MAT4681 - Statistique pour les sciences

Arthur Charpentier

03 - Probabilités

été 2022

Calcul de probabilités I

Dans une classe de 50 étudiants,
il y a 20 hommes, 25 personnes brunes
On tire une personne au hasard, A est l'évènement "la personne est un homme", B est l'évènement "la personne est brune".

Que vaut $\mathbb{P}(A \cup B)$?

$$\mathbb{P}(A \cup B) = \underbrace{\mathbb{P}(A)}_{20/50} + \underbrace{\mathbb{P}(B)}_{25/50} - \underbrace{\mathbb{P}(A \cap B)}_{?}$$

- $\mathbb{P}(A \cap B)$ est minimale si $A \cup B = \emptyset$ et $\mathbb{P}(A \cap B) = 0$
- $\mathbb{P}(A \cap B)$ est maximale si $A \subset B$ et $\mathbb{P}(A \cap B) = \mathbb{P}(A) = 40\%$

donc $\mathbb{P}(A \cup B) \in [50\%; 90\%]$.

Conditionnement I

Probabilité conditionnelle

Soient deux évènements A et B (dans $\mathcal{F}$) tels que $\mathbb{P}(B) \neq 0$, on note

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

Aussi, $\mathbb{P}(A \cap B) = \begin{cases} \mathbb{P}(A|B) \times \mathbb{P}(B) \\ \mathbb{P}(B|A) \times \mathbb{P}(A) \end{cases}$

Conditionnement II

On lance une pièce 4 fois, $\begin{cases} A = & \text{au moins 3 fois 'pile'} \\ B = & \text{'face' au premier tirage} \end{cases}$

$$\begin{cases} \Omega = \{FFFF, FFFP, \cdots, FPPP, PPPP\} \\ A = \{FPPP, PFPP, PPFP, PPPF, PPPP\} \\ B = \{FPPP, FPPF, FPFP, FFPP, FPFF, FFPF, FPFF, FFFF\} \\ A \cap B = \{FPPP\} \end{cases}$$

Conditionnement III

On lance une pièce 4 fois, $\begin{cases} A = \text{ au moins 3 fois 'pile'} \\ B = \text{ 'face' au premier tirage} \end{cases}$

Que vaut $\mathbb{P}(A|B)$?

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{8}$$

Que vaut $\mathbb{P}(B|A)$?

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{1}{5}$$

Formule des probabilités totales et formule de Bayes I

Si $B_1, \cdots, B_k$ forme une partition de Ω, i.e.

- $B_1 \cup \cdots \cup B_k = \Omega$
- $\forall i, j,\ B_i \cap B_j = \emptyset$

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A \cap B_i) = \sum_{i=1}^{k} \mathbb{P}(A|B_i) \times \mathbb{P}(B_i)$$

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

Indépendance I

Indépendance

Deux évènements A et B (dans $\mathcal{F}$) sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$$

On note $A \perp\!\!\!\perp B$

Il s'agit de la définition de l'indépendance (on verra d'autres caractérisations plus simples par la suite)

Notons que si $A \subset B$ avec $A \neq B$, $A \cap B = A \neq B$, et $B \neq \Omega$,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) > \mathbb{P}(A) \times \mathbb{P}(B)$$

donc $A \not\perp\!\!\!\perp B$

Indépendance II

On lance deux fois une même pièce, $\mathbb{P}(\{PP, FF\}) \overset{?}{\gtrless} \mathbb{P}(\{FP, PF\})$
Soit $p = \mathbb{P}(\{P\})$ (et $1 - p = \mathbb{P}(\{F\})$). Si les tirages sont indépendants,

$$\begin{cases} \mathbb{P}(\{PP\}) = \mathbb{P}(\{P\}) \cdot \mathbb{P}(\{P\}) = p^2 \\ \mathbb{P}(\{FF\}) = \mathbb{P}(\{F\}) \cdot \mathbb{P}(\{F\}) = (1-p)^2 \\ \mathbb{P}(\{FP\}) = \mathbb{P}(\{F\}) \cdot \mathbb{P}(\{P\}) = (1-p)p \\ \mathbb{P}(\{PF\}) = \mathbb{P}(\{P\}) \cdot \mathbb{P}(\{F\}) = p(1-p) \end{cases}$$

