Le révérend Thomas Bayes ministre du culte presbytérien est né en Angleterre en 1701 ou 1702. Il est censé avoir mené une vie paisible de célibataire studieux, comme souvent à cette époque (on peut en douter à voir le célibat marqué de Lewis Carroll), et publie de son vivant «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 », tout un programme. A sa mort, son vieux pote Richard Price soutient la publication en 1763 d'un texte posthume « Essai en vue de résoudre un problème de la doctrine des sciences » qui pose ce qu'il est désormais convenu d'appeler le théorème de Bayes.

Ce théorème redécouvert en 1774 par le Marquis de Laplace - lequel fût le premier à prévoir les mouvements de planètes et insistait lourdement auprès de Napoléon Bonaparte sur le fait qu'il n'avait nullement besoin d'invoquer Dieu pour ce faire – ce théorème donc dit de Bayes correspond à une équation d'apprentissage statistique dont les applications très larges dépassent le cadre des mathématiques. On l'utilise dès que l'information disponible est limitée ou bien difficile à rassembler. Rien d'étonnant à ce que l'industrie pharmaceutique (développement de molécules) et les big-oils (prospection pétrolière) en soient friands et figurent parmi les utilisateurs intensifs après que les mathématiciens dans les années 1980 aient reconnus des fondements « sains » à l'approche. Il faut noter également que la nouvelle économie avec les perspectives de gigantesques bases de données à traiter (Data Mining) constitue un champ d'application fabuleux pour une méthode capable d'apprentissage.

Pour les moins matheux d'entre nous:

Le théorème de Bayes prend en compte la vraisemblance d'un évènement conjointement à sa probabilité d'occurrence.

Il s'agit d'un raisonnement inverse où l'on cherche à déterminer la probabilité (vraisemblance) d'une hypothèse compte tenu d'un ensemble d'observations: **si l'on connait les conséquences d'une cause (ou d'un ensemble de causes), l'observation des effets produits permet de remonter aux causes probables.**

Pour les autres:

Le théorème de Bayes prend l'aspect d'une petite, toute petite formule qui permet d'évaluer la pertinence de ce que l'on croit savoir – une hypothèse (H) - à l'aune de l'information apportée par une observation (O).
On appelle :

- P(O) et P(H) les probabilités respectives de H et O
- P(O|H) la probabilité de O sachant H : si l'hypothèse H est vraie alors on devrait observer O avec une certaine probabilité
- P(H|O) la probabilité de H sachant O, résultat de la formule de Bayes constitue une mesure de la pertinence de l'hypothèse H ayant observé O – ceci permettra de classer les différentes hypothèses possibles en vue de déterminer la plus vraisemblable (et non pas la plus probable).

La formule de Bayes s'écrit ainsi $P(H|O) = \frac{P\,(H)*P(O|H)}{P(O)}$

Dans les probabilités classiques, on cherche à résoudre les problèmes directs :

⇨ étant donné une urne avec 6 balles blanches et 4 noires, quelle est la probabilité de tirer 3 noires.

Le problème que résout Bayes concerne l'inversion du raisonnement :

- ⇨ étant donné un tirage, que peut-on dire sur le contenu de l'urne ? Autrement dit, quelle est la probabilité des tirages suivants?

  Par exemple, pour une urne dont on ne sait rien :

- cinq tirages successifs d'une boule blanche conduisent à une vraisemblance FORTE d'un sixième tirage blanc sans qu'il soit possible de prouver que l'urne ne contienne pas de boule rose ou bleue…. ou verte.
- quatre tirages successifs d'une boule blanche suivis d'une cinquième noire conduisent à une vraisemblance un peu moins forte de blanc et une vraisemblance faible de noir On voit bien dans un tel contexte la capacité d'apprentissage à l'œuvre.

Pour aider un peu, un exemple emprunté à Stanislas Dehaene (professeur de Neuro Sciences):

Un jeune patient rend visite à son médecin en Novembre parce qu'il tousse (O).

Trois hypothèses

H1=il a la grippe.

H2=il a un cancer du poumon

H3=il a une gastro-entérite.

Théorème de Bayes: $p(H|O) = p(O|H) * p(H) / p(O)$

En l'occurrence, $p(O)= 1$ car le patient tousse, c'est un fait

	P(O|H) = P(tousse selon l'hypothèse) Vraisemblance/Causalité	P(H) = Probabilité (fréquence) de la pathologie chez les jeunes patients	Résultat P(H|O) : s'apparente au Diagnostic
H1 grippe	Forte (la grippe fait tousser)	Forte (fréquent en hiver)	Forte
H2 cancer	Forte (le cancer du poumon fait tousser)	Faible (patient jeune)	Faible
H3 gastro	Faible (la gastro fait rarement tousser)	Forte (communément répandue lorsqu'on ne se lave pas les mains☺)	Faible

Conclusion : le patient à la grippe

Cet exemple reste simple car il y a alignement entre la probabilité et la vraisemblance de l'évènement. Mais concernant des phénomènes plus complexes, l'application du théorème de Bayes devient subtile. Dans la pratique (et reprenant en cela une vieille idée de Turing), on travaille plutôt sur des logarithmes qui caractérisent l'évidence (mesurée en déciban tout comme le bruit en décibel), en fait une mesure de la vraisemblance d'une hypothèse par rapport aux autres.

Quelques champs d'application originaux :

- ce n'est plus un secret que la résolution des messages cryptés de la machine Enigma lors de la seconde guerre mondiale par Alan Turing reposait sur une application judicieuse du théorème de Bayes (ceci étant, Turing fut grandement remercié pour sa contribution sous forme de castration chimique - l'homosexualité affichée dans les années cinquante n'était pas bien vue chez nos cousins grand-bretons),
- les sciences cognitives commencent à percevoir le cerveau, tant du point de vue de l'apprentissage (capacité d'inférence Bayésiennes avérées chez l'enfant de huit mois), de la perception (capacité de notre système nerveux à ne nous donner « à voir » qu'une seule représentation du monde à un instant donné), que du point de vue de vieux débats (genre inné-acquis) comme reposant en grande partie sur des modes Bayésiens de traitement de l'information,
- la recherche du lieu de crash de l'AF 447 (mais en fait, la quasi-totalité des systèmes de recherche fonctionnent sur les mêmes principes),
- et même Henri Poincaré, en charge d'une énième expertise du fameux bordereau dans le cadre du procès en réhabilitation de Dreyfus, utilisa dans ses raisonnements (à l'inverse de ce plouc de Bertillon) ce que l'on pourrait appeler une méthode Bayésienne.

Reste l'extension du théorème de Bayes qui nous intéresse spécifiquement et sur lequel nous souhaitons vos avis – par mail ou par commentaire sur le blog.

Nous avons sélectionné un certain nombre d'assertions, issues de Sciences dures ou molles, qui illustrent une absence de preuve (probabilité faible ou nulle) mais une vraisemblance forte basée sur des observations répétées ou bien l'inverse.

Est-il valide d'utiliser une approche Bayésienne et si oui, quelle est votre conclusion ?

1 le dérèglement climatique est d'origine anthropique : Probabilité faible et vraisemblance forte.

2 dès 1935, l'Allemagne se prépare à envahir la France : Probabilité forte et vraisemblance (perçue par les gouvernements) faible.

3 les pesticides sont responsables de la disparition des abeilles : Probabilité faible et vraisemblance forte [nous traiterons spécifiquement des Abeilles dans le prochain post]

Nous sommes également intéressés de vos propres exemples…