Working Paper Series, N. 16, November 2010

Department of Statistical Sciences
University of Padua
Italy

A note on empirical likelihoods derived from pairwise score functions

Nicola Lunardon
Department of Statistical Sciences
University of Padua
Italy

Francesco Pauli
Department of Statistical Sciences
University of Padua
Italy

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: Pairwise likelihood functions are convenient surrogates for the ordinary likelihood, useful when the latter is too difficult or even impractical to compute. One drawback of pairwise likelihood inference is that, for a multidimensional parameter of interest, the pairwise likelihood analogue of the likelihood ratio statistic does not have the standard chi-square asymptotic distribution. Invoking the theory of unbiased estimating functions, this paper proposes and discusses a computationally and theoretically attractive approach based on the derivation of empirical likelihood functions from the pairwise scores. This approach produces alternatives to the pairwise likelihood ratio statistic, which allow reference to the usual asymptotic chi-square distribution useful when the elements of the Godambe information are troublesome to evaluate or in the presence of large datasets with relative small sample sizes. Monte Carlo studies are performed in order to assess the finite-sample performance of the proposed empirical pairwise likelihoods.

Keywords: Composite likelihood, Empirical likelihood, First-order asymptotics, Likelihood ratio statistic, Unbiased estimating function, Pairwise likelihood, Godambe information, Multivariate extreme values, Correlated binary data.

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Nicola Lunardon
tel: +39 049 827 4168
lunardon@unipd.it
http://www.stat.unipd.it/~lunardon

A note on empirical likelihoods derived from pairwise score functions

Nicola Lunardon
Department of Statistical Sciences
University of Padua
Italy

Francesco Pauli
Department of Statistical Sciences
University of Padua
Italy

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: Pairwise likelihood functions are convenient surrogates for the ordinary likelihood, useful when the latter is too difficult or even impractical to compute. One drawback of pairwise likelihood inference is that, for a multidimensional parameter of interest, the pairwise likelihood analogue of the likelihood ratio statistic does not have the standard chi-square asymptotic distribution. Invoking the theory of unbiased estimating functions, this paper proposes and discusses a computationally and theoretically attractive approach based on the derivation of empirical likelihood functions from the pairwise scores. This approach produces alternatives to the pairwise likelihood ratio statistic, which allow reference to the usual asymptotic chi-square distribution useful when the elements of the Godambe information are troublesome to evaluate or in the presence of large datasets with relative small sample sizes. Monte Carlo studies are performed in order to assess the finite-sample performance of the proposed empirical pairwise likelihoods.

Keywords: Composite likelihood, Empirical likelihood, First-order asymptotics, Likelihood ratio statistic, Unbiased estimating function, Pairwise likelihood, Godambe information, Multivariate extreme values, Correlated binary data.

1 Introduction

In various modern applications, such as models with a complex dependence structure, classical likelihood-based methods may encounter both theoretical and computational problems, due to the difficulty, or even impracticability, of specifying the full likelihood function. In these situations, it is possible to resort to alternative inferential methods that are based on approximate likelihoods derived by combining marginal distributions (see Cox and Reid, 2004; Varin, 2008; Varin *et al.*, 2011).

Let Y be a q-dimensional random vector with joint density $f(y;\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^d$, $d \geq 1$. Let $y = (y^1, \ldots, y^n)$ be a random sample of size n from Y. Suppose that there is a significant difficulty in evaluating $f(y;\theta)$, and the corresponding likelihood $L(\theta)$, but that we may compute likelihoods for pairs of observations (y_h^i, y_k^i) ($i = 1, \ldots, n$; $h, k = 1, \ldots, q$, $h \neq k$). From the bivariate marginal densities $f_{hk}(\cdot, \cdot; \theta)$, we can obtain the pairwise likelihood

$$pL(\theta) = pL(\theta; y) = \prod_{i=1}^{n} \prod_{h=1}^{q-1} \prod_{k=h+1}^{q} f_{hk}(y_h^i, y_k^i; \theta)^{w_{hk}^i} , \tag{1}$$

where w_{hk}^i are non-negative weights which do not depend on the parameter θ or on y. The pairwise likelihood is a particular instance of the general class of composite likelihoods (see Varin, 2008; Varin *et al.*, 2011, for recent reviews on composite likelihoods methods). Composite likelihood contains, and thus generalizes, the usual ordinary likelihood, as well as many other alternatives, such as the pseudo-likelihood of Besag (1974) and the partial likelihood of Cox (1975).

The validity of using the pairwise likelihood to perform inference about θ can be assessed from the standpoint of unbiased estimating functions or the Kullback-Leibler criterion (for details, see Lindsay, 1988; Cox and Reid, 2004; Varin, 2008; Varin *et al.*, 2011). Under regularity conditions (Molenberghs and Verbeke, 2005), the pairwise score is an unbiased estimating function, the pairwise maximum likelihood estimator is asymptotically normal and Wald type statistics and score-type statistics have the usual asymptotic null distribution.

