MAT 2777
Probabilités et statistique pour ingénieur.e.s

Chapitre 1
Introduction aux probabilités

P. Boily (uOttawa)

Hiver 2023

Aperçu

1.1 – Les espaces d'échantillonnage et les événements

Nous traiterons d'**expériences aléatoires** (par exemple, mesures de la vitesse/du poids, nombre et durée d'appels téléphoniques, etc.).

L'**espace d'échantillonnage** d'une expérience aléatoire, que l'on désigne par le symbole $\mathcal{S}$, est constitué de **tous** les résultats possibles.

Un espace d'échantillonnage peut être **discret** ou **continu**.

Un **événement** est une collection de résultats d'une expérience aléatoire. Les événements seront désignés par A, B, E_1, E_2, etc.

Exemples:

1. Lancez une pièce de monnaie équilibrée. L'espace d'échantillonnage (discret) est $\mathcal{S} = \{\text{Face, Pile}\}$.

2. Lancez un dé équilibré. L'espace d'échantillonnage (discret) est $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Voici quelques événements :
   - obtenir un nombre pair, $A = \{2, 4, 6\}$.
   - obtenir un nombre premier, $B = \{2, 3, 5\}$.

3. Nous mesurions le poids (en grammes) d'un échantillon de substance chimique. L'espace d'échantillonnage (continu) peut être représenté par $\mathcal{S} = (0, \infty)$. Voici quelques événements :
   - l'échantillon pèse moins de 1,5 g : $A = (0, 1.5)$;
   - l'échantillon pèse plus de 5 g : $B = (5, \infty)$;

Pour tout événement A et B de $\mathcal{S}$:

- l'**union** $A \cup B$ représente tous les résultats de $\mathcal{S}$ se retrouvant **soit dans A, soit dans B (ou dans les deux)** ;
- l'**intersection** $A \cap B$ représente tous les résultats de $\mathcal{S}$ se retrouvant **et dans A, et dans B** ;
- le **complément** A^c (parfois dénoté A^c ou $-A$) représente tous les résultats de $\mathcal{S}$ qui ne se retroube pas dans A ;
- si A et B n'ont aucun résultat en commun, ils sont dits **mutuellement exclusifs**, ce qui est dénoté par $A \cap B = \varnothing$ (l'ensemble vide).

En particulier, A et A^c sont mutuellement exclusifs (diagrammes de Venn).

Exemples:

1. Lancez un dé. Soient $A = \{2, 3, 5\}$ (# premiers) et $B = \{3, 6\}$ (multiples de 3). Ainsi $A \cup B = \{2, 3, 5, 6\}$, $A \cap B = \{3\}$, et $A^c = \{1, 4, 6\}$.

2. On analyse la résistance aux rayures et aux chocs d'échantillons de 100 échantillons de plastique.

		résistance aux chocs	
		élevée	faible
résistance aux rayures	élevée	40	4
	faible	1	55

Si A représente l'événement qu'un échantillon possède une résistance élevée aux chocs et B celui qu'un échantillon possède une résistance élevée aux rayures, alors $A \cap B$ est composé de 40 échantillons.

1.2 – Les techniques de dénombrement

Certains résultats combinatoires fondamentaux rendent certaines probabilités plus faciles à calculer.

On peut modéliser une **procédure en deux temps** en se servant de k sacs, contenant respectivement $m_1, \cdots, m_k$ articles.

La première étape consiste à choisir un des sacs, la seconde à choisir un des articles dans le sac choisi. Cela équivaut à sélectionner l'un des $n = m_1 + m_2 + \cdots + m_k$ articles disponibles.

Si tous les sacs contiennent le même nombre d'articles $(m_i = m)$, alors il y a $n = k \cdot m$ articles au total, et c'est le nombre total de façons dont la procédure peut se produire.

Exemples:

1. De combien de façons peut-on lancer un dé et tirer une carte d'un paquet (mélangé) de 52 cartes ?
   - On peut obtenir 6 résultats en lançant le dé, et pour chacun des résultats, il y a 52 façons de tirer une carte au hasard. Il y a donc $6 \times 52 = 312$ façons de réaliser cette suite d'opérations.

2. De combien de façons peut-on piger d'un sac deux billets numérotés de 1 à 100, le premier de la main droite et le second de la main gauche ?
   - Il y a 100 façons de choisir le premier billet ; pour *chacune de celles-ci* il y a 99 façons de piger le deuxième billet. Ainsi, il y a $100 \times 99 = 9900$ de réaliser cette suite d'opérations.

Procédures à plusieurs étapes

Une **procédure à k étapes** est une procédure pour laquelle :

- il y a n_1 possibilités à l'étape 1 ;
- quel que soit le premier résultat, il y a n_2 possibilités à l'étape 2 ;
- etc. ;
- **indépendamment** des résultats précédents, il y a n_k possibilités à l'étape k.

Il y a donc au total $n = n_1 n_2 \cdots n_k$ façons de réaliser la procédure.

1.3 – Les échantillons ordonnés

Supposons que nous ayons un sac contenant n boules de billard numérotées $1, 2, \ldots, n$. Nous tirons un **échantillon ordonné** de taille r en prélevant des boules au hasard, soit

- **avec remise**, ou
- **sans remise**.

Avec combien de collections différentes de r balles pouvons-nous nous retrouver dans chacun de ces cas (procédures à r étapes) ?

Supposition clé : tous les objets (comme les boules) peuvent être différenciés (à l'aide de chiffres, de couleurs, etc.)

Échantillonnage avec remise (l'ordre est important)

Si nous replaçons chaque boule dans le sac après qu'elle ait été choisie, alors chaque tirage d'une balle suit la même logique (il y a n façons de le faire).

Selon un résultat antérieur, il y a

$$\underbrace{n \times n \times \cdots \times n}_{r \text{ étapes}} = n^r$$

façons de sélectionner un échantillon ordonné de taille r **avec remise** à partir d'un ensemble comportant n objets $\{1, 2, \ldots, n\}$.

Échantillonnage sans remise (l'ordre est important)

Si on **ne remet pas** chaque boule dans le sac après qu'elle ait été tirée (n choix possibles), alors les choix pour le deuxième tirage dépendent du résultat du premier tirage, et il n'y a que $n-1$ résultats possibles.

Quel que soit le résultat des deux premiers tirages, il y a $n-2$ façons de tirer la troisième boule, et ainsi de suite.

Il y a ainsi

$$ {}_nP_r = \underbrace{n \cdot (n-1) \cdots (n-r+1)}_{r \text{ étapes}} \quad \text{(notation fréquemment utilisée)} $$

façons de sélectionner un échantillon ordonné de taille r **sans remise** à partir d'un ensemble comportant n objets $\{1, 2, \ldots, n\}$.

