STATISTIQUE ET ANALYSE DES DONNÉES

ERIC TEROUANNE

Inférence bayésienne dans les problèmes de décision composée et de l'empirique bayésien

Statistique et analyse des données, tome 4, nº 3 (1979), p. 33-44
<http://www.numdam.org/item?id=SAD_1979__4_3_33_0>

Statistique et Analyse des Données
3 - 1979 pp. 33-44

INFERENCE BAYESIENNE DANS LES PROBLEMES DE DECISION COMPOSEE ET DE L'EMPIRIQUE BAYESIEN

par Eric TEROUANNE
U.E.R. Mathématiques Appliquées aux Sciences Humaines
Université Paul Valéry, Montpellier

1. INTRODUCTION

1.1 Historique

Les travaux de H. ROBBINS (1951,55) ont indroduit les expressions de "problème de décision composée" et "empirique bayésien" pour désigner deux techniques de décision concernant, l'une et l'autre, plusieurs expériences statistiques, simultanées ou non. Lui même (62,63) et ses successeurs dans les années 65 à 75, CLEMMER, HANNAN, KRUTCHKOFF, MARTZ, MARITZ, MIYASAWA, OATEN, SAMUEL, VAN RYZIN, WIND etc..., ont développé parallèlement ces deux techniques qui ont atteint un certain degré de sophistication. Une revue assez complète peut en être trouvée dans MARITZ (70) et GHOZALI (76), et une revue comme les Annals of Statistics continue à publier ce genre de travaux.

Le but avoué de ROBBINS dans ses premiers papiers était de trouver une issue à l'affrontement entre les points de vue "bayésien" et "non-bayésien" en inférence statistique. Beaucoup ont cru que l'empirique Bayésien était cette issue. C'est à ce titre que J. NEYMAN (62) a salué les travaux de ROBBINS comme l'aube d'un jour nouveau en décision statistique. A l'opposé, LINDLEY (72, pp.64sq) estime qu'il y a là bien peu de nouveauté et s'élève contre l'usage proposé par ROBBINS pour ses techniques. La discussion sur les fondements de ces méthodes n'est pas close, et a peu de chance de se clore rapidement à notre avis puisqu'elle se rattache à la querelle entre les diverses écoles de décision statistique, et notamment entre Bayésiens et non bayésiens.

1.2 Buts et plan de ce papier

L'auteur pense, avec BASU (69, p.450) entre autres, que l'espoir d'une réconciliation entre les approches Bayésiennes et non-Bayésiennes dans l'analyse des données était et demeure illusoire. C'est cependant cette illusion qui se manifeste chez les empiristes bayésiens. Elle nous semble découler d'un malentendu sur la nature profonde de la démarche Bayésienne, qui n'est que rarement explicitée par les auteurs. Parmi les nombreuses interprétations possibles du qualificatif de statisticien bayésien, nous avons une option précise que nous décrivons au § 1.3. Cette option est personnelle, par nature même, et admet bien sûr l'existence d'autres options. Mais tout ce travail est basé sur elle et c'est à ce sens qu'il faut entendre ici le mot bayésien.

Ceci étant posé, ce papier voudrait contribuer à éclairer le débat sur les problèmes posés par ROBBINS, en exposant de façon précise comment un statisticien bayésien peut les traiter. Ceci n'a, à notre connaissance, pas encore été fait. Ce faisant, nous montrons qu'il n'y a rien de bayésien dans les techniques de l'empirique bayésien et que pour le bayésien les deux problèmes de ROBBINS n'en font qu'un et relèvent du schéma classique de l'inférence bayésienne.

Après avoir précisé ce que nous entendons ici par démarche bayésienne (§ 1.3), nous exposons sur un exemple les deux problèmes (§ 2) et fixons nos notations (§ 3). Le § 4 décrit la solution bayésienne des deux problèmes, et pour cela développe une version "à étages" du schéma bayésien classique. Le § 5 envisage divers type de situations pouvant relever de ce traitement. Enfin nous montrons au § 6 en quoi les espoir soulevés par l'empirique bayésien nous semblent relever d'une réconciliation illusoire.