As it is well known, Wald-type statistics lack invariance under reparametrization and force confidence regions to have an elliptical shape. Under this respect, likelihood ratio type statistics are more appealing. However, one drawback with pairwise likelihood methods is that the null distribution of the pairwise likelihood ratio statistic does not converge to the standard chi-square distribution, but to a linear combination of independent chi-square variates with coefficients given by the eigenvalues of a matrix related to the Godambe information (Kent, 1982). Several adjustments of the pairwise likelihood ratio to approximate the usual chi-square distribution have been proposed, which typically require the computation of the elements of the Godambe information (see Satterthwaite, 1946; Wood, 1989; Geys *et al.*, 1999; Lindsay *et al.*, 2000; Chandler and Bate, 2007; Pace *et al.*, 2011).

By invoking the theory of unbiased estimating functions, in this paper we propose a computationally and theoretically attractive alternative approach which is based upon empirical likelihoods derived from the pairwise score function. In particular, we propose two versions of the empirical pairwise likelihood ratio statistic. The first one, termed hereafter $pw_e(\theta)$, is derived following Adimari and Guolo (2010) and has a standard chi-square asymptotic null distribution. The statistic $pw_e(\theta)$ is particularly appealing when the elements of the Godambe information are computationally expensive to estimate (see Varin *et al.*, 2011). The second statistic, denoted with $\overline{pw_e}(\theta)$, shares the same asymptotic distribution of the pairwise likelihood ratio statistic and can be appealing since the rate of convergence to its null distribution is, in general, fast.

The paper is organized as follows. In Section 2 pairwise likelihood methods are briefly reviewed. In Section 3 the proposed approach is discussed and in Section 4 three simulation studies are analyzed in detail in the context of the equicorrelated normal distribution, of correlated binary data and of multivariate extreme values. Simulation results indicate that the proposed statistic allow quite accurate inferences. Some concluding remarks are given in Section 5.

2 Background on pairwise likelihood

The validity of inference about θ using the pairwise likelihood function (1) can be justified invoking the theory of unbiased estimating functions. Under broad assumptions assumed in this paper (Molenberghs and Verbeke, 2005), the maximum pairwise likelihood estimator $\hat{\theta}_p$ is the solution of the pairwise score function

$$pU(\theta) = \frac{\partial \log pL(\theta)}{\partial \theta} = \sum_{i=1}^{n} \sum_{h=1}^{q-1} \sum_{k=h+1}^{q} w_{hk}^i \, pU_{hk}^i(\theta) = \sum_{i=1}^{n} \eta^i(\theta) \ , \tag{2}$$

where $pU_{hk}^i(\theta) = \partial \log f_{hk}(y_h^i, y_k^i; \theta)/\partial\theta$, $\eta^i(\theta) = \sum_{h=1}^{q-1} \sum_{k=h+1}^{q} w_{hk}^i pU_{hk}^i(\theta)$. The pairwise score $pU(\theta)$ is unbiased, i.e. $E_\theta(pU(\theta)) = 0$, since it is a linear combination of valid score functions. Moreover, the maximum pairwise likelihood estimator $\hat{\theta}_p$ is consistent and asymptotically normal with mean θ and variance

$$V(\theta) = H(\theta)^{-1} J(\theta) H(\theta)^{-1},$$

with $H(\theta) = E_\theta(-\partial pU(\theta)/\partial\theta^\top)$ and $J(\theta) = \mathrm{var}_\theta(pU(\theta)) = E_\theta\left(pU(\theta) pU(\theta)^\top\right)$. The matrix $V(\theta)^{-1}$ is known as the Godambe information matrix (Godambe, 1960). The form of $V(\theta)$ is due to the failure of the second Bartlett identity since, in general, $H(\theta) \neq J(\theta)$.

Pairwise Wald-type or score-type test statistics based on $pL(\theta)$ are straightforward to derive using consistent estimates of the matrices $H(\theta)$ and $J(\theta)$ (see Varin, 2008, for a detailed discussion), and present the standard chi-square asymptotic distribution. In particular the the Wald type test is

$$pw_w(\theta) = (\hat{\theta}_p - \theta)^\top V(\theta)^{-1} (\hat{\theta}_p - \theta)$$

while the score type statistic is

$$pw_s(\theta) = pU(\theta)^\top J(\theta)^{-1} pU(\theta).$$

On the contrary, the pairwise likelihood ratio statistic $pw(\theta) = 2(p\ell(\hat{\theta}_p) - p\ell(\theta))$, with $p\ell(\theta) = \log pL(\theta)$, does not have the standard chi-square asymptotic distribution. Indeed, the null asymptotic distribution of $pw(\theta)$ is a linear combination of independent chi-square random variables, i.e.

$$pw(\theta) \xrightarrow{d} \sum_{j=1}^{d} \omega_j Z_j^2 \ , \tag{3}$$

where $Z_1, \ldots, Z_d$ are independent standard normal variates and the coefficients $\omega_1, \ldots, \omega_d$ are the eigenvalues of the matrix $H(\theta)^{-1}J(\theta)$. In the special case $d = 1$, we have $\omega_1 = J(\theta)/H(\theta)$, so that the adjusted pairwise likelihood ratio statistic $pw_1(\theta) = pw(\theta)/\omega_1$ is asymptotically χ^2_1. For $d > 1$ several authors (see Rotnitzky Jewell, 1990; Geys *et al.*, 1999; Molenberghs and Verbeke, 2005) propose to use first order moment matching, which gives $pw_1(\theta) = pw(\theta)/\bar{\omega}$, with $\bar{\omega} = \sum_{j=1}^{d} \omega_j/d = \mathrm{tr}(H(\theta)^{-1}J(\theta))/d$. A χ^2_d approximation is used for the null distribution of $pw_1(\theta)$.