Notation factorielle

Pour un entier positif n, on définit $n! = n \cdot (n-1)(n-2)\cdots 1$:

- lorsque $r = n$, ${}_nP_r = n!$, et la sélection ordonnée est appelée une **permutation** ;
- lorsque $r < n$, on peut écrire

$$ {}_nP_r = n\cdots(n-r+1) = \frac{n(n-1)\cdots(n-r+1)\,(n-r)\cdots 1}{(n-r)\cdots 1} = \frac{n!}{(n-r)!}. $$

Par convention, on pose $0! = 1$, de sorte que ${}_nP_r = \frac{n!}{(n-r)!}$ pour tout $r \leq n$.

Exemples:

1. De combien de façons peut-on tirer 6 boules au hasard provenant d'un sac de boules numérotées de 1 à 49, en tenant compte de l'ordre dans lesquelles elles sont pigées?

   Réponse: ${}_{49}P_6 = 49 \times 48 \times 47 \times 46 \times 45 \times 44 = 10,068,347,520$. C'est le nombre de façons dont le le tirage des boules peut se produire pour le Loto 6/49 en temps réel (boules tirées une par une).

2. Combien de codes NIP à 6 chiffres pouvez-vous créer à partir de l'ensemble de chiffres $\{0, 1, \ldots, 9\}$?

   - Avec répétitions: $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 10^6 = 1,000,000$;
   - Sans répétition: ${}_{10}P_6 = 10 \times 9 \times 8 \times 7 \times 6 \times 5 = 151,200$.

1.4 – Les échantillons non ordonnés

Supposons maintenant que nous ne pouvons pas faire la distinction entre différents échantillons ordonnés ; lorsque nous consultons les résultats du Lotto 6/49 dans le journal, par exemple, **nous n'avons aucun moyen de savoir dans quel ordre les boules ont été tirées**.

$1-2-3-4-5-6$ pourrait signifier que la première boule tirée était la boule $\#\ 1$, la deuxième boule tirée était la boule $\#\ 2$, etc., mais cela pourrait aussi signifier que la première boule tirée était la boule $\#\ 4$, la deuxième était la boule $\#\ 3$, etc., ou toute autre combinaison des 6 premières boules.

Dénotons le nombre (encore inconnu) d'échantillons **non ordonnés de taille** r prélevé à même un ensemble de taille n par ${}_nC_r$.

Nous pouvons dériver l'expression ${}_nC_r$ en notant que les deux processus suivants sont équivalents :

- prélever un échantillon ordonné de taille r (il y a ${}_nP_r$ façons de le faire) ;
- prélever un échantillon non ordonné de taille r (il y a ${}_nC_r$ façons de le faire), **pour ensuite** réorganiser (permuter) les objets de l'échantillon (il y a $r!$ façons de le faire).

Ainsi,

$$ {}_nP_r = {}_nC_r \times r! \quad \Longrightarrow \quad {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!\,r!} = \binom{n}{r}. $$

Cette dernière notation s'appelle le **coefficient binomial** (lu "r parmi n") et est couramment utilisée dans les manuels scolaires.

Exemple:

De combien de façons le tirage du loto 6/49 peut-il être présenté dans le journal (où il est toujours présenté en **ordre croissant**) ?

Solution: c'est le nombre de **échantillons non ordonnés** de taille 6 (différents réarrangements des mêmes 6 nombres sont indiscernables), donc

$$\begin{aligned} {}_{49}C_6 = \binom{49}{6} &= \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{10,068,347,520}{720} \\ &= 13,983,816\,. \end{aligned}$$

Identités de coefficients binomiaux $\Longrightarrow$ page suivante

$\binom{n}{k} = \frac{n!}{k!(n-k)!},\ \ k = 0, 1, 2, \ldots, n$

$\binom{n}{k} = \binom{n}{n-k},\ \ k = 0, 1, 2, \ldots, n$

$\binom{n}{0} < \binom{n}{1} < \cdots < \binom{n}{\lfloor n/2 \rfloor},\ \ n \geq 0$

$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k},\ \ k = 0, 1, 2, \ldots, n$

$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k},\ \ n \geq 0$

$\sum_{k=0}^{n} \binom{n}{k} = 2^n,\ \ n \geq 0$

$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0,\ \ n \geq 0$

$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n},\ \ n \geq 0$

$\left[\sum_{k=0}^{n} \binom{n}{k}\right]^2 = \sum_{k=0}^{2n} \binom{2n}{k},\ \ n \geq 0$

$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1},\ \ k \neq 0$

$\binom{n}{m}\binom{m}{k} = \binom{n}{k}\binom{n-k}{m-k},\ \ 0 \leq k \leq m \leq n$

$\sum_{k=0}^{m} \binom{n+k}{k} = \binom{n+m+1}{m},\ \ m, n \geq 0$

$\sum_{k=0}^{n-m} \binom{m+k}{m} = \binom{n+1}{m+1},\ \ n \geq m \geq 0$

$\sum_{k=0}^{r} \binom{m}{k}\binom{n}{r-k} = \binom{m+n}{r},\ \ m, n, r \geq 0$

$\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n-k}{k} = F_{n+1}$ (Fibonacci numbers), $n \geq 0$

$\sum_{k=0}^{n} k\binom{n}{k} = n2^{n-1},\ \ n \geq 0$

$\sum_{k=0}^{n} k^2\binom{n}{k} = n(n+1)2^{n-2},\ \ n \geq 0$

$\sum_{k=0}^{n} (-1)^k k\binom{n}{k} = 0,\ \ n \geq 0$

$\sum_{k=0}^{n} \frac{\binom{n}{k}}{k+1} = \frac{2^{n+1}-1}{n+1},\ \ n \geq 0$

$\sum_{k=0}^{n} (-1)^k \frac{\binom{n}{k}}{k+1} = \frac{1}{n+1},\ \ n \geq 0$

$\sum_{k=1}^{n} (-1)^{k-1} \frac{\binom{n}{k}}{k} = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n},\ \ n > 0$

$\sum_{k=0}^{n-1} \binom{n}{k}\binom{n}{k+1} = \binom{2n}{n-1},\ \ n > 0$

$\sum_{k=0}^{m} \binom{m}{k}\binom{n}{p+k} = \binom{m+n}{m+p},\ \ m, n, p \geq 0,\ n \geq p + m$

1.5 – La probabilité d'un événement

Dans le cas d'une expérience aléatoire pour laquelle il y a exactement N résultats possibles **mutuellement exclusifs** et **également probables**, nous pouvons attribuer une **probabilité** $P(A)$ à l'événement A en commençant par compter le nombre de résultats correspondant à A.

Si ce nombre est égal à a, alors

$$P(A) = \frac{a}{N}.$$

La probabilité de chaque **résultat individuel** est donc de $1/N$.

Exemples:

1. Lancez une pièce de monnaie équilibrée. L'espace d'échantillonnage (discret) est $\mathcal{S} = \{\text{Face, Pile}\}$, d'où $N = 2$. La probabilité d'observer Pile est $\frac{1}{2}$.

