

Statistiques et probabilités

Cours n°2

Guillaume Postic

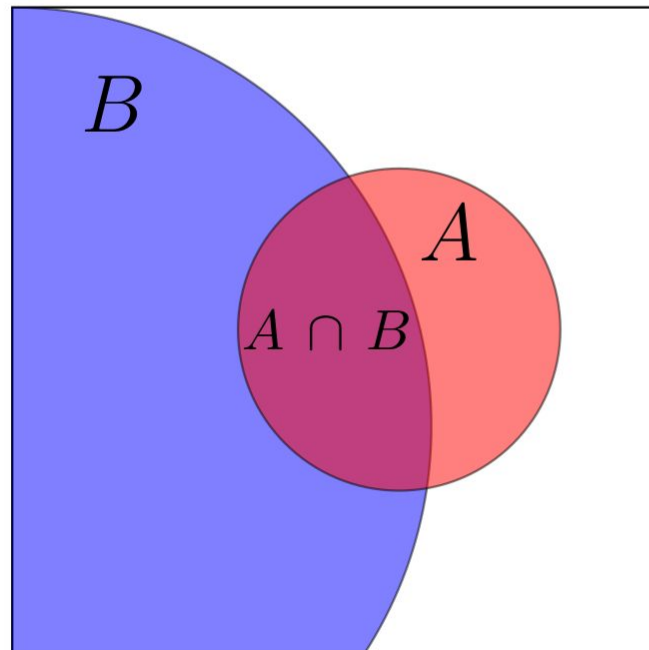
Université Paris-Saclay, Univ. Evry
Département informatique

Master 1 MIAGE - 2023/2024

Probabilité conditionnelle

« La probabilité de A sachant B »

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{avec } P(B) \neq 0$$



$A \cap B$		B			
HHH	HHT	THH	THT		
HTH	HTT	TTH	TTT		

▲ Exemple : pile (T) ou face (H)

◀ Représentation abstraite

Deux principes

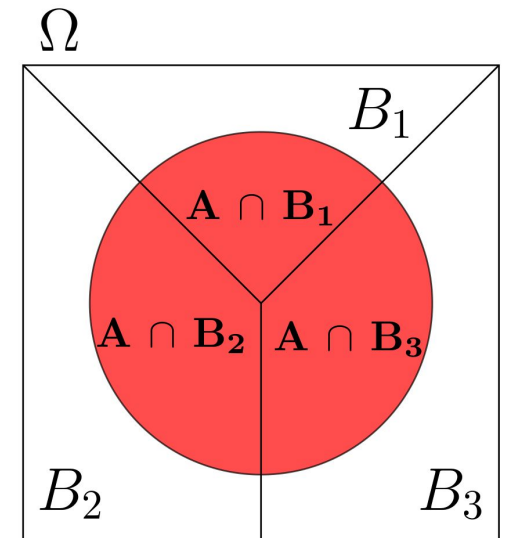
Principe de multiplication

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

Formule des probabilités totales

Si B_1 , B_2 et B_3 forment un **système exhaustif** (ou **partition**) de Ω (incompatibles deux-à-deux et réunion est l'univers tout entier), alors

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) \\ &= P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + P(A | B_3)P(B_3) \end{aligned}$$