First and second moment matching gives the Satterthwaite type (Satterthwaite, 1946) adjustment suggested in Varin (2008), i.e. $pw_2(\theta) = pw(\theta)/\kappa$, with $\kappa = \sum_j w_j^2/\sum_j w_j$. The null asymptotic distribution is χ^2_ν, with $\nu = (\sum_j w_j)^2/\sum_j w_j^2$. Matching of moments up to higher order have been also be considered (Wood, 1989; Lindsay *et al.*, 2000). These corrections to $pw(\theta)$ might be inaccurate because they correct only moments of the distribution.

The corrections proposed by Chandler and Bate (2007) and Pace *et al.* (2011) are alternative methods to moment based adjustments that aim to have a statistic with the usual χ^2_d asymptotic null distribution. In particular Chandler and Bate (2007) propose the so-called vertical scaling to $pw(\theta)$, given by

$$pw_{cb}(\theta) = \frac{(\hat{\theta}_p - \theta)^\top V(\theta)^{-1}(\hat{\theta}_p - \theta)}{(\hat{\theta}_p - \theta)^\top H(\theta)(\hat{\theta}_p - \theta)} pw(\theta).$$

Pace *et al.* (2011) proposed a parametrization invariant adjustment that takes the form

$$pw_{inv}(\theta) = \frac{pU(\theta)^\top J(\theta)^{-1} pU(\theta)}{pU(\theta)^\top H(\theta)^{-1} pU(\theta)} pw(\theta).$$

All these adjustments to $pw(\theta)$ require the evaluation of $J(\theta)$ and $H(\theta)$, which can be computationally troublesome or computationally demanding to compute in some situations, such as in the example of multivariate extreme values example considered in Section 4.

3 Empirical pairwise likelihood ratio statistics

This section develops a computationally and theoretically appealing approach, called empirical pairwise likelihood. This approach is attractive in those problems where the evaluation of the matrices $H(\theta)$ and $J(\theta)$ is computationally cumbersome or when dealing with large datasets with relatively few observations.

Empirical likelihood is a pseudo-likelihood function for θ derived from a very general estimating function (see Owen, 2001, as a general reference). It is a non-parametric tool which allow to obtain a pseudo-likelihood in several contexts, which include inference for dependent data(Owen, 1991; Kolaczyk, 1994; Kitamura, 1997; Nordman and Lahiri, 2006; Nordman, 2008). The main appeal of the empirical likelihood approach is that, under suitable regularity conditions, only unbiasedness of the estimating function is required to obtain a standard asymptotic chi-square distribution for the empirical likelihood ratio statistic or for its profile counterpart.

The drawback of the empirical likelihood ratio statistic is that it may lead to unsatisfactory inferences when the sample size is relative small, since the convergence to its null distribution may be slow.

An empirical pairwise likelihood ratio statistic for θ, derived from the pairwise score function (2) can be expressed as

$$pw_e(\theta) = 2\sum_{i=1}^{n} \log\left\{1 + \xi^\top \eta^i(\theta)\right\} \ , \tag{4}$$

where the Lagrangian multiplier $\xi = \xi(\theta)$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\eta^i(\theta)}{1 + \xi^\top \eta^i(\theta)} = 0 \ .$$

Note that, alternatively, the pairwise score function (2) may be expressed as

$$pU(\theta) = \sum_{r=1}^{m} pU_r(\theta) \tag{5}$$

when considering a set of measurable events $\{A_r : r = 1, \ldots, m\}$ defined in terms of pairs of observations (y_h^i, y_k^i), with $m = nq(q-1)/2$. With this formulation, $pU_r(\theta) = \partial \log f(y \in A_r; \theta)/\partial\theta$. In (5) it is highlighted that the pairwise score function has m terms and we have $pU_1(\theta) = pU_{12}^1(\theta)$, $pU_2(\theta) = pU_{13}^1(\theta), \ldots, pU_m(\theta) = pU_{(q-1)q}^n(\theta)$. An empirical likelihood ratio statistic for θ derived from (5), is

$$\overline{pw}_e(\theta) = 2\sum_{r=1}^{m} \log\left\{1 + \bar{\xi}^\top pU_r(\theta)\right\} \ , \tag{6}$$

where the Lagrangian multiplier $\bar{\xi}$ solves

$$\frac{1}{m}\sum_{r=1}^{m} \frac{pU_r(\theta)}{1 + \bar{\xi}^\top pU_r(\theta)} = 0 \ . \tag{7}$$

Expression (2) and (5) are alternative expressions of the same score function and hence they lead to the same estimator $\hat{\theta}_p$. The difference in grouping becomes relevant when (2) and (5) are used to obtain an empirical likelihood function. Indeed, the former maximizes a distribution function with n weights, while the second one a distribution functions with m weights. In order to state the asymptotic behaviour of the pairwise empirical likelihood ratio statistics $pw_e(\theta)$ and $\overline{pw}_e(\theta)$ two propositions are given.