2. Lancez un dé équilibré; il y a $N = 6$. L'espace d'échantillonnage est $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$.

   Si A correspond à un résultat qui est un multiple de 3, alors $A = \{3, 6\}$ et $a = 2$, de sorte que

$$P(\text{résultat est un multiple de } 3) = P(A) = \frac{2}{6} = \frac{1}{3}.$$

3. Lancez un dé équilibré. Les probabilités d'obtenir un nombre pair/un nombre premier sont :

   - $P(\{\text{résultat pair}\}) = P(\{2,4,6\}) = \frac{3}{6} = \frac{1}{2}$.
   - $P(\{\text{résultat premier}\}) = P(\{2,3,5\}) = 1 - P(\{1,4,6\}) = \frac{1}{2}$.

4. Dans un groupe de 1000 personnes, on sait que 545 souffrent de pression artérielle élevée. On choisit au hasard 1 personne. Quelle est la probabilité que cette personne souffre d'hypertension artérielle ?

   Solution: la **fréquence relative** des personnes souffrant d'hypertension artérielle est de 0.545. Avec la définition classique, c'est la probabilité que l'on recherche.

Ce n'est pas la seule approche $\Longrightarrow$ **analyse bayésienne**

Les axiomes de la probabilité

1. Pour tout événement A, $0 \leq P(A) \leq 1$.
2. Pour l'espace d'échantillonnage $\mathcal{S}$, $P(\mathcal{S}) = 1$.
3. Pour l'événement vide $\varnothing$, $P(\varnothing) = 0$.
4. Pour deux événements **mutuellement exclusifs** A et B, la probabilité que A ou B se produisent est $P(A \cup B) = P(A) + P(B)$.

Puisque $\mathcal{S} = A \cup A^c$, et puisque A et A^c sont mutuellement exclusifs, alors

$$1 \overset{\textbf{A2}}{=} P(\mathcal{S}) = P(A \cup A^c) \overset{\textbf{A4}}{=} P(A) + P(A^c) \Longrightarrow P(A^c) = 1 - P(A).$$

Exemples:

1. Lancez un dé équilibré et notez le nombre affiché. Soient A et B les événements où le nombre est un multiple de 3 ou < 3, respectivement. Alors $A = \{3, 6\}$, $B = \{1, 2\}$, et A et B sont mutuellement exclusifs puisque $A \cap B = \varnothing$. Alors

$$P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B) = \frac{2}{6} + \frac{2}{6} = \frac{2}{3}.$$

2. Une urne contient 4 boules blanches, 3 boules rouges, et 1 boule noire. On tire une boule au hasard; on note les événements suivants : $W = \{\text{boule blanche}\}$, $R = \{\text{boule rouge}\}$, et $B = \{\text{boule noire}\}$. Alors

$$P(W) = 1/2, \quad P(R) = 3/8, \quad P(B) = 1/8, \quad P(W \text{ ou } R) = 7/8.$$

Règle générale des additions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Exemple: Un gadget électronique a deux composantes A et B. Par expérience, nous savons que $P(A \text{ fait défaut}) = 0.2$, ainsi que $P(B \text{ fait défaut}) = 0.3$ et que $P(A \text{ et } B \text{ font défaut}) = 0.15$. Trouvez $P(\text{au moins l'une de } A \text{ ou } B \text{ fait défaut})$ et $P(\text{ni } A, \text{ ni } B \text{ ne font défaut})$.

Solution: posons A pour "A fait défaut" et de même pour B. Alors

$$P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.3 - 0.15 = 0.35\,;$$

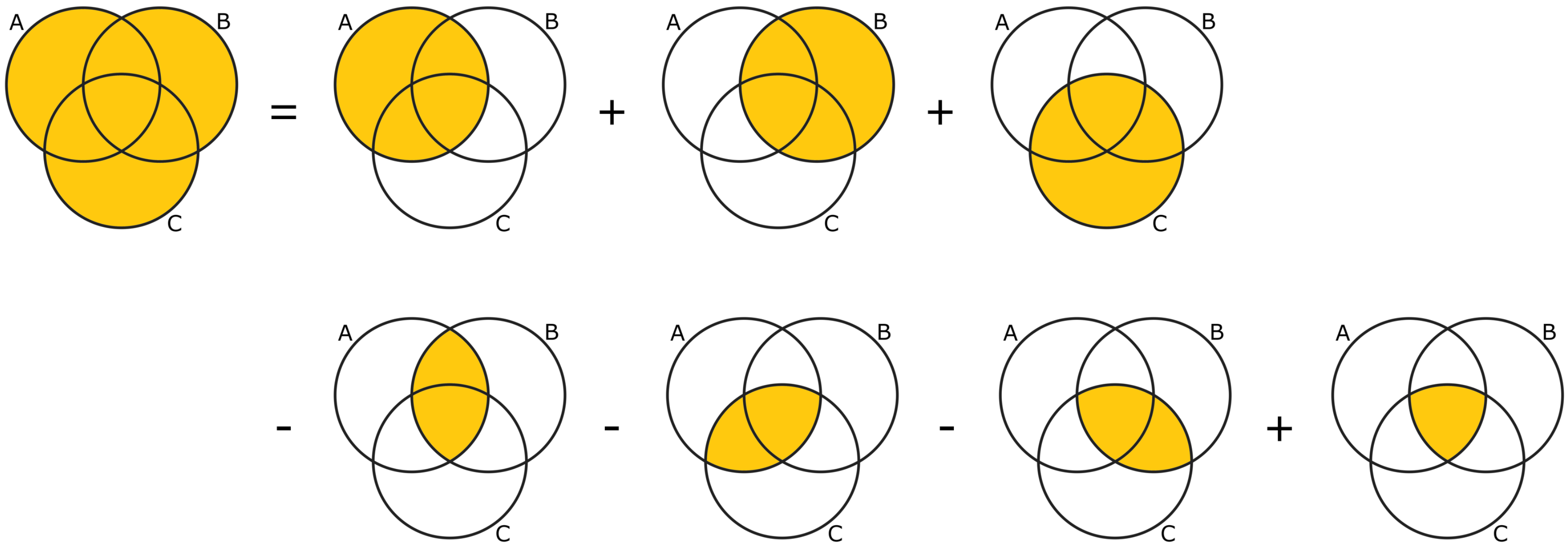
$$P(\text{ni } A, \text{ ni } B) = 1 - P(A \text{ ou } B) = 1 - 0.35 = 0.65\,.$$

Lorsque A et B sont mutuellement exclusifs, $P(A \cap B) = P(\varnothing) = 0$ et

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B).$$

S'il y a plus de deux événements, la règle se développe comme suit :

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$



1.6 – Les probabilités conditionnelles et l'indépendance

Deux évènements quelconques A et B satisfaisant

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

sont dits **indépendants**; il s'agit d'une définition mathématique, mais elle correspond à la notion intuitive d'indépendance dans des cas simples. Lorsque les événements ne sont pas indépendants, on dit qu'ils sont **dépendants** ou **conditionnels**.

