A Simple Density-Based Empirical Likelihood Ratio Test for Independence

Albert Vexler, Wan-Min Tsai, Alan D. Hutson

Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, U.S.A

We develop a novel nonparametric likelihood ratio test for independence between two random variables using a technique that is free of the common constraints of defining a given set of specific dependence structures. Our methodology revolves around an exact density-based empirical likelihood ratio test statistic that approximates in a distribution-free fashion the corresponding most powerful parametric likelihood ratio test. We demonstrate that the proposed test is very powerful in detecting general structures of dependence between two random variables, including non-linear and/or random-effect dependence structures. An extensive Monte Carlo study confirms that the proposed test is superior to the classical nonparametric procedures across a variety of settings. The real-world applicability of the proposed test is illustrated using data from a study of biomarkers associated with myocardial infarction.

KEY WORDS: Density-based empirical likelihood, Empirical likelihood, Independence test, Likelihood, Nonlinear dependence, Nonparametric test, Random effect.

1. INTRODUCTION

Independence and dependence are key concepts related to the implementation and validity of many statistical procedures. Research questions of interest often focus on the degree of dependence between two paired variables. For example, in studies related to lung cancer, a complex multifactorial disease, statistical tests for relationships between genetic factors and environmental factors play a vital role in understanding the mechanisms of disease (e.g., Gu *et al*., 2012). The functionality and limitations of classical tests of dependence for pairs of random

variables have received a great deal of attention. Our focus is on a new empirical likelihood test as it compares to tests based on the Pearson (r), Spearman (r_s), and Kendall (τ) correlation coefficients. Despite the fact that these measures of dependence are designed to capture specific structures of dependence between two random variables, statistical software packages often only provide procedures to conduct the r and/or r_s and/or τ based tests for independence. As is well-known, the Pearson correlation coefficient is a measure of the strength of the linear relationship between two random variables (e.g., Pearson, 1920; Hauke and Kossowski, 2011). The Spearman correlation coefficient is a measure of a monotonic association between two random variables (e.g., Spearman, 1904; Hauke and Kossowski, 2011). Both the Pearson and the Spearman correlation coefficients are not well-suited to analyze non-monotone forms of dependence between two random variables (e.g., Embrechts *et al.*, 2002). The Kendall correlation coefficient (τ) is a well-known measure of the concordance between two rankings associated with two sets of observations (Kendall, 1938, 1948). However, in many cases tests based on Kendall's correlation coefficient shows relatively low power as compared to the Pearson and Spearman correlation coefficients (e.g., Mudholkar and Wilding, 2003). Note also that when dealing with dependence between two random variables *Y* and *X* that is defined via conditional expectations, say $E(Y \mid X) \neq E(Y)$, a test of a nonparametric regression model can provide a test of independence $E(Y \mid X) = E(Y)$ (e.g., Christensen, 2002). In many practical data analysis settings, it is of interest to develop a general coefficient that can efficiently detect both linear and non-linear dependence between random variables.

Our goal is to propose a test for independence that is powerful in detecting a variety of complex dependence structures. In addition to linear/nonlinear dependence structures, the present applied biostatistical literature introduces random-effect-type associations between two sets of

observations. For example, in the context of a lung cancer study, Gu *et al*. (2012) proposed to examine the dependence relationship of two polymorphisms rs1051730 and rs8034191 in random-effect-type model formulation. We aim to develop a simple and highly efficient nonparametric test that can be applied to detect dependence, including complex random effect type associations between two random variables. Towards this end, we construct a novel exact nonparametric likelihood ratio type test.

The likelihood ratio approach provides a basis for many important procedures and methods in statistical inference. When functional forms of data distributions are completely specified under the hypothesis to be tested the parametric likelihood approach is unarguably a powerful tool that can provide optimal statistical inference. In such cases, by virtue of the Neyman-Pearson lemma, the likelihood ratio tests yield the most powerful decision rules (e.g., Lehmann and Romano, 2005; Vexler and Wu, 2009; Vexler *et al.*, 2010a). However, the parametric likelihood methods cannot be applied properly if assumptions on the forms of distributions of data do not hold. Often, in the context of likelihood applications the use of misspecified parametric forms of data distributions may result in wrong and/or inaccurate statistical conclusions. It is also well-known that when these key assumptions are not met, the parametric approach may be extremely biased and inefficient when compared to robust nonparametric counterparts. In this paper, we provide a nonparametric strategy to approximate an optimal parametric likelihood ratio test-statistic via an empirical likelihood methodology.

Empirical likelihood (EL) concepts were introduced as nonparametric alternatives to parametric likelihood methods. The EL method for testing has been dealt with extensively across a variety of settings (e.g., Qin and Lawless, 1994; Owen, 2001; Lazar and Mykand, 1998; Lazar, 2003; Vexler *et al.*, 2009, 2010b, 2011, 2012b). Commonly, the EL function has the form

$EL = \prod_{i=1}^{n} p_i$, where the probability weights, $p_i, i = 1,...,n$ satisfy the trivial assumptions

$\sum_{i=1}^{n} p_i = 1$, $0 < p_i < 1$, $i = 1,...,n$ and the values of $p_i, i = 1,...,n$, are derived by maximizing the

EL function under empirical constraints. For example, when we draw an *i.i.d.* sample $X_1, ..., X_n$

under the null hypothesis that $EX_1 = 0$, the corresponding empirical constraint is $\sum_{i=1}^{n} p_i X_i = 0$.

Recent uses of the EL technique have been developed for hypothesis tests regarding various

parameters of distributions, such as moments (e.g., Qin and Lawless, 1994; Owen, 2001, Vexler

and Yu, 2011 for other examples).

