

DATA CAMP - M1 SEP

Table des Matières

Algèbre Linéaire : niveau élémentaire	4
Matrices	4
Matrices usuelles	5
Opérations de base	5
Multiplication matricielle	6
Propriétés	7
Inverse	7
Systèmes linéaires	8
 Algèbre linéaire : niveau basique	 9
Déterminant	9
Diagonalisation	9
Matrice symétrique	10
Produit scalaire	10
Norme	10
Orthogonalité	11
Projection orthogonale	11
Matrice orthogonale	11
 Fondements de probabilité : niveau élémentaire	 12
Quelques définitions	12
Événements	12
Partitions	13
Tribus et boréliens	13
Mesure	15
Probabilités	15
Probabilités conditionnelles	15
Indépendance	16
Variable aléatoire	16
Fonction de répartition	17
Probabilité ponctuelle / Densité	17

Fondements de probabilités : niveau basique	19
Moments	19
Couples aléatoires	20
Propriétés	22
Vecteurs aléatoires	23
Lois usuelles	23
 Fondements de probabilités	 26
Vecteurs aléatoires	26
Notions de convergence	27
Loi faible des grands nombres	28
Loi forte des grands nombres	28
Théorème Central Limite	28
 Statistique inférentielle : niveau basique	 29
Echantillon / Estimateur	29
Estimateur sans biais	29
Estimateur convergent	29
Estimateur optimal	30
Construction d'un estimateur	31
Intervalles de confiance	32
Test d'hypothèses	33
Construction d'intervalles de confiance	34

Algèbre Linéaire : niveau élémentaire

Matrices

Une matrice est un tableau de nombres disposés en m lignes et n colonnes. Soit A de taille (m, n) :

$$A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

Le terme a_{ij} est situé à la i -ème ligne de la j -ème colonne de la matrice A .

Exemples :

$$A = \begin{bmatrix} -2 & 1 \\ 8 & -3 \\ 1 & 3 \end{bmatrix}; \quad B = \begin{bmatrix} -2 \\ -3 \\ 5 \end{bmatrix}; \quad C = [3 \quad -1 \quad 6]; \quad D = \begin{bmatrix} 3 & -1 \\ 4 & -3 \end{bmatrix}$$

A est une matrice de taille $(3, 2)$, B est une matrice de taille $(3, 1)$, C est une matrice de taille $(1, 3)$, D est une matrice de taille $(2, 2)$.

Une matrice $(m, 1)$ est dite matrice colonne. Une matrice $(1, n)$ est dite une matrice ligne. Une matrice (n, n) est dite une matrice carrée d'ordre n .

On appelle transposée de A (et on note A^t ou A'), la matrice dont les lignes sont les colonnes de A , et dont les colonnes sont les lignes de A .

Exemple :

$$A = \begin{bmatrix} -2 & 1 \\ 8 & -3 \\ 1 & 3 \end{bmatrix} \Rightarrow A^t = \begin{bmatrix} -2 & 8 & 1 \\ 1 & -3 & 3 \end{bmatrix}$$

Propriétés :

$$(A + B)^t = A^t + B^t$$

$$(AB)^t = B^t A$$

Matrices usuelles

Une matrice carrée d'ordre n est dite diagonale lorsque, pour tout $i \neq j$, on a $a_{ij} = 0$:

$$A = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

La matrice unité I_n est la matrice diagonale telle que $\forall i = 1, \dots, n, a_{ii} = 1$:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Opérations de base

Soient A et B deux matrices de même taille (m, n) de termes généraux respectifs a_{ij} et b_{ij} , et λ un nombre réel.

$A = B$ si et seulement si $a_{ij} = b_{ij}$ pour tous i, j .

La somme de A et B est la matrice de terme général $a_{ij} + b_{ij}$:

$$A + B = \begin{bmatrix} 5 + (-4) & 2 + 1 \\ 6 + 1 & 1 + (-3) \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 7 & -2 \end{bmatrix}$$

Le produit de A par λ est la matrice de terme général λa_{ij} :

$$\lambda A = 3 \times \begin{bmatrix} 5 & 2 \\ 6 & 1 \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ 18 & 3 \end{bmatrix}$$

Exemples :

$$A = \begin{bmatrix} 5 & 2 \\ 6 & 1 \end{bmatrix}; \quad B = \begin{bmatrix} -4 & 1 \\ 1 & -3 \end{bmatrix}; \quad \lambda = 3$$
$$A + B = \begin{bmatrix} 1 & 3 \\ 7 & -2 \end{bmatrix} \qquad \lambda A = \begin{bmatrix} 15 & 6 \\ 18 & -2 \end{bmatrix}$$

Multiplication matricielle

Le produit de matrices A et B (noté AB) n'est défini que si le nombre de colonnes de A est égal au nombre de lignes de B . C'est-à-dire A doit être de taille (m, p) et B de taille (p, n) . Alors AB est de taille (m, n) . De plus, soient a_{ij} et b_{ij} les termes généraux respectifs de A et B , alors le terme général de $C = AB$ est c_{ij} défini par :

$$c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}$$

Exemple :

$$A = \begin{bmatrix} 2 & -2 & 1 \\ -3 & 6 & 8 \\ 5 & 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} -1 & 5 \\ 1 & 2 \\ 3 & 4 \end{bmatrix}$$

Alors, la matrice $C = AB$ est donnée par :

$$C = \begin{bmatrix} -1 & 10 \\ 33 & 29 \\ 0 & 33 \end{bmatrix}$$

En effet, l'élément qui se trouve au croisement de la i -ème ligne et la j -ème colonne est :

$$c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}$$

Par exemple,

$$c_{2,1} = (-3) \times (-1) + 6 \times 1 + 8 \times 3 = 33$$

Le produit matriciel n'est pas commutatif.

Propriétés

Le produit est distributif par rapport à l'addition. C'est-à-dire :

$$A(B + C) = AB + AC \quad \text{et} \quad (B + C)A = BA + CA$$

Le produit est associatif, c'est-à-dire :

$$ABC = A(BC) = (AB)C$$

Si A est une matrice carrée d'ordre n , alors

$$AI_n = I_n A = A$$

Le produit de deux matrices peut être nul sans que l'une des deux matrices ne soit la matrice nulle, par exemple :

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \times \begin{bmatrix} -2 & 10 \\ 1 & -5 \end{bmatrix}$$

Inverse

Une matrice carrée A d'ordre n est dite inversible lorsqu'il existe une matrice B telle que :

$$AB = BA = I_n$$

B est alors notée A^{-1} , l'inverse de A .

