REVUE DE STATISTIQUE APPLIQUÉE

J. DE LEEUW

Discussion

Revue de statistique appliquée, tome 35, n° 3 (1987), p. 87-89

<http://www.numdam.org/item?id=RSA_1987__35_3_87_1>

© Société française de statistique, 1987, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) 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/

J. De LEEUW

Department of Data Theory
Faculty of Social Sciences, University of Leiden
4 Middelstegracht, 2312 TW Leiden, Pays Bas

The papers in this issue approach the relationship between data analysis and modeling, or between correspondence analysis (CA) and log linear analysis (LLA), in at least three different ways. In the papers by WORSLEY and in that by BACCINI, MATHIEU, and MONDOT CA is used to prepare the data for model fitting, either by straightforward data reduction or by using CA results to suggest an appropriate model. In the paper by AITKIN, FRANCIS, and RAYNAL CA and modeling, in this case latent class modeling, are treated as equals and the results of the two techniques are compared. We shall concentrate, in these remarks, on the third approach to the relationship between the two classes of methods. CAUSSINUS and De FALGUEROLLES and also De FALGUEROLLES and VAN DER HEIJDEN take modeling as their starting point, and use CA to decompose the residuals that are left after a LLA is carried out. They do this in various interesting special cases. It is the purpose of these remarks to indicate what the general idea behind this "complementary approach" is, and in how far it can be generalized to other models.

Let us suppose that the data are a random sample of size n from a discrete distribution taking m possible values. They can be displayed in the form of random frequencies $\underline{n}_j$, with $j = 1, \ldots, m$. Suppose $\mathbf{E}(\underline{n}_j) = n\pi_j$, and suppose we have a

model which says that $\pi \varepsilon \Omega$, with Ω a p-dimensional differentiable manifold in S^{m-1}, the unit simplex in R^m. The log-likelihood of an observed vector of frequencies $\{n_j\}$ is $L = n \sum p_j \ln \pi_j$, where $p_j = n_j/n$. Maximum likelihood estimates are found by maximizing L over all $\pi \varepsilon \Omega$. This means that at the maximum likelihood estimate $\mathbf{p}$ we have that δ, with elements $\delta_i = p_i/\mathbf{p}_i$, is orthogonal to the tangent space $T_\Omega(\mathbf{p})$ of Ω at $\mathbf{p}$. Suppose $\{u_0, \ldots, u_{m-p-1}\}$ is a basis for the orthogonal complement in R^m of $T_\Omega(\mathbf{p})$. Without loss of generality we can assume that u_0 has all elements equal to a constant, and that the basis is unit orthogonal in the metric defined by the diagonal matrix $\mathbf{P}$. Thus $u_s' \mathbf{P} u_t = \delta^{st}$. This means that we can write

$$p = \mathbf{P}\{1 + a_1 u_1 + \cdots + a_{m-p-1} u_{m-p-1}\}. \tag{1}$$

Also

$$\mathbf{P}^{-1/2}(p - \mathbf{p}) = a_1 \mathbf{P}^{+1/2} u_1 + \cdots + a_{m-p-1} \mathbf{P}^{+1/2} u_{m-p-1}, \tag{2}$$

and consequently

$$(p - \mathbf{p})' \mathbf{P}^{-1} (p - \mathbf{p}) = (a_1)^2 + \cdots + (a_{m-p-1})^2. \tag{3}$$

These simple geometrical facts are actually the basis of the proof that the Pearson goodness of fit statistic has a ξ^2 distribution with $m - p - 1$ degrees of freedom. How are these results related to CA ? First observe that (2) defines a decomposition of the normalized residuals in terms of the orthonormal vectors $v_s = \mathbf{P}^{+1/2} u_s$, just as (3) decomposes the chi square. But in the derivation of (2) and (3) there is still some freedom, because obviously the basis $\{u_s\}$ can be chosen in many different ways. In the independence model for an $R \times C$ table the likelihood equations are of the form $\sum_r (p_{rc} - \mathbf{p}_{rc}) = 0$ and $\sum_c (p_{rc} - \mathbf{p}_{rc}) = 0$. Thus rows and columns of the residual add up to zero. Using the singular value decomposition of the residuals gives

