Probabilité des causes

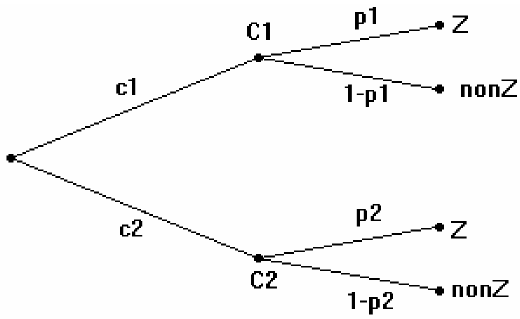
(Formule de Bayes)

En calcul des probabilités, le mot cause n'a pas du tout le sens que la langue commune lui attribue, ni son sens philosophique. Aussi, craignant quelque ambiguïté, certains auteurs ont proposé de parler de probabilité des antécédents : cela introduit dans la théorie des probabilités un paramètre chronologique qui a peu de raison de s'y trouver. Cournot a proposé « probabilité a posteriori », qui présente le même défaut.

Imaginons deux caisses remplies de pièces : l'une contient 999 pièces en chocolat et 1 pièce d'or. Dans l'autre, ces proportions sont inversées : 999 pièces d'or et 1 en chocolat. Quelqu'un tire au sort une caisse, puis dans cette caisse prend au hasard une pièce. Sans nous indiquer quelle caisse a été tirée, on nous informe que la pièce sortie est en chocolat. Certes, il est tout à fait possible que le choix de la seconde caisse en soit la « cause » : cependant, il semble clair que la caisse choisie est bien plus vraisemblablement la première. Est-ce vrai ?
Ce sont de telles intuitions que nous nous proposons d'étudier et de mesurer avec davantage de précision.
Posons le problème de façon abstraite, avec deux causes seulement pour commencer par un cas simple :
Deux causes C_1 et C_2 peuvent provoquer un événement Z avec des probabilités respectives p_1 et p_2. Les causes C_1 et C_2 ont des probabilités c_1 et c_2 d'intervenir ; si l'on suppose que Z s'est réalisé, quelles sont les probabilités respectives que ce soit par suite de l'intervention de la cause C_1 ou de la cause C_2 ?
La solution de ce problème est simple et n'exige que les règles élémentaires des probabilités.
Résolution en arbre :

$P(-, Z) = P[(C_1 \cap Z) \cup (C_2 \cap Z)] =$
$P(C_1 \cap Z) + P(C_2 \cap Z) = c_1 \cdot p_1 + c_2 \cdot p_2$;
d'où :

$$P(C_i / Z) = \frac{P(C_i \cap Z)}{P(-,Z)} = \frac{c_i \cdot p_i}{c_1 \cdot p_1 + c_2 \cdot p_2}$$



Cette formule est connue sous le nom de **formule de Bayes** : c'est en effet, dans un article posthume publié en 1763 dans les *Transactions philosophiques* que le révérend Thomas Bayes publia cette formule.

Avec trois causes ou davantage, on obtient de façon analogue une formule de Bayes de la forme :

$$P(C_i / Z) = \frac{P(C_i \cap Z)}{P(-,Z)} = \frac{c_i \cdot p_i}{c_1 \cdot p_1 + c_2 \cdot p_2 + c_3 \cdot p_3 + \ldots + c_n \cdot p_n} = \frac{c_i \cdot p_i}{\sum_{k=1}^{n} c_k \cdot p_k}$$

remarque : comme $p_i = P(Z / C_i)$, on a $P(C_i / Z) = \dfrac{P(C_i) \cdot P(Z / C_i)}{\sum_{k=1}^{n} c_k \cdot p_k}$ **Sa démonstration mathématique utilise exclusivement l'axiome des probabilités et le théorème 12.**

Problème résolu : On considère deux urnes C_1 et C_2 contenant des boules blanches et rouges.
La probabilité de tirer une boule rouge est 0,9 de l'urne C_1, de l'urne C_2 elle est de 0,3.
On tire au sort une urne et on en tire une boule ; on constate qu'elle est rouge. Quelle est la probabilité pour qu'elle provienne de l'urne C_1 ?
Réponse : Soit l'événement A : « tirer une boule rouge » :

on a $P(C_1 / A) = \dfrac{P(C_1 \cap A)}{P(-,A)} = \dfrac{\frac{1}{2} \cdot 0.9}{\frac{1}{2} \cdot 0.9 + \frac{1}{2} \cdot 0.3} = 0{,}75$; on interprète bien de façon plausible cette formule en parlant de « probabilités des causes », le choix de l'urne C_1 étant « cause » du tirage d'une boule rouge.

