Cours de Probabilités

Licence M.I.A.S.H.S. Première Année

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Année 2022-2023

Licence MIASHS : un très bon choix !

Pourquoi ? Et pourquoi les mathématiques ?

- Les maths formidable gymnastique de l'esprit : Jacobi "A quoi servent les mathématiques ? : pour l'honneur de l'esprit humain"
- Galilée : "Le livre de l'Univers est écrit en langue mathématique"
- La numérisation n'a jamais été aussi présente
- Les SHS mêlent complexité et humanités numériques
- Les métiers fondés sur les maths appliquées parmi les plus demandés, en particulier en sciences des données
- Mais les mathématiques appliquées requièrent des mathématiques "abstraites"
- La Licence ne doit être qu'un premier pas !

Organisation du cours

1. Cours de 2h, TD de 2h
2. 2 Contrôles Continus (CC1 et CC2) de 1h30 les 15 mars et 19 avril
3. Bonus de 0 pt à + 3pts en TD par participation, quizzs, DMs,...

   Mais **0 pt** à partir de 3 absences non justifiées !
4. Examen final en mai de 2h (Par)
5. Note finale $= \max\left(Par\,,\ \frac{1}{2}(CC + Par)\right)$ où $CC = \max(CC1\,,\ CC2) + \text{Bonus}$

Plan du cours

1. Quelques rappels de statistiques descriptives unidimensionnelles
   - Statistique unidimensionnelle
   - Statistique bidimensionnelle
2. Espace de probabilité, mesure de probabilité et probabilité conditionnelle
   - Espace de probabilité
   - Mesure de probabilité d'un événement
   - Cas particulier de l'équiprobabilité
   - Probabilité conditionnelle et indépendance
3. Variables aléatoires
   - Définitions et propriétés générales
   - Moments d'une variable aléatoire

Statistique

- Historique : "Status", l'état
- Probabilités et statistique : *a priori* et *a posteriori*
- Statistique et statistiques
- $\left\{ \begin{array}{l} \bullet \text{ Statistique descriptive} \\ \bullet \text{ Statistique inférentielle} \\ \bullet \text{ Analyse des données} \\ \bullet \text{ Apprentissage statistique} \\ \bullet \text{ Intelligence artificielle} \\ \bullet \text{ Big data} \end{array} \right.$

Plan du cours

Statistique unidimensionnelle

Exemple de base de données : Avec le logiciel R (gratuit `https://cran.r-project.org/`)

Définition

- *Variable* X *: caractéristique d'individus pouvant prendre plusieurs valeurs*

  $\Longrightarrow$ X_i *pour l'individu* i*.*

- *Variable quantitative : relation d'ordre entre les valeurs de* X

- *Variable qualitative : valeurs de* X *numériques ou alphanumériques ou... Mais pas de relation d'ordre entre les valeurs de* X *(même numériques !)*

  $\Longrightarrow$ *Les différentes valeurs prises par* X *sont des modalités.*

Statistique unidimensionnelle (2)

Définition

X variable et une population de n individus soit $(X_1, \dots, X_n)$ observé :

- *Répartition en k classes $C_1, \dots, C_k$:*
  une classe peut être une valeur alphanumérique, un intervalle de valeurs numériques, un groupe de valeurs alphanumériques,...
- *Effectif n_j d'une classe C_j : $n_j = Card\big(\{X_i \in C_j,\ i = 1, \dots, n\}\big)$*
- *Fréquence f_j d'une classe C_j : $f_j = n_j/n$*
- *Diagramme à bâtons : barre de taille proportionnelle à n_j pour C_j*
- *Diagramme circulaire ("camembert") : portion de $f_j * 360^0$ d'un cercle*

Statistique unidimensionnelle (3)

Définition

Soit X une variable quantitative et $(X_1, \ldots, X_n)$ observé.

- *Classes C_j de la forme $[x_j, x_{j+1}[$ (X continue) ou $\{x_j\}$ (X discrète)*
- *Mode de $(X_1, \ldots, X_n)$: $\frac{1}{2}(x_{j_0} + x_{j_0+1})$ ou x_{j_0}, avec $j_0 = \underset{1 \le j \le k}{\mathrm{Arg\,max}} \{n_j\}$*
- *Amplitude de la classe C_j : $x_{j+1} - x_j$*
- *Densité de fréquence de C_j : $d_j = \frac{1}{x_{j+1} - x_j} f_j$.*
- *Histogramme : on trace la fonction en escalier $f(x) = \sum_{j=1}^{k} d_j \, \mathbb{I}_{x \in C_j}$*

Remarque : Très souvent toutes les amplitudes des classes sont égales ! Les effectifs peuvent remplacer les densité pour construire l'histogramme

Statistique unidimensionnelle (4)

Définition

Soit X une variable quantitative et $(X_1, \ldots, X_n)$ observé.

- *Si $(X_1, \ldots, X_n)$ connu, la fonction de répartition empirique est :*

$$F(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{X_i \leq x} = \frac{1}{n} \, Card\big(\{X_i, \; X_i \leq x\}\big)$$

- *Si répartition par classes $[x_j, x_{j+1}[$ où $j = 1, \ldots, k$, approximation par le "polygône" des fréquences cumulées :*
  on relie les points $(x_j, f_0 + \cdots + f_{j-1})$ pour $j = 1, \ldots, k+1$ avec pour convention $f_0 = 0$.

Statistique unidimensionnelle (5)

Définition

Soit X une variable quantitative et $(X_1, \dots, X_n)$ observé et on note $\min_i(X_i) = X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)} = \max_i(X_i)$ *les X_i classés dans l'ordre.*

1. *Heuristiquement, la médiane empirique est "le nombre m tel qu'il y ait autant de X_i plus petit que m que de X_i plus grand que m"*
   - *Si $(X_1, \dots, X_n)$ connu,* $m = \begin{cases} X_{(\frac{n+1}{2})} & \text{si } n \text{ impair} \\ \frac{1}{2}\left(X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})}\right) & \text{si } n \text{ pair} \end{cases}$
   - *Si répartition par classes $[x_j, x_{j+1}[$, m antécédent de* $1/2$ *par le polygône*
2. *Par extension, les quartiles, sont les antécédents de* $1/4$ *et* $3/4$ *par le polygône, ou* $X_{([n/4])}$ *et* $X_{([3n/4])}$.
3. *Par extension, les déciles, sont les antécédents de* $1/10$, $2/10$,..., $9/10$ *par le polygône, ou* $X_{([n/10])}$, $X_{([2n/10])}$,...,$X_{([9n/10])}$.
4. *Un quantile d'ordre* $p \in]0, 1[$ *: antécédent de p par le polygône, ou* $X_{([pn])}$

Statistique unidimensionnelle (6)

Définition

Soit X une variable quantitative et $(X_1, \dots, X_n)$ observé

- *Si $(X_1, \dots, X_n)$ connu, on définit :*
  - *Moyenne empirique :* $\overline{X} = \frac{1}{n}\left(X_1 + \cdots + X_n\right)$
  - *Variance empirique :* $\overline{\sigma}_X^2 = \frac{1}{n}\left(X_1^2 + \cdots + X_n^2\right) - (\overline{X})^2$
  - *Ecart-type empirique :* $\overline{\sigma}_X = \sqrt{\overline{\sigma}_X^2}$
- *Si répartition par classes $[x_j, x_{j+1}[$*
  - *Moyenne empirique :* $\overline{X} \simeq \frac{1}{n}\left(n_1 \frac{x_1+x_2}{2} + n_2 \frac{x_2+x_3}{2} + \cdots + n_k \frac{x_k+x_{k+1}}{2}\right)$
  - *Variance empirique :* $\overline{\sigma}_X^2 \simeq \frac{1}{n}\left(n_1\left(\frac{x_1+x_2}{2}\right)^2 + \cdots + n_k\left(\frac{x_k+x_{k+1}}{2}\right)^2\right) - (\overline{X})^2$
  - *Ecart-type empirique :* $\overline{\sigma}_X \simeq \sqrt{\overline{\sigma}_X^2}$

Exercice

Exercice : Montrer que :

$$\overline{\sigma}_X^2 = \frac{1}{n}\left(X_1^2 + \cdots + X_n^2\right) - (\overline{X}_n)^2 = \frac{1}{n}\left((X_1 - \overline{X}_n)^2 + \cdots + (X_n - \overline{X}_n)^2\right)$$

Démonstration.

