Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Self-concordant profile empirical likelihood ratio tests for the population correlation coefficient: a simulation study

Thorsten Dickhaus[1]

submitted: August 4, 2014

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
E-Mail: Thorsten.Dickhaus@wias-berlin.de

No. 1989
Berlin 2014

2010 *Mathematics Subject Classification.* 62F03, 62F25, 62G15, 62G99.

Key words and phrases. Confidence region, elliptical model, Lagrange multiplier, nonparametric tests, type I error rate, Wilks phenomenon.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We present results of a simulation study regarding the finite-sample type I error behavior of the self-concordant profile empirical likelihood ratio (ELR) test for the population correlation coefficient. Three different families of bivariate elliptical distributions are taken into account. Uniformly over all considered models and parameter configurations, the self-concordant profile ELR test does not keep the significance level for finite sample sizes, albeit the level exceedance monotonously decreases to zero as the sample size increases. We discuss some potential modifications to address this problem.

1 Introduction

Empirical likelihood ratio (ELR) tests for multivariate means were introduced by Owen (1990). Although ELR tests are nonparametric tests for statistical functionals, they share an important property with parametric likelihood ratio tests. Namely, under regularity assumptions, the asymptotic distribution of twice the negative logarithmic (empirical) likelihood ratio statistic is chi-squared under the null, with degrees of freedom determined by the dimensionality of the parameter (functional) of interest. This result is commonly referred to as the "Wilks phenomenon", see Wilks (1938).

ELR tests for the population correlation coefficient ρ of a bivariate distribution constitute a particularly challenging application example, because the evaluation of the ELR in this case requires a nested optimization. In the inner level, optimization has to be performed with respect to a five-dimensional Lagrange multiplier, and in the outer level four nuisance parameters have to be profiled out. We will provide more details in Section 2.

As reported for instance in Table 1 of Hall and La Scala (1990), the original ELR test for ρ often does not keep the significance level accurately for finite sample sizes. Recently, several novel strategies have been proposed to address this problem. In particular, Tsao (2013), Tsao and Wu (2013, 2014), Wu and Tsao (2014) and Owen (2013) proposed to extend the parameter space over which the ELR is maximized beyond the convex hull of the observations. While Tsao and Wu (2013) assess the accuracy of their “extended empirical likelihood“ method for multivariate means rather systematically in their Section 4, the work of Owen (2013) does not contain numerical results.

In this work, we assess the type I error accuracy of the self-concordant profile ELR test for ρ according to Owen (2013) by means of computer simulations. We consider three families of multivariate elliptical distributions (see Gupta et al. (2013) for a comprehensive overview of such distributions). The choice of elliptical models is motivated by the fact that ρ is a meaningful measure of dependency only in such models. Elliptical models play an important role in many applications from the life sciences (cf., e. g., Part II of Dickhaus (2014)) and in portfolio theory in finance (see Part IV of Gupta et al. (2013)). ELR methods are particularly attractive in such a context, because the type of elliptical distribution can often not be specified exactly. For example, there may be lacking information about the degrees of freedom of a multivariate Student's t distribution, see Section 3 for a definition.

The rest of the paper is structured as follows. In Section 2.1, we briefly summarize the statistical methodology of ELR tests for ρ. Section 2.2 contains some remarks on the computational strategies employed. Our main contribution is Section 3, where simulation results under three different elliptical models are presented. We conclude with a discussion in Section 4.

2 Statistical methodology and implementation

2.1 Statistical methodology

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote an independent and identically distributed (i.i.d.) sample from a bivariate distribution, where (X_1, Y_1) is in distribution equal to (X, Y). Assume that the second moment of the joint distribution of (X, Y) (denoted by $\mathcal{L}(X, Y)$) exists. Let $\mathbf{Z} = (X, Y, X^2, Y^2, XY)^\top$ denote a random vector with values in $\mathbb{R}^5$. For the expectation of $\mathbf{Z}$, it holds that

$$\mathbb{E}[\mathbf{Z}] = (\mu_X, \mu_Y, \mu_X^2 + \sigma_X^2, \mu_Y^2 + \sigma_Y^2, \rho\sigma_X\sigma_Y + \mu_X\mu_Y)^\top. \tag{1}$$

In (1) and throughout the remainder, μ_W (σ_W^2) denotes the mean (variance) of the random variable W, and $\rho = \rho(X, Y)$ is Pearson's product-moment correlation coefficient of X and Y. We denote by $\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)^\top \in \Theta \subset \mathbb{R}^5$ the five-dimensional vector of the first two population moments of interest and define $h : \Theta \to \mathbb{R}^5$ as the function which maps θ onto $\mathbb{E}[\mathbf{Z}]$. Obviously, h possesses (partial) derivatives of any order.