Systèmes linéaires

Grâce au produit matriciel, on peut représenter un système linéaire par une équation matricielle. Soit le système linéaire de n équations et p inconnues :

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p = b_2 \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{np}x_p = b_n \end{cases}$$

On peut le représenter par $AX = B$ où :

$$A = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{np} \end{bmatrix}; \quad X = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}; \quad B = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Algèbre linéaire : niveau basique

Déterminant

Soit $A = (a_{ij})$ une matrice carrée d'ordre 2. Le déterminant de A est le réel noté $\det(A)$ tel que :

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Soient $A = (a_{ij})$ et $B = (b_{ij})$ deux matrices carrées d'ordre n et $\lambda \in \mathbb{R}$. On a les propriétés suivantes :

- $\det(AB) = \det(A) \cdot \det(B) = \det(BA)$
- $\det(A^T) = \det(A)$
- $\det(\lambda A) = \lambda^n \cdot \det(A)$
- Si A est diagonale, alors $\det(A) = a_{11}a_{22} \dots a_{nn}$
- $\det(I_n) = 1$
- A est inversible si et seulement si $\det(A) \neq 0$
- Si $\det(A) \neq 0$, alors $\det(A^{-1}) = \frac{1}{\det(A)}$

Diagonalisation

Valeur propre / Vecteur propre

Une valeur propre de A est un scalaire λ tel qu'il existe un vecteur colonne non nul V vérifiant : $AV = \lambda V$. V est alors appelé vecteur propre de A associé à λ .

Diagonalisation

Une matrice carrée A d'ordre n est diagonalisable lorsqu'il existe une matrice diagonale D et une matrice inversible P telles que : $A = PDP^{-1}$. D est constituée des valeurs propres de A . P est obtenue par la concaténation des vecteurs propres de A . Si les valeurs propres de A sont distinctes, alors A est diagonalisable (réciproque fausse).

Matrice symétrique

Une matrice carrée A d'ordre n est symétrique lorsque $A^T = A$. Si A est symétrique, alors :

- A a des valeurs propres réelles.
- A est diagonalisable.
- Il existe une matrice P telle que $P^{-1} = P^T$ et $A = PDP^{-1}$.

Produit scalaire

Soit x, y, z trois vecteurs de \mathbb{R}^n et λ un scalaire. On définit par produit scalaire (qu'on note $\langle \cdot, \cdot \rangle$) toute application qui vérifie les propriétés suivantes :

- $\langle x + \lambda y, z \rangle = \langle x, z \rangle + \lambda \langle y, z \rangle$
- $\langle x, y + \lambda z \rangle = \langle x, y \rangle + \lambda \langle x, z \rangle$
- $\langle x, y \rangle = \langle y, x \rangle$
- $\langle x, x \rangle \geq 0$
- $\langle x, x \rangle = 0 \Rightarrow x = 0$

Le produit scalaire canonique et usuel est défini comme :

$$\langle x, y \rangle = \sum_{i=1}^n x_i \cdot y_i$$

Norme

On appelle norme associée à un produit scalaire le réel $\|x\| = \sqrt{\langle x, x \rangle}$. Elle vérifie les propriétés suivantes :

- $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$ (Inégalité de Cauchy-Schwartz)
- $\|x + y\| \leq \|x\| + \|y\|$ (Inégalité triangulaire)
- $\|x\| \geq 0$ avec égalité si $x = 0$
- $\|\lambda x\| = |\lambda| \cdot \|x\|$
- $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle$

Un vecteur est dit unitaire ou normé si $\|x\| = 1$

Orthogonalité

Deux vecteurs x et y sont dits orthogonaux si $\langle x, y \rangle = 0$. On note $x \perp y$.

Une famille de vecteurs $\{x_i\}$ est dite orthogonale si tous ses vecteurs sont deux à deux orthogonaux. Toute famille orthogonale $\{x_i\}_{i=1}^p$ vérifie le théorème de Pythagore :

$$\left\| \sum_{i=1}^n x_i \right\|^2 = \sum_{i=1}^n \|x_i\|^2$$

Soit E un espace muni d'un produit scalaire et X une partie de E . On appelle orthogonal de X et on note X^\perp l'ensemble : $X^\perp = \{y \in E \mid \forall x \in X, \langle x, y \rangle = 0\}$.

On dit que $\{e_i\}_{i=1}^p$ est une base orthonormée de E si et seulement si : - Si $a_1 e_1 + a_2 e_2 + \dots + a_n e_n = 0$, alors $a_i = 0$ pour tout $i \in \{1, \dots, n\}$.

- Pour tout $x \in E$, il existe $a_1, a_2, \dots, a_n \in \mathbb{R}$ tels que $x = a_1 e_1 + a_2 e_2 + \dots + a_n e_n$.

- e_i est orthogonal à e_j pour tout $i \neq j$.

- Pour tout $i \in \{1, \dots, n\}$, $\|e_i\| = 1$.

Projection orthogonale

Soit x un vecteur d'un espace muni d'un produit scalaire E . Soit F un sous-espace vectoriel de E , x s'écrit de façon unique sous la forme : $x = f + f^\perp$ où $f \in F$ et $f^\perp \in F^\perp$. On dit que f est le projeté orthogonal de x sur F et on note $f = P_F(x)$.

Pour x et y de E , on a $\langle P_F(x), y \rangle = \langle x, P_F(y) \rangle$.

Si (e_1, e_2, \dots, e_n) est une base orthonormée, alors : $P_F(x) = \sum_{i=1}^n \langle x, e_i \rangle e_i$

Notons que : $\|x - P_F(x)\| = \inf_{f \in F} \|x - f\|$

Matrice orthogonale

Soit M une matrice carrée d'ordre n . On dit que M est une matrice orthogonale si elle vérifie :

$$M \cdot M^T = M^T \cdot M = I_n$$

Le déterminant d'une matrice orthogonale est égal à ± 1 . L'ensemble des valeurs propres de M est inclus dans l'ensemble $\{0, 1\}$.

Fondements de probabilité : niveau élémentaire

Quelques définitions

- On appelle épreuve E toute expérience probabiliste.
- On appelle univers de E l'ensemble, généralement noté Ω , de tous les résultats possibles de l'épreuve E (appelés “événements élémentaires”).

Lancer une paire de dés équilibrés et en retenir la somme est une épreuve. $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$



Evénements

Un événement est un sous-ensemble de Ω .