On a :

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2\,X_i\,\overline{X}_n + \overline{X}_n^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} 2\,X_i\,\overline{X}_n + \frac{1}{n}\sum_{i=1}^{n}\overline{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\,\overline{X}_n\,\frac{1}{n}\sum_{i=1}^{n} X_i + \overline{X}_n^2\,\frac{1}{n}\sum_{i=1}^{n} 1 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\,\overline{X}_n\,\overline{X}_n + \overline{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2
\end{aligned}$$

□

Remarque : Cette propriété serait fausse en remplaçant $\frac{1}{n}$ par $\frac{1}{n-1}$.

Plan du cours

Statistique bidimensionnelle

Définition

On observe 2 variables pour n individus : $((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n))$

- *Si les* X_i *(resp.* Y_i*) dans des classes* $(C_j^X)_{1 \leq j \leq k_X}$ *(resp.* $(C_j^Y)_{1 \leq j \leq k_Y}$*)*

  On définit les effectifs joints $n_{j,j'}$ *pour* $1 \leq j \leq k_X$ *et* $1 \leq j' \leq k_Y$

  Répartition des effectifs joints dans un tableau de contingence :

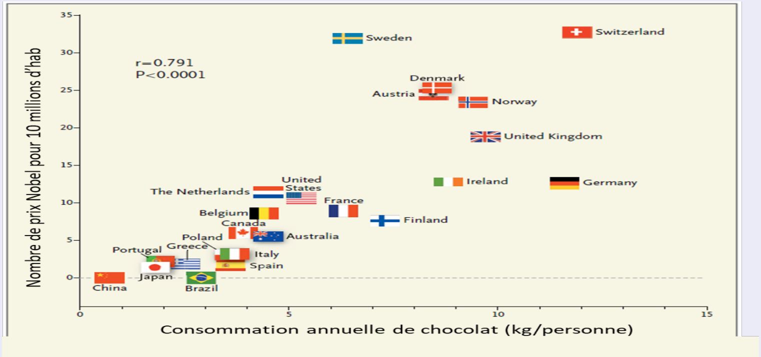
	C_1^X	C_2^X	$\cdots$	$C_{k_X}^X$
C_1^Y	*3*	*7*	$\cdots$	*6*
C_2^Y	*4*	*2*	$\cdots$	*1*
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
$C_{k_Y}^Y$	*11*	*0*	$\cdots$	*7*

Statistique bidimensionnelle (2)

Définition

On observe 2 variables pour n individus : $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$

- *Si* $((X_1, Y_1), \ldots, (X_n, Y_n))$ *connu, nuage de points* $((X_i, Y_i))_{1 \le i \le n}$ *:*



Statistique bidimensionnelle (3)

Définition

$((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ *connu*

- *Covariance empirique :* $\overline{\sigma}_{XY} = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$
- *Corrélation empirique :* $\overline{\rho}_{XY} = \dfrac{\overline{\sigma}_{XY}}{\overline{\sigma}_X \, \overline{\sigma}_Y}$

Remarque : $\overline{\sigma}_{XX} = \overline{\sigma}_X^2$.

Propriété

Pour tout $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$, *on a* $|\overline{\rho}_{XY}| \leq 1$

Remarque : Corrélation et causalité ?

Statistique bidimensionnelle (4)

Démonstration.

On note $X_i' = X_i - \overline{X}$ et $Y_i' = Y_i - \overline{Y}$. Alors $\overline{\sigma}_{XY} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i'\, Y_i'$. Alors pour tout $\lambda \in \mathbb{R}$,

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i' + \lambda\, Y_i'\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i')^2 + 2\,\lambda\,\frac{1}{n}\sum_{i=1}^{n} X_i'\, Y_i' + \lambda^2\,\frac{1}{n}\sum_{i=1}^{n}(Y_i')^2 = \overline{\sigma}_X^2 + 2\,\lambda\,\overline{\sigma}_{XY} + \lambda^2\,\overline{\sigma}_Y^2$$

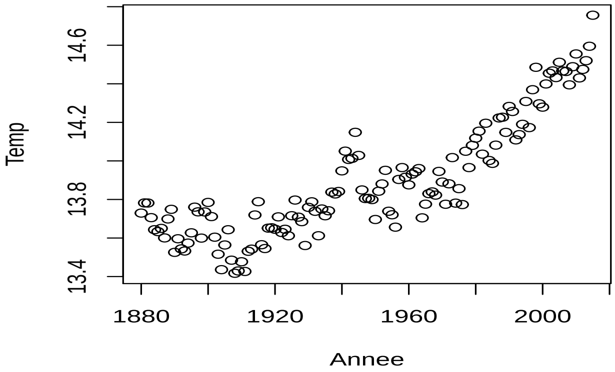
Remarquons que l'on a toujours $\frac{1}{n}\sum_{i=1}^{n}\left(X_i' + \lambda\, Y_i'\right)^2 \geq 0$. Donc le polynôme $P(\lambda) = \overline{\sigma}_X^2 + 2\,\lambda\,\overline{\sigma}_{XY} + \lambda^2\,\overline{\sigma}_Y^2$ est un polynôme positif ou nul pour tout $\lambda \in \mathbb{R}$. Donc le discriminant Δ de ce polynôme est forcément négatif ou nul. Or $\Delta = \left(2\,\overline{\sigma}_{XY}\right)^2 - 4\,\overline{\sigma}_X^2\,\overline{\sigma}_Y^2 \leq 0$. D'où $\left(\overline{\sigma}_{XY}\right)^2 \leq \overline{\sigma}_X^2\,\overline{\sigma}_Y^2$, soit encore $\left|\overline{\sigma}_{XY}\right| \leq \overline{\sigma}_X\,\overline{\sigma}_Y$, entraînant $|\rho_{XY}| \leq 1$. $\square$

Remarque : On montre que si (X_i) non tous nuls et (Y_i) non tous nuls, $\rho_{XY} = 1 \iff \exists\alpha > 0,\ X_i = \alpha\, Y_i$ et $\rho_{XY} = -1 \iff \exists\alpha < 0,\ X_i = \alpha\, Y_i$.

Regression linéaire simple par moindres carrés

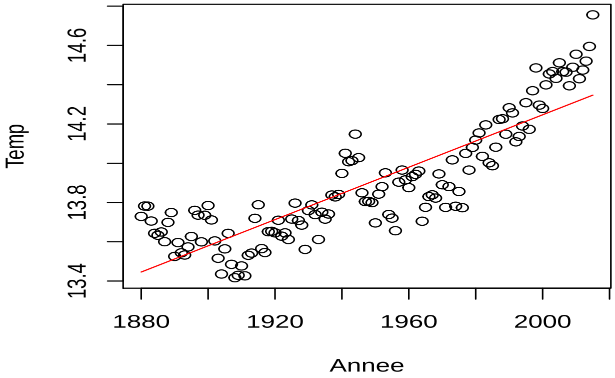
Soit X et Y deux variables quantitatives, $(X_1, Y_1), \ldots, (X_n, Y_n)$ observé

Example : Températures annuelles sur Terre de 1880 à 2015



Regression linéaire simple par moindres carrés (2)

$\Longrightarrow$ Une droite optimale pour l'évolution des températures



Regression linéaire simple par moindres carrés (3)

Définition

On définit une distance par moindres carrés entre $(X_i, Y_i)_{1\leq i\leq n}$ et la droite $y = ax + b$ par

$$\Delta(a,b) = \sum_{i=1}^{n} \big(Y_i - (aX_i + b)\big)^2$$

$$\Longrightarrow (\widehat{a}, \widehat{b}) = \underset{a,b\in\mathbb{R}}{\mathrm{Argmin}}\, \Delta(a,b) \text{ estimateur par moindres carrés de } a \text{ et } b$$

Remarque : Aussi possible $\underset{a,b\in\mathbb{R}}{\mathrm{Argmin}} \sum_{i=1}^{n} \big|Y_i - (aX_i + b)\big|$, estimateur par moindres valeurs absolues

Propriété

Si $\overline{\sigma}_X^2 > 0$, on a $\widehat{a} = \dfrac{\overline{\sigma}_{XY}}{\overline{\sigma}_X^2}$ et $\overline{Y}_n = \widehat{a}\,\overline{X}_n + \widehat{b} \Longrightarrow \widehat{b} = \overline{Y}_n - \widehat{a}\,\overline{X}_n$.