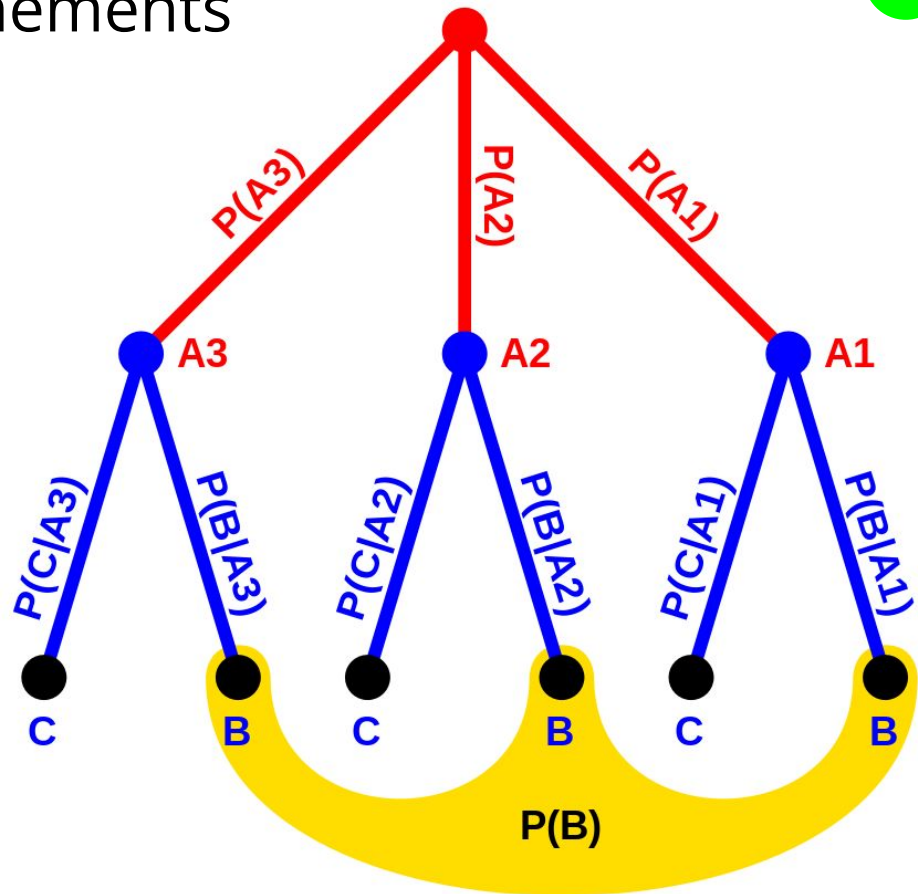


Formule des probabilités totales

(B_n) , système exhaustif d'évènements

$$P(A) = \sum_n P(A \cap B_n)$$

$$P(A) = \sum_n P(A | B_n)P(B_n)$$



Indépendance

Des évènements A et B sont indépendants si la probabilité que l'un survienne n'est pas affectée par la probabilité que l'autre soit survenu.

$$\text{Indépendance} \Leftrightarrow P(A | B) = P(A) \quad (\text{avec } P(B) \neq 0)$$

$$\Leftrightarrow P(B | A) = P(B) \quad (\text{avec } P(A) \neq 0)$$

(pour n'importe quel A et B)

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

Indépendance

Pour deux jets de dés, considérons les évènements suivants :

- $A = \text{« le premier jet fait 3 »}$
- $B = \text{« la somme fait 6 »}$
- $C = \text{« la somme fait 7 »}$

Question : A est indépendant de

- (a) B et C (b) B seulement
(c) C seulement (d) Ni B , ni C

Indépendance

Solution (dé n°1 et dé n°2)

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(A|B) = P(A \cap B) / P(B) = (1/36) / (5/36) = 1/5$$

$$P(A|C) = P(A \cap C) / P(C) = (1/36) / (6/36) = 1/6$$

Note :

$$P(B|A) = P(B \cap A) / P(A) = (1/36) / (1/6) = 1/6$$

$$= P(C|A)$$

Indépendance

Solution (dé n°1 et dé n°2)

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(A) = 1/6$, $P(A | B) = 1/5$. Pas égales, donc pas indépendantes

$P(A) = 1/6$, $P(A | C) = 1/6$. Égales, donc indépendantes

Autre façon :

$P(A \cap B) = 1/36 \neq P(A) P(B) = 1/6 \times 5/36$. Pas indépendantes

$P(A \cap C) = 1/36 = P(A) P(C) = 1/6 \times 6/36$. Indépendantes

Tableau de contingence (tableau croisé)

Représentation de données issues d'un comptage permettant d'estimer la dépendance entre deux caractères.

	Alice	Bob	Charlie	David	Classe
A	4	0	1	2	7
B	1	1	1	2	5
C	0	1	2	0	3
D	0	2	1	0	3
E	0	0	0	1	1
F	0	1	0	0	1
	5	5	5	5	20

- Probabilité **jointe** : $p(A \cap \text{Alice}) = 4/20 = 1/5$
- Probabilité **conditionnelle** :
 - $p(A | \text{Alice}) = p(A \cap \text{Alice}) / p(\text{Alice}) = (4/20) / (5/20) = 4/5$
- Probabilités **marginales** :
 - $p(A) = 7/20$ et $p(\text{Alice}) = 5/20 = 1/4$

Let's Make a Deal, avec Monty Hall (1)

