UNIVERSITÉ DE GENÈVE

Introduction à la Statistique

Concepts Clés en Probabilité

Stéphane Guerrier, Mucyo Karemera, Samuel Orso & Lionel Voirol

Data Analytics Lab

Contenu du cours

Ce cours propose une introduction aux notions fondamentales de la statistique.

Thèmes abordés :

- Règles de probabilité élémentaires
- Variables aléatoires discrètes et continues
- Estimateurs
- Intervalles de confiance
- Tests d'hypothèses
- Régression linéaire

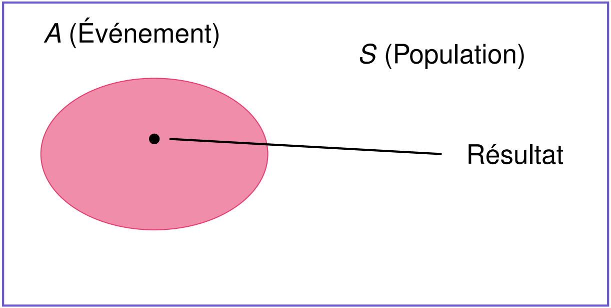
Statistique

Le contenu de ce cours se compose de trois parties :

1. Probabilité
   - Nous utilisons les règles de probabilités élémentaires afin de quantifier la probabilité associée à certains événements.
2. Statistiques descriptives
   - Nous résumons et visualisons un échantillon de données.
3. Statistiques inférentielles
   - Nous utilisons des modèles statistiques pour tirer des conclusions à partir d'un échantillon de données.

Modélisation des phénomènes aléatoires

- Résultats : résultats possibles d'une expérience.
- Population : ensemble de tous les résultats possibles (noté S).
- Événement : ensemble de résultats (par exemple, A dans l'illustration).



Exemple : *Lancer d'un dé équilibré*

- Univers :

$$S = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$$

- Exemples d'événements :

$$\begin{aligned} I &= \text{“\textit{obtenir un nombre impair}”} = \{1, 3, 5\} \\ P &= \text{“\textit{obtenir un nombre pair}”} = \{2, 4, 6\} \\ A &= \text{“\textit{obtenir un 5 et un 6}”} = \{5, 6\} \\ C &= \text{“\textit{obtenir un 5}”} = \{5\} \end{aligned}$$

Exemple : *Lancer d'un dé équilibré*

Si le dé est *équilibré*, alors la **probabilité d'obtenir chaque face est la même**, c'est-à-dire

Probabilité (un) = Probabilité (deux) = Probabilité (trois) =

Probabilité (quatre) = Probabilité (cinq) = Probabilité (six)

ou, en d'autres termes,

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\})$$

Exemple : *Lancer d'une pièce équilibrée*

- Univers :

$$S = \{F, P\}$$

- Si la pièce est *équilibrée*, alors la **probabilité d'obtenir Face ou Pile est la même**, c'est-à-dire

$$\text{Probabilité (Face)} = \text{Probabilité (Pile)},$$

ou, en d'autres termes,

$$\mathbb{P}(\{F\}) = \mathbb{P}(\{P\})$$

Probabilité

Si tous les résultats sont *équiprobables*, c'est-à-dire qu'ils ont la même probabilité de se produire

$$\mathbb{P}(\text{Événement}) = \frac{\#\text{ résultats dans l'événement}}{\text{Nombre total de résultats}}$$

Exemple : *Lancer d'un dé équilibré*

1. Quelle est la probabilité d'obtenir un 5 ?
2. Quelle est la probabilité que le dé tombe sur un nombre impair ?

Nous pouvons utiliser la *définition classique de la probabilité*, car tous les nombres d'un dé ont la même probabilité d'apparaître, puisque c'est un dé équilibré.

Exemple : *Lancer d'un dé équilibré*

(i) Déterminer l'univers, qui sont les résultats possibles d'un tirage de dé :

$$S = \{1, 2, 3, 4, 5, 6\}$$

(ii) Déterminer les événements qui nous intéressent :

$$C = \text{“}\textit{obtenir un 5}\text{”} = \{5\}$$

$$I = \text{“}\textit{obtenir un nombre impair}\text{”} = \{1, 3, 5\}$$

(iii) Calculer la probabilité en utilisant la définition :

$$\mathbb{P}(C) = \frac{\text{\# de résultat dans } C}{\text{Nombre total de résultats possibles}} = \frac{1}{6}$$

$$\mathbb{P}(I) = \frac{\text{\# de résultats dans } I}{\text{Nombre total de résultats possibles}} = \frac{3}{6} = \frac{1}{2}$$