Exclusivité mutuelle et indépendance:

$$\begin{aligned} 0 = P(A \cap B) = P(A) \times P(B) &\Longrightarrow P(A) = 0 \text{ ou } P(B) = 0 \\ &\Longrightarrow A = \varnothing \text{ ou } B = \varnothing. \end{aligned}$$

Exemples:

1. Lancez deux fois une pièce de monnaie équilibrée : les 4 résultats possibles sont tous également probables : $\mathcal{S} = \{\text{FF}, \text{FP}, \text{PF}, \text{PP}\}$. Soient A l'événement "Face au premier tour", B l'événement "Face au second tour". On note que $A \cup B \neq \mathcal{S}$ et que $A \cap B = \{\text{FF}\}$.

   Selon la règle générale des additions,

$$\begin{aligned} P(A) &= P(\{\text{FF}\} \cup \{\text{FP}\}) = P(\{\text{FF}\}) + P(\{\text{FP}\}) - P(\{\text{FF}\} \cap \{\text{FP}\}) \\ &= \frac{1}{4} + \frac{1}{4} - P(\varnothing) = \frac{1}{2} - 0 = \frac{1}{2}. \end{aligned}$$

   De même, $P(B) = P(\{\text{FF}\} \cup \{\text{PF}\}) = 1/2$, d'où $P(A)P(B) = 1/4$. Puisque $P(A \cap B) = P(\{\text{FF}\})$ vaut **également** $1/4$, on peut donc conclure que A et B sont **indépendants**.

2. Une carte est tirée d'un jeu de cartes nord-américain régulier (contenant 52 cartes) et bien mélangé. Soit A l'événement qu'il s'agisse d'un as et C l'événement qu'il s'agisse d'un carreau.

   Ces deux événements sont **indépendants** : il y a 4 as ($P(A) = \frac{4}{52} = \frac{1}{13}$) et 13 carreaux $P(C) = \frac{13}{52} = \frac{1}{4}$) dans un tel jeu, de sorte que

$$P(A)P(C) = \frac{1}{13} \times \frac{1}{4} = \frac{1}{52},$$

   et exactement 1 as de carreau dans le jeu, de sorte que $P(A \cap C) = 1/52$.

   De façon plus générale, la valeur de la carte et sa couleur sont indépendantes lorsque l'on tire une carte au hasard d'un paquet régulier.

3. Un dé à six faces est désiquilibré de telle manière que la probabilité d'obtenir chaque valeur est **proportionnelle** à cette valeur. Trouvez $P(3)$.

Solution: Soit $\mathcal{S} = \{1,2,3,4,5,6\}$ l'espace d'échantillonage correspondant à un seul lancer ; pour une certaine constante v, on a $P(k) = kv$, pour $k \in \mathcal{S}$.

Selon l'axiome **A2**, $P(\mathcal{S}) = P(1) + \cdots + P(6) = 1$, de sorte que

$$1 = \sum_{k=1}^{6} P(k) = \sum_{k=1}^{6} kv = v \sum_{k=1}^{6} k = v\frac{(6+1)(6)}{2} = 21v\,.$$

Donc $v = 1/21$ et $P(3) = 3v = 3/21 = 1/7$.

4. Maintenant, le même dé est jeté deux fois, le second lancer étant **indépendant** du premier. Trouvez $P(3_1, 3_2)$.

   Solution: l'expérience est telle que $P(3_1) = 1/7$ et $P(3_2) = 1/7$, comme dans l'exemple précédent. Puisque les lancers de dé sont indépendants, alors

$$P(3_1 \cap 3_2) = P(3_1)P(3_2) = 1/49 \,.$$

   Lancers indépendants $\Longrightarrow$ en pratique, si la même personne lance le même dé, les lancers ne sont peut-être pas indépendants. On peut alors imaginer que deux personnes différentes lancent le dé.

5. Quel avion est plus propice à s'écraser : un bimoteur ou un trimoteur ?

   Solution: Il est plus facile de répondre à cette question si l'on suppose que **les moteurs tombent en panne indépendamment** (réaliste ?). Soit p la probabilité qu'un moteur tombe en panne.

   Comment un avion peut-il s'écraser ? Supposons qu'un :

   - bimoteur s'écrase si ses deux moteurs tombent en panne (probabilité p^2)
   - trimoteur s'écrase si une paire de moteurs tombe en panne, ou si les 3 tombent en panne
     - probabilité qu'**exactement 1 paire de moteurs** tombe en panne indépendamment (c'est-à-dire que deux moteurs tombent en panne et un ne tombe pas en panne) est

$$p \times p \times (1 - p)$$

L'ordre dans lequel les moteurs tombent en panne n'a pas d'importance : il y a donc ${}_3C_2 = \frac{3!}{2!1!} = 3$ façons pour qu'exactement une paire de moteurs tombe en panne : pour 3 moteurs A, B, C, ce sont AB, AC, BC.

– probabilité qu'ils tombent **tous** en panne indépendamment est p^3.

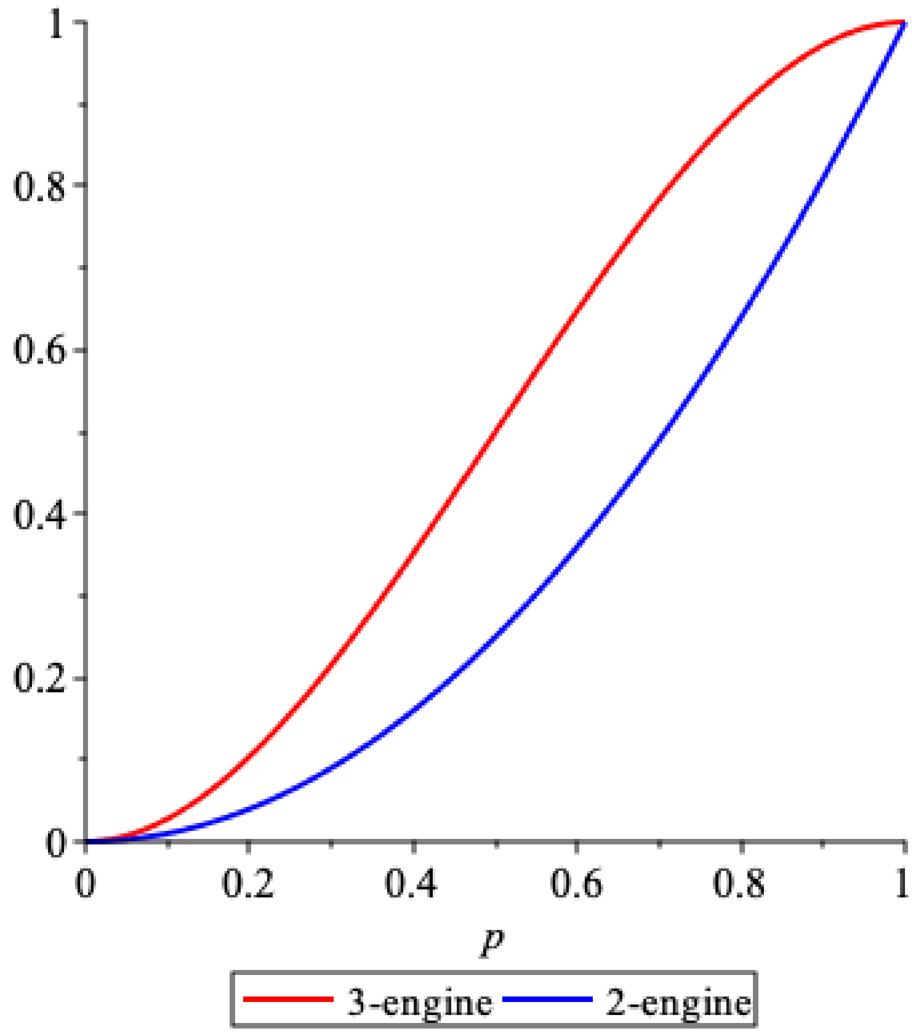
La probabilité qu'au moins 2 moteurs tombent en panne est ainsi :