1.3 Etre ou ne pas être bayésien

Nous appelons bayésien celui que SAVAGE (54) appelle "l'homme rationnel": considérant une expérience statistique dans laquelle on distingue :

- un espace des paramètres, ou états de la nature, F
- un espace des observables, E,
- une transition de probabilité P de F sur E, qui est la loi paramétrique des observations,
- un espace des actions ou décisions possibles A,

SAVAGE montre que, s'il veut respecter quelques axiomes simples, le décideur procèdera toujours au moyen de :

- sa probabilité personnelle, ou "a priori", sur les états de la nature, et
- une fonction de perte définie sur $F \times A$.

Pour nous, ce n'est pas l'utilisation du théorème de Bayes qui fait qu'une technique statistique est bayésienne, c'est le fait d'utiliser sa probabilité a priori personnelle qui fait qu'un statisticien a une démarche bayésienne.

2. EXPERIENCES STATISTIQUES LIEES : UN EXEMPLE.

Considérons le problème classique du statisticien qui reçoit un lot d'objets manufacturés et doit décider de le rejeter ou de l'accepter en fonction de la proportion "y" d'objets défectueux dans le lot. Il ne peut pas tester chacun des objets, et y restera donc inconnue. Par contre il peut tirer un échantillon du lot, dans lequel il déterminera le nombre "x" d'objets défectueux : x est une réalisation de la variable aléatoire X dont la loi de probabilité dépend du paramètre y.

Supposons de plus que les objets qui composent le lot ont été tirés d'un stock dans lequel la proportion de défectueux est "z", de telle sorte que y est une réalisation de la variable aléatoire Y dont la loi de probabilité dépend du paramètre z. Si z était connu, nous pensons avec BASU (69, p.451), que la majorité des statisticiens, Bayésiens ou non, estimerait acceptable la démarche suivante :
Connaissant la loi de probabilité de Y, on utilise le théorème de Bayes pour calculer la décision qui minimise le risque a postériori. On pourrait aussi, toujours dans une démarche non-bayésienne, utiliser cette connaissance de la loi de Y pour faire de l'estimation, par une méthode de maximum de vraisemblance par exemple. Mais ce n'est pas ce problème que nous nous posons ici :

Nous nous plaçons ici dans le cas où z est inconnu.

Supposons maintenant que chaque jour "i" ($i = 1,2,\dots,n$), notre statisticien reçoit, tiré du même stock, un lot caractérisé par une proportion y_i de défectueux, et que chaque jour il tire un échantillon du lot qu'il vient de recevoir et observe un nombre x_i de défectueux. Le $n^{\text{ième}}$ jour ("aujourd'hui"), le statisticien a observé les réalisations $x_1,\dots,x_n$ des variables aléatoires $X_1,\dots,X_n$; il sait comment la distribution de probabilité de X_i dépend du paramètre y_i inconnu ($i = 1,\dots,n$) ; il sait que $y_1,\dots,y_n$ sont des réalisations des variables aléatoires $Y_1,\dots,Y_n$ dont les distributions de probabilité dépendent du même paramètre inconnu z ; il doit prendre une décision sur le lot d'aujourd'hui, le $n^{\text{ième}}$.

Avant ROBBINS, le verdict des techniques classiques était : ignorant z , on ne peut lier entre elles les n expériences, donc la décision d'aujourd'hui ne dépend que de x_n.

L'idée de ROBBINS (55), et c'est ce qu'il appelle l'attitude "bayésienne empirique", c'est de prendre la décision sur y_n en tenant compte des observations "passées" $x_1,\dots,x_{n-1}$, aussi bien que de x_n.

Pour le Bayésien, au sens où nous l'entendons, les n expériences sont liées de manière naturelle dans le schéma bayésien classique, à partir du moment où il a explicité sa probabilité a priori sur l'espace du paramètre inconnu z.

Le problème de décision composée diffère du précédent en ce que la décision à prendre ne concerne plus le $n^{\text{ième}}$ lot seulement mais chacun des n lots : on peut imaginer par exemple que notre statisticien reçoit ses n lots simultanément, et doit décider d'accepter ou de rejeter chacun d'eux, après avoir observé $x_1,\dots,x_n$.