Exemple : *Lancer une pièce équilibrée trois fois*

Probabilité d'obtenir 3 faces :

$$\mathbb{P}(\{FFF\}) = \frac{1}{8} = 0.125$$

Résultats :

FFF	FFP	FPF	FPP
PFF	PFP	PPF	PPP

Exemple : *Lancer une pièce équilibrée trois fois*

Probabilité d'obtenir 2 faces :

$$D = \text{“}\textit{obtenir deux faces}\text{”} = \{PFF, FFP, FPF\}$$

$$\mathbb{P}(D) = \frac{3}{8} = 0.375$$

Résultats :

FFF	FFP	FPF	FPP
PFF	PFP	PPF	PPP

Exemple : *Lancer une pièce équilibrée trois fois*

Probabilité d'obtenir au moins 2 faces :

A = "*obtenir au moins deux faces*" = {FFF, PFF, FFP, FPF}

$$\mathbb{P}(A) = \frac{4}{8} = 0.5$$

Résultats :

FFF	FFP	FPF	FPP
PFF	PFP	PPF	PPP

Propriétés des probabilités

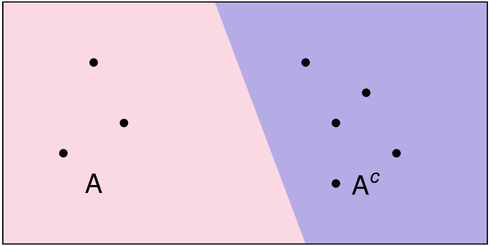
1. $0 \leq \mathbb{P}(A) \leq 1,$ pour tout événement A.

2. $\mathbb{P}(S) = 1$.
   Un événement qui a une probabilité de 1 est appelé *certain*.

3. $\mathbb{P}(A) = 0$.
   Un événement qui a une probabilité de 0 est appelé *impossible*.

Complément/Événements opposés

$$\mathbb{P}(A) = 1 - \mathbb{P}(\text{opposé de } A) = 1 - \mathbb{P}(A^c)$$



Exemple : Lancer un dé équilibré

$$A = \textit{“obtenir un nombre impair”} = \{1, 3, 5\}$$

$$A^c = \textit{“NE PAS obtenir un nombre impair”} = \textit{“obtenir un nombre pair”} = \{2, 4, 6\}$$

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

$$\mathbb{P}(\textit{“nombre impair”}) = 1 - \mathbb{P}(\textit{“nombre pair”})$$

Événements mutuellement exclusifs

- Lorsque les événements A et B (notés par $A \cap B$) ne peuvent pas se produire simultanément (noté par $A \cap B = \emptyset$), alors ils sont appelés *événements mutuellement exclusifs* .

Exemples :

1. Élections :

$$A = \text{“Harris élue présidente”}$$

$$B = \text{“Trump élu président”}$$

2. Lancer une pièce

$$A = \text{“obtenir face”} = \{F\}$$

$$B = \text{“obtenir pile”} = \{P\}$$

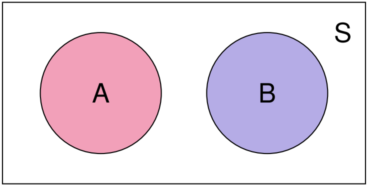
3. Lancer un dé équilibré

$$A = \text{“obtenir un nombre impair”} = \{1, 3, 5\}$$
$$B = \text{“obtenir un nombre pair”} = \{2, 4, 6\}$$

Événements mutuellement exclusifs

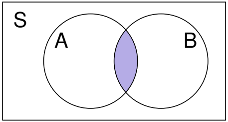
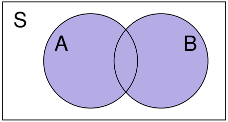
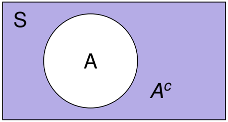
Si A et B ne peuvent pas se produire simultanément (*) :

$$\mathbb{P}(A \text{ OU } B) = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$



(*) A et B n'ont aucun élément en commun

Ensembles

Ensemble	Notation	Interprétation	Diagramme
A et B	$A \cap B$	Résultats dans A et B	
A ou B	$A \cup B$	Résultats dans A ou B, ou les deux	
Complément de A	A^c	Résultats qui ne sont pas dans A	

Probabilité Conditionnelle

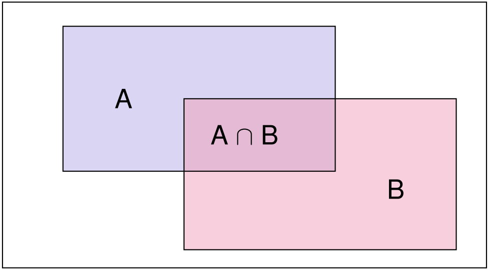
Lancez un *dé équilibré* et quelqu'un nous dit que *le résultat du dé est un nombre impair*.

