réalité défectueux ?

$$\mathbb{P}(\overline{B}/A) = \frac{\mathbb{P}(A \cap \overline{B})}{\mathbb{P}(A)} = \frac{\mathbb{P}(A/\overline{B}) \times \mathbb{P}(\overline{B})}{\mathbb{P}(A)} = \frac{0.04 \times 0.05}{0.9428} = 0.0021$$

C'est le « risque client ».

b) *Risque de deuxième espèce :* Quelle est la probabilité pour qu'un article soit bon, sachant qu'il a été refusé ?

$$\mathbb{P}(B/\overline{A}) = \frac{\mathbb{P}(\overline{A} \cap B)}{\mathbb{P}(\overline{A})} = \frac{\mathbb{P}(\overline{A}/B) \times \mathbb{P}(B)}{\mathbb{P}(\overline{A})} = \frac{0.96 \times 0.02}{0.0572} = 0.0336$$

C'est le « risque vendeur ».

2) Exemple d'un test :

Une population est atteinte par un virus. On dispose d'un test.

Pour un individu, on distingue les évènements :

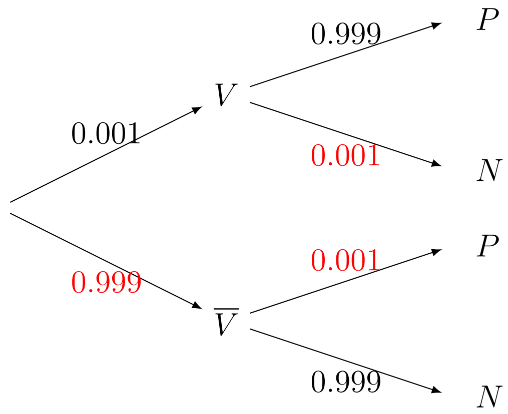
- V il est porteur du virus
- $\overline{V}$ il n'est pas porteur
- P son test est positif
- N son test est négatif

On envisage différentes situations selon :

- la probabilité (proportion) qu'une personne soit porteur du virus : $\mathbb{P}(V)$
- la probabilité qu'un test soit positif pour un porteur du virus $\mathbb{P}(P/V)$
- la probabilité qu'un test soit négatif pour un non-porteur du virus $\mathbb{P}(N/\overline{V})$

1^{er} cas :

$$\left.\begin{array}{l} \bullet\ \mathbb{P}(V) = 0.001 \\ \bullet\ \mathbb{P}(P/V) = 0.999 \\ \bullet\ \mathbb{P}(N/\overline{V}) = 0.999 \end{array}\right\}$$ Les données sont en noir, sur la figure, les compléments calculés en rouge.



$$\mathbb{P}(V/P) = \frac{\mathbb{P}(P/V) \times \mathbb{P}(V)}{\mathbb{P}(P)}$$

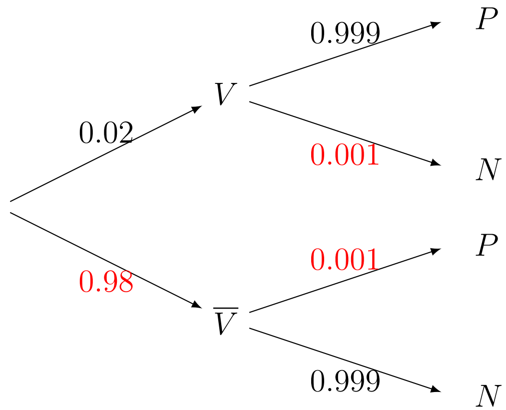
est la probabilité d'être porteur du virus, sachant que le test est positif.

$$\begin{aligned} \bullet\ \mathbb{P}(V/P) &= \frac{\mathbb{P}(P/V) \times \mathbb{P}(V)}{\mathbb{P}(P/V) \times \mathbb{P}(V) + \mathbb{P}(P/\overline{V}) \times \mathbb{P}(\overline{V})} \\ &= \frac{0.999 \times 0.001}{0.999 \times 0.001 + 0.001 \times 0.999} = \frac{1}{2} \end{aligned}$$ **Non concluant !**

Avec une probabilité seulement de 0.5 que le test soit fiable, il n'est pas plus concluant qu'un jeu équitable de « pile ou face ».

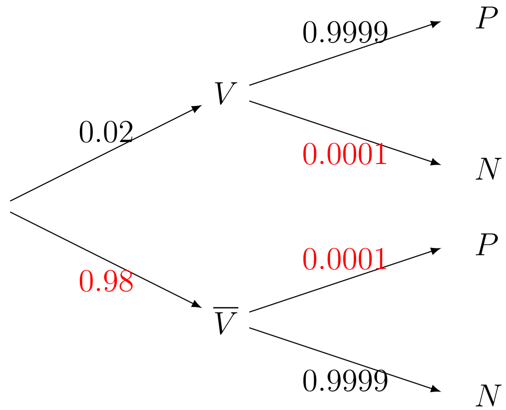
$2^{ème}$ cas :

- $\mathbb{P}(V) = 0.02$
- $\mathbb{P}(P/V) = 0.999$
- $\mathbb{P}(N/\overline{V}) = 0.999$
- $\mathbb{P}(V/P) = \dfrac{0.02 \times 0.999}{0.02 \times 0.999 + 0.98 \times 0.001} \approx 0.953$



$3^{ème}$ cas :

- $\mathbb{P}(V) = 0.02$
- $\mathbb{P}(P/V) = 0.9999$
- $\mathbb{P}(N/\overline{V}) = 0.9999$
- $\mathbb{P}(V/P) = \dfrac{0.02 \times 0.9999}{0.02 \times 0.9999 + 0.98 \times 0.0001} \approx 0.995$



Théorème 30 (Formule de Bayes, cas général)

• Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et un entier $n \geqslant 1$. Si $\left(B_i\right)_{1 \leqslant i \leqslant n}$ est un **système complet d'évènements**, avec $\forall i,\ \mathbb{P}(B_i) \neq 0$, alors, pour A tel que $\mathbb{P}(A) \neq 0$, on a : $\forall j \in \{1, \ldots, n\}$ $\quad \mathbb{P}(B_j/A) = \dfrac{\mathbb{P}(A/B_j) \times \mathbb{P}(B_j)}{\sum\limits_{1 \leqslant i \leqslant n} \mathbb{P}(A/B_i) \times \mathbb{P}(B_i)}$

2.6 Probabilité produit

Soient $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces probabilisables. On munit $\Omega \times \Omega'$ d'une tribu, la tribu engendrée par $\mathcal{A} \times \mathcal{A}'$.

Définition 31

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ et $(\Omega', \mathcal{A}', \mathbb{P}')$ deux espaces probabilisés, on définit la **probabilité produit** Π sur $\Omega \times \Omega'$ en étendant par σ-sommabilité à la tribu engendrée, la fonction définie sur $\mathcal{A} \times \mathcal{A}'$ par : $\forall \omega_i \in \mathcal{A},\ \forall \omega'_j \in \mathcal{A}', \quad \Pi\big((\omega_i, \omega'_j)\big) = \mathbb{P}(\omega_i) \times \mathbb{P}'(\omega'_j)$.
$(\Omega \times \Omega', \mathcal{A} \times \mathcal{A}', \Pi)$ est appelé **espace probabilisé produit**.

Pour Ω et Ω' discrets, (on l'admet dans le cas général), on peut vérifier que cela définit bien une probabilité sur $\Omega \times \Omega'$, puisque :
• $\forall \omega_i \in \mathcal{A}, \forall \omega'_j \in \mathcal{A}', \quad \Pi\big((\omega_i, \omega'_j)\big) \geqslant 0$,
• et en prenant les évènements élémentaires : $\displaystyle\sum_{(i,j)\in I\times J} \Pi\big((\omega_i, \omega'_j)\big) = \sum_{(i,j)\in I\times J} \mathbb{P}(\omega_i) \times \mathbb{P}'(\omega'_j) = \sum_{i\in I} \mathbb{P}(\omega_i) \sum_{j\in J} \mathbb{P}'(\omega'_j) = 1 \times 1 = 1$

Remarque : On peut étendre le procédé pour $n \geqslant 2$ espaces probabilisables, avec $\mathbb{P}\big(\{(\omega_1, \ldots, \omega_n)\}\big) = \prod \mathbb{P}(\{\omega_i\})$

Exemple d'un jet de deux dés : les dés ne sont pas pipés, les faces sont **équiprobables**.

- $1^{ère}$ **formalisation** : on jette les deux dés discernables en meme temps.

Le résultat est un couple (i, j) où $\begin{cases} 1 \leqslant i \leqslant 6 \\ 1 \leqslant j \leqslant 6 \end{cases}$

$\Omega = \{(i, j), 1 \leqslant i \leqslant 6, 1 \leqslant j \leqslant 6\}$ compte 36 éléments avec l'équiprobabilité : $\mathbb{P}\Big(\{(i, j)\}\Big) = \dfrac{1}{36}$, pour un **évènement élémentaire**.

- $2^{ème}$ **formalisation** : on jette un dé deux fois de suite.

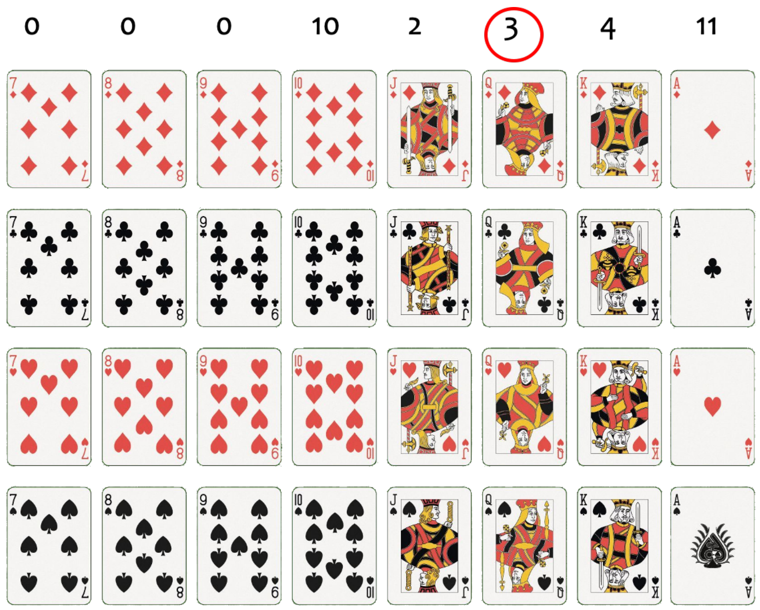
$\Omega_1 = \{i, 1 \leqslant i \leqslant 6\}$ compte 6 éléments avec l'équiprobabilité : $\mathbb{P}_1\Big(\{i\}\Big) = \dfrac{1}{6}$.
Le résultat de chaque jet est un entier i où $1 \leqslant i \leqslant 6$
On considère alors la probabilité produit sur $\Omega = \Omega_1 \times \Omega_1 \quad \mathbb{P}\Big(\{i\} \times \{j\}\Big) = \mathbb{P}_1\Big(\{i\}\Big) \times \mathbb{P}_1\Big(\{j\}\Big) = \dfrac{1}{6 \times 6} = \dfrac{1}{36}$.
On retrouve la même probabilité d'un évènement élémentaire que la première formalisation.

Chapitre 3

Variables aléatoires

Exemple d'introduction : Considérons un jeu de belote sans atout.
Chaque carte a la valeur X suivante :

$$\begin{array}{llll} 7 \longmapsto 0 & 8 \longmapsto 0 & 9 \longmapsto 0 & 10 \longmapsto 10 \\ \text{valet} \longmapsto 2 & \text{dame} \longmapsto 3 & \text{roi} \longmapsto 4 & \text{as} \longmapsto 11 \end{array}$$



Par exemple, les quatre dames ont pour valeur 3 : $X\big(\text{« Dame de coeur »}\big) = 3$

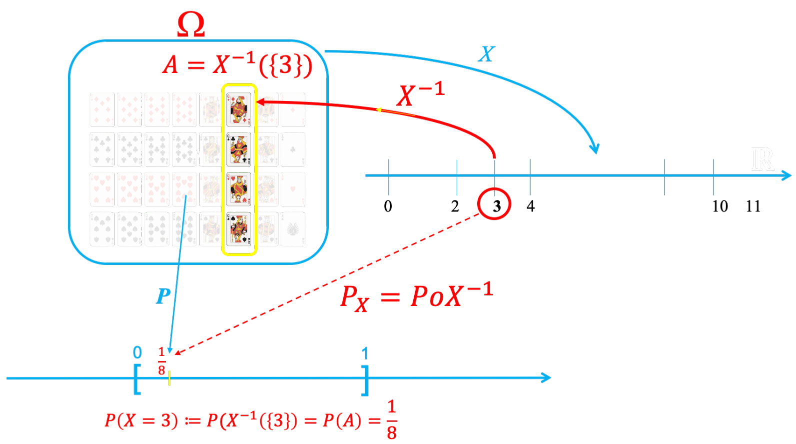
Soit Ω l'ensemble des tirages d'une des 32 cartes avec $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ espace probabilisé par **équiprobabilité** : $\mathbb{P}\big(\{\omega_i\}\big) = \dfrac{1}{32}$.

Par exemple, l'évènement : on a tiré la dame de coeur, noté « Dame de coeur », a une probabilité : $\mathbb{P}\big(\text{« Dame de coeur »}\big) = \dfrac{1}{32}$

On considère l'application $X : \Omega \to \mathbb{R}$ qui à un tirage associe sa valeur.
Son image est notée $X\langle\Omega\rangle = \{0, 2, 3, 4, 10, 11\}$.
On dira que X est une variable aléatoire sur Ω, à valeurs réelles.

On veut savoir quelle est la probabilité d'obtenir comme résultat du tirage une certaine valeur x de $\mathbb{R}$, par exemple $x = 3$. On notera cette probabilité : $\mathbb{P}\big(X = 3\big)$.

On a $X = 3$ dans le cas des quatre « dames », qui ont pour image 3 par l'application X. L'image réciproque de $\{3\}$ par X, que l'on note $\overset{-1}{X}\big(\{3\}\big)$, est constituée des quatre « dames ».



Chaque « dame » a une probabilité $\dfrac{1}{32}$ d'être tirée, et donc $\mathbb{P}(X = 3) = \dfrac{4}{32} = \dfrac{1}{8}$.

De même : $\quad \mathbb{P}(X = 0) = \dfrac{12}{32} \quad \mathbb{P}(X = 2) = \dfrac{4}{32} \quad \mathbb{P}(X = 3) = \dfrac{4}{32}$

$\qquad \mathbb{P}(X = 4) = \dfrac{4}{32} \quad \mathbb{P}(X = 11) = \dfrac{4}{32} \quad \mathbb{P}(X = 10) = \dfrac{4}{32}$

Et $\forall x \in \mathbb{R} \setminus X\langle\Omega\rangle, \quad \mathbb{P}(X = x) = 0$

3.1 Variable Aléatoire

Définition 32

Soient $(\Omega, \mathcal{A})$ et $(\Omega', \mathcal{A}')$ deux espaces probabilisables. Soit l'application $\quad X : \begin{cases} \Omega & \longrightarrow & \Omega' \\ \omega & \longmapsto & X(\omega) \end{cases}$

On dit que X est une **application mesurable**, ou que X est une **variable aléatoire** de $(\Omega, \mathcal{A})$ vers $(\Omega', \mathcal{A}')$ si et seulement si : $\quad \forall A' \in \mathcal{A}', \quad \overset{-1}{X}(A') \in \mathcal{A} \quad$ (l'image réciproque par X d'un évènement de Ω' est un évènement de Ω).

Notation : Pour une application f définie de E dans E', et pour un sous-ensemble $A' \subset E'$ de l'ensemble d'arrivée, on note $\overset{-1}{f}\big(A'\big)$ le sous-ensemble de E constitué de tous les éléments de E dont l'image est dans A' : $\overset{-1}{f}\big(A'\big) = \big\{x \in E, \quad f(x) \in A'\big\}$.

Remarque : Si $\mathcal{A} = \mathcal{P}(\Omega)$ (on parle de « tribu totale »), alors $\forall A' \in \mathcal{A}', \quad \overset{-1}{X}(A') \subset \Omega \quad$ donc $\overset{-1}{X}(A') \in \mathcal{P}(\Omega)$.
Toute application X de $\big(\Omega, \mathcal{P}(\Omega)\big)$ vers $\big(\Omega', \mathcal{A}'\big)$ définit, dans ce cas, une **application mesurable**.

3.2 Espace probabilisé image

Définition 33

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $(\Omega', \mathcal{A}')$ un espace probabilisable, et X une application mesurable de $(\Omega, \mathcal{A})$ vers $(\Omega', \mathcal{A}')$.

On définit l'application $\mathbb{P}_X : \left\{ \begin{array}{ccc} \mathcal{A}' & \longrightarrow & [0,1] \\ A' & \longmapsto & \mathbb{P}_X(A') = \mathbb{P}\left(\overset{-1}{X}(A')\right) \end{array} \right.$

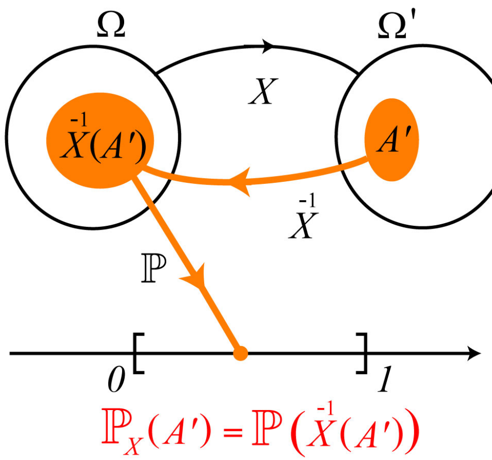
$\left(\Omega', \mathcal{A}', \mathbb{P}_X\right)$ est appelé **espace probabilisé image**.

Pour tout $A' \in \mathcal{A}'$, on a bien $\overset{-1}{X}(A') \in \mathcal{A}$, et puisqu'il est dans la tribu, la mesure de sa probabilité est connue : $\mathbb{P}\left(\overset{-1}{X}(A')\right)$ existe.

X est une Variable Aléatoire (VA), $\quad X : \Omega \longrightarrow \Omega'$

On peut définir $\mathbb{P}_X$: $\left\{ \begin{array}{ccccc} \mathcal{A}' & \xrightarrow{\overset{-1}{X}} & \mathcal{A} & \xrightarrow{\mathbb{P}} & [0,1] \\ A' & \longmapsto & \overset{-1}{X}(A') & \longmapsto & \mathbb{P}\left(\overset{-1}{X}(A')\right) \end{array} \right.$

$\left(\Omega', \mathcal{A}', \mathbb{P}_X\right)$ est l'espace probabilisé image



Vérifions que $\mathbb{P}_X$ définit bien une **probabilité** sur $(\Omega', \mathcal{A}')$:

- $\forall A' \in \mathcal{A}', \quad \mathbb{P}_X(A') \geqslant 0$ car $\mathbb{P}\left(\overset{-1}{X}(A')\right) \geqslant 0$.
- $\mathbb{P}_X(\Omega') = \mathbb{P}\left(\overset{-1}{X}(\Omega')\right) = \mathbb{P}(\Omega) = 1$.
- Soit une **famille dénombrable** d'évènements $\left(A'_i\right)_{i \in I}$ de $\mathcal{A}'$, **incompatibles 2 à 2** :

$$\mathbb{P}_X\left(\sum_i A'_i\right) = \mathbb{P}\left(\overset{-1}{X}\left(\sum_i A'_i\right)\right) = \mathbb{P}\left(\sum_i \overset{-1}{X}(A'_i)\right) = \sum_i \mathbb{P}\left(\overset{-1}{X}(A'_i)\right) = \sum_i \mathbb{P}_X(A'_i)$$

3.3 Variable aléatoire réelle

Définition 34

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. On considère $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ espace probabilisable, avec $\mathcal{B}_{\mathbb{R}}$ la tribu des boréliens.
Soit $X : \left\{ \begin{array}{l} \Omega \longrightarrow \mathbb{R} \\ \omega \longmapsto X(\omega) \end{array} \right.$ telle que $\forall B \in \mathcal{B}_{\mathbb{R}}$, on a $\overset{-1}{X}(B) \in \mathcal{A}$. Alors X est une application mesurable, appelée **variable aléatoire réelle**, qu'on notera **VAR**.
$(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P}_X)$ est l'espace probabilisé image avec la même construction qu'au paragraphe précédent.

Si $X\langle\Omega\rangle$ est **fini**, alors on peut l'écrire $X\langle\Omega\rangle = \{x_1, x_2, \dots, x_n\}$, avec $n = \operatorname{card}(\Omega)$.
Dans ce cas, pour montrer que X est une **VARD** (Variable Aléatoire Réelle Discrète), il faut et il suffit de montrer que :
$\forall x_i \in X\langle\Omega\rangle, \quad \overset{-1}{X}(\{x_i\}) \in \mathcal{A}$.

Exemple : Soient $\Omega = \{a, b, c\}$ et $\mathcal{B} = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$. C'est bien une tribu de Ω.

On définit deux applications : $X_1 : \left\{ \begin{array}{l} \Omega \to \mathbb{R} \\ a \longmapsto 0 \\ b \longmapsto 1 \\ c \longmapsto 1 \end{array} \right.$ et $X_2 : \left\{ \begin{array}{l} \Omega \to \mathbb{R} \\ a \longmapsto 0 \\ b \longmapsto 0 \\ c \longmapsto 1 \end{array} \right.$ On a $X_1\langle\Omega\rangle = X_2\langle\Omega\rangle = \{0, 1\}$.

- $\forall x \in \mathbb{R} \setminus \{0, 1\}, \overset{-1}{X_1}(\{x\}) = \emptyset$ et $\emptyset \in \mathcal{B}$.
- $\overset{-1}{X_1}(\{0\}) = \{a\}$ et $\{a\} \in \mathcal{B}$.
- $\overset{-1}{X_1}(\{1\}) = \{b, c\}$ et $\{b, c\} \in \mathcal{B}$.

Donc X_1 définit bien une VARD.

En revanche, $\overset{-1}{X_2}(\{1\}) = \{c\}$ et $\{c\} \notin \mathcal{B}$ donc X_2 ne définit pas une VARD.
Remarque : Pour qu'une application soit mesurable, il est nécessaire qu'elle soit constante sur chaque évènement élémentaire.

3.4 Fonction de répartition (*cumulative*) d'une VAR

3.4.1 Probabilité attachée à un intervalle

Avec les notations précédentes, on considère $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \mathbb{P}_X)$ l'espace probabilisé image et $]a,b] \in \mathcal{B}_{\mathbb{R}}$.
Pour $a < b$, on a, par réunion disjointe : $\mathbb{P}_X(]-\infty, b]) = \mathbb{P}_X(]-\infty, a]) + \mathbb{P}_X(]a,b])$ ou $\mathbb{P}(X \leqslant b) = \mathbb{P}(X \leqslant a) + \mathbb{P}(a < X \leqslant b)$.

Ainsi on a : $\boxed{\forall (a,b) \in \mathbb{R}^2, \quad \mathbb{P}(a < X \leqslant b) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X \leqslant a)}$

3.4.2 Définition de la fonction de répartition

Définition 35

On définit F la **fonction de répartition** de la VAR X par $F : \left\{ \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \mathbb{P}(X \leqslant x) \end{array} \right.$

Propriété 36

Soit X une VAR de fonction de répartition F.

1. $\forall x \in \mathbb{R}$, $0 \leqslant F(x) \leqslant 1$
2. $\forall (a,b) \in \mathbb{R}^2$, $\mathbb{P}(a < X \leqslant b) = F(b) - F(a)$
3. $\lim\limits_{x \to -\infty} F(x) = 0$, notée $F(-\infty) = 0$
4. $\lim\limits_{x \to +\infty} F(x) = 1$ notée $F(+\infty) = 1$
5. F est une application **croissante** sur $\mathbb{R}$
6. $\forall a \in \mathbb{R}$, F est **continue à droite** en a
7. L'ensemble des points de discontinuité de F est au plus dénombrable.

Démonstrations :

1. Vient du fait que la probabilité d'un évènement est un nombre compris entre 0 et 1.
2. Cela a été vu juste au-dessus.

3. $]-\infty, x] \to \emptyset$ quand $x \to -\infty$. Par continuité monotone, $\lim\limits_{x\to-\infty} \mathbb{P}_X(]-\infty, x]) = \mathbb{P}_X(\emptyset) = 0$ et $\lim\limits_{x\to-\infty} \mathbb{P}(X \leqslant x) = \lim\limits_{x\to-\infty} F(x) = 0$.
4. $\overline{]-\infty, x]} =]x, +\infty[\to \emptyset$ quand $x \to +\infty$. Ainsi $\lim\limits_{x\to+\infty} (]x, +\infty[) = 0$ et $\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqslant x) = 1 - F(x)$ d'où $\lim\limits_{x\to+\infty} F(x) = 1$.
5. $\forall (x, y) \in \mathbb{R}^2$ tel que $x \leq y$, on a : $F(y) = \mathbb{P}(X \leq y) = \mathbb{P}[(X \leq x) \cup (x < X \leq y)] = F(x) + \mathbb{P}(x < X \leq y) \geq F(x)$
6. Étudions la continuité en un point $a \in \mathbb{R}$.
 Continuité à gauche : $]x, a] \to \{a\}$ quand $x \to a^-$. Par l'axiome de continuité monotone, $\lim\limits_{x\to a^-} \mathbb{P}_X(]x, a]) = \mathbb{P}(X = a)$
 donc $\lim\limits_{x\to a^-} \mathbb{P}(x < X \leqslant a) = \mathbb{P}(X = a)$. Et ainsi $\lim\limits_{x\to a^-} (F(a) - F(x)) = \mathbb{P}(X = a)$ soit $\boxed{\lim\limits_{x\to a^-} F(x) = F(a) - \mathbb{P}(X = a)}$.
 - Si $\mathbb{P}(X = a) > 0$, alors $\lim\limits_{x\to a^-} F(x) \neq F(a)$ et F n'est pas continue à gauche en a.
 - Si $\mathbb{P}(X = a) = 0$, alors $\lim\limits_{x\to a^-} F(x) = F(a)$ et F est continue à gauche en a.

 Continuité à droite : $]a, x] \to \emptyset$ quand $x \to a^+$. Par l'axiome de continuité monotone, $\lim\limits_{x\to a^+} \mathbb{P}_X(]a, x]) = 0$ donc $\lim\limits_{x\to a^+} F(x) = F(a)$.
 Ainsi, $\forall a \in \mathbb{R}$, F est continue à droite en a.
7. L'ensemble $\{a \in \mathbb{R}, \mathbb{P}(X = a) > 0\}$ est l'ensemble des points de discontinuité de F. Cet ensemble est **au plus dénombrable**.

 En effet, il y a au plus : $\left\{\begin{array}{l} 1 \text{ point dont le saut a une amplitude } \geqslant 1 \\ 2 \text{ points dont le saut a une amplitude } \geqslant \dfrac{1}{2} \\ \cdots \quad \cdots \quad \cdots \\ n \text{ points dont le saut a une amplitude } \geqslant \dfrac{1}{n} \\ \cdots \quad \text{etc} \quad \cdots \end{array}\right.$

 Donc il y a un nombre au plus dénombrable de points où il y a un saut de discontinuité.

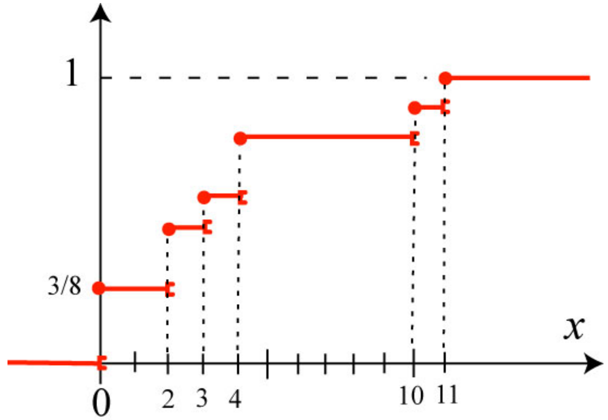
Exemple : On s'intéresse au jeu de belote sans atout. Alors $F : \left\{ \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \mathbb{P}(X \leqslant x) \end{array} \right.$ vérifie, selon les valeurs de x :

si $x < 0$, $\quad F(x) = 0$, $\quad$ et puis :

si $\; 0 \leqslant x < 2$: $\quad F(x) = \dfrac{12}{32} = \dfrac{3}{8}$, $\qquad$ si $\; 2 \leqslant x < 3$: $\quad F(x) = \dfrac{16}{32} = \dfrac{1}{2}$,

si $\; 3 \leqslant x < 4$: $\quad F(x) = \dfrac{20}{32} = \dfrac{5}{8}$, $\qquad$ si $\; 4 \leqslant x < 10$: $\quad F(x) = \dfrac{24}{32} = \dfrac{6}{8}$,

si $\; 10 \leqslant x < 11$: $\quad F(x) = \dfrac{28}{32} = \dfrac{7}{8}$, $\qquad$ si $\; 11 \leqslant x$: $\quad F(x) = 1$.