Autre exemple : **Le tricheur à l'écarté** : C'est à Henri Poincaré que l'on doit cet exemple, à la fois amusant et plein d'enseignements, d'utilisation de la formule de Bayes.
Pierre joue à l'écarté avec un inconnu qui, la première fois qu'il distribue les cartes , retourne le roi, qui, comme on sait, est à l'écarté la plus forte carte. Quelle est la probabilité pour que cet inconnu soit un tricheur ? Le problème ainsi posé est insoluble, il y manque de trop nombreuses données. Quelle est la probabilité pour qu'un tricheur retourne le roi ? Quelle est *à priori* la probabilité pour que cet inconnu avec qui Pierre engage la partie soit un tricheur ? Nous l'ignorons et il faut à ce sujet faire des hypothèses. Un tricheur ne saurait retourner trop souvent le roi sans courir le risque d'être trop vite démasqué. L'écarté se jouant avec un jeu de 32 cartes, un joueur honnête a pour probabilité 4/32, soit 1/8, de retourner un roi : admettons qu'un tricheur puisse le retourner deux fois plus souvent sans éveiller trop de soupçons ; nous prendrons donc 0,25 pour cette probabilité.
La probabilité *à priori* pour que l'inconnu soit un tricheur est bien plus difficile à évaluer : nous réservons cette question pour l'instant en notant *p* cette probabilité inconnue. La probabilité pour que l'inconnu ne soit pas un tricheur est alors *1-p* . Nous disposons maintenant de toutes les données nécessaires pour appliquer correctement la formule de Bayes : notons T l'événement « l'inconnu est un tricheur », H (pour « honnête ») l'événement contraire et R l'événement de la retourne du roi, la formule de Bayes s'écrit :

$$P(\ T\ /\ R) = \frac{P(T \cap R)}{P(-,R)} = \frac{\frac{1}{4}\cdot p}{\frac{1}{4}\cdot p + \frac{1}{8}\cdot(1\text{-}p)} = \frac{2p}{1+p}$$

. Ainsi, chaque fois que *p* est connu, on en déduit par cette formule très simple la probabilité *a posteriori* (après la retourne d'un roi) pour que l'inconnu soit un tricheur.

- Si p = 0, c'est à dire si Pierre a la certitude absolu que l'inconnu est honnête, P(T / R) est nulle également. En forme plaisante, on peut dire qu'une certitude absolue est inébranlable.
- Si Pierre est très légèrement méfiant, par exemple p = $\frac{1}{1000}$, alors à très peu près

  P(T / R) = $\frac{2}{1000}$, autrement dit, la retourne d'un roi « double » la méfiance de Pierre.
- Si enfin l'indécision de Pierre est totale, c'est à dire s'il estime que la probabilité p pour que son adversaire soit un tricheur est égale à $\frac{1}{2}$, la formule ci-dessus donne :

$$P(\ T\ /\ R) = = \frac{2p}{1+p} = \frac{2\cdot\frac{1}{2}}{1+\frac{1}{2}} = \frac{2}{3}\ ,$$

c'est que la première retourne augmente considérablement la méfiance de Pierre ; mais comme le fit observer Poincaré, il est bien léger d'engager une partie contre un adversaire dont on a si piètre opinion.
D'autres hypothèses sont possibles, conduisant à des résultats chaque fois différents : la formule de Bayes ne donne pas de probabilités des causes, elle les modifie à la lumière de l'expérience.
Cet exemple du tricheur nous introduit à une propriété très remarquable de la formule de Bayes : elle permet de modifier des jugements de probabilité, non d'en créer. Il faut, pour qu'on puisse l'appliquer, que les probabilités *a priori* d'interventions des diverses causes soient connues.
Exercices :
a) Résoudre le problème d'introduction sur les pièces en or et en chocolat.
b) **Examen ou loterie** ? Un étudiant répond à un questionnaire par oui ou non.
S'il connaît la réponse, sa marque est exacte ; s'il ignore la réponse, il joue la marque à pile ou face. Sachant qu'il connaît les trois cinquièmes du programme, quelle est la probabilité pour qu'une réponse juste soit due à ses connaissances et non au hasard ?
(Sauriez-vous donner une réponse intuitive ?)
c) Une boîte contient six boules dont on ignore si elles sont blanches ou noires.
On y effectue six tirages, en remettant chaque fois la boule tirée et observée dans la boîte ; on obtient toujours une boule blanche. Quels jugements de probabilité peut-on porter sur la composition de la boîte ? Résoudre le problème selon les deux hypothèses :
1) toutes les compositions possibles sont équiprobables ;
2) (hypothèse de Borel) La boîte a été composé en tirant à pile ou face, et indépendamment pour chaque boule, la couleur qu'on y peindrait.