Proposition 1. *Consider the pairwise score function* (2)*. Then*

$$pw_e(\theta) = 2\sum_{i=1}^{n} \log\left\{1 + \xi^\top \eta^i(\theta)\right\} \xrightarrow{d} \chi^2_d.$$

Proposition 1 states the same asymptotic results of Owen (1988, 1990) for the empirical pairwise likelihood ratio $pw_e(\theta)$. In particular the leading term in the expansion of $pw_e(\theta)$ is $O_p(n^{-1})$ and the error term is of order $O_p(n^{-1/2})$. The proof can be easily obtained following the theory in Adimari and Guolo (2010). The chi-square approximation still holds when θ is partitioned as $\theta = (\psi, \lambda)$, i.e. for the profile version of $pw_e(\theta)$. In particular $pw_{ep}(\psi) = \inf_\lambda w_e(\psi, \lambda)$ still converges in distribution to a chi-square random variable.

Proposition 2. *Consider the pairwise score function* (5). *Then*

$$\overline{pw}_e(\theta) = 2\sum_{r=1}^{m} \log\left\{1 + \xi^\top pU_r(\theta)\right\} \xrightarrow{d} \sum_{j=1}^{d} \omega_j Z_j^2.$$

The proof of Proposition 2 is given in the Appendix. Proposition 2 states that the asymptotic null distribution of $pw(\theta)$ and $\overline{pw}_e(\theta)$ are the same. Hence, when $d = 1$ it is possible to recover the usual asymptotic null distribution of $\overline{pw}_e(\theta)$ considering its scaled version $\overline{pw}_{e1}(\theta) = \overline{pw}_e(\theta)/\omega_1$ (see Adimari and Guolo, 2010). For $d > 1$, it is possible to use the scaled statistic $\overline{pw}_{e1}(\theta) = \overline{pw}_e(\theta)/\bar{\omega}$. In the presence of nuisance parameters, the profile version of $\overline{pw}_e(\theta)$, given by $\overline{pw}_{ep}(\psi) = \inf_\lambda \overline{pw}_e(\psi, \lambda)$, can be easily scaled in order to have a statistic distributed as chi-square random variable. In the Appendix some remarks about the accuracy of the chi-square approximation of $\overline{pw}_{e1}(\theta)$, which depends on n, q and the dependence structure of the observations, are given.

4 Monte Carlo studies

In this section three examples are discussed in order to compare the finite-sample behaviour of the inferential procedures based on the statistics presented in Sections 2 and 3. The first example deals with a equicorrelated multivariate normal distribution, the second one considers correlated binary data and the third one focuses on multivariate extreme values. The first example considers a vector parameter and is feasible to do closed form calculations both for complete and pairwise likelihood quantities. The second and the third examples provide a framework of practical interest, where the estimation of the matrices $H(\theta)$ and $J(\theta)$ is needed and can be computationally intensive to estimate.

4.1 Multivariate normal distribution

Let us focus on the mean μ, variance σ^2, and on the correlation coefficient ρ of a multivariate normal distribution. In this case, the full likelihood function $L(\theta)$, with $\theta = (\mu, \sigma^2, \rho)$, is available and it is possible to compare the full likelihood ratio statistic $w(\theta)$, based on $L(\theta)$, with the scaled versions of $pw(\theta)$ presented in Section 2 and with the proposed pairwise empirical likelihoods, $pw_e(\theta)$ and $\overline{pw}_e(\theta)$.

Let Y be a q-variate normal random variable with $\text{corr}(Y_r, Y_s) = \rho$, for $r, s = 1, \ldots, q$, $r \neq s$ and with $\mu_r = \mu$, and $\sigma_r^2 = \sigma^2$, for $r = 1, \ldots q$. Given a sample $y = (y^1, \ldots, y^n)$, the pairwise log-likelihood is

$$p\ell(\theta) = -\frac{nq(q-1)}{2}\log\sigma^2 - \frac{nq(q-1)}{4}\log(1-\rho^2) - \frac{q-1+\rho}{2\sigma^2(1-\rho^2)}SS_W +$$
$$-\frac{q(q-1)SS_B + nq(q-1)(\bar{y}-\mu)^2}{2\sigma^2(1+\rho)},$$

where

$$SS_B = \sum_{i=1}^{n}\sum_{h=1}^{q}(y_h^i - \bar{y}^i)^2 \ , \qquad SS_W = \sum_{i=1}^{n}(\bar{y}^i - \bar{y})^2,$$