Preuve de l'expression des estimateurs par MC

Démonstration.

Pour tout a et b, et avec $\widehat{a} = \frac{\overline{\sigma}_{XY}}{\overline{\sigma}_X^2}$ et $\widehat{b} = \overline{Y}_n - \widehat{a}\overline{X}_n$ on a :

$$\begin{aligned}
\sum_{i=1}^{n} (Y_i - (aX_i + b))^2 &= \sum_{i=1}^{n} \Big((Y_i - (\widehat{a}X_i + \widehat{b})) + (\widehat{a}X_i + \widehat{b} - (aX_i + b)) \Big)^2 \\
&= \sum_{i=1}^{n} (Y_i - (\widehat{a}X_i + \widehat{b}))^2 + \sum_{i=1}^{n} ((\widehat{a} - a)X_i + (\widehat{b} - b))^2 \\
&\quad + 2\sum_{i=1}^{n} \big((Y_i - \overline{Y}_n) - (\widehat{a}(X_i - \overline{X}_n))\big)\big((\widehat{a} - a)X_i + (\widehat{b} - b)\big).
\end{aligned}$$

En développant ce dernier terme, on $(\widehat{b} - b)\sum_{i=1}^{n} \big((Y_i - \overline{Y}_n) - (\widehat{a}(X_i - \overline{X}_n))\big) = 0$ car $\sum_{i=1}^{n}(Y_i - \overline{Y}_n) = \sum_{i=1}^{n} \widehat{a}(X_i - \overline{X}_n) = 0$. Par ailleurs,

$$(\widehat{a} - a)\sum_{i=1}^{n} \big((Y_i - \overline{Y}_n) - (\widehat{a}(X_i - \overline{X}_n))\big)X_i = (\widehat{a} - a)\big(n\overline{\sigma}_{XY} - \widehat{a}n\overline{\sigma}_X^2\big) = n(\widehat{a} - a)\big(\overline{\sigma}_{XY} - \widehat{a}\overline{\sigma}_X^2\big) = 0,$$

car $\sum_{i=1}^{n} X_i(Y_i - \overline{Y}_n) = \sum_{i=1}^{n}(X_i - \overline{X}_n)(Y_i - \overline{Y}_n) = n\overline{\sigma}_{XY}$ et de la même manière

$$\sum_{i=1}^{n}(X_i - \overline{X}_n)X_i = \sum_{i=1}^{n}(X_i - \overline{X}_n)(X_i - \overline{X}_n) = n\overline{\sigma}_X^2.$$

$\square$

Preuve de l'expression des estimateurs par MC (2)

Démonstration.

Ainsi : $\sum_{i=1}^{n}\left(Y_i-(aX_i+b)\right)^2=\sum_{i=1}^{n}\left(Y_i-(\widehat{a}X_i+\widehat{b})\right)^2+\sum_{i=1}^{n}\left((\widehat{a}-a)X_i+(\widehat{b}-b)\right)^2$. Le premier terme ne dépend pas de a et b, le second terme est ≥ 0, s'annule pour $a=\widehat{a}$ et $b=\widehat{b}$ et seulement dans ce cas (car $(\widehat{a}-a)X_i+(\widehat{b}-b)=0$ pour tout i) dès que deux X_i sont distincts. Donc $\widehat{a}$ et $\widehat{b}$ est l'unique minimum de $\Delta(a,b)$. □

Une mesure de l'adéquation

Définition

Soit $(X_i, Y_i)_{1\le i\le n}$ nuage observé, $(\widehat{Y}_i)_{1\le i\le n}$ valeurs prédites pour Y_i, i.e. $\widehat{Y}_i = g_i(X_1, \dots, X_n)$. Le coefficient de détermination R^2 de la modélisation vaut

$$R^2 = 1 - \frac{\frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2}{\overline{\sigma}_Y^2}.$$

$\Longrightarrow R^2 \to 1$ excellente adéquation.

Proposition

Soit $\widehat{Y}_i = \widehat{a}\, X_i + \widehat{b}$, avec $(\widehat{a}, \widehat{b})$ MC estimateurs. Alors $R^2 = \overline{\rho}_{XY}^2$.

Démonstration.

On a $\frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{a}\, X_i - \widehat{b})^2 = \frac{1}{n}\sum_{i=1}^n \left((Y_i - \overline{Y}_n) - \widehat{a}(X_i - \overline{X}_n)\right)^2$. Alors $\frac{1}{n}\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 = \overline{\sigma}_Y^2 - 2\widehat{a}\,\overline{\sigma}_{XY} + (\widehat{a})^2\overline{\sigma}_X^2 = \overline{\sigma}_Y^2(1 - \overline{\rho}_{XY}^2)$. $\square$

Prédiction et questions

Si X_{n+1} est connu mais pas Y_{n+1}.

$$\Longrightarrow \text{Valeur prédite : } \widehat{Y}_{n+1} = \widehat{a}\,X_{n+1} + \widehat{b}.$$

Questions :

- Régression parabolique plutôt que linéaire ?
- Régression polynomiale ou fonctionnelle plutôt que linéaire ?
- Autres méthodes que les moindres carrés ?
- Sélectionner parmi des modèles et les tester ?

Plan du cours

Espace fondamental

Définition

- *Expérience aléatoire = expérience avec résultat exact imprédictible*
- *Evénement élémentaire* $\{\omega_i\}$ = *résultat possible* ω_i *pris comme singleton*
- *Ensemble fondamental (ou univers)* Ω = *union de tous les événements élémentaires, d'où* $\Omega = \bigcup_{i \in I} \{\omega_i\}$

Exemples :

1. Lancer de deux dés : $\begin{cases} \omega_1 = (3,5) \text{ ou } \omega_2 = (1,1) \text{ résultat possible} \\ \{(2,6)\} \text{ est un événement élémentaire} \\ \Omega = \{(1,1),(1,2),\ldots,(6,6)\} = \{1,2,\ldots,6\}^2 \end{cases}$
2. Pluie (en h) un 5/01 à Paris : $\begin{cases} \omega_1 = 1.23 \text{ ou } \omega_2 = 0 \text{ résultat possible} \\ \{3.7325\} \text{ événement élémentaire} \\ \Omega = [0,24] \end{cases}$
3. touche `RAND` d'une calculatrice ou ordinateur ...

Tribu

Définition

Ω un ensemble fondamental. On appelle **tribu** *$\mathcal{A}$ associée à Ω un ensemble de sous-ensembles de Ω contenant tous les événements ?*

Un ensemble $\in \mathcal{A}$ est appelé un **événement** *$\Longrightarrow$ $\mathcal{A}$ ensemble des événements de l'expérience (dont l'événement impossible $\emptyset$).*

Tribu (2)

Premier exemple :

On lance 2 fois une pièce :

$\Longrightarrow$ Un résultat $\omega = (P, P)$

$\Longrightarrow \Omega = \big\{(P,P),(P,F),(F,P),(F,F)\big\}$

$\Longrightarrow$ Exemple d'événement : $A_1 =$"Au moins 1 P"$= \{(P,F),(F,P),(P,P)\}$

$\Longrightarrow$ Tribu contient tous les événements possibles $+ \emptyset$. D'où

$$\mathcal{A} = \mathcal{P}\big(\Omega\big) : \text{ ensemble de toutes les parties de } \Omega$$

$\mathcal{A} = \{\emptyset, \{(P,P)\}, \{(P,F)\}, \{(F,F)\}, \{(F,P)\}, \{(P,P),(P,F)\}, \{(P,P),(F,F)\}, \{(P,P),(F,P)\}, \{(P,F),(F,P)\}, \{(P,F)$
$\{(F,P),(F,F)\}, \{(P,P),(F,P),(P,F)\}, \{(P,P),(F,P),(F,F)\}, \{(P,P),(F,F),(P,F)\}, \{(P,F),(F,P),(F,F)\}, \Omega\}$

Tribu (3)

Exemples généraux :

1. Si $\Omega = \left\{\omega_1, \ldots, \omega_n\right\}$ famille finie d'événements élémentaires, $\mathcal{P}(\Omega)$ sera la tribu considérée sur Ω :

$$\mathcal{P}(\Omega) = \left\{\emptyset, \{\omega_1\}, \ldots, \{\omega_n\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \ldots, \{\omega_1, \omega_2, \omega_3\}, \ldots, \Omega\right\}$$

