ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE

HENRI CAUSSINUS

Some concluding observations

Annales de la faculté des sciences de Toulouse 6e série, tome 11, nº 4 (2002), p. 587-591

<http://www.numdam.org/item?id=AFST_2002_6_11_4_587_0>

Annales de la Faculté des Sciences de Toulouse Vol. XI, n° 4, 2002
pp. 587–591

Some concluding observations

HENRI CAUSSINUS [1]

First of all I wish to thank all the authors for their friendly and authoritative contributions to this special issue of the 'Annales de la Faculté des Sciences de l'Université de Toulouse', the editor of the 'Annales' for welcoming them, Stephen Fienberg and Peter van der Heijden who served as editors of this special issue. In addition to this special issue, Antoine de Falguerolles originated the idea of a web site devoted to quasi-symmetry. Alan Agresti, Murray Aitkin, Erling Andersen, Stephen Fienberg, Antonio Forcina, Maria Kateri and Takis Papaioannou, Tamás Rudas and Wicher Bergsma, Yoshio Takane, Sadao Tomizawa, Peter van der Heijden sent interesting contributions to this site: I would like to warmly thank all of them.

Since the topic and the development of my early work in the sixties may seem surprising (see the introduction to this issue by Stephen Fienberg and Peter van der Heijden), I shall start this reply by a few considerations on its origin. Actually, at the end of the fifties, there were few statisticians in the French universities. Fortunately one of them, Roger Huron, was professor at the Faculté des Sciences de Toulouse where I was studying mathematics. Since I was fascinated by the understanding of hazard (which, I presume, came from my childhood: my father was an insurance agent and I heard the word risk much more than my turn), I took his course 'probability and statistics' to learn probabilities – I was quite unaware that statistics had something to do with mathematics! Fortunately again, Professor Huron was a wonderful and stimulating teacher and, finally, he made an offer to me to become his assistant and to undertake a research work in his laboratory. Roger Huron's basic course was a 'classical' one in the spirit of statistical modeling and inference. This was consistent with his own research in the field of genetics which involved modeling, estimation, and testing, most

[1] Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France. Email: caussinu@cict.fr

often for categorical data: see, e.g., Huron (1955), Huron (1958), Huron and Ruffié (1959). For example, the second of these papers is devoted to maximum likelihood estimation of some models while the first considers the test of specific deviations in a multinomial distribution, that is the 'analysis' of the chi-square test (in a sense, this was also to be one of my goals when dealing with truncated contingency tables). I do not remember exactly why I decided (with the help of Roger Huron) to work on contingency tables: perhaps it was from the vague feeling that was a good time to do so (thanks to Stephen Stigler's paper I am now more aware of this fact). But the main reasons for which I continued are clear: the discovery of log-linear modeling and the reading of Leo Goodman's papers. I should mention also some papers by Maurice Fréchet (1959,1960) and Pierre Thionet (1963,1964) on the existence and construction of tables with given marginals.

Meanwhile, Jean-Paul Benzécri was developing his research in Rennes, then in Paris, but I was not at all aware of his work which remained hardly available up to, say 1965 or 1966: see, e.g., Benzécri (1977). The French atmosphere of the seventies, however, was certainly important for the further orientation of my work: prevalence of techniques related to the singular value decomposition and the associated graphical displays, new points of view on the role of models in statistics, etc. Actually, one of my main interests was then to bring together the concepts of statistical modeling and data analysis as an attempt to a better understanding of statistical methods: see, for example, Caussinus and Falguerolles (1987) or Besse, Caussinus, Ferré, and Fine (1986). It is indeed a great pleasure that several of my friends continue to go deeply into this approach as evidenced by the papers by Simplice Dossou-Gbété and Axel Grorud and by Antoine de Falguerolles and Peter van der Heijden in this issue.

Coming back to my 1966 paper, since it has been published in French, it would have been obviously overlooked has it not been for Leo Goodman who wrote a review in the Mathematical Reviews (vol. 39, number 3, March 1970) and sent me a very kind letter to initiate a contact, Stephen Fienberg who gave a thorough account in his authoritative book with Yvonne Bishop and Paul Holland, and Alan Agresti who, later on, kindly emphasized the importance of this work in his books. Additionally, all of you provided further interesting contributions to quasi-symmetry including your papers in this issue. It is quite amazing to see the variety of situations in which quasi-symmetry has been used as also exemplified in the web site.

Claude Thélot recalls some applications and some developments he did in the seventies (I had then the pleasure to write a paper with him) and explains why he gave up this model for his researches on social mobility.

As for me, I also gave up the topic progressively except for a few further contributions including the paper with Claude Thélot and another one with Antoine de Falguerolles (unfortunately all are still only available in French!). I also supervised the thesis of Arnold Izsak (1972). Incidentally, Arnold Izsak found a paper analyzing migration by means of an approach implicitly equivalent to the model of quasi-symmetry (Mossin, 1971), a paper which is worth mentioning as yet another equivalent model to the ones considered in this issue. But the main purpose of Izsak's thesis was to consider the identification of the parameters, a difficult question which does not have a general answer. I provided also some hints in Caussinus (1976). In fact, apart from the French atmosphere evoked above, I think that this difficulty (which I considered as a serious drawback of the model) was the main reason for my abandoning work this topic. I continued, however, to consider the question of identification from time to time. Let me add a few words about it, mostly inspired by my above-mentioned paper.

