

# LE RAISONNEMENT BAYESIEN

Spécialisé dans le traitement statistique des études cliniques,  
le Département Biométrie de FOVEA vous propose  
au travers de ce fascicule quelques notions simples  
concernant LE RAISONNEMENT BAYESIEN.

*en  
pratique*

*Traitement statistique*  
DES ETUDES CLINIQUES

NF EN ISO 9001 Ed. 2000



N°0071-2003-FR-COFRAC-QUA



*Le théorème des probabilités conditionnelles, établi par le Révérend Thomas BAYES (1702-1761), a été publié à titre posthume en 1763 sous le nom de « Probabilité des hypothèses ». De manière indépendante, LAPLACE apporte des résultats sur le même sujet dans son « Mémoire sur la probabilité des causes par les événements » publié en 1774.*

*Ce théorème peut s'illustrer ainsi :*

**Quelle est la probabilité *a posteriori***

**d'observer un événement A, en fonction de sa probabilité *a priori*, sachant qu'un événement B s'est produit ?**

*Le théorème de BAYES est à l'origine d'une partie de la **théorie moderne de la décision**. C'est un outil de modélisation de la demande diagnostique du médecin. Il est également à la base d'autres statistiques, notamment **des méthodes séquentielles et du test triangulaire**.*

## 1

### Probabilités *a priori* et *a posteriori*

La portée pratique du raisonnement bayésien peut être illustrée au travers des exemples ci-après.

Sachant que la prévalence de l'hépatite C est en France de 1% (1995), un médecin peut supposer qu'il y a une chance sur cent pour que le prochain patient entrant dans son cabinet soit VHC+, si bien sûr sa clientèle est à l'image de la population française. Dans le cas où ce patient est un enfant de 10 ans, la probabilité qu'il soit VHC+ tombe à environ 0,001%. Si par contre, ce patient est toxicomane, usage de drogues par voie intraveineuse depuis plus de 5 ans, les données épidémiologiques indiquent une probabilité d'être VHC+ supérieure à 50%. A l'inverse, s'il s'agit seulement d'un adulte asthmatique, le médecin peut maintenir son estimation du risque à 1%, en effet il n'y a

pas de relation entre l'asthme et le statut VHC.

Dans ces trois exemples, deux apportent des informations complémentaires - âge égal à 10 ans et toxicomanie - pouvant modifier la probabilité de 1% d'être VHC+. Le troisième exemple - adulte asthmatique - n'apporte pas ce type d'information.

Le calcul de la **probabilité conditionnelle** permet de déterminer de nouvelles probabilités d'être VHC+, **en tenant compte**, dans les deux premiers exemples, **des informations complémentaires modifiant la probabilité initiale**. Le troisième exemple illustre la **notion d'indépendance** de deux événements : asthme et statut VHC.

Le raisonnement se décline ainsi en **2 étapes** :

■ **1<sup>ère</sup> étape** : Elle comporte une certaine incertitude quantifiée par une **probabilité *a priori***. « *A priori* » signifie qu'il s'agit d'une **probabilité avant information complémentaire corrective** :  $P(\text{VHC}+) = 1\%$ .

■ **2<sup>ème</sup> étape** : Après recherche d'informations complémentaires, la probabilité peut être affinée. Cette **probabilité affinée** est la **probabilité *a posteriori*** :  $P(\text{VHC}+) = 0,001\%$  dans le cas d'un enfant de 10 ans,  $P(\text{VHC}+) = 50\%$  dans le cas d'un adulte toxicomane.

Dans ces exemples, on passe d'une **probabilité *a priori*, primaire ou pré-test**, équivalente à la **prévalence de la maladie observée dans la population générale**, à une **probabilité conditionnelle**. La probabilité *a priori* n'est **dépendante d'aucun facteur**, c'est une **probabilité non conditionnelle**. Croisée avec la probabilité d'un ou de plusieurs facteurs, cette probabilité *a priori* devient une **probabilité conditionnelle, *a posteriori*, secondaire ou post-test**.

Cette probabilité conditionnelle correspond par conséquent à une révision de la probabilité *a priori* à travers « le filtre » d'un ou de plusieurs facteurs, censée apporter une plus grande exactitude aux données étudiées.

# 2

## Probabilité conditionnelle

La probabilité conditionnelle de survenue d'un événement A, sachant que l'événement B s'est produit, est définie par :

$$P(A/B) = \frac{P(A \text{ et } B)}{P(B)}$$

$P(A/B)$  est la probabilité *a posteriori* de A sachant que B s'est produit.  $P(A \text{ et } B)$  est la probabilité que A et B se produisent simultanément.  $P(B)$  est la probabilité *a priori* de B.

La formule donnant  $P(A/B)$  peut également s'écrire :

$$P(A \text{ et } B) = P(A/B) \times P(B)$$

En échangeant A et B dans la formule ci-dessus, le 1<sup>er</sup> terme reste inchangé :  $(A \text{ et } B) = (B \text{ et } A)$  et la formule devient :

$$P(B \text{ et } A) = P(B/A) \times P(A)$$

$P(B/A)$  est la probabilité *a posteriori* de B sachant que A s'est produit.  $P(A)$  est la probabilité *a priori* de A avant que l'on sache si B s'est produit ou non.

