Journal of Classification 18:129-135 (2001)
DOI: 10.1007/s00357-001-0008-z

SOFTWARE REVIEW

GLIMMIX: Software for Estimating Mixtures and Mixtures of Generalized Linear Models

Michel Wedel

Universities of Groningen and Michigan

Abstract: GLIMMIX is a commercial WINDOWS-based computer program that implements the EM algorithm (Dempster, Laird and Rubin 1977) for the estimation of finite mixtures and mixtures of generalized linear models. The program allows for the specification of a number of distributions in the exponential family, including the normal, gamma, binomial, Poisson, and multinomial distributions. For each of those distributions, a variety of link functions can be specified to relate the expectation of the dependent variable to a linear predictor. Several statistics, including AIC, CAIC and BIC are computed to aid in model selection (cf. Akaike 1974; Bozdogan 1987), missing values are accommodated, and posterior membership probabilities are computed for cases, included or not included in the analysis. Simple discriminant type models dealing with concomitant variables to describe the classes are supported, and a random responder class can be added to the model. Various graphs are provided. A demonstration version of the program can be obtained from http://www/gamma.rug.nl. Before providing some details on the GLIMMIX software, a brief review of a few relevant issues in Mixture modelling are provided.

Author's Address: M. Wedel, Faculty of Economics University of Groningen, P.O. Box 800, 9700 AV Groningen, Netherlands, M.Wedel@eco.rug.nl; University of Michigan Business School, 701 Tappan Street, 48109 Ann Arbor, Michigan USA; e-mail: Wedel@umich.edu

1. Introduction

In finite mixture models, it is assumed that the observations of a sample arise from two or more unobserved classes, in unknown proportions, that are mixed. The purpose is to separate components of the sample and to identify the underlying classes. Mixture models present a model-based approach to clustering. They allow for hypothesis testing and estimation within the framework of standard statistical theory and provide a flexible class of models that can be tailored to a very wide range of substantive problems. Mixture regression models extend the traditional mixture approach in that they simultaneously allow for identifying classes, as well as for the estimation of a regression model within each of these classes. Details on these models are provided Titterington, Smith, and Makov (1985), McLachlan and Basford (1988), Wedel and Kamakura (1999), and McLachlan and Peel (2000). The models, statistics, and procedures below can all be found in those references.

2. Mixture And Mixture Regression Models

Assume a sample of N subjects, where on each subject K variables $y_n = (y_{nk}; n = 1,...,N; k = 1,...,K)$ are measured. The subjects are assumed to arise from a population that is a mixture of S unobserved classes, in (unknown) proportions $\pi_1,...,\pi_S$. It is not known in advance from which class a particular subject arises. Given that subject n comes from class s, the distribution function of y_n is represented by the general form $f_s(y_n \mid \theta_s)$, a member of the exponential family of distributions, where θ_s denotes the vector of all unknown parameters for class s. The unconditional distribution is obtained from the conditional distributions as:

$$f(y_n \mid \varphi) = \sum_{s=1}^{S} \pi_s f_s(y_n \mid \theta_s), \tag{1}$$

where $\varphi = (\pi', \theta')$ denotes all parameters.

Mixtures of generalized linear models (GLIMMIX) are extensions of mixture models. Here, the means of the observations in each class are to be predicted from a set of explanatory variables. Given class s the expectation of y_{nk} is denoted by μ_{nks}. Within each class the expectation is modelled as a function of a set of P explanatory variables:

$$g(\mu_{nks}) = \sum_{p=1}^{P} x_{nkp} \beta_{ps}, \tag{2}$$

where $g(\cdot)$ is a link-function, such as the identity, log, logit, and inverse functions for the normal, Poisson, binomial, and gamma distributions, respectively.

GLIMMIX models present a very flexible class of models. Applications of generalized linear models abound in the social and medical sciences, and the finite mixture version can be applied in cases where one expects the sample to be heterogeneous with respect to the regression coefficients. Various examples arise in the analysis of data from paired comparisons, count data modelling, structural equations modeling, conjoint analysis, cross sections of discrete- and continuous-time series, and the analysis of choice behaviour. Two particular extensions that arise frequently in practice are discussed below.

3. Random Responders

A problem that sometimes occurs, in particular in the analysis of preference and sensory data, is that some part of the respondents are unable or unwilling to provide responses based on careful judgment of the stimuli, but instead tend to respond in a random manner. Thus, for this subset of the respondents, there is no systematic relation between the attributes of the products and the overall judgment. To place this situation in a mixture model framework, instead of analysing the data with, say, S classes, an S+1th class is added. In this S+1st class, the model does not include a relationship between the independent and the dependent variables, but only estimates an intercept. Thus, if some part of the subjects are assigned to the random responders class, this strategy strengthens the relation of the dependent and the independent variables in the other classes.