$$p_{rc} = \mathbf{p}_{rc} + \sum_s \lambda_s x_s y_s,$$

which can be rewritten as

$$p_{rc} = p_{r+} p_{+c} \left\{1 + \sum_s \lambda_s \underline{x}_s \underline{y}_s\right\}, \tag{5}$$

with $\underline{x}_s$ and $\underline{y}_s$ suitable scaled versions of the orthonormal singular vectors x_s and y_s. This is exactly of the form (1). Thus we can decompose the residuals as in (4), and we can rescale the decomposition as in (5), which gives us a decomposition as in (1). This depends on the availability of the singular value decomposition, i.e. of a simple canonical form for two-way matrices, and on the particular form of the independence model.

Let us see what a similar analysis gives for GOODMAN'S RC-model. The likelihood equations are $\Sigma \mathbf{x}_r (p_{rc} - \mathbf{p}_{rc}) = 0$ and $\Sigma \mathbf{y}_c (p_{rc} - \mathbf{p}_{rc}) = 0$, where $\mathbf{x}_r$ and $\mathbf{y}_c$ are the maximum likelihood estimates of the scores. If follows that if we choose x_r and y_c orthogonal to these scores, then again the products $x_r y_c$ decompose the residuals of the RC-model. In fact we can choose x_r and y_c by computing the singular value decomposition of $p_{rc} - \mathbf{p}_{rc}$, because $\mathbf{x}_r$ and $\mathbf{y}_c$ are indeed singular vectors of this matrix, corresponding with singular value zero. Thus we can use the singular value decomposition of the residuals, which has rank not larger than min $(R, C) - 2$. Thus (4) is generalized very easily, but (5) becomes

$$p_{rc} = a_r b_c \left\{ \exp(\mathbf{x}_r \mathbf{y}_c) + \sum_s \lambda_s \underline{x}_s \underline{y}_s \right\}, \tag{6}$$

which is not of form (1). The products $\underline{x}_s \underline{y}_s$ are not orthogonal in the metric $\mathbf{p}_{rc}$, and thus the connection with chi square is not maintained.

For LLA, discussed in this issue by various authors, the likelihood equations are of the form $\sum g_{rcs}(p_{rc} - \mathbf{p}_{rc}) = 0$, where the G_s are known matrices. In the quasi-symmetry model, for instance, $R = C$, and there are $R(R-1)/2$ elementary symmetric matrices G_s (one upper diagonal element and the corresponding lower diagonal element $+ 1$). There are an additional $R + C$ matrices G_s which take care of the marginals, in the sense that they have one row or one column filled with $+ 1$ while all other elements are zero. It follows that the residuals are antisymmetric, in the sense that elements above and below the diagonal add to zero. Moreover the diagonal is filled in such a way that rows and columns add to zero. This already indicates one decomposition, but it is not a very satisfactory one in terms of separate scores for rows and columns. The more satisfactory decomposition, in this respect, is to make

$$p_{rc} = \mathbf{p}_{rc} + a_{rc}\delta^{rc} + b_{rc}, \tag{7}$$

where $b_{rc} + b_{cr} = 0$ for all c, r, and thus $b_{rr} = 0$ for all r. Residuals are decomposed in a diagonal matrix and an antisymmetric matrix. The antisymmetric part can then be decomposed by using the familiar Gower-decomposition of antisymmetric matrices. Again this generalizes the idea to use the singular value decomposition to study residuals, but it again does not preserve the close connection with chi-square of the independence model.

It seems that the complementary approach to modeling works especially nicely with the independence model, because of the availability of a simple canonical form, and because of the simple product form of the model which matches this canonical form. There are no other examples in which the complementary model works out so elegantly, except perhaps the quasi-independence model mentioned briefly by DE FALGUEROLLES and VAN DER HEIJDEN. In other cases the chi square geometry of the maximum likelihood method, and the unweighted Euclidean geometry of the singular value decomposition, cannot be matched.