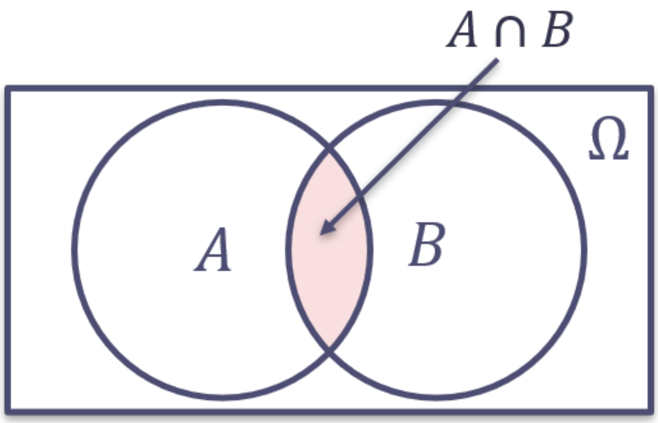
Probabilités conditionnelles

Rappel

Probabilité conditionnelle :

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



Formule de Bayes :

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

Indépendance mutuelle

Les événements $A_1, ..., A_n$ sont **mutuellement indépendants**, si $\forall k \in \{1, ..., n\}$ et $\forall (i_1, ..., i_k) \in \mathcal{N}^k\}$ tel que $1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq n$, on a :

$$P(A_{i_1} \cap ... \cap A_{i_k}) = P(A_{i_1}) \times ... \times P(A_{i_k})$$

c.à.d. pour le cas de 3 événements A, B, C, les conditions suivantes doivent être satisfaites :

1.indépendance 2 à 2 :

- $P(A \cap B) = P(A) \times P(B)$
- $P(A \cap C) = P(A) \times P(C)$
- $P(B \cap C) = P(B) \times P(C)$

2. $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

Remarque : des événements peuvent être 2 à 2 indépendants sans être mutuellement indépendants

Exercice 1

Une famille a deux enfants (consécutivement). Quelle est la probabilité que ce soient deux garçons, sachant que le premier est un garçon ? Quelle est la probabilité que ce soient deux garçons sachant que l'un des deux au moins est un garçon ?

Solution

Nous considérons que les issus possibles sont : $\Omega = \{GG, GF, FG, FF\}$, où GF siginifie que le 1er enfant est un garçon (G) et le 2ème enfant est une fille (F).

Nous pouvons définir le tableau suivant des options possibles :

enfant 1 \ enfant 2	F	G
F	FF	FG
G	GF	GG

Remarque : comme nous avons fait pour les dés

Deux garçons sachant que le 1er est un garçon En sachant que le 1er enfant parmi 2 soit un garçon, les possibilités sont : GF et GG. Donc, la probabilité d'avoir deux garçons a une chance sur deux : $P(GG) = 1/2 = 0.5$.

enfant 1 \ enfant 2	F	G
F	FF	FG
G	GF	GG

Une autre façon de représenter :

Nous pouvons représenter les possibilités sous forme d'un arbre suivant (G = garçon, F = fille) :

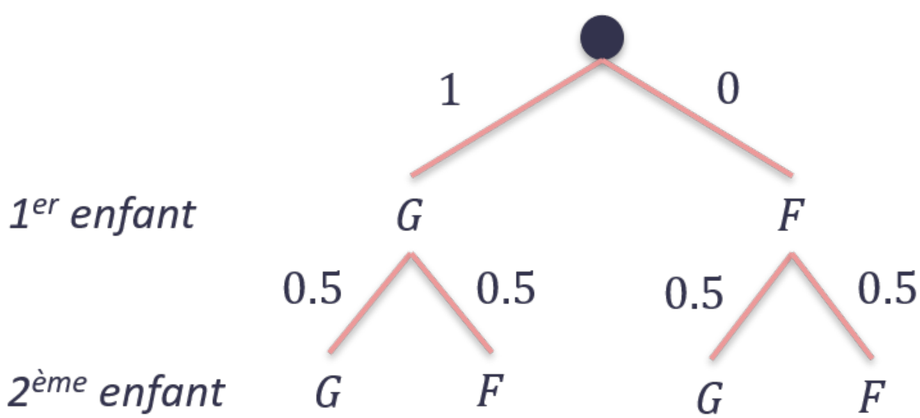


Figure 1: Fig.2. Cas 1: Graphe avec les probabilités

Lorsque nous savons que le 1er enfant est un garçon, la probabilité sur cette branche est 1.

