

Probabilités et statistique

Introduction

L'objet de la théorie des probabilités est de fournir des modèles mathématiques permettant l'étude d'expériences dont le résultat n'est pas connu ou ne peut pas être prévu avec une totale certitude. Voici quelques exemples de la vie courante :

Expérience	Résultat observable
Lancer d'un dé	Un entier $k \in \{1, \dots, 6\}$
Lancer d'une pièce jusqu'à l'obtention d'un pile	Un entier $k \in \mathbb{N}$: le temps d'attente du premier succès
Mise en service d'une ampoule	Durée de vie $T \in \mathbb{R}^+$
Lancer d'une flèche sur une cible	Point d'impact $M \in \mathbb{R}^2$
Évolution temporelle de la température d'une pièce pendant une journée	Une courbe réelle continue

Le résultat précis de ces expériences n'est en général pas prévisible. Toutefois, l'observation et/ou l'intuition amènent souvent à prévoir certains comportements. Par exemple, si on jette 6000 fois un dé à 6 faces, on s'attend à ce que le nombre total de 4 soit voisin de 1000. La théorie des probabilités permet de donner un sens mathématique rigoureux à ces constatations empiriques.

En aval des probabilités se trouve la statistique qui permet de confronter les modèles probabilistes à la réalité observée.

1 Espaces de probabilités

1.1 Expérience aléatoire

Définition 1 On appelle expérience aléatoire toute expérience \mathcal{E} conduisant, selon le hasard, à plusieurs résultats possibles. On appelle univers associé à \mathcal{E} , l'ensemble Ω de tous les résultats possibles de \mathcal{E} .

Exemple 1 L'expérience \mathcal{E} consiste à lancer un dé.
 $\Omega_1 = \{1, \dots, 6\}$ est fini.

Exemple 2 L'expérience \mathcal{E} consiste à jouer à pile ou face jusqu'à l'obtention d'un pile.
 $\Omega_2 = \{P, FP, FFP, FFFP, \dots\}$ est infini dénombrable.

Exemple 3 L'expérience \mathcal{E} consiste à évaluer la durée de vie d'une étoile dans notre galaxie.
 $\Omega_3 =]0, +\infty[$ est infini non dénombrable.

1.2 Événements aléatoires, tribus

1.2.1 Événements aléatoires

Définition 2 On appelle épreuve ou événement élémentaire, tout élément $\omega \in \Omega$. On appelle événement aléatoire, un ensemble A constitué d'épreuves pour lesquelles A est réalisé.

Exemple 4

$$A = \text{« On obtient un chiffre pair »} = \{2, 4, 6\} \subset \Omega_1$$

Exemple 5

$$B = \text{« On obtient un pile en au plus 4 lancers »} = \{P, FP, FFP, FFFP\} \subset \Omega_2$$

Exemple 6

$$C = \text{« La durée de vie de l'étoile est supérieure à 10 millions d'années »} =]10^7, +\infty[\subset \Omega_3$$

Définition 3 L'univers Ω s'appelle l'événement certain car il est toujours réalisé. \emptyset s'appelle l'événement impossible car il n'est jamais réalisé.

Un événement aléatoire est donc juste une sous-partie de Ω . Le but est ensuite de définir une fonction, que l'on appellera probabilité, qui notera chaque événement aléatoire entre 0 et 1 : plus la note est proche de 1, plus l'événement a de chance de se réaliser.

Néanmoins, pour des raisons mathématiques, on ne peut pas toujours considérer toute sous-partie de Ω comme un événement aléatoire. Il faudra parfois se restreindre à un sous ensemble de l'ensemble des parties de Ω .

1.2.2 Algèbres et tribus

On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Un sous-ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ est un ensemble de parties de Ω .

Définition 4 Soit \mathcal{A} un sous ensemble de $\mathcal{P}(\Omega)$. On dit que \mathcal{A} est une σ -algèbre ou tribu si

1. $\Omega \in \mathcal{A}$,
2. \mathcal{A} est stable par passage au complémentaire : si $A \in \mathcal{A}$, alors ${}^c A = \bar{A} \in \mathcal{A}$,
3. \mathcal{A} est stable par réunion dénombrable : si $(A_n)_{n \in \mathbb{N}}$ est une famille dénombrable d'ensembles de \mathcal{A} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Définition 5 Soient Ω un univers et \mathcal{A} une tribu associée. Le couple (Ω, \mathcal{A}) est appelé un espace mesurable ou espace probablisable.

Remarque 1

- Un espace probablisable est un espace sur lequel on va pouvoir considérer une mesure de probabilité.
- Dans la suite, l'ensemble des événements aléatoires associés à Ω est donné par une tribu \mathcal{A} . En particulier, si $\mathcal{A} \neq \mathcal{P}(\Omega)$, alors certaines parties de Ω ne sont pas des événements aléatoires et donc on ne pourra pas calculer leur probabilité.

Remarque 2

- $\{\emptyset, \Omega\}$ est une tribu appelée tribu triviale.
- $\mathcal{P}(\Omega)$ est toujours une tribu. C'est la plus grosse, elle contient toutes les autres.
- L'intersection d'une famille quelconque de tribus est une tribu.

Définition 6 Soit \mathcal{A} un sous-ensemble de $\mathcal{P}(\Omega)$. On appelle tribu engendrée par \mathcal{A} , notée $\sigma(\mathcal{A})$, l'intersection de toutes les tribus contenant \mathcal{A} . C'est également la plus petite tribu contenant \mathcal{A} .

Remarque 3 La famille des tribus contenant \mathcal{A} n'est pas vide car $\mathcal{P}(\Omega)$ en fait partie. Enfin, si \mathcal{A} est une tribu, alors $\sigma(\mathcal{A}) = \mathcal{A}$.

Exemple 7 Si A est une partie de Ω , alors

$$\sigma(\{A\}) = \{\emptyset, A, \bar{A}, \Omega\}, \quad \sigma(\{A, \bar{A}\}) = \{\emptyset, A, \bar{A}, \Omega\}.$$

Exemple 8 Soit (A_1, \dots, A_n) une partition de Ω , i.e. $\Omega = \bigcup_{i=1}^n A_i$ et $A_i \cap A_j = \emptyset$, pour tous les $i \neq j$. Alors

$$\sigma(\{A_1, \dots, A_n\}) = \{\emptyset, A_{i_1} \cup \dots \cup A_{i_k}, 1 \leq k \leq n, 1 \leq i_1 < \dots < i_k \leq n\}.$$

Cet ensemble est constitué de 2^n éléments (si les A_i sont non vides).

Définition 7 La tribu borélienne de \mathbb{R}^d , notée $\mathcal{B}(\mathbb{R}^d)$ est la tribu engendrée par les ensembles ouverts de \mathbb{R}^d .

Proposition 1 $\mathcal{B}(\mathbb{R}^d)$ coïncide avec

- la tribu engendrée par les ensembles fermés de \mathbb{R}^d ,
- la tribu engendrée par les pavés $[a_1, b_1] \times \dots \times [a_d, b_d]$,
- la tribu engendrée par les pavés $]a_1, b_1[\times \dots \times]a_d, b_d[$,
- la tribu engendrée par les pavés $]a_1, b_1] \times \dots \times]a_d, b_d]$,
- la tribu engendrée par les pavés $]a_1, b_1[\times \dots \times]a_d, b_d[$,
- la tribu engendrée par les ensembles $[a_1, +\infty[\times \dots \times [a_d, +\infty[$.

Remarque 4 $\mathcal{B}(\mathbb{R}^d) \neq \mathcal{P}(\mathbb{R}^d)$

1.3 Probabilités

1.3.1 Définition

Définition 8 Soit (Ω, \mathcal{A}) un espace mesurable ou probabiliste. On appelle probabilité sur (Ω, \mathcal{A}) toute application $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ telle que

1. $\mathbb{P}(\Omega) = 1$,
2. Pour toute famille $(A_n)_{n \in \mathbb{N}}$ d'ensembles de \mathcal{A} deux à deux disjoints, on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

Définition 9 Un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace probabilisé ou espace de probabilité.

Remarque 5 Une probabilité est une mesure positive bornée dont la valeur en Ω vaut 1.

Dans toute la suite on se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ donné.

1.3.2 Propriétés

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

Preuve. $\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset)$ car l'union est disjointe. \square

Proposition 3

- Soit $A \in \mathcal{A}$. Alors $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$.
- Plus généralement, si (A_1, \dots, A_n) est une partition de Ω , alors on a $\sum_{i=1}^n \mathbb{P}(A_i) = 1$.

Preuve. $1 = \mathbb{P}(\Omega) = \mathbb{P}(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i)$ car la réunion est disjointe. \square

Proposition 4 Soient $A, B \in \mathcal{A}$.

- Si $A \subset B$, alors $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$

Preuve. $B = (B \cap A) \cup (B \cap \bar{A})$ et l'union est disjointe. Donc

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap \bar{A}) = \mathbb{P}(A) + \mathbb{P}(B \cap \bar{A}) \geq \mathbb{P}(A).$$

De plus,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A \cup (B \cap \bar{A})) = \mathbb{P}(A) + \mathbb{P}(B \cap \bar{A}) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

car l'union est disjointe et $(B \cap \bar{A}) \subset B$. \square

1.3.3 Inégalités de Boole

Proposition 5 Soit $(A_n)_{n \in \mathbb{N}}$ une famille dénombrable d'ensembles de \mathcal{A} .

1. On a l'inégalité de Boole

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

2. De plus, si $(A_n)_{n \in \mathbb{N}}$ est croissante, i.e. $A_n \subset A_{n+1}$, alors on a

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

3. Enfin, si $(A_n)_{n \in \mathbb{N}}$ est décroissante, alors on a

$$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right) = \lim_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

Preuve.