2. Si $\Omega = \left\{(\omega_i)_{i \in \mathbb{N}}\right\}$, $\mathcal{P}(\Omega)$ sera la tribu considérée sur Ω

Exemples précédents :

1. Pour $\Omega = \left\{1, \ldots, 6\right\}^2$, on prendra

$$\mathcal{A} = \mathcal{P}\big(\Omega\big) = \Big\{\emptyset, \{(1,1)\}, \ldots, \{(1,1), (1,2)\}, \ldots, \Omega\Big\}$$

2. Pour $\Omega = [0, 24]$, on peut prendre aussi $\mathcal{P}\big([0, 24]\big)$. Mais on préférera une tribu plus "petite" (incluse) notée $\mathcal{B}\big([0, 24]\big)$ et qui contient tous les ensembles formés avec des intervalles de $[0, 24]$

$\Longrightarrow$ Si Ω intervalle de $\mathbb{R}$, on appelera $\mathcal{B}\big(\Omega\big)$ la tribu associée

Evénements

Définition

Pour A et B deux événements de $\mathcal{A}$ une tribu sur Ω alors :

- *On appelle $\overline{A}$ l'événement contraire de A.*
- *On appelle l'événement "A et B" l'ensemble $A \cap B$ qui appartient à $\mathcal{A}$.*
- *On appelle l'événement "A ou B" l'ensemble $A \cup B$ qui appartient à $\mathcal{A}$.*
- *On dit que A et B sont incompatibles lorsque $A \cap B = \emptyset$.*

Exemple :

1. Pour $\Omega = \{1, \ldots, 6\}^2$, considérons $\begin{cases} A = \text{"Les 2 dés sont égaux"} \\ B = \text{"Un des dés marque 4"} \end{cases}$
$\Longrightarrow A = \{(1,1), \ldots, (6,6)\}$ et $B = \{(4,1), \ldots, (4,6), (1,4), \ldots, (6,4)\}$
$\Longrightarrow$ "A et B"$= \{(4,4)\}$ et $\text{Card}(\text{"A ou B"}) = 16$
2. Pour $\Omega = [0, 24]$, on considère $A =$"Il a plu moins d'1h" et $B =$"Il a plu entre 13 et 22h" $\Longrightarrow$ A et B incompatibles

Plan du cours

1. Quelques rappels de statistiques descriptives unidimensionnelles
   - Statistique unidimensionnelle
   - Statistique bidimensionnelle
2. Espace de probabilité, mesure de probabilité et probabilité conditionnelle
   - Espace de probabilité
   - Mesure de probabilité d'un événement
   - Cas particulier de l'équiprobabilité
   - Probabilité conditionnelle et indépendance
3. Variables aléatoires
   - Définitions et propriétés générales
   - Moments d'une variable aléatoire

Probabilité

Définition

Une probabilité sur $(\Omega, \mathcal{A})$ *est une application* $\mathbb{P} : \mathcal{A} \to [0,1]$*, qui à un événement* $E \in \mathcal{A}$ *associe le réel* $\mathbb{P}(E) \in [0,1]$ *et telle que :*

- $\mathbb{P}(\Omega) = 1$.
- *Si* $(E_j)_{j \in J \subset \mathbb{N}}$ *événements incompatibles de* $\mathcal{A}$, $\mathbb{P}\Big(\bigcup_{j\in J} E_j\Big) = \sum_{j \in J} \mathbb{P}(E_j)$

Exemples génériques : 1. Pour $\Omega = \{\omega_1, \dots, \omega_n\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$, probabilité uniforme définie par $\mathbb{P}(A) = \mathrm{Card}(A)/n$ pour tout $A \in \mathcal{A}$.

2. Pour $\Omega = \mathbb{N}^*$, $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}\big(A\big) = \sum_{k \in A} p_k$ pour tout $A \in \mathcal{A}$ et $\sum_{k=1}^{\infty} p_k = 1$ définit une probabilité.

3. Pour $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$, on peut associer la probabilité uniforme définie par $\mathbb{P}([a', b']) = b' - a'$ pour tout $0 \leq a' < b' \leq 1$.

Probabilité (2)

Exemples concrets :

1. Lancer de 2 pièces équilibrées : $\mathcal{A} = \mathcal{P}\big(\{(P,P),(P,F),(F,P),(F,F)\}\big)$
$\Longrightarrow \mathbb{P}(A) = \mathrm{Card}(A)/4$. Exemple $\mathbb{P}($"Au moins 1 P"'$) = 3/4$

2. Au P/F, nombre d'essais avant P : $\mathcal{A} = \mathcal{P}(\mathbb{N}^*)$, $\mathbb{P}\big(\{k\}\big) = 2^{-k}$
$\Longrightarrow \mathbb{P}(\text{"Nombre d'essais} \geq 4\text{"}) = 1 - \mathbb{P}(\text{"Nombre d'essais} \leq 3\text{"})$
$= 1 - \mathbb{P}\big(\{1\}\big) - \mathbb{P}\big(\{2\}\big) - \mathbb{P}\big(\{3\}\big) = 1 - 1/2 - 1/4 - 1/8 = 1/8$

3. Choisir un réel "au hasard" entre 0 et 1 : Pour $\Omega = [0,1]$, $\mathcal{A} = \mathcal{B}([0,1])$
$\Longrightarrow \mathbb{P}(\text{"}Nombre = 5\text{"}) = 0$ et $\mathbb{P}(\text{"}\ln 2 < Nombre < 0.7\text{"}) = 0.7 - \ln 2$

Quelques définitions et propriétés

Définition

Pour Ω *un ensemble fondamental,* $\mathcal{A}$ *une tribu sur* Ω*,*

- $(\Omega, \mathcal{A})$ *est un espace probabilisable ;*
- *Si* $\mathbb{P}$ *probabilité sur* $(\Omega, \mathcal{A})$*,* $(\Omega, \mathcal{A}, \mathbb{P})$ *est un espace de probabilité.*

Propriété

Si $(\Omega, \mathcal{A}, \mathbb{P})$ *est un espace de probabilité,*

- $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$ *pour* $A \in \mathcal{A}$.
- $\mathbb{P}(\emptyset) = 0$.
- *Si* $A, B \in \mathcal{A}$*,* $A \subset B$*,* $0 \leq \mathbb{P}(A) \leq \mathbb{P}(B) \leq 1$.
- $\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B)$*, pour* $A, B \in \mathcal{A}$.
- *Pour* $(A_i)_{i \in \mathbb{N}}$ *une suite d'événements de* $\mathcal{A}$*, telle que* $A_i \subset A_{i+1}$ *pour tout* $i \in \mathbb{N}$*, alors* $\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \lim_{n \to \infty} \mathbb{P}(A_n)$.

Quelques définitions et propriétés (2)

Démonstration.

- A et $\overline{A}$ sont incompatibles ($A \cap \overline{A} = \emptyset$) et $A \cup \overline{A} = \Omega$, donc $\mathbb{P}(A) + \mathbb{P}(\overline{A}) = \mathbb{P}(\Omega) = 1$
- On prend $A = \emptyset$ et la propriété précédente
- On a $B = A \cup \left(B \cap \overline{A}\right)$ et comme A et $B \cap \overline{A}$ incompatibles, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap \overline{A})$, d'où $\mathbb{P}(B) \geq \mathbb{P}(A)$ car $\mathbb{P}(B \cap \overline{A}) \geq 0$
- Considérons les ensembles $A \cap \overline{B} = A \setminus A \cap B$ et $B \cap \overline{A} = B \setminus A \cap B$. Alors $A \cap B$, $A \cap \overline{B}$ et $B \cap \overline{A}$ sont trois événements incompatibles et $(A \cap B) \cup (A \cap \overline{B}) \cup (B \cap \overline{A}) = A \cup B$. D'où
$$\mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(B \cap \overline{A}) = \mathbb{P}(A \cup B). \tag{1}$$
Mais $(A \cap \overline{B})$ et $(A \cap B)$ sont incompatibles et $(A \cap \overline{B}) \cup (A \cap B) = A$. D'où $\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A) - \mathbb{P}(A \cap B)$. De même, $\mathbb{P}(B \cap \overline{A}) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$. En reportant dans (1), on obtient : $\mathbb{P}(A \cap B) + \mathbb{P}(A) - \mathbb{P}(A \cap B) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(A \cup B)$, d'où $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B)$.
- Pour $n \in \mathbb{N}^*$, on a $\bigcup_{i=0}^n A_i = A_n$ car $A_i \subset A_n$ pour $i \leq n$. Donc $\mathbb{P}\left(\bigcup_{i=0}^n A_i\right) = \mathbb{P}(A_n)$. Or la suite $\left(\mathbb{P}\left(\bigcup_{i=0}^n A_i\right)\right)_n$ est une suite croissante majorée par 1. Elle est donc convergente. Donc $\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=0}^n A_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$. Mais $\bigcup_{i\in\mathbb{N}} A_i = A_0 \bigcup \left(\bigcup_{i\in\mathbb{N}} (A_{i+1} \setminus A_i)\right)$ tous ces événements étant incompatibles. Donc $\mathbb{P}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \mathbb{P}(A_0) + \sum_{i=0}^\infty \mathbb{P}\left(A_{i+1} \setminus A_i\right) = \mathbb{P}(A_0) + \sum_{i=0}^\infty \mathbb{P}(A_{i+1}) - P(A_i)$. Ceci est une série télescopique qui vaut exactement $\lim_{n\to\infty} \mathbb{P}(A_n)$, puisque cette limite existe.