- L'intersection de A et B , notée $A \cap B$, est un événement. Il est réalisé uniquement si A et B se produisent.
- La réunion de A et B , notée $A \cup B$, est un événement. Il est réalisé si A ou B se produit. Deux événements remarquables sont à retenir :
 - L'événement certain Ω ;
 - L'événement impossible \emptyset ;
- Tous les éléments qui n'appartiennent pas à A appartiennent à un événement que l'on appelle le complémentaire de A . On le note \overline{A} ou A^C .
- On dit que deux événements A et B sont incompatibles s'ils ne peuvent pas être réalisés en même temps.

Si A , B et C sont des événements de Ω , les propriétés suivantes sont toujours vérifiées :

- $A \cup \overline{A} = \Omega$ et $A \cap \overline{A} = \emptyset$
- $\overline{\overline{A}} = A$ et $\overline{A \cap B} = \overline{A} \cup \overline{B}$ (lois de De Morgan)
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

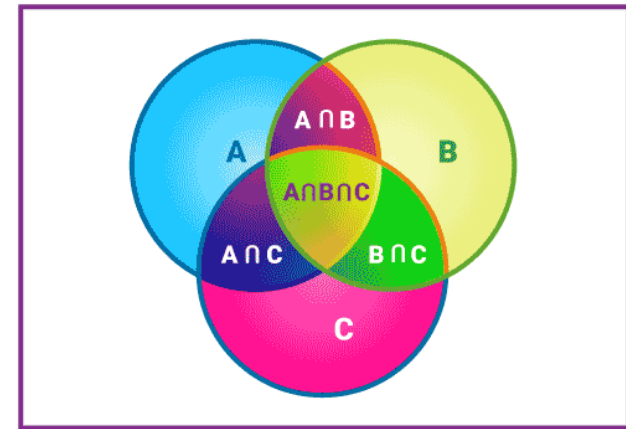
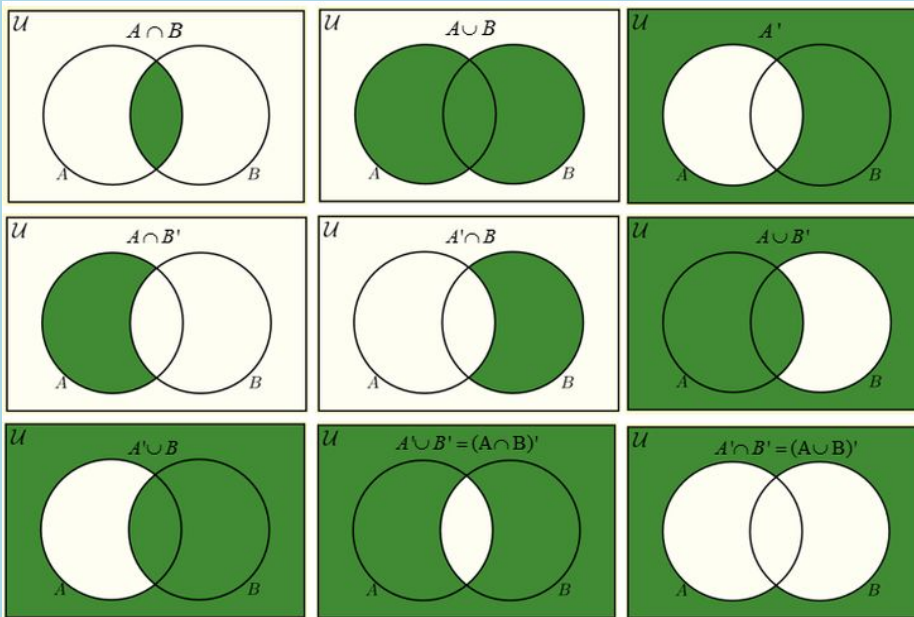
Partitions

La famille d'événements forme une partition de Ω si :

$$\bigcup_{i \in I} A_i = \Omega \quad \text{et} \quad A_i \cap A_j = \emptyset \quad \text{pour tout } i \neq j.$$

Une partition remarquable est la famille qui contient l'événement A et son complémentaire.

Tribus et boréliens



Comment pouvons-nous qualifier l'ensemble des événements ?

Une tribu est une famille T de parties de l'ensemble Ω qui vérifie les propriétés suivantes : - $\Omega \in T$

- Si (A_n) est une suite dénombrable d'éléments de T , alors $\bigcup A_n \in T$

— Si A est un élément de T , alors son complémentaire l'est aussi.

De plus, si T est une tribu, alors : - $\emptyset \in T$

- Si (A_n) est une suite d'éléments de T , alors $\bigcap A_n \in T$.

Exemples de tribus :

Pour le cas discret, on considère l'expérience "Lancer une pièce de monnaie équilibrée".

On notera : P pour "PILE apparaît" et F pour "FACE apparaît".

Dans ce cas, l'univers est l'ensemble $\{P, F\}$ et $T = \{\Omega, \emptyset, P, F\}$ est une tribu.

En général, l'ensemble des parties est une tribu classique.

Pour le cas continu, les intervalles du type $[a, +\infty[,]a, +\infty[,] - \infty, a[,] - \infty, a]$ sont des tribus. Nous les appelons DES BORÉLIENS.

Propriétés :

Soient A et B deux événements. Les propriétés suivantes sont toujours vraies :

1. $P(A^C) = 1 - P(A)$
2. $P(B) = P(A \cap B) + P(A^C \cap B)$
3. Si $A \subset B$, alors $P(A) \leq P(B)$
4. $0 \leq P(A) \leq 1$
5. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

De plus, en considérant une suite (A_k) d'événements, on a les relations suivantes :

$$P\left(\bigcup A_k\right) = \lim_{n \rightarrow +\infty} P\left(\bigcup_{k=1}^n A_k\right)$$

$$P\left(\bigcap A_k\right) = \lim_{n \rightarrow +\infty} P\left(\bigcap_{k=1}^n A_k\right)$$

$$P\left(\bigcup A_k\right) \leq \sum_{k=1}^{+\infty} P(A_k)$$

Et si $\bigcup A_k = \Omega$, alors :

$$P(B) = \sum_{k=1}^{+\infty} P(B \cap A_k)$$

Mesure

Soit E un ensemble muni d'une tribu T . On appelle mesure toute application $m : T \rightarrow \mathbb{R}^+$ telle que : - $m(\emptyset) = 0$ - Si (A_n) est une suite d'éléments de T deux à deux disjoints alors :

$$m\left(\bigcup_n A_n\right) = \sum_n m(A_n).$$

Probabilités

Soit E un ensemble muni d'une tribu T . On appelle probabilité toute application $P : T \rightarrow \mathbb{R}^+$ telle que : - $P(\emptyset) = 0$ - Si (A_n) est une suite d'éléments de T deux à deux disjoints alors :

$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n).$$

Probabilités conditionnelles

En théorie des probabilités, nous nous intéressons souvent au comportement d'un aléa, sachant qu'un autre événement est déjà passé. C'est ce que nous appelons Les Probabilités Conditionnelles.