- Commençons par le point 2. On définit une suite d'ensemble $(C_n)_{n \in \mathbb{N}}$ par

$$C_0 = A_0, \quad C_n = A_n \setminus A_{n-1} := A_n \cap \overline{A_{n-1}}, \quad \forall n \geq 1.$$

$(C_n)_{n \in \mathbb{N}}$ est ainsi une suite d'événements, deux à deux disjoints, vérifiant

$$A_n = \bigcup_{i=0}^n C_i \quad \text{et} \quad \mathbb{P}(A_n) = \sum_{i=0}^n \mathbb{P}(C_i).$$

On a alors

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right) &= \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} C_n\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(C_n) = \lim_{n \rightarrow +\infty} \sum_{i=0}^n \mathbb{P}(C_i) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}(A_n). \end{aligned}$$

- Concernant le point 3, il suffit de passer au complémentaire pour utiliser le point 2 : On pose $B_n = \overline{A_n}$ pour tout $n \in \mathbb{N}$. $(B_n)_{n \in \mathbb{N}}$ est une suite croissante donc on a $\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \lim_{n \in \mathbb{N}} \mathbb{P}(B_n)$. Or

$$\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} \overline{A_n}\right) = \mathbb{P}\left(\overline{\bigcap_{n \in \mathbb{N}} A_n}\right) = 1 - \mathbb{P}\left(\bigcap_{n \in \mathbb{N}} A_n\right)$$

et

$$\lim_{n \in \mathbb{N}} \mathbb{P}(B_n) = \lim_{n \in \mathbb{N}} \mathbb{P}(\overline{A_n}) = 1 - \lim_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

- Reste le point 1. On pose $B_n = \bigcup_{i=0}^n A_i$. Comme $B_n = B_{n-1} \cup A_n$ on a $\mathbb{P}(B_n) \leq \mathbb{P}(B_{n-1}) + \mathbb{P}(A_n)$ et par récurrence immédiate,

$$\mathbb{P}(B_n) \leq \sum_{i=0}^n \mathbb{P}(A_i) \leq \sum_{i=0}^{+\infty} \mathbb{P}(A_i).$$

Comme $(B_n)_{n \in \mathbb{N}}$ est croissante, le point 2 nous donne également

$$\lim_{n \in \mathbb{N}} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} B_n\right) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right).$$

□

1.3.4 Formule de Poincaré

Proposition 6 Soient $A, B \in \mathcal{A}$. On a $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Proposition 7 Pour $n \geq 2$, soit (A_1, \dots, A_n) une famille finie d'événements aléatoires.

1. On a la formule de Poincaré

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}).$$

2. En particulier, si $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$ ne dépend que de l'entier k , on a la formule de Poincaré simplifiée

$$\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^n (-1)^{k-1} C_n^k \mathbb{P}(A_1 \cap \dots \cap A_k).$$

Pour $n = 3$ on a par exemple

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

Preuve. On peut montrer la formule de Poincaré par récurrence sur n par exemple. \square

1.3.5 Probabilité uniforme

On suppose dans cette partie que Ω est fini ou dénombrable. On choisit alors pour tribu des événements \mathcal{A} l'ensemble $\mathcal{P}(\Omega)$.

Proposition 8

• La formule

$$\mathbb{P}(A) := \sum_{\omega \in A} p_\omega, \quad \forall A \in \mathcal{P}(\Omega),$$

définit une probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ dès que la famille de nombres réels $(p_\omega)_{\omega \in \Omega}$ vérifie

1. $p_\omega \in [0, 1]$ pour tous les $\omega \in \Omega$,
2. $\sum_{\omega \in \Omega} p_\omega = 1$.

- Inversement, toute probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ est de cette forme : il suffit de poser

$$p_\omega := \mathbb{P}(\{\omega\}), \quad \forall \omega \in \Omega.$$

Définition 10 Soit \mathbb{P} une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ avec Ω fini. On dit que \mathbb{P} est la loi uniforme sur Ω si toutes les épreuves $\omega \in \Omega$ sont équiprobables : i.e.

$$p_\omega = \frac{1}{\text{Card}(\Omega)}.$$

En particulier on a

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

Remarque 6 Le calcul des probabilités se ramène ici à un simple calcul de dénombrement. On a

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

2 Probabilités conditionnelles, indépendance

Dans toute la suite on se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

2.1 Probabilités conditionnelles

2.1.1 Définition

Définition 11 Soient $A, B \in \mathcal{A}$ tels que $\mathbb{P}(B) > 0$. On appelle probabilité conditionnelle de A sachant B , le nombre réel

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

Proposition 9 Soit $B \in \mathcal{A}$ tel que $\mathbb{P}(B) > 0$. L'application

$$\mathbb{P}(\cdot|B) : \begin{cases} \mathcal{A} & \rightarrow [0, 1] \\ A & \mapsto \mathbb{P}(A|B). \end{cases}$$

est une mesure de probabilité sur (Ω, \mathcal{A}) .

Preuve.

- Pour tout $A \in \mathcal{A}$, $A \cap B \subset B$ donc $\mathbb{P}(A|B) \in [0, 1]$.
- $\mathbb{P}(\Omega|B) = 1$.
- Si $(A_n)_{n \in \mathbb{N}}$ est une famille d'événements aléatoires deux à deux disjoints, on a

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n | B\right) &= \frac{\mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \frac{\mathbb{P}\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right)}{\mathbb{P}(B)} \\ &= \sum_{n \in \mathbb{N}} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n | B). \end{aligned}$$

□

Remarque 7 On a $\mathbb{P}(B|B) = 1$.

2.1.2 Formule des probabilités composées

Théorème 1 1. Soient $A, B \in \mathcal{A}$ avec $\mathbb{P}(A) > 0$. On a la formule des probabilités composées suivante :

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

2. Plus généralement, soient $n \geq 2$ et $A_1, \dots, A_n \in \mathcal{A}$ tels que

$$\mathbb{P}(A_1 \cap \dots \cap A_{n-1}) > 0.$$

Alors on a

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

Remarque 8 Comme pour tout $1 \leq k \leq n-1$, $(A_1 \cap \dots \cap A_{n-1}) \subset (A_1 \cap \dots \cap A_k)$, alors

$$\mathbb{P}(A_1 \cap \dots \cap A_k) > 0$$

et donc les probabilités conditionnelles précédentes sont bien définies.

Preuve. Montrons le second point par récurrence sur n , le premier point correspondant à $n = 2$.

Pour $n = 2$ le résultat découle directement de la formule des probabilités conditionnelles.

Supposons vrai le résultat pour $n \geq 2$. Alors on a

$$\mathbb{P}(A_1 \cap \dots \cap A_{n+1}) = \mathbb{P}(A_1 \cap \dots \cap A_n)\mathbb{P}(A_{n+1}|A_1 \cap \dots \cap A_n)$$

puis on conclut grâce à l'hypothèse de récurrence. \square

Le premier point du théorème peut sembler à première vue inutile car il s'agit d'une simple réécriture de la définition d'une probabilité conditionnelle. En fait il n'en est rien, les deux formules ont leur intérêt en pratique. Dans certaines situations, on connaît la probabilité de l'intersection de deux événements et on calcule la probabilité conditionnelle, tandis que dans d'autres situations on connaît la probabilité conditionnelle et on en déduit la probabilité de l'intersection de deux événements.

2.1.3 Formule des probabilités totales

Théorème 2 1. Soient $A, B \in \mathcal{A}$ tels que $0 < \mathbb{P}(A) < 1$, on a

$$\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(\bar{A})\mathbb{P}(B|\bar{A}).$$

2. Plus généralement, si $(A_n)_{n \in \mathbb{N}}$ est une partition de Ω avec $\mathbb{P}(A_n) > 0$ pour tout $n \in \mathbb{N}$, on a

$$\mathbb{P}(B) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B|A_n).$$

Preuve. On a $\bigcup_{n \in \mathbb{N}} A_n = \Omega$ et la réunion est disjointe, donc

$$\mathbb{P}(B) = \mathbb{P}(B \cap (\bigcup_{n \in \mathbb{N}} A_n)) = \mathbb{P}(\bigcup_{n \in \mathbb{N}} (B \cap A_n)) = \sum_{n \in \mathbb{N}} \mathbb{P}(B \cap A_n).$$

Il suffit alors d'utiliser la formule des probabilités composées pour conclure.

Le premier point du théorème se traite de la même façon car $\{A, \bar{A}\}$ est une partition de Ω . \square

2.1.4 Formule de Bayes

Théorème 3 1. Soient $A, B \in \mathcal{A}$ tels que $0 < \mathbb{P}(A) < 1$ et $\mathbb{P}(B) > 0$, on a

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

2. Plus généralement, si $(A_n)_{n \in \mathbb{N}}$ est une partition de Ω avec $\mathbb{P}(A_n) > 0$ pour tout $n \in \mathbb{N}$, et $\mathbb{P}(B) > 0$, on a

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k)\mathbb{P}(B|A_k)}{\sum_{n \in \mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B|A_n)} \quad \forall k > 0.$$

Preuve. Pour tout $k > 0$ on a

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

et on applique la formule des probabilités totales pour conclure.

Encore une fois, le premier point du théorème se traite de la même façon. \square

Exemple 9 On dispose de 4 jetons dont 2 possèdent 2 faces blanches, un a une face blanche et une face rouge, et le dernier a deux faces rouges. On tire un jeton au hasard et on ne voit que l'une des faces : elle est blanche. Quelle est la probabilité que l'autre face soit blanche ?

2.2 Indépendance

Définition 12 Soient $A, B \in \mathcal{A}$. On dit que A et B sont indépendants si

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

Remarque 9 La notion d'indépendance est différente de la notion d'incompatibilité. En particulier, si A et B sont incompatibles (i.e. disjoints), alors $\mathbb{P}(A \cap B) = 0$.

Remarque 10 Si $A, B \in \mathcal{A}$ sont indépendants avec $\mathbb{P}(B) > 0$, alors on a

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

Proposition 10 Soient $A, B \in \mathcal{A}$. Si A et B sont indépendants, alors A et \bar{B} , \bar{A} et B , \bar{A} et \bar{B} sont également indépendants.