Let us consider a square table whose entries have probabilities p_{ij}, $i, j = 1, \ldots, r$. Denote by P the $r \times r$ corresponding matrix and by T the matrix whose elements are $t_{ij} = p_{ij}/p_{i.}$ where $p_{i.}$ is the sum over j of the p_{ij}'s . In many problems (social mobility, migrations, etc.) it is natural and convenient to consider the r categories as the states of a discrete Markov chain with transition matrix T. (Note incidentally that there is equivalence between quasi-symmetry of P, or T , and reversibility of the associated Markov process: see Caussinus, 1976, McCullagh, 1982.) In practice, this chain is regular and there exists a column-vector, say ℓ, of stationary probabilities. This vector has clearly an intrinsic meaning and can be used to identify the parameters of a quasi-symmetric model in the following way. Assume that the matrix T is quasi-symmetric and write

$$t_{ij} = s_{ij}\ell_j \tag{1}$$

with $s_{ij} = s_{ji}$, which is a possible way to identify the parameters up to a multiplicative scalar. To achieve identification let us assume that the ℓ_j's add to one. Let L be the diagonal matrix of the ℓ_i's, S the matrix of the s_{ij} and $\mathbb{1}$ the column vector with coordinates equal to 1. Then $T = SL$, $T' = LS$ and $T'L\mathbb{1} = LSL\mathbb{1} = LT\mathbb{1} = L\mathbb{1}$, hence $\ell = L\mathbb{1}$. Thus, in equation (1), ℓ_j has a sensible interpretation in terms of attractivity since it is the stationary probability of category j. Coming back to P, $p_{ij} = p_{i.}s_{ij}\ell_j$ where $p_{i.}$ is the propensity to leave i at the time when the process is observed and $\ell_j/p_{j.}$ can be interpreted as the relative attractivity of j at this time. Furthermore, let us consider the transition matrix T^n for n steps. It is easy to prove that T^n is still quasi-symmetric and can be written as $T^n = S_nL$ where S_n is symmetric for all n and converges to $\mathbb{1}\mathbb{1}'$ when n tends to infinity (see Caussinus, 1976, for a proof in the more general framework of continuous time processes). If the elements of S (respectively S_n) are considered as distances (or rather as similarities) between the categories for one step (respectively n steps) this is consistent with the fact that the distance effects should decrease with the time and vanish – in practice become equal – for large n. Nevertheless, the previous approach to interpret the parameters suffers from a serious drawback: it depends both on the stayers and the movers, that is it depends on the diagonal elements of P. Of course, it can be used with a matrix P where the diagonal elements are set equal to zero, but there still remain a large degree of arbitrariness.

At this place, I would like to mention that I introduced the model of quasi-symmetry while looking for log-linear models beyond the more elementary ones. Defining such log-linear models with good properties did not seem a trivial question and Peter McCullagh's contribution to this issue brings interesting arguments that help us understand why it was not trivial. Finally, even if the model of quasi-symmetry was the main part of my early work, it is important to remember that I started from the description of association by means of log-linear models for ordinary as well as multidimensional contingency tables. Since that time, much progress has been realized, as evidenced, e. g., by the contribution of Wicher Bergsma and Tamás Rudas to this issue.

More generally, it is a great pleasure to realize that the problems I considered several years ago are now much better understood. Thanks again to you all for the best gift I could receive.

Bibliography

[1] BENZÉCRI (J.-P.). — Histoire et préhistoire de l'analyse des données, Partie V : l'analyse des correspondances, *Les Cahiers de l'Analyse des Données*, vol. II, 1 (1977), 9–40.

[2] BESSE (P.), CAUSSINUS (H.), FERRÉ (L.), FINE (J.). — Some guidelines for principal component analysis, in F. De Antoni, N. Lauro and A. Rizzi, editors, *COMPSTAT' 86*, 23–30, Physica-Verlag, Heidelberg, 1986.

[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.). — Quelques points de vue sur l'analyse des tableaux d'échange, *Annales de l'INSEE*, 22–23 (1976), 83–100.

[5] CAUSSINUS (H.), DE FALGUEROLLES (A.). — Tableaux carrés : modélisation et méthodes factorielles, *Revue de Statistique Appliquée*, 35 (1987), 35–52.

[6] FRÉCHET (M.). — Sur les tableaux de corrélation dont les marges et des cases vides sont données, *Compte Rendu à l'Académie des Sciences*, 249 (1959), 592.

[7] FRÉCHET (M.). — Sur les tableaux dont les marges et les bornes sont données, *Revue de l'Institut International de Statistique*, 28 (1960), 10–32.

[8] HURON (R.). — Loi multinomiale et test du χ^2, *Compte rendu à l'Académie des Sciences*, 240 (1955), 2047–2048.

[9] HURON (R.). — Méthode générale d'estimation de la fréquence des gènes, application aux groupes sanguins, *Annales de la Faculté des Sciences de Toulouse*, 22 (1958), 159–173.

[10] HURON (R.), RUFFIÉ (J.). — *Les méthodes en génétique générale et en génétique humaine*, Masson, Paris, 1959.

[11] IZSAK (A.). — *La quasi-symétrie dans les tables de contingence carrées*, PhD thesis, Université Paul Sabatier, Toulouse, 1972.

[12] MCCULLAGH (P.). — Some applications of quasisymmetry, *Biometrika*, 69 (1982), 303–308.

[13] MOSSIN (A.). — Mobility and equilibrium in regional migration. A theoretical model applied to Danish data, Cambridge, England, 1970, Second World Congress of the Econometric Society, Summary published in Econometrica, 39(4) (1971), 82.

[14] THIONET (P.). — Sur certaines variantes des projections du tableau d'échanges inter-industriels, *Bulletin de l'Institut International de Statistique*, XL, (1ère livraison) (1963), 431–446.

[15] THIONET (P.). — Note sur le remplissage d'un tableau à double entrée, *Journal de la Société de Statistique de Paris*, 10–11–12 (1964), 228–247.