

CONTRIBUTEURS

Rappels

Rappel : Série de Taylor

La série de Taylor de $(1 - t)^{-r}$ est $\sum_{i=0}^{\infty} \binom{r+i-1}{i} t^i$.

Rappel : Théorème du binôme

Pour $n \in \mathbb{N}$, $(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$.

Variables aléatoires

Notions aléatoires

1 Notion d'expérience aléatoire

Cadre dans lequel on observe différentes actions dues au hasard.

Notation

ω Le **résultat** d'une expérience aléatoire, alias *épreuve* ou *issue*.

Ω L'**ensemble des résultats possibles**.

- › Il s'ensuit que $\omega \in \Omega$.
- › Par exemple, pour le lancer d'un dé où l'on désire savoir le résultat $\Omega = \{\text{pile, face}\}$.
- › On dénote par $\mathcal{P}(\Omega)$ l'**ensemble de toutes les parties** de Ω .

2 Notion d'événement aléatoire

Événement lié à une certaine expérience aléatoire.

Un événement est tout **sous-ensemble** de Ω . Par exemple, pour l'expérience aléatoire de jeter un dé on a que l'ensemble des résultats possibles $\Omega = \{1, 2, 3, 4, 5, 6\}$. L'événement A « obtenir un nombre pair » s'écrit $A = \{2, 4, 6\}$. De ceci on déduit qu'à toute propriété définie sur Ω , on associe un sous-ensemble de Ω composé de tous les ω qui vérifient la propriété.

Algèbre de Boole des événements

 Algèbre de Boole (« *boolean algebra* ») des événements

La classe \mathcal{E} des événements est l'**algèbre de Boole de parties de Ω** , si elle contient Ω et est stable par intersection, réunion et complémentation.

Note On dit habituellement algèbre plutôt qu'algèbre de Boole.

✓ Opérations logiques

Les opérations logiques que l'on peut effectuer sur les événements sont :

1. Soit les événements $A \subset \Omega$ et $B \subset \Omega$, alors :
 - > $A \cup B$ est un événement réalisé ssi **au moins un** des deux est réalisé.
 - > $A \cap B$ est un événement réalisé ssi **les deux** sont réalisés simultanément.
2. \emptyset est un événement qui ne peut être réalisé appelé l'**événement impossible**. À chaque expérience, Ω est toujours réalisé et appelé l'**événement certain**.
3. $A \subset \Omega$ est un événement.
 - > Le complément A^c ou \bar{A} est appelé **événement contraire de A** et se réalise si $\omega \notin A$.
4. La **différence de deux événements** A et B est $A \setminus B = A \cap B^c$ se réalise si A est réalisé mais pas B .
5. La **différence symétrique** de A et B est $A \Delta B = (A \setminus B) \cup (B \setminus A)$ se réalise si l'un des deux événements est réalisé mais pas l'autre.
6. Si, $\forall n \in \mathbb{N}$, l'événement A_n représente « **gagner n matches** », alors
 - > $\bigcup_{n=1}^{\infty} A_n$ représente « **gagner au moins un match** ».
 - > $\bigcap_{n=1}^{\infty} A_n^c$ représente « **ne pas gagner de matches** ».
7. Deux événements sont **incompatibles** si $A_1 \cap A_2 = \emptyset$.
 - > On peut aussi dire que les parties de Ω représentées par A_1 et A_2 sont disjointes.
 - > Si deux événements sont incompatibles, on a une somme au lieu d'une réunion avec $A_1 \cup A_2 = A_1 + A_2$ si $A_1 \cap A_2 = \emptyset$.
8. Si les événements de la suite $(A_i)_{i \in \mathbb{I}}$ forment une **partition** de Ω , on dit que ses événements $(A_i)_{i \in \mathbb{I}}$ forment un **système exhaustif** de Ω .
9. La suite d'événements $(A_n)_{n \in \mathbb{N}^*}$ est :
 - croissante** ssi $A_1 \subset A_2 \subset \dots$
 - décroissante** ssi $A_1 \supset A_2 \supset \dots$
10. Si la suite $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements d'un ensemble Ω , on représente que :
 - une infinité de A_n est réalisé** en écrivant que, quel que soit le rang $k \in \mathbb{N}^*$, il existe des événements de rang supérieur (à k) qui sont réalisés : $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$.

un nombre fini de A_n est réalisé en écrivant qu'il existe un rang tel qu'à partir de ce rang, tous les événements réalisés sont les contraires des événements A_n : $\bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n^c$.

≡ Limites de suite d'événements

Soit $(A_n)_{n \in \mathbb{N}^+}$ une suite d'événements de Ω . On définit les limites inf et sup d'événements par :

$$A_* = \liminf A_n = \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$$

$$A^* = \limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

De plus, si les ensembles A_* et A^* coïncident, alors on écrit $A = A_* = A^* = \lim_{n \rightarrow \infty} A_n$.

✓ Propositions

Soit $(A_n)_{n \in \mathbb{N}^+}$ une suite d'événements de Ω .

i) Si $A_1 \subset A_2 \subset \dots$ alors $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

ii) Si $A_1 \supset A_2 \supset \dots$ alors $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

Espaces probabilisables

Power set \mathcal{P}

Le « power set » \mathcal{P} est l'ensemble de tous les sous-ensembles d'un ensemble Ω ; il est plus facile de donner un exemple que d'expliquer en mots.

Exemple de « power set »

Soit l'ensemble $\Omega = \{a, b, c\}$, alors

$$\mathcal{P}(\Omega) = \left\{ \begin{array}{l} \{\} \\ \{a\}, \{b\}, \{c\} \\ \{a, b\}, \{a, c\}, \{b, c\} \\ \{a, b, c\} \end{array} \right\}$$

Pour un ensemble de n éléments, il y aura 2^n sous-ensembles possibles. Ceci découle du binaire! Voir [cette page](#) pour plus d'information.

Note En anglais, on appelle la tribu \mathcal{A} composée des « *events* » le « *event space* » et l'ensemble Ω composé des « *outcomes* » le « *sample space* ».

