de Carvalho, M., and Marques, F. J. (2012), "Jackknife Euclidean Likelihood- Based Inference for Spearman's Rho," North American Actuarial Journal, 16, 487–492.

Jackknife Euclidean Likelihood-Based Inference for Spearman's Rho

M. de Carvalho*† and F. J. Marques†

Abstract

We discuss jackknife Euclidean likelihood-based inference methods, with a special focus on the construction of confidence intervals for Spearman's rho. We show that a Wilks' theorem holds for jackknife Euclidean likelihood, and based on it we construct confidence intervals for Spearman's rho. In a simulation study we examine the performance of our method, and a fire insurance claims database is used for its illustration.

1 INTRODUCTION

This short paper is a follow-up on a recent paper by Wang and Peng (2011), who introduced an innovative approach for obtaining confidence intervals for Spearman's rho—which is of wide interest for applied work in actuarial sciences and risk management (Embrechts et al., 2002; McNeil et al., 2005). In this paper we focus on describing a related strategy—which we call *jackknife Euclidean likelihood*—for obtaining confidence intervals for Spearman's rho, but which avoids the need to compute Lagrange multipliers. This allows our related approach to be computationally and theoretically appealing, leading to fast computations and a simple large sample theory—which generalizes to U-statistics (Kowalski and Tu, 2008). Our starting point is a simple one, and it is based on using Euclidean likelihood (Owen, 2001, p. 65) as an objective function, instead of empirical likelihood. In particular, we show that a Wilks' theorem still holds for jackknife Euclidean likelihood, and based on it we construct confidence intervals for Spearman's rho.

Since both empirical likelihood and Euclidean likelihood are members of the Cressie–Read family, our results can be thought as extensions of the arguments in Baggerly (1998)—but for the context of jackknife empirical likelihood—and as corollaries to the results of Jing et al. (2009) and Wang and Peng (2011).

2 JACKKNIFE EUCLIDEAN LIKELIHOOD

2.1 Confidence Intervals for U-Statistics

We start by describing jackknife Euclidean likelihood to one-sample U-statistics. Let $X_1, \dots, X_n$ be a random sample from a distribution function F. Formally, a one-sample U-statistic of degree

*Miguel de Carvalho is Assistant Professor at the Department of Statistics, Pontificia Universidad Católica de Chile (*mdecarvalho@uc.cl*) and Research Fellow at the Center for Mathematics and Applications, Universidade Nova de Lisboa, Portugal.

†Filipe Marques is Assistant Professor at the Faculty of Sciences and Technology, Universidade Nova de Lisboa (*fjm@fct.unl.pt*) and Research Fellow at the Center for Mathematics and Applications, Universidade Nova de Lisboa, Portugal.

$m \geq 2$ with a symmetric kernel h is defined as

$$U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \ldots, X_{i_m}),$$

where the parameter of interest is $\theta = Eh(X_1, \ldots, X_m)$. The corresponding jackknife pseudo-values are

$$Z_i = nU_n - (n-1)U_{n-1}^{(-i)}, \quad i = 1, \ldots, n,$$

where $U_{n-1}^{(-i)}$ denotes the estimate produced by the U-statistic computed using the sample of $n-1$ observations, constructed by deleting the ith observation. Note that by averaging the pseudo-values we get back the U-statistic, i.e.

$$U_n = \frac{1}{n} \sum_{i=1}^{n} Z_i = \overline{Z}.$$

The Euclidean loglikelihood function (Owen, 2001, §3.15) is defined as

$$\ell(p) = -1/2 \sum_{i=1}^{n} (np_i - 1)^2, \quad p = (p_1, \ldots, p_n) \in \mathbb{R}^n,$$

and it is often considered as an alternative to the empirical loglikelihood function

$$\ell^*(p) = -\sum_{i=1} \log(np_i), \quad p = (p_1, \ldots, p_n) \in \mathbb{R}_+^n.$$

Hence, we define the jackknife Euclidean loglikelihood function as

$$\mathscr{L}(\theta) = \sup \Big\{ -1/2 \sum_{i=1}^{n} (np_i - 1)^2 : p = (p_1, \ldots, p_n), \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Z_i = \theta \Big\}. \tag{1}$$

which should be understood as an analogous of the jackknife empirical loglikelihood in Jing et al. (2009). Using Lagrange multiplier-based procedures, it can be shown (Owen, 2001, p. 65) that

$$p_i = \frac{1}{n} \{1 - (Z_i - \theta)(\overline{Z} - \theta)/S(\theta)\}, \quad i = 1, \ldots, n, \tag{2}$$

where

$$S(\theta) = \frac{1}{n} \sum_{i=1}^{n} (Z_i - \theta)^2.$$

Simple computations detailed in the latter reference imply that

$$-2\mathscr{L}(\theta) = \frac{n(\overline{Z} - \theta)^2}{S(\theta)} = \left\{ \sqrt{n} \left(\frac{\overline{Z} - \theta}{\sqrt{S(\theta)}} \right) \right\}^2.$$

