# Prérequis des mathématiques

Ici, sera listé les pré-requis nécessaires pour le master SEP.

Nous allons commencer par les bases des probabilités, de l'algèbre linéaire et des statisitques.

## **Probabilités**

#### **Définition**

Une variable aléatoire est une fonction mesurable définie sur un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$  vers un espace mesurable  $(E, \mathcal{E})$ . Autrement dit, c'est une fonction  $X : \Omega \to E$  telle que pour tout ensemble mesurable  $B \in \mathcal{E}$ , l'ensemble des issues  $\{\omega \in \Omega : X(\omega) \in B\}$  appartient à  $\mathcal{F}$ .

#### Types de Variables Aléatoires

- Variable aléatoire discrète : Peut prendre un nombre fini ou dénombrable de valeurs. Par exemple, le résultat d'un lancer de dé.
- Variable aléatoire continue : Peut prendre une infinité de valeurs dans un intervalle. Par exemple, la température mesurée à un moment donné.

#### Loi de Probabilité

La loi de probabilité d'une variable aléatoire X décrit la distribution des probabilités associées aux différentes valeurs que X peut prendre.

## **Exemples**

- Lancer de dé : Si X représente le résultat d'un lancer de dé, alors X peut prendre les valeurs  $\{1, 2, 3, 4, 5, 6\}$ .
- **Temps d'attente** : Si Y représente le temps d'attente à un arrêt de bus, alors Y peut prendre n'importe quelle valeur positive.

La loi de probabilité d'une variable aléatoire peut être décrite via une fonction de densité.

#### Fonction de Densité

Une fonction de densité  $f_X$  d'une variable aléatoire continue X est une fonction telle que pour tout intervalle [a,b]:

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x) \, dx$$

La fonction de densité  $f_X$  est utilisée pour décrire la distribution des probabilités d'une variable aléatoire continue.

## **Propriétés**

- 1. Positivité :  $f_X(x) \ge 0$  pour tout x.
- 2. Intégrale égale à 1 :

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1$$

On calcule aussi la fonction de répartition lorsqu'on cherche la probabilité que la variable aléatoire (X) prenne une valeur inférieure ou égale à (x).

#### Fonction de Répartition

La fonction de répartition d'une variable aléatoire X est la fonction  $F_X$  définie pour tout réel x par :

$$F_X(x) = \mathbb{P}(X \leq x)$$

Cette fonction donne la probabilité que la variable aléatoire X prenne une valeur inférieure ou égale à x.

## **Propriétés**

- 1. Croissance :  $F_X$  est une fonction croissante.
- - $\begin{array}{ll} \bullet & \lim_{x \to -\infty} F_X(x) = 0 \\ \bullet & \lim_{x \to \infty} F_X(x) = 1 \end{array}$
- 3. Continuité à droite :  $F_X$  est continue à droite, c'est-à-dire que pour tout x, $\operatorname{lim}_{h\to 0^+}F_X(x+h)=F_X(x).$
- 4. Relation avec la fonction de densité :

$$F_X(x) = \int_{-\infty}^x f_X(t) \, dt$$

## Formules Importantes

• Espérance (ou moyenne) :

$$\mathbb{E}[X] = \sum_{i} x_{i} \mathbb{P}(X = x_{i}) \quad \text{(discrète)}$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx \quad \text{(continue)}$$

• Variance:

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

Ces concepts sont fondamentaux en probabilités et en statistiques pour modéliser et analyser des phénomènes aléatoires.

## Algèbre Linéaire

Nous allons vous rappeler les notions essentielles à connaître sur les matrices.

#### Trace d'une matrice

La trace d'une matrice carrée A est la somme de ses éléments diagonaux :

$$Tr(A) = \sum_{i=1}^{n} a_{ii}$$

Propriétés : - La trace est linéaire :

$$Tr(A + B) = Tr(A) + Tr(B)$$

- Invariance par similitude :

$$Tr(AB) = Tr(BA)$$

Note

La trace d'une matrice ne change pas lorsqu'on change de base.

#### Déterminant d'une matrice

Le déterminant d'une matrice  $A \in \mathcal{M}_{n \times n}$  peut être calculé en utilisant le développement par rapport à une ligne ou une colonne :

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij})$$

où  $(A_{ij})$  est la matrice obtenue en supprimant la (i)-ème ligne et la (j)-ème colonne de (A).

Tip

Dans certains cas, il peut être long de calculer le déterminant d'une matrice. Pour remédier à cela, il est utile de connaître par coeur le calcul d'un déterminant d'une matrice de taille  $2 \times 2$  et d'une matrice diagonale.

## **Déterminant d'une matrice** $(2 \times 2)$

Pour une matrice  $A \in \mathcal{M}_{2 \times 2}$  de la forme :

$$A = \left(\begin{array}{cc} a & b \\ c & d \end{array}\right)$$

le déterminant se calcule en utilisant la formule suivante :

$$det(A) = ad - bc$$

#### Déterminant d'une matrice diagonale

Pour une matrice diagonale de la forme :

$$D = \left( \begin{array}{cccc} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{array} \right)$$

le déterminant est le produit des éléments diagonaux :

$$\det(D) = d_1 \times d_2 \times \dots \times d_n$$

Les déterminants sont particulièrement utiles en algèbre linéaire pour déterminer si une matrice est inversible (une matrice est inversible si et seulement si son déterminant est non nul) et pour résoudre des systèmes d'équations linéaires.

