Statistical Methodology 18 (2014) 115–130

Contents lists available at ScienceDirect

Statistical Methodology

journal homepage: www.elsevier.com/locate/stamet

On the method of approximate Fisher scoring for finite mixtures of multinomials

Andrew M. Raim [a,*], Minglei Liu [b], Nagaraj K. Neerchal [a], Jorge G. Morel [c]

[a] *Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA*

[b] *Medtronic, Santa Rosa, CA, USA*

[c] *Biometrics & Statistical Sciences Department, Procter & Gamble, Cincinnati, OH, USA*

A R T I C L E   I N F O

Article history:
Received 20 December 2012
Received in revised form
9 October 2013
Accepted 9 October 2013

Keywords:
Multinomial
Finite mixture
Maximum likelihood
Fisher information matrix
Fisher scoring

A B S T R A C T

Finite mixture distributions arise naturally in many applications including clustering and inference in heterogeneous populations. Such models usually do not yield closed formulas for maximum likelihood estimates, hence numerical methods such as the well-known Fisher scoring or Expectation–Maximization (EM) algorithms are used in practice. This work considers an approximate Fisher scoring algorithm (AFSA) which has previously been used to fit the binomial finite mixture and a special multinomial finite mixture designed to handle extra variation. AFSA iterations are based on a certain matrix which approximates the Fisher information matrix. First focusing on the general finite mixture of multinomials, we show that the AFSA approach is closely related to Expectation–Maximization, and can similarly be generalized to other finite mixtures and other missing data problems. Like EM, AFSA is more robust to the choice of initial value than Fisher scoring. A hybrid of AFSA and classical Fisher scoring iterations provides the best of both computational efficiency and stable convergence properties.

1. Introduction

This paper considers an approximate Fisher scoring technique proposed by Morel and Nagaraj [12], and subsequently investigated in [14,15]. These authors used the technique to compute maximum likelihood estimates (MLEs) in the study of a multinomial model with extra variation. The model, now

* Corresponding author.
E-mail addresses: araim1@umbc.edu (A.M. Raim), nagaraj@umbc.edu (N.K. Neerchal).

1572-3127/$ – see front matter
http://dx.doi.org/10.1016/j.stamet.2013.10.002

known as the random-clumped multinomial (RCM) distribution, has made its way into mainstream use; for example, as an analytical tool in the SAS FMM procedure [18]. The RCM distribution can be written as a finite mixture of multinomials, an extension of [1,2], with specific constraints on parameters. Some details on RCM are given later in Example 3.1. Approximate Fisher scoring iterations were formulated in [12] using the observed score vector along with a certain matrix which is an approximation to the Fisher information matrix (FIM). The approximation is motivated by the difficulty in formulating the exact FIM, as it does not have an analytically tractable form and may be expensive to compute accurately by simulation (e.g. Monte Carlo). The matrix approximation has been justified by convergence results showing that the approximate FIM and exact FIM become close for large numbers of multinomial trials.

The present work shows that the approximate Fisher scoring algorithm (AFSA) is closely connected to the extremely popular Expectation–Maximization (EM) algorithm [5]. In a neighborhood of a solution, the solution is seen to be obtained by both algorithms at the same convergence rate. An explanation for the connection between the two algorithms is provided, in that the FIM approximation is actually a "complete data" information matrix. Closed-form iterations for both EM and AFSA are also obtained, giving expressions with related terms. This work focuses on the finite mixture of multinomials model, motivated by the work on RCM and noting that RCM can be obtained as a special case by enforcing some additional constraints. However, once it is established that AFSA is scoring with a complete data information matrix, its use can be justified for other finite mixture models and missing data problems. For the cases presented in this paper, an AFSA approach leads to practical procedures for computing MLEs.

A common complaint about EM in its basic form is the convergence rate, which can be slow depending on the proportion of missing data [5]. AFSA will be seen to have a similar convergence rate to EM. However, both algorithms possess a certain robustness to the initial value compared to faster methods such as Newton–Raphson or Fisher scoring, and are less likely to get stuck in neighborhoods of poor local maxima or to wander without any progress to a solution. We therefore recommend a hybrid algorithm, making use of both AFSA and exact Fisher scoring, where AFSA is used initially to progress to the neighborhood of a solution, and Fisher scoring is then used to give a fast convergence to that solution. We demonstrate that the proposed hybrid algorithm combines the best features of both AFSA and Fisher scoring.

Finite mixture models are widely used in practice and have long been studied in the statistical literature because of the analytical challenges they present. An overview of classical literature on finite mixtures is presented in [20], while [10,6] give more modern perspectives. Finite mixtures are often used to model the scenario where observations belong to one of several subpopulations, but it is unknown to which subpopulation each observation belongs. Hence, finite mixtures are a natural choice for use in clustering applications or in inference problems when overdispersion must be addressed [13]. The finite mixture of multinomials, which is the focus of this paper, has been applied to many areas including: clustering of internet traffic [8], text/topic analysis [7], item response theory for analysis of educational or psychological tests [3], and genetics [21]. Bayesian analysis of the finite mixture of multinomials is studied in [17].

The rest of the paper is organized as follows. In Section 2, the approximation to the Fisher information matrix is presented, along with some of its properties. This approximate information matrix is easily computed and has an immediate application in Fisher scoring, which is presented in Section 3. Simulation studies are presented in Section 4 to illustrate convergence properties of the approximate information matrix and approximate Fisher scoring. Concluding remarks are given in Section 5. A supplement provided along with this paper contains additional details as well as proofs for most of the results (see Appendix A).

2. An approximation to the Fisher information matrix

Consider the multinomial sample space with m trials placed into k categories at random,

$$\Omega = \left\{ (x_1, \ldots, x_k) : \; x_j \in \{0, 1, \ldots, m\}, \; \sum_{j=1}^{k} x_j = m \right\}.$$