3.5 Répartition absolument continue

3.5.1 Variable aléatoire continue

Définition 37

On dit qu'une VAR X est **continue**, (notée note VARC), lorsque sa fonction de répartition est continue sur $\mathbb{R}$.

$$\text{Alors : } \forall a \in \mathbb{R}, \quad \mathbb{P}(X = a) = 0$$

En effet : $\lim\limits_{x \to a^-} F(x) = F(a)$, donc $\mathbb{P}(X = a) = F(a) - \lim\limits_{x \to a^-} F(x) = 0$ (graphiquement il n'y a pas de saut)

3.5.2 Fonction de densité de probabilité

Définition 38

On appelle **fonction de densité** de probabilité sur $\mathbb{R}$ toute fonction $f : \mathbb{R} \longrightarrow \mathbb{R}$ telle que :

1. $\forall x \in \mathbb{R},\ f(x) \geqslant 0$ (positivité)
2. f est presque partout continue sur $\mathbb{R}$, autrement dit continue sauf sur un ensemble de points au plus dénombrable.
3. f est **sommable** sur $\mathbb{R}$ avec $\displaystyle\int_{\mathbb{R}} f(x)\mathrm{d}x = 1$.

Définition 39

Une VAR X est dite **absolument continue** s'il existe une densité de probabilité f telle que :

$$\forall B \in \mathcal{B}_{\mathbb{R}}, \quad \mathbb{P}_X(B) = \int_B f(t)dt.$$

Proposition 40

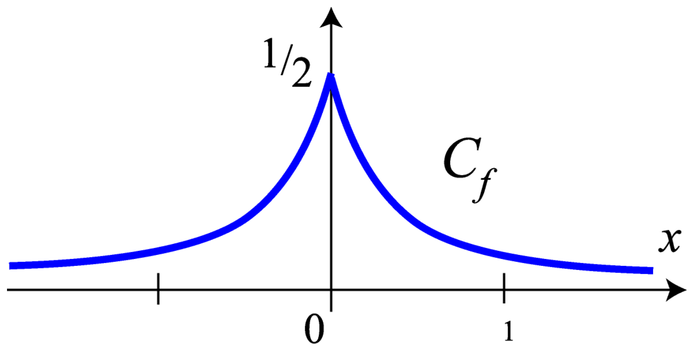
Si X est une VAR absolument continue admettant pour densité de probabilité f, celle-ci est alors liée à la fonction de répartition F de X : $\forall x \in \mathbb{R}, \quad \mathbb{P}(X \leqslant x) = F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$

Donc pour tout x où F est dérivable on a $F'(x) = f(x)$

3.5.3 Exemples

Exemple 1 :

X VARC suit la **loi de Laplace**, si elle admet une **fonction de densité** définie par : $\forall x \in \mathbb{R}, \quad f(x) = \dfrac{1}{2}\,\mathrm{e}^{-|x|}$



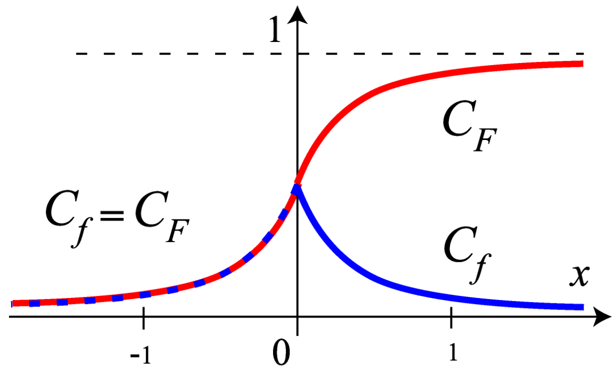
f définit bien une densité de probabilité car

- $\forall x \in \mathbb{R}, \quad f(x) \geqslant 0$,
- f est continue sur $\mathbb{R}$,
- $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 2 \times \left[-\mathrm{e}^{-x}\right]_0^{+\infty} = 1$.

La fonction de répartition est définie par $F : x \longmapsto \mathbb{P}(X \leqslant x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$. On distingue deux cas :

- Si $x < 0$, $F(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{2}\mathrm{e}^t dt = \frac{1}{2}\left[\mathrm{e}^t\right]_{-\infty}^{x} = \frac{1}{2}\,\mathrm{e}^x$.
- Si $x \geqslant 0$, $F(x) = \displaystyle\int_{-\infty}^{0} \frac{1}{2}\mathrm{e}^t dt + \int_0^x \frac{1}{2}\mathrm{e}^{-t} dt = 1 - \frac{1}{2}\,\mathrm{e}^{-x}$

Ainsi, on a : $\quad F(x) = \begin{cases} \dfrac{\mathrm{e}^x}{2} = f(x) & \text{si} \quad x < 0 \\ 1 - \dfrac{\mathrm{e}^{-x}}{2} & \text{si} \quad x \geqslant 0 \end{cases}$



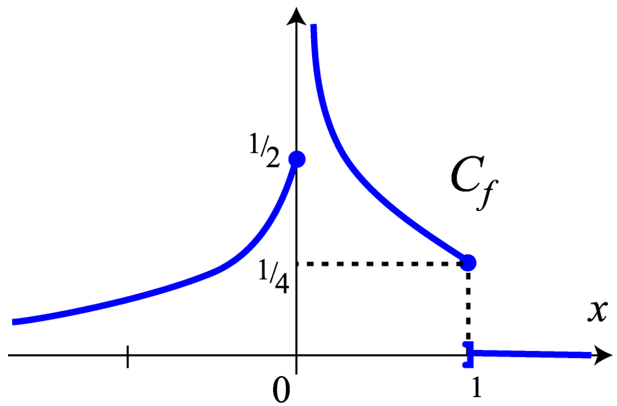
Exemple 2 : Une densité de probabilité peut admettre éventuellement des limites infinies aux points de discontinuité. On parlera alors de densité de probabilité généralisée (comme pour les intégrales).

Ainsi : $f : x \longmapsto \begin{cases} \frac{1}{2}\mathrm{e}^{x} & \text{si} \quad x \leqslant 0 \\ \frac{1}{4\sqrt{x}} & \text{si} \quad 0 < x \leqslant 1 \\ 0 & \text{si} \quad x > 1 \end{cases}$ est bien une **densité de probabilité**,

car f est **positive**, **continue par morceaux**,

et $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_{-\infty}^{0} f(x)\,\mathrm{d}x + \int_{0}^{1} f(x)\,\mathrm{d}x = \frac{1}{2}\Big[\mathrm{e}^{x}\Big]_{-\infty}^{0} + \frac{1}{2}\Big[\sqrt{x}\Big]_{0}^{1} = \frac{1}{2} + \frac{1}{2} = 1.$

On constate une limite infinie en 0^{+}.



3.6 Fonction aléatoire d'une VARC

Exemple 1 : Soit X une VARC dont la loi est connue par une fonction f densité de probabilité $f : x \longmapsto \begin{cases} 2x & \text{si} \quad x \in [0,1] \\ 0 & \text{si} \quad x \notin [0,1] \end{cases}$

$X\langle\Omega\rangle = [0,1]$ et on vérifie facilement que f est bien une fonction de densité de probabilité, car $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = \int_{0}^{1} 2x\,\mathrm{d}x = \Big[x^2\Big]_0^1 = 1.$

Soit φ la fonction de X définie par : $\varphi : X \longmapsto X^3 - 2$. **Déterminons la loi de probabilité** de la fonction aléatoire $Y = \varphi(X)$.
Notons F la fonction de répartition de X et G celle de Y.
Pour $0 \leqslant x \leqslant 1$, on a $-2 \leqslant x^3 - 2 \leqslant -1$ donc $Y\langle\Omega\rangle = [-2,-1]$.

$$\begin{aligned} \forall y \in [-2,-1], \quad G(y) &= \mathbb{P}(Y \leqslant y) &&= \mathbb{P}(X^3 - 2 \leqslant y) \quad = \mathbb{P}(X^3 \leqslant y+2) \\ &= \mathbb{P}\left(X \leqslant (y+2)^{\frac{1}{3}}\right) &&= F\left((y+2)^{\frac{1}{3}}\right) \end{aligned}$$

G, par composition, est donc dérivable sur $]-2,-1[$

$$\text{et } \forall y \in]-2,-1[, \quad g(y) = G'(y) = \frac{1}{3}F'\left((y+2)^{\frac{1}{3}}\right) \times (y+2)^{-\frac{2}{3}} = \frac{1}{3}f\left((y+2)^{\frac{1}{3}}\right) \times (y+2)^{-\frac{2}{3}}$$

Ainsi $\quad g : y \longmapsto \begin{cases} \frac{2}{3}(y+2)^{-\frac{1}{3}} & \text{si} \quad y \in]-2,-1[\\ 0 & \text{si} \quad y \notin]-2,-1[\end{cases}$ définit une densité de probabilité de la VARC $Y = \varphi(X)$.

Exemple 2 : Considérons X une VAR qui suit la **Loi Uniforme** sur $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$: $X \sim \mathcal{U}\left(\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]\right)$.

$X\langle\Omega\rangle = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, et une fonction de densité de probabilité de X est $f : x \longmapsto \begin{cases} \frac{1}{\pi} & \text{si} \quad x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ 0 & \text{sinon} \end{cases}$

Considérons $Y = \varphi(X) = X^2$. Alors $Y\langle\Omega\rangle = \left[0, \frac{\pi^2}{4}\right]$. Notons F la fonction de répartition de X, et G celle de Y.

$$\forall y \in \left[0, \frac{\pi^2}{4}\right] : \quad G(y) \begin{array}{ll} = \mathbb{P}(Y \leqslant y) & = \mathbb{P}(X^2 \leqslant y) \\ = \mathbb{P}(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) & = F(\sqrt{y}) - F(-\sqrt{y}) \end{array}$$

donc, par dérivation, $\forall y \in \left]0, \frac{\pi^2}{4}\right]$, $\quad g(y) = \frac{1}{2\sqrt{y}}\Big(f(\sqrt{y}) + f(-\sqrt{y})\Big) = \frac{1}{\pi\sqrt{y}}$

Ainsi $g : y \longmapsto \begin{cases} \frac{1}{\pi\sqrt{y}} & \text{si} \qquad y \in \left]0, \frac{\pi^2}{4}\right] \\ 0 & \text{sinon} \end{cases}$

Note : on a bien $\displaystyle\int_0^{\frac{\pi^2}{4}} \frac{1}{\pi\sqrt{y}} \mathrm{d}y = \frac{2}{\pi}\left[\sqrt{y}\right]_0^{\frac{\pi^2}{4}} = 1$

Chapitre 4

Couple de variables aléatoires réelles discrètes

4.1 Application mesurable, espace probabilisé image

On considère $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et l'espace probabilisable $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ muni de la tribu engendrée par les « borels » de $\mathbb{R}^2$ qui sont les quarts d'espace $]-\infty, x[\times]-\infty, y[$.

Définition 41

Une application $V : \Omega \to \mathbb{R}^2$ est **mesurable**, ou constitue un **couple aléatoire**, si pour tout $B \in \mathcal{B}_{\mathbb{R}^2}$, $\overset{-1}{V}(B) \in \mathcal{A}$.

On peut alors transporter la probabilité $\mathbb{P}$,

$$\begin{array}{rcccccc} \mathbb{P}_V : & \mathcal{B}_{\mathbb{R}^2} & \longrightarrow & \mathcal{A} & \longrightarrow & [0,1] \\ & B & \longmapsto & \overset{-1}{V}(B) & \longmapsto & \mathbb{P}\left(\overset{-1}{V}(B)\right) \end{array}$$

On obtient $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2}, \mathbb{P}_V)$ **espace probabilisé image**, et $V = (X, Y)$ définit un **couple de variables aléatoires réelles**.

Définition 42

Le couple aléatoire est dit **discret** si $V\langle\Omega\rangle$ est **fini ou dénombrable** dans $\mathbb{R}^2$.

Alors V définit une application **mesurable** si et seulement si $\forall\{(i,j)\} \in \mathcal{B}_{\mathbb{R}^2}, \quad \overset{-1}{V}(\{(i,j)\}) \in \mathcal{A}$
(pour les évènements élémentaires)

Notons $p_{i,j} = \mathbb{P}\left(\overset{-1}{V}(\{(i,j)\})\right) = \mathbb{P}(X = i, Y = j)$, alors : $\begin{cases} \forall (i,j) \in V\langle\Omega\rangle, p_{i,j} \geqslant 0 \\ \forall (i,j) \notin V\langle\Omega\rangle, p_{i,j} = 0 \end{cases}$

Exemple : On jette successivement 3 pièces non truquées. On désigne par X le nombre de piles obtenus avec les 2 premières pièces, et par Y le nombre de piles obtenus avec les 3 pièces.
On cherche la loi ou la distribution de probabilité du couple $V = (X, Y)$.

On a $X\langle\Omega\rangle = \{0, 1, 2\}$ et $Y\langle\Omega\rangle = \{0, 1, 2, 3\}$. Alors $V\langle\Omega\rangle \subset X\langle\Omega\rangle \times Y\langle\Omega\rangle$.

On remplit le tableau suivant, en déterminant les $p_{i,j}$.

$X=i$ \ $Y=j$	$j=0$	$j=1$	$j=2$	$j=3$	Loi de X
$i=0$	1/8	1/8	0	0	2/8
$i=1$	0	2/8	2/8	0	4/8
$i=2$	0	0	1/8	1/8	2/8
Loi de Y	1/8	3/8	3/8	1/8	1

$\overset{-1}{V}(\{(0,0)\}) = F_1F_2F_3$ donc $\mathbb{P}(F_1F_2F_3) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = p_{0,0} = \frac{1}{8}$
$\overset{-1}{V}(\{(0,2)\}) = \emptyset$ donc $p_{0,2} = 0$
$\overset{-1}{V}(\{(1,1)\}) = \{P_1F_2F_3, F_1P_2F_3\}$ donc $p_{1,1} = \frac{2}{8}$
etc.

Remarquons que l'on a $\sum_i \sum_j p_{i,j} = 1$

4.2 Lois marginales

Définition 43

On appelle **loi marginale** de X (resp. de Y) la loi de X (resp. de Y) qui découle de la loi du couple $V = (X, Y)$.

4.2.1 Loi marginale de X

Évaluons la probabilité que X prenne la valeur $i \in X\langle\Omega\rangle$. Notons $B_{X=i}$ l'évènement $X = i$.

$$\mathbb{P}(X = i) = \mathbb{P}_V(B_{X=i}) = \mathbb{P}\left(\bigcup_j \overset{-1}{V}(\{(i,j)\})\right) \overset{\text{disjoint}}{=} \sum_j \mathbb{P}\left(\overset{-1}{V}(\{(i,j)\})\right) = \sum_j \mathbb{P}_V(\{(i,j)\}) = \sum_j p_{i,j}$$

On somme donc sur les lignes du tableau précédent :
$\mathbb{P}(X = i) = \sum_j p_{i,j}$ que l'on note $p_{i,\bullet}$

On vérifie que $\sum_i p_{i,\bullet} = \sum_i \sum_j p_{i,j} = 1$.

X	0	1	2	Σ
$p_{i\bullet}$	2/8	4/8	2/8	1

4.2.2 Loi marginale de Y

De même pour tout $j \in Y\langle\Omega\rangle$ on a :
$\mathbb{P}(Y = j) = \sum_i p_{i,j}$ que l'on note $p_{\bullet,j}$.

Et de même $\sum_j p_{\bullet,j} = \sum_j \sum_i p_{i,j} = 1$.

Y	0	1	2	3	Σ
$p_{\bullet j}$	1/8	3/8	3/8	1/8	1

Remarque :
La connaissance de la loi du couple entraîne donc la connaissance de la loi de chaque variable du couple. La réciproque n'est pas vraie en général.

4.3 Couple indépendant

En général, $p_{i,j} \neq p_{i,\bullet} \times p_{\bullet,j}$.

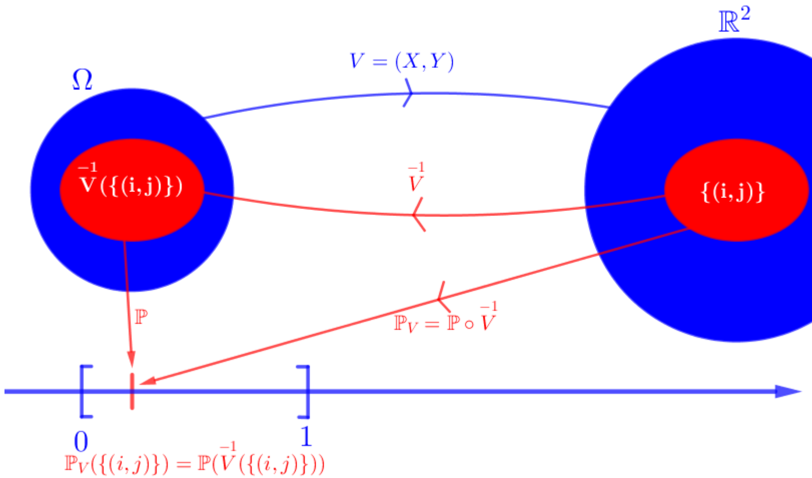
Définition 44

On dit que le couple $V = (X, Y)$ est **indépendant**, ou que les variables X et Y sont **indépendantes**, si et seulement si :

$$\forall (i,j) \in V\langle\Omega\rangle, \quad p_{i,j} = p_{i,\bullet} \times p_{\bullet,j}$$

Uniquement dans le cas d'un couple indépendant, la connaissance de la loi suivie par chaque variable aléatoire entraîne celle de la loi du couple (appelée **loi conjointe** de X et Y).

4.4 Cas d'un un évènement élémentaire



Pour un couple aléatoire $\quad V : \Omega \longrightarrow \mathbb{R}^2$

On a la composée :

$$\mathcal{B}_{\mathbb{R}^2} \xrightarrow{\overset{-1}{V}} \mathcal{A} \xrightarrow{\mathbb{P}} [0,1]$$

$$\{(i,j)\} \xmapsto{\overset{-1}{V}} \overset{-1}{V}(\{(i,j)\}) \xmapsto{\mathbb{P}} \mathbb{P}\left(\overset{-1}{V}(\{(i,j)\})\right) = \mathbb{P}_V(\{(i,j)\})$$

$$\begin{array}{ccccc} \mathcal{B}_{\mathbb{R}^2} & \xrightarrow{\overset{-1}{V}} & \mathcal{A} & \xrightarrow{\mathbb{P}} & [0,1] \\ \{(i,j)\} & \xmapsto{\overset{-1}{V}} & \overset{-1}{V}(\{(i,j)\}) & \xmapsto{\mathbb{P}} & \mathbb{P}\left(\overset{-1}{V}(\{(i,j)\})\right) = \mathbb{P}_V(\{(i,j)\}) \end{array}$$

Chapitre 5

Structure des lois de probabilité

5.1 Principaux paramètres

5.1.1 Paramètres de position

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et X une VAR.

Sous réserve d'existence des objets que l'on va définir (existence d'un maximum, d'une somme de série, d'une intégrale, ...)

Définition 45

- Un **mode** de X, noté $\widehat{x}$, est une valeur prise par la variable où se réalise la **probabilité maximale** pour une variable discrète, ou le maximum de la densité pour une VAR absolument continue.
- Une **médiane** de X, notée $\overline{\overline{x}}$ est un réel tel que $\mathbb{P}\left(X \leqslant \overline{\overline{x}}\right) \geqslant \dfrac{1}{2}$ **et** $\mathbb{P}\left(X \geqslant \overline{\overline{x}}\right) \geqslant \dfrac{1}{2}$.

Définition 46

- **L'espérance mathématique** (ou moyenne) de X est la moyenne arithmétique des valeurs prises par la variable aléatoire, pondérées par leur probabilité respective. Précisément :

– Si X est une VARD, $\mathbb{E}(X) = \overline{X} = \dfrac{\sum\limits_i x_i\, p_i}{\sum\limits_i p_i} = \sum\limits_i x_i\, p_i$ où $p_i = \mathbb{P}(X = x_i)$ et $\sum\limits_i p_i = 1$

– Si X est une VARC, $\mathbb{E}(X) = \overline{X} = \dfrac{\int_{\mathbb{R}} x\, f(x)\, \mathrm{d}x}{\int_{\mathbb{R}} f(x)\, \mathrm{d}x} = \displaystyle\int_{\mathbb{R}} x\, f(x)\, \mathrm{d}x$ car $\displaystyle\int_{\mathbb{R}} f(x)\, \mathrm{d}x = 1$.

5.1.2 Paramètres de dispersion

Définition 47

- **L'écart absolu moyen** est défini par $\mathbb{E}(|X - \overline{X}|)$.

– Si X est une VARD, on a $\mathbb{E}(|X - \overline{X}|) = \sum\limits_i |x_i - \overline{x}|\, p_i$

– Si X est VARC, on a $\mathbb{E}(|X - \overline{X}|) = \displaystyle\int_{\mathbb{R}} |x - \overline{x}|\, f(x)\, \mathrm{d}x$.

- **La variance** de X est définie par $\mathbb{V}(X) = \mathbb{E}\left((X - \overline{X})^2\right)$. Elle est positive.

– Si X est une VARD, $\mathbb{V}(X) = \sum\limits_i (x_i - \overline{x})^2\, p_i$.

– Si X est une VARC, $\displaystyle\int_{\mathbb{R}} (x - \overline{x})^2\, f(x)\, \mathrm{d}x$.

- **L'écart-type** de X est $\sigma(X) = \sqrt{\mathbb{V}(X)} \geqslant 0$.

Remarque : L'écart absolu moyen, bien qu'intéressant en terme de distance, ne possède pas de propriété algébrique contrairement à la **variance**.

5.1.3 Les moments

Définition 48

Sous réserve d'existence des sommes de séries ou des intégrales :

- Soit $k \in \mathbb{N}^*$, le **moment simple d'ordre** k est $m_k = \mathbb{E}(X^k)$.

  Pour une VARD, $m_k = \sum_i x_i^k \, p_i$. pour une VARC à densité, $m_k = \int_{\mathbb{R}} x^k \, f(x) \, \mathrm{d}x$.

- Soit $k \in \mathbb{N}^*$, le **moment centré** d'ordre k est $\mu_k = \mathbb{E}\left((X - \overline{X})^k\right)$.

  Pour une VARD, $\mu_k = \sum_i \left(x_i - \overline{X}\right)^k p_i$. pour une VARC à densité, $\mu_k = \int_{\mathbb{R}} \left(x - \overline{X}\right)^k f(x) \, \mathrm{d}x$.

Remarque : Pour l'interprétation, il est souhaitable d'associer à un paramètre de position, un paramètre de dispersion adequat. Par exemple (m_1, σ).

5.2 Propriétés relatives aux paramètres

Propriété 49 (de l'espérance)

1. Pour une VAR X constante égale à k, $k \in \mathbb{R}$, $\mathbb{E}(X) = k$.
2. Pour tout $(a, b) \in \mathbb{R}^2$, $\mathbb{E}(aX) = a\mathbb{E}(X)$, $\mathbb{E}(X + b) = \mathbb{E}(X) + b$ et donc $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

Démonstration de la propriété 2. On traite le cas discret, et on adapte la démonstration au cas continu.
Soit $(a, b) \in \mathbb{R}^2$, $\mathbb{E}(aX + b) = \sum_i (ax_i + b) \, p_i = a \sum_i x_i \, p_i + b \sum_i p_i = a\mathbb{E}(X) + b$ car $\sum_i p_i = 1$. □

Théorème 50 (König-Huygens)

Pour toute VAR X, sous réserve d'existence, $\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

Démonstration. On traite le cas discret, et on adapte au cas continu.

$$\begin{aligned}
\mathbb{V}(X) &= \sum_i (x_i - \overline{x})^2\, p_i \\
&= \sum_i (x_i^2\, p_i - 2\,\overline{x}\, x_i\, p_i + \overline{x}^2\, p_i) \\
&= \sum_i x_i^2\, p_i - 2\,\overline{x} \sum_i x_i\, p_i + \overline{x}^2 \sum_i p_i \\
&= m_2 - 2\,\mathbb{E}(X)\,\mathbb{E}(X) + \mathbb{E}(X)^2 \times 1 \\
&= \mathbb{E}(X^2) - (\mathbb{E}(X))^2
\end{aligned}$$

□

Propriété 51 (de la variance et de l'écart-type)

Soit X une VAR.

1. Si X est une VAR constante égale à k, $k \in \mathbb{R}$, $\mathbb{V}(X) = 0$.
2.
   - La variance est quadratique : $\forall a \in \mathbb{R}, \mathbb{V}(aX) = a^2\mathbb{V}(X)$
   - La variance est invariante par translation : $\forall b \in \mathbb{R}, \mathbb{V}(X + b) = \mathbb{V}(X)$
   - Ainsi $\forall (a, b) \in \mathbb{R}^2, \mathbb{V}(aX + b) = a^2\mathbb{V}(X)$
3. $\forall (a, b) \in \mathbb{R}^2, \sigma(aX + b) = |a| \cdot \sigma(X)$

Démonstration.

$$\begin{aligned}
\mathbb{V}(aX + b) &= \mathbb{E}\left((aX + b)^2\right) - (\mathbb{E}(aX + b))^2 \quad \text{(d'après König-Huygens)} \\
&= a^2\,\mathbb{E}(X^2) + 2ab\,\mathbb{E}(X) + b^2 - \left(a^2(\mathbb{E}(X))^2 + 2ab\,\mathbb{E}(X) + b^2\right) \\
&= a^2\left(\mathbb{E}(X^2) - (\mathbb{E}(X))^2\right) \\
&= a^2\,\mathbb{V}(X)
\end{aligned}$$

□

5.3 Propriétés relatives aux couples de variables aléatoires réelles

Définition 52

Sous réserve d'existence, l'espérance d'un couple est le couple des espérances : $\mathbb{E}\Big[(X,Y)\Big] = (\mathbb{E}(X), \mathbb{E}(Y))$.

5.3.1 Espérance de la somme de deux variables aléatoires réelles

Propriété 53

Pour toutes VAR X et Y, sous réserve d'existence, $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Démonstration. On démontre le cas discret.
$X\langle\Omega\rangle = \{x_i, i \in I\}$ et $Y\langle\Omega\rangle = \{y_j, j \in J\}$ sont finis ou dénombrables.
Notons $S = X+Y$, alors $S\langle\Omega\rangle = \{s_{i,j} = x_i + y_j, (i,j) \in I \times J\}$ et $p_{i,j} = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i \cap Y = y_j)$.
On a des systèmes complets d'évènements, donc d'après la formule des probabilités totales :

$$\sum_j p_{i,j} = \sum_j \mathbb{P}(X = x_i \cap Y = y_j) = \mathbb{P}\left(X = x_i \cap \left(\bigcup_{j \in J} Y = y_j\right)\right) = \mathbb{P}(X = x_i \cap \Omega) = p_i$$

De même, $\sum_i p_{i,j} = q_j$. Ainsi on a :

$$\begin{aligned}
\mathbb{E}(X+Y) &= \sum_{i,j} s_{i,j} p_{i,j} \\
&= \sum_i \sum_j (x_i + y_j) p_{i,j} \\
&= \sum_i \sum_j x_i p_{i,j} + \sum_i \sum_j y_j p_{i,j} \\
&= \sum_i x_i \left(\sum_j p_{i,j} \right) + \sum_j y_j \left(\sum_i p_{i,j} \right) \\
&= \sum_i x_i p_i + \sum_j y_j q_j \\
&= \mathbb{E}(X) + \mathbb{E}(Y)
\end{aligned}$$

□

Conséquence : $\mathbb{E}(X-\overline{X}) = \mathbb{E}(X) + \mathbb{E}(-\overline{X}) = \overline{X} - \overline{X} = 0$. On parle alors de **variable aléatoire centrée.**

5.3.2 Espérance du produit de deux variables aléatoires discrètes

Si $Z = XY$, alors $Z\langle\Omega\rangle = \{z_{i,j} = x_i y_j, (i,j) \in I \times J\}$ et $\mathbb{E}(XY) = \sum_{i,j} z_{i,j} p_{i,j} = \sum_i \sum_j x_i y_j p_{i,j}$.