## **Espace Euclidien**

Un **espace euclidien** est un espace vectoriel réel de dimension finie muni d'une forme bilinéaire symétrique définie positive, appelée **produit scalaire** :

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$$

Quelques propriétés importantes du produit scalaire : -  $\langle v,v \rangle = \|v\|_2^2 = \sum_{i=1}^n v_i^2$  -  $\langle v,w \rangle = v^t w = \operatorname{trace}(vw^t)$  -  $\langle v,w \rangle = \langle w,v \rangle$  -  $\langle v,w \rangle = 0 \iff v \perp w$ 

Quelques inégalités à connaître :

• Inégalité de Cauchy-Schwarz :

$$|\langle x, y \rangle| \le ||x|| ||y||$$

• Inégalité triangulaire :

$$||x + y|| \le ||x|| + ||y||$$

#### Orthogonalité

Soient v et w, deux vecteurs appartenant à un espace vectoriel E de dimension n, muni du produit scalaire  $\langle \cdot, \cdot \rangle$  associé à la norme euclidienne  $\| \cdot \|_2$ .On note  $v = (v_1, \dots, v_n)^t$  et  $w = (w_1, \dots, w_n)^t$  où  $v = v_i$  sont des scalaires (i.e. des réels).

## Normalisation d'un vecteur

La normalisation d'un vecteur v est donnée par :

$$x = \frac{v}{\|v\|_2}$$

Ainsi,  $||x||_2 = 1$ . Un espace vectoriel engendré par  $v_1, \dots, v_p$  est identique à celui engendré par leurs versions normalisées.

Deux vecteurs x et y sont **orthogonaux** si leur produit scalaire est nul:

$$\langle x, y \rangle = 0$$

Une base orthonormale est une base où tous les vecteurs sont orthogonaux et de norme 1.

## Matrice de Projection Orthogonale

Soit E un espace vectoriel muni d'un produit scalaire  $\langle \cdot, \cdot \rangle$  et W un sous-espace vectoriel de E. La projection orthogonale d'un élément B de E sur W est définie par :

$$\Pi_W B = \underset{a \in W}{argmin} \|B - a\|_2$$

La matrice de projection orthogonale de E sur W est notée  $\Pi_W$ .

## **Propriétés**

- $\begin{array}{ll} \bullet & \Pi_W^t = \Pi_W \\ \bullet & \Pi_W^2 = \Pi_W \end{array}$
- $\operatorname{trace}(\Pi_W) = \dim(\Pi_W)$
- $\bullet \quad \Pi_W^\perp = I_E \Pi_W$

Pour tout B, élément de E:

$$\Pi_{W^\perp}B=B-\Pi_WB$$

## Construction de matrice de projection Orthogonale

Soient  $v_1,\ldots,v_p,\,p$  vecteurs de  $\mathbb{R}^n$  avec  $p\leq n.$  Si  $(v_1,\ldots,v_p)$  est une base orthogonale de  $\mathbb{R}^p$  et  $W = \text{vect}(v_1, \dots, v_p)$ , alors la matrice de projection W de  $\mathbb{R}^n$  sur W est :

$$W = V(V^tV)^{-1}V^t$$

où V est la matrice dont les colonnes sont les vecteurs  $v_1, \dots, v_p$ .

Si  $(v_1, \dots, v_p)$  est une base orthonormale de W, alors :

$$W = VV^t$$

## Procédé d'orthogonalisation de Gram-Schmidt

Le but du procédé de Gram-Schmidt est de prendre un ensemble de vecteurs linéairement indépendants  $(v_1, v_2, \dots, v_n)$  et de produire un ensemble de vecteurs orthogonaux  $(u_1, u_2, \dots, u_n)$ qui engendrent le même sous-espace.

Tip

Avant de recourir au procédé de Gram-Schmidt, il faut s'assurer que les vecteurs ne soient pas déjà orthogonaux, cela serait une perte de temps de les recalculer. Ce calcul est long, il faut bien prendre son temps pour le faire et ne pas se précipiter.

## Étapes du Procédé

Le premier vecteur orthogonal (u\_1) est le premier vecteur de l'ensemble original :

$$u_1 = v_1$$

Pour chaque vecteur  $(v_k)$  (où  $(k \ge 2)$ ), on soustrait les projections orthogonales de  $(v_k)$  sur les vecteurs orthogonaux précédemment calculés  $(u_1, u_2, \dots, u_{k-1})$ :

$$u_k = v_k - \sum_{j=1}^{k-1} \frac{\langle v_k, u_j \rangle}{\langle u_j, u_j \rangle} u_j$$

Si l'on souhaite obtenir une base orthonormale, chaque vecteur  $(u_k)$  est normalisé pour obtenir  $(e_k)$ :

$$e_k = \frac{u_k}{||u_k||}$$

### Base duale

Pour une base  $\{e_1,e_2,\dots,e_n\}$  d'un espace vectoriel E, la base duale  $(e^1,e^2,\dots,e^n)$  est définie par :

$$e^i(e_i) = \delta_{ij}$$

où  $\delta_{ij}$  est le symbole de Kronecker.