Pour le 2ème enfant, le choix qui reste est entre G (garçon) ou F (fille) avec la même probabilité de 1/2.

Le cas recherché est GG :

Multiplions les probabilités sur la brache qui mène à GG : $P(GG|1er = G) = 0.5 \times 1 = 0.5$

Deux garçons sachant que l'un des deux au moins est un garçon En sachant qu'au moins 1 est un garçon, les possibilités pour les deux enfants sont : GG, GF, FG. Donc, $P(GG) = 1/3$.

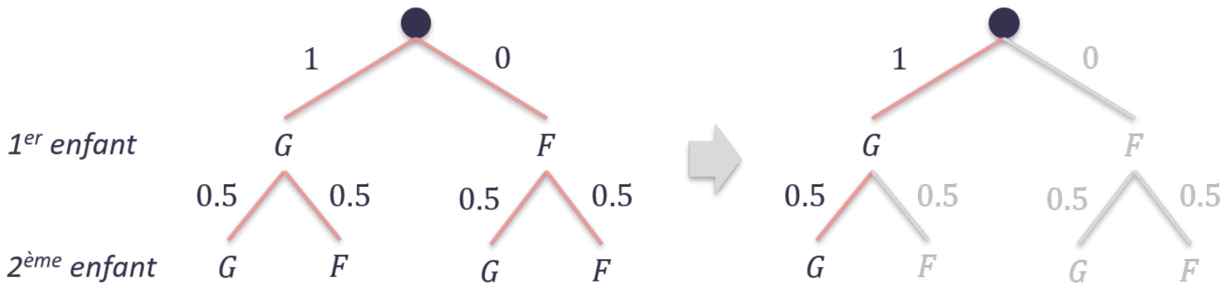


Figure 2: Fig.3. Cas 1: Graphe avec les probabilités pour GG

enfant 1 \ enfant 2	F	G
F	FF	FG
G	GF	GG

En écrivant la formule de Bayes :

$$P(GG|\geq 1G) = \frac{P(\geq 1G|GG) \cdot P(GG)}{P(\geq 1G)}$$

Notons que $P(\geq 1G|GG) = 1$ (la probabilité d'avoir au moins 1 garçon en sachant qu'ils sont deux).

$P(\geq 1G)$ correspond aux cas suivants : GG, GF, FG parmi 4 possibles (pour rappel $\Omega = \{GG, GF, FG, FF\}$). Donc, $P(\geq 1G) = 3/4$.

$P(GG) = 1/4$ car correspond à un cas parmi 4 possibles.

Donc :

$$P(GG|\geq 1G) = \frac{P(\geq 1G|GG) \cdot P(GG)}{P(\geq 1G)} = \frac{1 \cdot 1/4}{3/4} = 1/3$$

Graphiquement, nous pourrions représenter la solution de la façon suivante :

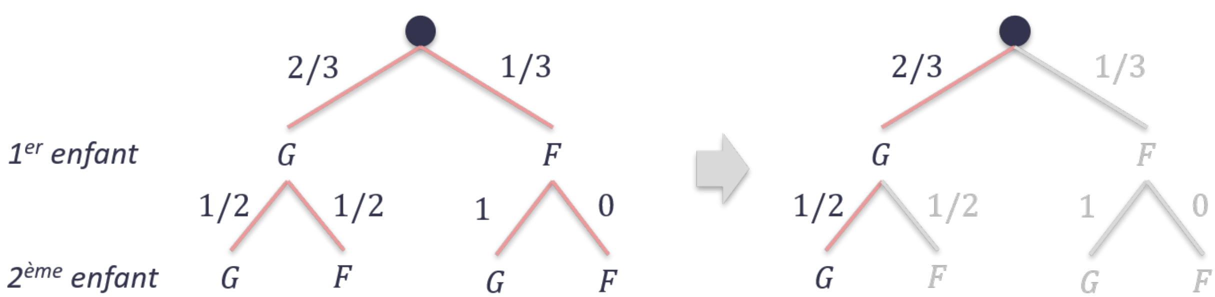


Figure 3: Fig.4. Cas 2 : Graphe avec les probabilités

La probabilité 2/3 sur la branche G pour le 1er enfant vient du fait que cette branche donne 2 possibilités sur 3 pour satisfaire la condition "au moins 1 garçon".

Multiplions les probabilités de la brache qui amène à GG : $P(GG|\geq 1G) = 2/3 \cdot 1/2 = 1/3$.