Considérant deux événements de probabilité non nulle, A et B , la probabilité conditionnelle de A sachant que B est réalisé (couramment dit A sachant B) est donnée par :

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

Par commutativité de l'intersection, nous avons :

$$P(A \cap B) = P(B \cap A)$$

En utilisant la formule ci-dessus, nous pouvons également exprimer la probabilité conditionnelle de B sachant A :

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

C'est ce que nous appelons **LA FORMULE DE BAYES**.

Indépendance

Deux événements A et B sont dits indépendants si et seulement si :

$$P(A \cap B) = P(A) \cdot P(B)$$

En termes courants, deux événements sont indépendants si le résultat de l'un n'influence aucunement l'aboutissement de l'autre. Sous condition d'indépendance de A et B , la notion de la probabilité conditionnelle tombe à l'eau, car les événements évoluent l'un sans se soucier de l'autre. Ceci se traduit par :

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

Notons que si A est indépendant de B , il le sera par rapport à son complémentaire également, et vice versa. En général, pour une suite (A_n) d'événements indépendants, on a :

$$P\left(\bigcap A_i\right) = \prod P(A_i) = P(A_1) \cdot \dots \cdot P(A_n)$$

Cette formule est largement utilisée en statistique.

NE PAS CONFONDRE INDÉPENDANCE ET INCOMPATIBILITÉ DES ÉVÉNEMENTS

Variable aléatoire

Une variable aléatoire est un nombre qui dépend du résultat d'une expérience aléatoire. Chaque exécution de l'expérience génère une réalisation de la variable aléatoire.

Mathématiquement, on définit une variable aléatoire X comme une fonction $X : T \rightarrow \mathbb{R}$ qui associe à chaque événement s , un réel $X(s)$.

Par exemple, dans une queue pour la caisse d'un magasin, le nombre de clients est une variable aléatoire. La durée de traitement de chaque requête aussi. Remarquons que la première est un nombre entier. On dit qu'elle est à support discret. Alors que la deuxième est une durée (un nombre réel). On dit qu'elle est à support continu.

Qu'est- ce qui caractérise une variable aléatoire ?

Fonction de répartition

Une variable aléatoire traduit le résultat d'une expérience aléatoire en nombre réel. La fonction de répartition transporte le calcul des probabilités concernant les réalisations de la variable aléatoire. C'est la fonction définie par :

$$F_X(x) = P(X \leq x)$$

Propriétés :

Pour tout x , $0 \leq F_X(x) \leq 1$ F_X est une fonction croissante. $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow \infty} F_X(x) = 1$

Probabilité ponctuelle / Densité

CAS DISCRET : Probabilité ponctuelle

La probabilité ponctuelle est la fonction qui décrit les sauts de la fonction de répartition :

$$P(X = k) = P(X \leq k) - P(X \leq k - 1) = p_k$$

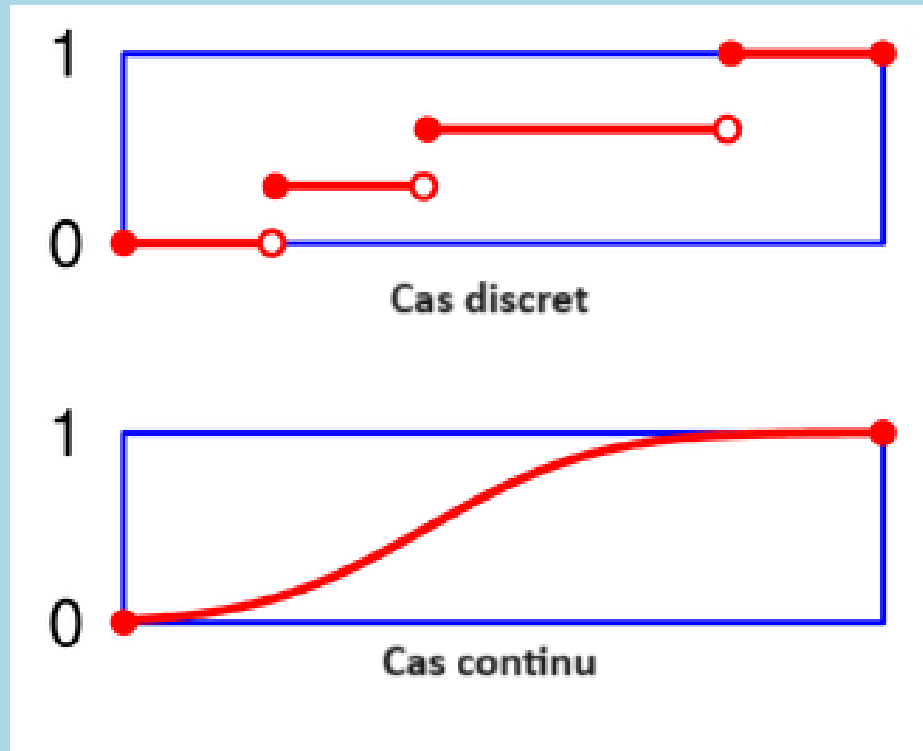
$$\sum p_i = 1$$

CAS CONTINU : densité de probabilité

La densité est la fonction qui décrit les variations de la fonction de répartition :

$$f(x) = \frac{dF_X}{dx}(x)$$

$$\int f(x)dx = 1$$



Fondements de probabilités : niveau basique

Moments

ESPÉRANCE

L'espérance d'une variable aléatoire est sa valeur attendue. C'est une mesure de localisation de la distribution.

Dans le cas discret :

$$E(X) = \sum_{k \in X(\Omega)} k \cdot P(X = k)$$

Alors que dans le cas continu :

$$E(X) = \int_{x \in X(\Omega)} x \cdot f_X(x) dx$$

THÉORÈME DE TRANSFERT

$$E(g(X)) = \sum_{\forall k \in X(\Omega)} g(k) \cdot P(X = k)$$

$$E(g(X)) = \int_{\forall x \in X(\Omega)} g(x) \cdot f_X(x) dx$$

VARIANCE

La variance d'une variable aléatoire décrit la dispersion de la variable aléatoire autour de sa valeur moyenne (son espérance).