La base duale permet de définir des formes linéaires et de travailler avec des espaces vectoriels de manière plus abstraite.

## Diagonalisation d'une matrice

Une matrice carrée A est dite **diagonalisable** s'il existe une matrice diagonale D et une matrice inversible P telles que :

$$A = PDP^{-1}$$

D est une matrice diagonale contenant les valeurs propres de A

## Diagonalisation d'une matrice symétrique réelle

Pour les matrices symétriques réelles, on peut toujours trouver une base orthonormale de vecteurs propres, ce qui permet de les diagonaliser par une matrice orthogonale Q:

$$A = QDQ^T$$

où Q est une matrice orthogonale.

## Warning

Pour que la propriété soit vérifiée, il est essentiel que la matrice soit réelle, c'est-à-dire sans composantes complexes. En effet, la symétrie d'une matrice complexe n'implique pas nécessairement la propriété en question

## Statistiques Descriptives Univariées

• Moyenne : La moyenne arithmétique d'une série de valeurs  $x_1, x_2, \dots, x_n$  est donnée par:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Médiane : La médiane est la valeur qui sépare la série en deux parties égales. Pour une série ordonnée, si n est impair, la médiane est la valeur centrale. Si n est pair, c'est la moyenne des deux valeurs centrales.
- Variance: La variance mesure la dispersion des valeurs autour de la moyenne:

$$Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

• Écart-type : L'écart-type est la racine carrée de la variance :

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



Plus l'écart-type est grand, plus les données sont dispersées autour de la moyenne.

Les statistiques descriptives univariées sont essentielles pour résumer et comprendre les caractéristiques principales d'une seule variable. Elles sont largement utilisées en analyse de données pour obtenir une vue d'ensemble rapide et efficace. Elles permettent aussi d'identifier rapidement des valeurs extrêmes.

## Statistiques Descriptives Bivariées

## **Définitions**

 $\bullet$  Covariance : La covariance entre deux variables X et Y mesure la manière dont deux variables varient ensemble. Elle est donnée par deux formules :

1. Pour un échantillon de données :

$$\mathrm{Cov}(X,Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

où  $x_i$  et  $y_i$  sont les valeurs des échantillons, et  $\bar{x}$  et  $\bar{y}$  sont les moyennes des échantillons.

2. Pour des variables aléatoires :

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

où  $\mathbb{E}[X]$  et  $\mathbb{E}[Y]$  sont les espérances (moyennes théoriques) des variables aléatoires X et Y.

## • Régression linéaire:

La régression linéaire est une méthode statistique qui permet de modéliser la relation entre deux variables en ajustant une ligne droite à un ensemble de données. L'équation de la régression linéaire est :

$$y = a + bx$$

où y est la variable dépendante, x est la variable indépendante, a est l'ordonnée à l'origine (l'interception de la ligne avec l'axe des ordonnées), et b est la pente de la ligne (le taux de changement de y par rapport à x).

## Fondements de Probabilité



#### Quelques définitions

- On appelle épreuve E toute expérience probabiliste.
- On appelle univers de E l'ensemble, généralement noté  $\Omega$ , 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  $\Omega$ .

- 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 B, est un événement. Il est réalisé si A ou B se produit. Deux événements remarquables sont à retenir:
- L'événement certain  $\Omega_i$ .
- L'événement impossible  $\emptyset_i$ .

Tous les éléments qui n'appartiennent pas à A appartiennent à un événement que l'on appelle le **complémentaire** de A. On le note  $A^c$  ou  $\overline{A}$ .

On dit que deux événements  $\bf A$  et  $\bf B$  sont **incompatibles** s'ils ne peuvent pas être réalisés enmême temps.

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

$$A \cup \overline{A} = \mathbf{\Omega}$$

$$A \cap \overline{A} = \emptyset$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \ et \ \overline{A \cup B} = \overline{A} \cap \overline{B} \ (lois \ de \ Morgan)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \ \cup \ (B \ \cap \ C) \ = (A \ \cup B) \ \cap \ (A \ \cup \ C)$$

#### **Partitions**

La famille d'événements forme une partition de si :

$$\cup_i A_i = \Omega \ et \ A_i \ \cap \ A_j = \emptyset; \ \forall i \neq j; \ i \in I$$

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  $\Omega$  qui vérifie les propriétés suivantes:

- $\Omega \in T$
- Si  $(A_n)_n$  est une suite dénombrable d'éléments de  $T_i$  alors  $\cup$   $A_n$   $\in$  T

Si A est un élément de  ${\cal T}_i$  alors son complémentaire l'est aussi

De plus, si T est une tribu, alors:

- $\emptyset \in T$
- Si  $(A_n)_n$  est une suite d'éléments de  $T_i$  alors  $\cap$   $A_n$   $\in$  T.

Exemple de Tribus:

Commençons par le cas discret.

On considère l'expérience "Lancer une pièce de monnaie équilibrée".

On notera: P "Pile apparait" et F "Face apparait".

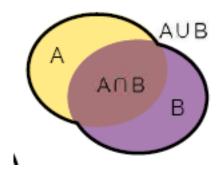
Dans ce cas, l'univers est l'ensemble {P,F} et  $T=\{~\Omega,~\emptyset,~\mathrm{P},~\mathrm{F}~\}$  est une tribu.

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

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

Nous les appelons des Boréliens.