Preuve. Par symétrie, il suffit de montrer que A et \bar{B} sont indépendants. On a $\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B})$. Donc

$$\mathbb{P}(A \cap \bar{B}) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) = \mathbb{P}(A)\mathbb{P}(\bar{B}).$$

□

Proposition 11 Soient $(A_n)_{n \in \mathbb{N}}$ une famille dénombrable d'événements de \mathcal{A} deux à deux disjoints et soit $B \in \mathcal{A}$. Si, pour tout $n \in \mathbb{N}$, A_n et B sont indépendants, alors $\bigcup_{n \in \mathbb{N}} A_n$ et B sont également indépendants.

Preuve.

$$\begin{aligned} \mathbb{P}\left(\left(\bigcup_{n \in \mathbb{N}} A_n\right) \cap B\right) &= \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} (A_n \cap B)\right) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n \cap B) \\ &= \sum_{n \in \mathbb{N}} \mathbb{P}(A_n)\mathbb{P}(B) = \mathbb{P}\left(\bigcup_{n \in \mathbb{N}} A_n\right)\mathbb{P}(B). \end{aligned}$$

□

On va maintenant prolonger la notion d'indépendance de deux événements au cas des familles finies ou dénombrables d'événements.

Définition 13 Soient $(A_n)_{n \in \mathbb{N}}$ des événements aléatoires (donc des éléments de \mathcal{A}).

1. $(A_n)_{n \in \mathbb{N}}$ est une famille d'événements 2 à 2 indépendants si pour tout $n, m \in \mathbb{N}$ avec $n \neq m$, on a

$$\mathbb{P}(A_n \cap A_m) = \mathbb{P}(A_n)\mathbb{P}(A_m).$$

2. $(A_n)_{n \in \mathbb{N}}$ est une famille d'événements indépendants dans leur ensemble (ou mutuellement indépendants ou bien encore indépendants) si, pour tout $k \geq 2$ et pour toute sous-suite finie $(A_{i_1}, A_{i_2}, \dots, A_{i_k})$ d'événements distincts, on a

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k \mathbb{P}(A_{i_j}).$$

Remarque 11

- L'indépendance mutuelle implique l'indépendance 2 à 2. La réciproque est fausse, c.f. TD.
- Pour une famille finie d'événements aléatoires, l'indépendance 2 à 2 impose $C_n^2 = \frac{n(n-1)}{2}$ conditions tandis que l'indépendance mutuelle impose

$$C_n^2 + C_n^3 + \dots + C_n^n = 2^n - C_n^0 - C_n^1 = 2^n - n - 1$$

conditions.

3 Variables aléatoires, variables aléatoires discrètes

Dans tout ce chapitre, on considère un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

3.1 Variables aléatoires

Définition 14 Soit (F, \mathcal{B}) un espace mesurable. On appelle variable aléatoire à valeurs dans F , toute application mesurable $X : \Omega \rightarrow F$, c'est à dire satisfaisant pour tout $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{A}$ avec

$$X^{-1}(B) = \{\omega \in \Omega \text{ tels que } X(\omega) \in B\}.$$

On notera $\{X \in B\} := X^{-1}(B)$.

Définition 15

- On appelle variable aléatoire réelle une variable aléatoire à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- On appelle vecteur aléatoire réel une variable aléatoire à valeurs dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Proposition 12 • Soit X est un vecteur aléatoire réel et $h : \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ une fonction mesurable. Alors $h(X)$ est un vecteur aléatoire réel.

- $X = (X_1, \dots, X_n)$ est un vecteur aléatoire réel si et seulement si chacune de ses composantes X_i est une variable aléatoire réelle.

Remarque 12 Soit X une variable aléatoire réelle. On utilisera les notations suivantes :

- $\{X = a\} = X^{-1}(\{a\}) = \{\omega \in \Omega \text{ tels que } X(\omega) = a\},$
- $\{X \in [a, b]\} = \{a \leq X \leq b\} = X^{-1}([a, b]) = \{\omega \in \Omega \text{ tels que } X(\omega) \in [a, b]\},$
- $\{X \in]-\infty, b]\} = \{X \leq b\} = X^{-1}(]-\infty, b]) = \{\omega \in \Omega \text{ tels que } X(\omega) \in]-\infty, b]\},$
- ...

Remarque 13 On a vu que $\mathcal{B}(\mathbb{R})$ coïncide avec la tribu engendrée par les intervalles $[a, b]$ avec $a, b \in \mathbb{R}$ et $a < b$. Ainsi X est une variable aléatoire réelle si pour tout intervalle $[a, b]$, $\{X \in [a, b]\} \in \mathcal{A}$, i.e. $\{X \in [a, b]\}$ est un événement aléatoire.

3.1.1 Loi de probabilité

Définition 16 Soit X une variable aléatoire réelle. On appelle loi de probabilité de X la mesure de probabilité \mathbb{P}_X définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ par

$$\mathbb{P}_X(B) := \mathbb{P}(X \in B), \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Remarque 14 Puisque $\mathcal{B}(\mathbb{R})$ est engendré par les intervalles (ouverts, fermés,...) et les ensembles du type $]-\infty, t]$, la loi d'une variable aléatoire réelle est caractérisée par la donnée des valeurs

- $\mathbb{P}(X \in [a, b])$, pour tous $a < b$,
- ou $\mathbb{P}(X \in]a, b])$, pour tous $a < b$,
- ou $\mathbb{P}(X \leq t)$, pour tous $t \in \mathbb{R}$,
- ...

Lorsque l'on considère la loi d'une variable aléatoire on s'intéresse aux valeurs que prend la variable aléatoire et aux probabilités associées. Par contre on oublie l'univers Ω : Pour toute variable aléatoire réelle on peut calculer sa loi, par contre on ne peut pas reconstruire une variable aléatoire à partir de sa loi. En particulier, deux variables aléatoires réelles qui ne sont même pas nécessairement définies sur le même univers, peuvent avoir une même loi ! (c.f. exemples plus loin).

3.1.2 Fonction de répartition

Définition 17 Soit X une variable aléatoire réelle. On appelle fonction de répartition de X , l'application

$$F_X : \begin{cases} \mathbb{R} & \rightarrow [0, 1] \\ t & \mapsto \mathbb{P}(X \leq t) = \mathbb{P}_X(]-\infty, t]) \end{cases}$$

Proposition 13 La fonction de répartition d'une variable aléatoire réelle X caractérise sa loi.

Cette proposition découle directement de la remarque 14.

Proposition 14 Soit X une variable aléatoire réelle de fonction de répartition F_X . Alors on a

1. F_X est une fonction croissante. Pour tous $a, b \in \mathbb{R}$ avec $a < b$, on a

$$F_X(b) - F_X(a) = \mathbb{P}(a < X \leq b).$$

2. F_X est continue à droite et possède une limite à gauche en tout point.
3. $\lim_{t \rightarrow -\infty} F_X(t) = 0$ et $\lim_{t \rightarrow +\infty} F_X(t) = 1$.

3.1.3 Espérance, variance

Proposition et définition 1 1. Soit X une variable aléatoire réelle positive, alors l'espérance de X est définie par la quantité, éventuellement infinie,

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x).$$

2. Soit X une variable aléatoire réelle non nécessairement positive, alors l'espérance de X est définie uniquement si $\mathbb{E}[|X|] < +\infty$. Dans ce cas on dit qu'elle est intégrable et elle vaut

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x).$$

Remarque 15 L'espérance d'une variable aléatoire ne dépend que de sa loi.

Proposition 15 Soit X une variable aléatoire réelle intégrable.

- $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.
- Pour tout $p > 1$, $\mathbb{E}[|X|] \leq \mathbb{E}[|X|^p]^{1/p}$ (Jensen ou Hölder).
- Soient $a, b \in \mathbb{R}$, alors $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.
- Soit Y une autre variable aléatoire réelle intégrable, alors $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ et, si $X \leq Y$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Définition 18 Soit X une variable aléatoire de carré intégrable. On appelle variance de X , notée $\text{Var}(X)$, le nombre positif

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

On appelle écart-type de X le nombre positif

$$\sigma := \sqrt{\text{Var}(X)}.$$

Proposition 16 Soit X une variable aléatoire de carré intégrable. Alors on a

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

Proposition 17 Soit X une variable aléatoire de carré intégrable et $a, b \in \mathbb{R}$. On a

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Définition 19 Soit $r \in \mathbb{N}^*$. Si $\mathbb{E}[|X|^r] < +\infty$, on appelle moment d'ordre r la quantité $\mathbb{E}[X^r]$.

Définition 20 Soit X un vecteur aléatoire. On dit que X est intégrable si chacune de ses composantes est intégrable. L'espérance de X , notée $\mathbb{E}[X]$, est alors le vecteur de \mathbb{R}^n défini par

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}.$$

De plus si X est de carré intégrable, i.e. ses composantes sont de carré intégrables, sa matrice de covariance $(\Gamma_{i,j})_{1 \leq i,j \leq n}$ est définie par

$$\Gamma_{i,j} = \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j] := \text{Cov}(X_i, X_j),$$

ou, de manière équivalente,

$$\Gamma = \mathbb{E}[(X - \mathbb{E}[X])^\top (X - \mathbb{E}[X])] = \mathbb{E}[X^\top X] - \mathbb{E}[X]^\top \mathbb{E}[X].$$

Remarque 16 la matrice de covariance est une matrice carrée de taille $n \times n$, symétrique et semi-définie positive, i.e.

$$u^\top \Gamma u \geq 0, \quad \forall u \in \mathbb{R}^n.$$

3.1.4 Indépendance

Définition 21 Soit X une variable aléatoire réelle. On appelle tribu engendrée par X , notée $\sigma(X)$, la sous-tribu de \mathcal{A} définie par

$$\sigma(X) := \{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\}.$$

Remarque 17

- $\sigma(X)$ est la plus petite sous-tribu de \mathcal{A} pour laquelle X soit mesurable.
- $\sigma(X)$ est la tribu engendrée par les événements $\{X \in [a, b]\}$ pour tous les réels $a < b$.

