Making the EM algorithm for NPML estimation less sensitive to tuning parameters

Jochen Einbeck and John Hinde

Department of Mathematics, National University of Ireland, Galway, Ireland.

The nonparametric maximum likelihood (NPML) approach was introduced in Aitkin (1996a,1996b) as a tool to fit overdispersed generalized linear mixed models. The idea is to approximate the unknown and unspecified distribution of the random effect by a discrete mixture, leading to a simple expression of the marginal likelihood which can then be maximized using a standard EM algorithm. Specifically, assume there is given a set of explanatory vectors $x_1, \dots, x_n$ and a set of observations $y_1, \dots, y_n$ sampled from an exponential family distribution $f(y_i|z_i, \beta)$, where a random effect Z_i with distribution $g(z)$ is included into the linear predictor

$$\eta_i = \beta' x_i + Z_i.$$

The marginal likelihood can now be written as

$$L(\beta, g(z)) = \prod_{i=1}^{n} \int f(y_i|z_i, \beta) g(z_i)\, dz_i$$

and can be approximated by

$$\prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} f(y_i|z_k, \beta) \pi_k \right\},$$

where z_k are the mass points and π_k their masses. The parameters β, z_k and π_k can now be simultaneously estimated by the EM algorithm. A standard setting for starting values of z_k and π_k is to choose the values which would be required for Gaussian quadrature, which turns out not be the best solution in many cases. The choice of K is yet not completely automatized, usually it is augmented until the likelihood stabilizes and a further rise of K would not yield a significant improvement.

Though being acknowledged as to be generally "impressively" stable (Aitkin, 1996a), the performance and the results of the EM algorithm still depend heavily on the choice of tuning parameters, in particular the starting points. We try to solve or facilitate these problems, on the one hand by giving advices on how to choose these tuning parameters, and on the other hand by making the EM algorithm itself less sensitive to the choice of these parameters. The first approach is based on exploiting a possibly multimodal structure of the random effect distribution, the second approach on installing a damping procedure in the first iterations of the EM algorithm.

References

[1] AITKIN, M. (1996a): A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing, 6, 251–262.*

[2] AITKIN, M. (1996b): Empirical Bayes shrinkage using posterior random efect means from nonparametric maximum likelihood estimation in general random effect models. *Statistical Modelling: Proceedings of the 11th IWSM 1996, 87-94.*

Keywords

EM algorithm, Nonparametric maximum likelihood, random effects, generalized linear mixed models, overdispersion.