Using Lemma A.3 in Jing et al. (2009) we have that

$$-2\mathscr{L}(\theta) = \left\{ \sqrt{n} \left(\frac{\overline{Z} - \theta}{2\sigma_g + o_p(1)} \right) \right\}^2, \quad \sigma_g = \mathrm{var}\{g(X_1)\}, \tag{3}$$

where $g(x) = Eh(x, X_2, \ldots, X_m) - \theta$. Lemma A.2 in Jing et al. (2009) states that $\sqrt{n}(\overline{Z} - \theta)/(2\sigma_g)$ converges in distribution to a standard normal distribution, which together with Slutsky theorem implies that $-2\mathscr{L}(\theta)$ converges in distribution to a chi-square distribution with one degree of freedom. The same principle extends for two-sample U-statistics, and using Lemmas A.5 and A.7 in Jing et al. (2009), we obtain again convergence in distribution to a chi-square distribution with one degree of freedom. These results hold under the same technical assumptions required by Theorems 1–2 in Jing et al. (2009).

A jackknife Euclidean likelihood confidence interval for the parameter θ with level α, can thus be obtained through the set-valued function $I : (0,1) \rightrightarrows \mathbb{R}$, which is defined as

$$I_\alpha = \left\{\theta : -2\mathscr{L}(\theta) \leq \chi^2_{1,\alpha}\right\} = \left\{\theta : \frac{n(U_n - \theta)^2}{S(\theta)} \leq \chi^2_{1,\alpha}\right\},$$

where $\chi^2_{1,\alpha}$ denotes the α quantile of a chi-square distribution with one degree of freedom.

2.2 Confidence Intervals for Spearman's Rho

The principles above can be extended to obtain confidence intervals for Spearman's rho. To simplify the comparison with the results obtained by Wang and Peng (2011), we redefine some of their notations. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors with distribution function H and continuous marginals $F(x) = H(x, \infty)$ and $G(y) = H(\infty, y)$. We define the Spearman's rho as

$$\rho^s = 12\mathbb{E}[(F(X_1) - 1/2)(G(Y_1) - 1/2)],$$

and to estimate it from data, we use its corresponding empirical estimator

$$\hat{\rho}^s_n = \frac{12}{n}\sum_{i=1}^{n}\{F_n(X_i) - 1/2\}\{G_n(Y_i) - 1/2\},$$

where

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i \leq x), \quad G_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(Y_i \leq x).$$

Following Wang and Peng (2011), we assume the following condition on the partial derivatives of the copula function $C(x,y) = \mathbb{P}(F(X_1) \leq x, G(Y_1) \leq y)$:

$$\begin{cases} \frac{\partial}{\partial x}C(x,y) & \text{exists and is continuous on the set } \{(x,y) : 0 < x < 1, 0 \leq y \leq 1\}, \\ \frac{\partial}{\partial y}C(x,y) & \text{exists and is continuous on the set } \{(x,y) : 0 \leq x \leq 1, 0 < y < 1\}. \end{cases} \tag{4}$$

Beyond avoiding the need to compute the asymptotic variance, our related approach also avoids the need to compute a Lagrange multiplier—λ in the notation of Wang and Peng (2011). The starting point of our approach is exactly the same, and we also define

$$F_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\mathbf{1}(X_j \leq x), \quad G_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\mathbf{1}(Y_j \leq x),$$

$$\hat{\rho}^s_{n,i} = \frac{12}{n-1}\sum_{j=1,j\neq i}^{n}\{F_{n,i}(X_j) - 1/2\}\{G_{n,i}(Y_j) - 1/2\}, \quad Z_i = n\hat{\rho}^s_n - (n-1)\hat{\rho}^s_{n,i},$$

for $i = 1, \ldots, n$. Our approach differs from the one proposed by Wang and Peng (2011), since instead of considering a jackknife empirical likelihood function as in Jing et al. (2009), we consider a jackknife Euclidean loglikelihood function as in (1), for $\theta = \rho^s$. Using Lagrange multiplier-based

Table 1
Coverage Probabilities for I_α at Levels $\alpha = 0.9, 0.95, 0.99$
Reported for $n = 100, 300$ and $\rho = 0, \pm 0.2, \pm 0.8$