Propriété 54

Si (X, Y) est **indépendant**, alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ (la réciproque est fausse)

Démonstration. Du fait de l'indépendance de X et Y, on a $p_{i,j} = p_i q_j$. Alors :

$$\mathbb{E}(XY) = \sum_i \sum_j x_i y_j p_i q_j = \left(\sum_i x_i p_i \right) \times \left(\sum_j y_j q_j \right) = \mathbb{E}(X)\mathbb{E}(Y)$$

5.3.3 Covariance d'un couple de variables aléatoires réelles discrètes

Définition 55

On définit la covariance du couple (X, Y) par $\mathbb{C}\text{ov}(X, Y) = \mathbb{E}\left[\left(X - \overline{X}\right) \cdot \left(Y - \overline{Y}\right)\right]$

Propriété 56

Pour toutes VAR X et Y, sous réserve d'existence, $\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Démonstration. On développe.

$$\begin{aligned}
\mathbb{C}\text{ov}(X, Y) &= \mathbb{E}\left[\left(X - \overline{X}\right) \cdot \left(Y - \overline{Y}\right)\right] \\
&= \mathbb{E}(XY - X\overline{Y} - \overline{X}Y + \overline{XY}) \\
&= \mathbb{E}(XY) - \mathbb{E}(X\overline{Y}) - \mathbb{E}(\overline{X}Y) + \mathbb{E}(\overline{XY}) \\
&= \mathbb{E}(XY) - \overline{Y}\mathbb{E}(X) - \overline{X}\mathbb{E}(Y) + \overline{XY} \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

□

Proposition 57

Si X et Y sont indépendantes, alors $\mathbb{C}\text{ov}(X, Y) = 0$. (la réciproque est fausse)

5.3.4 Coefficient de corrélation linéaire

Définition 58

On définit le **coefficient de corrélation linéaire** entre X et Y par :

$$\rho(X,Y) = \frac{\mathbb{C}\mathrm{ov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

C'est un réel, sans unité. S'il n'y a pas d'ambiguïté, on le note plus simplement ρ.

Proposition 59

Sous réserve d'existence, $\quad |\rho| \leqslant 1$

Démonstration. C'est l'inégalité de Cauchy-Schwarz :
Soient U et V deux VAR, et $\lambda \in \mathbb{R}$.
Alors $\mathbb{E}((U+\lambda V)^2) \geqslant 0$ donc le trinôme $T : \lambda \longmapsto \mathbb{E}((U+\lambda V)^2) = \lambda^2\mathbb{E}(V^2) + 2\lambda\mathbb{E}(UV) + \mathbb{E}(U^2)$ reste toujours positif. Son discriminant est donc négatif ou nul :
$\Delta = 4\left[\mathbb{E}(UV)\right]^2 - 4\mathbb{E}(U^2)\mathbb{E}(V^2) \leqslant 0$. Avec $U = X - \overline{X}$ et $V = Y - \overline{Y}$, cela donne :

$$\left[\mathbb{E}((X-\overline{X})(Y-\overline{Y}))\right]^2 \leqslant \mathbb{E}\left[(X-\overline{X})^2\right]\cdot\mathbb{E}\left[(Y-\overline{Y})^2\right]$$

Ainsi $\mathbb{C}\mathrm{ov}(X,Y)^2 \leqslant \mathbb{V}(X)\cdot\mathbb{V}(Y)$ et donc $\rho^2 \leqslant 1$, d'où $|\rho| \leqslant 1$. $\square$

Proposition 60 (totale corrélation)

$\rho = \pm 1 \Leftrightarrow X$ et Y sont liées par une relation affine $X = aY + b$.

Cas particuliers : Si $Y = X$, on obtient $\rho(X, X) = 1$. Si $Y = -X$, $\rho(X, -X) = -1$.

Démonstration. On s'intéresse au cas d'égalité dans l'inégalité de Cauchy-Schwarz. $\Delta = 0 \Leftrightarrow \exists \lambda_0 \in \mathbb{R} : \mathbb{E}\left((U + \lambda_0 V)^2\right) = 0$.
Or $(U + \lambda_0 V)^2$ est positive, son espérance est nulle si et seulement si les évènements où elle est non nulle sont de probabilité nulle (on dit qu'elle est presque sûrement nulle).
On va considérer la variable aléatoire nulle $U + \lambda_0 V = (X - \overline{X}) + \lambda_0(Y - \overline{Y}) = 0$.
Cela signifie que X et Y sont liées par une relation affine : $X = aY + b$.
Ainsi, si $X = aY + b$, alors

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(aY^2 + bY) - \mathbb{E}(aY + b)\mathbb{E}(Y) = a\mathbb{E}(Y^2) + b\mathbb{E}(Y) - a\mathbb{E}(Y)^2 - b\mathbb{E}(Y) = a\mathbb{V}(Y)$$

Et $\mathbb{V}(X) = \mathbb{V}(aY + b) = a^2\mathbb{V}(Y)$, donc $\rho = \dfrac{\mathbb{C}\text{ov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} = \dfrac{a}{|a|} = \pm 1$. □

5.3.5 Variance d'une somme de deux variables aléatoires réelles

Proposition 61

Pour toutes VAR X et Y, sous réserve d'existence,

1. $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\,\mathbb{C}\text{ov}(X, Y)$
2. En particulier, si X et Y sont deux VARD indépendantes, on obtient $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y)$. (la réciproque est fausse)

Démonstration. □

$$\begin{aligned}\mathbb{V}(X + Y) &= \mathbb{E}\left[(X + Y)^2\right] - \mathbb{E}(X + Y)^2 \\ &= \mathbb{E}(X^2 + 2XY + Y^2) - [\mathbb{E}(X) + \mathbb{E}(Y)]^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)^2 \\ &= \mathbb{V}(X) + \mathbb{V}(Y) + 2\mathbb{C}\text{ov}(X, Y)\end{aligned}$$

5.3.6 Exemple

On jette un dé non pipé. On a $\Omega = \{1,2,3,4,5,6\} = \{\omega_i, i \in \{1,\dots,6\}\}$ avec $\mathbb{P}(\omega_i) = \dfrac{1}{6}$.

$$X : \begin{cases} \Omega & \longrightarrow & \mathbb{R} \\ \text{Pair} & \longmapsto & 2 \\ \text{Impair} & \longmapsto & 1 \end{cases} \quad X\langle\Omega\rangle = \{1,2\} \quad ; \quad Y : \begin{cases} \Omega & \longrightarrow & \mathbb{R} \\ 6 & \longmapsto & 3 \\ 2 & \longmapsto & 3 \\ 1,3,4,5 & \longmapsto & 0 \end{cases} \text{ et } Y\langle\Omega\rangle = \{0;3\}.$$

Calculons les espérances et variances de ces VARD.

$Y=j$ / $X=i$	$j=0$	$j=3$	$p_{i\bullet}$
$i=1$	$1/2$	0	$1/2$
$i=2$	$1/6$	$1/3$	$1/2$
$p_{\bullet j}$	$2/3$	$1/3$	1

$$\mathbb{E}(X) = 1 \times \frac{1}{2} + 2 \times \frac{1}{2} = \frac{3}{2}$$

$$\mathbb{E}(Y) = 0 \times \frac{2}{3} + 3 \times \frac{1}{3} = 1$$

$$\mathbb{E}(XY) = 1\times 0\times \frac{1}{2} + 2\times 0\times \frac{1}{6} + 1\times 3\times 0 + 2\times 3\times \frac{1}{3} = 2$$

$$\mathbb{V}(X) = 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{2} - \frac{9}{4} = \frac{1}{4}$$

$$\mathbb{V}(Y) = 0 + 3^2 \times \frac{1}{3} - 1 = 2$$

$$\mathbb{C}\text{ov}(X,Y) = 2 - \frac{3}{2} \times 1 = \frac{1}{2}$$

$$\rho(X,Y) = \frac{\sqrt{2}}{2}.$$

5.4 Inégalité de Bienaymé-Tchebychev

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et X une VAR définie sur Ω dont on ne connaît que $m = \mathbb{E}(X) \in \mathbb{R}$ et $\sigma = \sqrt{\mathbb{V}(X)} \in \mathbb{R}_+^*$.
On s'intéresse à la probabilité d'un intervalle centré en m : $\mathbb{P}(X \in]\, m-\varepsilon\, ,\, m+\varepsilon\, [)$.
ε est une « proportion » de l'écart-type σ, nous noterons donc $\varepsilon = t\sigma$ avec $t > 0$.

Théorème 62 (Inégalité de Bienaymé-Tchebychev)

Soit $\Omega_1 = \overset{-1}{X}(]-\infty, m-\varepsilon\,[\cup]\, m+\varepsilon\, ,\, +\infty\,[) = \{\omega \in \Omega, |\, X(\omega) - m\,| > t\sigma\}$.

Alors, $\mathbb{P}(\Omega_1) = \mathbb{P}(|X-m| > t\sigma) \leqslant \dfrac{1}{t^2}$ ou $\mathbb{P}(|X-m| \leqslant t\sigma) \geqslant 1 - \dfrac{1}{t^2}$.

Autrement dit, $\mathbb{P}(|X-m| > \varepsilon) \leqslant \dfrac{\sigma^2}{\varepsilon^2}$ ou $\mathbb{P}(|X-m| \leqslant \varepsilon) \geqslant 1 - \dfrac{\sigma^2}{\varepsilon^2}$

Démonstration. $\Omega_1 \in \mathcal{A}$ et $\Omega = \Omega_1 \cup \overline{\Omega_1}$.

- Pour une VARD, $\sigma^2 = \displaystyle\sum_i (x_i - \overline{x})^2 p_i$.
- Pour une VARC, $\sigma^2 = \displaystyle\int_{\mathbb{R}} (x-\overline{x})^2 f(x)\mathrm{d}x$.

Dans les deux cas, on note $\mathbb{V}(X) = \sigma^2 = \mathbb{E}[(X-\overline{X})^2] = \displaystyle\int_\Omega (X-\overline{X})^2 d\mathbb{P}_\Omega$ (notation de Stieltjes)

La fonction est positive : $\sigma^2 \geqslant \displaystyle\int_{\Omega_1} (X-\overline{X})^2 d\mathbb{P}_\Omega$.

De plus, sur Ω_1, on a $(X-\overline{X})^2 = |X-m|^2 \geqslant \varepsilon^2 = t^2\sigma^2$ donc $\sigma^2 \geqslant \displaystyle\int_{\Omega_1} t^2\sigma^2 d\mathbb{P}_\Omega$.

Ainsi : $1 \geqslant \displaystyle\int_{\Omega_1} t^2 d\mathbb{P}_\Omega$ et $\displaystyle\int_{\Omega_1} d\mathbb{P}_\Omega = \mathbb{P}(\Omega_1) \leqslant \dfrac{1}{t^2}$, d'où le résultat. □

Remarques :

- Pour $t \leqslant 1$, l'inégalité ne nous donne rien de nouveau.
- Pour $t > 1$, prenons par exemple $t = 2$, alors $\mathbb{P}(|X - m| \leqslant 2\sigma) \geqslant 1 - \dfrac{1}{4}$.

  Pour $t = 3$, on obtient $\mathbb{P}(|X - m| \leqslant 3\sigma) \geqslant 1 - \dfrac{1}{9}$.

  On peut affirmer que, pour toute probabilité et toute VAR X, 75% (resp. 90%) des valeurs prises par X sont à une distance inférieure à 2σ (resp. 3σ) de la moyenne.
- L'inégalité est très générale dans son domaine d'application mais peu performante. Il est préférable, si possible, d'utiliser la fonction de répartition appropriée pour calculer la probabilité attachée à un intervalle.

  Par exemple si X est une VARC de fonction de répartition F :

$$\mathbb{P}(|X - m| \leqslant 2\sigma) = \mathbb{P}(m - 2\sigma \leqslant X \leqslant m + 2\sigma) = F(m + 2\sigma) - F(m - 2\sigma)$$

- Cependant, l'application du théorème est fondamentale pour la loi faible des grands nombres qui constitue le lien historique entre Statistiques et Probabilités.

Chapitre 6

Fonctions génératrices, produit de convolution

6.1 Première fonction génératrice

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une VAR, définie sur Ω.

6.1.1 Définition

Définition 63

On appelle **première fonction génératrice de** X, notée G_X, la fonction définie par :

$$G_X : \left\{ \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ u & \longmapsto & \mathbb{E}(u^X) \end{array} \right.$$

On parle aussi de **fonction génératrice des probabilités**. On a toujours $\boxed{G_X(1) = \mathbb{E}(1) = 1}$ et $\forall u \in [-1, 1],\ |G_X(u)| \leqslant 1$.

Dans ce cours, nous l'utiliserons principalement dans le cas où X est une VARD à valeurs dans $\mathbb{N}$.

Mais, selon les cas, sous réserve d'existence, $G_X(u)$ est :

- Dans le cas discret avec $X\langle\Omega\rangle \subset \mathbb{N}$, $G_X(u) = \mathbb{E}\big(u^X\big) = \sum_{n\geqslant 0} p_n\, u^n$ avec $p_n = \mathbb{P}(X = n)$.
  - Si $X\langle\Omega\rangle$ est **fini**, G_X est un **polynôme**.
  - Si $X\langle\Omega\rangle$ est **dénombrable** G_X est la somme totale d'une **série entière**. De plus $\sum_{n\in\mathbb{N}} p_n = 1$ donc elle est absolument convergente en $u = 1$, et son rayon de convergence vérifie $R \geqslant 1$.
- Dans le cas continu, $G_X(u) = \mathbb{E}\big(u^X\big) = \int_{\mathbb{R}} u^x\, f(x)\mathrm{d}x$ où f est une densité de X. En général, on préférera utiliser la deuxième fonction génératrice.

Proposition 64

Pour qu'une fonction soit génératrice d'une loi de probabilité, il faut et il suffit qu'elle soit développable en série entière, que les coefficients du développement soient positifs ou nuls, et que leur somme totale soit égale à 1.

Théorème 65

La première fonction génératrice d'une VARD caractérise sa loi de probabilité.

Démonstration. Dans le cas discret, c'est l'unicité des coefficients d'une série entière qui assure que deux variables aléatoires réelles discrètes X et Y, définies sur un même Ω, et ayant la même fonction génératrice vérifient $\forall n \in \mathbb{N}, \mathbb{P}(X = n) = \mathbb{P}(Y = n)$ donc elles ont même loi. $\square$

6.1.2 Calcul des moments factoriels

Proposition 66

Sous réserve d'existence :

- $\forall k \geqslant 1, {G_X}^{(k)}(1) = \mathbb{E}(X(X-1)(X-2)\cdots(X-k+1)) = \mu_{[k]}$ le **moment factoriel d'ordre** k.
- $\mathbb{E}(X) = {G_X}'(1) = \mu_{[1]}$
- $\mathbb{V}(X) = {G_X}'(1) + {G_X}''(1) - ({G_X}'(1))^2 = \mu_{[1]} + \mu_{[2]} - \mu_{[1]}^2$.

Démonstration. Dans le cas discret :

- Sur $]-R, R[$, on a ${G_X}'(u) = \sum_{n \geqslant 1} n\, p_n\, u^{n-1}$ et sous condition que la série dérivée converge en $u=1$, on a ${G_X}'(1) = \sum_{n \in \mathbb{N}} np_n = \mathbb{E}(X)$.

  De même, ${G_X}''(u) = \sum_{n \in \mathbb{N}} n(n-1)\, u^{n-2}$ et sous condition d'existence, ${G_X}''(1) = \sum_{n \geqslant 2} n(n-1)p_n = \mathbb{E}(X(X-1)) = \mu_{[2]}$.

  Par récurrence évidente, on généralise à l'ordre k pour obtenir :
  ${G_X}^{(k)}(u) = \sum_{n \geqslant k} n(n-1)\cdots(n-k+1)\, p_n\, u^{n-k}$ d'où ${G_X}^{(k)}(1) = \mathbb{E}(X(X-1)\cdots(X-k+1)) = \mu_{[k]}$.
- D'après ce qui précède, $\mathbb{E}(X) = {G_X}'(1) = \mu_{[1]}$
- $\mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}(X) = {G_X}''(1)$ donc $\mathbb{E}(X^2) = \mathbb{E}(X) + {G_X}''(1) = \mu_{[1]} + \mu_{[2]}$
  et ainsi $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mu_{[1]} + \mu_{[2]} - \mu_{[1]}^2$

On retrouve la formule annoncée.

Dans le cas continu, on dérive l'intégrale à paramètre $G_X(u) = \mathbb{E}\big(u^X\big) = \int_{\mathbb{R}} u^x\, f(x) \mathrm{d}x$ où f est une densité de X,

d'où ${G_X}'(u) = \int_{\mathbb{R}} x\, u^{x-1}\, f(x)\mathrm{d}x$ et ${G_X}'(1) = \int_{\mathbb{R}} x\, f(x)\mathrm{d}x = \mathbb{E}(X)$, de même pour les dérivées successives. □

6.1.3 Développement au voisinage de 1 et application

Proposition 67

Dans le cas discret, pour v au voisinage de 0^-, on a :

$$G_X(1+v) = \sum_{n\geqslant 0} \frac{v^n}{n\,!}\,\mu_{[n]}$$

Démonstration. On pose $u = 1+v$ et on développe en série entière pour u au voisinage de 1^- (donc pour v au voisinage de 0^-) :

$$G_X(u) = G_X(1+v) = G_X(1) + \frac{v}{1!}\,{G_X}'(1) + \frac{v^2}{2!}\,{G_X}''(1) + \cdots + \frac{v^n}{n!}\,{G_X}^{(n)}(1) + \cdots = \sum_{n\geqslant 0} \frac{v^n}{n\,!}\,{G_X}^{(n)}(1)$$

Exemple d'application : Soit X une variable aléatoire discrète telle que pour tout $k \in \mathbb{N}, \mu_{[k]} = m^k$ avec $m > 0$.
Déterminons la loi de X.

Pour les v pour lesquels cela a un sens (ce qui est assuré pour $v \in [-1, 0]$), on a :

$$\begin{aligned} G_X(1+v) &= \sum_{n\geqslant 0} \frac{v^n}{n!}\mu_{[n]} && = \sum_{n\geqslant 0} \frac{v^n}{n!} m^n && = \exp(mv) \\ &= \exp\big(m(u-1)\big) && = \sum_{n\geqslant 0} \frac{\mathrm{e}^{-m} m^n}{n!} u^n && = \sum_{n\geqslant 0} \mathbb{P}(X=n) u^n \end{aligned}$$

Par unicité des coefficients, on obtient $\forall n \in \mathbb{N},\ \ \mathbb{P}(X=n) = \dfrac{\mathrm{e}^{-m} m^n}{n!}$ donc X suit une loi de Poisson de paramètre m (voir chapitre 9).

Remarquons que $G_X(u) = \mathrm{e}^{-m}\mathrm{e}^{mu}$ a un rayon de convergence infini.

6.2 Deuxième fonction génératrice

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit X une VAR, définie sur Ω.

Définition 68

On appelle **deuxième fonction génératrice de** X, notée g_X, quand elle existe, la fonction : $g_X : \left\{ \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ t & \longmapsto & \mathbb{E}\big(\mathrm{e}^{tX}\big) \end{array} \right.$

On parle aussi de **fonction génératrice des moments**. On a toujours $\boxed{g_X(0) = 1}$.

- Dans le cas discret, $g_X(t) = \mathbb{E}\big(\mathrm{e}^{tX}\big) = \displaystyle\sum_{n \geqslant 0} p_n \, \mathrm{e}^{t\,n}$ avec $p_n = \mathbb{P}(X = n)$.
- Dans le cas continu, $g_X(t) = \mathbb{E}\big(\mathrm{e}^{tX}\big) = \displaystyle\int_{\mathbb{R}} \mathrm{e}^{t\,x} f(x) \,\mathrm{d}x$ où f est une densité de X.

Proposition 69

- Cas discret : si X est une VARD, alors (sous réserve d'existence des dérivées) :

$$\forall k \in \mathbb{N}^*, \quad {g_X}^{(k)}(t) = \sum_{n \geqslant 0} n^k \, p_n \, \mathrm{e}^{tn} \quad \text{et donc} \quad {g_X}^{(k)}(0) = \sum_{n \geqslant 0} n^k \, p_n = \mathbb{E}(X^k) = m_k$$

- Cas continu : si X est une VARC de densité f, alors :

$$\forall k \in \mathbb{N}^*, \quad {g_X}^{(k)}(t) = \int_{\mathbb{R}} x^k \, \mathrm{e}^{tx} \, f(x) \,\mathrm{d}x \quad \text{et donc} \quad {g_X}^{(k)}(0) = \int_{\mathbb{R}} x^k \, f(x) \,\mathrm{d}x = \mathbb{E}(X^k) = m_k$$

- Dans tous les cas $\mathbb{E}(X) = m_1 = {g_X}'(0)$ et $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = m_2 - m_1^2 = {g_X}''(0) - {g_X}'(0)^2$

Cette proposition justifie l'appellation de **génératrice des moments (simples)**.

6.3 Produit de convolution

6.3.1 Définition

Considérons deux variables aléatoires réelles **indépendantes** X_1 et X_2, qui suivent deux lois de probabilité $X_1 \sim \mathcal{P}_1$ et $X_2 \sim \mathcal{P}_2$.

Définition 70

Le **produit de convolution** des deux lois de probabilité $\mathcal{P}_1$ et $\mathcal{P}_2$, noté $\mathcal{P}_1 * \mathcal{P}_2$, est la loi de probabilité associée à $X_1 + X_2$ (somme indépendante). Autrement dit :

$$X_1 + X_2 \sim \mathcal{P}_1 * \mathcal{P}_2$$

6.3.2 Lien avec les fonctions génératrices

Remarque : Pour l'indépendance des VARC, voir chapitre 17.

Proposition 71 (première fonction génératrice)

Si X_1 et X_2 sont deux variables aléatoires réelles **indépendantes** : $\quad G_{X_1+X_2}(u) = G_{X_1}(u) \times G_{X_2}(u)$
là où elles sont définies simultanément.

Démonstration. $G_{X_1+X_2}(u) = \mathbb{E}\big(u^{X_1+X_2}\big) = \mathbb{E}\left(u^{X_1} \times u^{X_2}\right) \overset{\text{indépendantes}}{=} \mathbb{E}\big(u^{X_1}\big) \times \mathbb{E}\big(u^{X_2}\big) = G_{X_1}(u) \times G_{X_2}(u).$ □

Proposition 72 (deuxième fonction génératrice)

Si X_1 et X_2 sont deux variables aléatoires réelles **indépendantes** : $\quad g_{X_1+X_2}(t) = g_{X_1}(t) \times g_{X_2}(t)$
quand elles sont définies.

Démonstration. $g_{X_1+X_2}(t) = \mathbb{E}\left(\mathrm{e}^{t(X_1+X_2)}\right) = \mathbb{E}\left(\mathrm{e}^{tX_1} \times \mathrm{e}^{tX_2}\right) \overset{\text{indépendantes}}{=} \mathbb{E}\left(\mathrm{e}^{tX_1}\right) \times \mathbb{E}\left(\mathrm{e}^{tX_2}\right) = g_{X_1}(t) \times g_{X_2}(t).$ □

Deuxième partie

Lois de probabilité discrètes

Chapitre 7

Lois de Bernoulli

7.1 Loi alternative

7.1.1 Cadre de la loi

Elle régit les phénomènes à caractère dichotomique (pile/face, pair/impair, etc). Ω est un ensemble d'éventualités s'excluant mutuellement. Nous pouvons traduire cette dichotomie suivant la partition : $A + \overline{A} = \Omega$.
Considérons $\mathcal{B} = \{\emptyset, A, \overline{A}, \Omega\}$ et notons $p = \mathbb{P}(A) \in [0, 1]$ et $q = 1 - p = \mathbb{P}(\overline{A}) \in [0, 1]$. Nous obtenons ainsi $(\Omega, \mathcal{B}, \mathbb{P})$ comme espace probabilisé.

On utilise la **loi alternative** pour une VARD X ne **prenant que deux valeurs distinctes** :
$X\langle\Omega\rangle = \{B_1, B_2\}$ avec $B_1 < B_2$ et on choisit $A = \overset{-1}{X}(B_2)$ et donc $\overline{A} = \overset{-1}{X}(B_1)$.

7.1.2 Distribution ou loi de probabilité

x_i	B_1	B_2	Σ
$\mathbb{P}(X = x_i)$	q	p	1

Proposition 73

Si X suit une loi alternative avec $X\langle\Omega\rangle = \{B_1, B_2\}$, alors on a :
- $\mathbb{E}(X) = qB_1 + pB_2$
- $\mathbb{V}(X) = pq(B_2 - B_1)^2$ et $\sigma(X) = \sqrt{pq}|B_2 - B_1|$

Démonstration.
- $\mathbb{E}(X) = \sum_{i=1}^{2} x_i p_i = qB_1 + pB_2.$
- $$\begin{aligned}\mathbb{V}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= qB_1^2 + pB_2^2 - (qB_1 + pB_2)^2 \\ &= q(1-q)B_1^2 + p(1-p)B_2^2 - 2pqB_1B_2 \\ &= pq(B_1^2 + B_2^2 - 2B_1B_2) \\ &= pq(B_1 - B_2)^2\end{aligned}$$ □

7.1.3 Nature des informations fournies par les paramètres

- L'écart $(B_2 - B_1)$ est un facteur de dispersion.
- $\mathbb{E}(X) = (1-p)B_1 + pB_2$ donc $B_1 \leqslant \mathbb{E}(X) \leqslant B_2$.
- $p(1-p)$ admet un maximum en $p = \frac{1}{2}$ (équidistribution). Dans ce cas, $\mathbb{V}_{max} = \frac{1}{4}(B_2 - B_1)^2$.
- **Dans le cas où $\boldsymbol{X\langle\Omega\rangle}$** $\subset \mathbb{N}$, on peut définir la première fonction génératrice polynomiale : $\boxed{\forall u \in \mathbb{R}, \quad G_X(u) = \mathbb{E}(u^X) = q\,u^{B_1} + p\,u^{B_2}}$ et retrouver ainsi les principaux paramètres calculés supra (par dérivation).

7.2 Loi de Bernoulli ou loi alternative simple

7.2.1 Cas particulier : loi de Bernoulli

Définition 74

La **loi de Bernoulli** est un cas particulier de loi alternative avec $B_1 = 0, B_2 = 1$.

Notations : La loi de Bernoulli de paramètre p se note $\mathcal{B}(1, p)$. Si X suit cette loi, on écrira $X \sim \mathcal{B}(1, p)$.

A est l'évènement ensemble des « succès » et $\overline{A}$ est l'évènement ensemble des « échecs ».
$X\langle\Omega\rangle = \{0, 1\}$ et si $\omega \in A$, alors $X(\omega) = 1$ et si $\omega \notin A$, alors $X(\omega) = 0$. (On rappelle que ω désigne une éventualité)
Ainsi, $X = \varphi_A$ la **fonction caractéristique (au sens ensembliste)** de A, on dit que X est un **indicateur de probabilité**.

7.2.2 Loi et principaux paramètres

x_i	0	1	Σ
$\mathbb{P}(X = x_i)$	q	p	1

Proposition 75

Si $X \sim \mathcal{B}(1, p)$, alors :
- $\mathbb{E}(X) = p$
- $\mathbb{V}(X) = pq$ et $\sigma(X) = \sqrt{pq}$

7.2.3 Première fonction génératrice

Proposition 76

Si $X \sim \mathcal{B}(1,p)$, alors $\forall u \in \mathbb{R}$, $G_X(u) = \mathbb{E}(u^X) = q + p\,u$.

On peut retrouver les principaux paramètres (par dérivation en 1).

Chapitre 8

Loi binomiale et aléa géométrique

8.1 Loi binomiale

8.1.1 Qualification de la loi binomiale

Dans une production de pièces, avec une proportion $p \in [\,0, 1\,]$ de pièces défectueuses, on prélève un échantillon de 3 pièces par tirage bernoullien (avec remise assurant l'indépendance des tirages).
On désigne par X la VARD prenant pour **valeur le nombre de pièces défectueuses** de l'échantillon. On note $q = 1 - p$.

Distribution ou loi de probabilité :

Avec $X\langle\Omega\rangle = \{\,0, 1, 2, 3\,\}$ on passe en revue tous les cas : D pour défectueux, $\overline{D}$ pour non défectueux.