Tribu d'événements

La tribu (ou σ -algèbre) sur un ensemble Ω est un ensemble \mathcal{A} de parties de Ω tel que :

- i) $\Omega \in \mathcal{A}$.
- ii) Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.
- iii) $\forall (A_n)_{n \in \mathbb{N}^*}$ une suite d'éléments de \mathcal{A} , alors l'événement $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Espace probabilisable (ou mesurable)

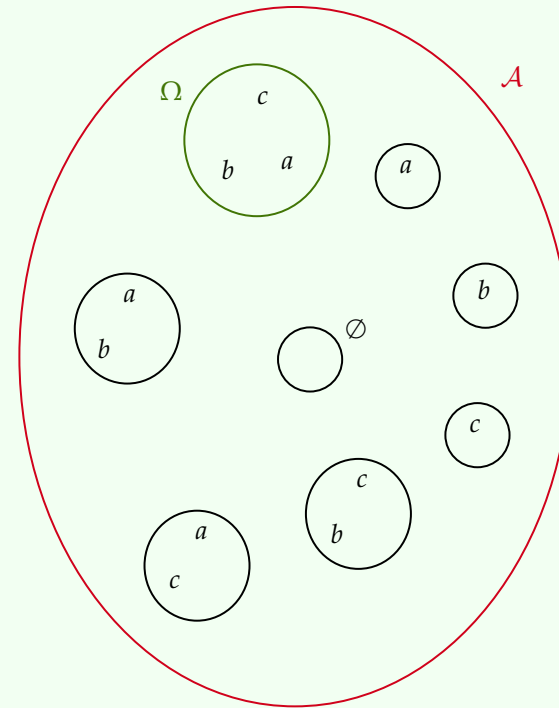
Le couple (Ω, \mathcal{A}) composé d'un ensemble Ω et une tribu \mathcal{A} sur Ω .

Les éléments de Ω sont appelés *éventualités* (« *outcomes* ») et les éléments de \mathcal{A} *événements* (« *events* »).

> En anglais, on dit « *measurable space* ».

Visualisation

Voici une visualisation de ce que représente l'espace mesurable :



On peut donc visualiser les 3 conditions dans la définition de la tribu. L'ensemble Ω est contenu, tous les événements possibles (alias toutes les combinaisons de $\{a, b, c\}$ possibles) sont contenus et tous leurs compléments sont contenus. Finalement, toute union d'événements sera contenue dans la tribu!

Propriétés de la tribu

Soit \mathcal{A} une tribu sur Ω . Alors :

- a) $\emptyset \in \mathcal{A}$.
- b) $\forall A_1, \dots, A_k \in \mathcal{A}$, alors $\bigcup_{i=1}^k A_i \in \mathcal{A}$ et $\bigcap_{i=1}^k A_i \in \mathcal{A}$.

c) $\forall (A_n)_{n \in \mathbb{N}^*}$ suite d'événements de \mathcal{A} , alors $\bigcap_{n \in \mathbb{N}^*} A_n \in \mathcal{A}$.

d) $\forall (A_n)_{n \in \mathbb{N}^*}$ suite d'événements de \mathcal{A} , alors $\liminf A_n \in \mathcal{A}$.

e) $\forall (A_n)_{n \in \mathbb{N}^*}$ suite d'événements de \mathcal{A} , alors $\limsup A_n \in \mathcal{A}$.

Note Voir la page 19 des notes de cours du chapitre 1 pour les preuves.

Variables aléatoires

Variable aléatoire

On définit une **variable aléatoire** comme une *fonction mesurable*. Pour ce faire, on définit 2 espaces mesurables :

1. On pose que le premier est (Ω, \mathcal{A}) .
2. On pose que le deuxième est tout ensemble E et sa tribu $\mathcal{E} : (E, \mathcal{E})$.
 › Habituellement, on pose que $E = \mathbb{R}$ et que $\mathcal{E} = \mathcal{B}(\mathbb{R})$.

Une fonction mesurable est une fonction qui associe les éléments de Ω aux éléments de E avec quelques propriétés additionnelles. On note qu'en associant les éléments de Ω , la fonction associe les *éventualités* et non les *événements* aux éléments de E .

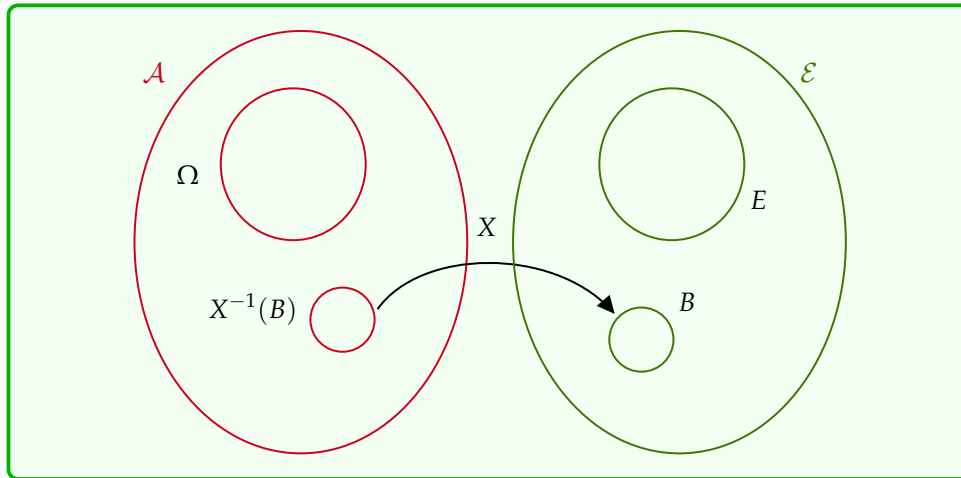
On désire avoir une correspondance entre les événements réalisés de \mathcal{A} et l'ensemble transformé d'événements \mathcal{E} . Pour ce faire, on impose que la variable aléatoire (alias, la fonction mesurable) $X : \Omega \rightarrow E$ est définie telle que l'image réciproque $X^{-1}(B)$ sur Ω de tout ensemble $B \in \mathcal{E}$ sur E :

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}, \forall B \in \mathcal{E}.$$

Donc, une **variable aléatoire réelle** est toute application à valeurs réelles $X : \Omega \rightarrow \mathbb{R}$ telle que, \forall intervalle B de \mathbb{R} , $\{X \in B\} = X^{-1}(B)$ soit un événement de la tribu \mathcal{A} .

Visualisation

On peut visualiser que l'événement B , où $B \in \mathcal{E}$, a un réciproque $X^{-1}(B)$ où $X^{-1}(B) \in \mathcal{A}$.



En posant $E = \mathbb{R}$ et $\mathcal{E} = \mathcal{B}(\mathbb{R})$, on obtient la *tribu borélienne*.

≡ Tribu borélienne

La tribu borélienne $\mathcal{B}_{\mathbb{R}}$ est la plus petite tribu de \mathbb{R} qui contient tous ses intervalles. Les éléments de $\mathcal{B}_{\mathbb{R}}$ sont appelés les *boréliens* de \mathbb{R} .

Pour une v.a. réelle X , $\forall B \in \mathcal{B}_{\mathbb{R}}$ on a $X^{-1}(B) \in \mathcal{A}$.

Bref, $(\Omega, \mathcal{A}) \xrightarrow{X} (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. On dit que la tribu $X^{-1}(\mathcal{B}_{\mathbb{R}})$ sur Ω est la *tribu des événements engendrés par X* .