(n, ρ)	$I_{0.9}$	$I_{0.95}$	$I_{0.99}$
(100,0)	0.9028	0.9502	0.9892
(100,0.2)	0.8969	0.9523	0.9895
(100,-0.2)	0.9021	0.9480	0.9876
(100,0.8)	0.9035	0.9462	0.9825
(100,-0.8)	0.9025	0.9510	0.9845
(300,0)	0.9014	0.9522	0.9906
(300,0.2)	0.9013	0.9514	0.9902
(300,-0.2)	0.8962	0.9474	0.9875
(300,0.8)	0.8986	0.9522	0.9877
(300,-0.8)	0.9041	0.9478	0.9871

procedures we obtain the expression for p_i as in (2), with $\theta = \rho^s$, which does not depend on the Lagrange multiplier λ. Hence, simple calculations can be used to show that (Owen, 2001, p. 65)

$$-2\mathscr{L}(\rho^s) = \frac{n(\overline{Z} - \rho^s)^2}{S(\rho^s)} = \frac{\{\sum_{i=1}^n (Z_i - \rho^s)\}^2}{\sum_{i=1}^n (Z_i - \rho^s)^2}.$$

Lemmas 1 and 2 in Wang and Peng (2011) allow us to establish the following Wilks' theorem.

Theorem 1
Assume condition (4) holds. Then $-2\mathscr{L}(\rho^s)$ converges in distribution to a chi-square distribution with one degree of freedom as $n \to \infty$.

Using this theorem we can construct the following alternative confidence intervals for ρ^s with level α

$$I_\alpha = \left\{\theta : \frac{n(\overline{Z} - \theta)^2}{S(\theta)} \leq \chi^2_{1,\alpha}\right\} = \left\{\theta : \frac{\{\sum_{i=1}^n (Z_i - \theta)\}^2}{\sum_{i=1}^n (Z_i - \theta)^2} \leq \chi^2_{1,\alpha}\right\}.$$

3 APPLICATIONS

3.1 Simulated Data

We replicate the simulation exercise by Wang and Peng (2011), so that direct comparisons can be made; for completeness we report in the appendix their numerical results. Thus, we consider 10,000 random samples of size $n = 100, 300$ from a bivariate normal distribution with correlation ρ and with standard normal marginal distributions. As it can be observed from Tables 1–2 the performance of the jackknife Euclidean likelihood method is comparable to the one proposed by Wang and Peng, in terms of coverage probability and average interval length.

3.2 Revisiting the Danish Fire Insurance Claims Database

We now revisit the Danish fire insurance claims database, which includes 2,167 industrial fire losses gathered from the Copenhagen Reinsurance Company over the period 1980–1990. For comparison purposes we focus on the same pair of variables as the Wang and Peng (2011), viz.: loss to buildings and loss to contents. We obtain the point estimate $\hat{\rho}^s_n = 0.1411$ with confidence intervals $I_{0.95} = [0.0870, 0.1952]$ and $I_{0.90} = [0.0957, 0.1865]$, which are similar to the ones obtained by Wang

Table 2

Average Interval Lengths for I_α at Levels $\alpha = 0.9, 0.95, 0.99$

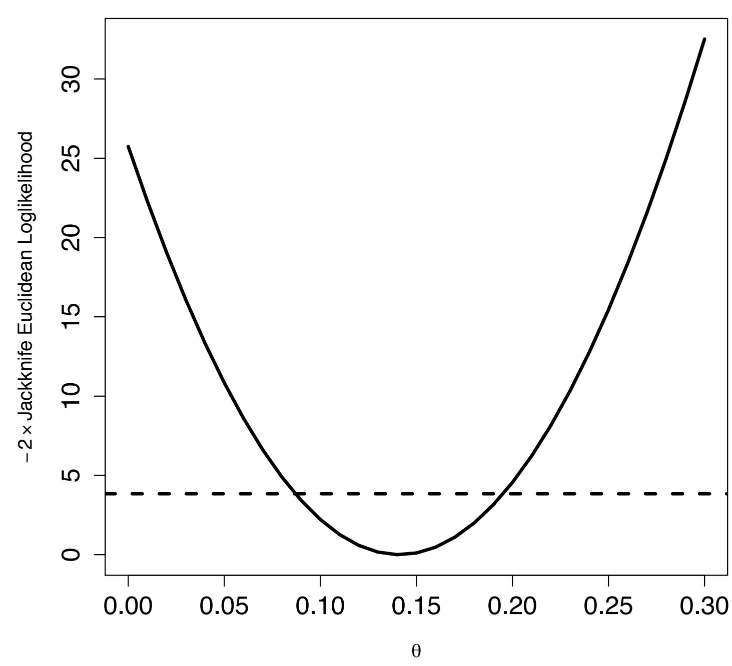
Reported for $n = 100, 300$ and $\rho = 0, \pm 0.2, \pm 0.8$