with $\bar{y}^i = \sum_{h=1}^{q} y_h^i/q$ and $\bar{y} = \sum_{i=1}^{n}\sum_{h=1}^{q} y_h^i/nq$. In order to assess the behaviour of the proposed test statistics, we ran a simulation experiment, with $n = 15, 30$ and $q = 30$, for three values of ρ, ranging from a moderate to a strong correlation. Table 1 reports the empirical coverages of equitailed confidence regions. Note that both $pw_1(\theta)$ and $\overline{pw}_{e1}(\theta)$ are multiplied by the same scale factor $1/\bar{\omega}$, which is evaluated at the pairwise maximum likelihood estimate. The results in Table 1 show that the proposed pairwise empirical likelihood statistic $\overline{pw}_{e1}(\theta)$ shows a reasonably performance in terms of coverage and is close to $w(\theta)$, $pw_1(\theta)$, $pw_s(\theta)$ and $pw_{inv}(\theta)$ when the correlation is less than 0.9. For all considered n, $\overline{pw}_{e1}(\theta)$ outperforms $pw_e(\theta)$, $pw_w(\theta)$ and $pw_s(\theta)$.

4.2 Binary data

The pairwise likelihood is particularly useful for modelling correlated binary outcomes, as discussed in Le Cessie and Van Houwelingen (1994). This kind of data arises, e.g., in the context of repeated measurements on the same subject, where a maximum likelihood analysis involve multivariate integrals whose dimension equals the cluster sizes.

Let us focus on a multivariate probit model and constant cluster sizes. In this case, the pairwise log-likelihood is

$$p\ell(\theta) = \sum_{i=1}^{n}\sum_{h=1}^{q-1}\sum_{k=h+1}^{q} \log \mathrm{pr}(Y_h^i = y_h^i, Y_k^i = y_k^i; \theta) \qquad (8)$$

(see Le Cessie and Van Houwelingen, 1994; Kuk and Nott, 2000). Pairwise likelihood inference is much simpler than using the full likelihood since it involves only bivariate normal integrals. For instance (see also Renard *et al.*, 2004), we have $\mathrm{pr}(Y_h^i = 1, Y_k^i = 1; \theta) = \Phi_2(\gamma_h^i, \gamma_k^i; \rho)$, where $\Phi_2(\cdot,\cdot;\rho)$ denotes the standard bivariate normal distribution function with correlation coefficient ρ and $\gamma_h^i = x_h^{i\top}\beta/\sigma$, with β unknown regression coefficient, σ known scale parameter and x_h^i fixed constants $(i = 1, \dots, n; h, k = 1, \dots, q)$.

Simulation results for the overall parameter $\theta = (\beta_0, \beta_1, \rho)$ are summarized in Table 2, which gives the empirical coverages for equitailed confidence regions for θ. The derivatives of (8) are not available in closed form and numerical evaluation of all the likelihood quantities involved in the simulation study has been used. The results in Table 2 show that, as in Example 1, the empirical likelihood statistic $\overline{pw}_e(\theta)$ gives quite good results for moderate sample sizes improves on $pw_w(\theta)$, $pw_s(\theta)$, $pw_{cb}(\theta)$, $pw_{pss}(\theta)$.

Table 1: Correlation coefficient: empirical coverage probabilities of equitailed confidence regions for θ based on 20.000 Monte Carlo trials

$q = 30$	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
$n = 15$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$w(\theta)$	0.891	0.943	0.987	0.889	0.941	0.987	0.888	0.941	0.987
$pw_1(\theta)$	0.838	0.890	0.949	0.839	0.892	0.952	0.845	0.899	0.959
$pw_2(\theta)$	0.865	0.919	0.972	0.863	0.919	0.972	0.869	0.924	0.976
$pw_w(\theta)$	0.809	0.860	0.924	0.776	0.831	0.900	0.715	0.767	0.837
$pw_s(\theta)$	0.906	0.947	0.983	0.906	0.947	0.983	0.905	0.948	0.983
$pw_{cb}(\theta)$	0.831	0.884	0.944	0.820	0.876	0.941	0.762	0.818	0.891
$pw_{inv}(\theta)$	0.907	0.953	0.989	0.898	0.948	0.989	0.890	0.941	0.986
$pw_e(\theta)$	0.886	0.930	0.976	0.884	0.935	0.949	0.856	0.870	0.888
$pw_{e1}(\theta)$	0.904	0.953	0.990	0.907	0.949	0.989	0.848	0.871	0.880
$n = 30$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$w(\theta)$	0.892	0.944	0.987	0.896	0.944	0.988	0.894	0.945	0.988
$pw_1(\theta)$	0.855	0.905	0.961	0.855	0.906	0.967	0.868	0.919	0.974
$pw_2(\theta)$	0.882	0.931	0.980	0.879	0.933	0.982	0.891	0.940	0.985
$pw_w(\theta)$	0.850	0.900	0.955	0.824	0.879	0.941	0.709	0.763	0.831
$pw_s(\theta)$	0.901	0.947	0.986	0.902	0.947	0.984	0.902	0.948	0.985
$pw_{cb}(\theta)$	0.861	0.914	0.967	0.852	0.908	0.963	0.743	0.796	0.869
$pw_{inv}(\theta)$	0.900	0.949	0.989	0.898	0.947	0.989	0.893	0.942	0.986
$pw_e(\theta)$	0.815	0.876	0.937	0.826	0.883	0.941	0.855	0.903	0.951
$pw_{e1}(\theta)$	0.900	0.950	0.990	0.900	0.946	0.976	0.871	0.923	0.958