Exemple 10 Si $X = c$, alors $\sigma(X) = \{\emptyset, \Omega\}$. Inversement, si $\sigma(X) = \{\emptyset, \Omega\}$ alors X est une variable aléatoire constante.

Définition 22 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles. On dit que $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes si la famille de sous-tribus $(\sigma(X_n))_{n \in \mathbb{N}}$ est indépendante dans son ensemble.

Proposition 18 $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes si et seulement si, pour tous $k \geq 2$, toute sous-suite finie $(X_{i_1}, \dots, X_{i_k})$ est constituée de variables aléatoires indépendantes.

Proposition 19 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires. Pour tout $n \geq 2$, X_1, \dots, X_n sont indépendantes si et seulement si pour toutes fonctions mesurables $h_1, \dots, h_n : \mathbb{R} \rightarrow \mathbb{R}$ positives ou bien bornées on a

$$\mathbb{E}\left[\prod_{i=1}^n h_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[h_i(X_i)].$$

Remarque 18 La proposition précédente implique que si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes et si $(h_n)_{n \in \mathbb{N}}$ est une suite de fonctions mesurables réelles, alors $(h_n(X_n))_{n \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes

Proposition 20 Si X et Y sont indépendantes et de carré intégrables, alors $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

3.2 Variables aléatoires discrètes

3.2.1 Définition, propriétés

Définition 23 On dit qu'une variable aléatoire X est discrète si l'ensemble des observables

$$X(\Omega) := \{X(\omega), \omega \in \Omega\}$$

est fini ou infini dénombrable.

Remarque 19 Si X est une variable aléatoire discrète, alors on note

$$X(\Omega) = \{x_1, x_2, \dots, x_n\}$$

si l'ensemble des observables est fini, et

$$X(\Omega) = \{x_0, x_1, \dots\}$$

si l'ensemble des observables est infini.

Proposition 21 La loi de probabilité d'une variable aléatoire discrète est caractérisée par la donnée des informations suivantes :

1. Les valeurs prises par la variable aléatoire, à savoir $X(\Omega)$.

2. Les probabilités associées, i.e. les valeurs $\mathbb{P}(X = x)$ pour tous les $x \in X(\Omega)$.

Preuve. Il suffit juste de remarquer que pour tous $a < b$ on a

$$\mathbb{P}(X \in [a, b]) = \sum_{x \in X(\Omega) \cap [a, b]} \mathbb{P}(X = x).$$

□

Proposition 22 La fonction de répartition d'une variable aléatoire discrète X est une fonction en escalier donnée par

$$F_X(t) = \sum_{x \in X(\Omega) \cap]-\infty, t]} \mathbb{P}(X = x).$$

Les abscisses des sauts sont données par $X(\Omega)$ et les hauteurs de saut par

$$F_X(x) - F_X(x^-) = \mathbb{P}(X = x), \quad \forall x \in X(\Omega).$$

Remarque 20 Si X est discrète, $\sigma(X)$ est la tribu engendrée par les événements $\{X = x\}$ pour tous les $x \in X(\Omega)$. De plus on a

$$\sigma(X) = \{\{X \in E\}, E \subset X(\Omega)\}.$$

Proposition 23 Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires discrètes. Pour tout $n \geq 2$, X_1, \dots, X_n sont indépendantes si et seulement si pour tous $x_1 \in X_1(\Omega), \dots, x_n \in X_n(\Omega)$ on a

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i).$$

3.2.2 Espérance et variance

Proposition 24 Soit X une variable aléatoire discrète. X est intégrable si et seulement si

$$\mathbb{E}[|X|] = \sum_{x \in X(\Omega)} |x| \mathbb{P}(X = x) < +\infty$$

et dans ce cas on a

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

Proposition 25 Soit X une variable aléatoire discrète et h une fonction mesurable réelle. On suppose que

$$\sum_{x \in X(\Omega)} |h(x)| \mathbb{P}(X = x) < +\infty.$$

Alors, si on pose $Y = h(X)$, Y est intégrable et

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

Remarque 21 Y est une variable discrète et on peut déterminer sa loi :

- $Y(\Omega) = h(X(\Omega))$,
- $\mathbb{P}(Y = y) = \sum_{x \in X(\Omega), y=h(x)} \mathbb{P}(X = x)$.

L'intérêt de la proposition précédente vient du fait que l'on peut calculer l'espérance de Y sans calculer sa loi.

Exemple 11

$$\mathbb{E}[X^2] = \sum_{x \in X(\Omega)} x^2 \mathbb{P}(X = x).$$

Proposition 26 Soit X une variable aléatoire discrète de carré intégrable. Alors on a

$$\text{Var}(X) = \sum_{x \in X(\Omega)} x^2 \mathbb{P}(X = x) - \left(\sum_{x \in X(\Omega)} x \mathbb{P}(X = x) \right)^2.$$

3.2.3 Lois discrètes classiques

Loi de Bernoulli

Définition 24 Soit $p \in [0, 1]$. X suit la loi de Bernoulli de paramètre p si

$$X(\Omega) = \{0, 1\}, \quad \mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

On note $X \sim \mathcal{B}(p)$.

Proposition 27 $\mathbb{E}[X] = p$ et $\text{Var}(X) = p(1 - p)$.

Loi binomiale

Définition 25 Soit $n \in \mathbb{N}$, $n \geq 1$ et $p \in [0, 1]$. On note X le nombre de succès lorsque l'on réalise n expériences indépendantes, chaque expérience ayant une même probabilité p de succès. On dit alors que X suit la loi binomiale de paramètres (n, p) et on note $X \sim \mathcal{B}(n, p)$.

Proposition 28 $X(\Omega) = \{0, \dots, n\}$ et

$$\mathbb{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \quad 0 \leq k \leq n.$$

De plus X a même loi que $\sum_{i=1}^n X_i$ avec X_1, \dots, X_n i.i.d. de loi $\mathcal{B}(p)$.

Proposition 29 $\mathbb{E}[X] = np$ et $\text{Var}(X) = np(1-p)$.

Loi géométrique

Définition 26 Soit $p \in]0, 1[$. On note X le rang du premier succès lorsque l'on réalise de façon indépendante une même expérience de probabilité p de succès. On dit que X suit une loi géométrique de paramètre p et on note $X \sim \mathcal{G}(p)$.

Proposition 30 $X(\Omega) = \mathbb{N}^*$ et

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

Proposition 31 $\mathbb{E}[X] = \frac{1}{p}$ et $\text{Var}(X) = \frac{1-p}{p^2}$.

Loi de Poisson

Définition 27 Soit $\lambda > 0$. X suit la loi de Poisson de paramètre λ si $X(\Omega) = \mathbb{N}$ et

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

Proposition 32 $\mathbb{E}[X] = \lambda$ et $\text{Var}(X) = \lambda$.

3.3 Vecteurs aléatoires discrets

Définition 28 On dit qu'un vecteur aléatoire X est discret si et seulement si chacune de ses composantes est une variable aléatoire discrète.

Proposition 33 La loi d'un vecteur aléatoire discret $X = (X_1, \dots, X_n)$ est caractérisée par les quantités

- $X_1(\Omega), \dots, X_n(\Omega)$,
- $\mathbb{P}(X = x) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n)$, pour tous $x = (x_1, \dots, x_n) \in X_1(\Omega) \times \dots \times X_n(\Omega)$.

Définition 29 La loi du vecteur X est appelée loi jointe tandis que les lois de chacune des composantes X_i sont appelées lois marginales.

Proposition 34 On a pour tout $1 \leq i \leq n$, et tout $x_i \in X_i(\Omega)$,

$$\mathbb{P}(X_i = x_i) = \sum_{x_j \in X_j(\Omega), j \neq i} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

Exemple 12

$X_1 \backslash X_2$	1	2	
1	1/3	1/6	1/2
2	1/2	0	1/2
	5/6	1/6	

3.4 Estimateur du maximum de vraisemblance pour des variables aléatoires discrètes

3.4.1 Introduction

Dans un tiroir se trouvent des pièces truquées et des pièces non truquées. On choisit une pièce sans savoir si elle est truquée et on la lance plusieurs fois en notant X_i le résultat du lancer i : $X_i = 1$ si elle fait pile et $X_i = 0$ sinon. On suppose que les X_i sont i.i.d. et $X_i \sim \mathcal{B}(p)$. On sait également par expérience que $p \in \{1/3, 1/2\}$ avec $p = 1/2$ si la pièce est équilibrée et $p = 1/3$ si la pièce est truquée. On cherche à estimer le paramètre p . Supposons que l'on observe pour les deux premiers lancers : 1 puis 0.

- Si $p = 1/2$, $\mathbb{P}_{p=1/2}(X_1 = 1, X_2 = 0) = \frac{1}{2} \frac{1}{2} = 0,25$.
- Si $p = 1/3$, $\mathbb{P}_{p=1/3}(X_1 = 1, X_2 = 0) = \frac{1}{3} \frac{2}{3} \simeq 0,222$.

Donc, au vu des deux premiers lancers, il est plus vraisemblable que p soit égal à $1/2$. Si on observe 0 au troisième lancer, on a alors

- si $p = 1/2$, $\mathbb{P}_{p=1/2}(X_1 = 1, X_2 = 0, X_3 = 0) = \frac{1}{2} \frac{1}{2} \frac{1}{2} = 0,125$.
- si $p = 1/3$, $\mathbb{P}_{p=1/3}(X_1 = 1, X_2 = 0, X_3 = 0) = \frac{1}{3} \frac{2}{3} \frac{2}{3} \simeq 0.149$.

Dans ce cas, il est plus vraisemblable que p soit égal à $1/3$.

