Illustration : Noémie Ross

Thomas Bayes
1702-1761

Le mathématicien britannique et pasteur de l'Église presbytérienne, Thomas Bayes est surtout connu pour avoir formulé le théorème de Bayes utilisé de nos jours en classement automatique et dans la lutte contre les pourriels, par la méthode dite d'*inférence bayésienne*[1] utilisée en intelligence artificielle.

Thomas Bayes

Le père de Thomas Bayes, Joshua est un des premiers pasteurs non-conformistes[2] à être ordonné en Angleterre. Thomas est l'aîné d'une famille de sept enfants, quatre garçons et trois filles. Étant issu d'une famille non-conformiste, il ne pouvait s'inscrire dans une école régulière et on croit qu'il a reçu une éducation privée. À l'époque, un non-conformiste ne pouvait être admis aux universités d'Oxford et de Cambridge. Thomas choisit donc d'étudier la théologie à l'Université d'Édimbourg où il est admis en 1719, il part pour l'université d'Édimbourg, afin d'étudier la théologie. Ordonné pasteur non-conformiste, il travaille d'abord comme assistant de son père dans la paroisse de Horborn. En 1733, il devient ministre de la Chapelle presbytérienne de Tunbridge Wells, près de Londres, poste qu'il occupe de ces occupations jusqu'à sa retraite en 1752. Il demeure cependant à Tunbridge Wells, jusqu'à son décès, le 7 avril 1761. Durant ce ministère, il publie *La Bienveillance divine, ou une tentative de preuve que la fin première de la Providence divine et du Gouvernement est le Bonheur de ses créatures*, en 1731. Il rédige aussi *Une introduction à la doctrine des fluxions, et une défense des mathématiciens contre les objections faites à l'auteur de l'Analyse*[3], publié anonymement en 1736. Dans ce texte, il défend les bases du calcul infinitésimal établies par Isaac Newton.

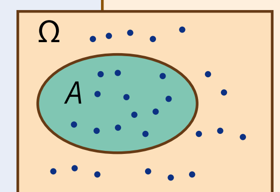
Parallèlement à ses activités de pasteur presbytérien, Bayes poursuit des travaux en mathématiques. Même s'il n'a publié aucun article dans ce domaine au cours de sa vie, sauf anonymement, il devient membre de la Royal Society en 1742.

Les travaux de Bayes en probabilités ont été résumées dans son *Essai sur la manière de résoudre un problème dans la doctrine des risques* (*Essay Towards Solving a Problem in the Doctrine of Chances*) publié en 1763 à titre posthume dans les comptes-rendus de l'Académie royale de Londres (*The Philosophical Transactions of the Royal Society*). Cette publication est l'œuvre de l'un de ses amis, Richard Price à qui il avait légué ses écrits mathématiques. Ce texte comporte un résultat de base en théorie des probabilités, appelé *théorème de Bayes*. L'encadré ci-dessous explique de quoi il s'agit.

1. L'inférence bayésienne est une méthode d'inférence par laquelle on calcule les probabilités de diverses causes hypothétiques à partir de l'observation d'événements connus.
2. En Angleterre, les *non-conformistes*, appelés aussi *Dissidents*, refusaient de suivre la doctrine de l'Église anglicane. C'est le cas des puritains, des presbytériens ou calvinistes, des anabaptistes et plus tard des quakers. Ce vocable est apparu vers 1566, durant le règne d'Élisabeth I[re], lorsque l'archevêque de Cantorbéry, Matthew Parker, voulut forcer les ecclésiastiques à porter des ornements liturgiques. Ne pouvant occuper un emploi public, civil ou militaire, plusieurs non-conformistes ont choisi d'émigrer vers les colonies, ce qui fut encouragé par le pouvoir royal anglais afin de restreindre leur influence et pour peupler les colonies. Cette politique a permis d'augmenter rapidement la démographie de la Nouvelle-Angleterre et d'assurer la domination anglo-saxonne en Amérique du nord.
3. L'*Analyste* est une critique des fondations de la science et tout particulièrement du calcul infinitésimal publiée en 1734, par l'évèque anglican irlandais George Berkeley. Cette critique a amené les mathématiciens à consolider les fondements de leur science et en particulier ceux du calcul différentiel et intégral, ou calcul infinitésimal.