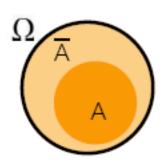
Soient A et B deux événements. Les propriétés suivantes sont toujours vraies:



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

2. 
$$P(B) = P(A \cap B) + P(\bar{A} \cap B)$$

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



De plus, Considérant une suite  $(A_n)_n$  d'événements. On a alors :

$$P(\bigcup_{k=1}^{+\infty}A_k) \ = \lim_{x\to +\infty}(P(\bigcup_{k=1}^nA_k))$$

$$P(\bigcap_{k=1}^{+\infty}A_k) \ = \lim_{x\to +\infty}(P(\bigcap_{k=1}^nA_k))$$

$$P\Bigl(\bigcup_{k=1}^{+\infty}A_k\Bigr)\leq \ \sum_{k=1}^{+\infty}\ P(A_k)$$

 $Et\ si$ :





$$\bigcup_{k=1}^{n} A_k = \Omega$$

Alors:

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

#### Mesure

Soit E un ensemble muni d'une tirbu T . On appelle mesure toute application m :  $T \to R^+$  telle que:

- $m(\emptyset) = 0$ .
- Si  $(A_n)_n$  est une suite d'éléments de T deux à deux disjoints alors:

$$m(\bigcup_n A_n) = \sum_n m(A_n)$$

#### Probabilité

Soit E un ensemble muni d'une tribu T On appelle probabilité toute m :  $T \to R^+$  telle que:

- $P(\emptyset) = 0$
- Si  $(A_n)_n$  est une suite d'éléments de T deux à deux  ${\it disjoints}$  alors:

$$P(\cup_n A_n) = \Sigma_n P(A_n)$$

#### 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 proba non nulles A et B, la probabilité conditionnelle de A sachant que B est réalisé (couramment dit A sachant B) est :

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

Par commutativité de l'intersection nous avons :  $P(A \cap B) = P(B \cap A)$ 

Et donc en utilisant la formule ci-dessus :

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

D'où alors:

$$P(B \setminus A) = \frac{P(A \setminus B)P(B)}{P(A)}$$

C'est ce que nous appelons : La formule de BAYES

## Indépendance

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

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

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 :

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

Notons que si A est indépendant de B, il le sera par rapport à son coplémentaire également et vice versa.

En général, pour une suite  $(A_n)_n$  d'événements indépendants, on a :

$$P(\bigcap_{i=1}^n A_i) \ = \ \prod_{i=1}^n P(A_i) \ = \ P(A_1) \ \dots \ P(A_n)$$

Cette formule est largement utilisée en statistique.

#### Remarque importante:

Il ne faut pas confondre l'indépendance et l'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  $X: T \to R$  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 requte 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 VA traduit le résultat d'une expérience aléatoire en nombre réel. La fontion de répartition transporte le calcul des probabilités concernant les réalisations de la VA.

C'est la fonction définie par :

$$F_r(x) = P(X \le x)$$

Propriétés:

$$\forall x; \ 0 \le F_x(x) \le 1$$

 $F_x$  est une fonction croissante.

$$\lim_{x \rightarrow -\infty} F_x(x) \ = \ 0 \lim_{x \rightarrow \infty} F_x(x) \ = \ 1$$



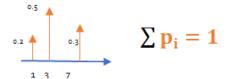
#### Cas discret Cas continu

## Probabilité ponctuelle / Densité

## Cas discret : Probabilité ponctuelle

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

$$P(X = K) = P(X \le K) - P(X \le K - 1) = P_K$$



## Cas continu : densité de probabilité

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

$$f_x(x) = \frac{\delta F_x}{\delta x}(x) \int f_x = 1$$

#### **Moments**

#### Espérance

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

Dans le cas discret :

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

Alors que dans le cas continu:

$$E(X) = \int_{x \in X(\Omega)} x.f_x(x).dx$$

Thérorème de Transfert :

$$E(\phi(X)) = \sum_{k \in X(\Omega)} \phi(k).P(X=k)E(\phi(X)) = \int_{x \in X(\Omega)} \phi(x).f_x(x).dx$$

#### 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 :

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

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

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

### 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 :

$$\tilde{\mu_r} = E((X - E(X))^r)$$

#### Couples aléatoires

La fonction  $F_{x,y}(x,y) = P(X \le x \ \cap \ Y \le y)$  est dite distribution conjointe de X et de Y.

Dans le cas continu, la fonction définie par :

$$f_{x,y}(x,y) = \frac{\delta^2}{\delta_x \delta_y} F_{x,y}(x,y)$$

Est une densité conjointe du couple (X,Y). On a donc :

$$F_{x,y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{x,y}(t,u) dt du$$

Dans le cas discret, on définit la fonction de fréquences conjointes :

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

Et on a donc:

$$F_{x,y} = \sum_{i: x_i \leq x} \sum_{j: y_j \leq y} p_{ij}$$

## Loi marjinale

On définit la loi marginale de X:

$$f_x(x) = \int_{-\infty}^{+\infty} f_{x,y}(x,y) dy$$

Dans le cas continu, ou encore:

$$f_x(x_i) = \sum_j p_{ij}$$

Dans le cas discret :

(De meme on peut définir la densité marginale de Y)

Si X et Y sont indépendants, alors :

$$f_{x,y}(x,y) = f_x(x)f_y(y)$$

#### Covariance

La covariance mesure l'intensité de la relation linéaire entre deux variables aléatoires X et Y.

