EXPONENTIAL GEOMETRIC MODELS

JAN DE LEEUW

ABSTRACT. Meet the abstract. This is the abstract.

1. FRAMEWORK

The general problem is to minimize a *Poisson deviance* of the form

$$\mathscr{D}(\mu,\alpha,\beta,\eta)=\sum_{i=1}^{n}\sum_{j=1}^{m}\{\lambda_{ij}(\mu,\alpha,\beta,\eta)-n_{ij}\log\lambda_{ij}(\mu,\alpha,\beta,\eta)\},$$

where

$$\lambda_{ij}(\mu,\alpha,\beta,\eta)=\mu\alpha_i\beta_j\eta_{ij}.$$

For the Exponential Geometric Model or EGM we assume in addition that the η_{ij} are of the form

$$\eta_{ij}(X,Y)=\exp(\phi(x_i,y_j)).$$

where x_i and y_j are vectors in $\mathbb{R}^p$, and where the choice for the *EGM geometry* ϕ is any one of

$$\phi(x_i,y_j)=\begin{cases}x_i'y_j,\\ -\frac{1}{2}\|x_i-y_j\|^2,\\ -\|x_i-y_j\|.\end{cases}$$

These define, respectively, the inner product (EGM-IP), the negative squared distance (EGM-SD), and the negative distance (EGM-ND).

Remark 1. For square tables in which rows and columns refer to the same objects, we often require $X=Y$. For square symmetric tables we could also require $\alpha=\beta$. In any case, it is easy to see that unidimensional and multidimensional Rasch models, exponential models for social networks, quasi-symmetry models, exponential correspondence analysis models, ideal-point discriminant analysis models, Bradley-Terry-Luce models, and Shepard-Luce confusion matrix models are all special cases of EGM (in some cases with constraints on X and Y).

Date: May 7, 2008 — 23h 33min — Typeset in TIMES ROMAN.

Remark 2. An important class of submodels requires that α and/or β are equal to vectors with ones. To code this we distinguish, for example, the EGM-IP-$(\mu\alpha\beta)$ from the EGM-IP-$(\mu 1\beta)$ and the EGM-IP-(111) models. For given dimensionality p this means there are a total of $3\times 2\times 2\times 2=24$ models in the EGM class. Many of them are equivalent, for instance EGM-??-(1??) is equivalent to EGM-??-(μ??), except for EGM-??-(111) and EGM-??-$(\mu 11)$. There are various more subtle equivalences such as EGM-SD-(?$\alpha\beta$) being equivalent to EGM-IP-(?$\alpha\beta$).

Remark 3. For any EGM-??-(μ??) model we can minimize $\mathscr{D}$ over μ. The minimum is attained for

$$\hat\mu=\frac{n_{\bullet\bullet}}{\sum_{i=1}^n\sum_{j=1}^m\alpha_i\beta_j\eta_{ij}(X,Y)},$$

with bullets indicating sums. The minimum is equal to

$$\min_\mu\mathscr{D}(\mu,\alpha,\beta,\eta(X,Y))=$$
$$=(n_{\bullet\bullet}-n_{\bullet\bullet}\log n_{\bullet\bullet})-n_{\bullet\bullet}\sum_{i=1}^n\sum_{j=1}^m p_{ij}\log\frac{\alpha_i\beta_j\eta_{ij}(X,Y)}{\sum_{k=1}^n\sum_{\ell=1}^m\alpha_k\beta_\ell\eta_{k\ell}(X,Y)}$$

where $p_{ij}=n_{ij}/n_{\bullet\bullet}$.

For any EGM-??-($\mu\alpha$?) model we can minimize $\mathscr{D}$ over μ and α. First use the equivalence of EGM-??-($\mu\alpha$?) to EGM-??-(1α?). Now

$$\min_\alpha\mathscr{D}(1,\alpha,\beta,\eta(X,Y))=$$
$$=\sum_{i=1}^n\{n_{i\bullet}-n_{i\bullet}\log n_{i\bullet}\}-\sum_{i=1}^n n_{i\bullet}\sum_{j=1}^m p_{j|i}\log\frac{\beta_j\eta_{ij}(X,Y)}{\sum_{\ell=1}^m\beta_\ell\eta_{i\ell}(X,Y)}$$

where $p_{j|i}=n_{ij}/n_{i\bullet}$. The minimum is attained for

$$\hat\alpha_i=\frac{n_{i\bullet}}{\sum_{j=1}^m\beta_j\eta_{ij}(X,Y)}$$

Similar reasoning applies to eliminating β in EGM-??-(μ?β) models. Thus we see that the Poisson deviance functions with free row and columns and main effects can be turned into multinomial deviance functions, and consequently the EGM framework and the EGM algorithms cover these multinomial deviance functions as well.

Remark 4. As an important generalization we also introduce partitioned EGM's. In a column-partitioned EGM the index set $\{1,\cdots,m\}$ is partitioned into K subsets $\mathscr{J}_k$. The EGM is

$$\lambda_{ij}=\mu\alpha_{ij}\beta_j\eta_{ij}(X,Y),$$

where the $\alpha_{ij} = \alpha_{i\ell}$ if columns j and ℓ are in the same subset of the partitioning. Now

$$\min_{\alpha} \mathscr{D}(1,\alpha,\beta,\eta(X,Y)) =$$
$$= \sum_{i=1}^{n}\sum_{k=1}^{K}\{n_{ik} - n_{ik}\log n_{ik}\} - \sum_{i=1}^{n}\sum_{k=1}^{K} n_{ik} \sum_{j\in\mathscr{J}_k} p^k_{j|i} \log \frac{\beta_j \eta_{ij}(X,Y)}{\sum_{j\ell\in\mathscr{J}_k} \beta_\ell \eta_{i\ell}(X,Y)},$$

where $n_{ik} = \sum_{j\in\mathscr{J}_k} n_{ij}$ and $p^k_{j|i} = n_{ij}/n_{ik}$. This is also called a GGL model, where GGL stands for "Gifi Goes Logistic". If the data are K complete indicator matrices, as in the Gifi System, then all n_{ik} are equal to one. There is exactly one $j \in \mathscr{J}_k$ for which $p^k_{j|i} = 1$, for all other $j \in \mathscr{J}_k$ we have $p^k_{j|i} = 0$.

In the same way we could define row-partitioned models, or even row-column-partitioned models, but these seem less interesting given the way data are usually organized (in data frames with rows as objects and columns as variables).

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095-1554

E-mail address, Jan de Leeuw: `deleeuw@stat.ucla.edu`

URL, Jan de Leeuw: `http://gifi.stat.ucla.edu`