(n, ρ)	$I_{0.9}$	$I_{0.95}$	$I_{0.99}$
(100,0)	0.338	0.405	0.541
(100,0.2)	0.328	0.393	0.524
(100,-0.2)	0.328	0.392	0.524
(100,0.8)	0.147	0.175	0.235
(100,-0.8)	0.147	0.176	0.234
(300,0)	0.192	0.229	0.302
(300,0.2)	0.186	0.222	0.293
(300,-0.2)	0.186	0.222	0.293
(300,0.8)	0.082	0.098	0.129
(300,-0.8)	0.082	0.098	0.129

and Peng (2011) ($I^*_{0.95} = [0.0882, 0.1952]$, $I^*_{0.90} = [0.0962, 0.1862]$). In Figure 1 we show how to graphically construct a 95% confidence interval using our approach, with the bounds of the confidence interval simply being given by the roots of the function $-2\mathscr{L}(\theta) - \chi^2_{1,0.95} \approx -2\mathscr{L}(\theta) - 3.84$. Our confidence intervals suggest that Spearman's rho is positive, and hence reinforce the view that loss to contents is positively correlated with loss to buildings.

Figure 1

Graphical Construction of Jackknife Euclidean Likelihood Intervals



Note: The dashed line represents the critical value given by the 95% quantile of chi-square distribution with one degree of freedom

ACKNOWLEDGEMENTS

We thank the editor and an anonymous referee for helpful comments on an earlier version of this paper. We thank Professors Wang and Peng for helpful discussions and for sending us the Danish fire insurance claims database. This work was partially supported by the *Fundação para a Ciência e a Tecnologia*, through PEst-OE/MAT/UI0297/2011 (CMA).

REFERENCES

Baggerly, K. A. 1998. Empirical Likelihood as a Goodness-of-fit Measure. *Biometrika* 85: 535–547.

Embrechts, P., A. J. McNeil, and D. Straumann. 2002. Correlation and Dependency in Risk Management: Properties and Pitfalls. In *Risk Management: Value at Risk and Beyond*, edited by M. Dempster, pp. 176–223. Cambridge: Cambridge University Press.

Jing, B-Y., J. Yuan, and W. Zhou. 2009. Jackknife Empirical Likelihood. *Journal of the American Statistical Association* 104: 1224–1232.

Kowalski, J., and X. M. Tu. 2008. *Modern Applied U-Statistics*. New York: John Wiley & Sons.

McNeil, A. J., R. Frey, and P. Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton: Princeton University Press.

Owen, A. 2001. *Empirical Likelihood*. Boca Raton: Chapman and Hall.

Wang, R., and L. Peng. 2011. Jackknife Empirical Likelihood Intervals for Spearman's Rho. *North American Actuarial Journal* 15: 475–486.

APPENDIX

COMPARATIVE NUMERICAL ANALYSIS—JACKKNIFE EMPIRICAL LIKELIHOOD

For completeness, we report in this appendix the numerical results of Wang and Peng (2011), which are based on 10,000 random samples of size $n = 100, 300$ from a bivariate normal distribution with correlation ρ and with standard normal marginal distributions.

Table 1
Coverage Probabilities for the Wang–Peng approach (I^*_α) at Levels $\alpha = 0.9, 0.95, 0.99$
Reported for $n = 100, 300$ and $\rho = 0, \pm 0.2, \pm 0.8$

(n, ρ)	$I^*_{0.9}$	$I^*_{0.95}$	$I^*_{0.99}$
(100,0)	0.9024	0.9524	0.9898
(100,0.2)	0.9016	0.9524	0.9900
(100,-0.2)	0.9003	0.9513	0.9896
(100,0.8)	0.9013	0.9473	0.9850
(100,-0.8)	0.8926	0.9390	0.9818
(300,0)	0.9055	0.9530	0.9915
(300,0.2)	0.9035	0.9513	0.9906
(300,-0.2)	0.9073	0.9529	0.9908
(300,0.8)	0.9037	0.9529	0.9900
(300,-0.8)	0.9008	0.9505	0.9899

Table 2
Average Interval Lengths for Wang–Peng approach (I^*_α) at Levels $\alpha = 0.9, 0.95, 0.99$
Reported for $n = 100, 300$ and $\rho = 0, \pm 0.2, \pm 0.8$

(n, ρ)	$I^*_{0.9}$	$I^*_{0.95}$	$I^*_{0.99}$
(100,0)	0.337	0.403	0.529
(100,0.2)	0.327	0.391	0.515
(100,-0.2)	0.327	0.390	0.515
(100,0.8)	0.148	0.177	0.235
(100,-0.8)	0.147	0.176	0.234
(300,0)	0.192	0.229	0.302
(300,0.2)	0.186	0.222	0.293
(300,-0.2)	0.186	0.222	0.293
(300,0.8)	0.083	0.099	0.130
(300,-0.8)	0.083	0.099	0.130