Elle est définie par :

$$Var(X) = E(X^2) - (E(X))^2 = E((x - E(X))^2)$$

Sa racine carrée est appelée écart-type et notée généralement :

$$\sigma(X) = \sqrt{Var(X)}$$

CENTRAGE ET RÉDUCTION

Le centrage consiste à localiser la distribution autour de l'origine et la réduction consiste à normaliser la dispersion. La technique est simple :

$$Y = \frac{X - E(X)}{\sigma(X)}$$

MOMENTS D'ORDRE r

Le moment d'ordre r est défini par :

$$\mu_r = E(X^r)$$

Le moment centré d'ordre r est défini ainsi :

$$\mu_r = E((X - E(X))^r)$$

Couples aléatoires

La fonction conjointe

$$F_{X,Y}(x, y) = P(X \leq x \cap Y \leq y)$$

est appelée la distribution conjointe de X et Y .

Dans le cas continu, la fonction définie par :

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

est la densité conjointe du couple (X, Y) . On a donc :

$$F_{X,Y}(x, y) = \int \int f_{X,Y}(t, u) dt du, \text{ où } -\infty < x, y < +\infty,$$

Dans le cas discret, on définit la fonction de probabilité conjointe :

$$P(X = x_i, Y = y_j) = p_{ij}$$

On a donc :

$$F_{X,Y}(x, y) = \sum \sum p_{ij}, \text{ où } x_i \leq x \text{ et } y_j \leq y$$

LOI MARGINALE

La loi marginale de X est définie comme suit :

$$f_X(x) = \int f_{X,Y}(x,y) dy, \text{ où } -\infty < x < \infty,$$

dans le cas continu, ou encore :

$$f_X(x_i) = \sum p_{ij}, \text{ où } j \text{ tel que } y_j \leq y$$

Si X et Y sont indépendants, alors :

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

COVARIANCE

La covariance mesure l'intensité de la relation linéaire entre deux variables aléatoires X et Y. Elle est définie comme suit :

$$Cov(X,Y) = E(XY) - E(X) \cdot E(Y)$$

Si X et Y sont indépendants, alors :

$$Cov(X,Y) = 0$$

Il est important de noter que la réciproque n'est pas vraie : la covariance n'implique pas nécessairement l'indépendance entre X et Y.

Propriétés

ESPÉRANCE

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$$

$$\mathbb{E}(a) = a$$

VARIANCE

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

$$\text{Var}(a) = 0$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$$

COVARIANCE

$$\text{Cov}(X, Y) = \text{Cov}(Y, X)$$

$$\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$$

$$\text{Cov}(aX + bY, U) = a\text{Cov}(X, U) + b\text{Cov}(Y, U)$$

$$\text{Cov}(X, cU + dV) = c\text{Cov}(X, U) + d\text{Cov}(X, V)$$

$$\text{Cov}(aX + bY, cU + dV) = ac\text{Cov}(X, U) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, U) + bd\text{Cov}(Y, V)$$

Vecteurs aléatoires

Pour un vecteur aléatoire

$$(X_1, X_2, \dots, X_n)$$

, l'espérance est toujours linéaire. Pour une suite

$$(a_i)_{i \in \{1, \dots, n\}}$$

de réels, on a :

$$\mathbb{E}(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2) + \dots + a_n \mathbb{E}(X_n)$$

Si les variables aléatoires

$$X_1, X_2, \dots, X_n$$

sont indépendantes, alors la variance de leur somme est égale à la somme de leurs variances individuelles :

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$$

Lois usuelles

Ces tableaux récapitulent les lois usuelles que vous pourrez rencontrer dans différents cours du master.

Nom	Notation	$X(\Omega)$	$P(X = k)$	$E[X]$	$V(X)$
Uniforme	$X \sim U(\{1, 2, \dots, n\})$	$\{1, 2, \dots, n\}$	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Bernouilli	$X \sim B(p), 0 < p < 1$	$\{0, 1\}$	$P(X = 1) = p$ $P(X = 0) = 1 - p$	p	$p(1 - p)$
Binomiale	$X \sim B(n, p), 0 < p < 1$	$\{1, 2, \dots, n\}$	$C_k^n p^k (1 - p)^{n-k}$	np	$np(1 - p)$
Géométrique	$X \sim G(p), 0 < p < 1$	\mathbb{N}	$p(1 - p)^k$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson	$X \sim P(\lambda), \lambda > 0$	\mathbb{N}	$\frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ

Nom	Notation	$X(\Omega)$	$f_X(x)$	$E[X]$	$V(X)$
Uniforme	$X \sim U([a, b]), a < b$	$[a, b]$	$\frac{1}{b-a} 1_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
Exponentielle	$X \sim \mathcal{E}(\lambda), \lambda > 0$ $\mathcal{E}(\lambda) = \gamma(1, \lambda)$	\mathbb{R}^+	$\lambda e^{-\lambda x} 1_{\mathbb{R}^+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normale ou Gaussienne	$X \sim N(m, \sigma^2), \sigma > 0$	\mathbb{R}	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$	m	σ^2
Gamma	$X \sim \gamma(\alpha, \theta), \alpha > 0, \theta > 0$ $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$	\mathbb{R}^+	$\frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} 1_{\mathbb{R}^+}(x)$	$\frac{\alpha}{\theta}$	$\frac{\alpha}{\theta^2}$
Khi-2	$X \sim \chi^2(n), n \in \mathbb{N}^+$ Y_1, Y_2, \dots, Y_n indépendantes, $Y_i \sim \mathcal{N}(0, 1), X = \sum_{i=1}^n Y_i^2$	\mathbb{R}^+	$\gamma\left(\frac{n}{2}, \frac{1}{2}\right)$	n	$2n$
Bêta	$X \sim B(\alpha, \theta), \alpha > 0, \theta > 0$ $B(\alpha, \theta) = \int_0^1 x^{\alpha-1} (1-x)^{\theta-1} dx$ $B(\alpha, \theta) = \int_0^1 x^{\alpha-1} (1-x)^{\theta-1} dx$ $X = \frac{Z}{1+Z}, Z \sim B'(\alpha, \theta)$	$[0, 1]$	$\frac{x^{\alpha-1} (1-x)^{\theta-1}}{B(\alpha, \theta)} 1_{[0,1]}(x)$	$\frac{\alpha}{\alpha+\theta}$	$\frac{\alpha\theta}{(\alpha+\theta)^2(\alpha+\theta+1)}$
Bêta (prime)	$Z \sim B'(\alpha, \theta), \alpha > 0, \theta > 0$ $X \sim \gamma(\alpha, 1), Y \sim \gamma(\theta, 1), X \perp\!\!\!\perp Y$ $Z = \frac{X}{Y}$ $B(\alpha, \theta) = \int_0^\infty \frac{x^{\alpha-1}}{(1+x)^{\alpha+\theta}} dx$	\mathbb{R}^+	$\frac{z^{\alpha-1}}{B(\alpha, \theta) \cdot (1+z)^{\alpha+\theta}} \cdot 1_{\mathbb{R}^+}(z)$	$\frac{\alpha}{\theta-1}$ $\theta > 1$	$\frac{\alpha(\alpha+\theta-1)}{(\theta-1)^2(\theta-2)}$ $\theta > 2$
Student T	$T \sim T(n), n \in \mathbb{N}^*$ $X \sim \mathcal{N}(0, 1), Y \sim \chi^2(n), X \perp\!\!\!\perp Y$ $T = \frac{X}{\sqrt{Y/n}}$ $T^2/n \sim B'(1/2, n/2)$	\mathbb{R}	$\frac{\left(1+\frac{t^2}{n}\right)^{-(n+\frac{1}{2})}}{\sqrt{n} \cdot B(\frac{1}{2}, \frac{n}{2})}$	0	$\frac{n}{n-2}$ $n > 2$
Fisher	$X \sim F(n, m)$ $N \sim \chi^2(n), n \in \mathbb{N}^*$ $M \sim \chi^2(m), m \in \mathbb{N}^*$ $N \perp\!\!\!\perp M, X = \frac{N}{n} / \frac{M}{m}$ $\frac{n}{m} X \sim B'\left(\frac{n}{2}, \frac{m}{2}\right)$	\mathbb{R}^+	$\frac{K x^{n/2-1}}{(m+nx)^{(n+m)/2}} 1_{\mathbb{R}^+}(x)$ $K = \frac{n^{\frac{n}{2}} \cdot m^{\frac{m}{2}}}{B(\frac{n}{2}, \frac{m}{2})}$	$\frac{m}{m-2}$ $m > 2$	$\frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$ $m > 4$

$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$: désigne la fonction Gamma d'Euler

$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$: désigne la fonction Bêta

Nous allons souvent rencontrer les lois grisées dans les Tests statistiques. Là encore, connaître les densités ne servirait pas à grand chose, mais ceci nous évitera de parler de lois dont nous ne connaissons pas la tête.

Fondements de probabilités

Vecteurs aléatoires

INDEPENDANCE DEUX À DEUX

Les variables X_1, \dots, X_n sont deux à deux indépendantes si et seulement si : $\forall i \neq j, X_i$ et X_j sont indépendantes.

INDEPENDANCE MUTUELLE

Les variables X_1, \dots, X_n sont mutuellement indépendantes si et seulement si :

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \times \dots \times P(X_n = x_n)$$

L'INDÉPENDANCE MUTUELLE IMPLIQUE L'INDÉPENDANCE DEUX À DEUX. LA RÉCIPROQUE EST FAUSSE.

Si X_1, \dots, X_n sont mutuellement indépendantes, alors pour toute famille de fonctions réelles f_i , on a : $(f_1(X_1), \dots, f_n(X_n))$ sont indépendantes.

ESPÉRANCE ET VARIANCE-COVARIANCE

Soit $X = (X_1, \dots, X_n)^T$ un vecteur aléatoire. Dans le cas multidimensionnel, l'espérance scalaire est remplacée par un vecteur espérance.

$$E(X) = (E(X_1), \dots, E(X_n))^T$$

La variance unidimensionnelle est remplacée par la matrice symétrique de variance-covariance. Elle contient les variances en diagonale et les covariances ailleurs. On la note généralement Σ_X .

$$\Sigma_X = \begin{bmatrix} V(X_1) & \dots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \dots & V(X_n) \end{bmatrix}$$

Notions de convergence

Si l'on pense à des données, vues comme réalisation de variables aléatoires X_1, \dots, X_n , il serait intéressant de se poser la question de savoir comment évolue cette suite lorsque n tend vers l'infini.

Convergence presque sûre

On dit que (X_n) converge presque sûrement vers X et on note $X_n \xrightarrow{\text{p.s.}} X$ si et seulement si : $P(\lim_{n \rightarrow +\infty} X_n = X) = 1$

Convergence en probabilité

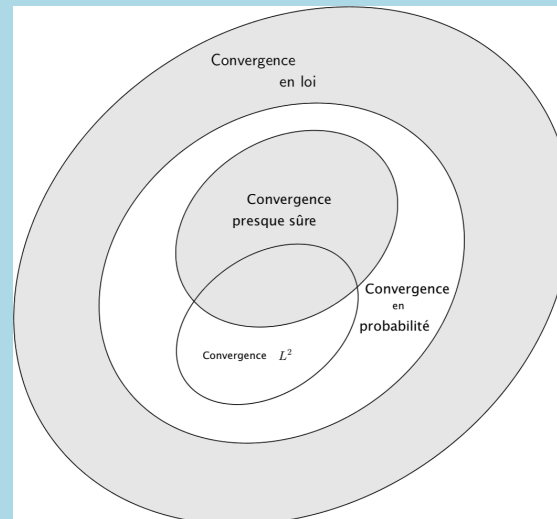
On dit que (X_n) converge en probabilité vers X et on note $X_n \xrightarrow{P} X$ si et seulement si : $\forall \varepsilon > 0, P(|X_n - X| > \varepsilon) \rightarrow 0$

Convergence en Loi

On dit que (X_n) converge en loi vers X et on note $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si : $F_{X_n} \xrightarrow{n \rightarrow +\infty} F_X$ Où F_X dénote la fonction de répartition de X .

Convergence en Moyenne quadratique

On dit que (X_n) converge en moyenne quadratique vers X et on note $X_n \xrightarrow{m.q.} X$ si et seulement si : $\mathbb{E}((X_n - X)^2) \rightarrow 0$ Cette définition peut se généraliser jusqu'à l'ordre n , mais nous n'en aurons pas besoin.