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

Si X et Y sont indépendants, alors :

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

Attention: La réciproque n'est pas vraie.

## À mémoriser

Soient U, V, X et Y des variables aléatoires et a, b, c et d des constantes réelles.

## Espérance

$$E(aX + bY) = aE(X) + bE(Y)E(a) = a$$

#### Variance

$$V(aX) = a^2 V(X)$$

$$V(A) = 0$$

$$V(X+Y) = V(X) + V(Y) - 2Cov(X,Y)$$

$$V(X - Y) = V(X) + V(Y) - 2Cov(X, Y)$$

## Covariance

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

$$Cov(aX + b, cY + d) = ac.Cov(X, Y)$$

$$Cov(aX + bY, U) = aCov(X, U) + bCov(Y, U)$$

$$Cov(X, cU + dV) = cCov(X, U) + dCov(X, V)$$

$$Cov(aX + bY, cU + dV) = ac.Cov(X, U) + adCov(X, U) + bcCov(Y, U) + bdCov(Y, V)$$

## Vecteurs aléatoires

Un vecteur aléatoire est un n-uplet formé de variables aléatoires. On note  $(X_1,X_2,...,X_n)^t$  L'espérance est toujours linéaire. Pour une suite  $(a_i)_{i\in\{1,...,n\}}$  de réels, on a :

$$E(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$$

Pour les variables indépendantes, on a :

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + \dots + V(X_n)$$

## Lois usuelles

## Lois discètes

distribution	loi de probabilité	$\mathbb{E}(X)$	$\operatorname{var}\left(X\right)$	fonction génératrice $\mathbb{E}(z^X)$
Bernoulli	$\mathbb{P}(X=0) = q, \ \mathbb{P}(X=1) = p$ $q = 1 - p$	p	pq	pz + q
Binomiale $\mathcal{B}(n,p)$	$\mathbb{P}(X=k) = C_n^k p^k q^{n-k}$ $q = 1 - p,  k = 0, 1, \dots, n$	np	npq	$(pz+q)^n$
Poisson $\mathcal{P}(\lambda)$	$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $k = 0, 1, \dots$	λ	λ	$e^{\lambda(z-1)}$
Géométrique $\mathcal{G}(p)$	$\mathbb{P}(X = k) = pq^{k-1}$ $q = 1 - p,  k = 1, 2, \dots$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pz}{1-qz}$
Hypergéométrique $\mathcal{H}(N,n,p)$	$\mathbb{P}(X=k) = \frac{C_{Np}^k C_{Nq}^{n-k}}{C_N^n}$ $q = 1 - p$ $\max(0, n - Nq) \leqslant k \leqslant \min(Np, n)$	np	$npq\frac{N-n}{N-1}$	$\frac{C_{Nq}^n}{C_N^n}F(-n,-Np;Nq-n+1;z)$
Binomiale négative	$\mathbb{P}(X = k) = C_{k+r-1}^{r-1} p^r q^k$ $q = 1 - p,  k = 0, 1, \dots$	$\frac{rq}{p}$	$rac{rq}{p^2}$	$\left(\frac{p}{1-qz}\right)^r$
Pascal	$\mathbb{P}(X = k) = C_{k-1}^{r-1} p^r q^{k-r}$ $q = 1 - p,  k = r, r + 1, \dots$	$\frac{r}{p}$	$rac{rq}{p^2}$	$\left(\frac{pz}{1-qz}\right)^r$

#### $Lois\ absolument\ continues$

distribution	loi de probabilité	$\mathbb{E}(X)$	$\operatorname{var}\left(X\right)$	fonction caract. $\mathbb{E}(e^{itX})$
Uniforme $\mathcal{U}(a,b)$	$\frac{1}{b-a}1_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{ibt} - e^{iat}}{i(b-a)t}$
Exponentielle $\mathcal{E}(\lambda)$	$\lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - it}$
Normale $\mathcal{N}(m, \sigma^2)$	$\frac{1}{\sqrt{2\pi} \ \sigma} \ \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$	m	$\sigma^2$	$e^{imt-\frac{1}{2}\sigma^2t^2}$
Weibull $\mathcal{W}(\lambda,a)$	$\lambda a x^{a-1} e^{-\lambda x^a} 1_{]0,+\infty[}(x)$	$\lambda^{-\frac{1}{a}}\Gamma\left(\frac{1}{a}+1\right)$	$\lambda^{-\frac{2}{a}} \left[\Gamma\left(\frac{2}{a} + 1\right) - \Gamma\left(\frac{1}{a} + 1\right)^{2}\right]$	
Cauchy $C(a,b)$	$\frac{a}{\pi(a^2 + (x-b)^2)}$	non définie	non définie	$e^{ibt-a t }$
Gamma $\Gamma(a,\lambda)$	$\frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} 1_{]0,+\infty[}(x)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - it}\right)^a$
Bêta $B(a,b)$	$\frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} 1_{]0,1[}(x)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	M(a,a+b;it)
Khi-Deux $\chi^2(n)$	$\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{]0,+\infty[}(x)$	n	2n	$(1-2it)^{-n/2}$
Student $\mathcal{T}(n)$	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$	0  si  n > 1	$\frac{n}{n-2} \text{ si } n > 2$	$\frac{2}{\Gamma(\frac{n}{2})} \left(\frac{ t \sqrt{n}}{2}\right)^{\frac{n}{2}} K_{\frac{n}{2}}( t \sqrt{n})$
Fisher $\mathcal{F}(m,n)$	$\frac{m^{\frac{m}{2}}n^{\frac{n}{2}}}{B(\frac{m}{2},\frac{n}{2})} \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}} \mathbb{1}_{]0,+\infty[}(x)$	$\frac{n}{n-2}$ si $n > 2$	$\frac{2n^2(m+n-2)}{m(n-4)(n-2)^2}$ si $n > 4$	$M\left(\frac{m}{2};-\frac{n}{2};-\frac{n}{m}it\right)$