Notons que

$$\mathbb{P}(\{PP, FF, FP, PF\}) = p^2+(1-p)^2+2p(1-p) = p^2+1-2p+p^2+2p-2p^2$$

Or $(a-b)^2 \geq 0$ implique $a^2 + b^2 \geq 2ab$, donc

$$p^2 + (1-p)^2 \geq 2p(1-p) \text{ donc } \mathbb{P}(\{PP, FF\}) \geq \mathbb{P}(\{FP, PF\})$$

Indépendance III

Probabilité conditionnelle et indépendance

Deux évènements A et B (dans $\mathcal{F}$) sont indépendants

$$\begin{cases} \text{si } \mathbb{P}(A) \neq 0, \ \mathbb{P}(B|A) = \mathbb{P}(B) \\ \text{si } \mathbb{P}(B) \neq 0, \ \mathbb{P}(A|B) = \mathbb{P}(A) \end{cases}$$

On lance deux pièces. $A = \{$ face pour la première pièce $\}$ et $B = \{$ face pour la seconde pièce $\}$, $\mathbb{P}[B|A] = \mathbb{P}[B]$

Indépendance IV

On lance deux dés (à 6 faces), $\begin{cases} A = & \text{le premier dé est un 3} \\ B = & \text{la somme des dés vaut 6} \\ C = & \text{la somme des dés vaut 7} \end{cases}$

$$\begin{cases} A = \{31, 32, 33, 34, 35, 36\} \\ B = \{15, 24, 33, 42, 51\} \\ C = \{16, 25, 34, 43, 52, 61\} \\ A \cap B = \{33\} \text{ et } A \cap C = \{34\} \end{cases}$$

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{5} \neq \mathbb{P}(A), \ A \not\perp\!\!\!\perp B$$

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{1}{6} = \mathbb{P}(A), \ A \perp\!\!\!\perp C$$

Indépendance V

Parmi les évènements suivants, quand a-t-on $A \perp\!\!\!\perp B$?

On lance deux dés, $A = \{$ premier dé est 6 $\}$ et $B = \{$ somme > 6 $\}$

On pige deux cartes, $A = \{$ première $\heartsuit\}$ et $B = \{$ seconde $\clubsuit\}$

On pige deux cartes, $A = \{$ première $\heartsuit\}$ et $B = \{$ seconde est 10$\}$

Indépendance VI

Parmi les évènements suivants, quand a-t-on $A \perp\!\!\!\perp B$?

On lance deux dés, $A = \{$ premier dé est 6 $\}$ et $B = \{$ somme > 6 $\}$

Non indépendants ($A \subset B$)

On pige deux cartes, $A = \{$ première $\heartsuit\}$ et $B = \{$ seconde $\clubsuit\}$

Non indépendants $\mathbb{P}[A \cap B] = \dfrac{1}{4} \cdot \dfrac{13}{51} > \dfrac{1}{4} \cdot \dfrac{13}{52} = \mathbb{P}[A] \cdot \mathbb{P}[B]$

On pige deux cartes, $A = \{$ première $\heartsuit\}$ et $B = \{$ seconde est 10$\}$

Indépendants $\mathbb{P}[A \cap B] = \dfrac{12}{52} \cdot \dfrac{4}{51} + \dfrac{1}{52} \cdot \dfrac{3}{51} = \dfrac{1}{52} = \mathbb{P}[A] \cdot \mathbb{P}[B]$

Probabilité et cote I

Dans un jeu de 52 cartes, soit A l'évènement 'en tirant une carte au hasard, on a un nombre pair'. Que vaut $\mathbb{P}[A]$?
Il y a 5 cartes associés à un nombre pair (2, 4, 6, 8, 10) et 4 couleurs, donc 20 cartes en tout. Donc

$$\mathbb{P}[A] = \frac{20}{52} = \frac{5}{13}$$

Cote

La cote d'un évènement A (ou en faveur de A) est $a : \overline{a}$ où $\mathbb{P}[A] = a/(a+\overline{a})$, ou parfois $\mathbb{P}[A]/\mathbb{P}[\overline{A}]$.

- $1 : 1$ correspond à un évènement de probabilité $1/2$.
- $2 : 1$ correspond à un évènement de probabilité $2/3$.
- $1 : 2$ correspond à un évènement de probabilité $1/3$.

Probabilité et cote II

Dans l'exemple précédant (carte paire), quelle est la cote de A ?
Notons que $\mathbb{P}[A] = \dfrac{20}{52} = \dfrac{5}{13}$ et $\mathbb{P}[\overline{A}] = 1 - \dfrac{5}{13} = \dfrac{8}{13}$

La cote est alors $5 : 8$.

