Nonparametric saddlepoint test and pairwise likelihood inference

Nicola Lunardon and Elvezio Ronchetti

Abstract The asymptotic distribution of the pairwise log likelihood ratio is a linear combination of independent chi-square random variables with coefficients depending on the elements of the Godambe information. Adjusted versions of the pairwise log likelihood statistic have been proposed, but they still depend on the Godambe information matrix. Approximated p-values for testing a composite hypothesis may be obtained by refferring the observed value of such statistics to a critical value. The asymptotic theory can be used to approximate the desired quantile, but the approximation may be inaccurate. In this work we provide a nonparametric saddlepoint statistic derived from the pairwise score function. This statistic enjoys some desirable properties: it is asymptotically chi-square distributed and the approximation has a relative error of second order. Thereby our proposal claims a high level of accuracy with no need to estimate the Godambe information.

Key words: pairwise likelihood, saddlepoint approximation, Godambe information

1 Introduction

Let $Y \in \mathbb{R}^q$ be a random vector with probability distribution $F(y;\theta)$, $\theta \subseteq \mathbb{R}^p$, density function $f(y;\theta)$, and associated full log likelihood function $\ell(\theta) = \log f(y;\theta)$. Consider marginal density functions specified for pairs of observations of the form $f_{Y_r,Y_s}(y_r,y_s;\theta) = f(y_r,y_s;\theta)$, $r \neq s = 1,\dots,q$. Having observed a random sample of size n from Y, $y = (y_1,\dots,y_n)$, the pairwise log likelihood is [2]

Nicola Lunardon
Department of Statistical Sciences, Univeristy of Padova, via Cesare Battisti 241-243, e-mail: lunardon@stat.unipd.it

Elvezio Ronchetti
Research Center for Statistics and Department of Economics, University of Geneva, Blv. Pont d'Arve 40, e-mail: elvezio.ronchetti@unige.ch

$$p\ell(\theta) = p\ell(\theta; y) = \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \log f(y_{ir}, y_{is}; \theta).$$

The maximum pairwise likelihood estimator $\hat{\theta}_p$ is defined as the solution of the pairwise score equation, i.e.

$$ps(\theta) = ps(\theta; y) = \sum_{i=1}^{n} ps(\theta; y_i) = \sum_{i=1}^{n} \sum_{r=1}^{q-1} \sum_{s=r+1}^{q} \frac{\partial \log f(y_{ir}, y_{is}; \theta)}{\partial \theta} = 0.$$

Under regularity conditions [5], $\hat{\theta}_p$ is consistent and asymptotically normal with mean θ and covariance matrix $V(\theta)$ given by the inverse of the Godambe information $H(\theta)^{-1}J(\theta)H(\theta)^{-1}$, where $J(\theta) = \mathbb{E}(ps(\theta;Y)ps(\theta;Y)^\top)$ and $H(\theta) = \mathbb{E}(-\partial ps(\theta;Y)/\partial\theta^\top)$. Here the symbol $\mathbb{E}(\cdot)$ is used to take expectation with respect to $F(\cdot;\theta)$. The pairwise likelihood counterparts of the Wald and score statistics are, respectively, $pw_w(\theta) = (\hat{\theta}_p - \theta)^\top V(\theta)^{-1}(\hat{\theta}_p - \theta)$ and $pw_s(\theta) = ps(\theta)^\top J(\theta)^{-1} ps(\theta)$, and both converge to a chi-square random variable with p degrees of freedom. The pairwise log likelihood ratio statistic is

$$pw(\theta) = 2\left\{p\ell(\hat{\theta}_p) - p\ell(\theta)\right\} \xrightarrow{d} \sum_{j=1}^{p} \lambda_j(\theta) Z_j^2,$$