On peut faire exactement les mêmes remarques pour ce problème que pour le précédent, concernant les techniques classiques, l'apport de ROBBINS (51) et l'attitude bayésienne.

Dans les deux problèmes, la liaison entre les n expériences statistiques est au niveau du "stock" commun, c'est à dire, mathématiquement parlant, au niveau du paramètre z des distributions de probabilité des Y_i. C'est à propos du caractère bayésien ou non de cette liaison que s'est établie une certaine confusion.

3. NOTATIONS ET RAPPELS

3.1 Notations.

Tous les ensembles utilisés ici sont discrets sauf l'espace probabilisé "au sommet", Ω. Il serait facile mais fastidieux d'écrire la même chose en continu. Les mesures utilisées étant discrètes, il n'y a pas de restriction à prendre pour tribus sur les ensembles discrets, leurs ensembles de parties. C'est ce que nous ferons sans le préciser à chaque fois.

On appelle schéma bayésien un quadruplet (Q,F,P,E), où E et F sont des ensembles, Q une probabilité sur F et P une transition de probabilité de F sur E.

On note $P(/y)$ la probabilité sur E associée par P à $y \in F$, et $P \circ Q$ la probabilité "marginale" sur E :

$$P \circ Q(x) = \sum_{y \in F} Q(y)\, P(x/y) \quad , \; x \in E .$$

Si P' et P sont des transitions de probabilité, respectivement de G sur F et de F sur E, on note $P \circ P'$ la transition de G sur E qui à tout $z \in G$ associé la probabilité $P \circ P'(\;/z)$ sur E.

Si $P_1, P_2, \dots, P_n$ sont des transitions de probabilité de F sur $E_1, E_2, \dots, E_n$ respectivement, on note $\prod_{i=1}^{n} P_i$ la transition de F sur $\prod_{i=1}^{n} E_i$ qui à tout $y \in F$ associe la probabilité sur $E_1 \times \dots \times E_n$ produit des probabilités $P_i(\;/y)$:

$$\prod_{i=1}^{n} P_i \;(x_1,\dots,x_n/y) = \prod_{i=1}^{n} P_i\,(x_i/y) \quad , \; x_1 \in E_1, \dots, x_n \in E_n .$$

Soient $(\Omega, \mathcal{A}, P)$ un espace probabilisé, E et F des ensembles, X et Y des variables aléatoires à valeurs dans E et F respectivement. On notera :

P_F la probabilité sur F, loi de la variable Y,

$P_F(\;/X)$ la transition de probabilité de E sur F qui à toute valeur x de la variable X associe :

$P_F(\;/x)$ la loi conditionnelle de Y sachant $X = x$.

De même P_E, $P_E(\;/Y)$, $P_E(\;/y)$ pour $y \in F$.

Ainsi "P" désignera toutes les probabilités ou transitions de probabilité, l'indice désignant l'ensemble sur lequel elle est définie. Cet indice pourra être omis s'il n'y a pas de confusion possible.

3.2 modèle probabiliste de l'inférence bayésienne

L'inférence bayésienne consiste en la transformation par l'observation de l'a priori du statisticien en un a posteriori. Le modèle probabiliste représente donc simultanément un statisticien bayésien et une expérience statistique. On y distingue :

- un espace probabilisé $(\Omega, \mathcal{A}, P)$
- une variable aléatoire Y appelée paramètre, ou état de la nature, à valeur dans un ensemble F,
- et une variable aléatoire X appelée observation, à valeur dans un ensemble E.

On a alors, pour tout $(y,x) \in F \times E$:

$P_{F \times E}(y,x) = P_F(y)\, P_E(x/y) = P_E(x)\, P_F(y/x)$.

- P_F exprime l'a priori du statisticien
- $P_E(\ /Y)$ est la loi paramétrique des observations
- $P_F(\ /X)$ est la trasition a posteriori.
- Le schéma bayésien correspondant est le quadruplet $(P_F,\ F,\ P_E(\ /Y),\ E)$.