La cote est aussi parfois simplement définie comme le rapport $\mathbb{P}[A]/\mathbb{P}[\overline{A}]$
Pour passer d'une cote $a : \overline{a}$ à $\mathbb{P}[A]$, notons que

$$\mathbb{P}[A] \propto a \text{ et } \mathbb{P}[\overline{A}] \propto \overline{a}$$

et donc $\mathbb{P}[A] = \dfrac{a}{a + \overline{a}}$

Arbres et ensembles

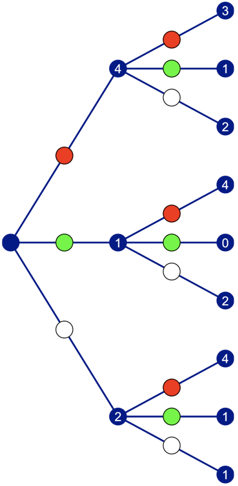
Une urne avec 10 boules, 5 rouge ● 2 vertes ● et 3 blanches ○.

La probabilité conditionnelle de A sachant que B est survenu est

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \text{ et } B]}{\mathbb{P}[B]}$$

ou $\mathbb{P}[A \text{ and } B] = \mathbb{P}[A|B] \cdot \mathbb{P}[B]$
(on parle de chain rule)

On pige 2 boules, sans remise, quelle est la probabilité d'avoir au moins une boule verte ?



Arbres et ensembles

On pige 2 boules, sans remise, quelle est la probabilité d'avoir au moins une boule verte ?

$$p = \mathbb{P}\big[X_1 = \bullet \text{ ou } X_2 = \bullet\big]$$

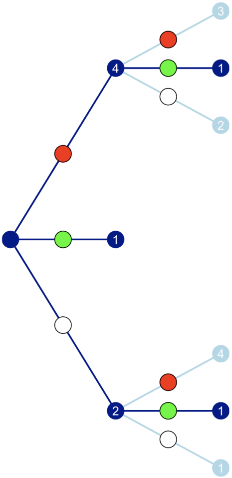
$$p = \mathbb{P}\big[X_1 = \bullet\big] + \mathbb{P}\big[X_2 = \bullet \text{ et } X_1 \neq \bullet\big]$$

$$p = \mathbb{P}\big[X_1 = \bullet\big] + \mathbb{P}\big[X_2 = \bullet | X_1 = \bullet\big] \cdot \mathbb{P}\big[|X_1 = \bullet\big]$$
$$+\mathbb{P}\big[X_2 = \bullet | X_1 = \circ\big] \cdot \mathbb{P}\big[|X_1 = \circ\big]$$

$$p = \frac{2}{10} + \frac{2}{9} \cdot \frac{5}{10} + \frac{2}{9} \cdot \frac{3}{10} = \cdots = \frac{17}{45}$$

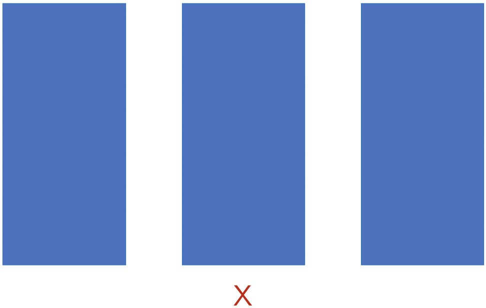
ou encore plus simple

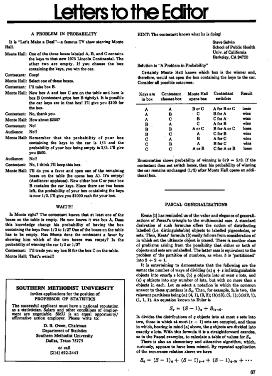
$$p = 1 - \mathbb{P}\big[X_1 \in \{\bullet, \circ\} \text{ et } X_2 \in \{\bullet, \circ\}\big] = 1 - \frac{8}{10} \cdot \frac{7}{9} = \frac{17}{45}$$



Monty Hall

Trois portes : l'une d'elles a un coffre au trésor derrière elle et les deux autres ont des chèvres. Vous choisissez une porte et l'indiquez à Monty. Il ouvre l'une des deux autres portes pour révéler une chèvre. Maintenant, devez-vous vous en tenir à votre choix initial ou passer à l'autre porte non ouverte ?
cf wikipedia





Letters to the Editor

A PROBLEM IN PROBABILITY

It is "Let's Make a Deal"—a famous TV show starring Monte Hall.