The ELR for a given parameter value θ^* is given by

$$\mathcal{R}(\theta^*) = \max\left\{\prod_{i=1}^n np_i \,|\, 0 \leq p_i \leq 1, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\mathbf{Z}_i = h(\theta^*)\right\}, \tag{2}$$

where $\mathbf{Z}_i$ is calculated from (X_i, Y_i) as $\mathbf{Z}$ from (X, Y), for $1 \leq i \leq n$. The (asymptotic) ELR test for the null hypothesis $H_0' : \theta = \theta^*$ rejects H_0' at significance level $\alpha \in (0, 1)$, if $\ell(\theta^*) = -2\log(\mathcal{R}(\theta^*))$ exceeds $\chi^2_{5;1-\alpha}$, where $\chi^2_{\nu;1-\alpha}$ denotes the $(1-\alpha)$-quantile of the central chi-square distribution with ν degrees of freedom.

In this work, we are concerned with the more general hypothesis $H_0 : \{\rho = \rho^*\}$ for a given value $\rho^* \in [-1, 1]$. For testing H_0, we profile out the nuisance parameters μ_X, μ_Y, σ_X^2, and σ_Y^2. More specifically, let $\Theta(\rho^*) = \{\theta \in \Theta : \rho = \rho^*\}$. The test for H_0 can then be described by the following algorithm.

Algorithm 2.1.

1. *Maximize $\mathcal{R}$ over $\theta^* \in \Theta(\rho^*)$. Denote the maximizer by $\theta(\rho^*)$.*

2. *Reject H_0 at significance level α, if $\ell(\rho^*) = -2\log(\mathcal{R}(\theta(\rho^*))) > \chi^2_{1;1-\alpha}$.*

Remark 2.1. *The testing method based on $\mathbf{Z}$ has been outlined in Section 3.4 of Owen (2001). It appears more convenient than the method originally proposed in Section 6.2 of Owen (1990), because it avoids iterated re-centering of the observations when step 1 of Algorithm 2.1 is performed.*

2.2 Implementation

Notice that Algorithm 2.1 involves a nested double optimization. Namely, the inner optimization is given by the maximization in (2) for given θ^*, and the outer optimization is given by the maximization over the nuisance parameters μ_X, μ_Y, σ_X^2, σ_Y^2 as described in the first step of Algorithm 2.1.

For the outer optimization, a general-purpose (constrained) optimizer can be employed. For the simulations in Section 3, we utilized the `optim` function in `R` with method `"L-BFGS-B"`. This routine implements the box-constrained optimization algorithm by Byrd et al. (1995). Constraints are required in our context, because σ_X^2 and σ_Y^2 are necessarily non-negative.

More crucial is the inner optimization problem (2). To this end, Owen (2013) introduced an algorithm based on self-concordance. In a nutshell, the negative empirical log-likelihood ratio is approximated by a quartic polynomial, leading to a convex constrained optimization problem which can be solved by the method of Lagrange multipliers. Owen (2013) also contributed the `R` program `scel.R` on which our simulations in Section 3 rely.

3 Simulation results

In this section, we present simulation results under three different elliptical models for $\mathcal{L}(X, Y)$. In general, the probability density function (pdf) of an elliptically contoured distribution on $\mathbb{R}^d$ is of the form

$$f(\mathbf{t}) = C_d |\det \Sigma|^{-1/2} g(\mathbf{t}^\top \Sigma^{-1} \mathbf{t}), \quad \mathbf{t} \in \mathbb{R}^d,$$

where C_d is a normalizing constant which depends on d. The positive definite matrix $\Sigma \in \mathbb{R}^{d\times d}$ is called the dispersion matrix. It captures the (elliptical) dependencies among the components of a random vector with pdf f. The scalar function g is called the density generator of the elliptical distribution. Three well-known families of elliptical distributions are

(a) normal distributions on $\mathbb{R}^d$ with $g(u) = \exp(-u/2)$,

(b) Student's t distributions on $\mathbb{R}^d$ with $g(u) = (1 + u/\nu)^{-(d+\nu)/2}$, $\nu \in \mathbb{N}$,

(c) double-exponential (Laplace) distributions on $\mathbb{R}^d$ with $g(u) = \exp(-|u|)$.

