Journal of Multivariate Analysis 144 (2016) 68–80

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Reliable inference for complex models by discriminative composite likelihood estimation

Davide Ferrari *, Chao Zheng

School of Mathematics and Statistics, University of Melbourne, Australia

ARTICLE INFO

Article history:
Received 22 November 2014
Available online 11 November 2015

AMS subject classifications:
62F10
62F30
62H99
62F07

Keywords:
Composite likelihood estimation
Model selection
Exponential tilting
Stability
Robustness

ABSTRACT

Composite likelihood estimation has an important role in the analysis of multivariate data for which the full likelihood function is intractable. An important issue in composite likelihood inference is the choice of the weights associated with lower-dimensional data sub-sets, since the presence of incompatible sub-models can deteriorate the accuracy of the resulting estimator. In this paper, we introduce a new approach for simultaneous parameter estimation by tilting, or re-weighting, each sub-likelihood component called discriminative composite likelihood estimation (D-McLE). The data-adaptive weights maximize the composite likelihood function, subject to moving a given distance from uniform weights; then, the resulting weights can be used to rank lower-dimensional likelihoods in terms of their influence in the composite likelihood function. Our analytical findings and numerical examples support the stability of the resulting estimator compared to estimators constructed using standard composition strategies based on uniform weights. The properties of the new method are illustrated through simulated data and real spatial data on multivariate precipitation extremes.

1. Introduction

While likelihood-based inference is central to modern statistics, for many multivariate problems the full likelihood function is impossible to specify or its evaluation involves a prohibitive computational cost. These limitations have motivated the development of composite likelihood approaches, which avoid the full likelihood by compounding a set of low-dimensional likelihoods into a surrogate criterion function. Composite likelihood inference has proved useful in a number of fields, including geo-statistics, analysis of spatial extremes, statistical genetics, and longitudinal data analysis. See [19] for a comprehensive survey of composite likelihood theory and applications. Larribe and Fearnhead [13] review several applications in genetics.

Let X be a $d \times 1$ random vector and $f(x|\theta)$ be the assumed density model for X, indexed by the parameter $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$. Suppose that the full likelihood function, $L(\theta|x) \propto f(x|\theta)$, is difficult to specify or compute, but we can specify low-dimensional distributions with one, two, or more variables. Specifically, let $\{Y_j, j = 1, \ldots, m\}$ be a set of marginal or conditional low-dimensional variables constructed from X with associated likelihoods $L_j(\theta|y_j) \propto f_j(y_j|\theta)$, where $f_j(\cdot|\theta)$, $\theta \in \Theta$ denotes the jth low-dimensional density model for Y_j. The low-dimensional variables $\{Y_j\}$ are user-defined and could be constructed by taking marginal models, like $X_1, \ldots, X_d$, pairs like (X_1, X_2), or conditional variables like $(X_1, X_2)|X_2$. The overall structure of such lower-dimensional models is sometimes referred to as composite likelihood design [16] and its choice is often driven by computational convenience. For example, if X follows a d-variate normal distribution $N_d(0, \Sigma)$,

* Correspondence to: Richard Berry Building, University of Melbourne, Parkville, 3010, VIC, Australia.
E-mail address: dferrari@unimelb.edu.au (D. Ferrari).

http://dx.doi.org/10.1016/j.jmva.2015.10.008

the full likelihood is hard to compute when d is large due to inversion of Σ, which involves $O(d^3)$ operations. In contrast, using sub-models for variable pairs $(X_k, X_{k'})$, $1 \le k < k' \le d$, can reduce the computational burden since it involves simply inverting 2×2 partial covariance matrices.

Following [15], we define the composite likelihood function by

$$CL(\theta|w, x) = \prod_{j=1}^{m} f_j(y_j|\theta)^{w_j}, \tag{1}$$

where $\{w_j, j = 1, \ldots, m\}$ are non-negative weights, possibly depending on θ. A well-known issue in composite likelihood estimation is the selection of the weights, as their specification plays a crucial role in determining both efficiency and reliability of the resulting composite likelihood estimator [15,12,6,19,20]. Despite the importance of the weights, many statistical and computational challenges still hinder their selection [16].