Monte Hall: One of the three boxes labeled A, B, and C contains the keys to that new 1975 Lincoln Continental. The other two are empty. If you choose the box containing the keys, you win the car.
Contestant: Gasp!
Monte Hall: Select one of these boxes.
Contestant: I'll take box B.
Monte Hall: Now box A and box C are on the table and here is box B (contestant grips box B tightly). It is possible the car keys are in that box! I'll give you $100 for the box.
Contestant: No, thank you.
Monte Hall: How about $200?
Contestant: No!
Audience: No!!
Monte Hall: Remember that the probability of your box containing the keys to the car is 1/3 and the probability of your box being empty is 2/3. I'll give you $500.
Audience: No!!
Contestant: No, I think I'll keep this box.
Monte Hall: I'll do you a favor and open one of the remaining boxes on the table (he opens box A). It's empty! (Audience: applause). Now either box C or your box B contains the car keys. Since there are two boxes left, the probability of your box containing the keys is now 1/2. I'll give you $1000 cash for your box.

WAIT!!!!

Is Monte right? The contestant knows that at least one of the boxes on the table is empty. He now knows it was box A. Does this knowledge change his probability of having the box containing the keys from 1/3 to 1/2? One of the boxes on the table has to be empty. Has Monte done the contestant a favor by showing him which of the two boxes was empty? Is the probability of winning the car 1/2 or 1/3?

Contestant: I'll trade you my box B for the box C on the table.
Monte Hall: That's weird!!

HINT: The contestant knows what he is doing!

Steve Selvin
School of Public Health
Univ. of California
Berkeley, CA 94720

Solution to "A Problem in Probability"

Certainly Monte Hall knows which box is the winner and, therefore, would not open the box containing the keys to the car. Consider all possible outcomes:

Keys are in box	Contestant chooses box	Monte Hall opens box	Contestant switches	Result
A	A	B or C	A for B or C	loses
A	B	C	B for A	wins
A	C	B	C for A	wins
B	A	C	A for B	wins
B	B	A or C	B for A or C	loses
B	C	A	C for B	wins
C	A	B	A for C	wins
C	B	A	B for C	wins
C	C	A or B	C for A or B	loses

Enumeration shows probability of winning is 6/9 = 2/3. If the contestant does not switch boxes, then his probability of winning the car remains unchanged (1/3) after Monte Hall opens an additional box.

PASCAL GENERALIZATIONS

Keats [1] has reminded us of the value and elegance of generalizations of Pascal's triangle to the multinomial case. A standard derivation of such formulae offers the notion of distributing labelled (i.e. distinguishable) objects to labelled pigeonholes, or sets. Thus, Keats' formula (3) easily follows from consideration of in which set the ultimate object is placed. There is another class of problems arising from the possibility that either or both of objects and sets are unlabelled. The latter case is equivalent to the problem of the partition of numbers, as when 5 is 'partitioned' into 2 + 2 + 1.

It is entertaining to demonstrate that the following are the same: the number of ways of dividing (a) $q + s$ indistinguishable objects into *exactly* s lots, (b) q objects into *at most* s lots, and (c) q objects into any number of lots, but with no more than s objects in each. Let us select a notation in which the common answer to these questions is S_q. Then, for example, 2_3 is two, the relevant partitions being (a):(4, 1), (3, 2); (b):(3), (2, 1); (c):(2, 1), (1, 1, 1). An equation known to Euler is

$$S_q = (S-1)_q + S_{(q-s)}.$$

It divides the distributions of q objects into at most s sets into two, those in which at most $(s-1)$ sets are occupied, and those in which, bearing in mind (a) above, the q objects are divided into exactly s lots. With this formula it is a straightforward exercise, as in the Pascal examples, to calculate a table of values for S_q.