4. INFERENCE BAYESIENNE DANS LES EXPERIENCES STATISTIQUES LIEES.

4.1 Schéma bayésien à étages

On considère le modèle probabiliste suivant :

- un espace probabilisé $(\Omega, \mathcal{A}, P)$,
- une variable aléatoire Z à valeurs dans un ensemble G,
- des v.a. $Y_1,\dots,Y_n$ à valeurs dans $F_1,\dots,F_n$ respectivement,
- des v.a. $X_1,\dots,X_n$ à valeurs dans $E_1,\dots,E_n$ respectivement,

tels que :

- conditionnellement à Z, les Y_i sont indépendantes

$$P_{F_1 \times F_2 \times \dots \times F_n}(\ /Z) = \prod_{i=1}^{n} P_{F_i}(\ /Z)\ ,$$

- conditionnellement à Y_i, X_i est indépendante des autre v.a., $Z, Y_1,\dots,Y_{i-1}, Y_{i+1},\dots,Y_n, X_1,\dots,X_{i-1}, X_{i+1},\dots,X_n$, et ceci quel que soit $i = 1,\dots,n$.

On a alors, pour tout $z \in Z$, $y_i \in Y_i$, $x_i \in X_i$ $(i = 1,\dots,n)$:

$$P(z,y_1,\dots,y_n,x_1,\dots,x_n) = P(z) \prod_{i=1}^{n} P(y_i/z)\ P\,(x_i/y_i)$$

Un tel schéma permet de représenter un statisticien bayésien et n expériences statistiques liées. On y distingue en effet :

- les n expériences statistiques : pour tout $i = 1,\dots,n$,

F_i est l'espace états de la nature particulier à la $i^{\text{ème}}$ exp.
E_i est l'espace des observations particulier à la $i^{\text{ème}}$ exp.
$P_{E_i}(\ /Y_i)$ est la loi paramétrique particulier à la $i^{\text{ème}}$ exp.

- la liaison entre les expériences : les états de la nature particuliers à chacune des n expériences ont des lois $P_{F_i}(\ /Z)$ qui dépendent du même paramètre Z qu'on peut appeler état général de la nature.

- enfin l'a priori du statisticien : c'est la loi P_G de Z.

Remarquons au passage une propriété de ce schéma :

<u>Lemme</u> : conditionnellement à Z, les X_i sont indépendantes.
En effet, pour tout $(z,x_1,\dots,x_n) \in Z \times X_1 \times \dots \times X_n$,

on a
$$P(z,x_1,\dots,x_n) \sum_{y_1 \in F_1} \dots \sum_{y_n \in F_n} P(z) \prod_{i=1}^{n} P(y_i/z)\ P(x_i/y_i)$$
$$= P(z) \prod_{i=1}^{n} \sum_{y_i \in F_i} P(y_i/z)\ P(x_i/y_i)$$
$$= P(z) \prod_{i=1}^{n} P(x_i/z)\ .$$

Dans le cas particulier, qui est celui de notre exemple au § 2, où les n expériences $(F_i, P_{E_i}(\ /Y_i), E_i)$ sont n répétitions de la même expérience $(F, P_E(\ /Y), E)$, les observations $X_1,\dots,X_n$ sont, conditionnellement à $z \in G$, indépendantes et équidistribuées selon la loi "marginale" $P_E(\ /Y) \circ P_F(\ /z)$.

<u>4.2 Inférence bayésienne dans le problème "bayésien empirique".</u>

La $n^{\text{ième}}$ jour, dans un problème bayésien empirique, il faut prendre une décision concernant l'état particulier de la nature ce jour là, Y_n, au vu des observations passées : $X_1,\dots,X_{n-1}$, et présente : X_n.

Il faut donc calculer la transition a posteriori $P_{F_n}(\ / X_1,\dots,X_n)$.

