Statistics and Probability Letters 110 (2016) 34–38

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

On the equivalence between conditional and random-effects likelihoods in exponential families

Riccardo De Bin *

Department of Medical Informatics, Biometry and Epidemiology, University of Munich, Germany

ARTICLE INFO

Article history:
Received 20 July 2015
Received in revised form 3 November 2015
Accepted 2 December 2015
Available online 11 December 2015

Keywords:
Modified profile likelihood
Neyman–Scott problem
Stratified models

ABSTRACT

In the exponential families framework, we provide a mixing distribution which assures the equivalence between the conditional and the random-effects likelihoods, two widely used tools to make inference on a parameter of interest in the case of many nuisance parameters.

1. Introduction

The treatment of nuisance parameters is a central problem in statistical inference. When the number of nuisance parameters grows with the sample size, for example in case of stratified models with stratum-dependent nuisance parameters, some difficulties arise and the usual likelihood approaches are known to provide inconsistent results (Neyman and Scott, 1948). Among the solutions developed in literature to face this issue, the conditional likelihood and the mixture models are likely the most used inferential tools (Lindsay, 1980). It has been noted that these two approaches provide very similar results in many situations (Rice, 2008), encouraging the investigation of their relationship and of the common properties. Papers which deal with this topic are, for example, Lindsay et al. (1991), Neuhaus et al. (1994), Rice (2004, 2008).

Starting from Rice's contributions (Rice, 2004, 2008) and focusing on the exponential family framework, here we show that the use of a particular mixing distribution based on moment generating functions leads to the equivalence between the random-effects (or marginal) likelihood used in the mixture models approach and the conditional likelihood. We also show that an approximation of the mixing function leads to the equivalence with the modified profile likelihood (Barndorff-Nielsen, 1983), a higher order asymptotics tool which may be used as an approximation of either marginal or conditional likelihoods when the latter do not exist; for further information on the modified profile likelihood, see, e.g., Severini (2000, Chapter 9).

After a short methodological introduction in Section 2, we show the aforementioned equivalences in Section 3. Three examples are finally provided in Section 4.

2. Methods

Let y be a n-dimensional vector containing realizations of a random variable Y with density $p_Y(y;\psi,\lambda)$, where ψ denotes the parameter of interest and λ is a nuisance parameter. If a statistic s sufficient for λ exists, then $p_Y(y;\psi,\lambda)$ can be

* Tel.: +49 089 440077754.
E-mail address: debin@ibe.med.uni-muenchen.de.

http://dx.doi.org/10.1016/j.spl.2015.12.002
0167-7152/

rewritten as

$$p_Y(y;\psi,\lambda) = p_{Y|S=s}(y;\psi,s)p_S(y;\psi,\lambda), \tag{1}$$

where $p_{Y|S=s}(y;\psi,s)$ is called the conditional density. The conditional likelihood,

$$L_C(\psi) = p_{y|S=s}(y;\psi,s),$$

does not depend on the nuisance parameter and can be used to make inference on ψ.

A different approach for the treatment of the nuisance parameters is given by the mixture models. The mixture models are generated by considering the nuisance parameter as a random observation from an unknown distribution: inference on ψ can be then based on the likelihood

$$L_M(\psi) = \int_\Lambda p_Y(y;\psi,\lambda)g(\lambda;\psi)d\lambda, \tag{2}$$

where $g(\lambda;\psi)$ is the mixing function and Λ is the parametric space of λ. Using Eq. (1), we can rewrite $L_M(\psi)$ as

$$L_M(\psi) = p_{Y|S=s}(y;\psi,s)\int_\Lambda p_S(s;\psi,\lambda)g(\lambda;\psi)d\lambda, \tag{3}$$

where the conditional part, which does not depend on ψ, is taken out of the integral.

In this note we focus on the exponential families framework, in which

$$p_Y(y;\psi,\lambda) = \exp\{t\psi + s\lambda - K(\psi,\lambda)\}h(t,s),$$

with $K(\psi,\lambda)$ being the cumulant generating function of Y and t, s functions of y. In this case, the marginal part of Eq. (1) can be written as

$$p_S(s;\psi,\lambda) = \exp\{\lambda s - K(\psi,\lambda)\}M_s(\psi)p_0(s), \tag{4}$$

where $p_0(s)$ and $p_0(t|s)$ are the marginal densities of S and the conditional density of T given $S = s$, respectively, when $(\psi,\lambda) = (0,0)$. Moreover, $M_s(\psi)$ is the conditional moment generating function of t given $S = s$.

To obtain the equivalence between the random-effects and the conditional likelihoods, Rice (2004, 2008) suggests forcing the integral term of Eq. (3), namely $\int_\Lambda p_S(s;\psi,\lambda)g(\lambda;\psi)d\lambda$, to be independent of ψ. In particular, in the context of pair-matched case-control studies investigated in those two papers, the equivalence is achieved by identifying the mixing function $h(\lambda;\psi)$ which solves the system of equations

$$\sum_{t\in T_s} e^{\psi t}\int_\Lambda e^{\lambda s}h(\lambda;\psi)d\lambda = \text{constant} \tag{5}$$

for any value of s, i.e., that forces all the marginal probabilities of S to be independent of ψ. Here $h(\lambda;\psi) = g(\lambda;\psi)/\exp\{K(\psi,\lambda)\}$ and $\sum_{t\in T_s} e^{\psi t}$ is the expression of the moment generating function $M_s(\psi)$ in case of discrete observations, with T_s denoting the support of $p_0(t|s)$. Note that, to reach this goal, the mixing distribution must depend on ψ, making Rice's strategy different from the usual mixture models approach. This dependence upon ψ is supported by the work of Severini (2007).

3. Equivalence

Although the transformation into a moment problem as seen in Eq. (5) could seem attractive, it limits the possibility of using this technique for the cases in which the sufficient statistic s is discrete and with a finite support. Through the previous section, it is clear that the goal is to drop out the dependence on ψ from the integral term in the right-hand side of Eq. (3); but Rice (2004, 2008) exceeded that, dropping out the dependence on s as well (the integral part of Eq. (3) becomes a constant, see Eq. (5)). This is not necessary, since there is no problem if the mixing function depends on the data (Severini, 2007). In the following, we suggest a formulation for the mixing function that also assures the equivalence between the random-effect and the conditional likelihoods in the case of continuous sufficient statistics.

Let us consider the integral term of $L_M(\psi)$ in the light of Eq. (4). It is

$$\int_\Lambda \exp\{\lambda s - K(\psi,\lambda)\}M_s(\psi)p_0(s)g(\lambda;\psi)d\lambda,$$

where the only quantities that depend on ψ are $e^{-K(\psi,\lambda)}$ and $M_s(\psi)$. Therefore, it is straightforward to see that this term can be forced to be independent of ψ by choosing the mixing function

$$g(\lambda;\psi) = \frac{e^{K(\psi,\lambda)}}{M_s(\psi)}p_\lambda(\lambda) = \frac{M(\psi,\lambda)}{M_s(\psi)}p_\lambda(\lambda), \tag{6}$$

where $M(\psi,\lambda)$ is the moment generating function of Y and $p_\lambda(\lambda)$ is any distribution independent of ψ.