## Théorème de Fisher :

Soient  $\sigma>0,\ m\in\mathbb{R}$  et  $X_1,...,X_n$  des variables aléatoires indépedantes et de même loi  $N(m,\sigma^2)$ . Alors, si  $X=(X_1,...,X_n)$  :

- $(n-1)S_n^2 / \sigma^2 \sim \chi_{n-1}^2$ ;
- $\bullet \ \sqrt{n}(\bar{X}_n-m)\ /\ S_n(X) \sim \tau_{n-1};$

## Théorème de Cochran:

Soient  $\sigma>0, X\sim N_n(0,\sigma^2)$  et  $V_1\oplus\ldots\oplus V_p$  une décomposition de  $\mathbb{R}^n$  en sous-espaces vectoriels orthogonaux de dimensions  $r_1,\ldots,r_p$ .

Alors les projectons orthogonales  $\pi_1,...,\pi_p$  de X sur  $V_1,...,V_p$  sont des vecteurs gaussiens indépendants et pour chaque i=1,...,p:

$$\frac{1}{\sigma^2} ||\pi_i||^2 \sim \chi_{r_i}^2$$

## Statistique inférentielle

## L'échantillonnage

Soit X une v.a. sur  $\Omega$  . Un échantillon de X de taille n est un n-uplet  $(X_i,...,X_n)$  de v.a. iid.

Une réalisation de cet échantillon est un n-uplet de réels  $(x_1,...,x_n)$  où  $X_i(\omega)=x_i.$ 

#### **Estimateur**

Un estimateur de  $\theta$  est une statistique  $\hat{\theta}$  (donc une fonction de  $(X_1,...,X_n)$ ) qui ne dépend pas de  $\theta$  et dont la réalisation est envisagée comme une "bonne valeur" du paramètre  $\theta$ .

#### Risque quadratique

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

$$R_{\hat{\theta}}(\theta) = V(\hat{\theta}) + b_{\hat{\theta}}^2(\theta)$$

#### **Biais**

On appelle biais d'un estimateur  $\hat{\theta}$  pour  $\theta$  la valeur :

$$b_{\hat{\theta}}(\theta) = E(\hat{\theta}) - \theta$$

Un estimateur T est dit sans biais si :

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

#### **Statistique**

On appelle statistique sur un n-échantillon une fonction mesurable de  $(X_1,...,X_n)$  ne dépendant pas de  $\theta$ .

#### Consistance

Un estimateur  $\hat{\theta}$  est dit fortement consistant s'il converge en presque-sûrement vers  $\theta$  lorsque  $n \to +\infty$ .

### Moyenne empirique

On appelle moyenne de l'échantillon ou moyenne empirique, la statistique notée  $\bar{X}$  définie par :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

### Variance empirique

On appelle variance empirique non biaisée, la statistique notée  $S^2_n$  définie par :

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

## Estimation par la méthode du maximum de vraisemblance

On appelle fonction de vraisemblance de  $\theta$  d'un n-échantillon, la fonction suivante :

$$V_{x_1,...,x_n}(\theta) = f_{x_1,...,x_n}(x_1,...,x_n) = \prod_{i=1}^n f_{x_1,\theta}(x_i)$$

avec

$$f_{x_1,\theta}(x_i) = \left\{ \begin{array}{l} f_{X,\theta}(x) \ lorsque \ X \ est \ une \ v.a. \ continue \ de \ densit \ f_{X,\theta} \\ P_{X,\theta}(x) \ lorsque \ X \ est \ une \ v.a. \ discrte \ de \ probabilit \ ponctuelle \ P_{X,\theta} \end{array} \right.$$

La méthode consistant à estimer  $\theta$  par la valeur qui maximise V (vraisemblance) s'appelle méthode du maximum de vraisemblance.

Les étapes à suivre sont les suivantes :

- Calculer la fonction de vraisemblance ci-dessus;
- Calculer le log de la fonction de vraisemblance noté L;
- Calculer la dérivée de la log-vraisemblance obtenue par rapport à  $\theta$ ;
- Trouver la valeur  $\hat{\theta}$  qui annule la dérivée;
- Vérifier que la dérivée seconde par rapport à  $\theta$  est négative en  $\hat{\theta}$ .

En résumé, si la dérivée première s'annulle en  $\theta = \hat{\theta}$  et que la dérivée seconde est négative en  $\theta = \hat{\theta}$ , alors  $\hat{\theta}$  est un maximum local de  $V_{x_1,...,x_n}(\theta)$ .