Loi faible des grands nombres

Soit X_1, \dots, X_n une suite de variables aléatoires indépendantes et de même loi telles que : $\mathbb{E}(X_i) = \mu$ et $\text{Var}(X_i) = \sigma^2$ alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p} \mu$$

Loi forte des grands nombres

Soit X_1, \dots, X_n une suite de variables aléatoires indépendantes et de même loi telles que : $\mathbb{E}(X_i) = \mu$ et $\text{Var}(X_i) = \sigma^2$, alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mu$$

Théorème Central Limite

Soit X_1, \dots, X_n une suite de variables aléatoires indépendantes et de même loi telles que : $\mathbb{E}(X_i) = \mu$ et $\text{Var}(X_i) = \sigma^2$, alors :

$$\sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}oi} \mathcal{N}(0, 1)$$

Statistique inférentielle : niveau basique

Echantillon / Estimateur

Le point de départ est un vecteur (ou un tableau dans le cas multidimensionnel) de données. Ces données peuvent être vues comme les réalisations (x_1, x_2, \dots, x_n) d'une variable aléatoire X qui dépend d'un certain paramètre θ que nous allons chercher à estimer. Pour ce faire, nous allons construire un échantillon de cette variable. Un échantillon (X_1, X_2, \dots, X_n) est un n-uplet de variables aléatoires indépendantes qui suivent toutes la même loi (celle de X). Un estimateur de θ est une fonction $\hat{\theta} = f(X_1, X_2, \dots, X_n)$ de notre échantillon, qui possède une loi de probabilité. Lorsque l'aléa est réalisé, $\hat{\theta}(\omega) = f(x_1, x_2, \dots, x_n)$ est une estimation de θ . Le but de ce cours est de construire le meilleur estimateur possible de θ .

Estimateur sans biais

Pour que l'estimation soit bonne, il faut que $\hat{\theta}$ soit proche de θ . Comme $\hat{\theta} = f(X_1, X_2, \dots, X_n)$ est une variable aléatoire, on ne peut imposer de condition qu'à sa valeur moyenne.

On définit ainsi le biais :

$$b_n(\hat{\theta}, \theta) = \mathbb{E}(\hat{\theta}_n) - \theta$$

Un estimateur est dit sans biais si $b_n(\hat{\theta}, \theta) = 0$, c'est-à-dire :

$$\mathbb{E}(\hat{\theta}_n) = \theta$$

Estimateur convergent

Un estimateur est dit convergent s'il converge en probabilité vers le paramètre à estimer :

$$\hat{\theta}_n \xrightarrow{P} \theta$$

En pratique, tout estimateur sans biais et dont la variance tend vers 0 est convergent.

Estimateur optimal

Qualité d'un estimateur

La qualité d'un estimateur est mesurée à travers son erreur quadratique moyenne définie par :

$$EQM(\hat{\theta}_n) = (b_n(\hat{\theta}, \theta))^2 + V(\hat{\theta}_n)$$

Comme nous cherchons tout le temps (presque) des estimateurs sans biais, il reste à comparer les variances.

Un estimateur $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ si :

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

Inégalité de Rao-Cramer/ Efficacité

On définit la quantité d'information apportée par l'estimateur par :

$$I(\hat{\theta}_n) = - \left(\mathbb{E} \left(\frac{\partial L}{\partial \theta} \right) \right)^2$$

Où $(,) = (,)$ (nous reviendrons sur sa définition)

L'inégalité de Rao-Cramer postule que la variance d'un estimateur ne peut pas aller en delà d'un certain seuil :

$$V(\hat{\theta}_n) \geq \frac{1}{I(\hat{\theta}_n)}$$

Un estimateur est optimal (ou efficace) si sa variance vérifie le cas d'égalité.

Construction d'un estimateur

Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance consiste à affecter θ la valeur qui maximise la probabilité d'observer (x_1, x_2, \dots, x_n) lorsque l'aléa du vecteur (X_1, X_2, \dots, X_n) tombe. Sans trop rentrer dans la théorie de la vraisemblance, nous allons présenter un algorithme en cinq étapes pour calculer cet estimateur (qui présente des propriétés assez séduisantes) :

Etape 1 : Calculer la fonction de vraisemblance

Dans le cas continu :

$$L(\mathbf{x}, \theta) = \prod_{i=1}^n f(x_i)$$

Dans le cas discret :

$$L(\mathbf{x}, \theta) = \prod_{i=1}^n P(X_i = x_i)$$

Etape 2 : Calculer le log-vraisemblance

Il s'agit de calculer un maximum, ce qui revient à dériver. Il s'agit ici d'un produit de n facteurs, ce qui rend la dérivation assez coriace. La fonction logarithmique présente des propriétés assez sympas pour faciliter cette tâche.

Etape 3 : Calculer la dérivée de la log-vraisemblance

Etape 4 : Résoudre l'équation d'inconnue θ

$$\frac{\partial(\ln(L))}{\partial\theta} = 0 \Rightarrow \theta = \theta_0$$

Etape 5 : Vérifier qu'il s'agit d'un maximum.

En s'assurant que :

$$\frac{\partial^2(\ln(L))}{\partial\theta^2} < 0$$

Méthode des moments

Comme le paramètre à estimer intervient dans la densité de probabilité, les moments théoriques sont souvent en fonction de ce paramètre. Ainsi, la méthode des moments consiste à égaliser les moments théoriques (espérance, variance) à leurs équivalents empiriques et à en dégager une estimation ponctuelle.

En pratique, il faut résoudre l'(les) équation(s) :

$$\mathbb{E}(X) = \overline{X} \text{ et } \text{Var}(X) = S_n^2$$

avec :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$$

Méthode des moindres carrés ordinaires

Lorsqu'il s'agit de prendre une mesure avec un appareil doté d'une imprécision ε , alors le problème d'estimation peut s'écrire : $X = \theta + \varepsilon$. La méthode des moindres-carrés ordinaires consiste à trouver le paramètre θ qui minimise la somme des carrées des erreurs :

$$\theta_{MCO} = \arg \min \left(\sum_{i=0}^n \varepsilon_i^2 \right) = \arg \min \left(\sum_{i=0}^n (X_i - \theta)^2 \right)$$

Intervalles de confiance

Un intervalle de confiance $[A, B]$ de niveau $1 - \alpha$ est un intervalle aléatoire qui a la probabilité $1 - \alpha$ de contenir le paramètre à estimer θ . Formellement, on écrit : $P(t_1(\theta) \leq f(X_1, \dots, X_n) \leq t_2(\theta)) = P(A \leq \theta \leq B) = 1 - \alpha$

Test d'hypothèses

Dans le cadre d'un test d'hypothèse, nous cherchons à faire valoir une hypothèse en dépit d'une autre, qui lui est contradictoire.

On appellera la première (celle dont le rejet à tort sera le plus préjudiciable) « Hypothèse nulle » et la deuxième « Hypothèse alternative ».