Nous montrons dans ce paragraphe comment se décompose le calcul de cette transition :

<u>Proposition</u> $P_{F_n}(\ /X_1,\dots,X_n)$ est la transition qui a tout $(x_1,\dots,x_n) \in E_1 \times \dots \times E_n$ associe la probabilité a posteriori sur F_n quand on a observé x_n, dans le schéma bayésien $(P_{F_n}(\ /x_1,\dots,x_{n-1}), F_n, P_{E_n}(\ / Y_n), E_n)$, où l'a priori $P_{F_n}(/x_1,\dots,x_{n-1})$ est la composée de la transition $P_{F_n}(\ /Z)$ et de l'a posteriori sur G quand on a observé $x_1,\dots,x_{n-1}$:
$$P_G(\ /x_1,\dots,x_{n-1}).$$

Les formules utiles au calcul sont donc : pour tout $y_n \in E_n$,

$$P(y_n/x_1,\dots,x_n) = \frac{P(y_n/x_1,\dots,x_{n-1})\ P(x_n/y_n)}{P(x_n/x_1,\dots,x_{n-1})}$$

et $P(y_n/x_1,\dots,x_{n-1}) = \sum_{z \in G} P(y_n/z)\, P(z/x_1,\dots,x_{n-1})$.

On a en effet $\quad P(y_n/x_1,\dots,x_n) =$

$$= \frac{P(x_1,\dots,x_n,y_n)}{P(x_1,\dots,x_n)} =$$

$$= \frac{P(x_n/x_1,\dots,x_{n-1},y_n)\, P(x_1,\dots,x_{n-1},y_n)}{P(x_n/x_1,\dots,x_{n-1})\, P(x_1,\dots,x_{n-1})}$$

$$= \frac{P(x_n/y_n)\, P(y_n/x_1,\dots,x_{n-1})}{P(x_n/x_1,\dots,x_{n-1})}$$

et $$P(y_n,x_1,\dots,x_{n-1}) = \sum_{z \in Z} \sum_{y_1 \in F_1} \dots \sum_{y_{n-1} \in F_{n-1}} P(y_n/z)\, P(z) \prod_{i=1}^{n-1} P(x_i/y_i)\, P(y_i/z)$$

$$= \sum_{z \in Z} P(z)\, P(y_n/z) \prod_{i=1}^{n-1} P(x_i/z)$$

$$= \sum_{z \in Z} P(z)\, P(y_n/z)\, P(x_1,\dots,x_{n-1}/z)$$

$$= \sum_{z \in Z} P(y_n/z)\, P(z/x_1\dots x_{n-1})\, P(x_1,\dots,x_{n-1})$$

d'où $$P(y_n/x_1,\dots,x_{n-1}) = \sum_{z \in Z} P(y_n/z)\, P(z/x_1,\dots,x_{n-1})$$

Enfin, pour calculer $P_G(\ /X_1,\dots,X_{n-1})$, on utilise le résultat suivant qui découle immédiatement de l'indépendance des X_i conditionnellement à M (il ne fait que généraliser à des expériences différentes le processus Markovien classique des a posteriori lors de n répétitions indépendantes d'une même expérience).

__Lemme__ pour tout $(x_1,\dots,x_n) \in \prod_{i=1}^{n} E_i$ et pour tout $k = 1,\dots,n$, $P_G(\ /x_1,\dots,x_k)$ est l'a posteriori sur G quand on a observé x_k, dans le schéma $(P_G(/x_1,\dots,x_{k-1}),G,P_{E_k}(/Z),\ E_k)$, où l'a priori est l'a posteriori sur G quand on a observé $x_1,\dots,x_{k-1}$.

4.3 Inférence bayésienne dans les problèmes de décision composée.

Dans un problème de décision composée concernant n expériences statistiques liées, il faut prendre une décision concernant simultanément tous les états particuliers de la nature $Y_1,\dots,Y_n$, au vu des observations $X_1,\dots,X_n$.

Il faut donc calculer la transition a posteriori

$$P_{Y_1 \times \dots \times Y_n}(\ /\ X_1,\dots,X_n)\ ,$$

c'est à dire pour tout $(x_1,\dots,x_n,\ y_1,\dots,y_n)\quad E_1 \times \dots \times E_n \times F_1 \times \dots \times F_n$

$P(y_1,\ldots,y_n/x_1,\ldots,x_n)$ qui est proportionnel à

$$P(x_1,\ldots,x_n,y_1,\ldots,y_n) = \sum_{z \in Z} P(z) \prod_{i=1}^{n} P(y_i/z)\, P(x_i/y_i)$$

$$= \prod_{i=1}^{n} P(x_i/y_i) \sum_{z \in Z} P(z) \prod_{j=1}^{n} P(y_j/z)$$

$$\text{et non pas} \quad \prod_{i=1}^{n} P(x_i/y_i) \sum_{z \in Z} P(z)\, P(y_i/z)$$

comme on trouve dans LINDLEY (72)