where $\lambda_j(\theta)$ are the eigenvalues of the matrix $J(\theta)H(\theta)^{-1}$ and Z_j are standard normal random variables, $j = 1, \ldots, p$. Several adjustments to $pw(\theta)$ have been proposed in order to recover, or at least to approximate, the usual chi-square distribution. The statistic proposed by [3] is $pw_1(\theta) = pw(\theta)/\kappa_1$, where $\kappa_1 = \sum_j \lambda_j(\theta)/p$. A χ^2_p approximation is used for the distribution of $pw_1(\theta)$. Recently [1] and [6] propose the statistics $pw_{cb}(\theta) = [pw(\theta)/(\hat{\theta}_p - \theta)^\top H(\hat{\theta}_p)(\hat{\theta}_p - \theta)] pw_w(\theta)$ and $pw_{pss}(\theta) = [pw(\theta)/ps(\theta)^\top H(\theta)^{-1} ps(\theta)] pw_s(\theta)$, respectively. The asymptotic distribution of $pw_{cb}(\theta)$ and $pw_{pss}(\theta)$ is chi-square with p degrees of freedom.

However, the chi-square approximation for the distribution of $pw_w(\theta)$, $pw_s(\theta)$, $pw_1(\theta)$, $pw_{cb}(\theta)$, and $pw_{pss}(\theta)$ may be inaccuarate, especially when n is small, since it takes into account the uncertainty of the elements of the Godambe information. Indeed, in the pairwise likelihood framework the joint distribution $F(\cdot;\theta)$ is not specified, therefore analytic expressions for $J(\theta)$ and $H(\theta)$ are not available, and the empirical counterparts

$$\hat{J}(\theta) = \frac{1}{n} \sum_{i=1}^{n} ps(\theta; y_i) ps(\theta; y_i)^\top \qquad \hat{H}(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial ps(\theta; y_i)}{\partial \theta^\top}$$

need to be used.

2 Nonparametric saddlepoint statistic

In order to overcome the problem of possible lack of accuracy of pairwise likelihood statistics presented in the previous section, we propose a nonparametric saddlepoint statistic derived following [4]. Our proposal claims some desirable properties: i) despite saddlepoint approximations were originally proposed in a fully parametric setting, the proposed test statistic is based on a nonparametric saddlepoint approximation, hence there is no need to specify $F(\cdot;\theta)$; ii) it has a standard asymptotic behaviour and the error in the approximation is, under suitable regularity conditions [4], relative and of order $O_p(n^{-1})$; iii) there is no need to estimate the elements of the Godambe information.

The proposed nonparametric saddlepoint statistic is

$$pw_{sp}(\theta) = -2n\log\left[\sum_{i=1}^{n} w_i(\theta)\exp\left\{\lambda(\hat{\theta}_p)^\top ps(\hat{\theta}_p;y_i)\right\}\right] \xrightarrow{d} \chi^2_p,$$

where $w_i(\theta) = \exp\left\{\beta(\theta)^\top ps(\theta;y_i)\right\}/\sum_i \exp\left\{\beta(\theta)^\top ps(\theta;y_i)\right\}$ is a suitable nonparametric estimate of $F(\cdot;\theta)$, and the Lagrange multiplier $\beta(\theta)$ is the root of the equation $\sum_i w_i(\theta)ps(\theta;y_i) = 0$. Finally, the saddlepoint $\lambda(\hat{\theta}_p)$ satisfies the equation $\sum_i \exp\left\{\lambda(\hat{\theta}_p)^\top ps(\hat{\theta}_p;y_i)\right\} ps(\hat{\theta}_p;y_i) = 0$.

3 Numerical example: multivariate normal distribution

Let Y be a normally distributed random vector, with vector of means $(\mu,\ldots,\mu)^\top \in \mathbb{R}^q$ and compound symmetric covariance matrix Σ having diagonal elements σ^2 and off-diagonal ones $\sigma^2\rho$, $\rho \in (-1/(q-1),1)$. The pairwise log likelihood for the parameter $\theta = (\mu,\sigma^2,\rho)$ is

$$p\ell(\theta) = -\frac{nq(q-1)}{2}\left[\log\sigma^2 + \frac{\log(1-\rho^2)}{2}\right] - \frac{1}{2\sigma^2(1-\rho^2)}\sum_i (y_{i\cdot}-\mu)^\top \Gamma(\theta)(y_{i\cdot}-\mu),$$

with $y_{i\cdot} = \sum_j y_{ij}$, $\Gamma_{jj}(\theta) = (q-1)$, $\Gamma_{jk}(\theta) = -\rho$, $j \neq k = 1,\ldots,q$. Note that for this model the the full log likelihood $\ell(\theta)$ is available and this allows us to set the log likelihood ratio $w(\theta)$ as a benchmark.