Quelques définitions et propriétés (3)

Définition

Soit Ω un ensemble et $J \subset \mathbb{N}$. On dit que $(E_i)_{i\in J}$ famille de $\mathcal{A}$ forme une **partition** *de Ω dans $\mathcal{A}$ si :*

- *Les E_i sont incompatibles deux à deux soit $E_i \cap E_j = \emptyset$ pour $i \neq j$.*
- *L'ensemble des E_i couvre Ω soit $\bigcup_{i\in J} E_i = \Omega$.*

Exemple : Si $A \in \mathcal{A}$, alors A et $\overline{A}$ partition de Ω dans $\mathcal{A}$.

Proposition

(Formule des probabilités totales) *Soit $(\Omega, \mathcal{A}, P)$ espace de probabilité et $(E_j)_{j\in J}$ partition de Ω dans $\mathcal{A}$. Alors, pour tout $A \in \mathcal{A}$,*

$$\mathbb{P}(A) = \sum_{j\in J} \mathbb{P}(A \cap E_j).$$

Quelques définitions et propriétés (4)

Démonstration.

Comme $A \cap E_i \subset E_i$ et $A \cap E_j \subset E_j$ et $E_i \cap E_j = \emptyset$ pour $i \neq j$, alors on a $(A \cap E_i) \cap (A \cap E_j) = \emptyset$ pour $i \neq j$. Ainsi $\mathbb{P}\Big(\bigcup_{j \in J} (A \cap E_j)\Big) = \sum_{j \in J} \mathbb{P}(A \cap E_j)$. De plus, par distributivité de l'intersection et l'union d'ensembles (à l'égal de celle dans $\mathbb{R}$, l'union jouant le rôle de $+$ et l'intersection de $\times$), on a :

$$\bigcup_{j \in J} (A \cap E_j) = A \bigcap \Big(\bigcup_{j \in J} E_j\Big) \qquad \text{(comme une factorisation !).}$$

Comme $\bigcup_{j \in J} E_j = \Omega$ (partition !), alors $\bigcup_{j \in J} (A \cap E_j) = A \cap \Omega = A$. Par conséquent, on a bien $\mathbb{P}(A) = \sum_{j \in J} \mathbb{P}(A \cap E_j)$. □

Exemples d'utilisation :

- On lance n fois une pièce équilibrée. $\mathbb{P}\big(\text{"Nombre de } P = k\text{"}\big)$?
  $\Omega = \{P, F\}^n$, $\mathcal{A} = \mathcal{P}(\Omega)$ et probabilité uniforme. Il y a $\mathrm{Card}(\Omega) = 2^n$ tirages possibles, d'où $\Omega = \{\omega_i\}_{1 \leq i \leq 2^n}$. Mais $(\{\omega_i\})_{1 \leq i \leq 2^n}$ partition de Ω dans $\mathcal{A}$. Donc $\mathbb{P}(A) = \sum_{\omega_i \in A} 2^{-n} = \mathrm{Card}(A)/2^n = C_n^k/2^n$

Quelques définitions et propriétés (5)

Exemples d'utilisation de la formule des probabilités totales :

- Pour $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et A et B deux événements de $\mathcal{A}$, on a toujours :

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

- Soit $\Omega = \{\omega_i\}_{i \in I}$ où $I \subset \mathbb{N}$ et $\mathbb{P}\big(\{\omega_k\}\big) = p_k$ pour tout $k \in I$ alors $\mathbb{P}$ est entièrement définie par les p_k car pour tout $A \in \mathcal{A}$,

$$\mathbb{P}(A) = \sum_{k \in I} \mathbb{P}\big(A \cap \{\omega_k\}\big) = \sum_{k,\ \{\omega_k\} \subset A} p_k.$$

Exemple concret : On jette une pièce. Quelle est la probabilité que le nombre d'essais avant d'avoir un Pile soit pair ?
On a vu $\mathbb{P}("k") = p_k = 2^{-k}$ pour tout $k \in \mathbb{N}^*$. Alors :

$$\mathbb{P}(\text{"Nombre d'essais pair"}) = \sum_{k=1}^{\infty} p_{2k} = \sum_{k=1}^{\infty} \left(\frac{1}{4}\right)^k = \frac{1}{4} \frac{1}{1 - 1/4} = \frac{1}{3}.$$

Plan du cours

1. Quelques rappels de statistiques descriptives unidimensionnelles
   - Statistique unidimensionnelle
   - Statistique bidimensionnelle
2. Espace de probabilité, mesure de probabilité et probabilité conditionnelle
   - Espace de probabilité
   - Mesure de probabilité d'un événement
   - Cas particulier de l'équiprobabilité
   - Probabilité conditionnelle et indépendance
3. Variables aléatoires
   - Définitions et propriétés générales
   - Moments d'une variable aléatoire

Equiprobabilité

Définition

Soit Ω *un ensemble fini et la tribu* $\mathcal{A} = \mathcal{P}(\Omega)$ *associée. On dit que* $\mathbb{P}$ *est la mesure uniforme sur* $(\Omega, \mathcal{A})$ *si* $\forall \omega,\ \omega' \in \Omega,\ \mathbb{P}(\{\omega\}) = \mathbb{P}(\{\omega'\})$ *: Equiprobabilité.*

Propriété

Si Ω *fini,* $\mathbb{P}$ *probabilité uniforme sur* $(\Omega, \mathcal{A})$*, alors* $\forall A \in \mathcal{A},\ \mathbb{P}(A) = \dfrac{card(A)}{card(\Omega)}$.

Cas particulier : Pour tout $\omega \in \Omega,\ \mathbb{P}(\{\omega\}) = 1/\mathrm{card}(\Omega)$.

Exemple : Dé équilibré avec $\Omega = \{1, 2, 3, 4, 5, 6\}$

Equiprobabilité (2)

Remarque : Si équiprobabilité, alors :

$$\text{(Calculer une probabilité)} \iff \text{(Calculer le cardinal d'un ensemble)}$$

$\Longrightarrow$ Résultats combinatoires : on tire k éléments dans un ensemble de n

- S'il y a remise, et que l'ordre compte, un tirage est un **k-uplet**, et le nombre total de k-uplets est : n^k.
- S'il n'y a pas remise, et que l'ordre compte, un tirage est un **arrangement**, et le nombre total d'arrangements est :

$$A_n^k = n(n-1) \times \cdots \times (n-k+1) = n!/(n-k)!$$

- S'il n'y a pas remise, et que l'ordre ne compte pas, un tirage est une **combinaison**, et le nombre total de combinaisons est :

$$C_n^k = A_n^k/k! = n!/(k!(n-k)!)$$

Plan du cours

Probabilité conditionnelle

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace de probabilité. Si* A, $B \in \mathcal{A}$ *et* $\mathbb{P}(A) \neq 0$, *la* **probabilité conditionnelle** *de* B *sachant* A *est* $\mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$

Remarque : Calculer des probabilités sachant $A \Longrightarrow$ travailler avec A nouvel espace fondamental, une tribu associée et une nouvelle probabilité

Exemple : Pile ou face 2 fois avec pièce équilibrée. Calculer la probabilité d'avoir (P, P) sachant qu'on a obtenu au moins un P ?