Probabilités

📄 Mesure

Pour un espace mesurable (Ω, \mathcal{A}) , une fonction $\mathbb{P} : \mathcal{A} \rightarrow [0, \infty]$ s'appelle une **mesure** sur (Ω, \mathcal{A}) si :

1. Elle attribue une masse de zéro à l'ensemble vide : $\mathbb{P}(\emptyset) = 0$.
2. Elle est « *countably additive* » :
$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i), \forall A_i \in \mathcal{A}.$$

≡ Mesure de probabilité

Pour un espace probabilisable (Ω, \mathcal{A}) , on appelle **probabilité** sur (Ω, \mathcal{A}) toute application $P : \mathcal{A} \rightarrow [0, 1]$ telle que :

- i) $P(\Omega) = 1$.
- ii) $\forall (A_n)_{n \in \mathbb{N}^*}$ d'événements deux à deux disjoints,
$$P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n \in \mathbb{N}^*} P(A_n).$$

La mesure de probabilité est donc une mesure qui est **restreint** sur $[0, 1]$.

📄 Espace probabilisé

Le triplet (Ω, \mathcal{A}, P) s'appelle un **espace probabilisé** et est composé de :

Ω Le « *sample space* ».

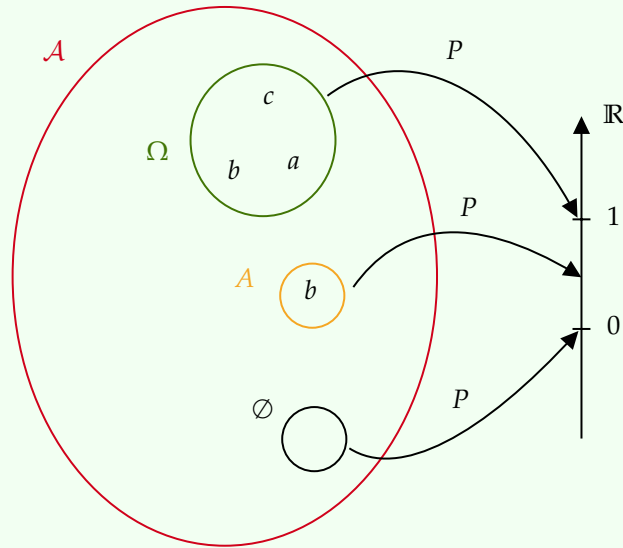
\mathcal{A} Le « *event space* ».

P La **mesure de probabilité**.

> En anglais, on dit « *probability space* ».

Visualisation

Voici une visualisation de ce que représente l'application P :



On peut donc visualiser les 2 conditions dans la définition de l'espace probabilisé. La probabilité d'observer l'ensemble Ω est de 1 car il contient tous les événements possibles. La probabilité d'un événement sera contenue entre 0 et 1 ce qui veut dire que la probabilité de quelques événements *disjoints* correspond à la somme des probabilités.

On complète la notion précédente sur l'espace borélien avec $(\Omega, \mathcal{A}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P_X)$ où P_X est appelée loi de probabilité de X . On définit $P_X(B) = P(X \in B) = P(X^{-1}(B))$.

f) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements quelconques, alors

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

g) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements tels que $A_n \downarrow \emptyset$, alors

$$P(A_n) \downarrow 0.$$

h) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements tels que $A_n \downarrow A$, alors

$$P(A_n) \downarrow P(A).$$

i) Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements tels que $A_n \uparrow A$, alors

$$P(A_n) \uparrow P(A).$$

✓ Propriétés des probabilités

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Alors :

- $P(\emptyset) = 0$.
- Si A et B sont des événements disjoints, alors $P(A \cup B) = P(A) + P(B)$.
- Si A et B sont des événements quelconques, alors $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- Si A et B sont des événements tels que $A \subset B$, alors $P(B \setminus A) = P(B) - P(A)$ et $P(A) \leq P(B)$.
- $\forall A \in \mathcal{A}, P(A^c) = 1 - P(A)$.

Lemmes de Borel-Cantelli

Lemme de Borel-Cantelli (1ère partie)

Si $(A_n)_{n \in \mathbb{N}^*}$ est une suite d'événements telle que : $\sum_{n=1}^{\infty} P(A_n) < \infty$, alors

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

Probabilité conditionnelle

Formule de Bayes (2 événements)

Soit un espace probabilisé (Ω, \mathcal{A}, P) , A et B deux événements de \mathcal{A} tels que $\Pr(A) \neq 0$, $\Pr(A^C) \neq 0$ et $\Pr(B) \neq 0$. Alors :

$$\Pr(A \setminus B) = \frac{\Pr(B|A) \Pr(A)}{\Pr(B|A) \Pr(A) + \Pr(B|A^C) \Pr(A^C)}$$

Théorème des probabilités totales

Soit un espace probabilisé (Ω, \mathcal{A}, P) et $(A_i)_{i \in \mathbb{N}}$ une partition de Ω telle que $\forall i, \Pr(A_i) \neq 0$. Alors, $\forall B \in \mathcal{A}$, $\Pr(B) = \sum_{i \in \mathbb{N}} \Pr(B|A_i) \Pr(A_i)$.

Formule de Bayes (n événements)

Soit un espace probabilisé (Ω, \mathcal{A}, P) et $(A_i)_{i=1,2,\dots,n}$ une partition **finie** de Ω telle que $\forall i, \Pr(A_i) \neq 0$. Alors, $\forall B \in \mathcal{A}$ tel que $\Pr(B) \neq 0$,

$$\Pr(A_i|B) = \frac{\Pr(B|A_i) \Pr(A_i)}{\sum_{j=1}^n \Pr(B|A_j) \Pr(A_j)}$$

Indépendance

Indépendance (2 événements)

Soit un espace probabilisé (Ω, \mathcal{A}, P) , A et B deux événements de \mathcal{A} . Alors A et B sont indépendants pour la probabilité P **si et seulement si** $\Pr(A \cap B) = \Pr(A) \Pr(B)$.

Il est important de bien saisir que la notion d'indépendance n'est pas intrinsèque aux événements, mais *dépend* de la probabilité P choisie sur (Ω, \mathcal{A}) . Deux événements peuvent être indépendants pour une probabilité, mais être dépendants pour une autre.

Propriétés (2 événements)

- 1 A^C et B sont indépendants.
- 2 A et B^C sont indépendants.
- 3 A^C et B^C sont indépendants.

Indépendance (n événements)

Soient (A_1, \dots, A_n) un n -uplet d'événements. On dit qu'ils sont **indépendants**, ou **mutuellement indépendants**, **si et seulement si** $\forall k = 1, \dots, n$, si \forall sous-ensemble $(A_{i_1}, \dots, A_{i_k})$ de k événements choisis parmi les (A_1, \dots, A_n) , on a $\Pr(A_{i_1} \cap \dots \cap A_{i_k}) = \Pr(A_{i_1}) \times \dots \times \Pr(A_{i_k})$.