Plus généralement, si on observe x_1, x_2, \dots, x_n pour les n premiers lancers, le p le plus vraisemblable est celui qui maximise

$$\mathbb{P}_p(X_1 = x_1, \dots, X_n = x_n).$$

3.4.2 La méthode du maximum de vraisemblance

Soient X_1, \dots, X_n n variables i.i.d. de loi discrète dépendant d'un paramètre $\theta \in \Theta \subset \mathbb{R}^d$.

Définition 30 On suppose que l'on observe les réalisations x_1, \dots, x_n des variables aléatoires X_1, \dots, X_n . La fonction

$$\theta \mapsto \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i)$$

est appelée la vraisemblance de θ pour la réalisation (x_1, \dots, x_n) et est notée $L(\theta; x_1, \dots, x_n)$ ou simplement $L(\theta)$ s'il n'y a pas d'ambiguïté.

Définition 31 Supposons que pour toute observation (x_1, \dots, x_n) des v.a. (X_1, \dots, X_n) il existe une valeur θ , notée $\hat{\theta}_n^{MV}(x_1, \dots, x_n)$ telle que la vraisemblance soit maximale :

$$L(\hat{\theta}_n^{MV}(x_1, \dots, x_n); x_1, \dots, x_n) = \max_{\theta} L(\theta; x_1, \dots, x_n) \quad \text{i.e.} \quad \hat{\theta}_n^{MV}(x_1, \dots, x_n) = \operatorname{argmax}_{\theta} L(\theta; x_1, \dots, x_n).$$

Alors on dit que $\hat{\theta}_n^{MV}(x_1, \dots, x_n)$ est une estimation du maximum de vraisemblance de θ .

La variable aléatoire correspondante $\hat{\theta}_n^{MV}(X_1, \dots, X_n)$, notée $\hat{\theta}_n^{MV}$ ou $\hat{\theta}^{MV}$, est appelée estimateur du maximum de vraisemblance de θ (E.M.V.).

Remarque 22 On peut facilement généraliser l'estimateur du maximum de vraisemblance lorsque les variables ne sont plus indépendantes.

3.4.3 Détermination pratique de l'E.M.V.

Très souvent les calculs sont plus simples en considérant la log-vraisemblance.

Définition 32 On définit la log-vraisemblance comme le logarithme de la vraisemblance :

$$\ln L(\theta; x_1, \dots, x_n) = \sum \ln \mathbb{P}_\theta(X_i = x_i).$$

Maximiser la vraisemblance ou la log-vraisemblance revient au même puisque la fonction \ln est strictement croissante.

Souvent la recherche de l'E.M.V. passe par le calcul différentiel. Lorsque l'on est en dimension 1, la recherche de l'E.M.V. revient juste à étudier une fonction réelle à valeurs réelles. En dimension plus grande que 1, on a des conditions nécessaires et des conditions suffisantes pour obtenir l'E.M.V.

Proposition 35 On suppose que Θ est un ouvert de \mathbb{R}^d et $\ln L : \Theta \mapsto \mathbb{R}$ est deux fois différentiable sur \mathbb{R}^d . Si $\hat{\theta}$ vérifie

1.

$$\nabla \ln L(\theta; x_1, \dots, x_n)|_{\theta=\hat{\theta}} = \left(\frac{\partial}{\partial \theta_i} \ln L(\theta; x_1, \dots, x_n) \right)_{1 \leq i \leq d} \Big|_{\theta=\hat{\theta}} = 0$$

2. pour tout $\theta \in \Theta$,

$$H(\theta; x_1, \dots, x_n) = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln L(\theta; x_1, \dots, x_n) \right)_{1 \leq i, j \leq d}$$

est définie négative,

alors $\hat{\theta}$ est l'unique E.M.V.

Rappels sur la notion de matrice définie négative : Soit A une matrice réelle carrée de dimension $d \times d$.

- La matrice A est dite définie négative si pour tout $u \in \mathbb{R}^d$ non nul, $u^\top A u < 0$.
- Si A est symétrique alors A est diagonalisable et on a : A est définie négative si et seulement si toutes les valeurs propres de A sont strictement négatives.
- Si $d = 1$ A est définie positive si et seulement si $A < 0$.

Proposition 36 On suppose que Θ est un ouvert de \mathbb{R}^d et $\ln L : \Theta \mapsto \mathbb{R}$ est deux fois différentiable sur \mathbb{R}^d . Si $\hat{\theta}$ est un E.M.V alors il vérifie

1.

$$\nabla \ln L(\theta; x_1, \dots, x_n)|_{\theta=\hat{\theta}} = \left(\frac{\partial}{\partial \theta_i} \ln L(\theta; x_1, \dots, x_n) \right)_{1 \leq i \leq d} \Big|_{\theta=\hat{\theta}} = 0$$

2. $H(\theta; x_1, \dots, x_n)|_{\theta=\hat{\theta}}$ est définie négative.

En pratique on pourra se contenter de cette condition nécessaire si la condition suffisante n'est pas vérifiée.

Remarque 23

- *Parfois, il n'est pas possible de calculer explicitement l'E.M.V. On peut faire alors appel à une approximation numérique (cf cours d'optimisation).*
- *Pour l'instant, tout cela reste heuristique. Les propriétés de bonne approximation de θ seront étudiées plus tard.*

4 Variables aléatoires à densité

Dans tout ce chapitre, on considère un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

4.1 Définition, propriétés

Définition 33 Soit X une variable aléatoire réelle. X est une variable aléatoire à densité (ou continue) s'il existe une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable, appelée densité de X , telle que pour tout intervalle $[a, b] \subset \mathbb{R}$, on a

$$\mathbb{P}(X \in [a, b]) = \int_a^b f(x)dx.$$

Remarque 24 • Si X est une variable aléatoire à densité, de densité f , alors on a

$$\mathbb{P}(X \in B) = \int_B f(x)dx, \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

- Pour tout $x \in \mathbb{R}$, $\mathbb{P}(X = x) = 0$. En particulier une variable aléatoire à densité n'est pas discrète et une v.a. discrète n'est pas à densité.
- La densité f d'une variable aléatoire à densité X caractérise la loi \mathbb{P}_X .

Proposition 37 Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable est une densité si et seulement si

1. $f \geq 0$ p.p.
2. $\int_{\mathbb{R}} f(x)dx = 1$.

Preuve.

1. On commence par considérer une variable aléatoire X de densité f . S'il existe $B \in \mathcal{B}(\mathbb{R})$ tel que $f(x) < 0$ pour tous les $x \in B$, alors $0 \leq \mathbb{P}(X \in B) = \int_B f(x)dx \leq 0$ et donc $\int_B f(x)dx = 0$ ce qui signifie que B est de mesure (de Lebesgue) nulle. De plus on a

$$1 = \mathbb{P}(X \in \mathbb{R}) = \int_{\mathbb{R}} f(x)dx.$$

2. Inversement on montre facilement que si l'on prend une fonction mesurable f vérifiant les deux points de la proposition, alors

$$\mathbb{P}_X(B) := \int_B f(x)dx, \quad \forall B \in \mathcal{B}(\mathbb{R})$$

définie bien une mesure de probabilité.¹

□

4.1.1 Fonction de répartition

Proposition 38 Soit X une variable aléatoire à densité, de densité f . On note F_X sa fonction de répartition.

- F_X est la primitive de f nulle en $-\infty$; i.e.

$$F_X(t) = \int_{-\infty}^t f(x)dx, \quad \forall t \in \mathbb{R},$$

- F_X est continue,
- si f est continue en un point $t \in \mathbb{R}$, alors F_X est dérivable en ce point et $F'_X(t) = f(t)$.

Preuve. Il suffit juste de remarquer que

$$F_X(t) = \mathbb{P}(X \in]-\infty, t]) = \int_{-\infty}^t f(x)dx.$$

□

Proposition 39 Soit X une variable aléatoire réelle de fonction de répartition F_X vérifiant :

1. F_X est continue,
2. F_X est dérivable sauf éventuellement en un nombre fini (ou dénombrable) de points.

Alors X est une variable à densité et sa densité vaut F'_X .

Remarque 25 Il existe des variables aléatoires dont la fonction de répartition est continue et qui ne sont pas à densité...

Proposition 40 Soit X une variable aléatoire à densité, de densité f . Il y a équivalence entre les assertions suivantes :

- $\mathbb{P}(X \in [a, b]) = 1$,
- $f(x) = 0$ p.p. sur $\mathbb{R} \setminus [a, b]$,
- $F_X(a) = 0$ et $F_X(b) = 1$.

Remarque 26 La première et la dernière assertion sont équivalentes pour n'importe quelle variable aléatoire, i.e. non nécessairement à densité.

1. L'existence d'une variable aléatoire X ayant une telle loi reste toutefois à démontrer...

4.2 Lois à densité classiques

4.2.1 Loi uniforme

Définition 34 Soient $a, b \in \mathbb{R}$ avec $a < b$. X suit la loi uniforme sur l'intervalle $[a, b]$ si sa densité vaut

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

On note $X \sim \mathcal{U}([a, b])$.

4.2.2 Loi exponentielle

Définition 35 Soient $\lambda > 0$. X suit la loi exponentielle de paramètre λ si

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

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

4.2.3 Loi de Cauchy

Définition 36 Soient $c > 0$. X suit la loi de Cauchy de paramètre c si

$$f_X(x) = \frac{1}{\pi} \frac{c}{c^2 + x^2}.$$

On note $X \sim \mathcal{C}(c)$.

4.2.4 Loi normale

Définition 37 Soient $m \in \mathbb{R}$ et $\sigma > 0$. X suit la loi normale de paramètres m et σ^2 si

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

On note $X \sim \mathcal{N}(m, \sigma^2)$.

