TO GET THE BEST, TAME THE BEAST:
ROBUST ML ESTIMATION FOR MIXTURE MODELS
Francesca Greselin

1 Department of Statistics and Quantitative Methods, University of Milano Bicocca,
(e-mail: francesca.greselin@unimib.it)

ABSTRACT: This paper presents a brief review of constrained maximization of the likelihood, in combination with data-driven trimming, as a powerful technique for achieving robust classification and clustering in mixture models. Trimming is the simpler way to achieve robustness, being also highly intuitive. Originally developed for Gaussian components, this methodology has been successfully extended to various scenarios, including parsimonious mixtures, mixtures of factor analyzers, mixtures of regression and cluster-weighted mixtures, as well as to mixtures of skew and functional data. By effectively taming the complexities associated with parameter estimation, this approach yields an estimator which exists and is strongly consistent to the corresponding solution of the population optimum under widely general conditions.

KEYWORDS: Model-based classification, robustness, trimming, constrained estimation, outliers.

1 Introduction

Mixture models offer a highly flexible approach for statistical modeling of diverse random phenomena, especially when we posit that the observations arise from unobserved groups within the population. However, estimating Gaussian (and related) mixture models using the Maximum Likelihood (ML) approach introduces two significant challenges: i) the unboundedness of the likelihood function, that sets the ML as a mathematically ill-posed problem, and ii) the presence of contaminating data (background noise, pointwise contamination, unexpected minority patterns, etc.) that could severely affect the model fitting.

To tame the likelihood, researchers have adopted two essential techniques for parameter estimation in mixture models: eigenvalue constraints and trimming. The foundation of this methodology can be traced back to García-Escudero et al., 2008, which has since evolved into a paradigm for robust model-based classification and clustering. The approach is well-regarded for its desirable theoretical properties and the availability of feasible EM algorithms for its implementation. Constrained estimation prevents convergence
towards degenerate solutions, and mitigates the occurrence of non-interesting (spurious) local maximizers associated with complex likelihood surfaces. From Hathaway’s seminal paper (Hathaway, 1985), these approaches, when coupled with impartial trimming, find applications in the context of robust statistical methods. Impartial trimming consists in excluding a small percentage of the less plausible observations, during the EM iterations, from contributing to model estimation. So doing, it protects the inferential results from the harmful effects of outliers.

Within the realm of this research stream, notable contributions include robust mixtures of factor analyzers (García-Escudero et al., 2017) and the robust cluster-weighted model (García-Escudero et al., 2016b). The introduction of their fuzzy versions (García-Escudero et al., 2018b) revealed an intriguing interplay between the fuzzifier parameter and the scale. Additionally, advancements have been made in the treatment of skew components (García-Escudero et al., 2016a). In the context of the semisupervised setting, when label noise interferes with the learning process, and whenever variable selection could be beneficial, the development of a specific robust approach is needed (Cappozzo et al., 2020b, Cappozzo et al., 2021).

However, in all such models, hyperparameter tuning remains an essential part of the inferential process, and ongoing research is focused on determining optimal settings for critical parameters such as the percentage of trimming, the number of components in the mixture, and the value for the eigenvalue constraint (Riani et al., 2019). To support practitioners in this delicate task, researchers have proposed graphical and computational tools based on the combination of two exploratory steps (Cappozzo et al., 2023).

Indeed, the literature presents several alternative methodologies for achieving robust model-based classification. Some of these approaches involve substituting Gaussian components with other elliptical distributions that possess heavier tails, such as the Student $t$ (e.g., Greselin & Ingrassia, 2010) or the contaminated normal distribution (Punzo & McNicholas, 2016). These approaches withstand the presence of mild outliers. One key concept used to assess the robustness of an estimate in the presence of outliers is the breakdown point (Hampel, 1971). Its finite sample version is the maximum fraction of outliers which a given sample may contain without spoiling the estimate completely (Donoho, 1982). Among the models with good breakdown properties, we may mention a method for cluster detection and clustering with random start forward searches (Atkinson et al., 2018), the optimally tuned robust improper maximum likelihood estimator, which uses an improper constant density for modeling outliers and noise (Coretto & Hennig, 2017), and
the weighted likelihood approach, aimed at downweighting outliers (Greco & Agostinelli, 2020). Here the weights are based on Pearson residuals stemming from robust Mahalanobis-type distances.

To conclude, the reviewed models deliver more reliable and stable estimation in the presence of outliers and noisy data, significantly enhancing model performance and facilitating more accurate statistical inference. The field of robust model-based clustering and classification is continually progressing, built upon solid results and theoretical foundations, while still presenting numerous intriguing challenges that await exploration. Exciting possibilities lie ahead, and the best is yet to come for researchers in this evolving domain.

References