Cependant, dans tous les exemples que nous avons rencontrés, la fonction de perte se décompose en une somme de fonctions dépendant chacune d'un seul y_i et de la "sous décision" concernant seulement la $i^{\text{ème}}$ expérience. Dans de tels cas, le risque a posteriori qu'il faut minimiser est la somme de risques partiels dont chacun dépend de l'a posteriori sur un des espaces de paramètres F_i. Il faut donc calculer, pour tout i,

$$P_{F_i}(\ /X_1,\ldots,X_n),$$

On voit ainsi que le problème de décision composée, du moins dans les cas exposés dans la littérature, se traite dans l'attitude bayésienne par n applications du calcul décrit au § 4.2 pour l'empirique bayésien. Dans le cas général où la fonction de perte dépend de tous les y_i et d'une décision globale $d \in D$, on calcule le risque a posteriori de la décision d en intégrant sur $Y_1 \times Y_2 \times \ldots \times Y_n$ selon la loi $P(\ /x_1,\ldots,x_n)$, et la décision Bayésienne est celle qui minimise cette quantité.

5. SUR LA NATURE DE LA LIAISON ENTRE LES EXPERIENCES.

5.1 Quand peut-on dire que n expériences sont liées ?

La question de savoir à quelles situations expérimentales pouvaient s'appliquer ces techniques areçu des réponses très variées. Pour ROBBINS (51) "aucun lien n'est supposé entre les paramètres Y_i ; X_1 peut être une observation sur un papillon en Equateur, X_2 sur une huître du Maryland, X_3 la température d'une étoile etc...". A l'opposé, pour LINDLEY (72), on doit se restreindre aux cas où "il y a une raison subjective de penser que les paramètres Y_i sont interchangeables". Dans la démarche bayésienne telle que nous la concevons ici, des expériences sont liées pour un statisticien à partir du moment où elles le sont dans son esprit, et plus pragmatiquement elles relèvent d'un traitement commun à partir du moment où il est capable de représenter ces expériences et son a priori dans un même schéma bayésien tel que celui que nous avons construit au § 4. Des problèmes techniques, dus entre autres causes à la pauvreté de l'outil mathématique, limiteront donc l'expression de ces liaisons. Quand à la nature extra-mathématique de la liaison, elle peut être variée et nous en donnons ici quelques exemples :

5.2 Liaison intuitive

Quand ROBBINS veut traiter comme liées des expériences sur les huîtres, les étoiles et les papillons, il est amené à utiliser un argument fallacieux, comme nous le montrons au § 6.2, parce qu'il abandonne en cours de route la démarche bayésienne. Mais pour le bayésien l'idée n'est pas absurde : toute expérience modifiant l'expérimentateur, on peut concevoir que son a priori sur l'huître n'est plus la même après qu'il a observé une étoile. Cependant cette modification, si elle était sensible, se passerait à un niveau, intuitif, très éloigné de ce que nous sommes capables de formaliser dans un modèle probabiliste !

5.3 Liaison fréquentiste

C'est le cas de n expériences similaires qui ne diffèrent que par l'état particulier de la nature, les y_i étant n réalisations d'une même variable aléatoire Y de loi paramétrique $P_F(\ /Z)$.

Nous qualifions cette liaison de fréquentiste parce que tout état général de la nature, $z \in G$, peut être identifié à une probabilité sur F , $P_F(\ /z)$, et que cette probabilité a une interprétation fréquentiste. Ce sont de tels cas que traitent, de façon bayésienne, la plupart des auteurs "bayésiens empiriques".