$$P(\text{au moins } 2 \text{ moteurs défaillants}) = 3p^2(1-p) + p^3 = 3p^2 - 2p^3.$$

Il est plus sûr d'utiliser un bimoteur qu'un trimoteur : ce dernier s'écrasera plus souvent, en supposant qu'il ait besoin de 2 moteurs.

C'est "logique" : le bimoteur n'a besoin que 50% de ses moteurs fonctionnent, tandis que c'est 66% pour le trimoteur.

À votre avis, que pourrait être une valeur réaliste de p ?


1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
p
3-engine
2-engine

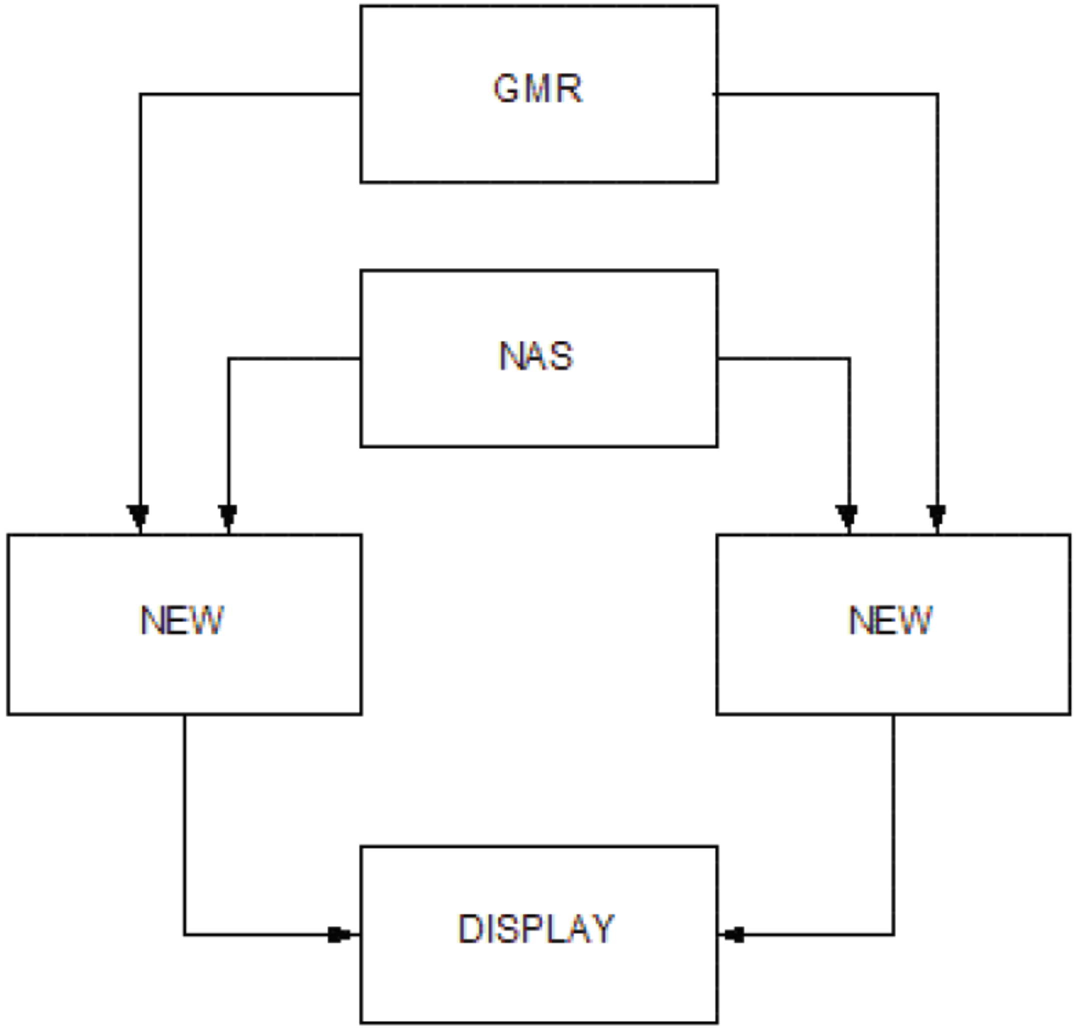
Scénario : Contrôle du trafic aérien

Le contrôle du trafic aérien est une activité liée à la sécurité.

Chaque pièce de l'équipement est conçue selon les normes de sécurité les plus strictes et, dans de nombreux cas, des équipements en double sont fournis, de sorte que si un élément tombe en panne, un autre prend le relais.

Un nouveau système doit être mis en place pour transmettre les informations de l'aéroport d'Heathrow au contrôle du terminal à West Drayton. Dans le cadre de la conception du système, une décision doit être prise quant à la nécessité de prévoir une duplication.

Le nouveau système utilise les données du *Ground Movements Radar* (GMR) de Heathrow, les combine avec celles du *National Airspace System* (NAS), et envoie les résultats à un écran au *Terminal Control*.


GMR
NAS
NEW
NEW
DISPLAY

Pour tous les systèmes existants, des registres des défaillances sont conservés et une **probabilité expérimentale de défaillance** est calculée.

La **fiabilité** d'un système est définie par $R = 1 - P$, où $P = P(\text{défaillance})$.

Données: $R_{\text{GMR}} = R_{\text{NAS}} = 0.9999$ (soit 1 panne par $10,000$ heures).

Hypothèse: les probabilités de défaillance sont indépendantes.