		Réalité		
		H_0 vraie	H_0 fausse	
Décision	Garder H_0	$1 - \alpha$	β	→ Risque de deuxième espèce
	Rejeter H_0	α	$1 - \beta$	
		↓ Risque de première espèce	↓ Puissance du test	

Les calculs qui se cachent derrière le choix de l'hypothèse à garder sont compliqués. Mais BONNE NOUVELLE, la machine fera tout à notre place. Il suffit juste de suivre correctement la méthode :

Etape 1 : Choisir judicieusement les hypothèses à évaluer et fixer le risque α

Etape 2 : Choisir le test adapté à la procédure

Etape 3 : Rentrer la commande correspondante sur R et exécuter

Etape 4 : Lire dans les sorties la p-value. si elle est supérieure à α on accepte H_0 . Si elle lui est inférieure, on rejette H_0

Construction d'intervalles de confiance

Les intervalles de confiance sont des outils essentiels en statistique pour estimer des paramètres inconnus tout en mesurant l'incertitude associée à cette estimation. Ci-dessous, vous trouverez un tableau présentant la construction des intervalles de confiance pour différents paramètres.

Construction d'intervalles de confiance usuels				Pour une proportion p
Pour les paramètres d'une loi $N(\mu, \sigma^2)$				
Pour μ		Pour σ^2		
Avec σ connu	Avec σ inconnu	Avec μ connu	Avec μ inconnu	
L'EMV de μ est :	L'EMV de μ est :	L'EMV de σ^2 est :	L'EMV de σ^2 est :	L'EMV de p est $\hat{p} = \frac{N}{n}$, où $N \sim B(n, p)$ (loi binomiale).
$\bar{X}_n = \frac{1}{n} \sum_{i=0}^n X_i, \bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$	$\bar{X}_n = \frac{1}{n} \sum_{i=0}^n X_i, \bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$	$s_n^2 = \frac{1}{n} \sum_{i=0}^n (X_i - \mu)^2$	$s_n'^2 = \frac{1}{n-1} \sum_{i=0}^n (X_i - \bar{X}_n)^2$	Pour n assez grand ($n \geq 30$), on a : $N \rightarrow N(np, np(1-p))$
On a : $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim N(0, 1)$	Comme σ inconnu, on l'estime avec son équivalent empirique. $s_n'^2 = \frac{1}{n-1} \sum_{i=0}^n (X_i - \bar{X}_n)^2$	On a : $\frac{ns_n^2}{\sigma^2} \sim \chi_n^2$	On a : $\frac{(n-1)s_n'^2}{\sigma^2} \sim \chi_{n-1}^2$	Donc :
Nous allons donc utiliser les fractiles d'une loi normale centrée réduite.	On a : $\sqrt{n} \frac{\bar{X}_n - \mu}{s_n'} \sim St(n-1)$	Nous allons donc utiliser les fractiles d'une loi de Khi-deux à n degrés de liberté.	Nous allons donc utiliser les fractiles d'une loi de Khi-deux à n-1 degrés de liberté.	$\sqrt{n} \frac{\hat{p} - p}{\sqrt{p(1-p)}} \sim N(0, 1)$
$P\left(-u_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$	Nous allons donc utiliser les fractiles d'une loi de Student à n-1 degrés de liberté.	$P\left(c_1 \leq \frac{ns_n^2}{\sigma^2} \leq c_2\right) = 1 - \alpha$	$P\left(c_1 \leq \frac{(n-1)s_n'^2}{\sigma^2} \leq c_2\right) = 1 - \alpha$	$P\left(-u_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\hat{p} - p}{\sqrt{p(1-p)}} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$
Ceci donne :	$P\left(-t_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{s_n'} \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$	Ceci donne :	Ceci donne :	Ceci donne :
$P\left(\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$	Ceci donne :	$P\left(\frac{ns_n^2}{c_2} \leq \sigma^2 \leq \frac{ns_n^2}{c_1}\right) = 1 - \alpha$	$P\left(\frac{(n-1)s_n'^2}{c_2} \leq \sigma^2 \leq \frac{(n-1)s_n'^2}{c_1}\right) = 1 - \alpha$	$P\left(\hat{p} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \leq p \leq \hat{p} + u_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$
Et donc un intervalle de confiance pour μ de niveau de confiance $1 - \alpha$ est :	$P\left(\bar{X}_n - t_{1-\frac{\alpha}{2}} \frac{s_n'}{\sqrt{n}} \leq \mu \leq \bar{X}_n + t_{1-\frac{\alpha}{2}} \frac{s_n'}{\sqrt{n}}\right) = 1 - \alpha$	Et donc un intervalle de confiance pour σ^2 de niveau de confiance $1 - \alpha$ est :	Et donc un intervalle de confiance pour σ^2 de niveau de confiance $1 - \alpha$ est :	Comme p (à l'intérieur de l'intervalle) est inconnu, on l'estime avec son équivalent empirique \hat{p} .
$[\bar{X}_n - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$	Et donc un intervalle de confiance pour μ de niveau de confiance $1 - \alpha$ est :	$[\frac{ns_n^2}{c_2}, \frac{ns_n^2}{c_1}]$	$[\frac{(n-1)s_n'^2}{c_2}, \frac{(n-1)s_n'^2}{c_1}]$	Un intervalle de confiance pour p de niveau de confiance $1 - \alpha$ est :
Où $u_{1-\frac{\alpha}{2}}$ est le fractile d'ordre $1 - \frac{\alpha}{2}$ d'une loi normale centrée réduite.	$[\bar{X}_n - t_{1-\frac{\alpha}{2}} \frac{s_n'}{\sqrt{n}}, \bar{X}_n + t_{1-\frac{\alpha}{2}} \frac{s_n'}{\sqrt{n}}]$	Où c_1 (respectivement c_2) est le fractile d'ordre $1 - \frac{\alpha}{2}$ (respectivement $\frac{\alpha}{2}$) d'une loi de Khi-deux à n degrés de liberté.	Où c_1 (respectivement c_2) est le fractile d'ordre $1 - \frac{\alpha}{2}$ (respectivement $\frac{\alpha}{2}$) d'une loi de Khi-deux à n-1 degrés de liberté.	$[\hat{p} - u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + u_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}]$
	Où $t_{1-\frac{\alpha}{2}}$ est le fractile d'ordre $1 - \frac{\alpha}{2}$ d'une loi de Student à n-1 degrés de liberté.			Où $u_{1-\frac{\alpha}{2}}$ est le fractile d'ordre $1 - \frac{\alpha}{2}$ d'une loi normale centrée réduite.