There is also an elementary and attractive algorithm, which, curiously, appears to have been missed. By repeated application of the recurrence relation above we have

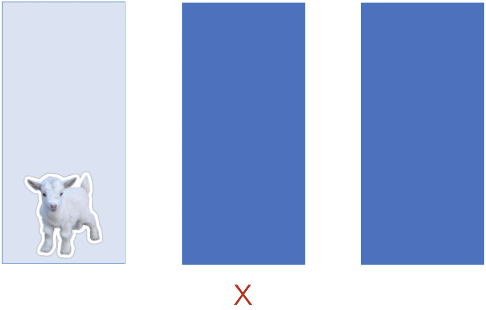
$$S_q = (S-1)_q + (S-1)_{q-s} + (S-1)_{q-2s} + \cdots$$

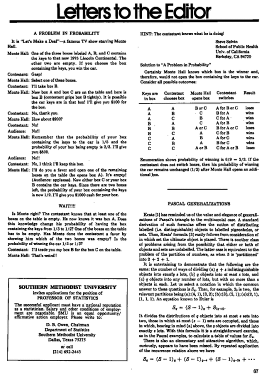
67

Monty Hall

Trois portes : l'une d'elles a un coffre au trésor derrière elle et les deux autres ont des chèvres. Vous choisissez une porte et l'indiquez à Monty. Il ouvre l'une des deux autres portes pour révéler une chèvre. Maintenant, devez-vous vous en tenir à votre choix initial ou passer à l'autre porte non ouverte ?

cf wikipedia





Letters to the Editor

A PROBLEM IN PROBABILITY

It is "Let's Make a Deal"—a famous TV show starring Monte Hall.

Monte Hall: One of the three boxes labeled A, B, and C contains the keys to that new 1975 Lincoln Continental. The other two are empty. If you choose the box containing the keys, you win the car.
Contestant: Gasp!
Monte Hall: Select one of these boxes.
Contestant: I'll take box B.
Monte Hall: Now box A and box C are on the table and here is box B (contestant grips box B tightly). It is possible the car keys are in that box! I'll give you $100 for the box.
Contestant: No, thank you.
Monte Hall: How about $200?
Contestant: No!
Audience: No!!
Monte Hall: Remember that the probability of your box containing the keys to the car is 1/3 and the probability of your box being empty is 2/3. I'll give you $500.
Audience: No!!
Contestant: No, I think I'll keep this box.
Monte Hall: I'll do you a favor and open one of the remaining boxes on the table (he opens box A). It's empty! (Audience: applause). Now either box C or your box B contains the car keys. Since there are two boxes left, the probability of your box containing the keys is now 1/2. I'll give you $1000 cash for your box.

WAIT!!!!

Is Monte right? The contestant knows that at least one of the boxes on the table is empty. He now knows it was box A. Does this knowledge change his probability of having the box containing the keys from 1/3 to 1/2? One of the boxes on the table has to be empty. Has Monte done the contestant a favor by showing him which of the two boxes was empty? Is the probability of winning the car 1/2 or 1/3?

Contestant: I'll trade you my box B for the box C on the table.
Monte Hall: That's weird!!

HINT: The contestant knows what he is doing!

Steve Selvin
School of Public Health
Univ. of California
Berkeley, CA 94720

Solution to "A Problem in Probability"

Certainly Monte Hall knows which box is the winner and, therefore, would not open the box containing the keys to the car. Consider all possible outcomes:

Keys are in box	Contestant chooses box	Monte Hall opens box	Contestant switches	Result
A	A	B or C	A for B or C	loses
A	B	C	B for A	wins
A	C	B	C for A	wins
B	A	C	A for B	wins
B	B	A or C	B for A or C	loses
B	C	A	C for B	wins
C	A	B	A for C	wins
C	B	A	B for C	wins
C	C	A or B	C for A or B	loses

Enumeration shows probability of winning is 6/9 = 2/3. If the contestant does not switch boxes, then his probability of winning the car remains unchanged (1/3) after Monte Hall opens an additional box.

PASCAL GENERALIZATIONS

Keats [1] has reminded us of the value and elegance of generalizations of Pascal's triangle to the multinomial case. A standard derivation of such formulae offers the notion of distributing labelled (i.e. distinguishable) objects to labelled pigeonholes, or sets. Thus, Keats' formula (3) easily follows from consideration of in which set the ultimate object is placed. There is another class of problems arising from the possibility that either or both of objects and sets are unlabelled. The latter case is equivalent to the problem of the partition of numbers, as when 5 is 'partitioned' into 2 + 2 + 1.