Indépendance (suite d'événements)

Soit un espace probabilisé (Ω, \mathcal{A}, P) , et une suite d'événements indépendants $(A_n)_{n \in \mathbb{N}^*}$ de \mathcal{A} . Alors on a

$$\Pr\left(\bigcap_{n \in \mathbb{N}^*} A_n\right) = \lim_{k \rightarrow \infty} \prod_{n=1}^k \Pr(A_n).$$

Lemme de Borel-Cantelli (2ème partie)

Soit (Ω, \mathcal{A}, P) un espace probabilisé, et une suite d'événements $(A_n)_{n \in \mathbb{N}^*}$ indépendants de \mathcal{A} telle que $\sum_{n=1}^{\infty} \Pr(A_n) = \infty$, alors $\Pr(\limsup_n A_n) = 1$

Fonction de répartition

Mesure image

La mesure image, ou « *pushforward measure* » en anglais, est obtenue « *by pushing* » une mesure d'un espace mesurable à un autre avec une fonction mesurable.

Loi de probabilité

La mesure image de P par X , notée P_X , s'appelle la **loi de probabilité** de X .

Fonction de répartition

Pour une mesure de probabilité P_X , on a que $\forall x \in \mathbb{R}$ la fonction F est définie comme $F(x) = P_X([-\infty, x])$. Cette fonction a les propriétés suivantes :

- 1 F est croissante au sens large.
- 2 F est continue à droite.
- 3 $\lim_{x \rightarrow \infty} F(x) = 1$ et $\lim_{x \rightarrow -\infty} F(x) = 0$.

Note Il y a une relation biunivoque entre les mesures de probabilités sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et les fonctions de répartition.

Classification des lois de probabilité sur la tribu borélienne

Pour une probabilité P sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, on classifie les lois de probabilités en 2 groupes :

1 Diffuse

On dit que P est diffuse si $\forall x \in \mathbb{R}$, $P(x) = 0$.

2 Discrète

On dit que P est discrète s'il existe un ensemble au plus dénombrable S tel que $P(S) = 1$.

Cependant, si P désigne une probabilité sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ ni diffuse ni discrète, alors $\exists \alpha \in]0, 1[$, P_1 une loi discrète et P_2 une loi diffuse tel que $P = \alpha P_1 + (1 - \alpha) P_2$.

Variable aléatoire discrète

Toute variable aléatoire X telle qu'il existe un sous-ensemble fini ou dénombrable S_X (ou tout simplement S) de \mathbb{R} vérifiant $P(\{X \in S\}) = 1$. On peut donc définir $S = \{x \in \mathbb{R} : P_X(\{x\}) = P(\{X = x\}) > 0\}$. Donc, on note $p_x = P(\{X = x\}) = P_X(\{x\})$.

Loi continue

Une mesure de probabilité absolument continue est une mesure de probabilité de la forme $P(B) = \int_B f(x) dx \forall B \in \mathcal{B}_{\mathbb{R}}$ où f est une densité de probabilité. C'est-à-dire, une fonction définie sur \mathbb{R} satisfaisant aux conditions :

- 1 $f(x) \geq 0 \forall x \in \mathbb{R}$,
- 2 $\int_{-\infty}^{\infty} f(x) dx = 1$

Toute variable aléatoire X telle qu'il existe un sous-ensemble fini ou dénombrable S_X (ou tout simplement S) de \mathbb{R} vérifiant $P(\{X \in S\}) = 1$. On peut donc définir $S = \{x \in \mathbb{R} : P_X(\{x\}) = P(\{X = x\}) > 0\}$. Donc, on note $p_x = P(\{X = x\}) = P_X(\{x\})$.

Moments et transformations de variables

Cas discret

Espérance

Sous réserve d'existence, l'**espérance mathématique** ou la *moyenne* de X est le nombre $E[X] = m_X = \sum_k x_k P(X = x_k)$.

✓ Invariance à la translation ou multiplication par un scalaire

Pour $a, b \in \mathbb{R}$, $E[aX + b] = aE[X] + b$.

✓ Espérance d'une puissance s

Sous réserve d'existence, $E[X^s] = \sum_{x \in S} x^s P_X(x)$.

Variance

Sous réserve d'existence, la **variance** de X est le nombre $\text{Var}(X)$ ou σ_X^2 où $\text{Var}(X) = E[(X - E[X])^2] = \sum_{x \in S} (x - m_X)^2 P(X = x)$. On peut également réécrire $E[(X - E[X])^2] = E[X^2] - (E[X])^2$.

✓ Translation ou multiplication par un scalaire

Pour $a, b \in \mathbb{R}$, $\text{Var}(aX + b) = a^2 \text{Var}(X)$. Donc, contrairement à l'espérance, la variance n'est **pas** invariante à la translation ou multiplication par un scalaire.

Moment centré et réduit de puissance s

Sous réserve d'existence, le **moment centré s** de X est $E[(X - E[X])^s]$.

≡ Variable aléatoire réelle centrée

Sous réserve d'existence de la moyenne, toute variable dont la moyenne est nulle.

≡ Variable aléatoire réelle réduite

Sous réserve d'existence de la moyenne, toute variable de variance 1.

Donc, la variable aléatoire réelle **centrée et réduite** est, sous réserve d'existence, toute variable de moyenne nulle et de variance 1.

Covariance

La **covariance** entre deux variables X et Y est $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$.

Vecteur

Espérance d'un vecteur

Soit un vecteur aléatoire de n v.a. discrètes $\mathbf{X} = (X_1, \dots, X_n)$. Sous réserve d'existence de $E[X_i]$ pour $i = 1, 2, \dots, n$, l'**espérance mathématique** de \mathbf{X} est le n -uplet $(E[X_1], \dots, E[X_n])$.

✓ Espérance de la somme

Si l'espérance mathématique du vecteur \mathbf{X} existe, alors l'espérance mathématique de la somme $X_1 + \dots + X_n$ est $E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$.

✓ Espérance du produit

Si l'espérance mathématique du vecteur X existe, et que les **composantes du vecteur sont indépendantes**, alors l'espérance mathématique du produit $X_1 \times \cdots \times X_n$ est $E[X_1 \times \cdots \times X_n] = E[X_1] \times \cdots \times E[X_n]$.

Matrice de variances-covariances d'un vecteur

Si elle existe, la matrice de variances-covariances d'un vecteur aléatoire (X_1, \dots, X_n) est définie par le terme général $\forall i, j$ t.q. $1 \leq i, j \leq n$:

$$\text{Cov}(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])].$$

✓ Variance de la somme

Si le vecteur aléatoire X est composé de v.a.r. discrètes **indépendantes**, dont le moment d'ordre 2 existe, alors la variance de la somme $\sigma_{X_1 + \cdots + X_n}^2 = \sigma_{X_1}^2 + \cdots + \sigma_{X_n}^2$.