$\Omega = \big\{(P,P),(P,F),(F,P),(F,F)\big\}$ et $\mathcal{A} = \mathcal{P}(\Omega)$, $\mathbb{P}$ uniforme
$\Longrightarrow A =$"Au moins 1 P"$= \{(P,F),(F,P),(P,P)\}$ et $B = \{(P,P)\}$
$\Longrightarrow \mathbb{P}(B \mid A) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(B)}{\mathbb{P}(A)} = \dfrac{1}{3}$

Exercice

Un exercice revu :

- On lance un dé équilibré, puis on lance le nombre du dé fois une pièce équilibrée. On note le nombre de Piles. Calculer $\mathbb{P}(A)$ avec $A =$ "4 Piles" ?

  On note D_i l'événement : "le dé a montré i", pour $i = 1, \ldots, 6$.
  Alors $(D_i)_{1 \leq i \leq 6}$ est une partition d'où :
  $$\Longrightarrow \mathbb{P}(A) = \sum_{i=1}^{6} \mathbb{P}(A \cap D_i) = \sum_{i=4}^{6} \mathbb{P}(A \cap D_i)$$
  Par définition, on a $\mathbb{P}(A \cap D_i) = \mathbb{P}(D_i)\,\mathbb{P}(A \mid D_i)$.
  Mais pour $i \geq 4$, $\mathbb{P}(A \mid D_i) = C_i^4\, 2^{-i}$, d'où $\mathbb{P}(A \cap D_i) = \frac{1}{6}\, C_i^4\, 2^{-i}$.
  Ainsi $\mathbb{P}(A) = \frac{1}{6}\,(2^{-4} + 5 \cdot 2^{-5} + 15 \cdot 2^{-6}) = \frac{1}{6}\,(4 + 10 + 15)\, 2^{-6} \simeq 0.076$.

Indépendance

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace de probabilité.* A, $B \in \mathcal{A}$, *avec* $\mathbb{P}(A) \neq 0$ *sont* **indépendants** *si* $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.

Conséquence : Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, A, $B \in \mathcal{A}$) :

$$\Big(A \text{ et } B \text{ indépendants}\Big) \Longleftrightarrow \Big(\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Big)$$

Remarque : Ne pas confondre indépendance et incompatibilité !

Exemple : P/F 2 fois pièce équilibrée $A = \{(P, F), (F, F)\}$,
$B = \{(F, P), (F, F)\}$

$\Longrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(\{(F, F)\}) = \frac{1}{4} = \mathbb{P}(A)\ \mathbb{P}(B)$: A et B indépendants mais pas incompatibles !

Indépendance (2)

Définition

$(A_i)_{i\in I}$, $I \subset \mathbb{N}$, *famille d'événements de* $(\Omega, \mathcal{A}, \mathbb{P})$. $(A_i)_{i\in I}$ *est une famille d'événements (mutuellement) indépendants si et seulement si pour tout* $k \in \mathbb{N}^*$, *pour tout* $j_1, \cdots, j_k \in I^k$ *distincts,*
$\mathbb{P}(A_{j_1} \cap \cdots \cap A_{j_k}) = \mathbb{P}(A_{j_1}) \times \cdots \times \mathbb{P}(A_{j_k})$.

Remarque : Etre mutuellement indépendant est plus contraignant que d'être indépendant deux à deux !

Exemple : P/F 2 fois pièce équilibrée $\begin{cases} A : \text{premier lancer P} \\ B : \text{second lancer P} \\ C : \text{deux lancers identiques} \end{cases}$.

On a A et B indépendants, A et C également, de même que B et C, et pourtant A, B et C ne sont pas mutuellement indépendants... car
$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{(P,P)\}) = \frac{1}{4} \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$

Formule de Bayes

Proposition

(Formule de Bayes) *Soit* $(\Omega, \mathcal{A}, \mathbb{P})$ *espace de probabilité et* $(E_j)_{j\in J}$ *famille d'événements de* $\mathcal{A}$ *et partition de* $(\Omega, \mathcal{A})$*. Pour tout* $A \in \mathcal{A}$*, si on connaît* $\mathbb{P}(E_j)$ *et* $\mathbb{P}(A \mid E_j)$ *pour tout* $j \in J$*, alors*

$$\mathbb{P}(E_k \mid A) = \frac{\mathbb{P}(A \mid E_k)\,\mathbb{P}(E_k)}{\sum_{j\in J} \mathbb{P}(A \mid E_j)\,\mathbb{P}(E_j)} \quad \textit{pour } k \in J$$

Démonstration.

On a $\mathbb{P}(E_k \mid A) = \frac{\mathbb{P}(E_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid E_k)\,\mathbb{P}(E_k)}{\mathbb{P}(A)}$. Mais $\mathbb{P}(A) = \sum_{j\in J} \mathbb{P}(E_j \cap A)$ par la formule des probabilités totale, d'où $\mathbb{P}(A) = \sum_{j\in J} \mathbb{P}(A \mid E_j)\,\mathbb{P}(E_j)$. D'où le résultat. $\square$

Formule de Bayes (2)

Exemple : Equipe de foot qui joue autant à domicile qu'à l'extérieur, gagne 1 fois sur 2 à domicile, 1 fois sur 3 à l'extérieur, fait nul 1 fois sur 4 à domicile, 1 fois sur 3 à l'extérieur. Déterminer la probabilité qu'elle gagne, puis la probabilité d'être à domicile sachant qu'elle a perdu.

Démonstration.

On note les événements :

- "M" : jouer à domicile (maison), "E" : jouer à l'extérieur
- "V" : victoire, "N" : nul, "D" : défaite

Alors : $\mathbb{P}(M) = \mathbb{P}(E) = 1/2$, $\mathbb{P}(V \mid M) = 1/2$, $\mathbb{P}(N \mid M) = 1/4$, $\mathbb{P}(V \mid E) = 1/3$ et $\mathbb{P}(N \mid E) = 1/3$. D'où $\mathbb{P}(D \mid M) = 1/4$ et $\mathbb{P}(D \mid E) = 1/3$.

• $\mathbb{P}(V) = \mathbb{P}(V \cap M) + \mathbb{P}(V \cap E) = \mathbb{P}(V \mid M)\,\mathbb{P}(M) + \mathbb{P}(V \mid E)\,\mathbb{P}(E) = (1/2 + 1/3)/2 = 5/12.$

• $\mathbb{P}(M \mid D) = \dfrac{\mathbb{P}(M \cap D)}{\mathbb{P}(D)} = \dfrac{\mathbb{P}(D \mid M)\,\mathbb{P}(M)}{\mathbb{P}(D \mid M)\mathbb{P}(M) + \mathbb{P}(D \mid E)\mathbb{P}(E)} = \dfrac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{3}} = \dfrac{3}{7}$ □

Plan du cours

Premières définitions

Définition

Soit $(\Omega, \mathcal{A})$ *un espace probabilisable.* X *variable aléatoire à valeurs dans* $I \subset \mathbb{R}$*, si* X *application de* $\Omega \to I$ *telle que pour tout* $x \in \mathbb{R}$*, l'événement*

$$\{X \leq x\} = \{\omega \in \Omega \text{ tel que } X(\omega) \leq x\} = X^{-1}(]-\infty, x])$$

soit un événement de $\mathcal{A}$*.*

Exemple : $\Omega = \{(P,P),(P,F),(F,F),(F,P)\}$, $\mathcal{A} = \mathcal{P}(\Omega)$, X nombre de P.
Pour chaque $\omega \in \Omega$, $X(\omega) \in \{0,1,2\} = I$, par exemple $X(\{(P,F)\}) = 1$.

$$\text{On a } \begin{cases} \{X \leq x\} = \emptyset & \text{pour } x < 0 \\ \{X \leq x\} = \{(F,F)\} & \text{pour } 0 \leq x < 1 \\ \{X \leq x\} = \{(F,F),(P,F),(F,P)\} & \text{pour } 1 \leq x < 2 \\ \{X \leq x\} = \Omega & \text{pour } 2 \leq x \end{cases}$$

$\Longrightarrow$ X variable aléatoire.