1^{er} tirage	2^{eme} tirage	3^{eme} tirage	x_i	probabilité
$\overline{D}$	$\overline{D}$	$\overline{D}$	0	q^3
D	$\overline{D}$	$\overline{D}$	1	pq^2
$\overline{D}$	D	$\overline{D}$	1	pq^2
D	D	$\overline{D}$	2	p^2q
$\overline{D}$	$\overline{D}$	D	1	pq^2
D	$\overline{D}$	D	2	p^2q
$\overline{D}$	D	D	2	p^2q
D	D	D	3	p^3

En regroupant suivant les valeurs x_i, on obtient la loi de probabilité de la VARD X :

x_i	0	1	2	3	Σ
$\mathbb{P}(X = x_i)$	q^3	$3q^2p$	$3qp^2$	p^3	$(p+q)^3 = 1$

8.1.2 Cadre et énoncé de la loi binomiale

Il s'agit d'une succession d'épreuves :

- Répétées n fois, où $n \in \mathbb{N}^*$,
- Identiques,
- Indépendantes et à caractère dichotomique.

On s'intéresse ici au **nombre** k **de succès parmi ces** n **épreuves**.

Soit $\Omega_1 = \{E, S\}$ pour « échec » et « succès », avec $p = \mathbb{P}(S) \in [\,0, 1\,]$ et $q = 1 - p = \mathbb{P}(E)$. Soit $\Omega = \Omega_1^n$, alors $\mathcal{P}(\Omega) = \mathcal{P}(\Omega_1^n)$ muni de la probabilité produit Π. Ainsi $(\Omega, \mathcal{P}(\Omega), \Pi)$ est un espace probabilisé.
On peut définir une VARD X qui compte le nombre de succès parmi ces n épreuves.

Définition 77

La VARD X suit la **loi binomiale de paramètres** n **et** p : $X \sim \mathcal{B}(n,p)$ avec $\begin{cases} n = \text{nombre d'épreuves,} \\ p = \text{probabilité d'un succès} \end{cases}$ avec $X\langle\Omega\rangle = [\![0,n]\!]$. Cette loi est définie par :

$$\forall k \in [\![0,n]\!], \quad \mathbb{P}(X=k) = \binom{n}{k} p^k q^{n-k} \ \text{ noté } p_k$$

avec k succès et $n-k$ échecs sur n épreuves.

Les p_k sont positifs, et on a $\displaystyle\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+q)^n = 1^n = 1$ donc les p_k définissent une loi de probabilité.

Remarque : Les p_k sont les différents termes du développement de $(p+q)^n$ par la formule du binôme de Newton, d'où l'appellation de loi binomiale.

8.2 Loi binomiale et loi de Bernoulli

Pour tout $i \in [\![1,n]\!]$, $X_i \sim \mathcal{B}_i(1,p)$ et X compte le nombre k de succès parmi les n épreuves indépendantes de Bernoulli (donc $X \sim \mathcal{B}(n,p)$). Alors $X = \displaystyle\sum_{i=1}^{n} X_i$ (somme **indépendante**) donc, par construction, la loi de X est le produit de convolution des lois des X_i :

$$\mathcal{B}(n,p) = \mathcal{B}_1(1,p) * \mathcal{B}_2(1,p) * \cdots * \mathcal{B}_n(1,p)$$

8.2.1 Principaux paramètres

Proposition 78

Si $X \sim \mathcal{B}(n,p)$ alors :
- $\mathbb{E}(X) = np$
- $\mathbb{V}(X) = npq$ et $\sigma(X) = \sqrt{npq}$

Démonstration.
- $\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np.$
- $\mathbb{V}(X) = \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{indépendance}}{=} \sum_{i=1}^{n} \mathbb{V}(X_i) = npq$ $\square$

8.2.2 Première fonction génératrice

Proposition 79

Si $X \sim \mathcal{B}(n,p)$, alors par convolution :

$$\forall u \in \mathbb{R}, G_X(u) = \mathbb{E}(u^X) = \prod_{i=1}^{n} G_{X_i}(u) = (p\,u + q)^n$$

On peut ainsi retrouver les principaux paramètres.

8.2.3 Convolution de lois binomiales

Supposons $X_1 \sim \mathcal{B}(n,p)$ et $X_2 \sim \mathcal{B}(m,p)$ deux variables aléatoires réelles discrètes indépendantes. Alors $X_1 + X_2 \sim \mathcal{B}(n,p) * \mathcal{B}(m,p)$ par définition du produit de convolution.

Proposition 80

Si $X_1 \sim \mathcal{B}(n,p)$ et $X_2 \sim \mathcal{B}(m,p)$ alors :

1. $\mathcal{B}(n,p) * \mathcal{B}(m,p) = \mathcal{B}(n+m,p)$
2. $X_1 + X_2 \sim \mathcal{B}(n+m,p)$ (une somme de 2 compteurs est un compteur)

Démonstration. On utilise les fonctions génératrices :
$G_{X_1+X_2}(u) = G_{X_1}(u) \times G_{X_2}(u) = (p\,u+q)^n \times (p\,u+q)^m = (p\,u+q)^{m+n}$ qui est la fonction génératrice d'une $\mathcal{B}(n+m,p)$ donc $X_1 + X_2 \sim \mathcal{B}(n+m,p)$ puisque la fonction génératrice caractérise la loi. $\square$

8.2.4 La fréquence binomiale

Définition 81

Si $X \sim \mathcal{B}(n,p)$ avec $p \in [0,1]$, on appelle **fréquence binomiale** la VARD $F = \dfrac{X}{n}\cdot$

Alors $F\langle\Omega\rangle = \left\{0\,,\frac{1}{n}\,,\frac{2}{n}\,,\,\dots,\frac{k}{n}\,,\,\cdots,1\right\}$ et $F\langle\Omega\rangle \subset \mathbb{Q}$.

Proposition 82 (principaux paramètres de la fréquence binomiale)

Si $X \sim \mathcal{B}(n,p)$ et $F = \dfrac{X}{n}$ alors : $\mathbb{E}(F) = p$, $\mathbb{V}(F) = \dfrac{p\,q}{n}$ *et* $\sigma(F) = \sqrt{\dfrac{p\,q}{n}}$

Démonstration. $\forall k \in [\![0,n]\!]$, $\mathbb{P}\big(F = \dfrac{k}{n}\big) = \mathbb{P}(X = k)$ donc $\mathbb{E}(F) = \mathbb{E}\left(\dfrac{X}{n}\right) = \dfrac{1}{n}\mathbb{E}(X) = p$ et $\mathbb{V}(F) = \mathbb{V}\left(\dfrac{X}{n}\right) = \dfrac{1}{n^2}\mathbb{V}(X) = \dfrac{npq}{n^2} = \dfrac{pq}{n}\cdot$ □

8.3 Généralisation

8.3.1 Distribution multinomiale

Pour $s \geqslant 2$, on considère $(A_1, A_2, \dots, A_s)$ un système complet d'évènements, ainsi $\Omega = \displaystyle\sum_{i=1}^{s} A_i$.

On note $p_i = \mathbb{P}(A_i), i \in \{1, \dots, s\}$ les probabilités de chacun de ces évènements.
On effectue n épreuves répétées indépendantes à caractère multinomial. La probabilité pour que A_1 se produise k_1 fois, A_2 se produise k_2 fois, ..., et A_s se produise k_s fois avec $k_1 + k_2 + \cdots + k_n = n$ est :

$$\mathbb{P}(X_1 = k_1, X_2 = k_2, \dots, X_s = k_s) = \frac{n!}{k_1!k_2!\cdots k_s!}p_1^{k_1}p_2^{k_2}\cdots p_s^{k_s}$$

Note : ce n'est pas une loi de probabilité d'une VARD (mais d'un vecteur aléatoire).

8.3.2 Exemple d'application

Les ampoules colorées produites par une usine sont rouges, bleues et vertes dans les proportions $50\%, 30\%$ et 20%. A partir d'un échantillon de 5 ampoules, calculons la probabilité d'avoir 2 rouges, 2 bleues et 1 verte :

$$\frac{5!}{2!2!1!}0,5^2 \cdot 0,3^2 \cdot 0,2 = 0,027$$

8.4 Aléa géométrique ou loi de Pascal

8.4.1 Cadre et énoncé de la loi géométrique

Il s'agit d'une succession d'épreuves :
- Répétées n fois, avec $n \in \mathbb{N}^*$
- Identiques
- Indépendantes à caractère dichotomique, avec une proportion de succès $p \in]\,0, 1\,[$.

On s'intéresse ici au **nombre k d'épreuves jusqu'au premier succès (succès compris dans le compte).**

Par exemple, pour l'évènement succession $E\ E\ E\ E\ S$ on a $X = 5$.

On répète l'épreuve **au moins une fois**, éventuellement **une infinité de fois**, et on considère $\Omega = \{E, S\}^{\mathbb{N}^*}$.

Définition 83

Soit X une VARD suivant la **loi géométrique (ou loi de Pascal) de paramètre** $p : X \sim \mathcal{G}(p)$ avec $p \in]\,0, 1\,[$. Alors $X\langle\Omega\rangle = \mathbb{N}^*$. Cette loi est définie par :

$$\forall k \in \mathbb{N}^*, \mathbb{P}(X = k) = q^{k-1}p$$

$\forall k \in \mathbb{N}^*$, $\mathbb{P}(X = k) \geq 0$, et $\displaystyle\sum_{k \in \mathbb{N}^*} \mathbb{P}(X = k) = p \left(\sum_{k \in \mathbb{N}^*} q^{k-1} \right) = \frac{p}{1-q} = 1$. Cela définit bien une loi de probabilité.

8.4.2 Première fonction génératrice, principaux paramètres

Proposition 84

1. $G_X(u) = \dfrac{pu}{1-qu}$ et le rayon de convergence est $R = \dfrac{1}{q}$ avec $\dfrac{1}{q} > 1$.
2. $\mathbb{E}(X) = \dfrac{1}{p}$, $\mathbb{V}(X) = \dfrac{q}{p^2}$ et $\sigma(X) = \dfrac{\sqrt{q}}{p}$

Démonstration.

1. $G_X(u) = \mathbb{E}(u^X) = \displaystyle\sum_{k\in\mathbb{N}^*} u^k p_k = \sum_{k\in\mathbb{N}^*} u^k q^{k-1} p = pu \sum_{s\in\mathbb{N}} u^s q^s = \frac{pu}{1-qu}$ pour $|uq| < 1$ c'est-à-dire $|u| < \dfrac{1}{q}$ donc $R = \dfrac{1}{q}$.
2. Pour tout $u \in \left]-\dfrac{1}{q}, \dfrac{1}{q}\right[$:

$$G'_X(u) = \frac{p(1-qu)+pqu}{(1-qu)^2} = \frac{p}{(1-qu)^2} \text{ donc } G'_X(1) = \frac{p}{p^2} = \frac{1}{p} = \mu_{[1]}$$

donc $\mathbb{E}(X) = \mu_{[1]} = \dfrac{1}{p}$

$$G''_X(u) = \frac{2pq}{(1-qu)^3} \text{ donc } G''_X(1) = \frac{2pq}{p^3} = \frac{2q}{p^2} = \mu_{[2]}$$

donc $\mathbb{V}(X) = \mu_{[1]} + \mu_{[2]} - \mu_{[1]}^2 = \dfrac{1}{p} + \dfrac{2q}{p^2} - \dfrac{1}{p^2} = \dfrac{p+q+q-1}{p^2} = \dfrac{q}{p^2}$ □

8.4.3 Seconde loi de Pascal

Il existe un **second aléa de Pascal** Y comptabilisant le nombre d'échecs jusqu'au premier succès : $Y = X - 1$.

Définition 85

Avec les notations précédentes :
$Y\langle\Omega\rangle = \mathbb{N}$. $\forall k \in \mathbb{N}, \mathbb{P}(Y = k) = q^k p$.

Proposition 86

1. $\forall u \in \left]-\frac{1}{q}, \frac{1}{q}\right[, G_Y(u) = \dfrac{p}{1 - qu}$
2. Ainsi $\mathbb{E}(Y) = \dfrac{q}{p}$; $\quad \mathbb{V}(Y) = \dfrac{q}{p^2}$ et $\sigma(Y) = \dfrac{\sqrt{q}}{p}$ (car $Y = X - 1$)

8.5 Loi binomiale négative

On reproduit plusieurs fois le second aléa de Pascal de façon indépendante.

Plus précisément : pour $i \geqslant 1$, on note X_i la VARD qui compte le nombre d'échecs compris entre le $(i-1)^{\text{ème}}$ succès et le $i^{\text{ème}}$ succès.
Pour $n \geqslant 1$, on note $Y_n = \displaystyle\sum_{i=1}^{n} X_i$ somme indépendante de n seconds aléas de Pascal.
Y_n compte le nombre d'échecs jusqu'au $n^{\text{ème}}$ succès, et « $Y_n = k$ » est l'évènement où il y a eu k échecs et n succès avec un succès à la dernière épreuve. On dit que Y_n suit la **loi binomiale négative de paramètres** $n \in \mathbb{N}^*$ **et** $p \in]\,0, 1\,[$.

Définition 87

$Y_n\langle\Omega\rangle = \mathbb{N}$ et

$$\forall k \in \mathbb{N}, \mathbb{P}(Y_n = k) = \binom{n+k-1}{n-1} p^{n-1} q^k p = \binom{n+k-1}{n-1} p^n q^k$$

Proposition 88

1. Par construction, la loi binomiale négative est la convolution de n secondes lois de Pascal, donc

$$\forall u \in \left] -\frac{1}{q}, \frac{1}{q} \right[, G_{Y_n}(u) = \left(\frac{p}{1-uq} \right)^n$$

2. Ainsi, $\mathbb{E}(Y_n) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = n\mathbb{E}(X_i) = \dfrac{nq}{p}$; $\mathbb{V}(Y_n) = \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{indépendance}}{=} n\mathbb{V}(X_i) = \dfrac{nq}{p^2}$.

Chapitre 9

Loi de Poisson

9.1 Loi de probabilité

9.1.1 Cadre et énoncé de la loi de Poisson

Il s'agit d'une **succession illimitée d'épreuves** :

- Répétées
- Identiques
- Indépendantes à caractère dichotomique, avec une proportion de succès p « petit », c'est-à-dire pour un **phénomène rare**.

On s'intéresse ici au **nombre k de succès sur un nombre illimité d'épreuves**.

Définition 89

Soit X une VARD suivant la **loi de Poisson de paramètre réel** $\lambda > 0$, on note $X \sim \mathcal{P}(\lambda)$.
Cette loi est définie par :

$$X\langle\Omega\rangle = \mathbb{N} \text{ avec } \forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = p_k = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k\,!}$$

Pour tout $k \in \mathbb{N}$, on a $p_k > 0$ et $\displaystyle\sum_{k\in\mathbb{N}} p_k = \mathrm{e}^{-\lambda}\sum_{k\in\mathbb{N}}\frac{\lambda^k}{k\,!} = \mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda} = 1$. Cela définit bien une loi de probabilité.

9.1.2 Première fonction génératrice, principaux paramètres

Proposition 90

Si $X \sim \mathcal{P}(\lambda)$ alors :

1. $\forall u \in \mathbb{R}, G_X(u) = \mathrm{e}^{\lambda(u-1)}$.
2. $\mathbb{E}(X) = \mathbb{V}(X) = \lambda$ et $\sigma(X) = \sqrt{\lambda}$.
3. $\forall k \in \mathbb{N}^*, \mu_{[k]} = \lambda^k$

Démonstration.

1. $\displaystyle G_X(u) = \mathbb{E}(u^X) = \sum_{k\in\mathbb{N}} u^k p_k = \mathrm{e}^{-\lambda}\sum_{k\in\mathbb{N}} u^k\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k\in\mathbb{N}}\frac{(\lambda u)^k}{k!}$ est convergente pour tout u (série exponentielle), donc le rayon de convergence est $R = +\infty$ et $\forall u \in \mathbb{R}, G_X(u) = \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda u} = \mathrm{e}^{\lambda(u-1)}$.
2. $G'_X(u) = \lambda\mathrm{e}^{\lambda(u-1)}$ donc $\mathbb{E}(X) = \mu_{[1]} = G'_X(1) = \lambda$.
   $G''_X(u) = \lambda^2\mathrm{e}^{\lambda(u-1)}$ donc $\mu_{[2]} = \lambda^2$. Ainsi, $\mathbb{V}(X) = \mu_{[1]} + \mu_{[2]} - \mu_{[1]}^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$.
3. Par récurrence immédiate. □

9.1.3 Convolution de deux lois de Poisson

Proposition 91

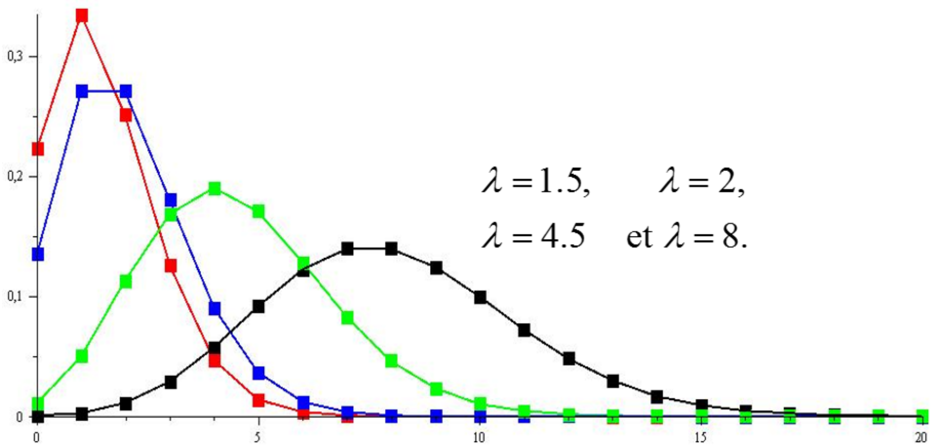
Soient $X_1 \sim \mathcal{P}(\lambda_1)$ et $X_2 \sim \mathcal{P}(\lambda_2)$ deux VARD **indépendantes**, avec $\lambda_1 > 0$ et $\lambda_2 > 0$. Alors :

1. $\mathcal{P}(\lambda_1) * \mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_1 + \lambda_2)$
2. $\forall u \in \mathbb{R}, G_{X_1+X_2}(u) = \mathrm{e}^{(\lambda_1+\lambda_2)(u-1)}$

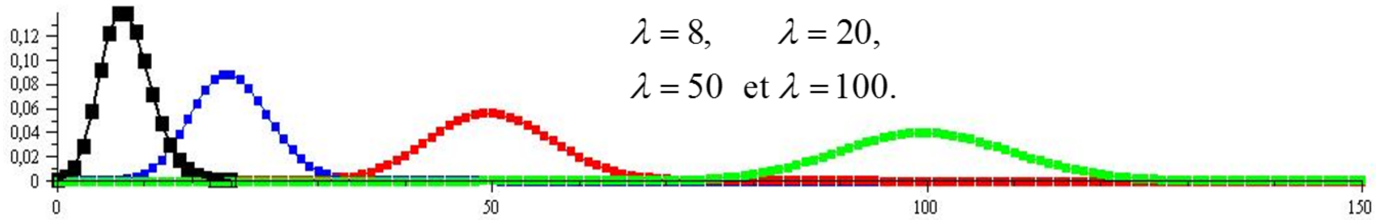
Démonstration. $G_{X_1+X_2}(u) = G_{X_1}(u) \times G_{X_2}(u) = \mathrm{e}^{\lambda_1(u-1)} \times \mathrm{e}^{\lambda_2(u-1)} = \mathrm{e}^{(\lambda_1+\lambda_2)(u-1)}$.
Comme les fonctions génératrices caractérisent la loi, $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$. □

9.2 Représentation graphique de la loi de Poisson

Selon la valeur de λ, on obtient :



La représentation graphique a un maximum (le mode) pour $k = [\lambda]$ (partie entière) et si λ est entier, on a deux réalisations du maximum en $k = \lambda - 1$ et $k = \lambda$.

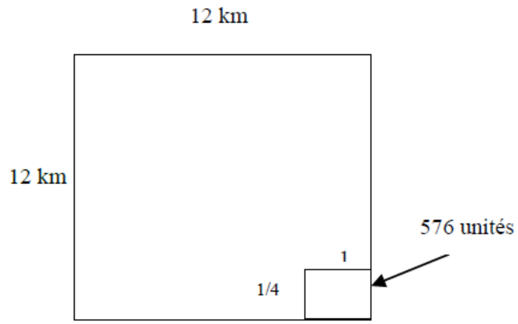


Rappel : Le paramètre λ est à la fois **le mode, l'espérance et la variance** de cette loi de probabilité.

9.3 Un cas concret

On considère la chute des V1 (bombes allemandes en 1941) sur Londres :

On délimite un espace de 12 km sur 12 km, que l'on partage en rectangles de 1/4km fois $1km$. Il y a $576 = 4 \times 12 \times 12$ rectangles « unités d'observation ».



On compte le nombre de V1 tombées dans chaque unité, il y en a X entre 0 et 5 donc $X\langle\Omega\rangle = \{0, 1, 2, 3, 4, 5\}$ et on note $x_i = i$. On compte le nombre n_i d'unités ayant reçu x_i bombes. On constate les résultats du tableau ci-dessous.

x_i = V1 par unité	0	1	2	3	4	5	Somme
n_i : Effectifs enregistrés	229	211	93	35	7	1	576
Fréquence de chaque cas : $f_i = \frac{n_i}{576}$	0,398	0,366	0,161	0,061	0,012	0,002	1
$f_i \cdot x_i$	0	0,3663	0,3229	0,1823	0,0486	0,0087	0,9288
$f_i \cdot x_i^2$	0	0,3663	0,6458	0,5469	0,1944	0,0434	1,7969

On obtient donc une moyenne $m = 0,929$ et une variance $\mathbb{V} = 1,7969 - 0,9288^2 = 0,934$.
La quasi égalité entre la moyenne et la variance calculées sur les données empiriques nous amènent à envisager que la VARD
« X suit la loi de Poisson de paramètre $\lambda = 0,93$ »

Le tableau ci-dessous présente les effectifs empiriques et théoriques calculés à partir de cette loi :

x_i	n_i effectif constaté	p_i théorique	e_i effectif théorique
0	229	0,395	227,5
1	211	0,367	211,4
2	93	0,171	98,4
3	35	0,052	30,6
4	7	0,012	6,9
5	1	0,002	1,1

Vous verrez en cours de Statistiques comment estimer la proximité entre les valeurs théoriques et les valeurs constatées, et donc **comment valider une hypothèse** , ici « cette situation suit une loi de Poisson de paramètre $\lambda = 0,93$ ».

Chapitre 10

Loi hypergéométrique

10.1 Cadre et énoncé de la loi hypergéométrique

Il s'agit d'une succession d'épreuves :

- Répétées n fois, où $n \in \mathbb{N}^*$,
- Identiques,
- Exhaustives (dépendantes) à caractère dichotomique.

On s'intéresse ici au **nombre k de succès parmi ces n épreuves.**

On tire **sans remise** n boules d'une urne contenant N boules indiscernables, où $N = N_1 + N_2$ avec :

- $N_1 = N \times p$ boules **blanches**, où p est la proportion de boules blanches, $p \in [0, 1]$
- $N_2 = N \times q$ boules **noires**, où $p + q = 1$.

X est la VARD prenant pour valeur le nombre de boules blanches tirées. Ω est l'ensemble des combinaisons ω de n boules tirées parmi les N, avec $0 \leqslant n \leqslant N$. $(\Omega, \mathcal{P}(\Omega))$ est un espace probabilisable avec **équiprobabilité**.

$\text{card}(\Omega) = \dbinom{N}{n}$, ainsi $\mathbb{P}(\{\omega\}) = \dfrac{1}{\binom{N}{n}}$ pour un évènement élémentaire et $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est un espace probabilisé.

En prenant k boules parmi N_1 blanches et $n-k$ parmi N_2 noires on a :

Définition 92

Soit X une VARD suivant la **loi Hypergéométrique de paramètres** N, n, p, on note $X \sim \mathcal{H}(N, n, p)$ avec :

$$\begin{cases} N = N \times p + N \times q & : \text{taille de la population totale} \\ n & : \text{taille de l'échantillon avec } 0 \leqslant n \leqslant N \\ p \in [0,1] & : \text{probabilité du succès, et } q = 1-p \end{cases}$$

$$X\langle\Omega\rangle = \left[\!\left[\sup(0, n - N \times q)\,,\, \inf(N \times p, n) \right]\!\right] \text{ et } \forall k \in X\langle\Omega\rangle \;\; \mathbb{P}(X=k) = \frac{\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}} = \frac{\binom{pN}{k}\binom{qN}{n-k}}{\binom{N}{n}}$$

Justification du support : $\displaystyle\binom{N}{n} = \binom{N_1+N_2}{n} = \sum_{k=0}^{n} \binom{N_1}{k}\binom{N_2}{n-k}$ car pour constituer un sous-ensemble à n éléments pris parmi $N = N_1 + N_2$, on en prend k parmi N_1 puis $n-k$ parmi N_2, où k varie de 0 à n.
On a $0 \leqslant k \leqslant N_1 = N \times p$, $k \leqslant n$, et $0 \leqslant n-k \leqslant N_2 = N \times q$. Donc $\sup(0, n - N \times q) \leqslant k \leqslant \inf(N \times p, n)$.

10.2 Principaux paramètres

Proposition 93

$$\text{Si } X \sim \mathcal{H}(N,n,p) \text{ alors : } \mathbb{E}(X) = np \qquad \mathbb{V}(X) = npq\frac{N-n}{N-1}.$$

Démonstration. Rappelons que $k\dbinom{n}{k} = \dfrac{k \times n!}{k!(n-k)!} = \dfrac{n(n-1)!}{(k-1)!(n-k)!} = n\dbinom{n-1}{k-1}$.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k \times p_k \\
&= \frac{1}{\binom{N}{n}} \sum_{k=1}^{n} k \binom{N\cdot p}{k}\binom{N\cdot q}{n-k} \\
&= \frac{1}{\binom{N}{n}} \sum_{k=1}^{n} N\cdot p\cdot \binom{N\cdot p-1}{k-1}\binom{N\cdot q}{n-k} \\
&= \frac{N\cdot p}{\binom{N}{n}} \sum_{k=1}^{n} \binom{N\cdot p-1}{k-1}\binom{N\cdot q}{n-k} \\
&= \frac{n\cdot p}{\binom{N-1}{n-1}} \sum_{k=1}^{n} \binom{N\cdot p-1}{k-1}\binom{N\cdot q}{n-k} \\
&\overset{k'=k-1}{=} \frac{n\cdot p}{\binom{N-1}{n-1}} \underbrace{\sum_{k'=0}^{n-1} \binom{N\cdot p-1}{k'}\binom{N\cdot q}{n-1-k'}}_{=\binom{N-1}{n-1}} \\
&= n\cdot p
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}(X(X-1)) &= \sum_{k=0}^{n} k(k-1)p_k \\
&= \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} k(k-1)\binom{N\cdot p}{k}\binom{N\cdot q}{n-k} \\
&= \frac{N\cdot p}{\binom{N}{n}} \sum_{k=2}^{n} (k-1)\binom{N\cdot p-1}{k-1}\binom{N\cdot q}{n-k} \\
&= \frac{N\cdot p\cdot (N\cdot p-1)}{\binom{N}{n}} \sum_{k=2}^{n} \binom{N\cdot p-2}{k-2}\binom{N\cdot q}{n-k} \\
&= \frac{N\cdot p\cdot (N\cdot p-1)}{\frac{N}{n}\times\frac{N-1}{n-1}\binom{N-2}{n-2}} \sum_{k=2}^{n} \binom{N\cdot p-2}{k-2}\binom{N\cdot q}{n-k} \\
&\overset{k'=k-2}{=} \frac{n\cdot p\cdot (N\cdot p-1)}{\frac{N-1}{n-1}\binom{N-2}{n-2}} \sum_{k'=0}^{n-2} \binom{N\cdot p-2}{k'}\binom{N\cdot q}{n-2-k'} \\
&= \frac{n(n-1)\cdot p\cdot (N\cdot p-1)}{N-1}
\end{aligned}$$

Ainsi, on obtient :

$$\begin{aligned}
\mathbb{V}(X) &= \mu_{[1]} + \mu_{[2]} - \mu_{[1]}^2 &&= n\cdot p + \frac{n(n-1)\cdot p\cdot (N\cdot p-1)}{N-1} - n^2p^2 \\
&= n\cdot p\left(\frac{N-1+(n-1)\cdot(N\cdot p-1) - n\cdot p\cdot (N-1)}{N-1}\right) &&= n\cdot p\left(\frac{(n-N)\cdot(p-1)}{N-1}\right)
\end{aligned}$$

D'où $\boxed{\mathbb{V}(X) = npq\frac{N-n}{N-1}\cdot}$ $\left(\frac{N-n}{N-1}\right)$ est appelé le **coefficient d'exhaustivité**. Quand $N \to +\infty$, alors $\frac{N-n}{N-1} \to 1$. □

Remarque : Si l'on compare à une VARD suivant une loi Binomiale, notée $X_{\mathcal{B}(n,p)}$, alors on a :

$$\mathbb{E}\left(X_{\mathcal{H}(N,n,p)}\right) = \mathbb{E}\left(X_{\mathcal{B}(n,p)}\right) \quad \text{et} \quad \mathbb{V}\left(X_{\mathcal{H}(N,n,p)}\right) = \mathbb{V}\left(X_{\mathcal{B}(n,p)}\right) \times \frac{N-n}{N-1}$$

Application : Dans la pratique, le contrôle métrologique industriel priviligie la loi binomiale (tirages avec remise) à la loi hypergéométrique (tirage sans remise) car la population est "grande".