5.4 Liaison ensembliste

Supposons que l'on connaisse les valeurs des états particuliers de la nature $y_1,\ldots,y_n$, mais qu'on ne sache pas à laquelle des expériences s'applique chacune d'elle : On a $F_1 = F_2 = \ldots = F_n = \{y_1,\ldots,y_n\} = F$, et chaque état général $z \in G$ peut être assimilé à une permutation de $\{ 1,\ldots,n \}$, la loi de probabilité de Y_i étant alors :

$$P_{F_i}(y_k/z) = \begin{cases} 0 & \text{si } z(i) \neq k \\ 1 & \text{si } z(i) = k \end{cases}$$

Ce cas assez curieux est souvent traité, de façon non-bayésienne, sous le nom de problème de décision composée (voir par exemple SAMUEL, 67). En qualifiant cette liaison d'ensembliste, nous avouons un certain manque d'intuition à son sujet.

6. CRITIQUE DES METHODES CLASSIQUES.

6.1 Bayésien empirique et décision composée classiques.

Dans les techniques développées par les auteurs que nous avons cités au § 1, il n'y a jamais la place pour une probabilité a priori personnelle, intuitive, du statisticien. Ce ne sont donc, pour nous , pas des techniques bayésiennes. Dans la plupart de ces techniques, on considère qu'il existe une vraie valeur, inconnue, de z , et on développe des estimateurs de cette valeur, puis on utilise l'estimation $\tilde{z}$ pour calculer une "estimation de la décision bayésienne" dans le schéma

$$(P_F(\ /\ \tilde{z}),\ F,\ P_X(\ /Y),\ X).$$

Tous ces travaux se rattachent donc aux divers courants de la statistique inférentielle non bayésienne, et comme tels conservent toute leur valeur.

Cependant nous croyons que seule la démarche bayésienne peut assurer une parfaite cohérence. Aussi n'est-il pas étonnant de relever dans la littérature "bayésienne empirique" plusieurs anomalies épistémologiques que nous décrivons ci-dessous. Ces anomalies ne peuvent s'expliquer que par le manque de fondements dont souffre la statistique. Elles sont la base de cet espoir illusoire de réconciliation entre les points de vue bayésien et non bayésien.

6.2 l'huître et le papillon.

La première anomalie consiste en l'absence fréquente d'identification de la liaison entre les expériences, et elle est particulièrement visible dans la phrase de ROBBINS citée au § 5.1 : Aucune raison, intuitive ou expérimentale, n'est requise pour considérer les expériences comme liées ; il suffit pour l'établir que la liaison soit techniquement réalisable, au niveau mathématique, formel. Le seul argument qu'avance ROBBINS pour lier l'huître au papillon tient dans l'identité formelle des modèles probabiliste $(F_i\ ,\ P_{E_i}\ (\ /Y_i),\ E_i)$ qu'il suppose gaussiens: $F_i = E_i = \mathbb{R}$ et $P_{E_i}(\ /y_i) = \mathcal{N}(y_i\ ;\ 1)$. Quand on connait [1] la pauvreté en modèles de l'arsenal probabilistes, et combien de variables sont facilement supposées gaussiennes, il n'y a pas de doute que cette condition est très permissive, ce qui a justifié la réaction en sens inverse de LINDLEY.

6.3 La tour de Babel

Une autre anomalie est l'illusion que rajouter un étage change la nature du problème. Le but avoué de l'empiriste bayésien est de justifier l'usage du schéma bayésien aux yeux de ceux qui se refusent à employer une probabilité a priori personnelle : pour que l'a priori sur F devienne objective il faut l'estimer, d'où l'idée de faire plusieurs expériences liées par le même état général $z \in G$. Devant ce nouvel espace de paramètre, la démarche bayésienne consiste à placer sur G l'a priori du statisticien, et le problème se pose dans les mêmes termes : si on voulait "estimer" P_G il faudrait rajouter un nouvel étage, et ainsi de suite. Ainsi dans notre exemple du § 2, on pourrait supposer que le stock est l'un des nombreux stocks constitués dans le pays à partir de la production d'une même usine ; la proportion de défectueux à la sortie de l'usine serait le nouvel état généralissime de la nature ; chacun des m stocks fournirait chaque jour un lot dont serait tiré un échantillon ... Nous laissons au lecteur le plaisir de rajouter un acheteur à un bout et une multinationale à l'autre. Le problème reste fondamentalement le même, et sans la clef de voûte de l'a priori au sommet du modèle, une telle construction n'a pas plus de cohésion que la célèbre tour.