We will consider these three distributional families in the remainder of this section.

3.1 Bivariate normal distribution

We assume that $\mathcal{L}(X, Y) = \mathcal{N}_2(0, \Sigma)$, where $\sigma_X^2 = \sigma_Y^2 = 1$ without loss of generality. The off-diagonal element of $\Sigma \in [-1, 1]^{2\times 2}$ is the parameter ρ of interest. Pseudo-random samples were generated by utilizing the routine `rmvnorm` from the `R` package `mvtnorm`, cf. Genz and Bretz (2009). Table 1 summarizes our simulation results under this Gaussian model for $\mathcal{L}(X, Y)$.

Table 1: Relative rejection frequencies of the self-concordant profile empirical likelihood ratio test for the population correlation coefficient in case of bivariate Gaussian data. The nominal significance level was set to $\alpha = 5\%$ in all simulations. Results are based on $10{,}000$ Monte Carlo repetitions for each parameter configuration (ρ: true underlying correlation coefficient, n: sample size).

ρ	n	Relative rejection frequency	ρ	n	Relative rejection frequency
-0.9	10	0.1669	0.25	10	0.1612
-0.9	20	0.1030	0.25	20	0.1085
-0.9	50	0.0681	0.25	50	0.0716
-0.9	100	0.0593	0.25	100	0.0558
-0.75	10	0.1668	0.5	10	0.1705
-0.75	20	0.1069	0.5	20	0.1066
-0.75	50	0.0758	0.5	50	0.0743
-0.75	100	0.0588	0.5	100	0.0586
-0.5	10	0.1645	0.75	10	0.1649
-0.5	20	0.1089	0.75	20	0.1077
-0.5	50	0.0688	0.75	50	0.0737
-0.5	100	0.0624	0.75	100	0.0605
-0.25	10	0.1655	0.9	10	0.1662
-0.25	20	0.1102	0.9	20	0.1015
-0.25	50	0.0737	0.9	50	0.0716
-0.25	100	0.0593	0.9	100	0.0625
0	10	0.1669			
0	20	0.1106			
0	50	0.0697			
0	100	0.0623			

3.2 Bivariate Student's t distribution

In this section, we assume that $\mathcal{L}(X, Y)$ is a centered bivariate Student's t distribution with $\nu > 2$ degrees of freedom and dispersion matrix $\Sigma \in \mathbb{R}^{2\times 2}$, denoted as $t_2(\nu, \Sigma)$. In analogy to Section 3.1, we may without loss of generality assume that $\Sigma_{11} = \Sigma_{22} = 1$. In this case, Σ is the correlation matrix of (X, Y) (see Section 1.7 of Kotz and Nadarajah (2004)); hence, its off-diagonal element equals again the parameter ρ. Pseudo-random samples were generated by utilizing the routine `rmvt` from the `R` package `mvtnorm`. Since it is well-known that $t_2(\nu, \Sigma)$ converges weakly to $\mathcal{N}_2(0, \Sigma)$ with increasing degrees of freedom ν, we restrict our attention to small values $\nu \in \{5, 10\}$ in Table 2.

3.3 Bivariate double-exponential distribution

Here, we consider centered bivariate double-exponential (Laplace) distributions for $\mathcal{L}(X, Y)$. To this end, it is convenient to notice that in this case (X, Y) possesses the stochastic representation

$$\begin{pmatrix} X \\ Y \end{pmatrix} \stackrel{d}{=} \sqrt{E}\mathbf{G}, \tag{3}$$

where E follows a univariate exponential distribution with intensity parameter $\lambda > 0$, and $\mathbf{G}$ denotes a centered bivariate Gaussian random vector with covariance matrix Σ; see, e. g., equation (6) in Eltoft (2006). Letting $\Sigma_{11} = \Sigma_{22} = 1$, we again obtain that $\rho(X, Y) = \Sigma_{12}$. Since the latter property holds regardless of the value of λ, we can restrict our attention to $\lambda = 1$. Based on (3), pseudo-random samples were generated by generating independent realizations of E with the `R` function `rexp` and independent realizations of $\mathbf{G}$ with the `rmvnorm` function as described in Section 3.1. Simulation results for this model are presented in Table 3.