Deux cas particuliers importants sont à distinguer :

Définition

- *Si* $I = \{x_j\}_{j \in J}$ *avec* $J \subset \mathbb{N}$ *(par exemple* $I = \{0, 1\}$*,* $I = \mathbb{Z}$*,...),* X *est appelée* **variable aléatoire discrète**.
- *Si* I *est une union dénombrable de "vrais" intervalles de* $\mathbb{R}$ *(par exemple* $I = [0, 1]$*,* $I = \mathbb{R}^+$*,...),* X *peut être une* **variable aléatoire continue**.

Exemples :

- Soit $\Omega = \{1, \ldots, 6\}^2$, $\mathcal{A} = \mathcal{P}(\Omega)$ et $X\big((i,j)\big) = i + j$ pour $(i,j) \in \Omega$.
  X est à valeurs dans $I = \{2, \ldots, 12\}$.
  On a $\begin{cases} \{X \leq x\} = \emptyset & \text{pour } x < 2 \\ \{X \leq x\} = \big\{(1,1)\big\} & \text{pour } 2 \leq x < 3 \\ \{X \leq x\} = \big\{(1,1),(1,2),(2,1)\big\} & \text{pour } 3 \leq x < 4 \\ \ldots\ldots\ldots & \ldots\ldots \end{cases}$
  $\Longrightarrow X$ variable aléatoire.
- Soit $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$ et X tel que $X(\omega) = 2\omega - 1$ pour $\omega \in \Omega$.
  X est à valeurs dans $I = [-1, 1]$.
  On a $\begin{cases} \{X \leq x\} = \emptyset & \text{pour } x < -1 \\ \{X \leq x\} = [0, (x+1)/2] & \text{pour } -1 \leq x < 1 \\ \{X \leq x\} = [0, 1] & \text{pour } 1 \leq x \end{cases}$
  $\Longrightarrow X$ variable aléatoire.

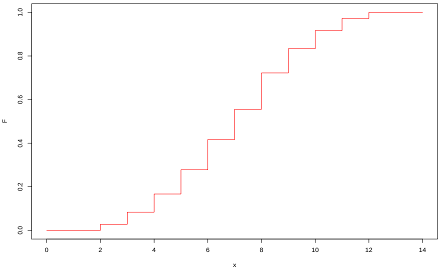
Fonction de répartition

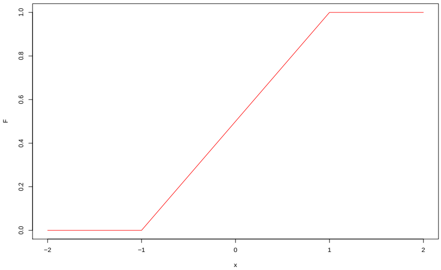
Remarque : Si $\mathcal{A} = \mathcal{P}(\Omega)$, $\{X \leq x\} \in \mathcal{A}$ pour tout $x \in \mathbb{R}$: toute application est une v.a.

Définition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ *un espace de probabilités et* X *une variable aléatoire sur* $(\Omega, \mathcal{A}, \mathbb{P})$ *à valeurs dans* I. *On appelle* **fonction de répartition** *de* X *la fonction* $F_X : \mathbb{R} \to [0,1]$ *telle que* $F_X(x) = \mathbb{P}(X \leq x)$, *pour* $x \in \mathbb{R}$.

Exemple : Tracés des fonctions de répartition pour 2 exemples précédents :





Propriété

1. F_x *est une fonction croissante sur* $\mathbb{R}$.
2. $\lim\limits_{x\to-\infty} F_X(x) = 0$ *et* $\lim\limits_{x\to+\infty} F_X(x) = 1$.
3. $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$, *pour tout* $-\infty \leq a < b \leq +\infty$.

Démonstration.

1. Si $x \leq y$, $\{X \leq x\} \subset \{X \leq y\}$, d'où $\mathbb{P}(\{X \leq x\}) \leq \mathbb{P}(\{X \leq y\})$
$$\Longrightarrow \quad F_X(x) \leq F_X(y).$$
2. On a montré que si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événement de $\mathcal{A}$ tel que $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$, alors $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right)$.
Soit $(x_n)_{n\in\mathbb{N}}$ une suite croissante de réels quelconque telle que $\lim\limits_{n\to\infty} x_n = \infty$. Si $A_n = \{X \leq x_n\}$, $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = \lim\limits_{n\to\infty} F_X(x_n) = \mathbb{P}(\Omega) = 1$. D'où $\lim\limits_{x\to+\infty} F_X(x) = 1$.
On a $F_X(x) = 1 - \mathbb{P}(X > x)$. Même preuve pour (x_n) suite décroissante tendant vers $-\infty$ et $A_n = \{X > x_n\} \Longrightarrow \lim_{x\to-\infty} \mathbb{P}(X > x) = 1 \Longrightarrow \lim\limits_{x\to-\infty} F_X(x) = 0$.
3. $A = \{X \leq a\}$, $B = \{a < X \leq b\}$ et $C = \{X > b\}$. Alors A, B et C partition de Ω. D'où $\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = 1$. Comme $\mathbb{P}(A) = F_X(a)$, $1 - \mathbb{P}(C) = F_X(b)$ on obtient $\mathbb{P}(B) = F_X(b) - F_X(a)$.

Fonction de répartition (suite)

Propriété

Si $X : \Omega \to (x_j)_{j\in J}$, $J \subset \mathbb{N}^*$ *est une v.a. discrète, alors*

- F_X *est une fonction en escalier sur* $\mathbb{R}$*, avec sauts en les* x_j.
- $\mathbb{P}(X = x_j) = F_X(x_j) - \lim_{x\to x_j^-} F_X(x)$ *pour tout* $j \in J$.

Démonstration.

- pour tout $x \in \mathbb{R}$,

$$\begin{cases} \text{si } \inf_{j\in J} x_j > -\infty \text{ et } x < \inf_{j\in J} x_j & \Longrightarrow \quad F_X(x) = 0 \\ & \text{ou} \\ \exists j_1 \in J,\ x_{j_1} \le x < \min(x_j,\ x_j > x_{j_1}) & \Longrightarrow \quad F_X(x) = \sum_{x_j \le x_{j_1}} \mathbb{P}(X = x_j) \\ & \text{ou} \\ \text{si } \sup_{j\in J} x_j < +\infty \text{ et } x > \sup_{j\in J} x_j & \Longrightarrow \quad F_X(x) = 1 \end{cases}$$

Ce sont les seules valeurs pouvant être prises par F_X.

- On a $F_X(x_j) = \sum_{x_k \le x_j} \mathbb{P}(X = x_k)$, $\lim_{x\to x_j^-} F_X(x) = \sum_{x_k < x_j} \mathbb{P}(X = x_k)$... $\square$

Définition

Soit X une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un intervalle ou une union d'intervalles. S'il existe $f_X : \mathbb{R} \to [0, +\infty[$ telle que $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ $\forall x \in \mathbb{R}$, X est appelée **v.a. continue** *et f_X* **densité de probabilité** *de X.*

Remarque : termes plus précis : v.a. absolument continue et f_X densité par rapport à la mesure de Lebesgue.

Important : $([0,1], \mathcal{B}([0,1]))$ X nombre pris uniformément dans $[0,1]$.

Par exemple $X(\omega) = \omega$ pour tout $\omega \in [0,1]$.

$\mathbb{P}$: même probabilité pour tout intervalle de même taille

$$\Longrightarrow \mathbb{P}(X \in [a,b]) = b - a \text{ pour } 0 \leq a \leq b \leq 1$$

$$\Longrightarrow F_X(x) = x \text{ pour } 0 \leq x \leq 1,\ 0 \text{ si } x \leq 0,\ 1 \text{ si } x \geq 1$$

$$\Longrightarrow f_X(x) = 1 \text{ pour } 0 \leq x \leq 1,\ 0 \text{ si } x \leq 0,\ 0 \text{ si } x \geq 1$$

Propriété

Si X est une v.a. continue de densité de probabilité f_X alors :

- F_X *continue sur* $\mathbb{R}$ *et* $F_X'(x) = f_X(x)$ *pour presque tout* $x \in \mathbb{R}$.
- $\mathbb{P}(a < X \leq b) = \int_a^b f_X(t)\,dt$, *pour tout* $-\infty \leq a < b \leq +\infty$.
- $\mathbb{P}(X = x) = 0$ *pour tout* $x \in \mathbb{R}$.

Démonstration.