Troisième partie

Lois de probabilité continues

Chapitre 11

Loi uniforme sur $[0, 1]$

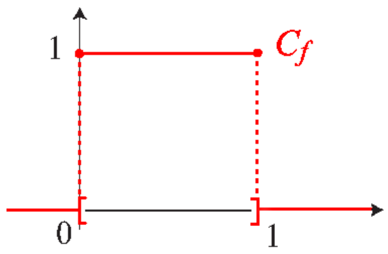
11.1 Définition

Définition 94

On dit qu'une VARC X suit la **loi uniforme** sur $[0, 1]$ si elle admet une fonction densité de probabilité f définie par
$f : x \longmapsto \begin{cases} f(x) = 1 & \text{si } x \in [0, 1] \\ f(x) = 0 & \text{si } x \notin [0, 1] \end{cases}$
On note $X \sim \mathcal{U}\Big([0, 1]\Big)$ avec $X\langle\Omega\rangle = [0, 1]$.

f définit bien une densité de probabilité car :

- f est positive sur $\mathbb{R}$
- f est continue sur $\mathbb{R} \setminus \{0,1\}$
- f est sommable sur $\mathbb{R}$ et $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$



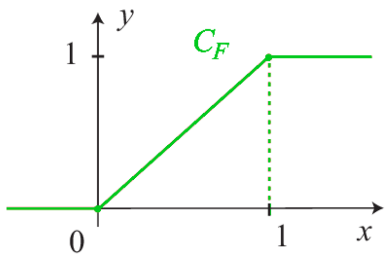
11.2 Fonction de répartition et principaux paramètres

11.2.1 Fonction de répartition

Par définition, la **fonction de répartition** est donnée par

$$F : \left\{ \begin{array}{ccl} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & F(x) = \mathbb{P}\big(X \leqslant x\big) \end{array} \right.$$

$$\boxed{\text{On a ainsi : } \left\{ \begin{array}{ll} \text{si } x < 0 & F(x) = 0 \\ \text{si } 0 \leqslant x \leqslant 1 & F(x) = x \\ \text{si } x > 1 & F(x) = 1 \end{array} \right.}$$



F est affine par morceaux, continue sur $\mathbb{R}$, et strictement croissante sur $[0,1]$.

11.2.2 Principaux paramètres

Théorème 95

Si $X \sim \mathcal{U}\Big([0,1]\Big)$, alors $\mathbb{E}\big(X\big) = \dfrac{1}{2}$, $\mathbb{V}\big(X\big) = \dfrac{1}{12}$ et $\sigma(X) = \dfrac{\sqrt{3}}{6}$.

Démonstration. $\mathbb{E}\big(X\big) = \displaystyle\int_{\mathbb{R}} x\, f(x)\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \frac{1}{2}\left[x^2\right]_0^1 = \frac{1}{2}$.

$\mathbb{E}\big(X^2\big) = \displaystyle\int_{\mathbb{R}} x^2\, f(x)\, \mathrm{d}x = \int_0^1 x^2\, \mathrm{d}x = \frac{1}{3}\Big[x^3\Big]_0^1 = \frac{1}{3}$

Par conséquent $\mathbb{V}\big(X\big) = \mathbb{E}\big(X^2\big) - \Big[\mathbb{E}\big(X\big)\Big]^2 = \dfrac{1}{3} - \dfrac{1}{4} = \dfrac{1}{12}$.

Enfin, $\sigma(X) = \sqrt{\mathbb{V}(X)} = \sqrt{\dfrac{1}{12}} = \dfrac{\sqrt{3}}{6}$. $\square$

11.3 Deuxième fonction génératrice

$\forall t \in \mathbb{R} : g_X(t) = \mathbb{E}\Big(\mathrm{e}^{t\,X}\Big) = \displaystyle\int_{\mathbb{R}} \mathrm{e}^{t\,X}\, f(x)\, \mathrm{d}x = \int_0^1 \mathrm{e}^{t\,X}\, \mathrm{d}x$

$\boxed{\text{Donc } g_X(t) = \begin{cases} \dfrac{\mathrm{e}^t - 1}{t} & \text{pour } t \neq 0 \\ 1 & \text{pour } t = 0 \end{cases}}$

g_X est donc définie, continue sur $\mathbb{R}$, et même développable en série entière sur $\mathbb{R}$.
On peut retrouver ainsi, par dérivation, les principaux paramètres de X. (ainsi que les moments simples)

Chapitre 12

Loi uniforme rectangle sur $[a, b]$

12.1 Définition

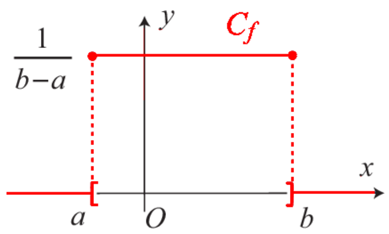
Définition 96

Soient a et b deux réels tels que $a < b$. On dit qu'une VARC X suit la **loi uniforme** sur $[a, b]$, si elle admet une fonction densité de probabilité f définie par $f : x \longmapsto \begin{cases} f(x) = \dfrac{1}{b-a} & \text{si } x \in [a, b] \\ f(x) = 0 & \text{si } x \notin [a, b] \end{cases}$

On note $X \sim \mathcal{U}\Big([a, b]\Big)$, avec $X\langle\Omega\rangle = [a, b]$.

f définit bien une densité de probabilité car :

- f est positive sur $\mathbb{R}$
- f est continue sur $\mathbb{R} \setminus \{a, b\}$
- f est sommable sur $\mathbb{R}$ et $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$



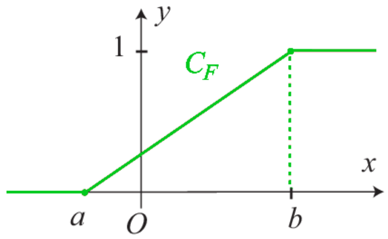
12.2 Fonction de répartition et principaux paramètres

12.2.1 Fonction de répartition

La **fonction de répartition** est :

$$F : \begin{cases} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & F(x) = \mathbb{P}(X \leqslant x) \end{cases}$$

On a ainsi : $\begin{cases} \text{si } x < a & F(x) = 0 \\ \text{si } a \leqslant x \leqslant b & F(x) = \dfrac{x-a}{b-a} \\ \text{si } x > b & F(x) = 1 \end{cases}$



F est affine par morceaux, continue sur $\mathbb{R}$, et strictement croissante sur $[a, b]$.

12.2.2 Principaux paramètres

Théorème 97

Si $X \sim \mathcal{U}\big([a, b]\big)$, alors $\mathbb{E}(X) = \dfrac{a+b}{2}$, $\mathbb{V}(X) = \dfrac{(b-a)^2}{12}$ et $\sigma(X) = \dfrac{(b-a)\sqrt{3}}{6}$.

Démonstration. $\mathbb{E}(X) = \displaystyle\int_{\mathbb{R}} x\, f(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x\,\mathrm{d}x = \frac{1}{2(b-a)}\Big[x^2\Big]_a^b = \frac{a+b}{2}$.

$$\mathbb{E}\big(X^2\big) = \int_{\mathbb{R}} x^2\, f(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b x^2\,\mathrm{d}x = \frac{1}{b-a}\,\frac{1}{3}\Big[x^3\Big]_a^b = \frac{1}{3}\Big(a^2 + a\,b + b^2\Big)$$

Par conséquent, $\mathbb{V}\big(X\big) = \mathbb{E}\big(X^2\big) - \Big[\mathbb{E}\big(X\big)\Big]^2 = \frac{1}{3}\Big(a^2 + a\,b + b^2\Big) - \frac{1}{4}\Big(a^2 + 2a\,b + b^2\Big) = \frac{1}{12}\Big(a^2 - 2a\,b + b^2\Big) = \frac{(b-a)^2}{12}$

Enfin, $\sigma(X) = \sqrt{\mathbb{V}(X)} = \frac{b-a}{\sqrt{12}} = \frac{(b-a)\sqrt{3}}{6}\cdot$ □

12.3 Deuxième fonction génératrice

$$\forall t \in \mathbb{R} : g_X(t) = \mathbb{E}\Big(\mathrm{e}^{t\,X}\Big) = \int_{\mathbb{R}} \mathrm{e}^{tx}\, f(x)\,\mathrm{d}x = \frac{1}{b-a}\int_a^b \mathrm{e}^{tx}\,\mathrm{d}x$$

$$\boxed{\text{Ainsi : } g_X(t) = \begin{cases} \dfrac{\mathrm{e}^{b\,t} - \mathrm{e}^{a\,t}}{t(b-a)} & \text{pour } t \neq 0 \\ 1 & \text{pour } t = 0 \end{cases}}$$

Cette deuxième fonction génératrice permet de retrouver, par dérivation, les principaux paramètres de X.(ainsi que les moments simples)

Chapitre 13

Loi exponentielle (négative)

13.1 Définition

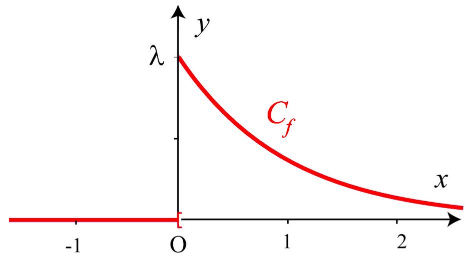
Définition 98

On dit qu'une variable aléatoire X suit la **loi exponentielle** (négative) de paramètre λ où $\lambda > 0$, si elle admet une fonction densité de probabilité f définie par $f : x \longmapsto \begin{cases} f(x) = \lambda\,\mathrm{e}^{-\lambda\,x} & \text{si } x \geqslant 0 \\ f(x) = 0 & \text{si } x < 0 \end{cases}$

On note : $X \sim \mathcal{E}\big(\lambda\big)$, avec $X\langle\Omega\rangle = [0, +\infty[$.

f définit bien une densité de probabilité car :

- f est positive sur $\mathbb{R}$
- f est continue sur $\mathbb{R}^*$
- f est sommable sur $\mathbb{R}$ et $\displaystyle\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$



En effet, $\displaystyle\int_{\mathbb{R}} f(t)\,\mathrm{d}t = \int_0^{+\infty} \lambda\,\mathrm{e}^{-\lambda\,t}\mathrm{d}t = \lim_{A\to+\infty}\int_0^{A} \lambda\,\mathrm{e}^{-\lambda\,t}\mathrm{d}t = \lim_{A\to+\infty}\Big[-\mathrm{e}^{-\lambda\,t}\Big]_0^A = \lim_{A\to+\infty}\Big(1-\mathrm{e}^{-\lambda\,A}\Big) = 1$ car $\lambda > 0$

13.2 Fonction de répartition et principaux paramètres

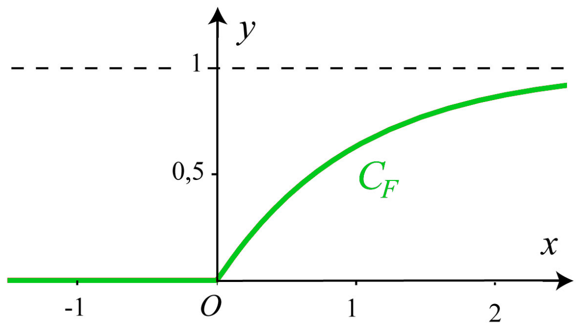
13.2.1 Fonction de répartition

La **fonction de répartition** est donnée par

$$F : \left\{\begin{array}{lcl} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & F(x) = \mathbb{P}\big(X \leqslant x\big) \end{array}\right.$$

On a ainsi : $\left\{\begin{array}{ll} \text{si } x < 0 & F(x) = 0 \\ \text{si } x \geqslant 0 & F(x) = \displaystyle\int_0^x \lambda\,\mathrm{e}^{-\lambda\,t}\mathrm{d}t = \Big[-\mathrm{e}^{-\lambda\,t}\Big]_0^x = 1-\mathrm{e}^{-\lambda\,x} \end{array}\right.$

F est continue sur $\mathbb{R}$, strictement croissante sur $[0,+\infty[$, et de classe C^∞ par morceaux.



13.2.2 Principaux paramètres

Théorème 99

Si $X \sim \mathcal{E}(\lambda))$, alors $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $\mathbb{V}(X) = \dfrac{1}{\lambda^2}$, et $\sigma(X) = \dfrac{1}{\lambda}$.

Démonstration. $\mathbb{E}(X) = \displaystyle\int_{\mathbb{R}} x\, f(x)\,\mathrm{d}x = \lambda \int_0^{+\infty} x\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x$ qu'on intègre par parties :

$$\int_0^A x\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \left[\frac{x\,\mathrm{e}^{-\lambda x}}{-\lambda}\right]_0^A + \frac{1}{\lambda}\int_0^A \mathrm{e}^{-\lambda x}\,\mathrm{d}x = -\frac{A\,\mathrm{e}^{-\lambda A}}{\lambda} + \frac{1}{\lambda}\left[-\frac{\mathrm{e}^{-\lambda x}}{\lambda}\right]_0^A = \frac{1}{\lambda^2} - \frac{A\,\mathrm{e}^{-\lambda A}}{\lambda} - \frac{\mathrm{e}^{-\lambda A}}{\lambda^2} \xrightarrow[A \to +\infty]{} \frac{1}{\lambda^2}$$

D'où, $\mathbb{E}(X) = \dfrac{1}{\lambda}$

Avec deux I.P.P. : $\mathbb{E}(X^2) = \displaystyle\int_{\mathbb{R}} x^2\, f(x)\,\mathrm{d}x = \frac{2}{\lambda^2}$ et par conséquent $\mathbb{V}(X) = \mathbb{E}(X^2) - \Big[\mathbb{E}(X)\Big]^2 = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$.

Enfin, $\sigma(X) = \sqrt{\mathbb{V}(X)} = \dfrac{1}{\lambda}$. □

13.3 Deuxième fonction génératrice

Sous réserve d'existence, $g_X(t) = \mathbb{E}\Big(\mathrm{e}^{t\,X}\Big) = \displaystyle\int_{\mathbb{R}} \mathrm{e}^{tx}\, f(x)\,\mathrm{d}x = \int_0^{+\infty} \mathrm{e}^{t\,x}\,\lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \lambda \int_0^{+\infty} \mathrm{e}^{(t-\lambda)\,x}\,\mathrm{d}x$.

Pour $t < \lambda$, on a : $g_X(t) = \dfrac{\lambda}{\lambda - t}$. Ainsi,

Théorème 100

Si $X \sim \mathcal{E}(\lambda)$, alors $g_X(t) = \dfrac{\lambda}{\lambda - t}$ pour $t \in]-\infty, \lambda[$.

g_X est donc définie et de classe C^{∞} sur $]-\infty, \lambda[$, et les moments simples d'ordre k sont obtenus par dérivation en 0 :

$$\boxed{\forall k \in \mathbb{N}^*, \quad m_k = \mathbb{E}\big(X^k\big) = g_X^{(k)}(0) = \frac{k\,!}{\lambda^k}}$$

On retrouve ainsi les principaux paramètres de X : $\boxed{\mathbb{E}\big(X\big) = \frac{1}{\lambda}, \ \mathbb{V}\big(X\big) = \frac{1}{\lambda^2} \text{ et } \sigma\big(X\big) = \frac{1}{\lambda}\cdot}$

13.4 Convolution de lois exponentielles

Soient $\lambda_1 > 0$, $\lambda_2 > 0$ et deux VARC $X_1 \sim \mathcal{E}(\lambda_1)$ et $X_2 \sim \mathcal{E}(\lambda_2)$ **indépendantes**.

Alors leurs fonctions génératrices sont définies par $g_{X_1} : t \longmapsto \dfrac{\lambda_1}{\lambda_1 - t}$ pour $t < \lambda_1$ et $g_{X_2} : t \longmapsto \dfrac{\lambda_2}{\lambda_2 - t}$ pour $t < \lambda_2$

Ainsi $g_{X_1+X_2} = g_{X_1} \times g_{X_2} : t \longmapsto \dfrac{\lambda_1 \cdot \lambda_2}{(\lambda_1 - t)\cdot(\lambda_2 - t)}$ pour $t < \lambda$ où $\lambda = \inf\big(\lambda_1, \lambda_2\big)$.

Ce n'est pas la fonction génératrice d'une loi exponentielle.

Cependant, si $\lambda_1 = \lambda_2 = \lambda$, alors $g_{X_1+X_2} : t \longmapsto \dfrac{\lambda^2}{(\lambda - t)^2}$ pour $t < \lambda$.

On obtient alors la deuxième fonction génératrice d'une **loi d'Erlang** de paramètre λ et d'ordre 2.

13.5 Applications : système sans mémoire

13.5.1 Phénomène dépendant de la durée, pas de l'instant initial

Théorème 101

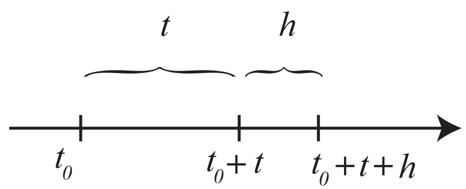
On étudie une VARC X, qui à un évènement E (fission d'un atome, défectuosité d'une ampoule, ...) associe l'instant t où il a lieu. Le système physique (atome, ampoule, ...) évolue avec le temps t, t parcourant $[0, +\infty[$ et l'évènement E peut se produire à un instant aléatoire.

On suppose que la probabilité de réalisation de E durant un intervalle de temps $[\,t_1, t_1 + t_2\,]$, avec $t_1 > 0$ et $t_2 > 0$ ne dépend que de la durée de celui-ci, t_2, mais pas de l'instant initial t_1.

On démontre qu'alors, il existe $\lambda > 0$ tel que la fonction de répartition de X soit définie par $\forall t \geqslant 0, \quad F(t) = 1 - \mathrm{e}^{-\lambda\, t}$, et $\forall t < 0, \quad F(t) = 0$. Ainsi X suit une **loi de probabilité exponentielle**.

Notons $F(t) = \mathbb{P}\big(E \in [\,0, t\,]\big) = \mathbb{P}\big(X \leqslant t\big)$, et la probabilité pour que E ne se produise pas durant $[\,0, t\,]$, est $G(t) = \mathbb{P}\Big(\overline{E \in [0, t]}\Big) = 1 - F(t)$.
On va supposer G dérivable sur $[\,0, +\infty[$. On a $G(0) = 1$, notons $\lambda = -G\,'(0^+)$, c'est un réel.

Pour des intervalles successifs, notons E_1 l'évènement E se produit pendant l'intervalle $[\,t_0, t_0{+}t\,]$ et E_2, pendant $[\,t_0 + t, t_0 + t + h\,]$.



Avec nos hypothèses : $\mathbb{P}\big(E_1\big) = F(t)$, $\mathbb{P}\big(E_2\big) = F(h)$, $\mathbb{P}\Big(\overline{E_1}\Big) = G(t)$ et $\mathbb{P}\Big(\overline{E_2}\Big) = G(h)$.
Ces évènements E_1 et E_2 sont indépendants, donc leurs complémentaires aussi.

Ainsi $G(t+h) = \mathbb{P}\Big(\overline{E_1 \cup E_2}\Big) = \mathbb{P}\Big(\overline{E_1} \cap \overline{E_2}\Big) = \mathbb{P}\Big(\overline{E_1}\Big)\,\mathbb{P}\Big(\overline{E_2}\Big) = G(t)\,G(h)$

Relation caractéristique : $\boxed{\text{Pour tout } t > 0 \text{ et } h > 0,\ G(t+h) = G(t)\,G(h)}$ (traduction de nos hypothèses)

Alors : $\dfrac{G(t+h) - G(t)}{h} = G(t)\,\dfrac{G(h) - 1}{h} = G(t)\,\dfrac{G(h) - G(0)}{h}$,

Et donc $\lim\limits_{h \to 0^+} \left(\dfrac{G(t+h) - G(t)}{h}\right) = G(t) \lim\limits_{h \to 0^+} \left(\dfrac{G(h) - G(0)}{h}\right) = -\lambda\, G(t)$.

Ainsi : $\forall t > 0,\quad G'(t) = -\lambda\, G(t)$ et G est solution sur $]\,0, +\infty[$ d'une équation différentielle linéaire d'ordre un.

$\boxed{\text{D'où } G : t \longmapsto \mathrm{e}^{-\lambda\, t} \text{ sur } [\,0, +\infty[\text{, en prolongeant par continuité en } G(0) = 1.}$ Et le résultat annoncé pour F et X.

13.5.2 Absence de mémoire, d'usure ou de vieillissement

Etudions les VARC X vérifiant : $\boxed{\forall (t_1, t_2) \in]\,0, +\infty\,[^2,\quad \mathbb{P}\Big(X > t_1 + t_2 \ /\ X > t_1\Big) = \mathbb{P}\Big(X > t_2\Big).}$

Interprétation : la probabilité que l'évènement ait lieu au-delà de $t_1 + t_2$, sachant qu'il a lieu au delà de t_1, est égale à la probabilité qu'il ait lieu au delà de t_2. Ce qui s'est passé avant t_1, n'a pas d'influence sur ce qui se passe entre t_1 et $t_1 + t_2$, qui est comparable à ce qui se passe entre 0 et t_2. Le système n'a pas de mémoire, ou d'usure, ou de vieillissement.

Notons $G : t \longmapsto \mathbb{P}\big(X > t\big)$, on a $G = 1 - F$ où F est la fonction de répartition de X. Notons que $G(1) \in\ [\,0, 1\,]$.

On a $\forall (t_1, t_2) \in]\,0, +\infty\,[^2,\quad \dfrac{\mathbb{P}\big(X > t_1 + t_2 \ \text{ et } X > t_2\big)}{\mathbb{P}\big(X > t_1\big)} = \mathbb{P}\big(X > t_2\big)$ donc $\dfrac{\mathbb{P}\big(X > t_1 + t_2\big)}{\mathbb{P}\big(X > t_1\big)} = \mathbb{P}\big(X > t_2\big)$.

On obtient la même relation caractéristique : $\boxed{\forall (t_1, t_2) \in]\,0, +\infty\,[^2,\quad G\big(t_1 + t_2\big) = G\big(t_1\big) \times G\big(t_2\big).}$

Étudions cette relation avec une **nouvelle démonstration**.
On a : $\forall n \in \mathbb{N},\ \forall t > 0,\ G\big((n+1)\, t\big) = G\big(n\, t + t\big) = G\big(n\, t\big) \times G(t)$, et par récurrence simple : $\forall n \in \mathbb{N},\ \forall t > 0,\ G\big(n\, t\big) = G(t)^n$.

Donc : $\quad \forall n \in \mathbb{N},\ G(n) = G(1)^n$

$$\forall q \in \mathbb{N}^*,\ G(1) = G\big(q\,\tfrac{1}{q}\big) = G\big(\tfrac{1}{q}\big)^q \text{ donc } G\big(\tfrac{1}{q}\big) = G(1)^{\frac{1}{q}}$$

$$\forall p \in \mathbb{N},\ \forall q \in \mathbb{N}^*,\ G\big(p\,\tfrac{1}{q}\big) = G\big(\tfrac{1}{q}\big)^p = G(1)^{\frac{p}{q}}$$

et par densité des rationnels dans $[0, +\infty[$ et la continuité de G, on obtient : $\boxed{\forall x \geqslant 0,\quad G(x) = G(1)^x}$

Si $G(1) = 0$, on a la solution $G = 0$ constante, de même si $G(1) = 1$, avec $G = 1$ constante.
Si $G(1) \in]0, 1[$, alors en posant $\lambda = -\ln\big(G(1)\big)$, on a $\lambda > 0$ et $\forall x > 0,\quad G(x) = \mathrm{e}^{-\lambda\, x}$ d'où $F(x) = 1 - \mathrm{e}^{-\lambda\, x}$ pour $x > 0$.
On a bien une **loi exponentielle**, on retrouve le même résultat.

13.5.3 Applications

Un domaine usuel d'application de la loi exponentielle est le domaine de la **radioactivité**. Chaque atome radioactif possède une durée de vie qui suit une loi exponentielle. Le paramètre λ s'appelle alors la constante de désintégration.
La durée de vie d'un composant électronique peut aussi être modélisée par une loi exponentielle, par exemple pour une ampoule LED, dont on estime qu'elle ne vieillit pas : la probabilité qu'elle ne tombe pas en panne pendant le laps de temps $[t_1, t_1 + t_2]$ est la même que pendant $[0, t_2]$, elle ne dépend que de la durée. Ce n'est pas le cas d'une ampoule à filament qui vieillit avec la chaleur dégagée, et dont la durée de vie ne suit pas une loi exponentielle.
On utilise encore la loi exponentielle pour modéliser les files d'attente.

Chapitre 14

Loi normale ou loi de Laplace-Gauss

14.1 Définition et fonction densité de la loi normale

14.1.1 Définition d'une loi normale

Définition 102

Soit $m \in \mathbb{R}$ et $\sigma > 0$. On dit qu'une VARC X suit une **loi normale** de paramètres réels m et σ^2, si elle admet une fonction densité de probabilité g telle que : $\forall x \in \mathbb{R}, \quad g(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$

On note $X \sim \mathcal{N}\big(m, \sigma^2\big)$, avec $X\langle\Omega\rangle = \mathbb{R}$.

g définit bien une densité de probabilité car :
- Pour tout x dans $\mathbb{R}$, $g(x) \geqslant 0$.
- La fonction g est continue sur $\mathbb{R}$.

• On a : $\displaystyle\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}\,\mathrm{d}x.$

Par le changement de variable : $\left(t = \dfrac{x-m}{\sigma}\right) \iff \big(x = \sigma\, t + m\big)$, avec $\mathrm{d}x = \sigma\,\mathrm{d}t$ on obtient :

$\displaystyle\int_{-\infty}^{+\infty} g(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t = 1.$ car : $\displaystyle\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t = \sqrt{2\pi}$ (intégrale de Gauss)

14.1.2 Cas de la loi Normale $\mathcal{N}(0,1)$

Définition 103

Le cas particulier où $m = 0$ et $\sigma = 1$ conduit à la loi $\mathcal{N}(0,1)$, **loi « standard »** ou **loi « de référence »** des lois normales.

Dans le cas général :

$X \sim \mathcal{N}(m, \sigma^2)$, avec $m \in \mathbb{R}$ et $\sigma > 0$, on peut faire le changement de variable aléatoire : $\left(T = \dfrac{X-m}{\sigma}\right) \iff \big(X = \sigma\, T + m\big)$.

Si on note G (respectivement F) la fonction de répartition de X (respectivement T), on a :

$\forall t \in \mathbb{R},\quad F(t) = \mathbb{P}\big(T \leqslant t\big) = \mathbb{P}\left(\dfrac{X-m}{\sigma} \leqslant t\right) = \mathbb{P}\big(X \leqslant m + \sigma\, t\big) = G\big(m + \sigma t\big).$

Par dérivation de la fonction de répartition, on obtient une fonction de densité :

$\forall t \in \mathbb{R},\ f(t) = F'(t) = \sigma\, G'(m + \sigma\, t) = \sigma\, g(m + \sigma\, t) = \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}t^2}.$

Ainsi $f : t \longmapsto \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}t^2}$ est une fonction de densité de T, d'où $T \sim \mathcal{N}(0,1)$.

Le changement de variable aléatoire ramène toute loi normale, à la loi normale où $m = 0$ et $\sigma = 1$ qu'on qualifiera de centrée réduite (voir plus loin).

14.2 Deuxième fonction génératrice

14.2.1 Deuxième fonction génératrice de $T \sim \mathcal{N}(0,1)$

Théorème 104

Si $T \sim \mathcal{N}(0,1)$, alors la deuxième fonction génératrice est définie sur $\mathbb{R}$ par $g_T : t \longmapsto \mathrm{e}^{\frac{1}{2}\,t^2}$.

