

JSFS

## Comptes rendus de lecture

*Journal de la société française de statistique*, tome 147, n° 3 (2006),  
p. 107-109

[http://www.numdam.org/item?id=JSFS\\_2006\\_\\_147\\_3\\_107\\_0](http://www.numdam.org/item?id=JSFS_2006__147_3_107_0)

© Société française de statistique, 2006, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) 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/>

# COMPTES RENDUS DE LECTURE

## Le Modèle linéaire par l'exemple. Régression, analyse de la variance et plans d'expériences illustrés avec R, SAS et Splus

Jean-Marc Azaïs et Jean-Marc Bardet  
1 vol., 326 pages, Dunod Paris 2006  
ISBN 2 10 049559 3

Le monde du modèle linéaire et de ses applications est très vaste comme en témoigne l'abondante littérature anglo-saxonne (ancienne et récente) mais pas toujours adaptée aux enseignements français, pas seulement pour des raisons de langage. La littérature française est loin d'être aussi riche et très inégale sur ce sujet. Elle se répartit en gros en trois sortes : des ouvrages très abstraits sans pratiquement d'applications, des ouvrages très appliqués qui ne permettent pas vraiment de comprendre les méthodes utilisées ou bien des ouvrages très ciblés sur un des aspects du modèle linéaire. J.-M. Azaïs et J.-M. Bardet (JMA et JMB) ont voulu l'enrichir en proposant un ouvrage assez complet mais aussi synthétique destiné aux étudiants de niveau licence, masters, école d'ingénieurs et aux statisticiens d'entreprise, de formations diverses (au sens où les niveaux en mathématiques et les domaines d'application sont différents), avec pour ambition de couvrir un champ le plus large possible de façon fondamentale (théorie qui sous-tend les applications) tout en donnant un guide d'apprentissage pour l'utilisation de ce modèle dans des situations très diverses, ce qui suppose de savoir utiliser correctement les logiciels. Équation difficile qui impose une pédagogie prenant en compte les niveaux en mathématiques différents, et des choix. Les choix des auteurs sont respectables et correspondent bien aux besoins du public disparate visé. Ces choix les amènent à traiter les différents thèmes abordés plus ou moins en profondeur et à renvoyer le lecteur à d'autres ouvrages ou des articles en complément.

JMA et JMB organisent leur exposé avec une progression du simple au compliqué aussi bien pour les modèles que pour le formalisme mathématique. C'est ainsi qu'après avoir dans le chapitre 1 présenté la régression simple et l'analyse de variance à un facteur, ils introduisent le modèle linéaire au chapitre 3 avec son traitement dans le cas de modèles réguliers. Ce n'est qu'au chapitre 7 qu'ils signalent les difficultés inhérentes aux modèles non réguliers. Quelques compléments de mathématiques sont donnés au fur et à mesure et, dans une annexe, des rappels de probabilité.

L'ouvrage comporte quinze chapitres de longueurs inégales (de 6 à 41 pages, moyenne 20). C'est le chapitre 2 baptisé « Intermède méta statistique : pour une théorie pratique des tests » qui est le plus court. C'est une réflexion historique bienvenue sur la « sacralisation du 5 % ». La plupart des chapitres proposent des exemples traités par les trois logiciels R, SAS et Splus suivis de commentaires sur les résultats obtenus. Quelques exercices de niveau de difficulté variable et signalé sont proposés à la fin de chaque chapitre. La part consacrée au traitement des exemples représente environ 27 % de l'ouvrage, ceci pour montrer que le titre « par l'exemple » n'est pas usurpé. Les exercices sont souvent l'occasion d'étendre le champ exploré.

Le contenu est en partie seulement décrit dans le titre. En effet JMA et JMB abordent aussi le problème de la calibration, les séries chronologiques, les modèles mixtes et l'analyse de variance multivariable.

Quatre chapitres sont consacrés aux plans d'expérience, avec la notion d'optimalité, qui, rapidement certes, mais de façon qui me paraît plus que suffisante, donnent les définitions des plans les plus couramment utilisés avec quelques développements algébriques pour leur construction.

Deux chapitres s'adressent plus particulièrement aux étudiants de masters recherche : l'un sur le comportement asymptotique des estimateurs (dans les cas gaussiens et non gaussiens), l'autre sur la sélection de modèles. Ce dernier, le plus important en nombre de pages, est particulièrement bien introduit et traité de façon approfondie.