Exercice 2

Soient A, B et C les événements correspondant au lancer de deux pièces équilibrées et distinguables suivants :

- $A =$ "*La première pièce est tombée sur pile*",
- $B =$ "*La deuxième pièce est tombée sur pile*",
- $C =$ "*Les deux pièces sont tombées sur des faces différentes*".

Montrer que A, B et C sont deux à deux indépendants. Sont-ils (mutuellement) indépendants ?

Solution

Nous pouvons ainsi définir l'univers comme suit : $\Omega = \{PP, PF, FP, FF\}$, où PF correspond au cas où la 1ère îèce est tombée sur pile (P), et la 2ème sur face (F).

Notons que le nombre de résultas (issus) possibles est 4.

Nous pouvons ré-écrire les événements de la façon ensembliste comme suit :

- $A =$ "*La première pièce est tombée sur pile*" $= \{PP, PF\}$,
- $B =$ "*La deuxième pièce est tombée sur pile*" $= \{FP, PP\}$,
- $C =$ "*Les deux pièces sont tombées sur des faces différentes*" $= \{PF, FP\}$.

Calculons les probabilités de chaque événement :

- $P(A) = 2/4 = 1/2$
- $P(B) = 2/4 = 1/2$
- $P(C) = 2/4 = 1/2$

Indépendance 2 à 2 Pour vérifier l'indépendance 2 à 2, nous pouvons nous servir de la formule de probabilités conditionnelles. Aisi, deux événements A et B sont indépendants si

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

.

Calculons les probabilités des intersections :

- $A \cap B = \{PP\} \Rightarrow P(A \cap B) = 1/4$
- $A \cap C = \{PF\} \Rightarrow P(A \cap C) = 1/4$
- $B \cap C = \{FP\} \Rightarrow P(B \cap C) = 1/4$

Reprenons la formule :

- $1/4 = P(A \cap B) = P(A) \cdot P(B) = 1/2 \times 1/2 = 1/4$
- $1/4 = P(A \cap C) = P(A) \cdot P(C) = 1/2 \times 1/2 = 1/4$
- $1/4 = P(B \cap C) = P(B) \cdot P(C) = 1/2 \times 1/2 = 1/4$

Ainsi nous pouvons déduire l'indépendance deux à deux de ces trois événements.

Mutuelle Indépendance Pour la mutuelle indépendance, les conditions suivantes doivent être satisfaites :

1.indépendance 2 à 2 :

- $P(A \cap B) = P(A) \times P(B)$
- $P(A \cap C) = P(A) \times P(C)$
- $P(B \cap C) = P(B) \times P(C)$

2. $P(A \cap B \cap C) = P(A) \times P(B) \times P(C)$

La condition 1 est satisfaite (voir plus haut) car les événement A, B et C sont indépendants deux à deux. Il reste de vérifier la condition 2.

Calculons la probabilité de l'intersection $P(A \cap B \cap C)$:

$$A \cap B \cap C = \emptyset \Rightarrow P(A \cap B \cap C) = 0$$

$$0 = P(A \cap B \cap C) \neq P(A) \times P(B) \times P(C) = 1/2 \times 1/2 \times 1/2 = 1/8$$

Donc, les événements A, B et C ne sont pas mutuellement indépendants.

Remarque : Notons que si la pièce n'était pas équilibrée (on obtient "pile" avec une probabilité p, où $p \neq 1/2$, C n'aurait été indépendant ni de A ni de B.

Exercice 3

Un entrepôt est muni d'un dispositif d'alarme. Lorsqu'il y a tentative de cambriolage, le dispositif se déclenche avec une probabilité égale à 0.99. Lorsqu'il n'y a pas de tentative de cambriolage, le dispositif se déclenche tout de même par erreur au cours d'une journée avec une probabilité égale à 0.01. En supposant qu'une tentative de cambriolage au cours d'une journée ait lieu avec une probabilité égale à 0.001, quelle est la probabilité qu'une alarme déclenchée un jour donné le soit par une tentative de cambriolage ?

On notera D l'événement "$l'alarme\ se\ déclenche$" et T l'événement "$il\ y\ a\ une\ tentative\ de\ cambriolage$", le tout considéré au cours de la journée donnée.