Autres mesures

Coeficient de corrélation de Pearson

Soit le couple de v.a.r. (X, Y) , possédant des variances non nulles, le **coefficient de corrélation** de X et de Y est le nombre $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$.

Cas continu

Espérance

Si X est une v.a.r. à densité f , alors l'**espérance mathématique**, sous réserve d'existence, est le nombre $E[X] = \int_{\mathbb{R}} xf(x)dx$.

✓ Espérance d'une fonction g de X

Si la loi P_X de X est **absolument continue** et de densité f , alors, sous réserve d'existence, $E[g \circ X] = E[g(X)] = \int_{\mathbb{R}} g(x)f(x)dx$.

Variance

Par l'inégalité $|x| < x^2 + 1$, si $E[X^2] < \infty$ alors $E[X] = m$ est définie. Sous réserve d'existence, la **variance** de X est donc $\text{Var}(X) = \int_{\mathbb{R}} (x - m)^2 f(x)dx$.

Vecteur

on dénote $M = E[XX^T]$ et $\sigma_X = E[(X - E[X])(X - E[X])^T]$.

Lois conditionnelles continues

On dénote la distribution conditionnelle de X par Y comme $f_X^{\{Y=y\}}(x) = \frac{f(x,y)}{g(y)}$.

Fonction génératrice des moments

Pour la somme $S_n = X_1 + \cdots + X_n$, $M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t)$.

$$E[X^n] = \left. \frac{d^n}{dt^n} M_X(t) \right|_{t=0}.$$

Calcul de lois

Quelques inégalités classiques

Inégalité de Schwartz

Si $E[X^2]$ et $E[Y^2]$ existent, alors $E[XY]$ existe et $E[|XY|] \leq \sqrt{E[X^2]} \sqrt{E[Y^2]}$.

Inégalité de Tchebychev

Soit X une v.a.r. positive et g une application strictement croissante de \mathbb{R}^+ dans \mathbb{R}^+ telle que $E[g(X)]$ existe, alors $\Pr(X \geq \alpha) \leq \frac{E[g(X)]}{g(\alpha)} \quad \forall \alpha > 0$.

Inégalité de Markov

Soit X une v.a.r. positive et intégrable, alors $\Pr(X \geq \alpha) \leq \frac{E[X]}{\alpha} \quad \forall \alpha > 0$.

Inégalité de Bienaymé-Tchebychev

Soit X une v.a.r. de carré intégrable, alors $\Pr(|X - E[X]| \geq \alpha) \leq \frac{V(X)}{\alpha^2} \quad \forall \alpha > 0$.

Inégalité de Jensen (cas unidimensionnel)

Rappel : Fonction convexe

Une fonction g définie sur un intervalle ouvert I et à valeurs réelles est dite **convexe** ssi $\forall a, b \in I$ et $\lambda \in [0, 1]$, $g(\lambda a + (1 - \lambda)b) \leq \lambda g(a) + (1 - \lambda)g(b)$.

Soient I un intervalle ouvert et $g : I \rightarrow \mathbb{R}$, une fonction convexe. Soit X une v.a. à valeurs dans I . Alors, $g(E[X]) \leq E[g(X)]$.

Inégalité de Jensen (cas multidimensionnel)

Rappel : Fonction convexe multivariée

On pose que, $\forall \gamma \in (0, 1)$, pour deux points $x, y \in \mathbb{R}^k$: $\gamma x + (1 - \gamma)y = (\gamma x_1 + (1 - \gamma)y_1 + \dots + \gamma x_k + (1 - \gamma)y_k)^t$.

Un sous-ensemble $C \in \mathbb{R}^k$ est **convexe** si $\forall x, y \in C$ et $\gamma \in [0, 1]$: $\gamma x + (1 - \gamma)y \in C$. Une fonction réelle φ définie sur un ouvert convexe C est **convexe** si $\forall x, y \in C$ et $\gamma \in [0, 1]$: $\varphi(\gamma x + (1 - \gamma)y) \leq \gamma \varphi(x) + (1 - \gamma)\varphi(y)$.

Soit un vecteur aléatoire X à valeurs dans un ouvert convexe $C \in \mathbb{R}^k$ ayant une espérance $E[X]$. Soit φ une fonction convexe sur C , telle que $E[\varphi(X)]$ existe. Alors, $\varphi(E[X]) \leq E[\varphi(X)]$.

Également, sous les mêmes hypothèses, si $\mathcal{B} \subset \mathcal{A}$, alors $\varphi(E[X|\mathcal{B}]) \leq E[\varphi(X)|\mathcal{B}]$.

Inégalité de Hölder

Soient les nombres conjugués p et $q > 1$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Si $E[X^p]$ et $E[Y^q]$ existent, alors $E[XY]$ existe, et $E[|XY|] \leq E[|X|^p]^{1/p} E[|Y|^q]^{1/q}$.

Note Si $p = q = 2$, on retrouve l'inégalité de Schwartz.

Convergences stochastiques

Convergence L^r

La séquence $(X_n)_{n \in \mathbb{N}}$ converge dans L^r si $\lim_{n \rightarrow \infty} E[|X_n - X|^r] = 0$ où les moments absolus $E[|X_n|^r]$ et $E[|X|^r]$ existent. On dit que c'est la convergence dans le r^{e} moyenne. On appelle X la limite des X_n .

Convergence presque sûre

Convergence presque sûre

Soit $(X_n)_{n \geq 1}$ une suite de v.a. définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mu)$. On dit que la suite de v.a. $(X_n)_{n \geq 1}$ converge presque sûrement vers la v.a. X ssi $P(\{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}) = 1$.

On écrit habituellement $X_n \xrightarrow{p.s.} X$.

Conditions équivalentes

Soit $(X_n)_{n \geq 0}$ une suite de v.a. définies sur un espace probabilisé (Ω, \mathcal{A}, P) , X une v.a. définie sur (Ω, \mathcal{A}, P) . On suppose que (X_n) et X sont P -presque sûrement finies. Ces conditions sont équivalentes :

- 1 $(X_n) \rightarrow X$ P -presque sûrement (P -p.s.)
- 2 $P(X_n \rightarrow X) = 1$
- 3 $P\left(\bigcap_{\varepsilon > 0} \bigcup_{n \geq 0} \bigcap_{k \geq n} \{X_k - X\} < \varepsilon\right) = 1.$
- 4 $\forall \varepsilon > 0, P\left(\bigcup_{n \geq 0} \bigcap_{k \geq n} \{X_k - X\} < \varepsilon\right) = 1.$

$$5 \quad \forall \varepsilon > 0, \lim_{n \rightarrow \infty} P\left(\bigcap_{k \geq n} \{X_k - X\} < \varepsilon\right) = 1.$$

Théorème convergence

Soit $(Z_n)_{n \geq 0}$ une suite de v.a. définies sur un espace probabilisé (Ω, \mathcal{A}, P) .

$$Z_n \xrightarrow{p.s.} 0 \text{ si } \sum_{i=1}^n P(|Z_n| \geq \varepsilon) < \infty \quad \forall \varepsilon > 0.$$

Convergence en probabilité

Convergence en probabilité

Soit une suite $(X_n)_{n \geq 1}$ de v.a. définies sur un espace probabilisé (Ω, \mathcal{A}, P) . On dit que cette suite converge en probabilité vers X si $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0 \quad \text{ou} \quad \lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1.$$

On écrit habituellement $X_n \xrightarrow{P} X$.