Démonstration. Sous réserve d'existence, la deuxième fonction génératrice de T est définie par :

$$g_T(t) = \mathbb{E}\big(\mathrm{e}^{t\,T}\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{t\,x}\,\mathrm{e}^{-\frac{1}{2}\,x^2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{t\,x-\frac{1}{2}\,x^2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}\,\left[(x-t)^2-t^2\right]}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\frac{1}{2}\,t^2} \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}\,(x-t)^2}\,\mathrm{d}x$$

Par changement de variable : $\big(y = x - t\big) \Longleftrightarrow \big(x = y + t\big)$ avec $\mathrm{d}y = \mathrm{d}x$, on obtient :

$g_T(t) = \mathrm{e}^{\frac{1}{2}\,t^2}\,\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}\,y^2}\,\mathrm{d}y = \mathrm{e}^{\frac{1}{2}\,t^2}$. Au passage, on vient de démontrer que la fonction g_T est bien définie sur $\mathbb{R}$. □

Théorème 105

1. Si $T \sim \mathcal{N}(0,1)$, alors $\mathbb{E}\big(T\big) = 0$ et $\mathbb{V}\big(T\big) = 1$.
2. Si $X \sim \mathcal{N}(m,\sigma^2)$, alors $\mathbb{E}\big(X\big) = m$ et $\mathbb{V}\big(X\big) = \sigma^2$.

Démonstration. D'une part, g_T est infiniment dérivable sur $\mathbb{R}$, et ${g_T}'(t) = t\,\mathrm{e}^{\frac{1}{2}\,t^2}$, donc ${g_T}'(0) = 0$ et $\mathbb{E}\big(T\big) = 0$.

De même : ${g_T}''(t) = \big(1+t^2\big)\,\mathrm{e}^{\frac{1}{2}\,t^2}$, donc ${g_T}''(0) = 1$, $\mathbb{E}\big(T^2\big) = 1$ et donc $\mathbb{V}\big(T\big) = \mathbb{E}\big(T^2\big) - \Big(\mathbb{E}\big(T\big)\Big)^2 = 1$.

D'autre part, $X = \sigma\,T + m$, donc $\mathbb{E}\big(X\big) = \sigma\,\mathbb{E}\big(T\big) + m = m$, et $\mathbb{V}\big(X\big) = \sigma^2\,\mathbb{V}\big(T\big) = \sigma^2$. □

Ainsi, si $X \sim \mathcal{N}(m,\sigma^2)$, alors les paramètres m et σ sont bien l'**espérance** et l'**écart-type** de la VARC X.

14.2.2 Deuxième fonction génératrice de $X \sim \mathcal{N}(m, \sigma^2)$

Théorème 106

Si $X \sim \mathcal{N}(m, \sigma^2)$, alors la deuxième fonction génératrice est définie sur $\mathbb{R}$ par $g_X : t \longmapsto \mathrm{e}^{t\,m+\frac{1}{2}\,\sigma^2\,t^2}$.

Démonstration. Sous réserve d'existence, la deuxième fonction génératrice de X est définie par :
$g_X(t) = \mathbb{E}\big(\mathrm{e}^{t\,X}\big) = \mathbb{E}\big(\mathrm{e}^{t\,(\sigma\,T+m)}\big) = \mathbb{E}\big(\mathrm{e}^{t\,m} \times \mathrm{e}^{t\,\sigma\,T}\big) = \mathrm{e}^{t\,m}\,\mathbb{E}\big(\mathrm{e}^{t\,\sigma\,T}\big) = \mathrm{e}^{t\,m}\,\mathrm{e}^{\frac{1}{2}\,\sigma^2\,t^2} = \mathrm{e}^{t\,m+\frac{1}{2}\,\sigma^2\,t^2}$.
Au passage, on vient de démontrer que la fonction g_X est bien définie sur $\mathbb{R}$.

A partir de cette deuxième fonction génératrice, on peut retrouver les principaux paramètres de X (espérance et variance)

14.3 Variable aléatoire réelle centrée réduite

Définition 107

Pour une VAR X quelconque, admettant une espérance et un écart-type non nul, on lui associe :
- La variable $X - \overline{X}$ **centrée** : $\mathbb{E}\big(X - \overline{X}\big) = \mathbb{E}(X) - \overline{X} = 0$.
- La variable $\dfrac{X}{\sigma}$ **réduite** : $\mathbb{V}\left(\dfrac{X}{\sigma}\right) = \dfrac{1}{\sigma^2}\mathbb{V}(X) = \dfrac{\sigma^2}{\sigma^2} = 1$.
- Ainsi $T = \dfrac{X - \overline{X}}{\sigma}$ est la variable centrée réduite associée à X, avec $\mathbb{E}(X) = 0$, et $\mathbb{V}(T) = 1$.

En particulier, si $X \sim \mathcal{N}(m, \sigma^2)$, alors la VARC , centrée réduite associée est $T = \dfrac{X - \overline{X}}{\sigma}$ avec $T \sim \mathcal{N}(0, 1)$.

14.4 Moments simples d'ordre k de la VARC T $(k \in \mathbb{N}^*)$

Théorème 108

Pour $T \sim \mathcal{N}(0,1)$, on a les moments simples d'ordre k, avec :

$$\forall j \in \mathbb{N},\ \ m_{2j+1} = \mathbb{E}\big(T^{2j+1}\big) = 0 \text{ et } \forall j \in \mathbb{N}^*,\ \ m_{2j} = \mathbb{E}\big(T^{2j}\big) = 1 \times 3 \times \cdots \times (2j-1) = \frac{(2j)\,!}{2^j\ j\,!}.$$

Démonstration. Si $T \sim \mathcal{N}(0,1)$, alors sa fonction densité est $f : t \longmapsto \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}t^2}$. f est **paire** et vérifie : $f'(t) = -t\,f(t)$ sur $\mathbb{R}$.

Ainsi $\mathbb{E}\big(T^k\big) = \displaystyle\int_{-\infty}^{+\infty} t^k\,f(t)\,\mathrm{d}t$.

- Si k est **impair**, alors $k = 2j+1$ avec $j \in \mathbb{N}$, et $\mathbb{E}\big(T^k\big) = 0$ par imparité de l'intégrande.
- Si k est **pair**, alors $k = 2j$ avec $j \in \mathbb{N}^*$, et $\mathbb{E}\big(T^{2j}\big) = \displaystyle\int_{-\infty}^{+\infty} t^{2j}\,f(t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} t^{2j-1}\,t\,f(t)\,\mathrm{d}t = -\int_{-\infty}^{+\infty} t^{2j-1}\,f'(t)\,\mathrm{d}t$ car $f'(t) = -t\,f(t)$.

  En intégrant par parties : $m_{2j} = \mathbb{E}\big(T^{2j}\big) = -\Big[t^{2j-1}\,f(t)\Big]_{-\infty}^{+\infty} + (2j-1)\displaystyle\int_{-\infty}^{+\infty} t^{2j-2}\,f(t)\,\mathrm{d}t = (2j-1)\,m_{2(j-1)}$.

  Par récurrence sur j : $m_{2j} = (2j-1)\,m_{2(j-1)}$ assure que $m_{2j} = (2j-1) \times (2j-3) \times \cdots \times 3 \times m_2 = \dfrac{(2j)\,!}{2^j\ j\,!}$, car $m_2 = 1$. □

14.5 Courbe de la loi normale centrée réduite

On a vu que si $T \sim \mathcal{N}(0,1)$, sa fonction de densité est $f : t \longmapsto \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}t^2}$, qui est **positive et paire** et telle que $f'(t) = -t\,f(t)$ sur $\mathbb{R}$.

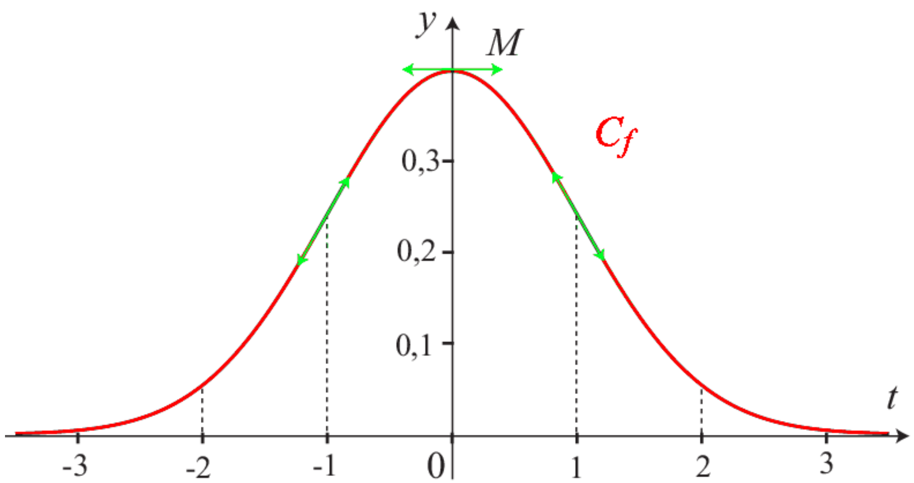
Donc f est croissante sur $]-\infty, 0]$ puis décroissante sur $[0, +\infty[$.

Elle admet un maximum en $t = 0$, avec $M = f(0) = \dfrac{1}{\sqrt{2\pi}} \approx 0,4$.

$\lim\limits_{t \to \pm\infty} f(t) = 0$, et la courbe admet l'axe des abscisses comme asymptote horizontale.

$f''(t) = \left(t^2 - 1\right) f(t)$ et on a donc deux points d'inflexion en $t = 1$ et $t = -1$, où f'' s'annule en changeant de signe.

On obtient la « **courbe en cloche** »,
dite aussi « courbe de Gauss ».



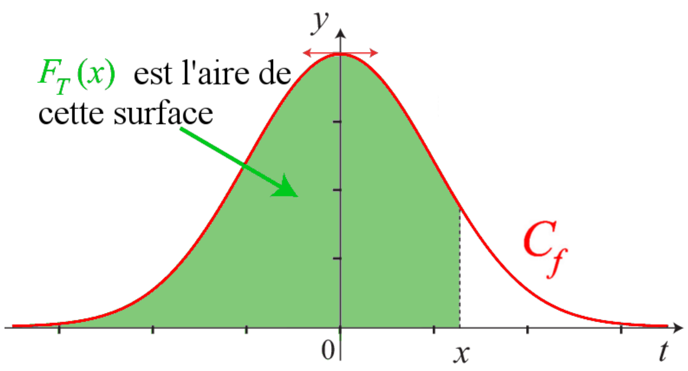
14.6 Fonction de répartition de T

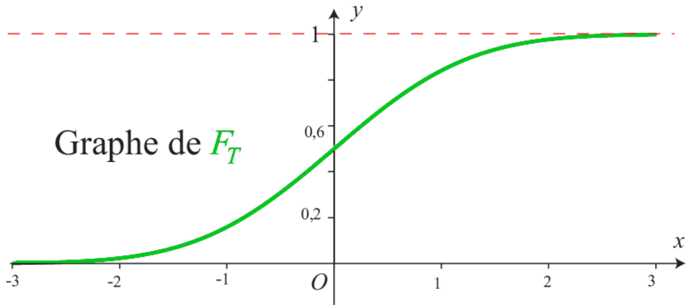
14.6.1 Tracé de la courbe

Par définition, la fonction de répartition de T est la fonction $F_T : \begin{cases} \mathbb{R} & \longrightarrow \quad \mathbb{R} \\ x & \longmapsto \quad F_T(x) = \mathbb{P}\big(T \leqslant x\big) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t \end{cases}$

On a $\lim\limits_{x \to -\infty} F_T(x) = 0$ et $\lim\limits_{x \to +\infty} F_T(x) = 1$.

F_T est continue et dérivable sur $\mathbb{R}$ de dérivée ${F_T}' = f > 0$, donc F_T est strictement croissante et bijective de $\mathbb{R}$ dans $]0,1[$.
Par parité de f, $F_T(0) = \dfrac{1}{2}\cdot$





14.6.2 Propriétés de F_T

Théorème 109

La fonction de répartition de la loi normale centrée réduite, vérifie les propriétés suivantes :

- $\forall x \in \mathbb{R},\ F_T(-x) = 1 - F_T(x)$.
- $\forall x \in \mathbb{R}_+,\ \mathbb{P}\big(|T| < x\big) = 2\,F_T(x) - 1$.

Les inégalités seront prises au sens large comme au sens strict puisque T est une VARC.

Démonstration. Par parité de f, on a : $\forall x \in \mathbb{R},\ \displaystyle\int_{-\infty}^{-x} f(t)\,\mathrm{d}t = \int_{x}^{+\infty} f(u)\,\mathrm{d}u$ avec le changement de variable $u = -t$.

Ainsi : $\forall x \in \mathbb{R},\ F_T(x) + F_T(-x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t + \int_{-\infty}^{-x} f(t)\,\mathrm{d}t = \int_{-\infty}^{x} f(t)\,\mathrm{d}t + \int_{x}^{+\infty} f(t)\,\mathrm{d}t = \int_{-\infty}^{+\infty} f(t)\,\mathrm{d}t = 1.$

De même : $\forall x \in \mathbb{R}_+,\ \mathbb{P}\big(|X| < x\big) = \mathbb{P}\big(-x < X < x\big) = F_T(x) - F_T(-x) = 2\,F_T(x) - 1.$

14.7 Fonction affine de la VARC normale centrée réduite

Pour $T \sim \mathcal{N}(0,1)$, on considère la VARC définie par $X = a\,T + b$, où $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$.
F_X et F_T sont respectivement les fonctions de répartition des VARC X et T. $\forall x \in \mathbb{R}$, $F_X(x) = \mathbb{P}\big(X \leqslant x\big) = \mathbb{P}\big(a\,T + b \leqslant x\big)$

$$\forall x \in \mathbb{R}, \begin{cases} \text{Si } a > 0 \quad F_X(x) = \mathbb{P}\Big(T \leqslant \dfrac{x-b}{a}\Big) = F_T\Big(\dfrac{x-b}{a}\Big) & \text{et } f_X(x) = {F_X}'(x) = \dfrac{1}{a}\, f_T\Big(\dfrac{x-b}{a}\Big) \\ \text{Si } a < 0 \quad F_X(x) = \mathbb{P}\Big(T \geqslant \dfrac{x-b}{a}\Big) = 1 - F_T\Big(\dfrac{x-b}{a}\Big) & \text{et } f_X(x) = {F_X}'(x) = -\dfrac{1}{a}\, f_T\Big(\dfrac{x-b}{a}\Big) \end{cases}$$

Théorème 110

La VARC $X = a\,T + b$, où $a \in \mathbb{R}^*$, $b \in \mathbb{R}$ et $T \sim \mathcal{N}(0,1)$ admet une fonction densité définie par :

$$f_X : x \longmapsto f_X(x) = \frac{1}{|a|}\, f_T\Big(\frac{x-b}{a}\Big) = \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{1}{2}\left(\frac{x-b}{a}\right)^2}$$

Donc si $T \sim \mathcal{N}(0,1)$, la VAR $X = a\,T + b$, où $a \in \mathbb{R}^*$, $b \in \mathbb{R}$ suit la loi $X \sim \mathcal{N}(b, |a|^2)$

De même, (en reprenant la démonstration) :
si $X \sim \mathcal{N}(m, \sigma^2)$, la VAR $Y = a\,X + b$, où $a \in \mathbb{R}^*$, $b \in \mathbb{R}$ suit la loi $Y \sim \mathcal{N}(a\,m + b, |a|^2\,\sigma^2)$

Ainsi il y a « stabilité » de la normalité par fonction aléatoire affine non constante

14.8 Convolution de lois normales, Théorème de Lévy-Cramer

Si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ sont des VARC **indépendantes**, alors :

leurs fonctions génératrices sont définies par $g_{X_1} : t \longmapsto \mathrm{e}^{t\,m_1 + \frac{1}{2}\,\sigma_1^2\,t^2}$ et $g_{X_2} : t \longmapsto \mathrm{e}^{t\,m_2 + \frac{1}{2}\,\sigma_2^2\,t^2}$,

donc $g_{X_1+X_2} = g_{X_1} \times g_{X_2} : t \longmapsto \mathrm{e}^{t\,(m_1+m_2) + \frac{1}{2}\,(\sigma_1^2+\sigma_2^2)\,t^2}$. Ainsi $\boxed{\mathcal{N}(m_1, \sigma_1^2) * \mathcal{N}(m_2, \sigma_2^2) = \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)}$

Théorème 111 (Théorème de Lévy-Cramer)

Si X_1 et X_2 sont deux variables aléatoires **indépendantes**, suivant des lois normales, $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$,
alors : $\quad X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.
$X_1 + X_2$ suit une loi normale, dont la moyenne est la somme des moyennes, et la variance, la somme des variances.

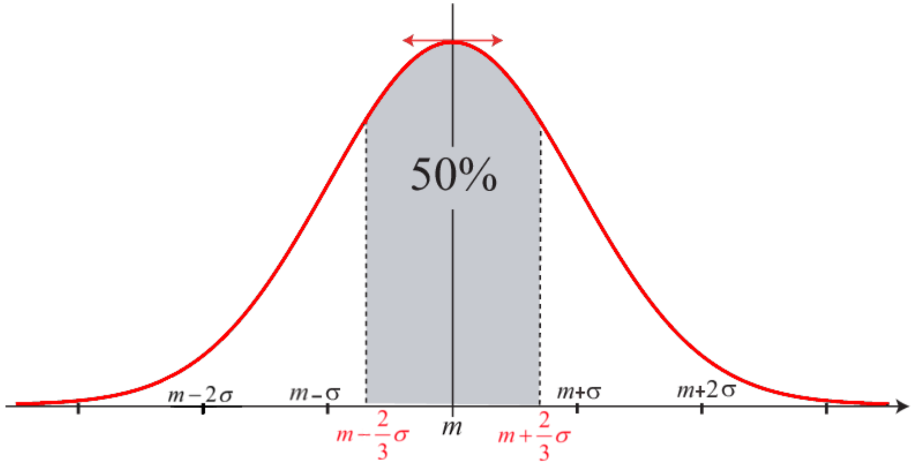
Ainsi il y a « stabilité » de la normalité par additivité indépendante de VARC Gaussiennes

14.9 Intervalles symétriques par rapport à la moyenne (fourchettes)

On se réfère fréquemment à des intervalles centrés sur la moyenne, dans lesquels va varier « probablement »une VARC X, en fonction de son écart-type, donc de sa dispersion.

Pour une variable X suivant une loi normale $\mathcal{N}(m, \sigma^2)$, par exemple, on se ramène à : $T \sim \mathcal{N}(0,1)$ en posant $T = \dfrac{X-m}{\sigma}$.

$$\begin{aligned}
\mathbb{P}\Big(m - \frac{2}{3}\sigma < X < m + \frac{2}{3}\sigma\Big) &= \mathbb{P}\Big(-\frac{2}{3} < \frac{X-m}{\sigma} < \frac{2}{3}\Big) \\
&= \mathbb{P}\Big(-\frac{2}{3} < T < \frac{2}{3}\Big) \\
&= \mathbb{P}\Big(\mid T \mid < \frac{2}{3}\Big) \\
&= 2F_T\left(\frac{2}{3}\right) - 1 \\
&= 0.5
\end{aligned}$$

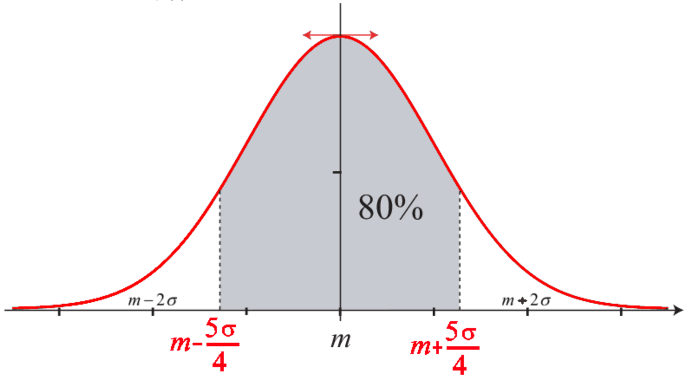


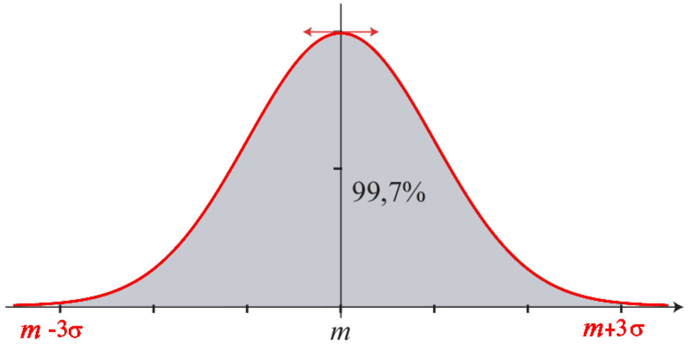
On estime ainsi qu'il y a 50% de chance que X soit dans une fourchette centrée sur m et de rayon les $2/3$ de l'écart-type σ, donc d'amplitude $\dfrac{4\sigma}{3}$.

De même : $\mathbb{P}\Big(m - \sigma < X < m + \sigma\Big) = \mathbb{P}\Big(-1 < T < 1\Big) \approx 0,68$ et $\mathbb{P}\Big(m - \frac{5}{4}\sigma < X < m + \frac{5}{4}\sigma\Big) = \mathbb{P}\Big(-1,25 < T < 1,25\Big) \approx 0,80$

$$\mathbb{P}\Big(m-2\,\sigma < X < m+2\,\sigma\Big) = \mathbb{P}\Big(-2 < T < 2\Big) \approx 0,95 \quad \text{et} \quad \mathbb{P}\Big(m-3\,\sigma < X < m+3\,\sigma\Big) = \mathbb{P}\Big(-3 < T < 3\Big) \approx 0,997$$

Ainsi « $6\,\sigma$ » est le surnom de ce dernier critère de qualité, avec une probabilité d'assurance de plus de 99, 7% donc une incertitude de moins de 3‰.





Chapitre 15

Loi Log-normale ou loi de Galton

15.1 Définition et fonction densité de la loi Log-normale

15.1.1 Définition

Définition 112

Soit $x_0 \in \mathbb{R}$, $m \in \mathbb{R}$ et $\sigma > 0$.
On dit qu'une VARC X définie sur $]x_0, +\infty[$ suit une **loi Log-normale** de paramètres m et σ si et seulement si la VARC $Y = \ln\big(X - x_0\big)$ suit une loi normale d'espérance m et de variance σ^2.
On note $X \sim \mathcal{L}\text{og}\mathcal{N}_{x_0}\big(m, \sigma^2\big)$.

Expression de X en fonction d'une VARC T suivant la loi normale standard :

$$\Big(Y = \ln\big(X - x_0\big)\Big) \Longleftrightarrow \Big(\mathrm{e}^Y = X - x_0\Big) \Longleftrightarrow \Big(X = x_0 + \mathrm{e}^Y\Big) \Longleftrightarrow \Big(X = x_0 + \mathrm{e}^{m+\sigma T}\Big) \quad \text{avec } \Big(T = \frac{Y - m}{\sigma}\Big) \Longleftrightarrow \Big(Y = m + \sigma T\Big)$$

Et donc $\Big(X \sim \mathcal{L}\text{og}\mathcal{N}_{x_0}(m,\sigma^2)\Big) \iff \Big(Y = \ln(X-x_0) \sim \mathcal{N}(m,\sigma^2)\Big)$

$\iff \Big(X = x_0 + \mathrm{e}^{m+\sigma T} \quad \text{où } T \sim \mathcal{N}(0,1)\Big)$

Ainsi $\boxed{(X \sim \mathcal{L}\text{og}\mathcal{N}_{x_0}(m,\sigma^2) \iff X = x_0 + \mathrm{e}^{m+\sigma T} \quad \text{où } T \sim \mathcal{N}(0,1)}$

15.1.2 Fonction densité de probabilité de X

On a : $\Big(Y = \ln(X-x_0)\Big) \iff \Big(X = x_0 + \mathrm{e}^{Y}\Big)$ avec $X\langle\Omega\rangle =]\,x_0, +\infty[$ et $Y\langle\Omega\rangle = \mathbb{R}$.

Étudions la **fonction de répartition** de X :

$\forall x \in]\,x_0,+\infty[,\quad F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}\Big(\mathrm{e}^Y + x_0 \leqslant x\Big) = \mathbb{P}\Big(\mathrm{e}^Y \leqslant x - x_0\Big) = \mathbb{P}\Big(Y \leqslant \ln(x-x_0)\Big)$

Donc $\forall x \in]\,x_0,+\infty[,\quad F_X(x) = F_Y\Big(\ln(x-x_0)\Big)$.

On obtient une **fonction densité** de probabilité en dérivant la fonction de répartition :

$$\begin{cases} \text{Si } x \leqslant x_0 \quad f_X(x) = {F_X}'(x) = 0 \\ \text{Si } x > x_0 \quad f_X(x) = {F_X}'(x) = \dfrac{1}{x-x_0}\cdot {F_Y}'\Big(\ln(x-x_0)\Big) \quad \text{avec } {F_Y}'(y) = f_Y(y) = \dfrac{1}{\sigma\sqrt{2\pi}}\cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{y-m}{\sigma}\right)^2} \end{cases}$$

Ainsi on obtient :

Théorème 113

Une VARC X suivant la loi $\mathcal{L}\text{og}\mathcal{N}_{x_0}(m,\sigma^2)$ admet une **fonction densité de probabilité** définie par :

$$f_X(x) = \begin{cases} \dfrac{1}{\sigma\sqrt{2\pi}}\dfrac{1}{(x-x_0)}\cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln(x-x_0)-m}{\sigma}\right)^2} & \text{Si } x > x_0 \\ 0 & \text{Si } x \leqslant x_0 \end{cases}$$

15.1.3 Principaux paramètres

Remarque : la deuxième fonction génératrice d'une loi Log-normale n'est définie que sur $\mathbb{R}_-$.
N'étant pas définie sur un voisinage ouvert autour de 0, on ne peut donc pas déterminer directement les moments d'ordre k à partir de celle-ci.