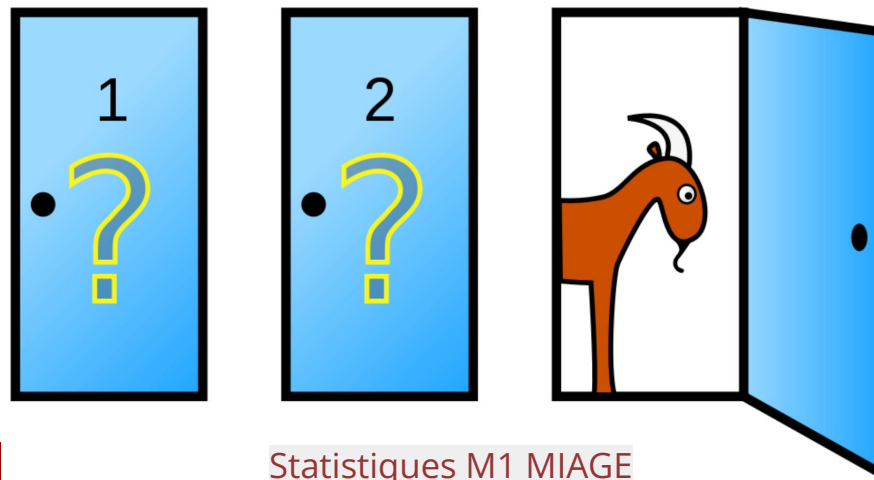
- Trois portes l'une cache une voiture, les deux autres une chèvre
- Le candidat choisit une porte
- Monty Hall ouvre une porte non-choisie et cachant une chèvre (il sait où est la voiture)
- Le candidat est alors autorisé à changer de porte

Quelle est la meilleure stratégie pour gagner ?

(a) Changer

(b) Ne pas changer

(c) Peu importe



Let's Make a Deal, avec Monty Hall (2)

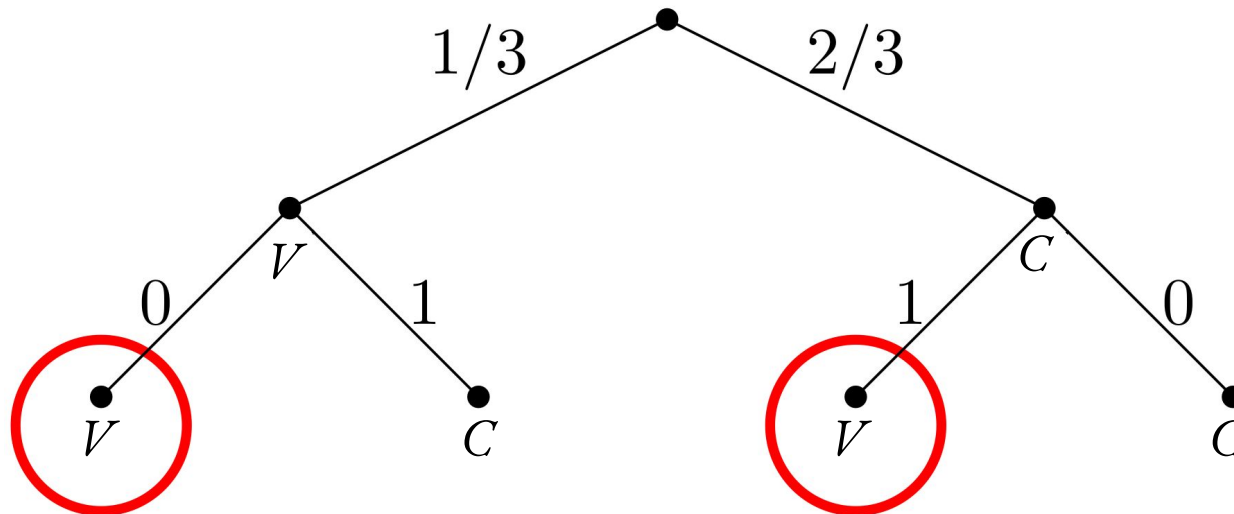
Q : Organisez le problème de Monty Hall sous la forme d'un arbre et calculez la probabilité de gagner si le candidat change de porte choisie quoi qu'il arrive.

$$P(\text{voiture} \mid \text{change}) = ?$$

Let's Make a Deal, avec Monty Hall (2)