4. Concomitant Variables

In social science research, the classes of subjects identified with GLIMMIX models are often described by background variables of the subjects (z_n) to obtain insights into the composition of the classes. Such profiling of classes is sometimes performed on the basis of the estimated a-posteriori class membership probabilities computed with Bayes rule:

$$\hat{p}_{ns} = \frac{\hat{\pi}_s f\left(y_n \mid \hat{\theta}_s\right)}{\sum_t \hat{\pi}_t f\left(y_n \mid \hat{\theta}_t\right)} \tag{3}$$

But that two-step procedure has several disadvantages. Models have been developed to accomplish this procedure in a single step, with the two most important being the concomitant variable mixture and the latent discriminant type mixture. In the first model the prior membership probability is specified directly as a function of the concomitant variables, e.g. as $\pi_{s|Z} = e^{Z\gamma_s} / \sum e^{Z\gamma_s}$. In formulating the latent discriminant type models, assumptions are made on the distribution of the concomitant variables so that they can be included in the core of the mixture, usually based on the assumption of conditional independence:

$$f(y_n, z_n \mid \varphi) = \sum_{s=1}^{S} \pi_s f_s(y_n \mid \theta_s) f_s(z_n \mid \omega_s) \ . \quad (4)$$

It can be shown that the two classes of models are closely related. While both types of models have advantages and disadvantages, the discriminant type models allow for inference on the concomitant variables, and missing observations are easier to deal with.

5. Estimation And Inference

Mixture and mixture regression models can be estimated in various ways, where most currently used methods are based on the likelihood function, which is obtained by summing equation (1) over n. Numerical maximisation of the likelihood function, for example with Newton or Quasi Netwon Methods, can be used to find the parameter estimates, but the EM algorithm (or the Stochastic or the Generalized EM algorithm) seem to be a more popular choice. More recently, researchers have used a Bayesian estimation, in which the Gibbs sampler is used to approximate the posterior distribution of the parameters (see for a review of EM and related algorithms, for example, McLachlan and Krishnan 1997). Because the likelihood is often multimodal, the investigation of local optima, based on several runs of the algorithms, is required. Statistical inference in a mixture model is based on the estimated information matrix. Likelihood ratio tests for nested models are possible, but the selection of the number of classes still awaits a fully satisfacory solution. Often, information statistics such as AIC, BIC and CAIC are used as heuristics (Akaike 1974; Bozdogan 1987).

6. Glimmix Software

GLIMMIX is a computer program for estimating exponential family mixture models and mixtures of generalized linear regression models that implements the EM algorithm for estimation. Version 2.0 includes several new features and improvements over version 1.0. GLIMMIX 2.0 runs under WINDOWS 3.1 and up, and WINDOWS NT. A demonstration version can be downloaded from the homepage of ProGAMMA: http://www.gamma.rug.nl/ catalog. The software comes with several test data sets that can be run to explore various possible options.

GLIMMIX operates in five major steps, which are activated from five buttons on the toolbar:

Step 1: Define the variables in the data set,
Step 2: Recode the variables,

Step 3: Make a selection of available cases,
Step 4: Provide the specifications for the analysis,
Step 5: View the results of the analysis.

In *Step 1*, GLIMMIX supports ASCII input data of various formats and allows one to define variables of different types, and label them for further use.

In *Step 2*, missing values can be defined and variables can be transformed, for example by taking a log, square root, and centering and/or scaling; nominal variables can be recoded.

Step 3 makes it possible to select a subset of the cases in the data, at random, on the basis of levels of categorical variables or using a text editor. This option may be useful for various types of applications. For example it is of use in *Hybrid Segmentation,* where segments are identified based on a GLIMMIX model within a-priori identified primary segments, in *Cross-Validation*, where a mixture model is to be cross-validated in two independent analyses to investigate stability, or in *Data-mining*, where because sample sizes are prohibitively large, an analysis of a random subsample is desired, or in *Data-fusion and other Missing Data Problems.* In the latter case, based on the analysis of subjects with complete data, posterior memberships are computed for all subjects. Those posterior membership probabilities serve to identify "imputation groups". For subjects with only partial data, the missing data may be imputed based on the posterior membership probabilities and class-level parameters.

Step 4 entails the specification of the mixture model (Figure 1). One may fit a standard mixture model (mixture clustering), or a regression mixture model. In the latter case one may select the independent variables to be included in the model. Several specifications of the distribution of the dependent variable are accommodated, where the options *normal, gamma, binomial, multinomial,* and *Poisson* are available and for each distribution, several Link functions can be chosen (*identity* and *log-links* for the normal and Poisson distributions, *inverse and log-links* for the gamma distribution, *identity, logit* and *probit-links* for the binomial distribution, and the *logit-link* for the multinomial distribution).
GLIMMIX 2.0 in addition allows one to estimate latent discriminant type concomitant variable models (where currently the distribution of the y and z variables most be the same), and to add a random responder class to the model to filter out subjects for whom the dependent variable is not affected by the independent variables. Since ratings scales occur very frequently in social science research, binomial rating scale models that assume a rank order logit model are accommodated. A range of numbers of classes to be estimated can be provided, as well as the number of starting values for each. Convergence parameters for the EM algorithm can be set.