This paper is concerned with the aspect of stability of composite likelihood selection. Stability occurs when the maximizer of the overall composite likelihood function $L(\theta|w)$ is not overly affected by the existence of locally optimal parameters that work only for a relatively small portion of such sub-sets, say $Y_1, \ldots, Y_{m^*}$, $m^* < m/2$. The presence of such local optima arises from the incompatibility between the assumed full-likelihood model and the m^* lower dimensional models. For example, suppose that the true distribution of X is a d-variate normal distribution with zero mean vector, unit variance and correlations $2\rho_0$ for all variable pairs, while the true correlation is ρ_0 for some small fraction of the $d(d-1)/2$ pairs. If one mistakenly assumes that all correlations are equal to ρ_0, both maximum likelihood and pair-wise likelihood estimators with uniform weight, $w_j = 1/m, j = 1, \ldots, m$, are not consistent for ρ_0 in this situation. Other examples of incompatible models are given in [20]. In applications, model compatibility is hard to detect, especially when m is large, so incompatible sub-models are often included in the composite likelihood function with detrimental effects on the accuracy of the global composite likelihood estimator.

Motivated by the above issues, we introduce the discriminative maximum composite likelihood estimator (D-McLE), a new methodology for reliable likelihood composition and simultaneous parameter estimation. The new approach computes smooth weights by maximizing the composite likelihood function for a sample of observations subject to moving a given distance, say ξ, from uniform weights. The D-McLE is regarded as a generalization of the traditional McLE. If $\xi = 0$ the D-McLE is exactly the common composite likelihood estimator with uniform weights. When $\xi > 0$, incompatible sub-models are down-weighted, thus resulting in estimators for θ with bounded worst-case bias. Our analytical findings and simulations support the validity of the proposed method compared to classic composite likelihood estimators with uniform weights. The new framework is illustrated through estimation of max-stable models, which have proved useful for describing extreme environmental occurrences as hurricanes, floods and storms [8].

The proposed procure would be useful in two respects. First, the resulting weights would be a valuable diagnostic tool for composite likelihood selection. Small weights would signal suspicious models, which could be further examined leading to improved assumptions. Conversely, the method can be employed to identify influential data sub-sets for many types of composite likelihood estimators. Second, the estimates obtained by such method would be trustworthy at least for the bulk of the data sub-sets models (which are compatible with model assumptions). Clearly, assigning the same weight to all the models including the ones in strong disagreement with the majority of data would lead to biased global estimates, which can be an untrustworthy representations of the entire data-set.

The proposed method is a type of data tilting, a general technique which involves replacing uniform weights with more general weights. To our knowledge, this is the first work that introduces tilting for lower-dimensional data sub-sets within the composite likelihood framework. In robust statistics, tilting has been typically employed to robustify parametric estimating equations, or to obtain natural data order in terms of their influence [4]. Tilting has also been used to obtain measures of outlyingness and influence of data-subsets; e.g., see [11,7,14,3]. Genton and Hall [10] use a tilting approach in the context of multivariate functional data to ranking influence of data subsets.

The rest of the paper is organized as follows. In Section 2, we describe the new methodology for simultaneous likelihood selection/estimation; we give an efficient algorithm and introduce the compatibility plot, a new graphical tool to assess the adequacy of the sub-models. In Section 3, we study the properties of the new estimator and give its limit distribution. In Section 4, we provide simulated examples in finite samples confirming our theoretical findings. In Section 5, we illustrate the new procedure to the Tasmanian rainfall spatial data on multivariate precipitation extremes. In Section 6, we conclude and discuss possible extensions for $m \to \infty$. Proofs of technical results are deferred to a separate Appendix.

2. Methodology

2.1. Composite likelihood selection

Given independent observations $X^{(1)}, \ldots, X^{(n)}$ from the true distribution $G(x)$, we construct the set of marginal or conditional low-dimensional observations $Y_j^{(1)}, \ldots, Y_j^{(n)}, j = 1, \ldots, m$, and define the weighted composite log-likelihood function

$$\ell_n(\theta|w) \equiv \sum_{j=1}^{m} w_j \ell_{nj}(\theta) \equiv \sum_{j=1}^{m} \frac{w_j}{n} \sum_{i=1}^{n} \log f_j(Y_j^{(i)}|\theta), \tag{2}$$