Note Pour le cas multivarié, on pose que les vecteurs aléatoires $(X_n)_{n \in \mathbb{N}}$ et X sont définis sur le même espace probabilisé (Ω, \mathcal{A}, P) , et à valeurs dans $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$. Puis, la suite de vecteurs $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X si pour toute composante $i : i = 1, \dots, p$, on a

$$X_{i,n} \xrightarrow{P} X_i.$$

Théorème de Slutsky

Soit les vecteurs aléatoires $(X_n)_{n \in \mathbb{N}}$ et X définis sur le même espace probabilisé (Ω, \mathcal{A}, P) , et à valeurs dans $(\mathbb{R}^p, \mathcal{B}_{\mathbb{R}^p})$. On suppose que la suite de vecteurs $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers X . Soit g une application continue de \mathbb{R}^p dans \mathbb{R}^q . On a alors

$$g(X_n) \xrightarrow{P} g(X).$$

Convergence en moyenne quadratique

Convergence en moyenne quadratique

La convergence en moyenne quadratique, alias la convergence au sens de L^2 , implique que

$$X_n \xrightarrow{m.q.} a \quad \text{ssi} \quad E[X_n] \xrightarrow{n \rightarrow \infty} a \quad \text{et} \quad \text{Var}(X_n) \xrightarrow{n \rightarrow \infty} 0.$$

Rapports

- › La convergence presque sûre entraîne la convergence en probabilité.
- › La convergence L^2 entraîne la convergence L^1 .
- › La convergence L^1 entraîne la convergence en probabilité.
- › La convergence en probabilité de (X_n) vers X entraîne la convergence en loi de (X_n) vers X .

Convergence des lois de probabilité

Convergence en loi

La suite (X_n) **converge en loi** vers X si, pour tout borélien A t.q.

$$P_X(\partial A) = P\{X \in \partial A\} = 0, \quad \lim_{n \rightarrow \infty} P\{X_n \in A\} = P\{X \in A\}.$$

$$X_n \xrightarrow{\mathcal{L}} X.$$

Théorème

Soit la suite de v.a. (X_n) qui converge en loi vers X et la suite (Y_n) telle que

$$(X_n - Y_n) \text{ converge vers } 0 \text{ en probabilité. Alors, } Y_n \xrightarrow{\mathcal{L}} X.$$

Loi des grands nombres et théorème centrale limite

Loi des grands nombres

Loi des grands nombres

On dit qu'une suite de v.a. (X_n) suite une LGN s'il existe deux suites réelles $(\alpha_n)_{n \in \mathbb{N}}$ et $(\beta_n)_{n \in \mathbb{N}}$ telles que : $\lim_{n \rightarrow \infty} \frac{1}{\alpha_n \sum_{i=1}^n (X_i - \beta_i)}$ existe en un « certain sens ».

LFGN de Bernoulli

Si la suite d'événements $(A_n)_{n \in \mathbb{N}}$ est formée d'événements *indépendantes* et de *même probabilité* p . Si on désigne par S_n le nombre de A_i réalisés parmi les événements A_1, \dots, A_n , alors : $\forall \varepsilon > 0, P\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \xrightarrow{n \rightarrow \infty} 0$.

LFGN

Soit la suite de v.a. discrètes iid $(X_n)_{n \geq 1}$, de moyenne m et variance σ^2 , définies sur un espace probabilisé (Ω, \mathcal{A}, P) , alors : $\frac{1}{n} \sum_{k=1}^n X_k = \bar{X}_n \xrightarrow[n \rightarrow \infty]{P} m$.

LFGN de Khintchine

Soit la suite de v.a. iid $(X_n)_{n \geq 1}$, de moyenne m et variance σ^2 , définies sur un espace probabilisé $L^1(\Omega, \mathcal{A}, P)$, alors : $\frac{S_n}{n} \sum_{k=1}^n X_k = \bar{X}_n \xrightarrow[n \rightarrow \infty]{P} m$.

LFGN dans L^2

Soit la suite de v.a.r. $(X_n)_{n \geq 1}$ de carré intégrable deux à deux non corrélées.

Une condition nettement suffisante (CNS) pour que : $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{L^2} a$ est que $\frac{1}{n} \sum_{i=1}^n E[X_i] \xrightarrow[n \rightarrow \infty]{} a$ et $\frac{1}{n^2} \sum_{i=1}^n V(X_i) \xrightarrow[n \rightarrow \infty]{} 0$

Théorème de Kolmogorov

Soit la suite de v.a. iid $(X_n)_{n \geq 1}$, de moyenne m , définies sur un espace probabilisé $L^1(\Omega, \mathcal{A}, P)$, alors : $\frac{S_n}{n} = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{p.s.} m$.

Théorème centrale limite

Théorème centrale limite (TCL)

Soit une suite de v.a.r. (X_n) iid, de moyenne m et variance σ^2 , alors :

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{L} \mathcal{N}(0, 1).$$

Note Alias le théorème de *Lindeberg-Lévy*.

Théorème de Moivre-Laplace

Soient les v.a. indépendantes (X_1, \dots, X_n) suivant la loi de Bernoulli de paramètre p , alors $\forall n$ et tous nombre réels $a < b$:

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Note Ce théorème est un cas spécial du TCL.

Théorème centrale limite sur \mathbb{R}^k

Soit la suite de vecteurs aléatoires (X_n) iid, de moyenne m et covariance Γ ,

alors $S_n \xrightarrow[n \rightarrow \infty]{L} \mathcal{N}(0, \Gamma)$.

Théorème centrale limite de Lyapounov

Soit les v.a. indépendantes (X_1, X_2, \dots) et $\delta > 0$ t.q. $\forall k \geq 1, E[|X_k|^{2+\delta}] < \infty$

et $\lim_{n \rightarrow \infty} \frac{1}{\sigma_{S_n}^{2+\delta}} \sum_{k=1}^n E[|X_k - m_k|^{2+\delta}] = 0$. Alors, $\frac{S_n - E[S_n]}{\sqrt{\text{Var}(S_n)}} \xrightarrow[n \rightarrow \infty]{L} \mathcal{N}(0, 1)$.