D'où :

$$P(A \text{ et } B) = P(A/B) \times P(B) \\ = P(B/A) \times P(A)$$

La symétrie de la notation correspond à des modes différents de recueil de l'information. Par exemple,  $P(\text{VHC+}/\text{Toxicomanie})$  peut

être déterminée dans la population de toxicomanes chez lesquels on détermine le statut VHC alors que  $P(\text{Toxicomanie}/\text{VHC+})$  sera calculée dans une population de patients VHC+ chez lesquels on recherche une éventuelle toxicomanie.

# 3

## Indépendance de l'information

### ■ 1<sup>ère</sup> définition

« Deux événements, A et B, sont dits indépendants lorsque la connaissance de la survenue de l'un d'entre eux n'apporte aucune d'information sur la probabilité de survenue de l'autre. »

Ainsi, un événement A est dit indépendant d'un événement B lorsque la connaissance de la survenue de l'événement B ne modifie pas la probabilité de l'événement A. Dans l'exemple 3 ci-avant, l'asthme et le statut VHC sont deux événements indépendants. En effet, savoir que le patient présente l'une de ces affections n'apporte aucune information sur sa probabilité de présenter l'autre.

Donc, la probabilité *a priori* de A est identique à la probabilité *a posteriori* de A.

### ■ 2<sup>ème</sup> définition

« Deux événements, A et B, sont indépendants

lorsque la probabilité de survenue de l'événement (A et B) est égale au produit des probabilités de survenue de A et de survenue de B. »

Ces définitions se traduisent toutes deux par une formule.

La 1<sup>ère</sup> définition de l'indépendance de A et B s'écrit :

$$P(A/B) = P(A)$$

La seconde définition s'écrit :

$$P(A \text{ et } B) = P(A) \times P(B)$$

En pratique, on en déduit que si l'on démontre que A n'apporte pas d'information sur B, on démontre alors par **réciprocité** que B n'apporte pas d'information sur A. Dans l'exemple VHC+, savoir que la probabilité d'être VHC+ est la même chez les asthmatiques, c'est automatiquement en déduire que la probabilité d'asthme est la même chez les patients VHC+ et les patients VHC-.

# 4

## Formulation du théorème de BAYES

Le théorème de BAYES découle directement de la définition de la probabilité conditionnelle. Il vise à calculer la probabilité *a posteriori* d'un événement en fonction de la probabilité *a priori* de ce dernier. « A priori » et « a posteriori » s'entendent par rapport à la connaissance d'une information.

L'idée de base du théorème de BAYES est de calculer la probabilité *a posteriori* d'un événement A, en fonction de sa probabilité *a priori*, sachant qu'un autre événement B s'est produit. On cherche donc à exprimer  $P(A/B)$  en fonction de  $P(A)$ .

Si A et B sont indépendants, on a :

$$P(A/B) = P(A)$$

Si A et B ne sont pas indépendants, c'est-à-dire si B apporte une information :

$$P(A/B) = P(A) \text{ et } P(A/\bar{B}) = P(A/\bar{B})$$

$\bar{B}$  : Non réalisation de l'événement B

La probabilité *a posteriori* de A sera différente selon que l'on dispose de l'information que B s'est produit ou que B ne s'est pas produit.

On peut écrire :

$$P(A/B) = \frac{P(A \text{ et } B)}{P(B)} = \frac{P(B/A) \times P(A)}{P(B)}$$

$P(B)$  peut s'écrire  $P(B \text{ et } A) + P(B \text{ et } \bar{A})$ .

En effet, si les éléments A et  $\bar{A}$  sont mutuellement exclusifs, la probabilité de B peut être la somme des probabilités composées de B et de A.

On peut donc écrire :

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B \text{ et } A) + P(B \text{ et } \bar{A})}$$

De plus :

$$P(B \text{ et } A) = P(B/A) \times P(A) \\ \text{et} \\ P(B \text{ et } \bar{A}) = P(B/\bar{A}) \times P(\bar{A})$$

Le théorème de BAYES s'écrit :

$$P(A/B) = \frac{P(B/A) \times P(A)}{P(B/A) \times P(A) + P(B/\bar{A}) \times P(\bar{A})}$$

## 5

### Illustration pratique du théorème de base

Supposons que A soit l'événement « Atteint de la maladie M » et que B soit l'événement « Positif pour le test diagnostique D », le théorème de BAYES permet de calculer la probabilité pour qu'un patient positif pour le test diagnostique D soit atteint de M, en fonction de la probabilité (ou fréquence) de la maladie M dans la population.