1. Quelle est la probabilité que le résultat soit un cinq *étant donné que le résultat du dé est un nombre impair* ?
2. Et la probabilité d'obtenir un 6 *étant donné que le résultat du dé est un nombre impair* ?

Probabilité Conditionnelle

Probabilité conditionnelle de l'événement B étant donné que l'événement A est survenu :

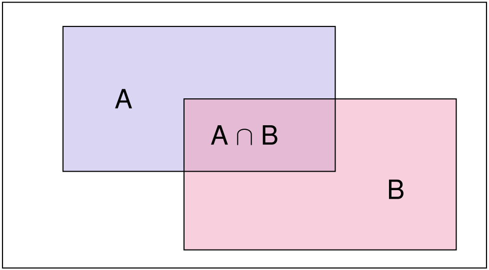
$$\mathbb{P}(B|A) = \text{\textit{“la probabilité de B sachant A”}} = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(A)}$$



Probabilité Conditionnelle

Probabilité conditionnelle de l'événement B étant donné que l'événement A est survenu. Si les résultats sont équiprobables :

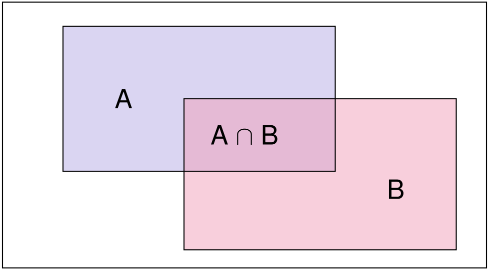
$$\begin{aligned}\mathbb{P}(B|A) &= \text{“\textit{la probabilité de B sachant A}”} \\ &= \frac{[\#\text{ de résultats dans } A \text{ et } B]/[\text{Nombre total de résultats}]}{[\#\text{ de résultats dans } A]/[\text{Nombre total de résultats}]}\end{aligned}$$



Probabilité Conditionnelle

Probabilité conditionnelle de l'événement B étant donné que l'événement A est survenu :

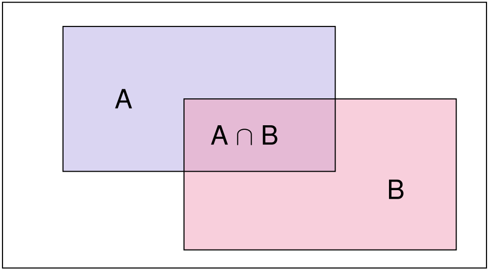
$$\begin{aligned}\mathbb{P}(B|A) &= \textit{la probabilité de } B \textit{ sachant } A \\ &= \frac{\text{\# de résultats dans } A \text{ et } B}{\text{\# de résultats dans } A}\end{aligned}$$



Probabilité Conditionnelle

La probabilité conditionnelle de l'événement A sachant que l'événement B s'est produit :

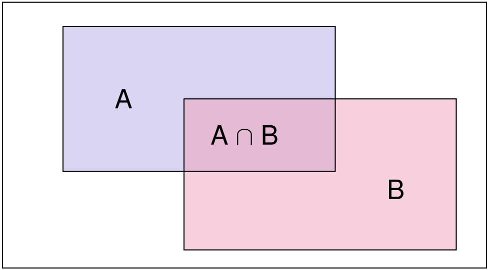
$$\mathbb{P}(A|B) = \textit{la probabilité de } A \textit{ sachant } B = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)}$$



Probabilité Conditionnelle

La probabilité conditionnelle de l'événement A sachant que l'événement B s'est produit. Si les résultats sont équiprobables :

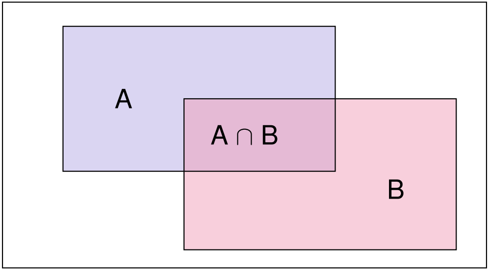
$$\mathbb{P}(A|B) = \textit{la probabilité de A sachant B}$$
$$= \frac{[\#\text{ de résultats dans } A \text{ et } B]/[\#\text{ total de résultats}]}{[\#\text{ de résultats dans } B]/[\#\text{ total de résultats}]}$$