Convergence faible et continuité

Convergence en probabilité et continuité

Proposition

$$(U_n, V_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} (c, d) \text{ ssi } U_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c \text{ et } V_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} d.$$

De plus, si $h : \mathbb{R} \rightarrow \mathbb{R}$ est continue au point c , alors $h(U_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} h(c)$. Sinon, si $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est continue au point (c, d) , alors $h(U_n, V_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} h(c, d)$.

Il s'ensuit que :

1. $U_n + V_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c + d.$
2. $U_n - V_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c - d.$
3. $U_n V_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} cd.$
4. $U_n / V_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c/d$ si $d \neq 0.$

Convergence en loi et continuité

1. Si $U_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U$ et si la fonction $h : \mathbb{R} \rightarrow \mathbb{R}$ est continue, alors $h(U_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} h(U)$.
2. Si $(U_n, V_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} (U, V)$ et si la fonction $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ est continue, alors $h(U_n, V_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} h(U, V)$.

Il s'ensuit que si $(U_n, V_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} (U, V)$, alors :

1. $U_n + V_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U + V.$
2. $U_n - V_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U - V.$
3. $U_n V_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} UV.$
4. $U_n / V_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} U/V$ si $P(V = 0) = 0.$

La méthode du delta

Classe C^d

Une fonction de classe C^d admet d dérivées. Par exemple, une fonction de classe C^1 admet une dérivée.

Pour X_1, \dots, X_n une suite iid de moyenne m et variable σ^2 et g une fonction de classe C^1 , $\sqrt{n} (g(\bar{X}_n) - g(m)) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, (g'(m))^2 \sigma^2)$.

Modèles statistiques

Modèle statistique

Modèle statistique

Un modèle statistique est tout triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ où :

\mathcal{X} Ensemble appelé espace des observations.

\mathcal{A} Tribu sur \mathcal{X} .

\mathcal{P} Famille de probabilités sur l'espace mesurable $(\mathcal{X}, \mathcal{A})$

Par exemple, pour la loi binomiale $(\mathcal{X}, \mathcal{A}, \mathcal{P}) = (\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n), \mathcal{P}_p^{\otimes n}; p \in (0, 1))$.

Modèle statistique paramétrique

Un modèle statistique paramétrique est un modèle statistique $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ tel que $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$, où $\Theta \subset \mathbb{R}^p$ et Θ est l'espace des paramètres. Donc, on définit la loi grâce à un paramètre.

Statistique

Soit un modèle statistique $(\mathcal{X}, \mathcal{A}, \mathcal{P})$. On appelle *statistique*, définie sur l'espace des observations $(\mathcal{X}, \mathcal{A})$ à valeurs dans $(\mathcal{X}', \mathcal{A}')$, toute application mesurable S de $(\mathcal{X}, \mathcal{A})$ dans $(\mathcal{X}', \mathcal{A}')$.

Donc, une statistique joue le rôle pour les statistiques qu'une variable aléatoire jouait pour les probabilités.

Indépendance

Deux statistiques S_1 et S_2 sont dites indépendantes si $\forall P \in \mathcal{P}$, S_1 et S_2 sont P -indépendantes où P est une loi quelconque.

Modèle d'échantillonnage

Soit une expérience aléatoire qui consiste à effectuer n expériences aléatoires iid. On suppose que le modèle statistique, dans le cas d'une seule observation, est $(\mathcal{Y}, \mathcal{B}, \tilde{\mathcal{P}})$. Pour l'expérience aléatoire, le modèle statistique $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ sera donc $(\mathcal{Y}, \mathcal{B}, \tilde{\mathcal{P}})^{\otimes n} = ((\mathcal{Y}^n, \mathcal{B}^{\otimes n}, \mathcal{P}^{\otimes n}); \mathcal{P} \in \tilde{\mathcal{P}})$.

Si les lois de $\tilde{\mathcal{P}}$ sont de densité f , alors les lois de \mathcal{P} sont de densité h où $h(y_1, \dots, y_n) = f(y_1) \times \dots \times f(y_n)$.

Modèle réguliers et homogènes

Modèle régulier

Un modèle paramétrique est régulier si :

1. les lois P_θ sont soit toutes absolument continues ou toutes discrètes, et
2. le support ne dépend pas de(s) paramètre(s) θ .

C'est-à-dire, $P_\theta(X \in D) = 1 \forall \theta \in \Theta$, pour un certain ensemble dénombrable $D \subset \mathbb{R}^n$ qui ne dépend pas de θ .

Modèle régulier homogène

Un modèle régulier $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ est dit homogène si toutes les densités sont **strictement** positives sur le même support.

Fonction de vraisemblance

La vraisemblance est la densité de l'observation et la fonction de vraisemblance le produit des vraisemblances : $\theta \rightarrow \frac{dP_\theta(x)}{d\mu} = \ell(x; \theta)$.

Habituellement, on pose ces hypothèses de base pour la théorie de la statistique :

- 1 Si $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$.
- 2 L'espace des paramètres Θ contient au moins un intervalle dans le cas unidimensionnel, ou un hypercube dans le cas multidimensionnel.
- 3 La fonction $\theta \rightarrow P_\theta$ est continue. Si donc $(\theta_n)_{n \in \mathbb{N}^*}$ est une suite de Θ , et si $\lim_{n \rightarrow \infty} \theta_n = \theta^*$, alors $P_{\theta_n} \xrightarrow{L} P_{\theta^*}$.
- 4 Le modèle est régulier.

Classes de modèles

Le modèle linéaire

Le modèle exponentiel

La famille exponentielle

La famille $\{P_\theta; \theta \in \Theta\}$ de lois sur un espace d'observations $(\mathcal{X}, \mathcal{A})$ est appelée **famille exponentielle** si ses lois P_θ ont des densités sur $(\mathcal{X}, \mathcal{A})$, de la forme :

$$f_\theta(x) = e^{[\sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta)] h(x)}$$

où les η_i et B sont des fonctions réelles définies sur Θ , où les T_i sont des fonctions réelles de l'observation x , et où h est une fonction positive de l'observation x . $T = (T_1, \dots, T_s)^t$ s'appelle la **statistique exhaustive** (privilegiée) du modèle.

Note Souvent, h est une fonction esthétique. Par exemple, la fonction indicatrice du domaine de la v.a.

L'ensemble des points $\eta = (\eta_1, \dots, \eta_s)^t$ pour lesquels $f_\eta(x)$ est une densité de probabilité est appelée **l'espace naturel des paramètres** H . Cet espace est **convexe** ce qui permet de déduire que pour $\eta_1, \eta_2 \in H$, alors $\forall \alpha \in (0, 1)$ que $\eta = \alpha \eta_1 + (1 - \alpha) \eta_2 \in H$.