4.3 Multivariate extreme values

As pointed out in Padoan *et al.* (2010), the general q-dimensional distribution function under a max-stable process representation does not permit an analytical tractable form. In this situation the specification of bivariate spatial models becomes crucial in order to write down a likelihood function to make inference with multivariate extreme values. The model used is the Gaussian extreme value model (Coles, 1993) that, for locations t_h, $t_k \in \mathbb{R}^2$ $(h, k = 1, \dots q)$, has probability function

$$\text{pr}\{Z(0) \le z_h, Z(t) \le z_k\} = \exp\left[-\frac{1}{z_h}\Phi\left(\frac{a(t)}{2} + \frac{1}{a(t)}\log\frac{z_k}{z_h}\right) + \right.$$
$$\left. - \frac{1}{z_k}\Phi\left(\frac{a(t)}{2} + \frac{1}{a(t)}\log\frac{z_h}{z_k}\right)\right],$$

Table 2: Binary data: empirical coverage probabilities of equitailed confidence regions based on 20.000 Monte Carlo trials, with $\beta_0 = 1/2$ and $\beta_1 = 1$

$q = 20$	$\rho = 0.25$			$\rho = 0.50$			$\rho = 0.75$		
$n = 50$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$pw_1(\theta)$	0.871	0.918	0.968	0.867	0.914	0.967	0.869	0.916	0.968
$pw_2(\theta)$	0.898	0.944	0.985	0.895	0.943	0.985	0.897	0.943	0.985
$pw_w(\theta)$	0.839	0.896	0.959	0.850	0.907	0.969	0.858	0.914	0.970
$pw_s(\theta)$	0.861	0.913	0.969	0.869	0.922	0.975	0.878	0.930	0.978
$pw_{cb}(\theta)$	0.843	0.902	0.965	0.863	0.914	0.972	0.863	0.920	0.979
$pw_{inv}(\theta)$	0.860	0.911	0.967	0.867	0.920	0.972	0.875	0.927	0.975
$pw_e(\theta)$	0.875	0.928	0.978	0.886	0.935	0.982	0.886	0.936	0.982
$\overline{pw}_{e1}(\theta)$	0.878	0.924	0.973	0.872	0.920	0.971	0.873	0.920	0.971
$n = 100$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$pw_1(\theta)$	0.876	0.922	0.969	0.877	0.923	0.971	0.872	0.920	0.972
$pw_2(\theta)$	0.900	0.947	0.985	0.903	0.948	0.987	0.900	0.947	0.987
$pw_w(\theta)$	0.867	0.923	0.976	0.878	0.932	0.983	0.880	0.932	0.981
$pw_s(\theta)$	0.875	0.929	0.978	0.886	0.938	0.983	0.887	0.940	0.983
$pw_{cb}(\theta)$	0.874	0.930	0.984	0.884	0.939	0.988	0.885	0.940	0.990
$pw_{inv}(\theta)$	0.873	0.928	0.976	0.884	0.935	0.980	0.885	0.938	0.981
$pw_e(\theta)$	0.894	0.945	0.987	0.894	0.948	0.990	0.896	0.947	0.988
$\overline{pw}_{e1}(\theta)$	0.880	0.925	0.971	0.879	0.925	0.972	0.874	0.922	0.973

where $\Phi(\cdot)$ is the standard gaussian distribution function, $t = (t_h - t_k)^\top$, $a(t) = (t^\top \Sigma t)^{1/2}$ and Σ is the covariance matrix with covariance σ_{hk} and standard deviations σ_h, $\sigma_k > 0$. The function $a(t)$ measures the strength of the extremal dependence between z_h and z_k: as $a(t)$ tends to zero we have complete dependence while as $a(t)$ tends to infinity we have complete independence. Data were simulated over regular grids of 8×8 and 30×30 locations and the couples considered in the pairwise likelihood function were all those satisfying $e(t) = (t^\top t)^{1/2} \leq d$ with $d > 0$ fixed constant. This condition implies that the pairwise likelihood function (1) has weights $w_{hk}^i = \{0, 1\}$ in accordance to $e(t)$. In our simulation study we set $\sigma_h = \sigma_1 = 2000$, $\sigma_k = \sigma_2 = 3000$, $\sigma_{hk} = \sigma_{12} = 1500$ and $d = 3$. The results for this example are summarized in Table 3, with $\theta = (\sigma_1, \sigma_2, \sigma_{12})$, and highlight a reasonable performance of the proposed statistics. The $\overline{pw}_e(\theta)$ seems to be more accurate when the dimension of the grid grows, in accordance with the proof given in the Appendix. As expected, $pw_e(\theta)$ needs relatively large sample sizes to be comparable with the scaled versions of $pw(\theta)$ and $\overline{pw}_e(\theta)$ but in this example a jack-knife estimate of $J(\theta)$ is cumbersome to obtain since the maximization step required to obtain $\hat{\theta}_p$ is time demanding: in the 30×30 grid framework with $n = 150$ the couples involved in the pairwise likelihood function are 2.975.400.