Solution

$D =$ "$l'alarme\ se\ déclenche$"

$T =$ "$il\ y\ a\ une\ tentative\ de\ cambriolage$"

Nous savons les probabilités suivantes :

1. Lorsqu'il y a tentative de cambriolage, le dispositif se déclenche avec une probabilité égale à 0.99 : $P(D|T) = 0.99$
2. Lorsqu'il n'y a pas de tentative de cambriolage, le dispositif se déclenche avec une probabilité égale à 0.01 : $P(D|\bar{T}) = 0.01$
3. Une tentative de cambriolage au cours d'une journée ait lieu avec une probabilité égale à 0.001 : $P(T) = 0.001$

Nous cherchons : $P(T|D) =?$

Nous pouvons utiliser la formule de Bayes :

$$P(T|D) = \frac{P(D|T) \cdot P(T)}{P(D)}$$

Il nous faut donc trouver la probabilité totale de l'événement D, $P(D)$.

Afin de la trouver, nous pouvons représenter le cas sous forme de l'arbre suivant :

Figure 4: Fig.5. Déclenchement de l'alarme

Sachant que $P(\bar{A}) = 1 - P(A)$, nous pouvons trouver $P(\bar{T}) = 1 - P(T) = 1 - 0.001 = 0.999$.

La probabilité totale de D est alors donnée par :

$$P(D) = P(D|T) \cdot P(T) + P(D|\bar{T}) \cdot P(\bar{T}) = \frac{99}{100} \cdot \frac{1}{1000} + \frac{1}{100} \cdot \frac{999}{1000} = \frac{99}{10000} + \frac{999}{10000} = \frac{1098}{10000}$$

Retournos à la formule de Bayes :

$$P(T|D) = \frac{P(D|T) \cdot P(T)}{P(D)} = \frac{\frac{1}{1000} \cdot \frac{99}{100}}{\frac{1098}{10000}} = \frac{\frac{99}{10000}}{\frac{1098}{10000}} = \frac{99}{1098} = 0.09016393 \approx 0.09$$

Exercice 4

On s'intéresse au report des voix dans le cadre du deuxième tour d'une élection à scrutin majoritaire à deux tours.

Supposons qu'il y avait 4 candidats au premier tour, A, B, C et D. Les scores des candidats au premier tour sont 35% pour A, 25% pour B, 20% pour C et 20% pour D parmi les suffrages exprimés. 23% des votants se sont abstenus.

Les candidats C et D ont été éliminés lors du premier tour de l'élection. Parmi l'électorat de C du premier tour,

- 75% voteront pour A au second tour,
- 20% voteront pour B au second tour,
- 5% s'abstiendront.

Parmi l'électorat de D du premier tour,

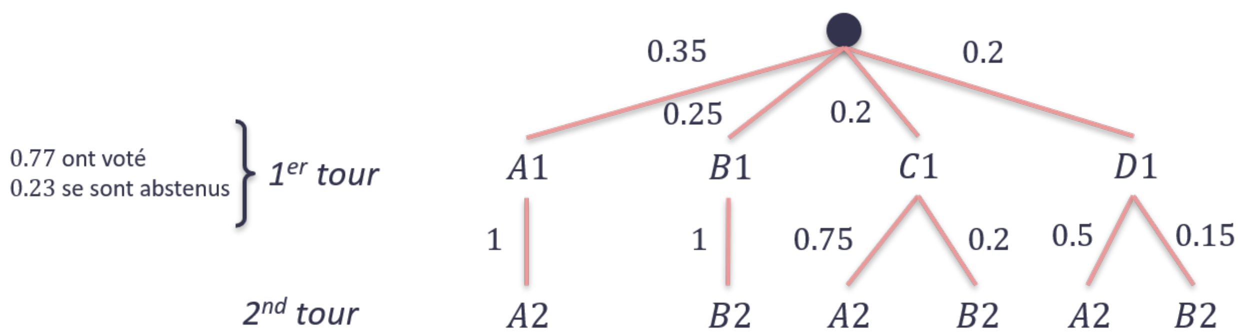
- 50% voteront pour A au second tour,
- 15% voteront pour B au second tour,
- 35% s'abtiendront.

On suppose que les personnes ayant voté pour A et pour B au premier tour ne changeront pas d'avis et ne s'abstiendront pas lors du second tour. On suppose également que les personnes s'étant abstenues au premier tour n'iront pas voter au second.

1. Si je prends un bulletin pour A au second tour, quelle est la probabilité qu'il ait été déposé par quelqu'un ayant voté pour C au premier tour ?
2. Qui va gagner l'élection ?
3. Quel sera le taux d'abstention au second tour ?