6.4 l'Oeuf et la Poule

Enfin le bayésien empirique se heurte au problème de l'origine du savoir : il veut utiliser la connaissance a priori mais la fonder objectivement sur les expériences passées. Mais que faire lors de la première expérience ? ou bien doit-il supposer que le passé est infini ? Se poser cette question (quelle est la première, de la connaissance ou de l'expérience), c'est comme se demander qui était le premier, l'oeuf ou la poule.

BIBLIOGRAPHIE

1. ATCHISON T.A. and MARTZ H.F. (1969), Proceedings of the Symposium on empirical Bayes Estimation and Computating in Statistics, Texas Tech. University Mathematics Series n° 6.

2. BASU D. (1969), Role of the sufficiency and likelihood principles in sample survey theory. Sankhya, A, 31, 441-454.

3. CLEMMER B.A. and KRUTCHKOFF R.G. (1968), The use of Empirical Bayes estimators in a linear regression model, Biometrika 55, 3, 525-534.

4. GHOZALI M.R. (1976), Problèmes statistiques à étages, Thèse, Université des Sciences et Techniques du Languedoc, Montpellier.

5. HANNAN J.F. and ROBBINS H. (1955), Asymptotic solutions of the Compound Decision problem for two completely specified distributions, Ann. of Math. Stat. 26, 37-51.

6. HANNAN J.F. and VAN RYZIN J.R. (1965), Rate of convergence in the Compound Decision problem for two completely specified distributions, Ann. of Math. Stat. 36, 1743-1752.

7. KRUTCHKOFF R.G. (1969), An introduction to Empirical Bayes, in ATCHINSON T.A. and MARTZ H.F., op. cit.

8. LINDLEY D.V. (1958), Professor Hogben's crisis - a survey of the foundations of statistics, Appl. Statist. 7, 186-198.

9. LINDLEY D.V. (1972), Bayesian statistics, a review, Regional conference series in applied mathematics, 2. SIAM, Philadelphia.

10. MARITZ J.S. (1970), Empirical Bayes methods, Methuen, London.

11. MARTZ H.F.Jr. (1969), Multivariate extended Empirical Bayes methods, in ATCHINSON T.A. and MARTZ H.F., op.cit.

12. MIYASAWA K. (1960), An empirical Bayes estimator of the mean of a normal population, Bull. Int. Inst. of Statist. 38, 181-188.

13. NEYMAN J. (1962), Two breakthroughs in the theory of statistical decision making, Rev. Int. Statist. Inst. 30, 11-27.

14. OATEN A. (1972), Approximation to Bayes risk in Compound Decision problem, Ann. of Math. Statist. 43, 1164-1184.

15. ROBBINS H. (1951), Asymptotically subminimax solutions of Compound statistical Decision problems, Proc.Sec. Berk. Symp. on Statist. and Prob., 131-148.

16. ROBBINS H. (1955), An Empirical Bayes approach to statistics, Proc. Third Berk. Symp. on Statist. and Prob., 157-163.

17. ROBBINS H. (1962), Some numerical results on a compound decision problem. Recent developments in information and decision processes, edited by MACHAL and GRAY, Mac Millan, 56-62.

18. ROBBINS H. (1963), The empirical Bayes approach to testing statistical hypotheses, Rev. Inter. Statist. Inst. 31, 195-208.

19. ROBBINS H. (1964), The empirical Bayes approach to Statistical decision problems, Ann. of Math. Statist. 35, 1-20.

20. SAMUEL E. (1967), The compound statistical decision problem, Sankhya, Ser. A. 29, 123-140.

21. SAVAGE L.J. (1954), The foundations of statistics, Wiley, New York.

22. SAVAGE L.J. (1961), The foundations of statistics reconsedered, Proc. Fourth Berk. Symp. 1, 575-586.

23. SAVAGE L.J. (1962), The foundations of statistical inference, Methuen, London.

24. VON MISES R. (1942), On the correct use of Bayes'formula, Ann. of Math.Statist. 13, 156-165.

25. WIND S.L. (1973), An empirical Bayes approach to multiple linear regression, Ann. of Statist. 1, 93-103.