Table 3: Multivariate extreme values: empirical coverage probabilities of equitailed confidence regions based on 1.000 Monte Carlo trials

n	50			100			150		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
	8×8 grid								
$pw_1(\theta)$	0.871	0.922	0.967	0.884	0.927	0.979	0.884	0.922	0.979
$pw_2(\theta)$	0.889	0.939	0.983	0.899	0.943	0.986	0.902	0.945	0.983
$pw_{cb}(\theta)$	0.829	0.899	0.969	0.860	0.928	0.979	0.862	0.926	0.979
$pw_{inv}(\theta)$	0.771	0.818	0.902	0.840	0.879	0.941	0.835	0.878	0.943
$pw_e(\theta)$	0.820	0.892	0.952	0.866	0.925	0.976	0.869	0.930	0.980
$\overline{pw}_{e1}(\theta)$	0.851	0.896	0.946	0.875	0.918	0.965	0.880	0.923	0.965
	30×30 grid								
$pw_1(\theta)$	0.854	0.909	0.969	0.875	0.932	0.985	0.891	0.930	0.971
$pw_2(\theta)$	0.871	0.930	0.981	0.887	0.947	0.992	0.903	0.941	0.985
$pw_{cb}(\theta)$	0.830	0.897	0.961	0.864	0.922	0.981	0.874	0.914	0.970
$pw_{inv}(\theta)$	0.798	0.850	0.912	0.829	0.887	0.945	0.855	0.902	0.945
$pw_e(\theta)$	0.827	0.879	0.954	0.857	0.918	0.981	0.873	0.921	0.978
$\overline{pw}_{e1}(\theta)$	0.865	0.904	0.951	0.896	0.932	0.974	0.895	0.939	0.973

5 Concluding remarks

In this paper, the possibility of deriving empirical likelihoods from a pairwise score function has been investigated. This approach offers a new attractive computational method to derive likelihood ratio type test statistics. The simulation results in Section 4 indicate that the proposed $\overline{pw}_{e1}(\theta)$ and $pw_e(\theta)$ can be useful to make inference in complex models. Moreover, for multidimensional θ and for large q, $\overline{pw}_{e1}(\theta)$ appears preferable than $pw_e(\theta)$, but it must be noticed that the latter can be used when both $\overline{pw}_{e1}(\theta)$ and $pw_1(\theta)$ are not available. For example, when analysing long sequences of genetic or spatial data, it may happen that the matrix $H(\theta)$ is too large and its inverse can be computational troublesome to compute. As a final remark, we note that the proposed pairwise empirical likelihoods may be readily extended to general composite score functions providing inferential tools alternative to composite likelihood functions.

Appendix

Proof of Proposition 2

Following Cox and Reid (2004) we formally expand $\eta^i(\hat{\theta}_p)$ around θ, up to the first order,

$$\eta^i(\theta) - (\hat{\theta}_p - \theta)^\top \partial\eta^i(\theta)/\partial\theta \doteq 0.$$

The second term is $O_p(q^2)$ while the order of the first term is $O_p(q^k)$. The constant $k \in [1, 2]$ accommodates for the dependence structure of the data (see Cox and Reid, 2004).

Since the $\eta^i(\theta)$ are independent we have

$$\sum_{i=1}^{n} \left\{\eta^i(\theta) - (\hat{\theta}_p - \theta)^\top \partial\eta^i(\theta)/\partial\theta\right\} = pU(\theta) + (\hat{\theta}_p - \theta)^\top \partial pU(\theta)/\partial\theta = \tag{9}$$

$$= O_p(n^{1/2}q^k) + O_p(n^{-1/2}q^{k-2})O_p(nq^2).$$

Let us denote the second moment of $pU(\theta)$ with $J_p(\theta)$ and minus the expected value of its first derivative with $H_p(\theta)$. The estimator of $J_p(\theta)$ supplied by (5) is $\hat{J}_p(\theta) = \sum_{r=1}^{m} pU_r(\theta)pU_r(\theta)^\top$, while that from (2) is $\tilde{J}_p(\theta) = \sum_{i=1}^{n} \eta^i(\theta)\eta^i(\theta)^\top$. Thus, $\hat{J}_p(\theta)$ converges to $H_p(\theta)$ and $\tilde{J}_p(\theta)$ converges to $J_p(\theta)$.

A McLaurin series expansion of (7) yields

$$\bar{\xi} = \hat{J}_p(\theta)^{-1}pU(\theta) + O_p(n^{-1}q^{2k-4}) = O_p(n^{-1/2}q^{k-2}),$$

and both $-\partial pU(\theta)/\partial\theta^\top$ and $\hat{J}_p(\theta)$ converge to $H_p(\theta)$. The order of $pU(\theta)$ and $\partial pU(\theta)/\partial\theta^\top$ can be derived from (9).