It is entertaining to demonstrate that the following are the same: the number of ways of dividing (a) $q + s$ indistinguishable objects into *exactly* s lots, (b) q objects into *at most* s lots, and (c) q objects into any number of lots, but with no more than s objects in each. Let us select a notation in which the common answer to these questions is S_q. Then, for example, 2_3 is two, the relevant partitions being (a):(4, 1), (3, 2); (b):(3), (2, 1); (c):(2, 1), (1, 1, 1). An equation known to Euler is

$$S_q = (S-1)_q + S_{(q-s)}.$$

It divides the distributions of q objects into at most s sets into two, those in which at most $(s-1)$ sets are occupied, and those in which, bearing in mind (a) above, the q objects are divided into exactly s lots. With this formula it is a straightforward exercise, as in the Pascal examples, to calculate a table of values for S_q.

There is also an elementary and attractive algorithm, which, curiously, appears to have been missed. By repeated application of the recurrence relation above we have

$$S_q = (S-1)_q + (S-1)_{q-s} + (S-1)_{q-2s} + \cdots$$

67

Monty Hall

$$\begin{aligned}
&\mathbb{P}(\text{trésor derrière l'autre porte})\\
&= \mathbb{P}(\text{trésor derrière l'autre porte}|X \text{ correct}) \cdot \mathbb{P}(X \text{ correct})\\
&+ \mathbb{P}(\text{trésor derrière l'autre porte}|X \text{ faux}) \cdot \mathbb{P}(X \text{ faux})\\
&= 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \frac{2}{3}
\end{aligned}$$

donc oui, il faut prendre l'autre porte.

Paradoxe des anniversaires

Considérons un ensemble de n personnes choisies au hasard. Si $n \geq 23$, il y a plus de 50% de chances qu'une paire d'entre elles ait le même anniversaire.

(en supposant que chaque jour de l'année a la même probabilité d'être un anniversaire)

$$A = \{\text{(au moins) une paire d'entre eux aura le même anniversaire}\}$$

$$\overline{A} = \{\text{aucune paire d'entre eux n'aura le même anniversaire}\}$$

$$\mathbb{P}(\overline{A}) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \frac{362}{365} \times \cdots \times \frac{343}{365} \simeq 49.2703\%$$

Approximation de Poisson, $\lambda = \dfrac{1}{365}\dbinom{23}{2} = \dfrac{253}{365} \simeq -0.6932$ i.e.

$$\mathbb{P}(X > 0) = 1 - \mathbb{P}(X = 0) \simeq 1 - e^{-0.6932} \simeq 0.500002$$

Test médicaux I

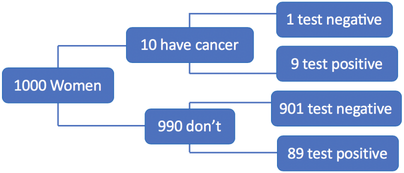
La probabilité qu'une femme ait un cancer du sein est de 1%. Si une femme a un cancer du sein, la probabilité que le test soit positif est de 90%. Si une femme n'a pas de cancer du sein, la probabilité qu'elle soit néanmoins positive est de 9%.

Une femme de 50 ans, sans symptômes, participe à un dépistage de routine par mammographie. Le test est positif, elle est alarmée et veut que vous lui disiez si elle est certaine d'avoir un cancer du sein ou quelles sont les chances qu'elle en ait un. En dehors des résultats du dépistage, vous ne savez rien d'autre sur cette femme. Combien de femmes dont le test est positif ont réellement un cancer du sein ? Quelle est la meilleure réponse ?

A) 9 sur 10
B) 8 sur 10
C) 1 sur 10
D) 1 sur 100

Test médicaux II

La probabilité qu'une femme ait un cancer du sein est de 1%. Si une femme a un cancer du sein, la probabilité que le test soit positif est de 90%. Si une femme n'a pas de cancer du sein, la probabilité qu'elle soit néanmoins positive est de 9%.



$$\mathbb{P}[\text{cancer}|\text{test positif}] = \frac{9}{9+89} \simeq \frac{1}{10}$$

See Do doctors understand test results? la moitié du groupe de 160 gynécologues a répondu que la probabilité pour la femme d'avoir un cancer était de neuf sur dix

Complément I

Ω peut être un ensemble relativement grand

- $\Omega = \{0, 1\}$ ou $\Omega = \{0, 1, \cdots, n\}$: fini
- $\Omega = \mathbb{N} = \{0, 1, \cdots, n, \cdots\}$ (entiers naturels) : dénombrable
- $\Omega = \mathbb{R}$ (réels) : non dénombrable