Quelques propriétés :

1. La somme de v.a. indépendantes de la famille exponentielle fait également partie de la famille exponentielle.
2. $A(\eta)$ est infiniment dérivable.
3. $\text{Cov}_\eta(T_j, T_i) = \frac{d^2}{d\eta_j d\eta_i} A(\eta)$.

Fonction de perte et de risque

Notation

d Décision.

D Espace de toutes les décisions $D = \{d_1, \dots, d_k\}$.

δ Règle de décision.

Δ Espace des règles de décisions.

Règle de décision δ

Une application mesurable δ de $(\mathcal{X}, \mathcal{A})$ à valeurs dans (D, \mathcal{D}) .

Fonction de perte

La fonction de perte est une application mesurable

$$L : (\Omega \times D, \mathcal{T} \otimes \mathcal{D}) \longrightarrow (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}).$$

Fonction de risque

La fonction de risque est $R(\theta, \delta) = E_\theta[L(\theta, \delta(X))]$ pour tout $\theta \in \Theta$, toute règle de décision $\delta \in \Delta$, et où X est une v.a.r. de loi P_θ .

Classement des règles de décisions

On dit que $\delta_1 \leq \delta_2$ (δ_1 est préférable à δ_2) si $\forall \theta \in \Theta, R(\theta, \delta_1) \leq R(\theta, \delta_2)$.

Exhaustivité, complétion et liberté

Exhaustivité

Exhaustivité

Soit un modèle paramétrique $(\mathcal{X}, \mathcal{A}; P_\theta, \theta \in \Theta)$ et l'espace \mathcal{Y} des valeurs prises par la statistique S . On suppose que l'observation est la valeur prise par une v.a.r. X définie sur un espace Ω :

$$(\Omega, \mathcal{C}, P) \xrightarrow{X} (\mathcal{X}, \mathcal{A}, P_\theta)_{\theta \in \Theta} \xrightarrow{S} (\mathcal{Y}, \mathcal{B})$$

Donc, la statistique S est exhaustive pour θ si la loi conditionnelle de X sachant que $S = s$ est indépendante de θ pour tout s .

Critère de factorisation

Soit un modèle paramétrique $(\mathcal{X}, \mathcal{A}; P_\theta, \theta \in \Theta)$ régulier et une statistique $S : (\mathcal{X}, \mathcal{A}) \xrightarrow{S} (\mathcal{Y}, \mathcal{B})$. Une CNS pour que la statistique S soit exhaustive est que les densités s'écrivent $\ell_\theta(x) = g_\theta(S(x))h(x)$ avec $h : (\mathcal{X}, \mathcal{A}) \rightarrow (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$ et $g_\theta : (\mathcal{Y}, \mathcal{B}) \rightarrow (\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+})$ mesurables.

Statistique exhaustive minimale

Une statistique S est exhaustive minimale, si elle est d'une part exhaustive, et si, pour toute autre statistique exhaustive T , il existe h telle que $S = h(T)$.

Exhaustif minimale

Soit un modèle statistique paramétrique régulier $(\Omega, \mathcal{A}, P_\theta)_{\theta \in \Theta}$. On suppose que $S = S(X)$ et $T = T(X)$ sont équivalentes, puis si S est exhaustive, alors T l'est également.

Soit un modèle paramétrique régulier et deux statistiques S_1 et S_2 telles que $S_1 = h^*(S_2)$. Si S_1 est exhaustive pour θ , S_2 l'est également. On dit que S_1 est préférable à S_2 , $S_1 \mathcal{R} S_2 \Leftrightarrow \exists h : S_1 = h(S_2)$. Donc, pour h qui n'est pas bijective on préfère employer S_1 qui permet d'écrire les observations sous format plus résumé.

Complétion

Statistique libre

Soit \mathcal{P} une famille de lois sur un espace d'observations $(\mathcal{X}, \mathcal{A})$, et S une statistique définie sur $(\mathcal{X}, \mathcal{A})$ à valeurs dans $(\mathcal{Y}, \mathcal{B})$. Pour toute loi $P \in \mathcal{P}$, on note P^S la loi image de P par S . On dit que S est libre si $\forall P, Q \in \mathcal{P}$

$$P^S = Q^S.$$

Dans un modèle paramétrique, si $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, on dit que S est libre si sa loi ne dépend pas du paramètre θ . Une statistique libre ne contient aucune information sur θ , mais des statistiques exhaustives minimales peuvent encore contenir une partie n libre z. Ce qui motive la définition de la statistique complète.

Statistique complète

Soit $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, une famille paramétrique de lois sur un espace d'observations $(\mathcal{X}, \mathcal{A})$, et soit S une statistique définie sur $(\mathcal{X}, \mathcal{A})$ à valeur dans $(\mathcal{Y}, \mathcal{B})$. On dira que S est complète ssi $\forall \theta \in \Theta$, $E_\theta[f(S)] = 0 \Rightarrow f = 0$ $P_\theta^S - p.s.$. C'est-à-dire implique que $P_\theta[f(S) = 0] = 1, \forall \theta \in \Theta$.

Estimation sans biais

Contexte

Bien que le théorème de Rao-Blackwell nous permet de trouver un estimateur de risque inférieur, il n'y existe pas un estimateur δ_0 absolu qui a uniformément le risque minimal. Il faut donc trouver d'autres façons de choisir un estimateur parmi un ensemble d'estimateurs. Pour ce faire, on débute par examiner des propriétés additionnelles que l'on pourrait imposer aux estimateurs.

Principes de réduction de l'ensemble des estimateurs

Nous observons 3 principes :

1 Principe d'invariance

Il est logique de vouloir retenir des estimateurs simples à manier. Donc, pour estimer des paramètres de type moyenne on utilise des fonctions linéaires des observations et pour estimer des paramètres de type variance on utilise des fonctions quadratiques des observations.

Ces principes s'interprètent donc comme une propriété *d'invariance*. La condition est que $\delta[f(x)] = f[\delta(x)]$ où f appartient au groupe des applications linéaires inversibles \mathbb{R}^n dans \mathbb{R}^n .

2 Principe de sans biais

Un estimateur δ est dit sans biais si, en moyenne, la valeur proposée est égale à la valeur recherchée. C'est-à-dire que δ est un estimateur sans biais de $g(\theta)$ ssi $E_\theta[\delta(X)] = g(\theta)$ pour $\forall \theta \in \Theta$.

3 Principes asymptotiques

La suite d'estimateurs $\{\delta_n, n \in \mathbb{N}\}$ est asymptotiquement sans biais ssi $\lim_{n \rightarrow \infty} E_\theta \delta_n(X) = g(\theta), \forall \theta \in \Theta$.

Un autre principe est d'imposer que l'estimateur δ_n s'approche g en un certain sens de la vraie valeur inconnue $g(\theta)$. On a la convergence en probabilité, convergence en moyenne quadratique, forte convergence.

Méthodes de recherche de bons estimateurs