

Editorial du numéro spécial à l'occasion des 12^{èmes} Journées Européennes "Agro-Industrie et Méthodes Statistiques"

Title: Editorial of the special issue on the occasion of the 12th European Symposium on Statistical Methods for the Food

Jean-Pierre Gauchi¹

Ce numéro spécial réunit des articles représentatifs des travaux exposés lors des 12^{èmes} Journées Européennes "Agro-Industrie et Méthodes Statistiques" qui se sont déroulées du 29 février au 2 mars 2012 dans les locaux d'AgroParisTech.

L'objectif de ce congrès, typiquement français par le passé, est devenu depuis quelques années totalement ouvert à nos collègues européens. Pour cette édition de 2012, cette volonté d'ouverture à l'Europe s'est traduite par la présence de cinq conférenciers, tous européens et non français.

Depuis quelques années, aux thèmes historiques, et fondateurs, du congrès – les méthodes statistiques en agronomie, proprement dite, et en sensométrie – se sont vus adjoindre les méthodes statistiques en agro-alimentaire (procédés, microbiologie et évaluation des risques) et en chimio-métrie (méthodes d'analyse et de contrôle, plans d'expériences). Le titre de ce numéro spécial – méthodes statistiques en agronomie – recouvre tous ces aspects. Il est donc clair que le terme agronomie doit être pris ici dans une acception moderne, assez large.

Neuf articles de statistique appliquée font l'objet de ce numéro spécial. Les trois premiers articles montrent des méthodes appliquées à des exemples de l'agronomie. Le quatrième est relatif à la sensométrie, le cinquième est typique des méthodes utilisées en chimio-métrie (mais exposant des exemples d'agronomie). Les sixième et septième, dans le domaine de la microbiologie alimentaire, même s'ils traitent des sujets très différents, ont pour but, in fine, l'amélioration de l'évaluation quantitative des risques alimentaires. Le huitième montre un apport au contrôle des procédés industriels, tout à fait applicable à l'industrie agro-alimentaire. Enfin, le neuvième propose une extension de la régression PLS pour des erreurs de loi Bêta, avec des applications.

Le premier article, rédigé par Jukka Ranta, Antti Mikkilä, PirkkoTuominen et Hélène Wahlström, s'intitule «L'évaluation bayésienne des risques de salmonellose, pour des élevages de poules pondeuses de prévalence apparente nulle et un test de sensibilité évoluant au cours du temps». C'est un article très dense sur cette problématique cruciale des risques alimentaires, la

¹ INRA, Unité de Mathématiques et Informatique Appliquées (MIA), Jouy-en-Josas, France.
E-mail : Jean-Pierre.Gauchi@jouy.inra.fr

salmonellose étant une des plus fréquentes maladies rencontrées dans la production industrielle des volailles, dans le monde. Pour décrire cette prévalence les auteurs utilisent un modèle à processus de Markov caché à deux états, en temps continu. L'état infectieux d'un cheptel est traité comme une variable cachée binaire susceptible d'être détectée comme étant positive à la présence de salmonelles uniquement par des tests microbiologiques imparfaits. La sensibilité du test dépend du type d'échantillonnage et de la méthode d'analyse employés mais aussi de la phase inconnue de l'épidémie parmi les poules du cheptel.

Le deuxième article, rédigé par Robert Sabatier, Myrtille Vivien et Christelle Reynès, s'intitule «Une nouvelle proposition, l'Analyse Discriminante Multitableaux : STATIS-LDA». Le cadre général est celui où les groupes d'observations sont identiques pour tous les tableaux. La méthode est basée sur une utilisation conjointe de l'analyse factorielle discriminante (AFD) usuelle (au sens de Fisher) et l'approche STATIS bien connue, d'où le nom proposé à la méthode. Son principe repose sur la décomposition inter/intra (propre à l'AFD) de chaque bloc et la quantification de l'importance de chaque bloc pour la discrimination. Un point fort de la méthode, mis en avant par les auteurs, est sa robustesse illustrée par deux exemples convaincants, le premier à partir de simulations et le second relatif aux vins de Loire.

Le troisième article, rédigé par Aida Eslami, El Mostafa Qannari, Achim Kohler et Stéphanie Bougeard, s'intitule «Analyses factorielles de données structurées en groupes d'individus». La spécificité de ces analyses factorielles est qu'elles prennent en compte des individus en groupes structurés a priori. L'objet de base est l'analyse en composantes principales multi-groupes de Krzanowski (1984) et l'apport de la publication est essentiellement un nouveau critère d'optimisation pour mettre en œuvre cette méthode. Dans un second temps la méthode est étendue au cas des données structurées à la fois en multi-groupes et multi-tableaux. Le cas particulier où les différents tableaux portent sur les mêmes variables est également considéré. Les deux cas réels traités concernent un jeu de données sur les vins portugais – pour mettre en évidence la typologie de deux catégories de vins – et un jeu sur les huiles d'olive en vue d'explorer les relations entre les variables physico-chimiques et sensorielles conditionnellement à la provenance géographique de trois groupes d'huiles d'olive (Grèce, Espagne, Italie).

Le quatrième article, de sensométrie, rédigé par Rune Haubo Bojesen Christensen et Per Bruun Brockhoff, s'intitule «Analyse des observations des évaluations sensorielles avec des modèles à fonction de lien cumulée». Au lieu d'analyser les données des études hédonique ou de préférence par la régression ordinale, comme il est d'usage dans la communauté des sensométriciens, les auteurs proposent d'améliorer ces analyses par l'usage des modèles logistiques à fonction de lien cumulée (« cumulative link models »). Ils montrent également le lien de ceux-ci avec les tests omnibus de Pearson et comment ils peuvent conduire à des tests plus puissants dans les situations sans répétitions. Plusieurs exemples issus de la littérature de sensométrie sont traités.