#### Intervalle de confiance

Un intervalle de confiance permet d'avoir une idée de la marge d'erreur de l'échantillon représentatif sélectionné. En estimant cette marge d'erreur, on est donc en mesure de faire une estimation assez précise de ce qu'aurait été le résultat réel.

Construction d'intervalles de confiance usuels				
Pour les paramètres d'une loi $N\left(\mu,\sigma^{2}\right)$				
Por	ur μ	Pour $\sigma^2$		Pour une proportion p
Avec $\sigma$ connu	Avec σ inconnu	Avec $\mu$ connu	Avec μ inconnu	
Avec $\sigma$ connu $ L'EMV \ de \ \mu \ est : $ $ \overline{X_n} = \frac{1}{n} \sum_{i=0}^n X_i, \overline{X_n} \sim N \left( \mu, \frac{\sigma^2}{n} \right) $ On a: $\sqrt{n} \ \frac{X_n - \mu}{\sigma} \sim N \left( 0, 1 \right) $ Nous allons donc utiliser les fractiles d'une loi normale centrée réduite. $ P \left( -u_{1 - \frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X_n} - \mu}{\sigma} \leq u_{1 - \frac{\alpha}{2}} \right) $ $ = 1 - \alpha $ Ceci donne : $ P \left( \overline{X_n} - u_{1 - \frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \right) $	L'EMV de $\mu$ est : $\overline{X_n} = \frac{1}{n} \sum_{i=0}^n X_i, \overline{X_n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ Comme $\sigma$ inconnu, on l'estime avec son équivalent empirique. $s_n'^2 = \frac{1}{n-1} \sum_{i=0}^n (X_i - \overline{X_n})^2$ On a: $\sqrt{n} \frac{X_{n-\mu}}{s_n'} \frac{1}{n} \sim St (n-1)$ Nous allons donc utiliser les fractiles d'une loi de student à n-1 degrés de liberté. $P\left(-t_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X_n} - \mu}{s_n'} \leq t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$ Ceci donne :	Pou		Pour une proportion p $ \begin{aligned} \text{L'EMV de p est } \hat{p} &= \frac{N}{n}, \text{ où } \\ N \sim & B(n,p) \text{ (loi binomiale)}. \end{aligned} $ Pour n assez grand $(n \geq 30)$ , on a : $N \rightarrow N$ $(np,np(1-p))$ Donc : $ \sqrt{n} \frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim & N(0,1) \\ P\left(-u_1\underline{ \alpha} \leq \sqrt{n} \frac{\hat{p}-p}{\sqrt{p(1-p)}} \leq u_1\underline{ \alpha} \right) = 1-\alpha $ Ceci donne : $ P\left(\hat{p}-u_1\underline{ \beta} = \frac{p(1-p)}{n} \leq p \right) $ $ \leq \hat{p}+u_1\underline{ \beta} = \frac{p(1-p)}{n} $ $ = 1-\alpha $
$\begin{split} &\leq \overline{X_n} + u_1 \tfrac{\sigma}{2} \frac{\sigma}{\sqrt{n}} \\ &= 1 - \alpha \end{split}$ Et donc un intervalle de confiance pour $\mu$ de niveau de confiance $1 - \alpha$ est : $[\overline{X_n} - u_1 - \frac{\sigma}{2} \frac{\sigma}{\sqrt{n}}, \overline{X_n} + u_1 - \frac{\sigma}{2} \frac{\sigma}{\sqrt{n}}]$ Où $u_1 - \frac{\alpha}{2}$ est le fractile d'ordre $1 - \frac{\alpha}{2}$ d'une loi normale centrée réduite.	$\begin{split} P\left(\overline{X_n} - t_1 \frac{\alpha}{2} \frac{S_n}{\sqrt{n}} \leq \mu \right. \\ &\leq \overline{X_n} + t_1 \underline{\alpha} \frac{S_n'}{2} \sqrt{n} \\ &= 1 - \alpha \end{split}$ Et donc un intervalle de confiance pour $\mu$ de niveau de confiance $1 - \alpha$ est : $\left[\overline{X_n} - t_1 \underline{\alpha} \frac{S_n'}{2} \sqrt{n}, \overline{X_n} + t_1 \underline{\alpha} \frac{S_n'}{2} \sqrt{n}\right] \\ \text{Où } t_1 \underline{\alpha} \text{ est le fractile d'ordre} \\ 1 - \frac{\alpha}{2} \text{ d'une loi de student à n-1} \\ \text{degrés de liberté.} \end{split}$	Et donc un intervalle de confiance pour $\sigma^2$ de niveau de confiance $1-\alpha$ est : $\left\lfloor \frac{ns_n^2}{c_2}, \frac{ns_n^2}{c_1} \right\rfloor$ 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.	Et donc un intervalle de confiance pour $\sigma^2$ de niveau de confiance $1-\alpha$ est : $\left[\frac{n-1)s_n'^2}{c_2},\frac{n-1)s_n'^2}{c_1}\right]$ 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.	Comme p (à l'intérieur de l'intervalle) est inconnu, on l'estime avec son équivalent empirique $\hat{p}$ . Un intervalle de confiance pour p de niveau de confiance $1-\alpha$ est : $[\hat{p}-u_1\underline{-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}},\hat{p}+u_1\underline{-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}]$ Où $u_1\underline{-\frac{\alpha}{2}}$ est le fractile d'ordre $1-\frac{\alpha}{2}$ d'une loi normale centrée réduite.

