

JSFS

Avant-propos

Journal de la société française de statistique, tome 146, n° 1-2 (2005),
p. 3-4

http://www.numdam.org/item?id=JSFS_2005__146_1-2_3_0

© Société française de statistique, 2005, 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/legal.php>). 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/>

Préface

L'idée de ce numéro spécial du Journal de la SFdS sur le thème de la bio-informatique et en particulier sur l'analyse des données issues des biopuces a germé lors de la préparation des 37^{ème} journées de Statistique organisées par la SFDS à Pau en 2005. En tant qu'éditeur associé de la revue, j'ai été chargé par son éditeur en chef de coordonner le projet et de recueillir un ensemble d'articles faisant un état de l'art destiné tant à des statisticiens qui voudraient se renseigner pour répondre aux questions de leurs collègues biologistes qu'à des biologistes qui utilisent des logiciels d'analyse sans en maîtriser toujours les options.

Au cours des vingt dernières années, on a vu se multiplier et s'accélérer, dans les sciences du vivant, des découvertes considérables et dont les conséquences de tous ordres n'ont peut-être pas d'équivalent dans l'histoire des sciences. Un des exemples les plus représentatifs est celui de la génétique. Longtemps simple science d'observation de la transmission des caractères, la génétique est devenue une science d'intervention sur le végétal et l'animal d'abord, puis dans tous les champs de la biologie et de la médecine, bouleversant ainsi les pratiques médicales et agronomiques, ainsi que les industries du médicament, de l'alimentation et de l'environnement. Les progrès dans les technologies et la recherche fondamentale, l'exploitation efficace des phénomènes observés et le développement de moyens d'analyse critique proprement finalisés nécessitent l'utilisation conjointe et le développement d'outils mathématiques variés (de statistique, de simulation numérique, de traitement d'images) et une collaboration étroite des statisticiens avec des médecins, des biologistes et des bio-informaticiens.

Dans ce numéro spécial je suis très heureux de vous présenter sept articles qui soulignent l'intérêt de la modélisation statistique et de l'analyse des données en bio-informatique. Cet ensemble reflète l'état actuel de la recherche statistique, nationale et internationale, dans ce domaine.

L'article de Baccini et ses collaborateurs donne un exemple de traitement de données biopuces par différentes approches statistiques : analyse en composantes principales et multidimensional scaling, classification non supervisée, analyse canonique, analyse de la variance, modélisation par forêts aléatoires. Les auteurs exposent ainsi un panorama assez complet du traitement de données issues des biopuces à ADN, illustrant parfaitement la démarche du statisticien pour l'analyse de telles données. Bar-Hen et ses collaborateurs ont choisi de se focaliser sur les problèmes de comparaisons multiples et de leurs applications dans la recherche de gènes différentiellement exprimés, par contrôle du taux des faux positifs (FDR). Le sujet traité est d'un grand intérêt tant du point de vue des développements théoriques récents que des applications aux données génomiques. L'estimation du FDR est également abordée dans l'article de Dalmasso et Broët. Les auteurs y présentent plusieurs

PRÉFACE

méthodes couramment utilisées pour cette estimation (basées sur la distribution des degrés de signification) en génomique fonctionnelle. Les procédures de tests multiples sont explorées de manière différente dans l'article de Pollard et ses collaborateurs, en analysant la loi conjointe des statistiques de tests dans le but d'obtenir des régions critiques qui permettent un contrôle précis des erreurs de première espèce. Par ailleurs, une des principales caractéristiques des données vectorielles issues de biopuces réside dans le fait qu'on dispose d'un nombre de vecteurs très faible devant leur dimension égale au grand nombre de gènes considérés. Fort et ses collaboratrices abordent dans ce contexte des méthodes de réduction de dimension pour l'analyse discriminante à but décisionnel. L'article de Wang *et al.* explore par des méthodes d'analyse mathématique appropriées la relation existant entre certains types de cancers et le fait qu'il existe des segments qui ne sont présents que sur une copie du génome des personnes concernées. Gentleman et ses collègues abordent enfin le problème important de l'agrégation de données issues d'expériences différentes mais concernant le même phénotype afin d'extraire d'une base de données l'information la plus riche.

Je tiens à remercier toutes les personnes qui de près ou de loin ont contribué à la réalisation de ce numéro spécial, en particulier tous les auteurs qui ont bien voulu y contribuer. Un grand merci s'adresse également à l'ensemble des rapporteurs qui ont accepté d'arbitrer les manuscrits soumis dans un délai raisonnable.

J'espère que vous apprécierez la lecture de ces travaux autant que moi-même.

Anestis ANTONIADIS, Editeur invité

Ce n'est certainement pas un hasard si des journaux de statistique importants nés au siècle dernier s'appellent *Biometrika* ou *Biometrics*. Tout au long du XX^{ème} siècle en effet, la réflexion des statisticiens a été puissamment stimulée par les questions issues des sciences de la vie dans leurs divers aspects. Et le mouvement se poursuit, voire s'amplifie, en ce début de XXI^{ème} siècle avec le formidable développement de la génétique qui commande de nouveaux développements de la statistique, aussi bien des techniques exploratoires que des techniques inférentielles. Il était impératif que le Journal de la Société Française de Statistique s'en fasse l'écho et je me réjouis de la parution de ce numéro spécial où l'essentiel des thèmes de cette actualité est présenté par quelques unes des équipes importantes de ce domaine en France et à l'étranger. Je remercie vivement Anestis Antoniadis qui a pris en charge la réalisation de ce dossier avec toute l'efficacité qu'on lui connaît et je me joins à lui pour remercier les rapporteurs qui l'ont aidé dans cette tâche et, bien entendu et en tout premier lieu, l'ensemble des auteurs.

Henri CAUSSINUS