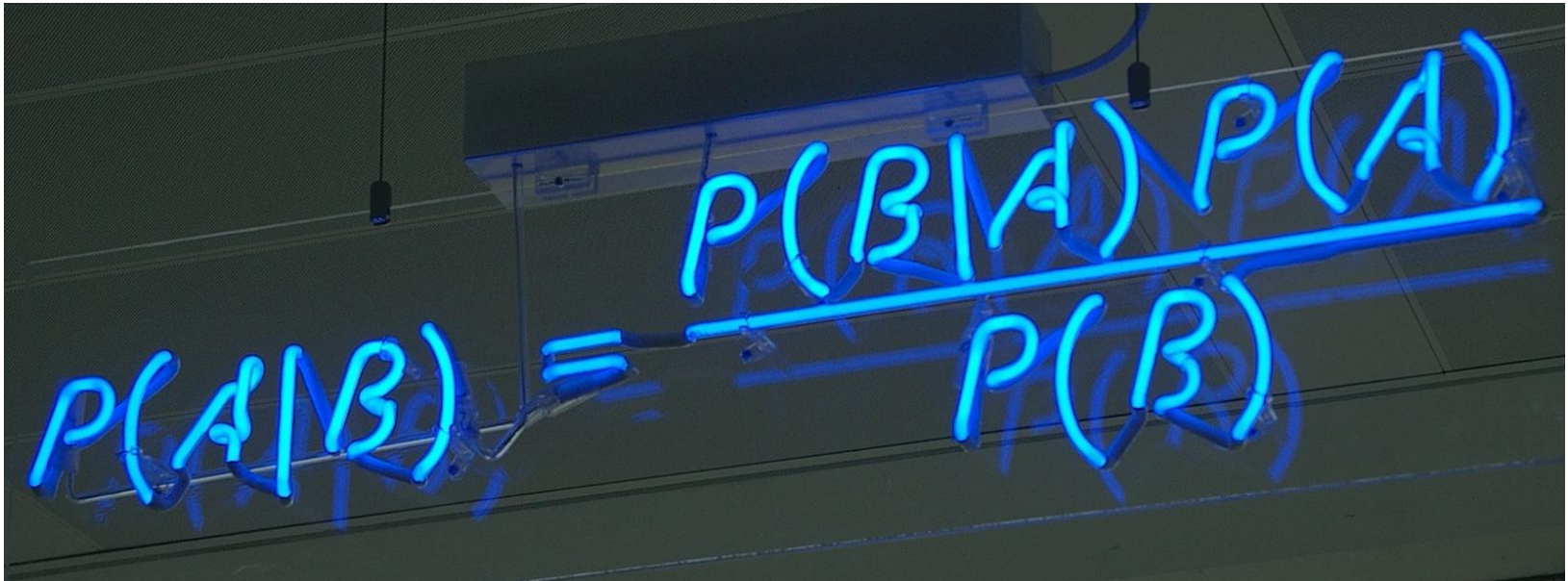
Q : Organisez le problème de Monty Hall sous la forme d'un arbre et calculez la probabilité de gagner si le candidat change de porte choisie quoi qu'il arrive.

$$P(\text{voiture} \mid \text{change}) = ?$$



La probabilité totale de V est $P(V \mid C) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$

Théorème de Bayes (1)


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

- Trouver $P(A|B)$ à partir de $P(B|A)$
- Corollaire
 - du principe de multiplication
 - et de la formule des probabilités totales

Théorème de Bayes (2)

Mise à jour Bayésienne de la probabilité d'une hypothèse H grâce à une donnée (observation) d :

Vraisemblance : si H vraie, probabilité d'observer d

Probabilité a posteriori

$$p(H | d) = \frac{p(d | H)}{p(d)} p(H)$$

Probabilité a priori

Vraisemblance marginale
ou preuve (*evidence*, en anglais)
ou encore « normalisation » (complexité du modèle réduite)

$$p(d) = \sum_i p(d | H_i) p(H_i) = \sum_i p(d \cap H_i)$$

Probabilités totales :
somme pour toutes les hypothèses

$$p(d_1, d_2, \dots, d_n | H) = \prod_i p(d_i | H)$$

Si **indépendance** mutuelle des événements (observations)

Notation

- $p(A, B) = p(A \cap B)$
- $p(A \mid B, C) = p(A \mid B \cap C) = p(A \cap B \cap C) / p(B \cap C)$
- $p(A, B \mid C) = p(A \cap B \mid C) = p(A \cap B \cap C) / p(C)$

Décisions et incertitudes



- Bob est décrit comme aimant la lecture, le calme et le rangement
- Quelle est la profession de Bob ?
 - Bibliothécaire
 - Agriculteur

Décisions et incertitudes



- Bob est décrit comme aimant la lecture, le calme et le rangement
- Quelle est la profession de Bob ?
 - Bibliothécaire
 - Agriculteur
- Il faut considérer la **proportion entre les populations** des deux professions : 20/1 en faveur des agriculteurs
- On met à jour avec l'information la personnalité décrite : la **probabilité de correspondre à cette description** est 4 fois supérieure pour un bibliothécaire

Décisions et incertitudes



- H_1 : être bibliothécaire
- H_2 : être agriculteur
- d : description de la personnalité

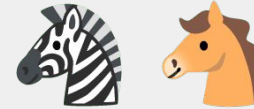
On calcule le rapport des probabilités *a posteriori* entre les deux modèles. Celui-ci implique le **facteur de Bayes** :

$$\frac{p(H_1|d)}{p(H_2|d)} = \boxed{\frac{p(d|H_1)}{p(d|H_2)}} \frac{p(H_1)}{p(H_2)} = \frac{4}{1} \times \frac{1}{20} = \frac{1}{5}$$

Note : inutile de calculer $p(d)$, la vraisemblance marginale

L'hypothèse que Bob soit agriculteur est 5 fois plus probable !

Décisions et incertitudes



*When you hear hoofbeats behind you,
don't expect to see a zebra.*

—Theodore Woodward, professor at the University of Maryland
School of Medicine

Let's Make a Deal, avec Monty Hall (3)

- Supposition : porte n°1 choisie

C (*car*) : porte cachant la voiture ($C = 1, 2, 3$)

H (*Hall*) : porte ouverte par Monty Hall (supposons que $H = 2$)

$$p(H = 2 \mid C = 1) = 0,5$$

$$p(H = 2 \mid C = 2) = 0$$

$$p(H = 2 \mid C = 3) = 1$$

Let's Make a Deal, avec Monty Hall (3)

- Calcul des probabilités voulues :

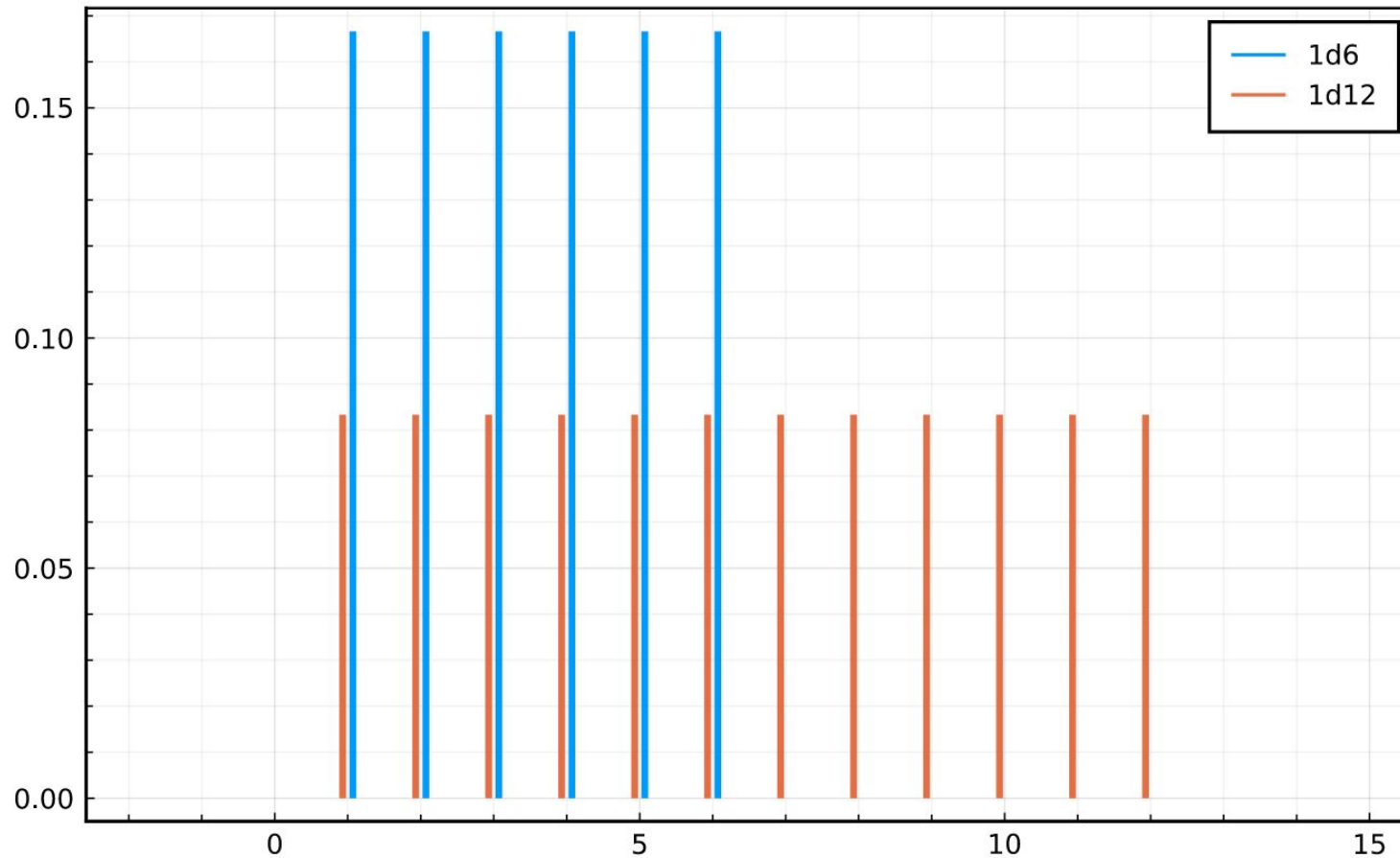
$$p(C = 3 \mid H = 2) = \frac{p(H=2, C=3)}{p(H=2)} = \frac{p(H=2 \mid C=3)p(C=3)}{\sum_{c=1}^3 p(H=2 \mid C=c)p(C=c)}$$

$$p(C = 3 \mid H = 2) = \frac{p(H=2 \mid C=3)p(C=3)}{p(H=2 \mid C=1)p(C=1) + p(H=2 \mid C=2)p(C=2) + p(H=2 \mid C=3)p(C=3)}$$

En supposant que $p(C) = 1/3$, on trouve :

$$p(C = 3 \mid H = 2) = 2/3$$

Lois de probabilité



Définition

- Une loi de probabilité, également appelée **distribution** de probabilité, est une description mathématique qui spécifie **comment les probabilités sont réparties sur l'ensemble des valeurs possibles d'une variable aléatoire**. En d'autres termes, elle donne la probabilité que la variable aléatoire prenne une certaine valeur ou se situe dans un certain intervalle.
- Une loi de probabilité peut être utilisée pour modéliser le comportement aléatoire d'une variable. Elle permet de quantifier les chances associées à différentes valeurs ou résultats possibles de cette variable dans un contexte probabiliste. Il existe de nombreuses lois de probabilité différentes, chacune adaptée à des types de variables et de phénomènes spécifiques.

Principaux descripteurs

1. **Mesures de tendance centrale** : elles incluent la moyenne (ou l'espérance), la médiane et le mode, qui donnent une idée de la valeur centrale des données.
2. **Mesures de dispersion** : elles comprennent la variance, l'écart-type, l'étendue et le coefficient de variation, qui indiquent la variation ou la dispersion des données autour de la valeur centrale.
3. **Graphiques et représentations visuelles** : les histogrammes, les diagrammes en boîte (box plots), les diagrammes en secteurs, les diagrammes à barres et les diagrammes de dispersion sont utilisés pour représenter graphiquement les données.
4. **Mesures de forme** : *skewness* (asymétrie) et *kurtosis* (aplatissement) sont utilisés pour décrire la forme de la distribution des données.

