

REVUE DE STATISTIQUE APPLIQUÉE

A. DE FALGUEROLLES

K. J. WORSLEY

A. BACCINI

J. R. MATHIEU

A. M. MONDOT

H. CAUSSINUS

M. AITKIN

B. FRANCIS

N. RAYNAL

Réponse des auteurs

Revue de statistique appliquée, tome 35, n° 3 (1987), p. 93-94

http://www.numdam.org/item?id=RSA_1987__35_3_93_0

© Société française de statistique, 1987, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

RÉPONSE DES AUTEURS

Lorsque les résultats d'un travail collectif sont présentés sous la forme de cinq articles distincts, il est difficile de faire ressortir une conclusion générale. Heureusement, les commentaires précédents pallient cette carence dans une très large mesure en apportant les divers ingrédients d'une telle conclusion : éléments de synthèse, précisions sur le degré d'achèvement (et, partant, d'inachèvement) du travail, pistes pour des réflexions nouvelles.

Un premier point que nous relèverons est que, même si chaque commentateur l'exprime avec ses nuances personnelles, un large consensus se dégage sur l'intérêt d'utiliser sans exclusive des techniques statistiques variées, dans une approche ouverte et flexible. La stratégie d'analyse peut varier selon les cas, et nous avons cherché à montrer comment sur des exemples de natures diverses; le schéma présenté par A. LECLERC est extrêmement utile pour synthétiser ces différentes possibilités.

Mais la discussion met aussi en lumière bien des limites de nos propositions, ce qui nous satisfait encore car l'un des objectifs était de sortir de divers conformismes et donc de soulever des interrogations. Nous allons maintenant passer en revue quelques-unes de ces questions.

Un point important est celui des rapports entre Maximum de Vraisemblance et Moindres Carrés. Est-il cohérent de faire cohabiter ces techniques ? On peut noter d'abord qu'elles sont moins éloignées qu'il peut paraître : par exemple, les estimateurs de Maximum de Vraisemblance et ceux des Moindres Carrés convenablement pondérés sont asymptotiquement équivalents. Remarquons aussi que les commentaires de J. DE LEEUW apportent quelques éléments de réponse (positive ou négative selon les cas) aux interrogations d'Y. ESCOUFIER. Mais, au delà de la technique des Moindres Carrés, c'est la décomposition en valeurs singulières elle-même qui pourrait être mise en cause comme unique méthode d'analyse des résidus proposée : elle ne s'applique directement qu'à des tableaux à deux entrées et elle cherche systématiquement des approximations en termes d'effets ligne et d'effets colonne. Ses limites sont donc évidentes, et c'est en partie pour cela qu'une analyse par rapport à un modèle, tenant compte d'effets conjoints, peut s'avérer utile. Mais il est clair que ces questions sont très ouvertes et fournissent un vaste champs de réflexions futures.

Il faut ensuite noter l'interrogation sur le rôle et le concept même de modèle qui se dégage d'une façon ou d'une autre de tous les commentaires. Sans vouloir revenir sur ce que certains d'entre nous ont dit à ce sujet ici ou ailleurs, soulignons cependant quelques points. Il semble en premier lieu que, pour beaucoup de statisticiens, le modèle soit de plus en plus considéré d'un point de vue purement utilitariste, comme un outil permettant une approche simplifiée, une représentation « à grands traits », d'une réalité sous-jacente dont l'appréhension complète est illusoire. Il est alors clair qu'une approche ne saurait être entièrement « confirmatoire » au sens fort du terme (comme celui qui est évoqué par le concept de tests d'hypothèses) puisqu'il s'agira au mieux de « confirmer des approximations », pratique qui reste à préciser. Il est donc naturel que la division entre approches confirmatoire et exploratoire ne puisse être que floue (A. LE-

CLERC) ou que la situation pratique soit le plus souvent intermédiaire (T. ALUJA et M. MARTI). En fait, la notion même de modèle, et d'adéquation d'un modèle, variera en fonction de la richesse des données disponibles et cette question est aujourd'hui particulièrement sensible car cette richesse a beaucoup crû ces dernières années, trop vite pour que les cadres de référence évoluent toujours de façon convenable (rappelons à ce sujet que l'essentiel de nos efforts a porté sur des exemples assez volumineux). D'autre part, il faut noter que la part de modélisation variera selon les buts de l'analyse (compréhension ou prévision, par exemple) comme le souligne justement J.-J. DAUDIN; on peut rappeler à ce propos que la qualité prévisionnelle d'un modèle est autre chose que sa bonne adéquation aux données.

Nous répondrons maintenant à quelques remarques plus spécifiques.

T. ALUJA et M. MARTI indiquent à juste titre que, dans certains cas particuliers d'interactions entre les variables, l'Analyse des Correspondances Multiples peut suggérer des liaisons artificielles. Il nous semble que c'est précisément dans de telles situations qu'une modélisation (log-linéaire ou autre) peut à sa suite permettre de corriger les erreurs susceptibles d'être commises en utilisant uniquement la première méthode. La modélisation joue alors typiquement un rôle confirmatoire au sens affaibli dont il est question ci-dessus.

Y. ESCOUFIER remarque que les étapes E et M de l'algorithme E.M. (article de AITKIN et al.) correspondent aux déterminations respectives des fonctions d'affectation et de représentation de la méthode de classification des Nuées Dynamiques; c'est exact, mais il faut préciser que le modèle des Classes Latentes conduit à calculer des probabilités d'affectation des individus aux classes (au moyen de l'algorithme E.M.), alors que les Nuées Dynamiques affectent les individus de façon déterministe : ce n'est donc que dans un cas particulier du modèle de Classes Latentes (affectations avec probabilités 0 ou 1) que l'on retrouve les Nuées Dynamiques (pour plus de détails, voir N. RAYNAL (1987)). A ce même sujet, deux individus aux profils identiques auront les mêmes probabilités d'affectation aux classes dans l'utilisation du modèle des Classes Latentes : la question d'Y. ESCOUFIER sur la possible différence de leur classification pose donc le principe même de la légitimité d'une affectation probabiliste (pour notre part, nous avons décidé de ne pas l'exclure).

Répondons pour terminer aux remarques de J.-J. DAUDIN et Y. ESCOUFIER concernant les logiciels utilisés. L'étude comparative que nous nous étions proposé de mener comportait un certain nombre d'a priori, en particulier l'idée d'utiliser des logiciels standards largement disponibles : SICLA, SPAD et des logiciels du même type pour les méthodes d'analyse des données, GLIM (essentiellement mais non uniquement) pour la modélisation. Il n'en reste pas moins que d'autres seraient mieux adaptés pour tel ou tel point et que le développement de l'aspect logiciel nous est apparu nécessaire : voir le rapport L.S.P.-C.A.S. (1986).

C'est finalement un grand plaisir de remercier très chaleureusement tous ceux qui ont bien voulu participer à cette discussion, sans oublier la Revue de Statistique Appliquée et P. CAZES qui l'ont organisée. Elle a été pour nous extrêmement bénéfique, et il est au plus haut point encourageant que la première étape de ce travail, entrepris sous le signe du dialogue, s'achève ainsi dans les mêmes conditions. Surtout quand on mesure tout ce qui reste à faire...