En notant M le fait d'être atteint de la maladie,  $\bar{M}$  le fait de ne pas en être atteint,  $D^+$  celui d'être positif pour le test diagnostique et  $D^-$  d'être négatif, on obtient :

$$P(M/D^+) = \frac{P(D^+/M) \times P(M)}{P(D^+/M) \times P(M) + P(D^+/\bar{M}) \times P(\bar{M})}$$

Il sera possible de calculer la probabilité recherchée si l'on connaît la fréquence de sujets positifs chez les malades [ $P(D^+/M)$ ] et les non malades [ $P(D^+/\bar{M})$ ], ce qui pose les limites du raisonnement bayésien.

## 6

### Les conditions du raisonnement bayésien

Le raisonnement bayésien s'applique à **des propositions dites « catégorielles »**, c'est-à-dire distinctes les unes des autres sans dimensions intermédiaires.

L'utilisation du théorème de BAYES implique le respect de **3 conditions**, qui en pratique limitent son usage.

■ **Exclusion réciproque des informations et des hypothèses diagnostiques**

La présence de la maladie M exclut l'absence de la maladie M et réciproquement.

■ **Exhaustivité des hypothèses**

La liste des événements possibles, par exemple celle des hypothèses diagnostiques proposées, doit inclure toutes les éventualités de sorte que la somme des probabilités respectives soit égale à 1.

■ **Indépendance mutuelle des informations dont les effets sont analysés par le théorème de BAYES**

Deux événements sont dits indépendants quand la probabilité de l'un n'est pas modifiée par la présence ou l'absence de l'autre.

## 7

**Limites au raisonnement bayésien**

Si l'intérêt du raisonnement bayésien est de rendre explicites des informations incertaines ou approximatives, certains auteurs en ont fixé les limites.

■ 1<sup>ère</sup> limite

Les limites du raisonnement bayésien relèvent de **l'imprécision de l'information clinique et épidémiologique** et plus encore de **l'insuffisance méthodologique** (Berwick et coll., 1981).

L'information dont dispose le médecin est incertaine ou supposée telle pour les 3 composantes de la formule bayésienne :

- La probabilité primaire,
- La probabilité du signe dans la maladie cherchée,
- Et surtout la probabilité du signe parmi les patients qui n'ont pas la maladie cherchée (Schwartz, 1979).

■ 2<sup>ème</sup> limite

**La condition d'indépendance des probabilités conditionnelles** quand plusieurs signes ou tests sont exploités conjointement est nécessaire.

■ 3<sup>ème</sup> limite

- **Raisonnement probabiliste** impropre à la décision médicale (Redelmeier et Tversky, 1990)

- La critique qui est faite de l'approche probabiliste est qu'elle n'offre **aucune certitude** alors que la décision prise ou proposée par le médecin doit être définie et précise (Kora, 1975 - Kassirer et Pauker, 1981)

- **Le concept de probabilité** dans le sens d'une estimation du degré de certitude ou d'incertitude n'est **pas d'usage courant dans le langage clinique ni dans le raisonnement décisionnel** (Sox, 1986)

## 8

**Conclusion**

En conclusion, malgré les contraintes d'utilisation et les réserves de ses détracteurs, l'intérêt du raisonnement bayésien est de **rendre explicites des informations manquantes, incertaines ou approximatives.**

Lorsque les adversaires du raisonnement bayésien invoquent que ce dernier est impropre à la décision médicale, dans la mesure où « toute approche probabiliste est dénuée de sens pour chaque patient dans sa singularité (Redelmeier et Tursky, 1990) », ils oublient bien sou-

vent que le médecin ne peut résoudre son incertitude diagnostique qu'au travers de l'expérience des autres.

L'expérience ainsi que les fréquences observées constituent les bases de son raisonnement et de sa décision. C'est en fonction de ces probabilités qu'il impose à son patient les risques d'un diagnostic et des prescriptions en découlant.

Dans la majorité des situations banales, les conséquences des décisions du médecin sont jugées mineures et les décisions intuitives ont le mérite de la rapidité et souvent celui de l'efficacité. Par contre, c'est dans les décisions graves et difficiles que l'analyse explicite du raisonnement prend toute sa valeur.

« Il n'en reste pas moins qu'aucune décision n'est anodine : même les décisions les plus « bénignes » en apparence gagnent à être éclairées par un raisonnement rationnel et explicite. » (Timmerman, 1994)

# LE RAISONNEMENT BAYESIEN EN PRATIQUE



## Bibliographie

1. **Baillargeon G.**  
Techniques statistiques  
Editions S.M.G. 1984
2. **Grenier B.**  
Evaluation de la  
décision médicale  
Collection Evaluation  
et Statistique  
Masson 1996
3. **Hymans S.H.**  
Probabilités, écono-  
métrie et décision  
Dunod 1969
4. **Jolion J.M.**  
Probabilités et statistiques  
Département Génie  
Industriel INSA  
Lyon 2004
5. **Meisel D.**  
L'histoire du calcul  
des probabilités  
Université  
de Leeds 1999
6. **Motulsky H.J.**  
Biostatistique -  
Une approche  
intuitive  
De Boeck Université  
2002
7. **Valleron A.J.**  
Introduction à la bio-  
statistique  
Collection Evaluation  
et Statistique  
Masson 1998