## Test d'hypothèses

Un test d'hypothèses sert à répondre à une question donnant 2 résultats alternatives complémentaires. Il faut alors définir :

- La question (des hypothèses);
- Une façon d'y répondre (une règle de décision).

## Hypothèse

Une hypothèse est un ensemble de valeurs des paramètres inconnus de la population.

Dans une question, on distingue en général deux hypothèses étant :

- Une hypothèse nulle, notée  ${\cal H}_o$  :

$$H_o: \theta \ \in \ \theta_o$$

• Une hypothèse alternative, notée  $H_1$ :

$$H_1:\theta\notin\theta_0$$

Avec  $\theta_o$  une valeur spécifiée pour un paramètre  $\theta$  de la population.

#### Test et test paramétrique

Un test est la donnée d'un jeu d'hypothèse et d'une règle de décision.

Un test peut être unilatéral si l'hypothèse  $H_1$  s'exprime sous forme d'inégalités strictes ou bilatéral si  $H_1$  s'exprime sous forme de différences  $(\neq)$ .

Un test paramétrique est un test pour lequel des hypothèses sur la distribution des populations sont requises.

#### Erreurs et risques

Lorsqu'on prend l'hypothèse nulle, la valeur estimée  $\theta_o$  pour un paramètre  $\theta$  de la population peut conduire à des erreurs. Ces erreurs sont habituellement classés en 2 catégories :

- L'erreur de première espèce;
- L'erreur de seconde espèce

Chaque erreur entraine un risque qui lui correspond:

- Le risque de première espèce, notée  $\alpha$ , est le risque de rejeter l'hypothèse  $H_o$  alors qu'en réalité cette hypothèse est vraie;
- Le risque de seconde espèce, notée  $\gamma$ , est le risque d'accepter l'hypothèse  $H_o$  alors qu'en réalité cette hypothèse est fausse

Le tableau suivant résume l'ensemble des couples (décisions/réalités) possibles :

		Réalité		
		H <sub>o</sub> est vraie	H <sub>o</sub> est fausse	
Décision	$\begin{array}{c} \text{Accepter} \\ \text{H}_{\text{o}} \end{array}$	1 – α	γ	
	Rejeter ${ m H_o}$	α	β	

La quantité  $\beta$  est une probabilité de bonne décision appelée puissance du test.

Lien utile : Récapitulatif des tests statistiques

#### **Econométrie**

#### Modèle linéaire

On cherche à expliquer / prédire une variable  $Y_i$  à l'aide de p variables aléatoires  $X_i^{(1)},...,X_i^{(p)}.$ 

On considère le modèle :

$$Y_i = X_i^{(1)} \beta_1 + \dots + X_i^{(p)} \beta_p + \epsilon_i, \ \forall i \in [1; n].$$

Pour déterminer les paramètres à estimer du modèle, il est plus simple d'écrire le modèle sous forme matricielle :

$$Y = X\beta + \epsilon$$

 $Y \in \mathbb{R}^n$  est appelée variable endogène, c'est à dire la variable à expliquer ou prédire.

 $X \in M_{n,p}(R)$  et contient les  $X_i^{(1)},...,X_i^{(p)}$  qui sont appelées variables exogènes, c'est à dire les variables explicatives.

 $\beta \in \mathbb{R}^p$  contient les paramètres à estimer du modèle.

 $\epsilon \in \mathbb{R}^p$  contient le terme d'erreur du modèle non observable.

## Hypothèses sur les erreurs

4 hypothèses sur les erreurs sont à vérifier dans le cadre d'un modèle linéaire OLS:

- $\forall i \in [1; n]; E(\epsilon_i) = 0$ : les erreurs sont centrées;
- $\forall i \in [1; n]; \ V(\epsilon_i) = \sigma^2$ : les erreurs sont homoscédastiques;
- $\forall i \neq j \in [1;n]; E(\epsilon_i,\epsilon_j) = 0$ : les erreurs ne sont pas auto-corrélées;
- $\forall i \in [1;n]; \epsilon_i \sim N(0,\sigma^2)$  : les erreurs sont normalement distribuées.

## Hypothèses structurelle

Dans le cadre d'un modèle linéaire OLS, on doit retrouver 3 hypothèses structurelles :

- Les variables  $X_1,...,X_p$  sont orthogonales à  $\epsilon;$  Les variables  $X_1,...,X_p$  forment une base de  $R^{p+1};$
- n > p+1.

## Décomposition de la variance

Source	Somme des carrés	Degré de liberté
Expliquée	$\text{SCE} = \textstyle \sum_{i=1}^n \; (\widehat{Y}_i - \overline{Y})^2$	p
Résiduelle	$SCR = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$	n-p-1
Totale	$SCT = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$	n – 1

L'équation d'analyse de la variance est : SCT = SCE + SCR, cela veut dire que la somme des carrées totale = somme des carrées expliquée + la somme des carrées résiduelle.

## Coefficient de détérmination

Le coefficient de détérmination noté  $\mathbb{R}^2$  est défini par :

$$R^2 = \frac{SCE}{SCR}.$$