- Si x telle que f_X continue en x on définit G une primitive de f_X alors $F_X(x) = \int_{-\infty}^x f_X(t)\,dt = \lim\limits_{u \to -\infty} \left[G(t)\right]_u^x = G(x) - \lim\limits_{u \to -\infty} G(u)$, d'où $F_X'(x) = f_X(x)$.
- $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^a f_X(t)\,dt + \int_a^b f_X(t)\,dt - \int_{-\infty}^a f_X(t)\,dt$ relation de Chasles, donc $\mathbb{P}(a < X \leq b) = \int_a^b f_X(t)\,dt$.
- On a $\mathbb{P}(X = x) = \mathbb{P}(X \leq x) - \mathbb{P}(X < x)$. Si $A_n = \{X \leq x_n\}$ avec (x_n) suite croissante telle que $x_n < x$ $\forall n \in \mathbb{N}$ et $x_n \to x$, alors $A_n \subset A_{n+1}$ et $\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mathbb{P}(X < x)$. Donc $\mathbb{P}(X = x) = F_X(x) - \lim_{n\to\infty} F_X(x_n)$. Comme F_X est continue, $\lim_{n\to\infty} F_X(x_n) = F_X(x)$ soit $\mathbb{P}(X = x) = 0$.

$\square$

Propriété

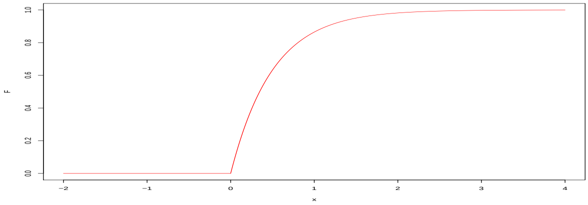
- *Si X v.a. discrète à valeurs dans $I = \{x_j\}_{i \in J}$ alors $\sum_{j \in J} \mathbb{P}(X = x_j) = 1$.*
- *Si X v.a. continue alors $\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$.*

Démonstration.

- Formule des probabilités totales : $\Omega = \bigcup_{i \in J} \{X = x_j\}$ et $(\{X = x_j\})_i$ partition...
- $\lim_{x \to \infty} F_X(x) = 1 = \lim_{x \to \infty} \int_{-\infty}^{x} f_X(t)\, dt$ d'où $\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$

$\square$

Exemple : Pour $f_X(t) = \lambda\, e^{-\lambda t}$ pour $t \geq 0$ et $f_X(t) = 0$ si $t < 0$: **v.a. exponentielle**
$\Longrightarrow F_X(x) = 0$ si $x \leq 0$, $F_X(x) = \int_0^x \lambda\, e^{-\lambda t} dt = 1 - e^{-\lambda x}$ pour $x \geq 0$.



Plan du cours

Moments d'une variable aléatoire

Définition

Si X une v.a. sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans I et $h : I \to \mathbb{R}$ continue par morceaux,

$$\mathbb{E}\big[h(X)\big] = \int_\Omega h(X(\omega))\, d\mathbb{P}(\omega) \qquad \textbf{si} \quad \int_\Omega \big|h(X(\omega))\big|\, d\mathbb{P}(\omega) < \infty.$$

Remarque : Cette formule nécessite des connaissances de L3...

Définition

- *Si X v.a. discrète à valeurs dans $I = \{x_j\}_{i\in J}$), l'espérance de X est*

$$\mathbb{E}[X] = \textstyle\sum_{j\in J} x_j\, \mathbb{P}(X = x_j) \quad \textbf{si} \ \textstyle\sum_{j\in J} |x_j|\, \mathbb{P}(X = x_j) < \infty.$$

- *Si X est une v.a. continue de densité f_X, l'espérance de X est*

$$\mathbb{E}[X] = \textstyle\int_{-\infty}^{+\infty} t\, f_X(t)\, dt \quad \textbf{si} \ \textstyle\int_{-\infty}^{+\infty} |t|\, f_X(t)\, dt < \infty.$$

Exemples (v.a. discrètes) :

1. **Loi de Bernoulli** $\mathcal{B}(p)$ $X : \Omega \to \{0,1\}$, $\left\{ \begin{array}{l} \mathbb{P}(X=1) = p \in [0,1] \\ \mathbb{P}(X=0) = 1-p \end{array} \right.$

$$\Longrightarrow \quad \mathbb{E}[X] = 0 * (1-p) + 1 * p = p.$$

2. **Loi Géométrique** $\mathcal{G}(p)$ $X : \Omega \to \mathbb{N}^*$, $\mathbb{P}(X=k) = p(1-p)^{k-1}$ $k \in \mathbb{N}^*$

$$\Longrightarrow \quad \mathbb{E}[X] = \textstyle\sum_{k=1}^{\infty} k\, p(1-p)^{k-1}.$$

$\Big[$ *Si* $S(x) = \sum_{k=0}^{\infty} x^k = (1-x)^{-1}$ *et* $|x| < 1$, $S'(x) = \sum_{k=1}^{\infty} kx^{k-1} = (1-x)^{-2}$ $\Big]$

$$\Longrightarrow \mathbb{E}[X] = pS'(1-p) = p^{-1} \text{ si } p \neq 0.$$

3. Si $X : \Omega \to \mathbb{N}^*$, $\mathbb{P}(X=k) = \dfrac{1}{k} - \dfrac{1}{k+1} = \dfrac{1}{k(k+1)}$, $k \in \mathbb{N}^*$

$$\Big[\sum_{k=1}^{n} \mathbb{P}(X=k) = (1-\frac{1}{2}) + (\frac{1}{2} - \frac{1}{3}) + (\frac{1}{3} - \frac{1}{4}) + ... = 1 - \frac{1}{n+1} \underset{n \to +\infty}{\longrightarrow} 1 \Big]$$

$$\Longrightarrow \quad \mathbb{E}[X] = \textstyle\sum_{k=1}^{\infty} \frac{1}{k+1} = +\infty.$$

$\Big[$ $\frac{1}{k+1} \geq \ln\left(1 + \frac{1}{k+1}\right) = \ln(k+2) - \ln(k+1)$ *car* $u \geq \ln(1+u)$

d'où $\mathbb{E}[X] \geq (\ln(2) - \ln(1)) + (\ln(3) - \ln(2)) + (\ln(4) - \ln(3)) + \ldots = +\infty$ $\Big]$

Exemples (v.a. continues) :

1. **Loi Uniforme** $\mathcal{U}([0,1])$ $X : \Omega \to [0,1]$, $\begin{cases} f_X(x) = 1 & x \in [0,1] \\ f_X(x) = 0 & \text{sinon} \end{cases}$

$$\Longrightarrow \quad \mathbb{E}[X] = \int_{-\infty}^{\infty} t\, f_X(t)\, dt = \int_0^1 t\, dt = \left[\frac{1}{2}\, t^2\right]_0^1 = \frac{1}{2}.$$

2. **Loi Exponentielle** $\mathcal{E}(\lambda)$ $X : \Omega \to [0, \infty[$, $\begin{cases} f_X(x) = \lambda\, e^{-\lambda\, x} & x \geq 0 \\ f_X(x) = 0 & \text{sinon} \end{cases}$

$$\begin{aligned} \Longrightarrow \quad \mathbb{E}[X] &= \int_{-\infty}^{\infty} t\, f_X(t)\, dt = \lambda \int_0^{\infty} t\, e^{-\lambda\, t}\, dt. \\ &= \Big[-t\, e^{-\lambda\, t}\Big]_0^{\infty} + \int_0^{\infty} e^{-\lambda\, t}\, dt \text{ (IPP)} \\ &= 0 - 0 + \Big[-\frac{1}{\lambda}\, e^{-\lambda\, t}\Big]_0^{\infty} = \frac{1}{\lambda}. \end{aligned}$$

3. **Loi de Cauchy** $\mathcal{C}(1)$ $X : \Omega \to \mathbb{R}$, $f_X(x) = \frac{1}{\pi}\, \frac{1}{1+x^2}$ pour $x \in \mathbb{R}$

$$\Longrightarrow \quad \mathbb{E}[X] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{t}{1+t^2}\, dt \text{ n'existe pas}$$

$$\left(\int_0^{\infty} \frac{1}{\pi}\, \frac{t}{1+t^2}\, dt = \left[\tfrac{1}{2} \ln(1+t^2)\right]_0^{\infty} = \infty \right)$$