4 Concluding remarks

Summarizing our findings we observe that the relative rejection frequencies obtained under joint normality of (X, Y) are very close to those reported in Table 1 of Hall and La Scala (1990). The ELR test does not keep the significance level accurately for finite sample sizes, but the level exceedance monotonously decreases to zero as the sample size increases. Qualitatively, this behavior of the ELR test is also reflected in our Tables 2 and 3 which correspond to two other elliptical models for $\mathcal{L}(X, Y)$.

Thus, if Gaussianity of (X, Y) can be assumed, it seems recommendable to carry out a parametric test as explained, e. g., in Section 4.2 of Anderson (1984). In the nonparametric setting, future research will consider Bartlett-corrected critical values (see DiCiccio et al. (1991)) in order to overcome the reported anti-conservativity of the considered ELR tests. On a more fundamental level, an interesting and challenging research direction would be to analyze the finite-sample properties of ELR-based inference by providing concentration inequalities in the spirit of Spokoiny (2012).

Finally, let us mention that our restriction to point hypotheses of the form $H_{\rho^*} : \{\rho = \rho^*\}$ is not a severe limitation. Namely, by duality of tests and confidence regions, a composite null hypothesis $H_{\text{comp.}}$ (associated with a subset of $[-1, 1] \ni \rho$) can be tested on the basis of a family of point hypothesis tests. Following Aitchison (1964), a $(1 - \alpha)$-confidence region $\mathcal{C}_\alpha$ is constituted by the set of all parameter values ρ^* for which H_{ρ^*} is not rejected. Then, $H_{\text{comp.}}$ can be rejected at level α if $H_{\text{comp.}} \cap \mathcal{C}_\alpha = \emptyset$. It is clear that the type I error accuracy of this test for $H_{\text{comp.}}$ depends on that of the tests for the H_{ρ^*}.

References

Aitchison, J., 1964. Confidence-region tests. J. R. Stat. Soc., Ser. B 26, 462–476.

Anderson, T., 1984. An introduction to multivariate statistical analysis. 2nd ed. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons, Inc.

Byrd, R. H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16 (5), 1190–1208.

DiCiccio, T., Hall, P., Romano, J., 1991. Empirical likelihood is Bartlett-correctable. Ann. Stat. 19 (2), 1053–1061.

Dickhaus, T., 2014. Simultaneous Statistical Inference with Applications in the Life Sciences. Springer-Verlag Berlin Heidelberg.

Eltoft, T., 2006. On the Multivariate Laplace Distribution. IEEE Signal Processing Letters 13 (5), 300–303.

Genz, A., Bretz, F., 2009. Computation of multivariate normal and t probabilities. Lecture Notes in Statistics 195. Berlin: Springer.

Gupta, A. K., Varga, T., Bodnar, T., 2013. Elliptically contoured models in statistics and portfolio theory, 2nd Edition. New York, NY: Springer.

Hall, P., La Scala, B., 1990. Methodology and algorithms of empirical likelihood. Int. Stat. Rev. 58 (2), 109–127.

Kotz, S., Nadarajah, S., 2004. Multivariate t distributions and their applications. Cambridge University Press.

Owen, A., 1990. Empirical likelihood ratio confidence regions. Ann. Stat. 18 (1), 90–120.

Owen, A. B., 2001. Empirical likelihood. Boca Raton, FL: Chapman & Hall/ CRC.

Owen, A. B., 2013. Self-concordance for empirical likelihood. Can. J. Stat. 41 (3), 387–397.

Spokoiny, V., 2012. Parametric Estimation. Finite Sample Theory. Ann. Stat. 40 (6), 2877–2909.

Tsao, M., 2013. Extending the empirical likelihood by domain expansion. Can. J. Stat. 41 (2), 257–274.

Tsao, M., Wu, F., 2013. Empirical likelihood on the full parameter space. Ann. Stat. 41 (4), 2176–2196.

Tsao, M., Wu, F., 2014. Extended empirical likelihood for estimating equations. Biometrika Advance Access published June 10, 2014.

Wilks, S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9, 60–62.

Wu, F., Tsao, M., 2014. Two-sample extended empirical likelihood. Stat. Probab. Lett. 84, 81–87.