4.3 Espérance et variance

Proposition 41 Soit X une variable aléatoire à densité, de densité f . X est intégrable si et seulement si

$$\int_{\mathbb{R}} |x|f(x)dx < +\infty.$$

Dans ce cas, l'espérance de X est alors égale à

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

Proposition 42 Soit X une variable aléatoire à densité, de densité f , et h une fonction mesurable réelle. On suppose que

$$\int_{\mathbb{R}} |h(x)|f(x)dx < +\infty.$$

Alors, si on pose $Y = h(X)$, Y est une variable aléatoire intégrable et

$$\mathbb{E}[Y] = \mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx.$$

Exemple 13

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x)dx.$$

Proposition 43 Soit X une variable aléatoire à densité, de densité f et de carré intégrable. Alors on a

$$\text{Var}(X) = \int_{\mathbb{R}} x^2 f(x)dx - \left(\int_{\mathbb{R}} xf(x)dx \right)^2.$$

Exemple 14 • Si $X \sim \mathcal{U}([a, b])$,

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(a-b)^2}{12}.$$

• Si $X \sim \mathcal{E}(\lambda)$,

$$\mathbb{E}[X] = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}.$$

• Si $X \sim \mathcal{C}(c)$, X n'est pas intégrable, elle n'a donc pas d'espérance et de variance.

• Si $X \sim \mathcal{N}(m, \sigma^2)$,

$$\mathbb{E}[X] = m, \quad \text{Var}(X) = \sigma^2.$$

4.4 Transformation

On considère X une variable aléatoire à densité, de densité f , h une fonction mesurable réelle et on pose $Y = h(X)$. Il est clair que Y est une variable aléatoire réelle.

- Est-ce que Y est une variable aléatoire à densité ?
- Si oui, peut-on calculer sa densité, en fonction de f et h ?

Concernant la première question, la réponse est clairement non : Il suffit par exemple de prendre la fonction nulle $h = 0$ pour obtenir $Y = 0$ qui est une variable aléatoire discrète...

Concernant la seconde question, cela revient à se demander comment montrer qu'une v.a. est à densité. Une réponse possible est de calculer sa fonction de répartition puis d'utiliser la Proposition 39. Une autre possibilité est d'utiliser la méthode de la fonction muette que l'on peut voir comme une réciproque de la Proposition 42.

Proposition 44 (Méthode de la fonction muette) *Soit X une variable aléatoire réelle. S'il existe une fonction mesurable $g : \mathbb{R} \rightarrow \mathbb{R}$ telle que, pour toute fonction mesurable positive ou bornée $\phi : \mathbb{R} \rightarrow \mathbb{R}$, on a*

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

alors X est une variable aléatoire à densité, de densité g .

Revenons à notre exemple $Y = h(X)$. Soit $\phi : \mathbb{R} \rightarrow \mathbb{R}$ une fonction mesurable positive ou bornée quelconque. Alors on a

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

que l'on veut écrire sous la forme $\int_{\mathbb{R}} \phi(y)g(y)dy$. Pour cela il suffit de faire le changement de variable $y = h(x)$, si cela est possible !

Exemple 15 • Si $X \sim \mathcal{E}(\lambda)$, X^2 à densité, de densité ...

- Si $X \sim \mathcal{N}(0, 1)$, X^2 à densité, de densité ...

Exemple 16 *Soit X une v.a. à densité, de densité f .*

- Si $Y = aX + b$ avec $b \neq 0$, alors Y de densité

$$f_Y(x) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right).$$

- Si $Y = \frac{1}{aX+b}$ avec $a \neq 0$, alors Y de densité

$$f_Y(x) = \frac{1}{|a|y^2} f\left(\frac{1}{ay} - \frac{b}{a}\right).$$

En particulier,

- Si $X \sim \mathcal{N}(0, 1)$ et $a \neq 0$, $aX + b \sim \mathcal{N}(b, a^2)$.
- Si $X \sim \mathcal{N}(m, \sigma^2)$, $\frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$.
- Si $X \sim \mathcal{C}(1)$, $1/X \sim \mathcal{C}(1)$.

4.5 Vecteurs aléatoires à densité

Définition 38 Soit X un vecteur aléatoire réel à valeurs dans \mathbb{R}^d . X est un vecteur aléatoire à densité (ou continu) s'il existe une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable, appelée densité de X , telle que pour tout borélien $B \in \mathcal{B}(\mathbb{R}^d)$, on a

$$\mathbb{P}(X \in B) = \int_B f(x) dx = \int_B f(x_1, \dots, x_d) dx_1 \dots dx_d.$$

Comme précédemment, on peut se contenter des pavés de \mathbb{R}^d en lieu et place des boréliens.

Proposition 45 Une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable est une densité si et seulement si

1. $f \geq 0$ p.p.
2. $\int_{\mathbb{R}^d} f(x) dx = 1$.

Proposition 46 Si $X = (X_1, \dots, X_d)$ est un vecteur aléatoire à densité, de densité f , alors les composantes X_i du vecteur sont des variables aléatoires à densités, de densités marginales

$$f_i(x_i) = \int_{\mathbb{R}^{d-1}} f(x_1, \dots, x_d) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

La réciproque est fautive ! Si X_1, \dots, X_d sont des v.a. à densité, (X_1, \dots, X_d) n'est pas nécessairement un vecteur aléatoire à densité.

On revient à la question initiale en la modifiant légèrement : on considère X un vecteur aléatoire à densité, de densité f , $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ une fonction mesurable et on pose $Y = h(X)$. Est-ce que Y est un vecteur aléatoire à densité et, si oui, que vaut sa densité ?

Proposition 47 (Méthode de la fonction muette) Soit X un vecteur aléatoire de \mathbb{R}^d . Alors X est un vecteur aléatoire à densité, de densité f , si et seulement si pour toute fonction mesurable $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ bornée ou positive on a

$$\mathbb{E}[\phi(X)] = \int_{\mathbb{R}^d} \phi(x)f(x)dx.$$

Soit $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction mesurable positive ou bornée quelconque. Alors

$$\mathbb{E}[\phi(Y)] = \mathbb{E}[\phi(h(X))] = \int_{\mathbb{R}^d} \phi(h(x))f(x)dx$$

et donc on cherche à nouveau à faire le changement de variable $y = h(x)$ pour obtenir une quantité sous la forme $\int_{\mathbb{R}^d} \phi(y)g(y)dy$ pour conclure.

Théorème 4 Soient U et V deux ouverts de \mathbb{R}^d , $h : U \rightarrow V$ un C^1 -difféomorphisme (bijective, C^1 , d'inverse C^1), X une variable aléatoire à valeurs dans U et de densité f , $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction mesurable positive ou bornée quelconque. On pose $Y = h(X)$. Alors on a

$$\int_U \phi(h(x))f(x)dx = \int_V \phi(y)f(h^{-1}(y))|detJ_{h^{-1}}(y)|dy = \int_{\mathbb{R}^d} \phi(y)f(h^{-1}(y))|detJ_{h^{-1}}(y)|\mathbb{1}_V(y)dy$$

et donc Y est un vecteur aléatoire à densité, de densité

$$y \mapsto f(h^{-1}(y))|detJ_{h^{-1}}(y)|\mathbb{1}_V(y).$$

4.6 Indépendance

Proposition 48 Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires réelles à densité. Pour tout $n \geq 2$, X_1, \dots, X_n sont indépendantes si et seulement si X est un vecteur à densité de densité f égale au produit de ses densités marginales, i.e.

$$f(x) = \prod_{i=1}^n f_i(x_i), \forall x = (x_1, \dots, x_n) \in \mathbb{R}^n$$

avec f_i la densité de X_i .

Remarque 27 Si $X = (X_1, \dots, X_n)$ admet une densité f de la forme

$$f(x_1, \dots, x_n) = \prod_{i=1}^n g_i(x_i)$$

où g_1, \dots, g_n sont des fonctions mesurables positives, alors X_1, \dots, X_n sont indépendantes et pour tout $1 \leq i \leq n$ X_i a une densité f_i proportionnelle à g_i . On trouve f_i en écrivant $\int_{\mathbb{R}} f_i(x_i) dx_i = 1$:

$$f_i = \frac{1}{\int_{\mathbb{R}} g_i(t) dt} g_i.$$

Proposition 49 Soient X et Y deux v.a. à densité, indépendantes, de densité f_X et f_Y . Alors $X + Y$ est une variable aléatoire à densité de densité $f_X \star f_Y$ donnée par

$$f_X \star f_Y(u) := \int_{\mathbb{R}} f_X(u-t) f_Y(t) dt.$$

4.7 Fonction caractéristique

Définition 39 Soit X une variable aléatoire réelle. On appelle fonction caractéristique de X , la fonction $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$, donnée par

$$\phi_X(t) = \mathbb{E}[e^{itX}], \quad \forall t \in \mathbb{R}.$$

Proposition 50 • ϕ_X est bien définie sur \mathbb{R} .

- $\phi_X(0) = 1$ et $|\phi_X(t)| \leq 1$ pour tout $t \in \mathbb{R}$.
- Si X est une v.a. discrète,

$$\phi_X(t) = \sum_{x \in X(\Omega)} e^{itx} \mathbb{P}(X = x).$$

- Si X est une v.a. de densité f ,

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx,$$

c'est la transformée de Fourier de f .

Exemple 17 • Si $X \sim \mathcal{B}(n, p)$, $\phi_X(t) = (1 - p + pe^{it})^n$.

- Si $X \sim \mathcal{P}(\lambda)$, $\phi_X(t) = e^{\lambda(e^{it}-1)}$.
- Si $X \sim \mathcal{G}(p)$, $\phi_X(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$.
- Si $X \sim \mathcal{U}([a, b])$, $\phi_X(t) = \frac{(e^{itb}-e^{ita})}{it(b-a)}$ si $t \neq 0$.
- Si $X \sim \mathcal{E}(\lambda)$, $\phi_X(t) = \frac{\lambda}{\lambda-it}$.
- Si $X \sim \mathcal{C}(c)$, $\phi_X(t) = e^{-c|t|}$.

- Si $X \sim \mathcal{N}(m, \sigma^2)$, $\phi_X(t) = e^{itm - \frac{t^2 \sigma^2}{2}}$.