Moment d'ordre $k \in \mathbb{N}^*$: $\mathbb{E}\Big[\big(X - x_0\big)^k\Big] = \displaystyle\int_{x_0}^{+\infty} (x - x_0)^k \times \frac{1}{x - x_0} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln(x-x_0)-m}{\sigma}\right)^2} \mathrm{d}x$

Par changement de variable : $\left(u = \dfrac{\ln(x - x_0) - m}{\sigma}\right) \Longleftrightarrow \left(x = x_0 + \mathrm{e}^{\sigma\, u + m}\right)$, avec $\mathrm{d}x = \sigma\, \mathrm{e}^{\sigma\, u+m}\, \mathrm{d}u$ on obtient :

$$\forall k \in \mathbb{N}^* \qquad \mathbb{E}\Big[\big(X - x_0\big)^k\Big] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{1}{2}u^2} \cdot \mathrm{e}^{(\sigma\, u+m)\, k}\, \mathrm{d}u = \frac{\mathrm{e}^{k\, m}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{k\, \sigma\, u - \frac{1}{2}u^2}\, \mathrm{d}u = \mathrm{e}^{k\, m + \frac{1}{2}\sigma^2\, k^2} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}(u-k\sigma)^2}\, \mathrm{d}(u - k\sigma)$$

En effet : $k\, \sigma\, u - \frac{1}{2}u^2 = \dfrac{k^2\, \sigma 2}{2} - \dfrac{1}{2}(u - k\, \sigma)^2$

Enfin : $\dfrac{1}{\sqrt{2\pi}} \cdot \displaystyle\int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{1}{2}v^2}\, \mathrm{d}v = 1$ avec $v = u - k\sigma$

Ainsi : $\boxed{\forall k \in \mathbb{N}^* \qquad \mathbb{E}\Big[\big(X - x_0\big)^k\Big] = \mathrm{e}^{k\, m + \frac{1}{2}\sigma^2\, k^2}}$

En particulier :

Pour $k = 1$: $\mathbb{E}\big[\big(X - x_0\big)\big] = \mathrm{e}^{m + \frac{1}{2}\sigma^2} = \mathbb{E}(X) - x_0$

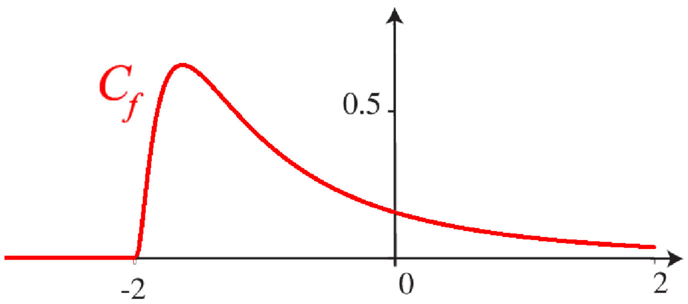
Pour $k = 2$: $\mathbb{E}\Big[\big(X - x_0\big)^2\Big] = \mathrm{e}^{2\, m + 2\sigma^2}$ et $\mathbb{V}(X) = \mathbb{V}\big(X - x_0\big) = E\Big[\big(X - x_0\big)^2\Big] - \Big[E\big(X - x_0\big)\Big]^2 = \mathrm{e}^{2m+2\sigma^2} - \mathrm{e}^{2m+\sigma^2} = \mathrm{e}^{2m+\sigma^2}\big(\mathrm{e}^{\sigma^2} - 1\big)$

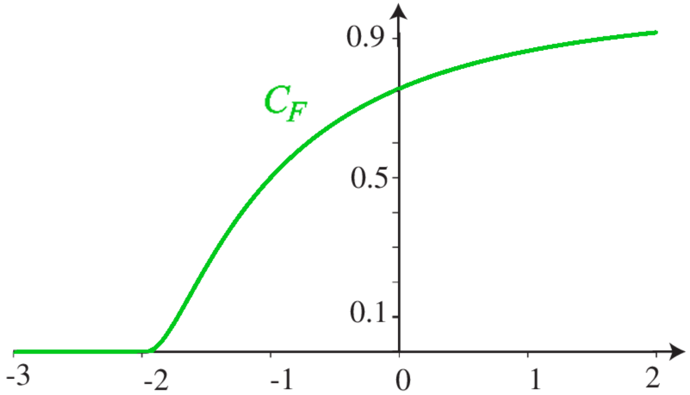
Théorème 114

Pour une VAR X suivant la loi $\mathcal{L}og\mathcal{N}_{x_0}\big(m, \sigma^2\big)$, on a : $\mathbb{E}\big(X\big) = x_0 + \mathrm{e}^{m + \frac{1}{2}\sigma^2}$ et $\mathbb{V}\big(X\big) = \mathrm{e}^{2m+\sigma^2}\big(\mathrm{e}^{\sigma^2} - 1\big)$

15.1.4 Représentations graphiques

Pour une VAR X suivant la loi $\mathcal{L}\text{og}\mathcal{N}_{x_0}(m, \sigma^2)$, avec $x_0 = -2$, $m = 0$ et $\sigma = 1$, on a :





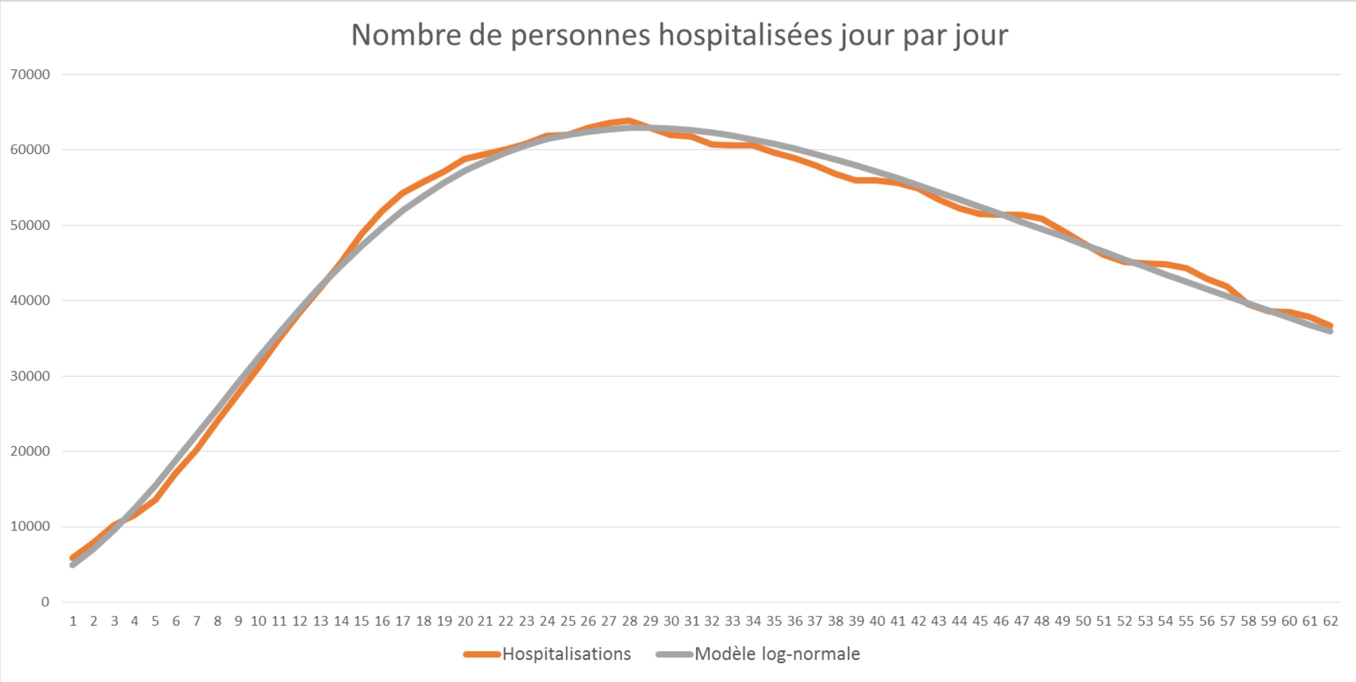
15.1.5 Exemples d'applications

Divers domaines d'application

• En biologie, on utilise la loi Log-normale pour modéliser la distribution des masses d'organismes vivants.
• En mécanique des fluides, pour modéliser la distribution des tailles de gouttes d'un jet pulvérisé.
• En épidémiologie, le délai entre l'infection d'une personne et son décès suit une loi Log-normale (*source Journal of Clinical Medicine*).

Exemple des hospitalisations dues à la Covid19 (groupe de Recherches du département Mathématiques IMT Mines Alès)

Voici deux courbes, l'une correspond au nombre d'hospitalisations, jour par jour, du 18 mars 2020 jusqu'au 18 mai 2020, et l'autre, est la fonction densité d'une loi de Galton. Une optimisation par réseau de neurones a permis de déterminer x_0 ($x_0 = -8,5$) et les deux paramètres de cette loi. L'erreur relative moyenne de l'adéquation sur cette période est inférieure à 3%.
C'est un outil d'aide à la décision car elle permet de prédire la date à partir de laquelle le nombre de personnes hospitalisées passera en dessous d'un certain seuil fixé. (capacité d'accueil des hôpitaux)



Le groupe de recherche ETE (Evolution Théorique et Expérimentale) de Montpellier a de son côté mis en évidence plusieurs résultats :
• Le coefficient force d'infection individuelle (nombre moyen d'infectés secondaires par cas) initialement considéré constant à $R_0 = 3$ suit une **loi gamma** de paramètre de forme $k = 0,6$ et de moyenne R_0 (hétérogénéité des individus quant à leurs contacts en société)
• Le nombre d'infections causées par une personne en un jour, suit une loi binomiale négative de paramètre de dispersion k, et moyenne R_0.

Source : `https://covid-ete.ouvaton.org/Rapport8_stochasticite.html#extinction_de_l'épidémie`

Dans cette étude, interviennent les lois de probabilités suivantes :
lois discrètes : Binomiale négative, Poisson
lois continues : Gamma, Log-normale

Quatrième partie

Développements en Probabilités

Chapitre 16

Notions de convergence

16.1 Convergence en probabilité

16.1.1 Définition et condition suffisante de convergence en probabilité

Définition 115

On dit qu'une suite $\left(X_i\right)_{i\in\mathbb{N}^*}$ de VAR **converge en probabilité** vers la VAR X si et seulement si :

$$\forall \varepsilon > 0,\ \lim_{i\to+\infty} \mathbb{P}\Big(\big|X_i - X\big| \geqslant \varepsilon\Big) = 0.$$

ou encore : $$\forall \varepsilon > 0,\ \lim_{i\to+\infty} \mathbb{P}\Big(X - \varepsilon < X_i < X + \varepsilon\Big) = 1.$$

Lorsqu'il y a convergence, on note : $X_i \xrightarrow[i\to+\infty]{\mathbb{P}} X$.

Remarque : La convergence en probabilité cherche à décrire la proximité des variables aléatoires réelles, elles doivent donc toutes être définies **sur un même espace probabilisé**.

Théorème 116 (Condition Suffisante)

Pour que $X_i \xrightarrow[i\to+\infty]{\mathbb{P}} X$, il suffit que : $\lim\limits_{i\to+\infty}\left(\mathbb{E}\left[\left(X_i - X\right)^2\right]\right) = 0$.

Démonstration. D'après l'inégalité de Bienaymé-Tchebychev, on a : $\forall \varepsilon > 0 \ : \ 0 \leqslant \mathbb{P}\left(\left|X_i - X\right| > \varepsilon\right) \leqslant \dfrac{\mathbb{E}\left[\left(X_i - X\right)^2\right]}{\varepsilon^2}$.
Donc si $\lim\limits_{i\to+\infty}\left(\mathbb{E}\left[\left(X_i - X\right)^2\right]\right) = 0$ alors $\forall \varepsilon > 0, \ \lim\limits_{i\to+\infty}\mathbb{P}\left(\left|X_i - X\right| \geqslant \varepsilon\right) = 0$. □

Remarque : On vient de démontrer que la convergence en moyenne quadratique entraîne la convergence en probabilité.

16.1.2 Loi faible des grands nombres

Théorème 117

Pour $m \in \mathbb{R}$ et $\sigma > 0$, on considère :
$\left(X_i\right)_{i\in\mathbb{N}^*}$ une suite de VAR **indépendantes**, suivant toutes une **même loi** ($\forall i \in \mathbb{N}^*, \ X_i \sim \mathcal{L}$), de **même espérance** ($\forall i \in \mathbb{N}^*, \ \mathbb{E}\left(X_i\right) = m$) et de **même variance** ($\forall i \in \mathbb{N}^*, \ \mathbb{V}\left(X_i\right) = \sigma^2$), alors :

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{\mathbb{P}} m \qquad \text{(au sens de la variable aléatoire constante égale à } m\text{).}$$

Démonstration. On pose $Z_n = \sum\limits_{i=1}^{n} X_i$, ainsi $\dfrac{Z_n}{n} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.
Alors $\mathbb{E}\left(\dfrac{Z_n}{n}\right) = m$ par linéarité, et $\mathbb{V}\left(\dfrac{Z_n}{n}\right) = \dfrac{\sigma^2}{n}$, car les X_i sont indépendantes.
Alors $\lim\limits_{n\to+\infty}\left(\mathbb{E}\left[\left(\dfrac{Z_n}{n} - m\right)^2\right]\right) = \lim\limits_{n\to+\infty}\mathbb{V}\left(\dfrac{Z_n}{n}\right) = \lim\limits_{n\to+\infty}\dfrac{\sigma^2}{n} = 0$, ce qui assure la condition suffisante précédente et le résultat escompté. □

16.1.3 Application en Statistiques

On peut préciser, toujours avec l'inégalité de Bienaymé-Tchebychev, que : $\forall \varepsilon > 0, \quad \mathbb{P}\left(\left|\frac{Z_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\,\varepsilon^2}$.

Si on se donne $\mu > 0$, pour obtenir $\mathbb{P}\left(\left|\frac{Z_n}{n} - m\right| \geqslant \varepsilon\right) \leqslant \mu$, $\boxed{\text{il suffit alors de choisir } n \geqslant n_0 = \left\lfloor \frac{\sigma^2}{\mu\,\varepsilon^2} \right\rfloor + 1.}$

Ainsi en **Statistiques inférentielles**, on est amené, dans les cas où σ est connu ou estimé, pour assurer une précision $\varepsilon > 0$ donnée, et un niveau de confiance $1 - \mu$ (avec $\mu \in]0, 1[$) fixé, à choisir $n \geqslant n_0 = \left\lfloor \frac{\sigma^2}{\mu\,\varepsilon^2} \right\rfloor + 1$ pour assurer : $\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - m\right| < \varepsilon\right) \geqslant 1 - \mu$

16.2 Convergence en loi

16.2.1 Définition et cas discret

Définition 118

On dit qu'une suite $\left(X_i\right)_{i\in\mathbb{N}^*}$ de VAR **converge en loi** vers la VAR X si et seulement si, en notant F_{X_i} et F_X leurs fonctions de répartition respectives, on a :

$$\forall x \in \mathbb{R} \ \text{ où } F_X \ \text{ est continue }, \ \lim_{i\to+\infty} F_{X_i}(x) = F_X(x).$$

Lorsqu'il y a convergence, on note : $X_i \xrightarrow[i\to+\infty]{L} X$.

Remarque : La convergence en loi cherche à décrire la proximité des lois suivies par les variables réelles, **ces dernières ne sont donc par forcément définies sur un même espace probabilisé**. On peut aussi remplacer toute VAR par une autre VAR sans modifier la convergence en loi tant que ces deux VAR ont la même loi.

Théorème 119 (Application directe)

Soient $\left(X_i\right)_{i\in\mathbb{N}^*}$ une suite de VAR et X est une VARC.
Si $X_i \xrightarrow[i\to+\infty]{L} X$, alors $\forall(a,b)\in\mathbb{R}^2,\quad \lim\limits_{i\to+\infty}\Big(\mathbb{P}\big(a<X_i\leqslant b\big)\Big)=\mathbb{P}\big(a<X\leqslant b\big)$

Démonstration. $\forall i\in\mathbb{N}^*$, $\mathbb{P}\big(a<X_i\leqslant b\big)=F_{X_i}(b)-F_{X_i}(a)$ et $\mathbb{P}\big(a<X\leqslant b\big)=F_X(b)-F_X(a)$
F_X étant continue sur $\mathbb{R}$, on obtient le résultat voulu. □

Théorème 120 (Cas discret à valeurs entières)

Si pour tout $i\in\mathbb{N}^*$, X_i et X sont des VARD à valeurs dans $\mathbb{Z}$, alors :
$X_i \xrightarrow[i\to+\infty]{L} X$ si et seulement si $\forall k\in\mathbb{Z},\quad \lim\limits_{i\to+\infty}\Big(\mathbb{P}\big(X_i=k\big)\Big)=\mathbb{P}\big(X=k\big)$.

Remarque : On veillera à ne pas généraliser ce théorème lorsque les X_i et X sont des VARC : la convergence en loi n'équivaut pas à la convergence des densités.

16.2.2 Théorème Central Limite

Pour $m\in\mathbb{R}$ et $\sigma>0$, on considère : $\left(X_i\right)_{i\in\mathbb{N}^*}$ une suite de VAR **indépendantes**, suivant une **même loi** ($\forall i\in\mathbb{N}^*,\ X_i\sim\mathcal{L}$), de **même espérance** ($\forall i\in\mathbb{N}^*,\ \mathbb{E}\big(X_i\big)=m$) et de **même variance** ($\forall i\in\mathbb{N}^*,\ \mathbb{V}\big(X_i\big)=\sigma^2$).
On dit que les variables aléatoires sont i.i.d.(indépendantes et identiquement distribuées)

On note Z_n la somme $Z_n=\sum\limits_{i=1}^{n}X_i$, et on considère la fréquence (ou moyenne des X_i) : $\dfrac{Z_n}{n}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$.
Leurs paramètres sont $\mathbb{E}(Z_n)=n\,m$ et $\mathbb{E}\left(\dfrac{Z_n}{n}\right)=m$ par linéarité, alors que $\mathbb{V}(Z_n)=n\,\sigma^2$ et $\mathbb{V}\left(\dfrac{Z_n}{n}\right)=\dfrac{\sigma^2}{n}$, car les X_i sont indépendantes.

On a ainsi les **variables centrées réduites** de Z_n et $\dfrac{Z_n}{n}$:

• $Y_n = \dfrac{Z_n - \mathbb{E}(Z_n)}{\sigma(Z_n)} = \dfrac{Z_n - n\,m}{\sigma\,\sqrt{n}}$, • $W_n = \dfrac{\dfrac{Z_n}{n} - \mathbb{E}\left(\dfrac{Z_n}{n}\right)}{\sigma\left(\dfrac{Z_n}{n}\right)} = \dfrac{\dfrac{Z_n}{n} - m}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{Z_n - n\,m}{\sigma\,\sqrt{n}}$ • On a donc $Y_n = W_n = \dfrac{Z_n - n\,m}{\sigma\,\sqrt{n}}$.

Théorème 121 (Théorème Central Limite : TCL)

Sous les hypothèses ci-dessus, la **somme centrée réduite** Y_n, ainsi que la **moyenne centrée réduite** W_n convergent en loi vers une VAR qui suit la loi normale centrée réduite $\mathcal{N}(0,1)$, lorsque $n \to +\infty$.

$$Y_n = W_n \xrightarrow[n\to+\infty]{L} T \text{ où } T \sim \mathcal{N}(0,1)$$

La démonstration, technique, est admise, mais ce **résultat fondamental** souligne le **caractère UNIVERSEL de la loi normale**.
En Statistiques, on estime en général que :

pour $n \geqslant 30$, on obtient une approximation légitime des lois de Y_n ou de W_n par la loi $\mathcal{N}(0,1)$.

Ainsi, pour $n \geq 30$: $\forall (a,b) \in \mathbb{R}^2$ avec $a < b$, $\mathbb{P}\left(a < Y_n \leqslant b\right) \approx \mathbb{P}(a < T \leqslant b)$ (idem pour W_n).

De même pour $n \geqslant 30$ on approche la loi de $\dfrac{Z_n}{n}$ par la loi $\mathcal{N}(m, \dfrac{\sigma^2}{n})$ et la loi de Z_n par la loi $\mathcal{N}(n\,m, n\,\sigma^2)$.

Dès qu'un échantillon d'une population, constitué d'individus indépendants et identiquement distribués, est **assez important**, même sans identifier la loi suivie, on peut, pour **la moyenne arithmétique**, dire qu'elle suit, à peu près, la **loi normale** centrée et réduite.
C'est parce que le TCL n'exige pas d'hypothèses sur la loi de probabilité suivie par chaque VAR, hormis celle d'une variance finie, qu'il se révèle si indispensable aux statistiques.
En Statistiques inférentielles, le TCL permet le calcul des **intervalles de confiance** autour des **estimateurs**.

16.3 Approximations usuelles entre lois

On peut démontrer un certain nombre de **résultats d'approximation** de certaines lois par d'autres, plus pratiques à calculer, ou qui permettent de nouvelles interprétations.
Ces approximations numériques, pour être pertinentes, ont des conditions de validité.
A l'aide de tables, elles permettaient alors d'établir des résultats difficiles à obtenir par un calcul direct.
Aujourd'hui, la performance des outils numériques minimise l'intérêt pratique de ces approximations.

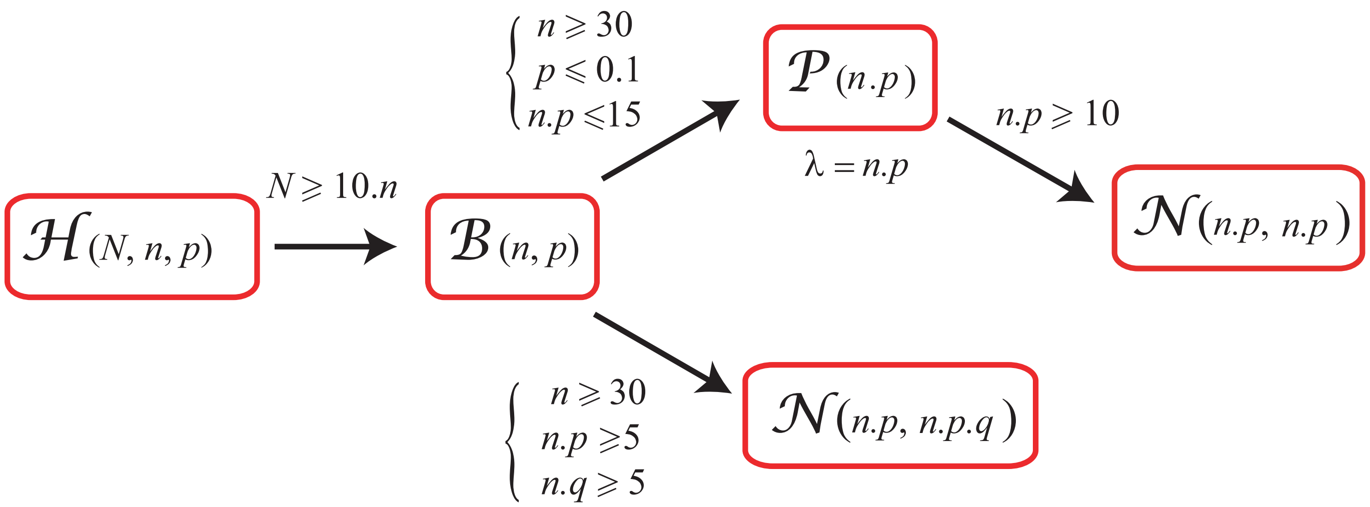
Il s'agit toujours d'approcher une VAR suivant une loi par une VAR de même espérance et de même variance.

16.3.1 Approximations usuelles

Théorème 122 (Approximations usuelles entre lois)

- Pour $N \geqslant 10\,n$, on peut approcher la loi hypergéométrique $\mathcal{H}(N, n, p)$ par la loi binomiale $\mathcal{B}(n, p)$.
- Pour $n \geqslant 30$, $p \leqslant 0.1$ et $n\,p \leqslant 15$, on peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de Poisson $\mathcal{P}(\lambda)$ où $\lambda = n\,p$.
- Pour $n\,p \geqslant 10$, on peut approcher la loi de Poisson $\mathcal{P}(\lambda)$ par la loi normale $\mathcal{N}(\lambda, \lambda)$.
- Pour $n \geqslant 30$, $n\,p \geqslant 5$ et $n\,q \geqslant 5$, on peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(n.p, n.p.q)$.

Remarque : les plages de convergence peuvent varier selon les référentiels sur lesquels on travaille.



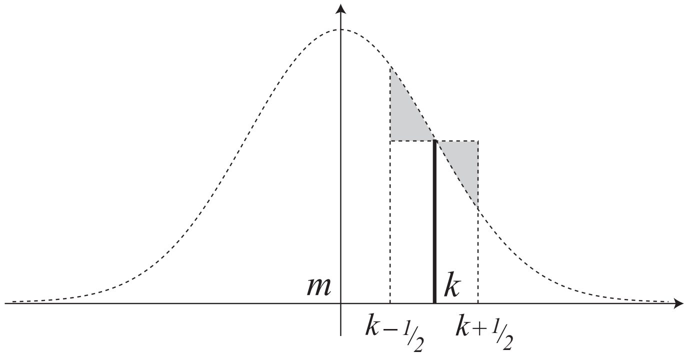
16.3.2 Correcteur de continuité

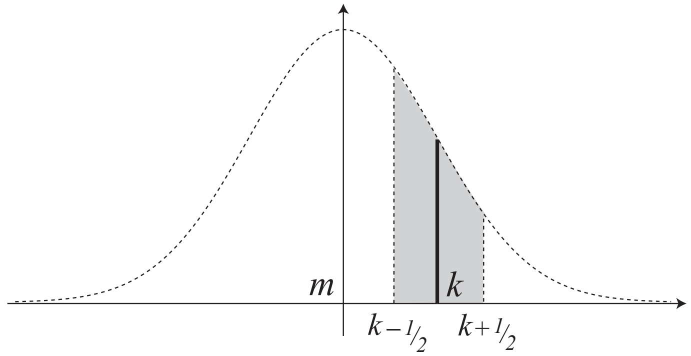
Lorsqu'on approche une loi discrète que nous noterons $\mathcal{D}$ (finie, comme une Loi Binomiale, ou dénombrable, comme une Loi de Poisson) par une loi continue que nous noterons $\mathcal{C}$ (souvent une Loi Normale), il est nécessaire de remplacer une valeur ponctuelle $\mathbb{P}_{\mathcal{D}}(X = k)$ par une probabilité sur un intervalle $\mathbb{P}_{\mathcal{C}}(a_k < X < b_k)$. Le choix de a_k et b_k en fonction de k est ce qu'on appelle un **correcteur de continuité**.

Théorème 123

On prend en général $\mathbb{P}\Big(X_{\mathcal{D}} = k\Big) \approx \mathbb{P}\Big(k - \frac{1}{2} < X_{\mathcal{C}} < k + \frac{1}{2}\Big)$.

On choisit un intervalle de longueur 1, ce qui amène à approcher la surface du rectangle de hauteur $\mathbb{P}\Big(X_{\mathcal{D}} = k\Big)$ par l'intégrale de la fonction de densité de $X_\mathcal{C}$ sur $\left[k-\dfrac{1}{2}, k+\dfrac{1}{2}\right]$.





Théorème 124

On aura donc, pour une VARD $X_\mathcal{D}$, à valeurs entières consécutives, approchée par une VARC $X_\mathcal{C}$:

- $\mathbb{P}\Big(X_\mathcal{D} < k\Big) \approx \mathbb{P}\Big(X_\mathcal{C} < k - \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(X_\mathcal{D} \leqslant k\Big) \approx \mathbb{P}\Big(X_\mathcal{C} < k + \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(X_\mathcal{D} > k\Big) \approx \mathbb{P}\Big(X_\mathcal{C} > k + \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(X_\mathcal{D} \geqslant k\Big) \approx \mathbb{P}\Big(X_\mathcal{C} > k - \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(a \leqslant X_\mathcal{D} \leqslant b\Big) \approx \mathbb{P}\Big(a - \dfrac{1}{2} < X_\mathcal{C} < b + \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(a < X_\mathcal{D} < b\Big) \approx \mathbb{P}\Big(a + \dfrac{1}{2} < X_\mathcal{C} < b - \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(a \leqslant X_\mathcal{D} < b\Big) \approx \mathbb{P}\Big(a - \dfrac{1}{2} < X_\mathcal{C} < b - \dfrac{1}{2}\Big)$
- $\mathbb{P}\Big(a < X_\mathcal{D} \leqslant b\Big) \approx \mathbb{P}\Big(a + \dfrac{1}{2} < X_\mathcal{C} < b + \dfrac{1}{2}\Big)$

Chapitre 17

Vecteurs aléatoires continus en dimension 2

17.1 Couple aléatoire de variables aléatoires réelles

17.1.1 Tribu borélienne de $\mathbb{R}^2$

On a vu, Chapitre 1, paragraphe 1. 2. 5., que $\mathcal{B}_{\mathbb{R}}$ la tribu borélienne de $\mathbb{R}$, est engendrée par les sous-ensembles de $\mathbb{R}$, demi-droites ouvertes inachevées : $\mathcal{B}_x =]-\infty, x[$ où $x \in \mathbb{R}$. On obtient la même tribu, en l'engendrant par les $]-\infty, x]$ ou les $[a, b]$.
La tribu borélienne de $\mathbb{R}^2$, notée $\mathcal{B}_{\mathbb{R}^2}$, est de même engendrée par les sous-ensembles $\mathcal{B}_{x,y} =]-\infty, x[\times]-\infty, y[$ où $(x, y) \in \mathbb{R}^2$, ou par les $]-\infty, x]\times]-\infty, y]$ ou les $[a, b] \times [c, d]$.

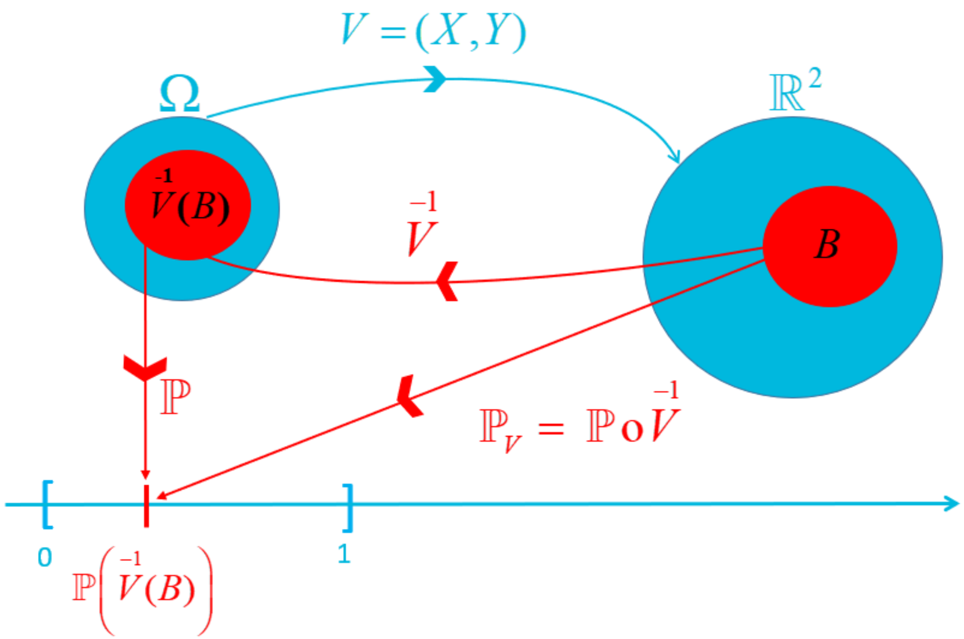
17.1.2 Couple aléatoire

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition 125

La fonction V définie de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}^2})$ est un **couple aléatoire** réel si et seulement si l'image réciproque de tout élément de la tribu borélienne de $\mathbb{R}^2$ est un élément de $\mathcal{A}$.