The expansion for $\overline{pw}_e(\theta)$ is

$$\begin{aligned}
\overline{pw}_e(\theta) &= 2\sum_{r=1}^{m} \log\left(1 + \bar{\xi}^\top pU_r(\theta)\right) = \\
&= 2\left(\bar{\xi}^\top pU(\theta) - \frac{1}{2}\xi^\top \hat{J}_p(\theta)\xi\right) + O_p\left(n^{-1/2}q^{3k-4}\right) = \\
&= pU(\theta)\hat{J}_p(\theta)^{-1}pU(\theta)^\top + O_p\left(n^{-1/2}q^{3k-4}\right) = \\
&= \left\{\tilde{J}_p(\theta)^{-1/2}pU(\theta)\right\}\tilde{J}_p(\theta)\hat{J}_p(\theta)^{-1}\left\{\left(\tilde{J}_p(\theta)^{-1/2}pU(\theta)\right)^\top\right\} + O_p\left(n^{-1/2}q^{3k-4}\right) ,
\end{aligned}$$

where $\tilde{J}_p(\theta)^{1/2}\tilde{J}_p(\theta)^{1/2} = \tilde{J}_p(\theta)$. The result stated in Proposition 2 follows since, in the last equality, we have a quadratic form in normal random variables.

In order to use the statistic $\overline{pw}_e(\theta)$ for inference it might be necessary to use its scaled version, i.e. $\overline{pw}_{e1}(\theta)$. It is easy to show that $\bar{\omega} = \mathrm{tr}(\tilde{J}_p(\theta)\hat{J}_p(\theta)^{-1})/d = O_p(q^{2k-2})$ and hence the remainder term of the scaled statistic $\overline{pw}_{e1}(\theta)$ is $O_p\left(n^{-1/2}q^{k-2}\right)$. This highlights that, as the correlation strengthen, hence k moves from 1 to 2, the convergence will be slower. Indeed, if $q = O(n^{1/2})$ and $k = 1$, the remainder term is $O_p(n^{-1})$, while if $k = 2$ it is $O_p(n^{-1/2})$.

References

Adimari, G., Guolo, A. (2010). A note on the asymptotic behaviour of empirical likelihood statistics. *Stat. Meth. Appl.*, **19**, 463–476.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B*, **4**, 192–236.

Chandler, R., Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167–183.

Coles, S. (1993). Regional modelling of extreme storms via max-stable processes. *J. Roy. Statist. Soc. Ser. B*, **55**, 797–816.

Cox, D. (1975). Partial likelihood. *Biometrika*, **62**, 269–276.

Cox, D., Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737.

Geys, H., Molenberghs, G., Ryan, L. (1999). Pseudolikelihood modelling of multivariate outcomes in developmental toxicology. *J. Amer. Statist. Assoc.*, **94**, 734–745.

Godambe, V. (1960). An optimum property of regular maximum likelihood equation. *Ann. Statist.*, **31**, 1208–1211.

Kent, J. (1982). Robust properties of likelihood ratio tests. *Biometrika*, **69**, 19–27.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084–2102.

Kolaczyk, E. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica*, **4**, 199–218.

Kuk, A., Nott, D. (2000). A pairwise likelihood approach to analyzing correlated binary data. *Stat. Prob. Lett.*, **47**, 329–335.

Le Cessie, S., Van Houwelingen, J. (1994). Logistic regression for correlated binary data. *Appl. Statist.*, **43**, 95–108.

Lindsay, B. (1988). Composite likelihood methods. *Contemp. Math.*, **80**, 221–240.

Lindsay, B., Pilla, R., Basak, P. (2000). Moment-based approximations of distributions using mixtures: theory and applications. *Ann. Inst. Statist. Math*, **52**, 215–230.

Molenberghs, G., Verbeke, G. (2005). *Models for Discrete Longitudinal Data.* New York: Springer.

Nordman, D. (2008). A blockwise empirical likelihood for spatial lattice data. *Statist. Sinica*, **18**, 1111–1129.

Nordman, D., Lahiri, S. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.*, **34**, 3019–3050.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Math. Statist.*, **18**, 90–120.

Owen, A. (1991). Empirical likelihood for linear models. *Ann. Statist.*, **19**, 1725–1747.

Owen, A. (2001). *Emprirical likelihood.* London: Chapman and Hall.

Pace, L., Salvan, A., Sartori, N. (2011). adjusting composite likelihood ratio statistics. *Statist. Sinica*, **21**, 129–148.

Padoan, S., Ribatet, M., Sisson, S. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.*, **105**, 263–277.

Renard, D., Molenberghs, G., Geys, H. (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Comput. Statist. Data Anal.*, **44**, 649–667.

Rotnitzky, A., Jewell, N. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. *Biometrika*, **77**, 485–497.

Satterthwaite, F. (1946). Approximate distribution of estimates of variance components. *Biometrika Bull.*, **69**, 110–114.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, **92**, 1–28.

Varin, C., Reid, N., Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica*, **21**, 5–42.

Wood, A. (1989). An f approximation to the distribution of a linear combination of chi-squared variables. *Commun. Statist. Ser. B*, **18**, 1439–1456.