In S*tep 5,* output can be inspected, i.e., estimates of the posterior proba-

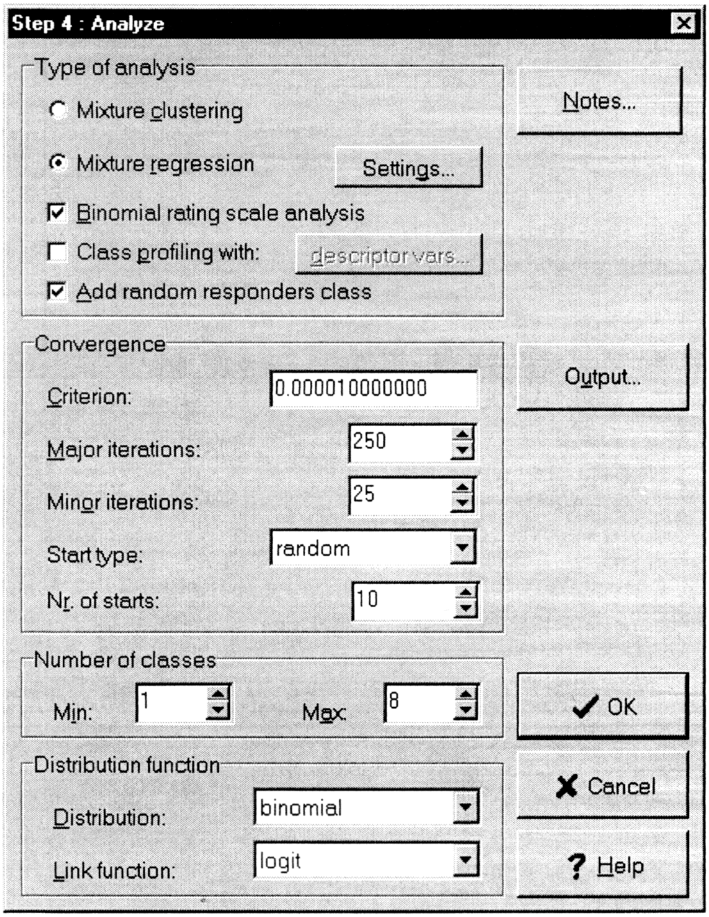

Figure 1. The GLIMMIX Analysis Screen

bilities, (where posteriors are also computed for subjects in the database that are not used in the estimation), coefficients and standard errors (the latter are currently based on an approximation), such information statistics as AIC, CAIC, and BIC. Several graphics can be displayed, e.g. histograms of the coefficient estimates, plots of the likelihood against the EM iterations, plots of residuals, and plots of the information statistics against the number of classes.

7. Conclusion

Standard mixture models can be seen as a statistical analogue to cluster methods, with the former offering the advantage of allowing for statistical inference and investigation of model fit within the framework of standard statistical theory. The application of mixtures of generalized linear models is useful in all cases where a (generalized linear) regression model is to be estimated, but the sample is suspected to be heterogeneous with respect to the coefficients of that model. The mixture regression model approach thus presents a very flexible class of models that can be tailored to a very wide range of substantive problems in the social sciences and other fields. A flexible and easy to use program such as GLIMMIX enables applied researchers in these fields to exploit fully the possibilities of this interesting class of models in practical applications.

References

AKAIKE, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control, AC-19,* 716-723.

BOZDOGAN, H. (1987), "Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions," *Psychometrika* 52, 345-370.

DEMPSTER, A. P., LAIRD, N. M., and RUBIN, R. B. (1977), "Maximum Likelihood From Incomplete Data Via the EM-algorithm," *Journal of the Royal Statistical Society, B39,* 1-38.

MCLACHLAN, G. (2000), *The EM algorithm and Extensions,* New York: Wiley.

MCLACHLAN, G.J., and BASFORD, K.E. (1988), *Mixture Models: Inference and Applications to Clustering,* New York: Marcel Dekker.

MCLACHLAN, G.J., and KRISHNAN, T. (1997), *The EM Algorithm and Extensions.* New York: Wiley..

MCLACHLAN, G.J., and PEEL, D. (2000), *Finite Mixture Models,* New York: Wiley.

TITTERINGTON, D. M., SMITH, A. F. M., and MAKOV, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions,* New York: Wiley.

WEDEL, M., and KAMAKURA, W.A. (1999), *Market Segmentation: Methodological and Conceptual Foundations* (2nd ed.), Dordrecht: Kluwer.