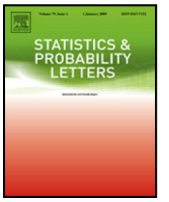
ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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



On the consistency of penalized MLEs for Erlang mixtures

Cuihong Yin [a], X. Sheldon Lin [b,*], Rongtan Huang [c], Haili Yuan [d]

[a] School of Insurance, Southwestern University of Finance and Economics, Chengdu, China
[b] Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada M5S 3G3
[c] School of Mathematical Sciences, Xiamen University, Xiamen, China
[d] School of Mathematics and Statistics, Wuhan University, Wuhan, China

ARTICLE INFO

Article history:
Received 13 July 2017
Received in revised form 29 July 2018
Accepted 5 August 2018
Available online xxxx

Keywords:
Erlang mixture
Insurance loss modeling
Penalized maximum likelihood estimates
Consistency

ABSTRACT

In Yin and Lin (2016), a new penalty, termed as iSCAD penalty, is proposed to obtain the maximum likelihood estimates (MLEs) of the weights and the common scale parameter of an Erlang mixture model. In that paper, it is shown through simulation studies and a real data application that the penalty provides an efficient way to determine the MLEs and the order of the mixture. In this paper, we provide a theoretical justification and show that the penalized maximum likelihood estimators of the weights and the scale parameter as well as the order of mixture are all consistent.

1. Introduction

The Erlang mixture model under consideration in this paper is of the following density:

$$h(x;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)=\sum_{j=1}^{m}\alpha_j g(x;\gamma_j,\theta),\ x>0, \tag{1.1}$$

where $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_m)$ is the mixing distribution and m is the number of components or the order of the mixture. Further, each component density is an Erlang of the form:

$$g(x;\gamma_j,\theta)=\frac{x^{\gamma_j-1}e^{-x/\theta}}{\theta^{\gamma_j}(\gamma_j-1)!},\ x>0, \tag{1.2}$$

with common scale parameter $\theta>0$ and positive integer shape parameter γ_j. To ensure the unique expression of density function (1.1), we assume that $\gamma_1<\gamma_2<\cdots<\gamma_m$ as we did in Yin and Lin (2016). The Erlang mixture model and its multivariate version have been widely used in modeling insurance losses due to its desirable distributional properties. For example, risk measures such as VaR and TVaR can be calculated easily. For more details on the applications, see Lee and Lin (2010), Cossette et al. (2013), Porth et al. (2014), Verbelen et al. (2015), Hashorva and Ratovomirija (2015), Verbelen et al. (2016), and references therein.

As a mixture model, an expectation–maximization (EM) algorithm is naturally used to fit the model to data by estimating the scale parameter and the mixing weights. However, the shape parameter of each of the Erlang components is not estimated. In order to include all possible Erlang distributions for component selection, one must start with a large number of components in an Erlang mixture when running the EM algorithm. Over-fitting could be a concern in this situation. To

* Corresponding author.
E-mail address: sheldon@utstat.utoronto.ca (X. Sheldon Lin).

https://doi.org/10.1016/j.spl.2018.08.004
0167-7152/

maintain the goodness of fit and to avoid over-fitting at the same time, an ad hoc method for shape parameter selection and BIC are used. See Lee and Lin (2010) and Verbelen et al. (2015). Several issues arise. First, the ad hoc method requires repeated runs of the EM algorithm, which can be computationally burdensome. Second, the chosen shape parameters are often suboptimal in terms of the order of the mixture. Third, using BIC often results in a poor fit of a model to the sparse right tail of the data, a major shortcoming in insurance loss modeling and risk measure calculation. Last, statistical properties of the corresponding estimators cannot be obtained under the ad hoc approach. Yin and Lin (2016) propose a new thresholding penalty function, termed as iSCAD, to penalize the likelihood when estimating the scale parameter and the mixing weights of the Erlang mixture. This approach is motivated by the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li, 2001) in regression analysis and the MSCAD introduced in Chen and Khalili (2008) for Gaussian mixtures. The thresholding feature of the proposed penalty ensures the sparsity of the mixture, which allows us to avoid over-fitting and maintain fitting accuracy at the same time. Moreover, the structure of the penalty results in the unbiasedness and continuity in estimation.