Probabilité Conditionnelle

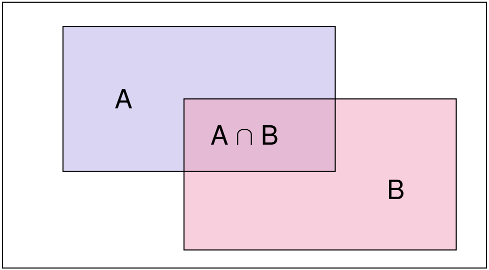
La probabilité conditionnelle de l'événement A sachant que l'événement B s'est produit :

$$\mathbb{P}(A|B) = \textit{la probabilité de } A \textit{ sachant } B$$
$$= \frac{\text{\# de résultats dans } A \text{ et } B}{\text{\# de résultats dans } B}$$



Probabilité jointe

$\mathbb{P}(A \text{ et } B)$ = *probabilité jointe* de A et B
la probabilité que les événements A et B se produisent **ensemble** ($A \cap B$).



Probabilité Conditionnelle

Quelle est la probabilité que le résultat d'un lancer de dé équilibré soit cinq *étant donné que le résultat est un nombre impair* ?

$$A = \text{“}\textit{obtenir un 5}\text{”} = \{5\}$$
$$B = \text{“}\textit{obtenir un nombre impair}\text{”} = \{1, 3, 5\}$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)} = \frac{\#\text{ de résultats dans } A \text{ et } B}{\#\text{ de résultats dans } B} = \frac{1}{3},$$

puisque A et $B = A \cap B = \{5\}$

Autre exemple

Si vous savez qu'une famille a *deux enfants* et que l'un d'eux est un garçon, quelle est la probabilité que l'autre enfant soit également un garçon ?

- Population = ensemble des paires possibles d'enfants

$$S = \{GG, GF, FG, FF\}$$

- La probabilité que les deux enfants soient des garçons étant donné qu'il y a au moins un garçon est

$$\mathbb{P}(\text{ deux garçons } | \text{au moins un garçon}) = \frac{\mathbb{P}(\text{ deux garçons } \textbf{et} \text{ au moins un garçon})}{\mathbb{P}(\text{au moins un garçon})}$$

Exemple

$$\begin{aligned}\mathbb{P}(\text{deux garçons } \textbf{et} \text{ au moins un garçon}) &= \mathbb{P}(\text{deux garçons})\\ &= \frac{\#\text{ de paires avec deux garçcons}}{\#\text{ total de paires}}\\ &= \frac{1}{4}\end{aligned}$$

$$\mathbb{P}(\text{au moins un garçon}) = \frac{\#\text{ de paires avec au moins un garçon}}{\#\text{ total de paires}} = \frac{3}{4}$$

$$\mathbb{P}(\text{deux garçons } |\text{au moins un garçon}) = \frac{1/4}{3/4} = \frac{1}{3}$$

En général, $\mathbb{P}(A|B) \neq \mathbb{P}(B|A)$

- $\mathbb{P}(A|B)$ = probabilité conditionnelle de A sachant B
- $\mathbb{P}(B|A)$ = probabilité conditionnelle de B sachant A

- Exemple :

$$A = \{\text{obtenir un } 5\}$$

$$B = \{\text{obtenir un nombre impair}\}$$

  - $\mathbb{P}(A|B) = \mathbb{P}(\text{obtenir un } 5|\text{obtenir un impair}) = 1/3$
  - $\mathbb{P}(B|A) = \mathbb{P}(\text{obtenir un impair}|\text{obtenir un } 5) = 1$

Règles de Probabilité

1. Probabilité conditionnelle de *A sachant B*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \textbf{ et } B)}{\mathbb{P}(B)}$$

   Probabilité conditionnelle de *B sachant A*

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \textbf{ et } B)}{\mathbb{P}(A)}$$

2. Règle de multiplication

$$\mathbb{P}(A \textbf{ et } B) = \mathbb{P}(B) \cdot \mathbb{P}(A|B)$$

$$\mathbb{P}(A \textbf{ et } B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$$

Règles de Probabilité

3. Règle générale d'addition

$$\mathbb{P}(A \textbf{ ou } B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \textbf{ et } B)$$

4. Probabilité totale

$$\mathbb{P}(A) = \mathbb{P}(A \textbf{ et } B) + \mathbb{P}(A \textbf{ et } B^c)$$