Proposition 51 Soit X une variable aléatoire réelle de fonction caractéristique ϕ .

- ϕ caractérise la loi de X .
- ϕ est uniformément continue sur \mathbb{R} .
- $\forall n \geq 1$, si $\mathbb{E}[|X|^n] < +\infty$, alors ϕ est dérivable jusqu'à l'ordre n et pour tout $1 \leq k \leq n$,

$$\phi^{(k)}(t) = i^k \mathbb{E}[X^k e^{itX}], \quad \phi^{(k)}(0) = i^k \mathbb{E}[X^k].$$

De plus, au voisinage de 0, on a

$$\phi(t) = \sum_{k=0}^n \frac{(it)^k}{k!} \mathbb{E}[X^k] + o(|t|^n).$$

- Si ϕ est intégrable, alors X admet une densité de probabilité f continue et bornée, donnée par la formule d'inversion de Fourier

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi(t) dt.$$

Définition 40 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans \mathbb{R}^n . On appelle fonction caractéristique de X la fonction $\phi : \mathbb{R}^n \rightarrow \mathbb{C}$ définie par

$$\phi(t) := \mathbb{E}[e^{i\langle t, X \rangle}] = \mathbb{E}[e^{i \sum_{k=1}^n t_k X_k}], \quad \forall t = (t_1, \dots, t_n) \in \mathbb{R}^n.$$

En particulier, si X est discret, on a

$$\phi(t) = \sum_{x \in X(\Omega)} e^{i\langle t, x \rangle} \mathbb{P}(X = x) = \sum_{x_1 \in X_1(\Omega)} \dots \sum_{x_n \in X_n(\Omega)} e^{i \sum_{k=1}^n t_k x_k} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n),$$

et si X est à densité de densité f , on a

$$\phi(t) = \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} f(x) dx = \int_{\mathbb{R}^n} e^{i \sum_{k=1}^n t_k x_k} f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Proposition 52 Comme dans le cas unidimensionnel, la fonction caractéristique d'un vecteur aléatoire caractérise sa loi.

Proposition 53 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire à valeurs dans \mathbb{R}^n . X_1, \dots, X_n sont indépendantes si et seulement si la fonction caractéristique de X est égale au produit des fonctions caractéristiques de X_1, \dots, X_n , c'est à dire que pour tout $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ on a

$$\phi(t) = \mathbb{E}[e^{i \sum_{k=1}^n t_k X_k}] = \mathbb{E}\left[\prod_{k=1}^n e^{it_k X_k}\right] = \prod_{k=1}^n \mathbb{E}[e^{it_k X_k}] = \prod_{k=1}^n \phi_{X_k}(t_k).$$

Proposition 54 Soient X_1, \dots, X_n des variables aléatoires réelles indépendantes et soit $S_n = \sum_{k=1}^n X_k$. Alors, pour tout $t \in \mathbb{R}$, on a

$$\phi_{S_n}(t) = \mathbb{E}[e^{itS_n}] = \prod_{k=1}^n \mathbb{E}[e^{itX_k}] = \prod_{k=1}^n \phi_{X_k}(t).$$

En particulier, si les variables sont i.i.d. alors $\phi_{S_n}(t) = (\phi_{X_1}(t))^n$.

Remarque 28 La réciproque de la Proposition 54 est fautive. Prenons l'exemple de $X_1 = X_2 = \dots = X_n = X$ où $X \sim \mathcal{C}(c)$ et $S_n = nX$. On a

$$\phi_{S_n}(t) = e^{-cn|t|} = (e^{-c|t|})^n$$

mais X_1, \dots, X_n ne sont pas indépendantes !

4.8 Estimateur du maximum de vraisemblance

Une variable à densité peut être vue comme une limite infinitésimale de variables discrètes. Ainsi, par analogie avec le cas discret on définit la vraisemblance de la façon suivante.

Soient X_1, \dots, X_n n variables i.i.d. de loi de densité f dépendant d'un paramètre θ .

Définition 41 On suppose que l'on observe les réalisations x_1, \dots, x_n des variables aléatoires X_1, \dots, X_n . La fonction

$$\theta \mapsto \prod_{i=1}^n f_{\theta}(x_i)$$

est appelée la vraisemblance de θ pour la réalisation (x_1, \dots, x_n) et est notée $L(\theta; x_1, \dots, x_n)$ ou simplement $L(\theta)$ s'il n'y a pas d'ambiguïté.

Définition 42 Supposons que pour toute observation (x_1, \dots, x_n) des v.a. (X_1, \dots, X_n) il existe une seule valeur θ , notée $\hat{\theta}_n^{MV}(x_1, \dots, x_n)$ telle que la vraisemblance soit maximale :

$$L(\hat{\theta}_n^{MV}(x_1, \dots, x_n); x_1, \dots, x_n) = \max_{\theta} L(\theta; x_1, \dots, x_n) \quad \text{i.e.} \quad \hat{\theta}_n^{MV}(x_1, \dots, x_n) = \operatorname{argmax}_{\theta} L(\theta; x_1, \dots, x_n).$$

Alors on dit que $\hat{\theta}_n^{MV}(x_1, \dots, x_n)$ est une estimation du maximum de vraisemblance de θ .

La variable aléatoire correspondante $\hat{\theta}_n^{MV}(X_1, \dots, X_n)$, notée $\hat{\theta}_n^{MV}$ ou $\hat{\theta}^{MV}$, est appelée estimateur du maximum de vraisemblance de θ .

Comme dans le cas discret, il est préférable de faire les calculs en considérant la log-vraisemblance.

Exemple : Soient X_1, \dots, X_n i.i.d. de loi $\mathcal{E}(\lambda)$.

$$\ln L(\lambda; x_1, \dots, x_n) = \sum_{i=1}^n \ln (\lambda e^{-\lambda x_i} \mathbb{1}_{x_i \geq 0}) = n \ln \lambda - \lambda \sum_{i=1}^n x_i.$$

alors

$$\frac{\partial}{\partial \lambda} \ln L(\lambda; x_1, \dots, x_n) = \frac{n}{\lambda} - \sum_{i=1}^n x_i$$

et

$$\frac{\partial^2}{\partial^2 \lambda} \ln L(\lambda; x_1, \dots, x_n) = -\frac{n}{\lambda^2} < 0.$$

Donc l'E.M.V. est donné par

$$\hat{\theta}_n^{MV} = \frac{n}{\sum_{i=1}^n X_i}.$$

5 Estimation paramétrique

Dans tout ce chapitre, on se place dans un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ sur lequel on considère (X_1, X_2, \dots, X_n) , n v.a. i.i.d. de loi inconnue. Le but est d'obtenir des informations sur la loi des $(X_i)_{1 \leq i \leq n}$ en utilisant les informations apportées par un tirage $(x_1, \dots, x_n) := (X_1(\omega), \dots, X_n(\omega))$.

5.1 Modèle statistique

5.1.1 Modèle statistique paramétrique / non paramétrique

Les v.a. $(X_i)_{1 \leq i \leq n}$ sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ et sont à valeur dans \mathbb{R}^d . On note \mathbb{P}_X la loi commune des (X_i) . On rappelle que \mathbb{P}_X est une mesure de probabilité définie sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ avec $\mathcal{B}(\mathbb{R}^d)$ l'ensemble des boréliens de \mathbb{R}^d .

Hypothèse : On suppose que la loi \mathbb{P}_X appartient à une famille de probabilités \mathcal{P} (sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$) indicées par un paramètre inconnu θ :

$$\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}.$$

Définition 43 La famille $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ est appelée modèle statistique.

Définition 44 1. Si le paramètre θ est de dimension finie : \mathcal{P} est un modèle paramétrique.

2. Si Θ n'est pas de dimension finie (sous ensemble d'un espace vectoriel de dimension infinie) : \mathcal{P} est un modèle non paramétrique.

Exemples :

- Modèles paramétriques : $\mathbb{P}_X \sim \mathcal{P}(\lambda)$ avec λ inconnu, $\mathbb{P}_X \sim \mathcal{N}(m, 1)$ avec m inconnu, ...
- Modèles non paramétriques : \mathbb{P}_X est une loi réelle à densité, \mathbb{P}_X est une loi réelle de moyenne nulle, ...

En pratique, si \mathbb{P}_X appartient à une famille de loi réelle usuelle (Bernoulli, normale, Poisson,...) mais dont certains paramètres sont inconnus alors le modèle est paramétrique.

Remarque 29 Le choix de la famille \mathcal{P} se fait **a priori** par le statisticien en fonction de ses connaissances du phénomène étudié. Exemple : loi exponentielle lorsque l'on veut modéliser un phénomène sans mémoire, par exemple l'arrivée de clients à la poste.

Dans toute la suite on ne considère que le cadre des modèles paramétriques.

5.1.2 Échantillon, réalisations

Définition 45 Soient (X_1, \dots, X_n) des v.a. i.i.d. de même loi que X , notée \mathbb{P}_X . On dit que (X_1, \dots, X_n) est un échantillon (ou n-échantillon) de la loi de X (ou du modèle $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$).

Définition 46 On appelle réalisation (ou observation) de l'échantillon (X_1, \dots, X_n) et on note (x_1, \dots, x_n) toute valeur possible prise par ces v.a. Autrement dit on a $(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega))$ pour un certain $\omega \in \Omega$ et $(x_1, \dots, x_n) \in X(\Omega)^n$.

5.2 Le cadre paramétrique

\mathbb{P}_X appartient au modèle paramétrique $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$. En particulier, il existe un entier $k \in \mathbb{N}^*$ connu tel que l'espace paramétrique Θ est inclu dans \mathbb{R}^k .

k est la dimension du paramètre inconnu θ :

- pour $k = 1$, on parle de paramètre uni-dimensionnel.
- pour $k > 1$, on parle de paramètre multi-dimensionnel : $\theta = (\theta_1, \dots, \theta_k)$.