Pour le système ci-dessus, si un seul module NEW est introduit, la fiabilité du système (modèle à simple fil) est

$$R_{\text{STD}} = R_{\text{GMR}} \times R_{\text{NEW}} \times R_{\text{NAS}}.$$

Si le module NEW est dupliqué, la fiabilité du modèle à double fil DTD est

$$R_{\text{DTD}} = R_{\text{GMR}} \times (1 - (1 - R_{\text{NEW}})^2) \times R_{\text{NAS}}.$$

La duplication du module NEW entraîne une amélioration de la fiabilité de

$$\rho = \frac{R_{\text{DTD}}}{R_{\text{STD}}} = \frac{(1-(1-R_{\text{NEW}})^2)}{R_{\text{NEW}}} \times 100\% \,.$$

Pour le module NEW, aucune donnée historique n'est disponible; nous calculons l'**amélioration obtenue** en utilisant DTD pour différentes valeurs de R_{NEW}.

R_{NEW}	0.1	0.2	0.5	0.75	0.99	0.999	0.9999	0.99999
ρ (%)	190	180	150	125	101	100.1	100.01	100.001

Si le module NEW est très peu fiable (c'est-à-dire que R_{NEW} est petit), alors il y a un avantage significatif à utiliser DTD (ρ est grand).

Mais pourquoi installer un module que nous savons ne pas être fiable ?

Si le nouveau module est aussi fiable que NAS et GMR, c'est-à-dire si

$$R_{\mathsf{GMR}} = R_{\mathsf{NEW}} = R_{\mathsf{NAS}} = 0.9999,$$

alors STD a une fiabilité combinée de 0.9997 (c'est-à-dire 3 pannes par $10,000$ heures), tandis que DTD a une fiabilité combinée de 0.9998 (c'est-à-dire 2 pannes par $10,000$ heures).

Si la probabilité de défaillance est indépendante pour chaque composante, nous pourrions en conclure que le gain de fiabilité du DTD ne justifie probablement pas le coût supplémentaire.

(Avez-vous bien compris tous les calculs ?)

Les probabilités conditionnelles

Nous pouvons mieux comprendre quand l'indépendance s'applique en définissant la **probabilité conditionnelle d'un événement B étant donné qu'un autre événement A s'est produit** comme suit

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

Cela n'a de sens que lorsque A peut "arriver", c'est-à-dire quand $P(A) > 0$. Si $P(A) \cdot P(B) > 0$, alors

$$P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B) = P(B \cap A);$$

A et B sont ainsi indépendants si $P(B \mid A) = P(B)$ et $P(A \mid B) = P(A)$.

Exemples:

1. On choisit 1 personne dans un groupe de 100. Quelle est la probabilité que cette personne souffre d'hypertension artérielle (HTA) ?

   Solition: si on ne sait rien d'autre sur la population, il s'agit d'une **probabilité (inconditionnelle)**, à savoir

$$P(\text{HTA}) = \frac{\#\text{individus avec HTA dans la population}}{100}.$$

2. Si au contraire, nous filtrons d'abord tout ceux ayant un faible taux de cholestérol, quelle est la probabilité qu'une personne soit atteinte d'HTA ? Il s'agit d'une **probabilité conditionnelle** $P(\text{HTA} \mid \text{cholestérol élevé})$; la probabilité de sélectionner une personne atteinte d'HTA, étant donné un taux de cholestérol élevé, est vraisemblablement différente de celle de $P(\text{HTA} \mid \text{cholestérol faible})$.

3. Un échantillon de 249 individus est prélevé et chaque personne est classée par groupe sanguin et statut tuberculeux (TB).

	O	**A**	**B**	**AB**	Total
TB	34	37	31	11	113
pas de TB	55	50	24	7	136
Total	89	87	55	18	249

La probabilité (inconditionnelle) qu'un individu aléatoire soit atteint de tuberculose est $P(\text{TB}) = \frac{\#\text{TB}}{249} = \frac{113}{249} = 0.454$. Parmi les individus dont le sang est de type **B**, la probabilité (conditionnelle) d'être atteint de tuberculose est la suivante

$$P(\text{TB} \mid \text{type } \mathbf{B}) = \frac{P(\text{TB} \cap \text{type } \mathbf{B})}{P(\text{type } \mathbf{B})} = \frac{31}{55} = \frac{31/249}{55/249} = 0.564.$$

4. Une famille a deux enfants (pas des jumeaux). Quelle est la probabilité que le plus jeune enfant soit une fille étant donné qu'au moins un des enfants est une fille ? Supposez que les garçons et les filles ont la même probabilité de naître.

Solution: Soit A et B les événements que le plus jeune enfant soit une fille et qu'au moins un enfant soit une fille, respectivement :

$$A = \{\text{GG}, \text{BG}\} \quad \text{et} \quad B = \{\text{GG}, \text{BG}, \text{GB}\}, \quad \text{de sorte que} \quad A \cap B = A.$$

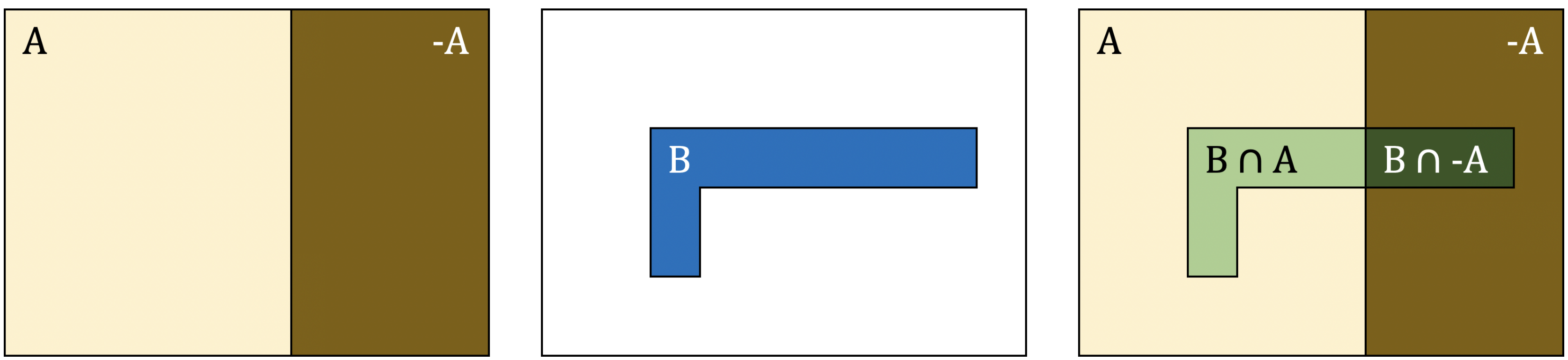
Alors $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{2/4}{3/4} = \frac{2}{3}$ (et non $\frac{1}{2}$).

Incidemment, $P(A \cap B) = P(A) \neq P(A) \times P(B)$ ce qui signifie que A et B sont des événements **dépendants**.