Simulations have been performed by generating 100.000 samples of size $n = 10$ from $Y \in \mathbb{R}^{30}$, after setting $\mu = 0$, $\sigma^2 = 1$ and ρ ranging from a moderate to a strong correlation.

For each sample the log likelihood ratio $w(\theta)$, the nonparametric saddlepoint statistic $pw_{sp}(\theta)$ as well as the statistics $pw_1(\theta)$, $pw_{cb}(\theta)$, and $pw_{pss}(\theta)$, introduced in Section 1, have been computed. Both the observed and the expected elements of the Godambe information (denoted in the following with the superscript e)

have been used to compute the pairwise likelihood-based statistics, the latter being available in this example [6].

Table 1 reports the empirical coverage probabilities of confidence regions for θ. As expected, the best results are obtained when the elements of the expected Godambe information have an analytical expression and, in particular, when $pw^e_{pss}(\theta)$ is used. However, it must be stressed that, in most of real applications, only the observed Godambe information is available. In this case, the pairwise likelihood-based statistics have coverage probabilities far from the nominal levels. Instead, the distribution of the nonparametric saddlepoint statistic $pw_{sp}(\theta)$ is approximated quite well by the χ^2_3 and the approximation is close to the one provided by the gold standard $w(\theta)$.

From simulation studies not reported here it is shown that all the pairwise likelihood statistics achieve the nominal levels either with increasing sample sizes or with resampling-based estimates of $J(\theta)$ and $H(\theta)$.

Table 1 Multivariate normal distribution: empirical coverage probabilities of confidence regions for $\theta = (\mu, \sigma^2, \rho)$.

	$\rho = 0.2$			$\rho = 0.5$			$\rho = 0.9$		
$1-\alpha$	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
$w(\theta)$	0.8802	0.9375	0.9858	0.8795	0.9367	0.9858	0.8800	0.9365	0.9859
$pw_{sp}(\theta)$	0.8644	0.9282	0.9820	0.8722	0.9300	0.9833	0.8650	0.9254	0.9809
$pw_1(\theta)$	0.7847	0.8442	0.9194	0.7505	0.8179	0.9058	0.7540	0.7823	0.8197
$pw_{cb}(\theta)$	0.5570	0.6250	0.7286	0.4201	0.4829	0.5906	0.1689	0.1991	0.2581
$pw_{pss}(\theta)$	0.7955	0.8950	0.9786	0.7980	0.8791	0.9516	0.9122	0.9462	0.9758
$pw^e_1(\theta)$	0.8133	0.8673	0.9336	0.8136	0.8692	0.9361	0.8407	0.8983	0.9613
$pw^e_{cb}(\theta)$	0.7885	0.8459	0.9126	0.7858	0.8463	0.9190	0.6296	0.6836	0.7610
$pw^e_{pss}(\theta)$	0.9080	0.9528	0.9883	0.8940	0.9477	0.9889	0.8699	0.9276	0.9802

References

1. Chandler, R., Bate, S.: Inference for clustered data using the independence loglikelihood. Biometrika **94**, 167–183 (2007)
2. Cox, D, Reid, N.: A note on pseudolikelihood constructed from marginal densities. Biometrika **91**, 729–737 (2004)
3. Geys, H. and Molenberghs, G., Ryan, L.: Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. J. Amer. Statist. Assoc. **94**, 734–745 (1999)
4. Ma, Y., Ronchetti, E.: Saddlepoint test in measurement error models. J. Amer. Statist. Assoc. **106**, 147–156 (2011)
5. Molenberghs, G., Verbeke, G.: Models for discrete longitudinal data. Springer, New York (2005)
6. Pace, L., Salvan, A., Sartori, N.: Adjusting composite likelihood ratio statistics. Statist. Sinica **21**, 129–148 (2011)