Autrement dit, $V = (X, Y) : \begin{cases} (\Omega, \mathcal{A}) & \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}^2}) \\ \omega & \longmapsto \big(X(\omega), Y(\omega)\big) \end{cases}$ est telle que $\forall B \in \mathcal{B}_{\mathbb{R}^2},\ \overset{-1}{V}(B) \in \mathcal{A}$



On peut alors définir la **loi de V** sous la probabilité $\mathbb{P}$:

$$\boxed{\forall B \in \mathcal{B}_{\mathbb{R}^2}, \mathbb{P}_V(B) = \mathbb{P}\left(\overset{-1}{V}(B)\right) = \mathbb{P}\Big(\big\{\omega \in \Omega,\ V(\omega) \in B\big\}\Big)}$$

Définition 126

La **fonction de répartition conjointe** de V (ou simultanée de X et Y), notée F_V, est ainsi définie par :

$$\forall (x,y) \in \mathbb{R}^2, \quad F_V(x,y) = \mathbb{P}_V\Big(]-\infty, x] \times]-\infty, y]\Big) = \mathbb{P}\Big(\overset{-1}{V}(]-\infty, x] \times]-\infty, y])\Big) = \mathbb{P}\Big(\big\{\omega \in \Omega,\ X(\omega) \leqslant x \text{ et } Y(\omega) \leqslant y\big\}\Big)$$

Définition 127

Les lois de X et de Y sont dites **lois marginales du couple** V, et on a :

$$\forall x \in \mathbb{R}, \quad F_X = \mathbb{P}_X\big(]-\infty, x]\big) = \mathbb{P}\Big(\big\{\omega \in \Omega,\ X(\omega) \leqslant x\big\}\Big) = \mathbb{P}\Big(\overset{-1}{V}(]-\infty, x] \times]-\infty, +\infty[)\Big) = \mathbb{P}_V\Big(]-\infty, x] \times]-\infty, +\infty[\Big)$$

$$\forall y \in \mathbb{R}, \quad F_Y = \mathbb{P}_Y\big(]-\infty, y]\big) = \mathbb{P}\Big(\big\{\omega \in \Omega,\ Y(\omega) \leqslant y\big\}\Big) = \mathbb{P}\Big(\overset{-1}{V}(]-\infty, +\infty[\times]-\infty, y])\Big) = \mathbb{P}_V\Big(]-\infty, +\infty[\times]-\infty, y]\Big)$$

Il en découle :

Proposition 128

Avec les notations précédentes, on a les **relations suivantes entre les fonctions de répartition :**

$$\forall x \in \mathbb{R}, \quad F_X(x) = \lim_{y \to +\infty} F_V(x,y)$$

$$\forall y \in \mathbb{R}, \quad F_Y(y) = \lim_{x \to +\infty} F_V(x,y)$$

17.1.3 Exemple 1

Exemple : On jette une fléchette au hasard sur une cible, le disque unité $D = \big\{(x,y) \in \mathbb{R}^2,\ x^2+y^2 \leqslant 1\big\}$.

ω est un jet, Ω l'ensemble des jets possibles.

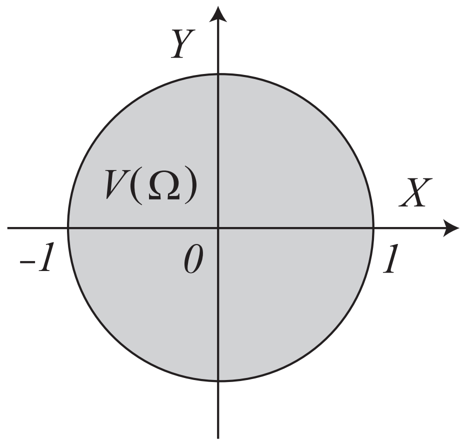
$$\begin{array}{rccl} V & : & \Omega & \longrightarrow & D \\ & & \omega & \longmapsto & V(\omega) = \big(X(\omega), Y(\omega)\big) \end{array}$$

$V(\Omega)$ est l'ensemble des impacts sur la cible : D.

On a $X(\Omega) = Y(\Omega) = [-1, 1]$

On distinguera $V(\Omega)$ et $X(\Omega) \times Y(\Omega)$ qui est un carré :

$V(\Omega) \subsetneq X(\Omega) \times Y(\Omega)$



17.2 Couple aléatoire à densité

17.2.1 Fonction densité

Définition 129

On appelle **fonction densité de probabilité** sur $\mathbb{R}^2$ toute application f de $\mathbb{R}^2$ dans $\mathbb{R}$ telle que :

1. $\forall (x,y) \in \mathbb{R}^2,\ \ f(x,y) \geqslant 0$,
2. f est continue sur $\mathbb{R}^2$, sauf éventuellement sur un ensemble de mesure nulle,
3. f est sommable (ou d'intégrale convergente) sur $\mathbb{R}^2$, avec $\displaystyle\iint_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\mathrm{d}y = 1$.

17.2.2 Couple aléatoire à densité de probabilité

Définition 130

Un couple de VARC, $V = (X, Y)$, est dit **couple aléatoire à densité** ou **absolument continu**, (ou encore **conjointement continu**) si et seulement s'il existe une fonction densité de probabilité f telle que :

$$\forall (a, b, c, d) \in \mathbb{R}^4, \quad \mathbb{P}\Big(a \leqslant X \leqslant b \text{ et } c \leqslant Y \leqslant d\Big) = \iint_{[a,b]\times[c,d]} f(x, y)\,\mathrm{d}x\mathrm{d}y$$

f est alors appelée une **densité de probabilité conjointe** de (X, Y).

Un tel couple est aussi appelé **couple aléatoire réel continu (CARC)**.

17.3 Propriétés des couples à densité

17.3.1 Fonction de répartition conjointe, densité de probabilité conjointe

Soit $V = (X, Y)$ un CARC admettant une fonction de répartition F_V et une densité de probabilité f :

Théorème 131

$$\forall (x, y) \in \mathbb{R}^2, \quad F_V(x, y) = F_{(X,Y)}(x, y) = \mathbb{P}\Big(X \leqslant x \text{ et } Y \leqslant y\Big) = \int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(u, v)\,\mathrm{d}v\right]\mathrm{d}u$$

Par dérivation d'une intégrale fonction de sa borne supérieure :

Théorème 132

$$\forall (x,y) \in \mathbb{R}^2, \quad f(x,y) = \frac{\partial^2 F_{(X,Y)}}{\partial x\, \partial y}\big(x,y\big) \quad \text{là où } f \text{ est continue.}$$

17.3.2 Lois marginales de X et Y

Soit $V = (X,Y)$ un CARC admettant une fonction de répartition $F_{(X,Y)}$ et une densité de probabilité f :

Théorème 133

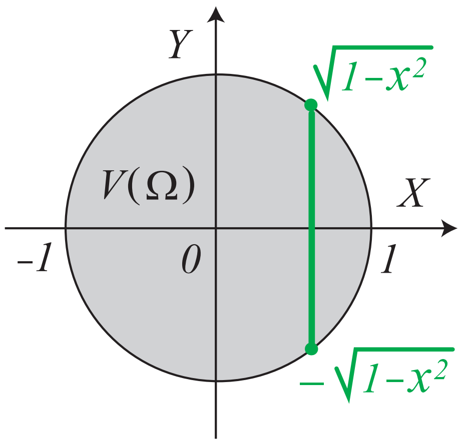
- X admet pour **fonction de répartition**, la fonction F_X définie par :
  $\forall x \in \mathbb{R}, \quad F_X(x) = \mathbb{P}\Big(X \leqslant x \text{ et } Y \in \mathbb{R}\Big) = F_{(X,Y)}(x,+\infty) = \displaystyle\int_{-\infty}^{x} \left[\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}v\right] \mathrm{d}u$
- X admet une **densité marginale de probabilité**, la fonction f_X définie par :
  $\forall x \in \mathbb{R}, \quad f_X(x) = F'_X(x) = \displaystyle\int_{-\infty}^{+\infty} f(x,v)\,\mathrm{d}v$
- $\mathbb{E}(X) = \displaystyle\int_{-\infty}^{+\infty} u\, f_X(u)\,\mathrm{d}u = \int_{-\infty}^{+\infty} u \left[\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}v\right] \mathrm{d}u = \iint_{\mathbb{R}^2} u\, f(u,v)\,\mathrm{d}u\mathrm{d}v$
- Y admet pour **fonction de répartition**, la fonction F_Y définie par :
  $\forall y \in \mathbb{R}, \quad F_Y(y) = \mathbb{P}\Big(X \in \mathbb{R} \text{ et } Y \leqslant y\Big) = F_{(X,Y)}(+\infty,y) = \displaystyle\int_{-\infty}^{y} \left[\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}u\right] \mathrm{d}v$
- Y admet une **densité marginale de probabilité**, la fonction f_Y définie par :
  $\forall y \in \mathbb{R}, \quad f_Y(y) = F'_Y(y) = \displaystyle\int_{-\infty}^{+\infty} f(u,y)\,\mathrm{d}u$
- $\mathbb{E}(Y) = \displaystyle\int_{-\infty}^{+\infty} v\, f_Y(v)\,\mathrm{d}v = \int_{-\infty}^{+\infty} v \left[\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}u\right] \mathrm{d}v = \iint_{\mathbb{R}^2} v\, f(u,v)\,\mathrm{d}u\mathrm{d}v$

17.3.3 Suite de l'exemple 1 : lancer de fléchettes sur une cible

- Pour obtenir une équiprobabilité, munissons $V(\Omega)$ d'une **densité constante** (loi uniforme sur D)
  Définissons alors f par : $\forall (x,y) \in D, \quad f(x,y) = K > 0$ et $\forall (x,y) \notin D, \quad f(x,y) = 0$.
  f vérifie les conditions suivantes :
  - $f(x,y) \geq 0$ pour tout (x,y) de $\mathbb{R}^2$.
  - f est continue sur $\mathbb{R}^2$ sauf sur le cercle unité (surface de mesure nulle)
  - $\left(\iint_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\mathrm{d}y = 1 \right) \iff \left(K \iint_D \mathrm{d}x\mathrm{d}y = 1 \right) \iff \left(K = \frac{1}{\pi} \right)$ (car l'aire du disque vaut π).

  Choisissons donc $\boxed{K = \frac{1}{\pi}.}$

- **Espérances** :
  $\mathbb{E}(X) = \iint_{\mathbb{R}^2} x\, f(x,y) \mathrm{d}x\mathrm{d}y = \frac{1}{\pi} \iint_D x\, \mathrm{d}x\mathrm{d}y = 0$ par imparité.
  De même $\mathbb{E}(Y) = 0$.
  Ainsi $\mathbb{E}(V) = \big(\mathbb{E}(X), \mathbb{E}(Y)\big) = (0,0)$ d'après la définition 52



- **Densités marginales** :
  pour $-1 \leqslant x \leqslant 1$:
  $$\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \mathrm{d}y \text{ donc } f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}$$
  De même, pour $-1 \leqslant y \leqslant 1$: $f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$.
  Enfin $f_X(x) = 0$ et $f_Y(y) = 0$ en dehors de $[-1,1]$.
  D'où $f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2} & \text{si } x \in [-1,1] \\ 0 & \text{sinon} \end{cases}$ et $f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2} & \text{si } y \in [-1,1] \\ 0 & \text{sinon} \end{cases}$

Remarque : X et Y **ne** sont **pas** des VARC à densités constantes, bien que $V = (X,Y)$ soit un CARC à densité constante.

17.3.4 Densités conditionnelles

Définition 134

Soit (X, Y) un CARC admettant une densité de probabilité f.
La **densité conditionnelle** de X sachant que $Y = y_0$ est la fonction définie sur $\mathbb{R}$ par :

$$f_{X\,/\,Y=y_0}\big(x, y_0\big) = \frac{f(x, y_0)}{f_Y(y_0)} \quad \text{dans les cas où } f_Y(y_0) \neq 0.$$

C'est une fonction de x, qui dépend, a priori, de y_0, même s'il est considéré comme fixé.
De même, la **densité conditionnelle** de Y sachant que $X = x_0$ est définie par : $f_{Y\,/\,X=x_0}\big(x_0, y\big) = \dfrac{f(x_0, y)}{f_X(x_0)}$ dans les cas où $f_X(x_0) \neq 0$.

Cas de l'exemple 1

$f_{X\,/\,Y=y_0}\big(x, y_0\big) = \dfrac{f(x, y_0)}{f_Y(y_0)} = \dfrac{1}{2\sqrt{1-y_0^2}}$ pour $y_0 \in]-1, 1[$, fonction constante pour $x \in \left[-\sqrt{1-y_0^2}, \sqrt{1-y_0^2}\right]$, et nulle sinon.

D'où, pour $y_0 \in]-1, 1[$, $\quad f_{X\,/\,Y=y_0}\big(x, y_0\big) = \begin{cases} \dfrac{1}{2\sqrt{1-y_0^2}} & \text{si } x \in \left[-\sqrt{1-y_0^2}, \sqrt{1-y_0^2}\right] \\ 0 & \text{sinon} \end{cases}$ (avec les notations précédentes.)

17.3.5 Indépendance

Définition

Définition 135

$V=(X,Y)$ est un **couple** $\mathbb{P}-$**indépendant**, ou les deux variables X et Y sont dites $\mathbb{P}-$**indépendantes**, si et seulement si pour tous les couples (A,B) de boréliens de $\mathbb{R}$, les évènements $\big\{X(\omega)\in A\big\}$ et $\big\{Y(\omega)\in B\big\}$ sont indépendants dans Ω :

$$\mathbb{P}\Big(\big\{X(\omega)\in A\big\} \text{ et } \big\{Y(\omega)\in B\big\}\Big)=\mathbb{P}\Big(\big\{X(\omega)\in A\big\}\Big)\times\mathbb{P}\Big(\big\{Y(\omega)\in B\big\}\Big)$$

A l'aide des fonctions de répartition et des fonctions densité on a le résultat suivant :

Théorème 136

1. (X,Y) un couple de VAR est $\mathbb{P}-$indépendant si et seulement si

$$\forall (x,y)\in\mathbb{R}^2,\quad F_{(X,Y)}(x,y)=F_X(x)\times F_Y(y)$$

2. (X,Y) un CARC de densité conjointe f est $\mathbb{P}-$indépendant si et seulement si

$$\forall (x,y)\in\mathbb{R}^2,\quad f(x,y)=f_X(x)\times f_Y(y)$$

Remarque : $\forall (x,y)\in\mathbb{R}^2,\quad f(x,y)=\dfrac{\partial^2 F_{(X,Y)}}{\partial x\,\partial y}\big(x,y\big)$.

Cas de l'exemple 1

Aux points (x,y) qui sont dans le carré, sans être dans le disque : $(x,y)\in\,]-1,1[^2\setminus D$, on a $f(x,y)=0\neq f_X(x)\times f_Y(y)$, car celles-ci sont non nulles.
Le couple $V=(X,Y)$ est donc **dépendant**.

17.3.6 Somme des deux VARC d'un couple à densité

Théorème 137

Si (X,Y) est un CARC de densité f, alors la variable aléatoire $Z = X+Y$ a une **fonction de répartition** donnée par :

$$\forall z \in \mathbb{R}, \quad F_Z(z) = \mathbb{P}\big(X+Y \leqslant z\big) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-y} f(x,y)\,\mathrm{d}x\right] \mathrm{d}y = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-x} f(x,y)\,\mathrm{d}y\right] \mathrm{d}x$$

.

Par dérivation de l'intégrale à paramètre par rapport à z, on obtient le théorème suivant.

Théorème 138

Si (X,Y) est un CARC de densité f, alors la variable aléatoire $Z = X+Y$ a une **densité de probabilité** donnée par :

$$\forall z \in \mathbb{R}, \quad f_Z(z) = {F'}_Z(z) = \int_{-\infty}^{+\infty} f(z-y,y)\,\mathrm{d}y = \int_{-\infty}^{+\infty} f(z-x,y)\,\mathrm{d}x$$

.

17.3.7 Covariance d'un couple (X,Y)

Définition 139

Sous réserve d'existence, on appelle **covariance** du couple (X,Y) le réel : $\mathbb{C}\mathrm{ov}\big(X,Y\big) = \mathbb{E}\Big[\big(X-\mathbb{E}(X)\big)\big(Y-\mathbb{E}(Y)\big)\Big]$.

Théorème 140

Sous réserve d'existence,

- pour tout couple (X,Y) : $\mathbb{C}\text{ov}\big(X,Y\big) = \mathbb{E}\big(X\,Y\big) - \mathbb{E}(X)\,\mathbb{E}(Y)$.
- Pour un **couple à densité** : $\mathbb{E}\big(X\,Y\big) = \displaystyle\iint_{\mathbb{R}^2} xy\, f(x,y)\,\mathrm{d}x\mathrm{d}y$
- Si (X,Y) est un **couple indépendant**, alors $\mathbb{C}\text{ov}\big(X,Y\big) = 0$ (mais la réciproque est fausse)

Remarque :
Les existences des moments d'ordre 1 assurent l'existence de $\mathbb{E}(X)$ et de $\mathbb{E}(Y)$, celles d'ordre 2 assurent l'existence de $\mathbb{E}(X\,Y)$, car $|X\,Y| \leqslant \dfrac{1}{2}\big(X^2+Y^2\big)$ chez les réels.

17.3.8 Coefficient de corrélation d'un couple

Définition 141

Sous réserve d'existence, on appelle **coefficient de corrélation** (linéaire) du couple (X,Y), le réel : $\rho\big(X,Y\big) = \dfrac{\mathbb{C}\text{ov}\big(X,Y\big)}{\sigma(X)\times\sigma(Y)}$

Par **l'inégalité de Cauchy-Schwarz**, quand il existe : $-1 \leqslant \rho(X,Y) \leqslant 1$

Si $\rho\big(X,Y\big) = 0$, les VAR X et Y sont dites **décorrélées**.

On retrouve des résultats similaires à ceux du chapitre 5.

17.3.9 Somme de deux VARC à densité indépendantes

Théorème 142

Si (X, Y) est un CARC **indépendant**, alors la variable aléatoire $Z = X + Y$ a une **fonction de répartition** donnée par :

$$\forall z \in \mathbb{R}, \quad F_Z(z) = \int_{-\infty}^{+\infty} F_X(z-y)\, f_Y(y)\, \mathrm{d}y.$$

F_Z est le produit de convolution de deux fonctions (en analyse) : $F_X * f_Y$.

Démonstration.

$$\begin{aligned} \forall z \in \mathbb{R}, \quad F_Z(z) &= \mathbb{P}\big(X + Y \leqslant z\big) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-y} f(x,y)\, \mathrm{d}x\right] \mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-y} f_X(x)\ f_Y(y)\, \mathrm{d}x\right] \mathrm{d}y = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{z-y} f_X(x)\, \mathrm{d}x\right] f_Y(y)\, \mathrm{d}y \end{aligned}$$

□

Théorème 143

Sous les mêmes hypothèses, la variable aléatoire $Z = X + Y$ admet une **densité de probabilité** donnée par :

$$\forall z \in \mathbb{R}, \quad f_Z(z) = \int_{-\infty}^{+\infty} f_X(z-y)\, f_Y(y)\, \mathrm{d}y.$$

f_Z est le produit de convolution de deux fonctions (en analyse) : $f_X * f_Y$.

17.4 Exemples d'application

17.4.1 Exemple 2

Considérons (X,Y) un CARC de densité f : $\begin{cases} f(x,y) = \mathrm{e}^{-(x+y)} & \text{si } x>0 \text{ et } y>0 \\ f(x,y) = 0 & \text{sinon.} \end{cases}$

- **Fonction de répartition** de X :

$$\begin{aligned}
\forall x > 0, \quad F_X(x) &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{x} f(u,v)\,\mathrm{d}u \right] \mathrm{d}v \\
&= \int_{0}^{+\infty} \left[\int_{0}^{x} \mathrm{e}^{-(u+v)}\,\mathrm{d}u \right] \mathrm{d}v \\
&= \int_{0}^{+\infty} \mathrm{e}^{-v} \left[\int_{0}^{x} \mathrm{e}^{-u}\,\mathrm{d}u \right] \mathrm{d}v \\
&= \int_{0}^{+\infty} \mathrm{e}^{-v} \left[1-\mathrm{e}^{-x} \right] \mathrm{d}v \\
&= \left(1-\mathrm{e}^{-x}\right) \int_{0}^{+\infty} \mathrm{e}^{-v}\mathrm{d}v \\
&= 1-\mathrm{e}^{-x}
\end{aligned}$$

$$\forall x \leqslant 0, \quad F_X(x) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{x} f(u,v)\,\mathrm{d}u \right] \mathrm{d}v = 0$$

Ainsi, $F_X(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1-\mathrm{e}^{-x} & \text{si } x > 0 \end{cases}$

- **Densité de probabilité** :

On a, $f_X(x) = {F'}_X(x) = \begin{cases} \mathrm{e}^{-x} & \text{si } x > 0 \\ 0 & \text{sinon} \end{cases}$ Ainsi $X \sim \mathcal{E}(1)$.

De la même manière, on obtient pour Y :

$$f_Y(y) = F'_Y(y) = \begin{cases} \mathrm{e}^{-y} & \text{si } y > 0 \\ 0 & \text{sinon} \end{cases}$$

Ainsi $\forall (x,y) \in \mathbb{R}^2, \quad f(x,y) = f_X(x) \times f_Y(y)$

Le couple (X,Y) est donc **indépendant**.

- **Densité conditionnelle** de X sachant $Y = y_0$.

Pour $y_0 > 0$, on a $f_Y(y_0) > 0$ et l'on a :

$$f_{X\,/\,Y=y_0}\big(x, y_0\big) = \frac{f(x,y_0)}{f_Y(y_0)} = \frac{\mathrm{e}^{-(x+y_0)}}{\mathrm{e}^{-y_0}} = \mathrm{e}^{-x} \text{ pour } x > 0 \text{ et } y_0 > 0$$

Elle ne dépend pas de y_0, ce qui est cohérent avec l'indépendance de X et Y.

17.4.2 Exemple 3

Quand on connaît la loi conjointe d'un couple (X,Y), on peut en déduire les lois marginales de X et Y.

Par contre, si on connaît les lois de X et de Y , on ne peut reconstituer la loi du couple (X,Y) que si les variables X et Y sont indépendantes.

Comparons les deux couples aléatoires suivants :

- $V = (X,Y)$ résultat d'un jet aléatoire d'une fléchette dans le carré $[0,1] \times [0,1]$, avec équiprobabilité.

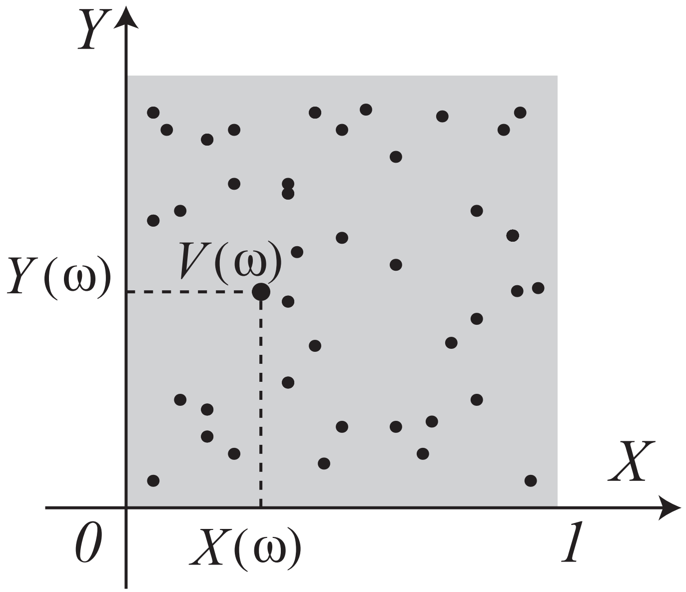
X admet une fonction de densité constante par morceaux f : $\begin{cases} f(x,y) = 1 & \text{si } (x,y) \in [0,1]^2 \\ f(x,y) = 0 & \text{sinon.} \end{cases}$

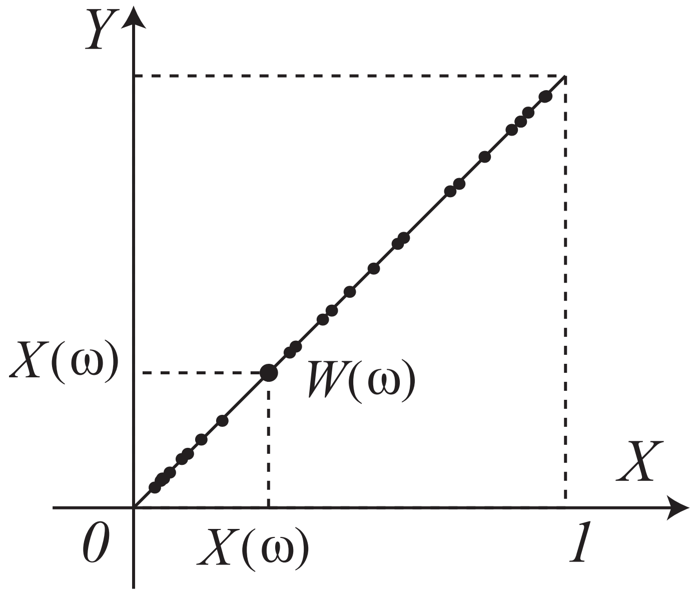
Alors $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)\,\mathrm{d}y = \int_0^1 \mathrm{d}y = 1$, si $x \in [0,1]$, et $f_X(x) = 0$ sinon.
La loi de X est donc la loi uniforme sur $[0,1]$, et de même celle de Y.

- $W = (X, X)$ avec X qui suit la loi uniforme sur $[0,1]$.

Le vecteur W a par construction, les **mêmes lois marginales** que le vecteur V précédent.

Mais ces fonctions vectorielles n'ont guère de rapport. Ainsi $V\langle\Omega\rangle = [0,1]^2$ est le carré, alors que $W\langle\Omega\rangle$ est la diagonale de ce carré. W n'est pas un couple à densité.





17.5 Loi normale en dimension 2

Définition 144

Un couple (X, Y) de VARC suit une **loi normale** à 2 dimensions, lorsqu'il admet une fonction densité définie par :
$\forall (x, y) \in \mathbb{R}^2, \quad f(x, y) = \dfrac{\sqrt{ac - b^2}}{2\pi} \, \mathrm{e}^{-\frac{1}{2}\left(a(x-m)^2 + 2b(x-m)(y-n) + c(y-n)^2\right)}$. Avec $a > 0, \ c > 0$ et $ac - b^2 > 0$.

Par calcul, on montre qu'alors, X et Y suivent des lois normales.

Théorème 145

Si (X, Y) suit une loi normale à 2 dimensions, alors X et Y suivent les lois normales :

$$X \sim \mathcal{N}\left(m, \frac{c^2}{ac - b^2}\right) \qquad \text{et} \qquad Y \sim \mathcal{N}\left(n, \frac{a^2}{ac - b^2}\right).$$

Proposition 146

Si (X, Y) est un couple de VARC **normales, centrées, réduites et indépendantes** alors :

$$\forall (x, y) \in \mathbb{R}^2, \quad f(x, y) = \frac{1}{2\pi} \, \mathrm{e}^{-\frac{1}{2}\left(x^2 + y^2\right)}$$

On a alors $a = 1, \ b = 0, \ c = 1, \ m = n = 0$,
et $f(x, y) = \dfrac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{x^2}{2}} \times \dfrac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-\frac{y^2}{2}}$.

La surface d'équation $z = f(x, y)$, au voisinage de $O = (0, 0)$: