Statistics and Probability Letters 97 (2015) 120–124

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On full efficiency of the maximum composite likelihood estimator

E.C. Kenne Pagui, A. Salvan *, N. Sartori

University of Padova, Italy

ARTICLE INFO

Article history:
Received 1 August 2014
Received in revised form 7 November 2014
Accepted 7 November 2014
Available online 24 November 2014

Keywords:
Composite likelihood
Efficiency
Exponential family
Score function
Multivariate normal
Pseudo-likelihood

ABSTRACT

We establish conditions for full efficiency of the maximum composite likelihood estimator, related to proportionality of the full and composite score functions. A major application is in exponential family models. An illustrative example is considered.

1. Introduction

The likelihood function plays a crucial role in several approaches to statistics, mainly due to the fact that it provides an efficient summary of the data. However, in many applications, the full likelihood may not be a practical solution, either because of computational burden or due to the inability of specifying the whole joint distribution of the data. An alternative inferential tool with properties similar to those of a proper likelihood is the composite likelihood (Lindsay, 1988).

The idea behind the composite likelihood is to construct a new objective function by adequately compounding likelihood objects based on appropriate lower dimensional events in the sample space. However, as for any pseudo-likelihood, related inference procedures are affected by a loss of efficiency with respect to the full likelihood. See Varin et al. (2011) for a recent review on composite likelihood methods.

Composite likelihood methods have been receiving growing interest in many application areas in recent years, such as spatial statistics (Hjort et al., 1994; Heagerty and Lele, 1998; Varin et al., 2005; Bevilacqua et al., 2012), statistical genetics (Fearnhead and Donnelly, 2002; Larribe and Lessard, 2008; Larribe and Fearnhead, 2011), high-dimensional data (Faes et al., 2008; Gao and Song, 2010), clustered and longitudinal data (Joe and Lee, 2009).

Although in most models drawing inference from composite likelihood entails a loss of efficiency with respect to the full likelihood, in some exceptional cases, there is no efficiency loss and in particular the estimator based on a composite likelihood is identical to the maximum likelihood estimator. Mardia et al. (2009) provide an explanation for this, by showing that such identity holds for exponential families that have a certain closure property. The authors give specific further conditions for full efficiency of marginal and conditional composite likelihood. However some notable examples of full efficiency are not accounted for in the theory developed by Mardia et al. (2009).

* Corresponding author.
E-mail addresses: kenne@stat.unipd.it (E.C. Kenne Pagui), salvan@stat.unipd.it (A. Salvan), sartori@stat.unipd.it (N. Sartori).

http://dx.doi.org/10.1016/j.spl.2014.11.003
0167-7152/

Our main result is to identify new sufficient conditions for the composite likelihood estimator to coincide with the maximum likelihood estimator. A major application is in exponential family models. The class of exponential models satisfying the new condition includes the class of Mardia et al. (2009) as well as notable instances that do not fall into this latter class.

In Section 2, we give notations and definitions of the composite likelihoods. Section 3 focuses on the main results establishing the conditions for full efficiency of the maximum composite likelihood estimator. In addition, the relationship with Mardia et al. (2009)'s conditions is considered.

2. Notation

Consider a multivariate random variable $Y = (Y_1, \ldots, Y_q)$ with density $f(y; \theta)$, where $Y_j = (Y_{1j}, \ldots, Y_{nj})^{\mathrm{T}}, j = 1, \ldots, q$, $\theta \in \Theta \subseteq \mathbb{R}^d$ is an unknown parameter and $y = (y_1, \ldots, y_q)$ is the observed data. Typically, Y has independent rows. The full likelihood and log-likelihood are, respectively, $L(\theta) = f(y; \theta)$ and $\ell(\theta) = \log L(\theta)$. The score function is denoted by $U(\theta)$, the maximum likelihood estimator by $\hat{\theta}$ and Fisher information by $\imath(\theta)$.

The composite likelihood is defined through a set of marginal or conditional events $\{\mathcal{A}_1(y), \ldots, \mathcal{A}_K(y)\}$, usually related to small subsets of the data, with component likelihoods given by $L_k(\theta; y) = L_k(\theta; \mathcal{A}_k(y))$. Following Lindsay (1988), the composite likelihood, obtained by compounding these component likelihoods, is defined as

$$cL(\theta) = cL(\theta; y) = \prod_{k=1}^{K} L_k(\theta; y). \tag{1}$$

The associated composite log-likelihood is $c\ell(\theta) = \log cL(\theta)$ and its maximizer is denoted by $\hat{\theta}_C$.

Three notable examples of $c\ell(\theta)$ are the pairwise log-likelihood

$$c\ell^P(\theta) = \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \log f(y_r, y_s; \theta),$$

obtained by using only two-dimensional margins, the pairwise and full conditional log-likelihoods

$$c\ell^{PC}(\theta) = \sum_{r \neq s}^{q} \log f(y_r | y_s; \theta), \qquad c\ell^{FC}(\theta) = \sum_{r=1}^{q} \log f(y_r | y_{-r}; \theta),$$

obtained by using conditional densities. Above, y_{-r} denotes the vector of all the observations but y_r.

The fundamental argument for consistency and asymptotic normality of the composite likelihood estimator lies on the standard theory of estimating equations. The composite score function denoted by $cU(\theta) = \partial c\ell(\theta)/\partial\theta$ is an unbiased estimating equation. Under standard regularity conditions (Molenberghs and Verbeke, 2005, Chap. 9), the maximum composite likelihood estimator is consistent and asymptotically normal as $n \to \infty$, $\hat{\theta}_C \dot{\sim} N(\theta, G(\theta)^{-1})$, where $G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$ is the Godambe or sandwich information, with $H(\theta) = \mathrm{E}_\theta\{-\partial cU(\theta)/\partial\theta^{\mathrm{T}}\}$ the sensitivity matrix and $J(\theta) = \mathrm{Var}_\theta\{cU(\theta)\}$ the variability matrix.

3. Full efficiency conditions

The following results give sufficient conditions for the maximum composite likelihood estimator to coincide with the full maximum likelihood estimator.

Lemma 1. *Let $U(\theta)$ and $cU(\theta)$ be the score functions based on the full and composite log-likelihoods, respectively. If for a non stochastic $d \times d$ matrix $A(\theta)$ of full rank, we have that*

$$cU(\theta) = A(\theta)U(\theta), \tag{2}$$

then

(i) $\hat{\theta}_C = \hat{\theta}$;
(ii) $G(\theta) = \imath(\theta)$, *even though* $J(\theta) \neq H(\theta)$;
(iii) $A(\theta) = J(\theta)H(\theta)^{-1}$.

Proof. Identity (i) is obtained directly from (2). From (2), we have $\mathrm{Var}_\theta\{cU(\theta)\} = A(\theta)\mathrm{Var}_\theta\{U(\theta)\}A(\theta)^{\mathrm{T}}$, that is

$$J(\theta) = A(\theta)\imath(\theta)A(\theta)^{\mathrm{T}} \tag{3}$$

and

$$\mathrm{E}_\theta\left\{-\frac{\partial cU(\theta)}{\partial\theta^{\mathrm{T}}}\right\} = B(\theta) + A(\theta)\mathrm{E}_\theta\left\{-\frac{\partial U(\theta)}{\partial\theta^{\mathrm{T}}}\right\},$$