STEPHEN E. FIENBERG
PETER G. M. VAN DER HEIJDEN

Introduction to special issue on quasi-symmetry and categorical data analysis

Introduction to special issue on quasi-symmetry and categorical data analysis [*]

STEPHEN E. FIENBERG [1], PETER G.M. VAN DER HEIJDEN [2]

We are pleased and honored to serve as editors of this special issue of *Annales de la Faculté des Sciences de Toulouse, Mathématiques* which appears approximately 35 years after the publication in this journal of Henri Caussinus' seminal paper on quasi-independence and quasi-symmetry, based on his doctoral dissertation at the *Faculté des Sciences de l'Université de Toulouse*, which is now available as a downloadable portable document format (pdf) file from

> `http://www.lsp.ups-tlse.fr/Projet_QS/index.html`

This issue brings together an array of internationally distinguished authors who take stock of some of the major methodological developments in the statistical analysis of categorical data which grew out of Caussinus' work, especially that dealing with the generalizations of symmetry and quasi-symmetry. On the occasion of Caussinus' retirement as Professeur in the Laboratoire de Statistique et Probabilités, Université Paul Sabatier, the authors reflect upon the influence of Caussinus' ideas and research.

The 1960s was a decade of considerable ferment and development in the analysis of categorical data, in part because of the emergence of computers to aid in computations and in part because of the development of an integrative theory based largely on log-linear models. At the time of the publication of Caussinus [4], few statisticians in France worked with what we now know as log-linear models, while others were focusing on the development of the roots of correspondence analysis, the school of exploratory data analysis associated with the French statistician, Jean-Paul Benzecri (e.g., see [1], [2]).

Because of its explicit use of what we now know as log-linear models (in their multiplicative form) and maximum likelihood methods, Henri Caussinus' 1966 paper had a major influence on the development of the integrative log-linear model methodology that emerged in the decade following its publication. But the paper also influenced the development of exploratory approaches to the analysis of square tables in important ways, a fact documented by several papers. Thus Caussinus [4] served in a remarkable way to help draw together what appeared to be two irreconcilable approaches to statistical data analysis.

Stigler, in his contribution, reviews some of the early history of the analysis of categorical data prior to 1900 and notes the influence on Karl Pearson and others in the 20th century especially in connection with the concept of quasi-independence. Goodman picks up on this development and explains the superiority of the Caussinus approach to quasi-symmetry to a related approach suggested by Pearson, and he demonstrates the linkage between the Caussinus models and both log-linear and log-bilinear models. The latter class of models is a formalization and extension of the exploratory approaches associated with correspondence analysis. McCullagh revisits the formulation of models for the analysis of square contingency tables and demonstrates that, while there are many group representations that one might think of using for such models, only a small number of them make statistical sense: these include symmetry, quasi-symmetry, and skew-symmetry. Thélot describes his own development of quasi-symmetry as a log-linear model (in a way that parallels that in Bishop, Fienberg, and Holland, [3]), and the idea of a functional inverse for quasi symmetry for decomposing quasi-independence in square tables. These ideas turn out to be closely linked to the much more formal mathematical structure that McCullagh develops here. Thélot also explains why he abandoned this approach as he turned to repeated measures over time in the study of social mobility.

Several papers in the issue, deal with formal statistical models for multi-way tables. Agresti and Erosheva, Fienberg and Junker both demonstrate the connections between log-linear generalizations of quasi-symmetry and item response models, but they develop these ideas in different directions. Agresti focuses on ordinal variables and repeated measurement and the role this linkage plays, thus suggesting some modeling pathways to respond to Thélot's abandonment of approaches tied to quasi-symmetry. Erosheva, Fienberg, and Junker consider alternatives for the analysis of large sparse multi-way contingency tables, including generalizations of quasi-symmetry, item response models for heterogeneity, and "partial membership" latent class models. Caussinus used quasi-symmetry as a bridge between full symmetry and marginal homogeneity in two-way tables, thus suggesting the

need for tying log-linear models for multi-way tables together with marginal models. This is the theme of the paper by Bergsma and Rudas, who discuss the role of marginal log-linear models and the linkages between marginal and conditional association.

Another pair of papers in this issue develop extensions to quasi-symmetry ideas for exploratory data analysis, thus demonstrating further the natural ties between the traditional correspondence analysis approaches and the modeling approaches associated with log-linear and generalized linear models. Dossou-Gbété and Grorud use a matrix decomposition approach to tables that separates out symmetry and from skew-symmetry components and explore the use of these ideas through the use of the correspondence analysis tool of biplots for the analysis of matched two-way tables. And Falguerolles and van der Heijden describe the generalized bi-linear models of reduced-rank quasi-symmetry and quasi-skew-symmetry, and their use via biplots in the analysis of square tables.

In the final paper in the issue Henri Caussinus reflects upon the his earlier work and the developments that followed.

A number of the contributors to this issue and others have prepared brief personal statements regarding the influence of Henri Caussinus and his 1966 paper on their own research work. These are available along with the downloadable pdf file of Caussinus [4] at:

   http://www.lsp.ups-tlse.fr/Projet_QS/index.html

We thank Antoine de Falguerolles for the organization and maintenance of this website and with his assistance and collaboration in the organization and editing of this special issue.

Bibliography

[1] BENZÉCRI (J.-P.). — *L'Analyse de Données, Tome 1, La Taxinomie*, Dunod, Paris, 1973.

[2] BENZÉCRI (J.-P.). — *L'Analyse de Données, Tome 2, L'Analyse des Correspondances*, Dunod, Paris, 1973.

[3] BISHOP (Y. M. M.), FIENBERG (S. E.), HOLLAND (P. W.). — *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge, MA, 1975.

[4] CAUSSINUS (H.). — Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de l'Université de Toulouse*, 29 (année 1965), (1966), 77–183.