Espérance (mathématique)

Pour X une variable aléatoire prenant les valeurs x_1, x_2, \dots, x_n , l'espérance de X est définie par :

$$E(X) = p(x_1)x_1 + p(x_2)x_2 + \dots + p(x_n)x_n = \sum_{i=1}^n p(x_i) x_i$$

- Calcul d'une **moyenne arithmétique pondérée**
- **Indicateur de tendance centrale**

Propriétés

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX + b) = aE(X) + b$$

$$E(h(X)) = \sum_i h(x_i) p(x_i)$$

Variance (1)

Indicateur de dispersion des valeurs d'un échantillon ou d'une distribution de probabilité. Elle exprime la moyenne des carrés des écarts à la moyenne (**écart quadratique moyen**) :

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

Sa racine carrée définit l'**écart type** σ : $\sigma^2 = V(X)$

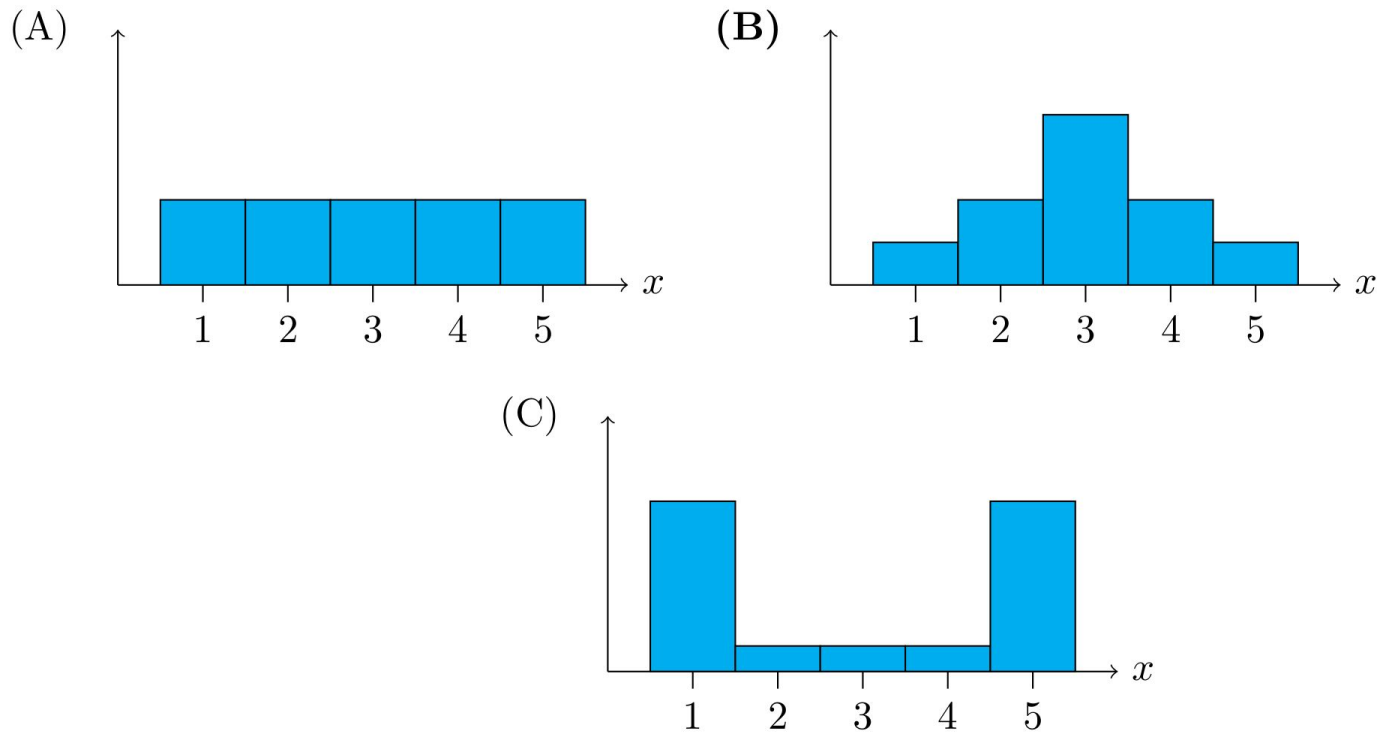
Elle est aussi égale à la **différence entre la moyenne des carrés** des valeurs de la variable et le **carré de la moyenne** : $V(X) = E(X^2) - E(X)^2$

Propriétés :

- $V(a+bX) = b^2 V(X)$
- $V(X+Y) = V(X-Y) = V(X) + V(Y)$, si X et Y indépendants

Variance (2)

La figure ci-dessous donne les fonctions de masse de 3 variables aléatoires. Classez-les par valeur d'écart-type, du plus grand au plus petit (on supposera que x a toujours la même unité).



Variance (3)

Remarque : si la variance d'une variable X est nulle ($V(X) = 0$), alors X est une constante.