La loi de la probabilité totale

Soient A et B deux événements. D'après la théorie des ensembles, nous avons

$$B = (A \cap B) \cup (A^c \cap B).$$



Notez que $A \cap B$ et $A^c \cap B$ sont **mutuellement exclusifs**.

Selon l'axiome **A4**, nous avons alors

$$P(B) = P(A \cap B) + P(A^c \cap B).$$

En supposant de plus que $\varnothing \neq A \neq \mathcal{S}$, alors

$$P(A \cap B) = P(B \mid A)P(A) \quad \text{et} \quad P(A^c \cap B) = P(B \mid A^c)P(A^c),$$

de sorte que $P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c)$.

Cela se généralise comme suit : si $A_1 \ldots, A_k$ sont **mutuellement exclusifs** et **exhaustifs** (c-à-d que $A_i \cap A_j = \varnothing$ pour tout $i \neq j$ et $A_1 \cup \cdots \cup A_k = \mathcal{S}$), alors pour tout événement B, nous avons :

$$P(B) = \sum_{j=1}^{k} P(B \mid A_j)P(A_j) = P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_k)P(A_k).$$

Exemple: utilisez la loi de la probabilité totale pour calculer $P(\text{TB})$ à l'aide des données de l'exemple précédent.

Solution: les groupes sanguins $\{\mathbf{O}, \mathbf{A}, \mathbf{B}, \mathbf{AB}\}$ forment une partition mutuellement exclusive de la population, avec

$$P(\mathbf{O}) = \frac{89}{249}, \quad P(\mathbf{A}) = \frac{87}{249}, \quad P(\mathbf{B}) = \frac{55}{249}, \quad \text{et} \quad P(\mathbf{AB}) = \frac{18}{249}.$$

Il est facile de voir que $P(\mathbf{O}) + P(\mathbf{A}) + P(\mathbf{B}) + P(\mathbf{AB}) = 1$. De plus,

$$P(\text{TB} \mid \mathbf{O}) = \frac{P(\text{TB} \cap \mathbf{O})}{P(\mathbf{O})} = \frac{34}{89}, \quad P(\text{TB} \mid \mathbf{A}) = \frac{P(\text{TB} \cap \mathbf{A})}{P(\mathbf{A})} = \frac{37}{87},$$
$$P(\text{TB} \mid \mathbf{B}) = \frac{P(\text{TB} \cap \mathbf{B})}{P(\mathbf{B})} = \frac{31}{55}, \quad P(\text{TB} \mid \mathbf{AB}) = \frac{P(\text{TB} \cap \mathbf{AB})}{P(\mathbf{AB})} = \frac{11}{18}.$$

Selon la loi de la probabilité totale,

$$\begin{aligned} P(\mathsf{TB}) = {} & P(\mathsf{TB} \mid \mathbf{O})P(\mathbf{O}) + P(\mathsf{TB} \mid \mathbf{A})P(\mathbf{A}) \\ & + P(\mathsf{TB} \mid \mathbf{B})P(\mathbf{B}) + P(\mathsf{TB} \mid \mathbf{AB})P(\mathbf{AB}), \end{aligned}$$

d'où

$$\begin{aligned} P(\mathsf{TB}) &= \frac{34}{89} \cdot \frac{89}{249} + \frac{37}{87} \cdot \frac{87}{249} + \frac{31}{55} \cdot \frac{55}{249} + \frac{11}{18} \cdot \frac{18}{249} \\ &= \frac{34 + 37 + 31 + 11}{249} = \frac{113}{249} = 0.454, \end{aligned}$$

ce qui correspond au résultat de l'exemple précédent.

1.7 – Le théorème de Bayes

Après qu'une expérience ait généré un résultat, nous sommes souvent intéressés par la probabilité qu'une certaine condition soit présente compte tenu du résultat (ou qu'une hypothèse particulière soit valide, par exemple).

Nous avons remarqué précédemment que si $P(A)P(B) > 0$, alors

$$P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A \mid B) = P(B \cap A);$$

ceci peut être ré-écrit comme le **théorème de Bayes** :

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}.$$

Ce dernier n'est ainsi qu'un simple corollaire des règles de la probabilité.

La question centrale de l'analyse des données

Compte tenu de tout ce qui était connu avant l'expérience, les données recueillies/observées confirment (ou infirment) l'hypothèse/la présence d'une certaine condition ?

Problème: il est presque toujours impossible de le calculer directement.

Solution: En utilisant le théorème de Bayes, nous pouvons réécrire la question comme suit :

$$\begin{aligned} P(\text{hypothèse} \mid \text{données}) &= \frac{P(\text{données} \mid \text{hypothèse}) \times P(\text{hypothèse})}{P(\text{données})}, \\ &\propto P(\text{données} \mid \text{hypothèse}) \times P(\text{hypothèse}) \end{aligned}$$

en espérant que les termes à droite soient plus faciles à calculer.

Le vernaculaire bayésien

Les termes suivants sont utilisés dans l'analyse bayésienne :

- $P(\text{hypothèse})$ est la probabilité que l'hypothèse soit vraie avant l'expérience (appelée **probabilité** *a priori*) ;
- $P(\text{hypothèse} \mid \text{données})$ est la probabilité que l'hypothèse soit vraie une fois les données expérimentales obtenues (**probabilité** *a posteriori*) ;
- $P(\text{données} \mid \text{hypothèse})$ est la probabilité que les données expérimentales soient observées si l'hypothèse est vraie (la **vraisemblance**)

Sous la forme **a posteriori** $\propto$ **vraisemblance** $\times$ **a priori**, ceci dit que les **croyances sont mises à jour en présence de nouvelles informations**.

Formulations

Si A et B sont des événements pour lesquels $P(A) \cdot P(B) > 0$, alors on peut ré-écrire le théorême de Bayes comme suit :

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}.$$

Cela se généralise comme suit : si $A_1, \ldots, A_k$ sont des événements **mutuellement exclusifs** et **exhaustifs**, alors pour tout événement B et pour chaque i,

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_k)P(A_k)}.$$

Exemples:

1. En 1999, Nissan vendait 3 modèles de voitures en Amérique : la Sentra (S), la Maxima (M) et le Pathfinder (Pa). Parmi les véhicules vendus, 50% étaient des S, 30% des M, et 20% des Pa. La même année, 12% des S, 15% des M, et 25% des Pa présentaient un défaut particulier (D).

   (a) Quelle est la probabilité qu'une Nissan 1999 présente ce défaut ?

   Solution: Dans le langage des probabilités conditionnelles,

$$
\begin{aligned}
P(\text{S}) = 0.5, \quad P(\text{M}) = 0.3, \quad P(\text{Pa}) = 0.2,\\
P(D \mid \text{S}) = 0.12, \quad P(D \mid \text{M}) = 0.15, \quad P(D \mid \text{Pa}) = 0.25, \text{ d'où}\\
P(D) &= P(D \mid \text{S})P(\text{S}) + P(D \mid \text{M})P(\text{M}) + P(D \mid \text{Pa})P(\text{Pa})\\
&= 0.12 \times 0.5 + 0.15 \times 0.3 + 0.25 \times 0.2 = 0.155 = 15.5\%
\end{aligned}
$$

(b) Une voiture possède le défaut D. Peut-on déduire son modèle ?