In this paper, we turn to the issue of consistency of the estimates including the order estimate when using the iSCAD penalized likelihood for the Erlang mixture model, as the consistency of the order is one of the most important statistical issues for mixture modeling. In the current statistics literature, most research focuses on Gaussian mixtures and a number of methods have been proposed. See Leroux (1992), James et al. (2001), Keribin (2000), Ciuperca et al. (2003), Ahn (2009), Chen et al. (2012) and references therein. However, few research have been done on non-negative non-Gaussian mixtures and few existing results may be directly applicable to the aforedescribed Erlang mixture.

In this paper, we examine the consistency of the estimators of the weight parameter and common scale parameter, as well as the order estimator, when using the iSCAD penalized likelihood. In Section 2 we introduce the iSCAD penalty, the corresponding penalized likelihood and the estimators obtained from the maximum penalized likelihood. Main results and their proofs are given in Section 3 in which we show that the estimators are consistent.

2. The iSCAD penalty

Yin and Lin (2016) propose a new penalty function termed as iSCAD, which penalizes individually the weights of an Erlang mixture. For each weight π_j, $j = 1, \ldots, m$, the iSCAD penalty function is defined as

$$\begin{aligned} P_\lambda(\pi_j) = {} & \lambda\{\log\frac{a\lambda+\varepsilon}{\varepsilon} + \frac{a^2\lambda^2}{2} - \frac{a\lambda}{a\lambda+\varepsilon}\}I(\pi_j > a\lambda) \\ & + \lambda\{\log\frac{\pi_j+\varepsilon}{\varepsilon} - \frac{\pi_j^2}{2} + (a\lambda - \frac{1}{a\lambda+\varepsilon})\pi_j\}I(\pi_j \le a\lambda), \end{aligned} \tag{2.1}$$

where λ is a tuning parameter that is a function of n with condition $\lambda \to 0$, as $n \to \infty$. $a = \frac{m}{m-\lambda} > 1$ is to ensure that the estimator $\hat{\pi}_j$ of π_j is continuous and parameter $\varepsilon = \lambda^{3/2}$ is to ensure that the range of π_j includes 0. These conditions are motivated by the conditions in Theorem 4 of Leroux (1992) to ensure to not overestimate the order of the Erlang mixture. The tuning parameter will ensure the sparsity of the mixture as it serves as a lower bound of the mixing weights. This property is crucial to avoid over-fitting and maintain fitting accuracy at the same time. Moreover, the structure of the iSCAD penalty and its derivative will result in the unbiasedness and continuity in estimation of the mixing distribution when an EM algorithm is used.

Insurance loss/claim data are mostly left truncated with known truncation points (in the form of a deductible or retention limit). See the data sets in Beirlant et al. (2006) and Verbelen et al. (2015). Suppose l to be a truncation point. Then, the probability density function of a left-truncated Erlang mixture is

$$\begin{aligned} h(x;\boldsymbol{\phi}) &= \frac{h(x;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)}{\overline{H}(l;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)} = \sum_{j=1}^{m}\alpha_j\frac{g(x;\gamma_j,\theta)}{\overline{H}(l;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)} \\ &= \sum_{j=1}^{m}\alpha_j\frac{\overline{G}(l;\gamma_j,\theta)}{\overline{H}(l;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)}\frac{g(x;\gamma_j,\theta)}{\overline{G}(l;\gamma_j,\theta)} = \sum_{j=1}^{m}\pi_j\, g_\theta(x;l,\gamma_j), \end{aligned} \tag{2.2}$$

where $\boldsymbol{\phi} = (\pi_1, \ldots, \pi_m, \theta)$. There, $\overline{H}(x;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)$ and $\overline{G}(x;\gamma_j,\theta)$ are the survival functions of $h(x;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)$ and $g(x;\gamma_j,\theta)$, respectively,

$$g_\theta(x;l,\gamma_j) = \frac{g(x;\gamma_j,\theta)}{\overline{G}(l;\gamma_j,\theta)},$$

and

$$\pi_j = \alpha_j\frac{\overline{G}(l;\gamma_j,\theta)}{\overline{H}(l;\boldsymbol{\alpha},\boldsymbol{\gamma},\theta)}. \tag{2.3}$$

Further, let $G_\theta(x;l,\gamma_j)$ be the cumulative distribution function of $g_\theta(x;l,\gamma_j)$.

Please cite this article in press as: Yin C., et al., On the consistency of penalized MLEs for Erlang mixtures. Statistics and Probability Letters (2018), https://doi.org/10.1016/j.spl.2018.08.004.