Le dernier chapitre est consacré au traitement plus complet de trois exemples très différents : données de l'Argus, étude de l'effet de la formulation sur les propriétés de surface du béton et étude de préférence sur des boissons « Cola »<sup>1</sup>. Ce chapitre est l'occasion de présenter une démarche complète d'analyse statistique incluant différentes méthodes de statistique descriptive permettant de guider le tâtonnement dans la recherche de modèles linéaires.

La bibliographie proposée est très importante et variée mais on peut s'étonner qu'elle ne mentionne pas le livre de Henry Sheffé, *The Analysis of Variance* (Wiley, New York, 1959). C'est un excellent ouvrage fondateur, l'un des plus complets, de la présentation moderne du modèle linéaire. Scheffé a été le premier à avoir introduit les aspects géométriques du modèle linéaire.

L'ensemble est écrit dans un style vivant, en particulier quant il s'agit de l'interprétation des résultats. Les remarques, mises en garde et conseils pertinents sont faits dans un style décontracté qui doit gagner la sympathie des étudiants. JMA et JMB montrent bien que, comme la médecine, la statistique est un art qui s'appuie sur un socle scientifique. L'ensemble est bien adapté au public visé. Je pense que ce livre sera utile à beaucoup d'étudiants (et d'enseignants en mal d'exemples!).

Camille Duby

---

1. (NDLR) Ce dernier exemple a fait l'objet d'un article dans ce journal : J.-M. Azaïs (2004), Illustration de la méthode des plans d'expériences sur la comparaison de boissons au cola, *Journal de la SFdS*, 145, 4, 69-78.

# Nonparametric Functional Data Analysis Theory and Practice

Frédéric Ferraty, Philippe Vieu

1 vol., 268 pages, Springer Series in Statistics, Springer New York 2006

ISBN 0 387 30369 3

L'étude statistique de données fonctionnelles joue un rôle de plus en plus important en Théorie et dans les Applications. Elle se développe dans deux directions principales : l'Analyse des données lorsque celles-ci sont des courbes, et l'inférence statistique pour des processus à valeurs dans des espaces fonctionnels. Jusqu'à présent ces deux types d'études ne s'étaient pas rencontrés, c'est pourquoi le livre de Ferraty et Vieu comble une lacune.

Cet ouvrage adopte un point de vue non paramétrique. On sait que les méthodes non paramétriques ont fait leurs preuves en petites dimensions, notamment dans les problèmes de prévision. En grandes dimensions et a fortiori en dimension infinie, leur emploi est plus délicat à cause de la raréfaction locale des observations. Les auteurs s'attaquent à ce problème crucial et le résolvent d'une façon convaincante en introduisant des hypothèses adéquates et réalistes.

Le livre comporte cinq parties. La première pose les problèmes envisagés à partir de données fonctionnelles réelles et discute les méthodes qui seront utilisées dans la suite. La deuxième partie est consacrée à la prévision non paramétrique de variables aléatoires scalaires à partir de variables observées fonctionnelles. Les auteurs étudient le comportement asymptotique de prédicteurs associés à l'espérance conditionnelle, la médiane conditionnelle et les quantiles conditionnels. Ils s'intéressent également à la construction effective des prédicteurs utilisés. La troisième partie traite de la classification des courbes par des méthodes non paramétriques. Les méthodes pratiques d'implémentation sont indiquées. Dans la quatrième partie les auteurs obtiennent des résultats asymptotiques lorsque les variables fonctionnelles observées sont corrélées. Enfin la cinquième partie précise le rôle crucial du choix des semi-métriques et des probabilités de « petites boules » dans la vitesse de convergence des estimateurs à noyau.

L'ensemble est rédigé d'une façon très claire et détaillée, avec le souci de garder un équilibre entre résultats théoriques et mise en œuvre pratique<sup>†</sup>. C'est pourquoi je recommande fortement sa lecture aux chercheurs, aux ingénieurs et aux doctorants.

Denis Bosq

---

<sup>†</sup> Toutes les procédures développées pour la mise en œuvre sont disponibles sur le site <http://www.lsp.ups-tlse.fr/staph/npfda>, ainsi que les manuels d'utilisation et les jeux de données.