Solution: nous comparons les probabilités postérieures $P(\mathsf{M} \mid D)$, $P(\mathsf{S} \mid D)$, et $P(\mathsf{Pa} \mid D)$ (et non les probabilités *a priori*), calculées à l'aide du théorème de Bayes :

$$P(\mathsf{S} \mid D) = \frac{P(D \mid \mathsf{S})P(\mathsf{S})}{P(D)} = \frac{0.12 \times 0.5}{0.155} \approx 38.7\%$$

$$P(\mathsf{M} \mid D) = \frac{P(D \mid \mathsf{M})P(\mathsf{M})}{P(D)} = \frac{0.15 \times 0.3}{0.155} \approx 29.0\%$$

$$P(\mathsf{Pa} \mid D) = \frac{P(D \mid \mathsf{Pa})P(\mathsf{Pa})}{P(D)} = \frac{0.25 \times 0.2}{0.155} \approx 32.3\%$$

Même si les Sentras sont moins susceptibles d'avoir le défaut D, leur prévalence globale dans la population leur permet de passer le cap.

2. Supposons qu'un test pour une condition ait un taux de réussite très élevé. Si un patient

   - est atteint de la maladie, le test est 'positif' 0.99% du temps ;
   - n'a pas la maladie, le test est 'négatif' 0.95% du temps.

   Supposons que 0.1% de la population ait la condition. Quelle est la probabilité qu'un patient dont le test est positif ne soit pas atteint ?

   Solution: Soit D l'événement que le patient est atteint, et A l'événement que le test est positif. La probabilité d'un **vrai positif** est

$$P(D \mid A) = \frac{P(A \mid D)P(D)}{P(A \mid D)P(D) + P(A \mid D^c)P(D^c)}$$
$$= \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999} \approx 0.019.$$

La probabilité d'un **faux positif** est donc de $1 - 0.019 \approx 0.981$.

Malgré la précision apparemment élevée du test, l'incidence de la condition est si faible (1 sur 1000) que la grande majorité des patients dont le test est positif (98 sur 100) n'ont pas la condition.

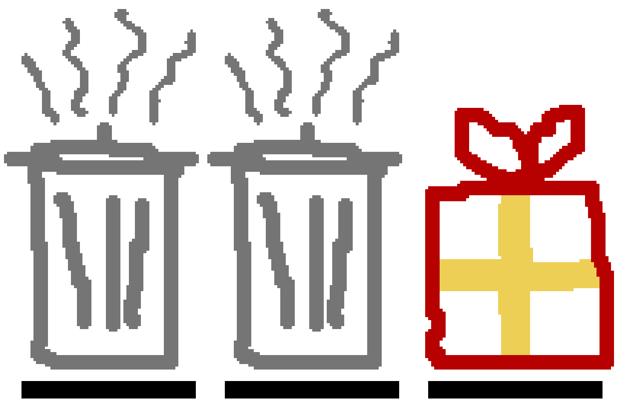
Les 2 sur 100 qui sont de vrais positifs représentent toujours 20 fois la proportion de positifs trouvés dans la population (avant que le résultat du test ne soit connu).

Il est **important de retenir** que lorsqu'on traite de probabilités, il faut tenir compte **à la fois** de la vraisemblance et de la prévalence !

Dans ce cas, même si le taux de réussite du test est très élevé, la **faible** prévalence de la condition fait en sorte qu'il y a de **fortes** chances qu'un test positif soit un **faux positif**.

3. (**Problème de Monty Hall**) Vous participez à un jeu télévisé et on vous donne le choix entre trois portes. Derrière une porte se trouve un prix ; derrière les autres, des détritus. Vous choisissez une porte, disons la 1e, et l'animatrice, qui sait ce qui se trouve derrière les portes, ouvre une autre porte, disons la 3e, derrière laquelle se trouve des détritus. Elle vous dit alors : "voulez-vous changer de la 1e porte 2e ?".

   Avez-vous intérêt à le faire ?



Solution: dans ce qui suit, soient S et D représentent les événements que le passage à une autre porte est une stratégie gagnante, et que le prix se trouvait derrière la porte originale, respectivement.

- Supposons d'abord que l'animatrice n'ouvre aucune porte. Quelle est la probabilité que le passage à une autre porte dans ce scénario s'avère être une stratégie gagnante ?

  Si le prix **se trouve** derrière la porte originale, le changement réussit 0% du temps : $P(\text{S} \mid \text{D}) = 0$ (et $P(\text{D}) = 1/3$). Si le prix **ne se trouve pas** derrière la porte d'origine, le changement réussira dans 50% des cas : $P(\text{S} \mid \text{D}^c) = 1/2$ (et $P(\text{D}^c) = 2/3$).

  Ainsi, $P(\text{S}) = P(\text{S} \mid \text{D})P(\text{D}) + P(\text{S} \mid \text{D}^c)P(\text{D}^c) = 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{2}{3} \approx 33\%.$

- Si l'animatrice ouvre l'une des deux autres portes pour montrer les détritus, quelle est la probabilité que le passage à une autre porte dans ce scénario s'avère être une stratégie gagnante ?

  Si le prix **se trouve** derrière la porte originale, le changement réussira 0% du temps : $P(\text{S} \mid \text{D}) = 0$ (avec $P(\text{D}) = 1/3$). Si le prix **ne se trouve pas** derrière la porte originale, le changement réussira 100% du temps : $P(\text{S} \mid \text{D}^c) = 1$ (avec $P(\text{D}^c) = 2/3$).

  Ainsi, $P(\text{S}) = P(\text{S} \mid \text{D})P(\text{D}) + P(\text{S} \mid \text{D}^c)P(\text{D}^c) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} \approx 67\%.$

Si aucune porte n'est ouverte, l'échange **n'est pas** une stratégie gagnante puisqu'il ne donne lieu à un succès que dans 33% des cas. En revanche, si une porte est ouverte, l'échange **devient** la stratégie gagnante, avec un taux de réussite de 67%.

Annexe – Les sophismes probabilistes

Soyez attentif à ces erreurs faciles à commettre :

- **l'erreur du joueur** : interprétations fréquentistes vs bayésiennes ;
- **erreur de taux de base** : faible incidence/précision élevée du test ;
- **faux raisonnement du procureur** : petites probabilités et grandes populations ;
- **confusion de l'inverse** : $P(A \mid B) \neq P(B \mid A)$, en général ;
- **Paradoxe de Simpson** : les schémas dans les sous-groupes peuvent ne pas se généraliser à la population entière.