Théorème de Bayes

La probabilité d'un événement A est le rapport du nombre de cas favorables au nombre de cas possibles. En associant les événements à des ensembles, l'ensemble de cas possibles est noté Ω, et l'ensemble des cas favorables est noté A, la probabilité de l'événement A est le nombre d'éléments de l'ensemble A divisé par le nombre d'éléments dans l'ensemble Ω, soit :



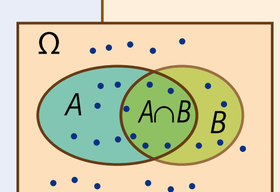
$$\Pr(A) = \frac{n(A)}{n(\Omega)}.$$

Ainsi, en lançant un dé bleu et un dé rouge, le nombre de cas possibles est $n(\Omega) = 36$. Si on s'intéresse à l'événement « la somme des faces obtenues est plus grande ou égale à 8 », le sous-ensemble associé à cet événement contient quinze éléments.

	1	2	3	4	5	6
1	(1;1)	(1;2)	(1;3)	(1;4)	(1;5)	(1;6)
2	(2;1)	(2;2)	(2;3)	(2;4)	(2;5)	(2;6)
3	(3;1)	(3;2)	(3;3)	(3;4)	(3;5)	(3;6)
4	(4;1)	(4;2)	(4;3)	(4;4)	(4;5)	(4;6)
5	(5;1)	(5;2)	(5;3)	(5;4)	(5;5)	(5;6)
6	(6;1)	(6;2)	(6;3)	(6;4)	(6;5)	(6;6)

Par conséquent, la probabilité d'obtenir une somme plus grande ou égale à 8 en lançant deux dés est :

$$\Pr(A) = \frac{n(A)}{n(\Omega)} = \frac{15}{36}.$$



Supposons maintenant que l'on vous informe que la somme plus grande ou égale à 8 et on vous demande de calculer la probabilité que les deux dés présentent la même face.

Le diagramme à droite permet de constater que le nombre de cas possibles est alors $n(A) = 15$ et l'ensemble des cas favorables est $n(A \cap B) = 3$, on a donc :

$$\Pr(B \mid A) = \frac{n(A \cap B)}{n(A)} = \frac{3}{15} = \frac{1}{5}.$$

	1	2	3	4	5	6
1	(1;1)	(1;2)	(1;3)	(1;4)	(1;5)	(1;6)
2	(2;1)	(2;2)	(2;3)	(2;4)	(2;5)	(2;6)
3	(3;1)	(3;2)	(3;3)	(3;4)	(3;5)	(3;6)
4	(4;1)	(4;2)	(4;3)	(4;4)	(4;5)	(4;6)
5	(5;1)	(5;2)	(5;3)	(5;4)	(5;5)	(5;6)
6	(6;1)	(6;2)	(6;3)	(6;4)	(6;5)	(6;6)

En divisant le numérateur et le dénominateur par $n(\Omega)$, on obtient:

$$\Pr(B \mid A) = \frac{n(A \cap B)/n(\Omega)}{n(A)/n(\Omega)} = \frac{\Pr(A \cap B)}{\Pr(A)}.$$

On en tire $\Pr(A \cap B) = \Pr(B \mid A)\Pr(A)$.

On peut aussi calculer la probabilité que la somme soit plus grande ou égale à 8 sachant que les deux dés présentent la même face. Dans ce cas, le nombre de cas possibles se restreint à 6 et le nombre de cas favorables est 3. Par conséquent :

$$\Pr(A \mid B) = \frac{n(A \cap B)}{n(B)} = \frac{3}{6} = \frac{1}{2}.$$

On a donc aussi $\Pr(A \cap B) = \Pr(A \mid B)\Pr(B)$.

La symétrie de ces deux expressions de $\Pr(A \cap B)$ donne l'énoncé du théorème de Bayes, soit :

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

En se servant du fait que $\Pr(B)$ s'écrit également :

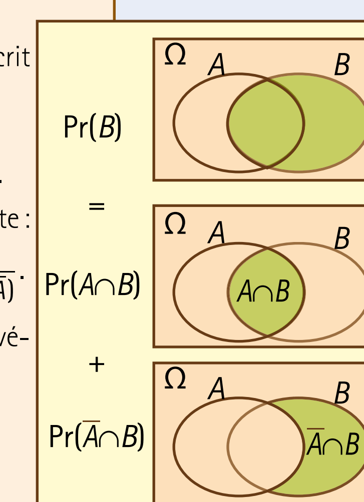
$$\Pr(B) = \Pr(A \cap B) + \Pr(\bar{A} \cap B)$$
$$= \Pr(B \mid A)\Pr(A) + \Pr(B \mid \bar{A})\Pr(\bar{A}).$$

On obtient une formulation équivalente :

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B \mid A)\Pr(A) + \Pr(B \mid \bar{A})\Pr(\bar{A})}.$$

On peut généraliser pour plusieurs événements A_i, ce qui donne :

$$\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i)\Pr(A_i)}{\sum_j \Pr(B \mid A_j)\Pr(A_j)}.$$



Retour, notes historiques