Table 2: Relative rejection frequencies of the self-concordant profile empirical likelihood ratio test for the population correlation coefficient in case of bivariate Student's t data. The nominal significance level was set to $\alpha = 5\%$ in all simulations. Results are based on $10{,}000$ Monte Carlo repetitions for each parameter configuration (ρ: true underlying correlation coefficient, ν: degrees of freedom, n: sample size).

ρ	ν	n	Relative rejection freq.	ρ	ν	n	Relative rejection freq.
-0.9	5	10	0.2069	0	10	10	0.1835
-0.9	5	20	0.1481	0	10	20	0.1327
-0.9	5	50	0.1091	0	10	50	0.0844
-0.9	5	100	0.0974	0	10	100	0.0738
-0.9	10	10	0.1762	0.25	5	10	0.2163
-0.9	10	20	0.1157	0.25	5	20	0.1561
-0.9	10	50	0.0869	0.25	5	50	0.1112
-0.9	10	100	0.0763	0.25	5	100	0.0888
-0.75	5	10	0.2088	0.25	10	10	0.1803
-0.75	5	20	0.1459	0.25	10	20	0.1215
-0.75	5	50	0.1080	0.25	10	50	0.0867
-0.75	5	100	0.0919	0.25	10	100	0.0699
-0.75	10	10	0.1741	0.5	5	10	0.2131
-0.75	10	20	0.1271	0.5	5	20	0.1492
-0.75	10	50	0.0838	0.5	5	50	0.1092
-0.75	10	100	0.0722	0.5	5	100	0.0916
-0.5	5	10	0.2083	0.5	10	10	0.1955
-0.5	5	20	0.1482	0.5	10	20	0.1248
-0.5	5	50	0.1068	0.5	10	50	0.0851
-0.5	5	100	0.0871	0.5	10	100	0.0692
-0.5	10	10	0.1856	0.75	5	10	0.2106
-0.5	10	20	0.1307	0.75	5	20	0.1425
-0.5	10	50	0.0853	0.75	5	50	0.1110
-0.5	10	100	0.0690	0.75	5	100	0.0884
-0.25	5	10	0.2094	0.75	10	10	0.1844
-0.25	5	20	0.1530	0.75	10	20	0.1224
-0.25	5	50	0.1053	0.75	10	50	0.0865
-0.25	5	100	0.0904	0.75	10	100	0.0639
-0.25	10	10	0.1888	0.9	5	10	0.2022
-0.25	10	20	0.1282	0.9	5	20	0.1478
-0.25	10	50	0.0835	0.9	5	50	0.1039
-0.25	10	100	0.0707	0.9	5	100	0.0942
0	5	10	0.2089	0.9	10	10	0.1830
0	5	20	0.1468	0.9	10	20	0.1220
0	5	50	0.1065	0.9	10	50	0.0825
0	5	100	0.0887	0.9	10	100	0.0679

Table 3: Relative rejection frequencies of the self-concordant profile empirical likelihood ratio test for the population correlation coefficient in case of bivariate double-exponential data. The nominal significance level was set to $\alpha = 5\%$ in all simulations. Results are based on $10{,}000$ Monte Carlo repetitions for each parameter configuration (ρ: true underlying correlation coefficient, n: sample size).

ρ	n	Relative rejection frequency	ρ	n	Relative rejection frequency
-0.9	10	0.2262	0.25	10	0.2363
-0.9	20	0.1585	0.25	20	0.1570
-0.9	50	0.1017	0.25	50	0.1045
-0.9	100	0.0823	0.25	100	0.0833
-0.75	10	0.2273	0.5	10	0.2281
-0.75	20	0.1575	0.5	20	0.1565
-0.75	50	0.1004	0.5	50	0.1067
-0.75	100	0.0840	0.5	100	0.0856
-0.5	10	0.2417	0.75	10	0.2293
-0.5	20	0.1603	0.75	20	0.1578
-0.5	50	0.1059	0.75	50	0.1005
-0.5	100	0.0812	0.75	100	0.0824
-0.25	10	0.2362	0.9	10	0.2279
-0.25	20	0.1632	0.9	20	0.1494
-0.25	50	0.1021	0.9	50	0.1023
-0.25	100	0.0846	0.9	100	0.0822
0	10	0.2351			
0	20	0.1644			
0	50	0.1022			
0	100	0.0848			