Solution

Nous pouvons représenter les données sous forme de l'arbre suivant :



Les chiffres 1 ou 2 sont utilisés pour distinguer les votes au 1er et 2nd tour, c.à.d. $A1$ désigne les votes pour le candidat A au premier tour et $A2$ au second tour.

Probabilité que le vote pour A au 2ème tour vient des gens qui ont voté C au 1er tour Nous recherchons $P(C1|A2)$.

Afin de répondre à cette question, nous pouvons utiliser la formule de Bayes :

$$P(C1|A2) = \frac{P(A2|C1) \cdot P(C1)}{P(A2)}$$

Calculons la probabilité totale de $A2$:

$$P(A2) = P(A2|A1){\cdot}P(A1){+}P(A2|C1){\cdot}P(C1){+}P(A2|D1){\cdot}P(D1) = 1{\cdot}0.35{+}0.75{\cdot}0.2{+}0.5{\cdot}0.2 = 0.35{+}0.15{+}0.1 = 0.6$$

Reprenons la formule de Bayes:

$$P(C1|A2) = \frac{P(A2|C1) \cdot P(C1)}{P(A2)} = \frac{0.75 \cdot 0.2}{0.6} = \frac{0.15}{0.6} = 1/4 = 0.25$$

Qui va gagner l'élection ? Le choix au second tour est entre 2 candidats : A et B.

Selon le calcul précédent, $P(A2) = 0.6$ ce qui repésente la majorité des votes. Ainsi, nous pouvons déduirre que le candidat A va gagner.

Le taux d'abstention au second tour Nous savons qu'au 1er tour le taux d'abstention est 23% et nous supposons également que les personnes s'étant abstenues au premier tour n'iront pas voter au second. (Il faut qu'on inclut ces 23% dans le calcul.)

C'est-à-dire qu'il y a 77% de population qui ont participer au second tour.

Parmi ces 77% de population, nous savons que 5% de ceux qui ont voté C au 1er tour et 35% de ceux qui ont voté D au 1er tour s'abtiendrent.

Ainsi le taux d'abstention au second tour peut être calculer comme suit :

$$taux = 0.23{+}0.77(P(abstention2|C1){\cdot}P(C1){+}P(abstention2|D1){\cdot}P(D1)) = 0.23{+}0.77(0.05{\cdot}0.2{+}0.35{\cdot}0.2) =$$

$$= 0.23 + 0.77(0.01 + 0.07) = 0.23 + 0.77 \cdot 0.08 = 0.23 + 0.0616 = 0.2916$$

Le taux d'abstention au second tour est 29.16%.

Paradoxe de Simpson

Dans un lycée, on a observé les résultats suivants à un examen :

Filière	Littéraire		Scientifique		Total	
Sexe	F	G	F	G	F	G
Succès	20	1	200	600	220	601
Echec	180	19	100	400	280	419
Taux de réussite	$1/10 > 1/20$		$2/3 > 3/5$		$0.44 < 0.59$	

Les filles sont-elles meilleures que les garçons à l'examen ?

Commentaires

Ce paradoxe décrit une situation où un phénomène observé (le taux de réussite à un examen) dans plusieurs groupes peut s'inverser quand on considère les données dans leur ensemble. Ce résultat est lié à des éléments qui ne sont pas pris en compte (e.g. la présence de variables non indépendantes ou de différences d'effectifs entre les groupes, etc.).

Nous pouvons réorganiser le tableau de la manière suivante :

	F	G
Littéraire	1/10 (20/200)	1/20 (1/19)
Scientifique	2/3 (200/300)	3/5 (600/1000)
Total	0.44 (220/500)	0.59 (601/1020)

Il existe les différences entre les groupes qui ne sont pas pris en compte.

Par exemple, notons que les tailles de groupes (le nombre de représentants de chaque groupe) sont très différentes : * 200 F vs. 20 G (littéraire) * 300 F vs. 1,000 G (scientifique) * 500 F vs 1,020 G (total) Les groupes selon la lanière ne sont pas équilibrées au nuveau F:G.

Nous notons que le nombre total de filles et 2 fois plus petit que le nombre de garçons ce qui joue significativement dans le calcul de pourcentage.

En plus, le résultat total pour les garçons est dominé par la filière "Scientifique" (1,000 vs 19), et un peu moins pour les filles (300 vs. 200).

En regardant les taux selon les filières, nous constatons aussi que les élèves ont rencontré plus de difficulté dans la filière Littéraire.

On remarque ainsi que le contexte est important pour qualifier la notion de succès.