Le cinquième article, de chimiométrie, rédigé par Lidwine Grosmaire, Christelle Reynès et Robert Sabatier, s'intitule «Sélection conjointe de régions de spectres MidIR et RAMAN et de variables en régression PLS à l'aide d'algorithmes génétiques». Avec ce type de sélection peu de techniques existent pour sélectionner des intervalles pour des spectres (RAMAN et Infra-Rouge moyen). La proposition des auteurs est basée sur un algorithme génétique pour la régression PLS1, spécialement adapté à ce type de données représentées par un multitableau (les deux types de spectres plus un tableau de variables physico-chimiques). La méthode est appliquée à la transformation du manioc.

Le sixième article, centré sur les plans d'expériences optimaux, rédigé par Dries Telen, Filip Logist, Eva Van Derlinden et Jan Van Impe, s'intitule «Sur le compromis entre l'effort expérimental et le contenu des informations dans les plans d'expériences optimaux pour la calibration d'un modèle de la microbiologie prédictive». En microbiologie prédictive, des modèles mathématiques et dynamiques sont développés pour décrire l'évolution microbienne en fonction des conditions environnementales. On sait que les techniques des plans d'expériences optimaux permettent d'obtenir une estimation "optimisée" des paramètres de ces modèles, mais une question complémentaire importante concerne le schéma optimal d'échantillonnage, c'est-à-dire l'effort expérimental à fournir. Dans ce travail, dans un premier temps, un schéma d'échantillonnage optimisé, basé sur une approche tout-ou-rien est implémentée. Celle-ci conduit à trouver un compromis entre l'effort expérimental et le contenu informatif dans les plans d'expériences optimaux. Dans un second temps une optimisation de type multi-objectif est mise en œuvre pour prendre en compte à la fois un critère d'optimalité de plan d'expérience et la quantification de l'effort expérimental. Des simulations sont réalisées à partir d'un modèle connu de la communauté des chercheurs en microbiologie alimentaire.

Le septième article, également d'application en microbiologie alimentaire mais orienté vers l'analyse quantitative des risques, rédigé par Laurent Guillier, Jean-Marc Kabunda, Jean-Baptiste Denis et Isabelle Albert, s'intitule «Élicitation pour l'évaluation des risques microbiologiques dans les aliments : vers une approche probabiliste de l'outil Risk Ranger». À partir du logiciel Risk Ranger mis en œuvre par une feuille de calcul Excel, les auteurs proposent de faire évoluer ce logiciel vers une version probabiliste. Cette amélioration est basée sur une procédure d'élicitation de la variabilité à partir de deux quantiles de la distribution sous-jacente à la variable d'intérêt pour chaque question posée aux experts. Les experts sont également interrogés sur leur degré de confiance pour chacun des quantiles, ce qui conduit à intégrer un niveau d'incertitude à la variable d'intérêt analysée. Ces nouveaux apports rendent la feuille Excel plus performante. Une application du logiciel ainsi amélioré est menée pour l'évaluation quantitative du risque d'intoxication par l'histamine, ce type d'intoxication étant le plus fréquent lors de la consommation des produits de la pêche en France.

Le huitième article, rédigé par Ndèye Niang, Flavio Fogliatto et Gilbert Saporta, s'intitule «Contrôle non paramétrique de procédés par lots basé sur STATIS et la classification». Les procédés par lots sont largement utilisés dans le secteur industriel notamment dans l'industrie agroalimentaire, chimique ou pharmaceutique. Les auteurs proposent une approche du contrôle de qualité des procédés par lots basée sur la méthode STATIS et des régions de contrôles non paramétriques obtenues à partir d'enveloppes convexes. Cette approche générale peut être utilisée pour le contrôle en fin de fabrication des procédés par lots ainsi que pour le contrôle en cours de fabrication après une étape de classification sous contrainte basée sur une extension multivariée de l'algorithme de W.D. Fisher. La méthode proposée est illustrée sur des données réelles issues d'un procédé par lots à temps fixe, concernant un problème de chimie industrielle.

Le neuvième article, rédigé par Frédéric Bertrand, Nicolas Meyer, Michèle Beau-Faller, Karim El Bayed, Izzie-Jacques Namer et Myriam Maumy-Bertrand, s'intitule «Régression Bêta PLS». Dans le contexte général de la régression, lorsque la réponse analysée s'exprime naturellement sous la forme de taux, de proportion ou d'indice, dont les valeurs sont nécessairement comprises entre zéro et un ou plus généralement deux valeurs fixes connues à l'avance, le cadre d'erreurs gaussiennes n'est plus adapté. Le contexte des lois Bêta semble plus propice, d'autant plus que les

fonctions de densité des lois Bêta pouvant prendre des formes très variées, une bonne souplesse est possible pour modéliser les données. En outre, si les prédicteurs présentent des problèmes de multicollinéarité (ou pire lorsqu'ils sont plus nombreux que les observations), alors la régression PLS (Partial Least Squares) est très performante. Les auteurs proposent d'utiliser conjointement les avantages des lois Bêta et de la régression PLS pour conduire à une nouvelle méthode dite de régression Bêta PLS. L'application de cette méthode de régression, peut s'avérer très utile dans de très nombreux domaines, de l'agronomie à la chimie, en passant par la médecine ou l'économétrie.