5. Règle du complément

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c)$$

$$\mathbb{P}(A|B) = 1 - \mathbb{P}(A^c|B)$$

Deux types de questions

1. $\mathbb{P}(A \text{ et } B)$
   - Déduire la probabilité conditionnelle $\mathbb{P}(B|A)$.
   - Nous calculons la probabilité jointe : $\mathbb{P}(A \text{ et } B) = \mathbb{P}(B|A)\mathbb{P}(A)$
2. $\mathbb{P}(B|A)$
   - Nous connaissons la probabilité jointe = $\mathbb{P}(A \text{ et } B)$
   - Nous calculons la probabilité conditionnelle : $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(A)}$

Exemple : Le problème de Monty Hall

Exemple : *Test COVID*

- 1 % des étudiants ont le COVID.
- Le test de dépistage est précis à 99 %. (Ainsi, il y a 1 % de chance que le test ne soit pas précis).
- Léa (une étudiante) est testée positive au COVID.

Quelle est la probabilité que Léa ait effectivement le COVID *étant donné qu'elle a été testée positive* ?

Exemple : *Test COVID*

- 1 % des étudiants ont le COVID.
- Le test de dépistage est précis à 99 % (et 1 % non précis).

- *Quelle est la probabilité que Léa ait effectivement le COVID, étant donné qu'elle a été testée positive ?*
  - Nous savons que :

$$\mathbb{P}(\textit{“Léa a le COVID”}) = 0.01$$
$$\mathbb{P}(\textit{“Test positif”} | \textit{“Léa a le COVID”}) = 0.99 \text{ (précis)}$$
$$\mathbb{P}(\textit{“Test positif”} \mid \textit{“Léa N'A PAS le COVID”}) = 0.01 \text{ (non précis)}$$

  - Mais, nous devons calculer

$$\mathbb{P}(\textit{“Léa a le COVID”} \mid \textit{“Test positif”}) = ???$$

Exemple : *Test COVID*

Commençons par la définition

$$\mathbb{P}(\textit{“Léa a le COVID”} \mid \textit{“Test positif”}) = \frac{\mathbb{P}(\textit{“Léa a le COVID”}\ \textbf{et}\ \textit{“Test positif”})}{\mathbb{P}(\textit{“Test positif”})}$$

Utilisons la règle de multiplication :

$$\begin{aligned}\mathbb{P}(\textit{“Léa a le COVID”}\ \textbf{et}\ \textit{“Test positif”}) &= \mathbb{P}(\textit{“Léa a le COVID”}) \times \\ &\quad \mathbb{P}(\textit{“Test positif”} \mid \textit{“Léa a le COVID”}) \\ &= (0.01)\,(0.99) = 0.0099\end{aligned}$$

Exemple : *Test COVID*

1. Utilisez la règle de la probabilité totale :

$$\mathbb{P}(\text{ “\textit{Test positif}”}) = \mathbb{P}(\text{ “\textit{Test positif}” } \textbf{et} \text{ “\textit{Léa a le COVID}”}) + \mathbb{P}(\text{ “\textit{Test positif}” } \textbf{et} \text{ “\textit{Léa N'A PAS le COVID}”})$$

2. Utilisez la règle de multiplication :

$$\mathbb{P}(\text{ “\textit{Test positif}” } \textbf{et} \text{ “\textit{Léa N'A PAS le COVID}”}) = \mathbb{P}(\text{“\textit{Léa N'A PAS le COVID}”})\, \mathbb{P}(\text{ “\textit{Test positif}” } \mid \text{ “\textit{Léa N'A PAS le COVID}”})$$

3. Utilisez la règle du complément :

$$\mathbb{P}(\text{“\textit{Léa N'A PAS le COVID}”}) = 1 - \mathbb{P}(\text{“\textit{Léa a le COVID}”}) = 1 - 0.01 = 0.99$$

Exemple : *Test COVID*

Enfin,

$$\mathbb{P}(\text{ \textit{“Test positif”}}) = 0.0099 + (0.99)(0.01) = 0.0198$$

$$\begin{aligned}\mathbb{P}(\text{\textit{“Léa a le COVID”}} \mid \text{\textit{“Test positif”}}) &= \frac{\mathbb{P}(\text{\textit{“Léa a le COVID”}}\,\textbf{et}\text{ \textit{“Test positif”}})}{\mathbb{P}(\text{ \textit{“Test positif”}})} \\ &= \frac{0.0099}{0.0198} \\ &= \frac{1}{2}\end{aligned}$$