Définition 47 On dit qu'un modèle $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est identifiable si

« $\theta_1 \neq \theta_2$ implique $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$ » soit encore « $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$ implique $\theta_1 = \theta_2$ ».

Autrement dit : le paramètre θ détermine de manière unique la loi \mathbb{P}_θ .

Remarque 30 • L'estimation correcte du paramètre θ permet d'obtenir une estimation correcte de la loi de X et donc une estimation de toutes les caractéristiques de cette loi : espérance, variance, médiane, ... Le paramètre θ n'est pas forcément une caractéristique usuelle pour la loi (ex : $\mathcal{G}(p)$). Par contre toutes les caractéristiques usuelles sont des fonctions de θ .

- Il n'existe pas qu'une seule façon de paramétrer une famille de lois.
- Le paramètre θ est inconnu mais il ne fluctue pas, il n'est pas aléatoire, c'est une constante. C'est un point de vue différent de la statistique bayésienne.

5.3 Estimateur

Étant donné un modèle paramétrique $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$, on estime θ ou une fonction de θ en utilisant les informations apportées par l'échantillon (X_1, \dots, X_n) par l'intermédiaire d'une fonction de (X_1, \dots, X_n) appelée statistique.

5.3.1 Définitions

Définition 48 On appelle statistique toute variable aléatoire T_n qui est une fonction, qui **ne dépend pas de** θ , de l'échantillon c'est à dire que T_n s'écrit : $T_n = t(X_1, \dots, X_n)$ où $t : (\mathbb{R}^d)^n \rightarrow \mathbb{R}^a$ est une fonction mesurable connue, ne dépendant pas de θ . Ainsi, T_n est une variable aléatoire : $T_n(\omega) = t(X_1(\omega), \dots, X_n(\omega))$.

Remarque 31 Une statistique T_n ne fait pas apparaître le paramètre θ mais sa loi va dépendre de θ !

Définition 49 Une statistique T_n qui est à valeur dans Θ est appelée estimateur de θ . Une réalisation de T_n sur l'échantillon, qui s'écrit $t_n = t(x_1, \dots, x_n)$ si (x_1, \dots, x_n) est la réalisation associée de l'échantillon, sera appelée estimation de θ .

Plus généralement, si on cherche à estimer une fonction $g(\theta)$ du paramètre inconnu θ alors un estimateur de $g(\theta)$ est une variable aléatoire à valeur dans $g(\Theta)$, qui ne dépend pas de θ mais qui est fonction de l'échantillon (X_1, \dots, X_n) .

Remarque 32 *estimateur* \neq *estimation* : un estimateur est une v.a. tandis qu'une estimation est une valeur déterministe.

Exemples :

- Lorsque l'on considère le modèle paramétrique $\mathcal{B}(p)$ on a vu que l'estimateur du maximum de vraisemblance est donné par

$$T_n(X_1, \dots, X_n) = \hat{p}_n^{MV} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n, \quad \text{moyenne empirique de l'échantillon.}$$

L'E.M.V. est bien un estimateur de p (attention tout de même on peut sortir de Θ ...). Étant donnée une réalisation (x_1, \dots, x_n) de l'échantillon, la valeur observée (ou réalisation) de \bar{X}_n est le réel : $\frac{1}{n} \sum_{i=1}^n x_i$ que l'on pourra noter naturellement \bar{x}_n (Attention aux notation!).

- On peut trouver beaucoup d'autres estimateurs de p : $t(X_1, \dots, X_n) = 1/2$, $t(X_1, \dots, X_n) = X_1$, $t(X_1, \dots, X_n) = \max_{1 \leq i \leq n} X_i, \dots$ Ces exemples semblent naïfs mais ils correspondent bien à la définition d'un estimateur.

La théorie de l'estimation consiste à :

- donner des méthodes pour construire un ou des estimateurs de θ (E.M.V., méthode des moments,...),
- étudier les propriétés des estimateurs (convergence lorsque n est grand,...)
- donner des critères de qualité pour les comparer et choisir le meilleur.

5.3.2 Biais

Intuitivement, on s'attend à ce qu'un estimateur T_n de θ soit un "bon estimateur" s'il est suffisamment proche, en un certain sens, de θ . Dans toute la suite on suppose que T_n est de carré intégrable et donc que $\mathbb{E}_\theta[T_n]$ et $\text{Var}_\theta[T_n]$ existent pour tout θ .

Définition 50 On appelle biais de l'estimateur T_n pour le paramètre θ la quantité :

$$b_\theta(T_n) = \mathbb{E}_\theta[T_n] - \theta.$$

Le biais mesure l'erreur systématique faite en estimant θ .

Par exemple, si $b_\theta(T_n) < 0$, cela signifie que T_n aura tendance à sous-estimer θ .

Définition 51 On dit que T_n est un estimateur sans biais de θ si

$$b_\theta(T_n) = 0, \quad \text{pour tout } \theta \in \Theta.$$

En abrégé on écrira que T_n est un E.S.B. de θ . Au contraire, s'il existe $\theta \in \Theta$ tel que $b_\theta(T_n) \neq 0$, on dit que l'estimateur est biaisé.

Définition 52 Si T_n est un estimateur de θ tel que

- i) T_n est biaisé,
- ii) pour tout $\theta \in \Theta$, $b_\theta(T_n) \rightarrow 0$ quand $n \rightarrow +\infty$,

alors on dit que T_n est asymptotiquement sans biais pour θ .

Exemple : Si la v.a. admet une espérance μ inconnue alors la moyenne empirique est un E.S.B. de μ . En effet on a par linéarité de l'espérance

$$\mathbb{E} \left[\frac{X_1 + \dots + X_n}{n} \right] = \frac{\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]}{n} = \mu.$$

5.3.3 Risque quadratique

On muni \mathbb{R}^k de la norme euclidienne $\|\cdot\|$ associée au produit scalaire usuel.

Définition 53 On appelle risque quadratique (ou erreur quadratique moyenne) de l'estimateur T_n pour θ la quantité notée $R_\theta(T_n)$ et définie pour tout $\theta \in \Theta$ par

$$R_\theta(T_n) = \mathbb{E}_\theta[\|T_n - \theta\|^2].$$

Proposition 55 (Décomposition biais-variance du risque)

$$R_\theta(T_n) = \|b_\theta(T_n)\|^2 + \text{Tr}(\text{Var}_\theta(T_n)), \quad \forall \theta \in \Theta.$$

En particulier, si T_n est un estimateur sans biais de θ alors $R_\theta(T_n) = \text{Tr}(\text{Var}_\theta(T_n))$.

Preuve.

$$\begin{aligned} R_\theta(T_n) &= \mathbb{E}_\theta[\|T_n - b_\theta(T_n) - \theta + b_\theta(T_n)\|^2] \\ &= \mathbb{E}_\theta[\|T_n - \mathbb{E}_\theta[T_n]\|^2] + 2\mathbb{E}_\theta[\langle T_n - b_\theta(T_n) - \theta, b_\theta(T_n) \rangle] + \mathbb{E}_\theta[\|b_\theta(T_n)\|^2] \\ &= \sum_{i=1}^k \text{Var}_\theta[T_n^i] + 0 + \|b_\theta(T_n)\|^2. \end{aligned}$$

□

Remarque 33 • En dimension 1, $\text{Tr}(\text{Var}_\theta(T_n)) = \text{Var}_\theta(T_n)$.

- Le risque quadratique mesure la distance au carré à laquelle T_n se situe en moyenne par rapport à θ . On peut comprendre le rôle du biais et de la variance en faisant l'analogie avec le tir sur une cible : le biais correspond à un décalage systématique et la variance mesure la dispersion des tirs.

Le risque quadratique est le critère généralement utilisé pour mesurer la qualité d'un estimateur et choisir entre deux estimateurs d'un même paramètre θ :

- La variance d'un estimateur mesure sa variabilité (dim 1). Si l'estimateur est sans biais, cette variabilité est autour de θ . Si on veut estimer correctement θ il ne faut pas que cette variabilité soit trop forte. En pratique : si on observe plusieurs jeux d'observations, on obtient une estimation de θ pour chacun d'entre eux. Alors si l'estimateur est de faible variance, ces estimations seront toutes proches les unes des autres, et s'il est sans biais leur moyenne sera très proche de θ .
- Pour sélectionner le meilleur estimateur, l'idée est de minimiser le risque quadratique, c'est à dire le biais et la variance. En particulier, entre deux E.S.B., le meilleur est celui qui à la variance la plus faible. Nous verrons que la variance d'un E.S.B. ne peut pas descendre en dessous d'une certaine borne.

Définition 54 Soient T_n^1 et T_n^2 deux estimateurs du paramètre θ . On dit que T_n^1 est préférable à T_n^2 (ou que T_n^1 domine T_n^2) si :

- Pour tout $\theta \in \Theta$, $R_\theta(T_n^1) \leq R_\theta(T_n^2)$,
- l'inégalité est stricte pour au moins une valeur de θ : il existe $\theta' \in \Theta$, $R_{\theta'}(T_n^1) < R_{\theta'}(T_n^2)$.

Un estimateur sans biais de variance minimale (E.S.B.V.M.) est un bon candidat pour être le meilleur estimateur possible de θ

Remarque 34 • *Un E.S.B.V.M. n'existe pas forcément.*

- *On a un critère suffisant pour avoir un E.S.B.V.M. grâce à l'information de Fischer (cf suite du cours).*
- *Un estimateur sans biais n'est pas gage de qualité : un estimateur biaisé peut avoir un risque quadratique plus faible.*

On peut généraliser à l'estimation de $g(\theta)$. Le risque quadratique devient :

$$R_{g(\theta)} := \mathbb{E}_\theta[|T_n - g(\theta)|^2].$$

Attention : « T_n bon estimateur de θ » n'implique pas forcément « $g(T_n)$ bon estimateur de $g(\theta)$ ». Par exemple,

$$\mathbb{E}_\theta[T_n] = \theta \not\Rightarrow \mathbb{E}_\theta[g(T_n)] = g(\theta).$$