EXEMPLES

1) Un élève répond à une question à choix multiple. De deux choses l'une :soit il connaît la réponse, soit il la devine. Soit *p* la probabilité que l'élève connaisse la réponse et donc *1– p* celle qu'il la devine. On admet que l'élève qui devine répondra correctement avec probabilité 1/*m* où *m* est le nombre de réponses proposées.
Quelle est la probabilité pour qu'un élève connaisse la réponse à une question s'il y a répondu correctement ?
Soient les événements E : « il connaît vraiment la question » et
F : « l'étudiant répond correctement à la question ».

Alors $P(E / F) = \dfrac{1 \cdot p}{1 \cdot p + \frac{1}{m} \cdot (1-p)} = \dfrac{1 \cdot p}{1 \cdot p + (m-1) \cdot p}$

En prenant par exemple m=5 et p=1/2, la probabilité qu'un élève connaisse la réponse à une question sachant qu'il a répondu correctement sera ainsi 5/6.
En prenant par exemple m=5 et p=3/5, la probabilité qu'un élève connaisse la réponse à une question sachant qu'il a répondu correctement sera ainsi 15/17.
En prenant par exemple m=5 et p = x, si la probabilité qu'un élève connaisse la réponse à une question sachant qu'il a répondu correctement est 9/10, alors x = 9/14 ≈ 0,643.
En prenant par exemple m=5 et p = x, si la probabilité qu'un élève connaisse la réponse à une question sachant qu'il a répondu correctement est 99/100, alors x = 99/104 ≈ 0,952.

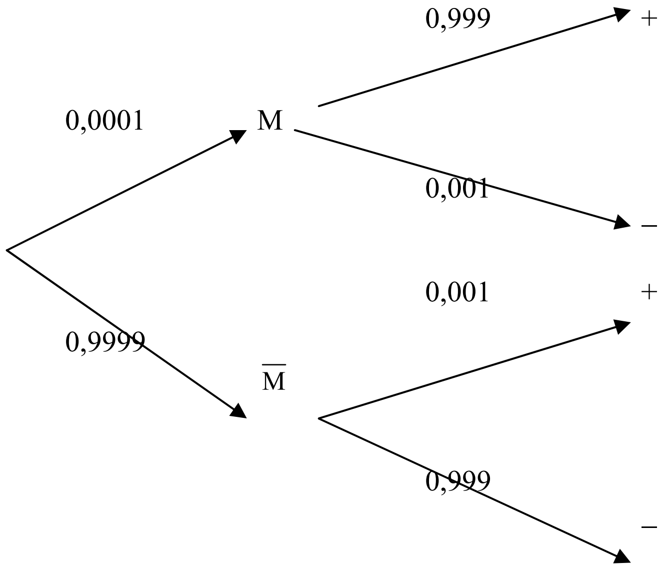
Probabilité des causes et Sida

(Formule de Bayes)

" Prenons par exemple une maladie rare qui atteint, disons 1 individu sur 10 000 (*ce peut être le sida ou la vache folle*). On met au point un test pour détecter si un individu est infecté par la maladie. Ce test donne des résultats fiables dans 99,9 % des cas: c'est donc à priori un excellent test. Supposons à présent qu'on pratique le test sur un individu pris au hasard et que ce test soit positif. Quelle est la probabilité que l'individu ainsi testé soit effectivement infecté ? Le calcul des probabilités nous répond sans hésiter 10 %
(9% exactement). Autrement dit, alors que le test est positif, l'individu a 90 % de chances d'être sain .
Si dans notre exemple, la maladie atteignait 1 % des individus, le test positif sur un individu pris au hasard l'informerait qu'il a plus de 90 % de chances d'être malade.
Les tests de dépistage ne sont efficaces que pour les maladies fréquentes ou les épidémies. Dans le cas de maladies rares, contrairement à ce que disent les bons esprits, ("Le test n'est pas parfait mais ça vaut mieux que rien"), il faut que les tests soient parfaits. Si on applique un test excellent mais imparfait pour dépister la maladie de la vache folle et qu'on abatte les vaches testées positives, on va déclencher un massacre bovin généralisé, ruineux et inutile ."

L'Express, 1 au 7 mars 2001, Claude Allègre

Explications :
Notons par M, + et – les événements respectifs : "être malade", "le test est positif", "le test est négatif" .



Le texte suppose le calcul de P((M,-) / (-,+)) :

$$P((M,-) / (-,+)) = \frac{P(M\cap +)}{P(+)} = \frac{P(+/M)\times P(M)}{P(+)}$$

$$= \frac{P(+/M)\times P(M)}{P(+/M)\times P(M) + P(+/\overline{M})\times P(\overline{M})} = \frac{0,999\times 0,0001}{0,999\times 0,0001 + 0,001\times 0,9999} \approx 0,0908$$

L'autre résultat s'obtient de même avec P(M) = 0,01 et P($\overline{M}$) = 0,99 et on obtient alors environ :

P((M,-) / (-,+)) = 0,9098.