Recently, several papers have introduced the density-based EL (dbEmpLike) approach

for creating nonparametric test statistics that approximate parametric Neyman-Pearson statistics

(e.g., Vexler and Gurevich, 2010; Gurevich and Vexler, 2011; Vexler *et al*., 2012a). The

dbEmpLike method proposes to consider the likelihood in the form of

$$L_f = \prod_{i=1}^{n} f(X_i) = \prod_{i=1}^{n} f_i, \ f_i = f(X_{(i)}),$$

where $f(\cdot)$ is a density function of observations $X_1,...,X_n$, and $X_{(1)} \leq ... \leq X_{(n)}$ are the order

statistics based on $X_1,...,X_n$. The dbEmpLike approach is a technique to approximate values of

f_j via maximization of L_f given a constraint related to the empirical version of the density

property of the form $\int f(X)dX = 1$. The dbEmpLike method was applied successfully to

construct powerful test procedures for goodness-of-fit (e.g., Vexler and Gurevich, 2010; Vexler

et al., 2011; Karagrigoriou, 2012) and two-sample problems (e.g., Gurevich and Vexler, 2011;

Vexler, Tsai *et al.*, 2012a). The dbEmpLike ratio tests have significantly improved power as

compared to the corresponding classical procedures, each with various forms of alternative

hypotheses. Miecznikowski *et al.* (2013) developed an R package (R Development Core Team) of statistical procedures based on the dbEmpLike methodology.

We introduce a simple dbEmpLike ratio test with high and stable power for detecting general cases of dependence. The proposed method is distribution-free, robust to model structures and highly efficient. Consider the simple example where we assume $X_1,\ldots,X_{15}$ are i.i.d. observations from a standard normal distribution and $Y_1,\ldots,Y_{15}$ are defined as $Y_i = X_i^2$, i $=1,\ldots,15$. In this case, the classical Pearson, Spearman, and Kendall tests based on $\{(X_i,Y_i), i$ $=1,\ldots,15\}$ for independence show powers of 0.36, 0.13, 0.21 at the significance level of 5%, whereas the proposed test provides power of 0.99. One advantage of the EL-based technique is that by applying the dbEmpLike approach we avoid having to specify a specific dependence structure, and hence we can efficiently detect linear, non-linear, and/or random-effect forms of dependence. An extensive Monte Carlo study is employed to support this conclusion. We observe that the proposed test either has relatively small power loss or is comparable to classical tests for independence when data has an exact linear dependence. On the other hand, in the cases of non-monotonic and/or random-effect type structures of the dependence, the proposed test is superior to the classical methods in terms of power gains. The asymptotic consistency of the proposed test is presented.

The paper is organized as follows. In Section 2, we introduce the dbEmpLike ratio test statistic. An extensive Monte Carlo comparison between the proposed test and the classical procedures is shown in Section 3. We discuss the performance of the tests under different alternative designs involving linear, non-linear and/or random-effect type forms of dependence. We show that the proposed test outperforms the classical tests in most of the scenarios we considered. In Section 4 the proposed test is applied to a biomarker study associated with

myocardial infarction (MI) disease. The data set was collected from a sample of randomly selected residents of Erie and Niagara counties at the age of 35 to 79 years. This study examined the diagnostic ability of the thiobarbituric acid-reactive substances (TBARS) biomarker for MI disease by indirectly testing for independence between the TBARS biomarker values and the other measurements related to vitamin E and cholesterol biomarkers. The epidemiological literature indicates significant associations between the biomarkers vitamin E, cholesterol, and the MI disease. We conclude with remarks in Section 5.

2. METHOD

In this section, we develop the dbEmpLike test of association between random variables X and Y. The asymptotic consistency of the proposed test is presented. The null distribution of the test statistic is examined in detail, and we prove that the test offers exact control of the Type I error rate.

2.1 Development of the test statistic

Assume we observe a sample of i.i.d. two-dimensional random vectors $(X_1, Y_1), ..., (X_n, Y_n)$. The joint distribution function of (X, Y) is denoted by F_{XY}, while $F_X(x)$ and $F_Y(y)$ denote the marginal distribution functions of X and Y, respectively. Consider the problem of testing the null hypothesis of bivariate independence

$$H_0 : F_{XY}(x, y) = F_X(x)F_Y(y), \text{ for all } x, y \in \mathbb{R}$$

against a wide class of alternatives, where F_{XY}, F_X, and F_Y are absolutely continuous with corresponding density functions f_{XY}, f_X, and f_Y, respectively. The forms of the distribution and density functions are assumed to be unknown. In the case of completely specified forms of the density functions, the likelihood ratio test statistic is given by

$$L = \prod_{i=1}^{n} \frac{f_{XY}(X_i, Y_i)}{f_X(X_i) f_Y(Y_i)} = \prod_{i=1}^{n} \frac{f_{XY}(X_{t(i)}, Y_{(i)})}{f_X(X_{t(i)}) f_Y(Y_{(i)})} = \prod_{i=1}^{n} \frac{f_{Y|X}(Y_{(i)} | X_{t(i)})}{f_Y(Y_{(i)})},$$

where $Y_{(1)} \leq ... \leq Y_{(n)}$ are the order statistics based on the observations $\{Y_1,...,Y_n\}$; $X_{t(i)}$ is the concomitant of the i-th order statistic (e.g., David and Nagaraja, 2003), i.e., the couple $(X_{t(i)}, Y_{(i)})$ belongs to $(X_1, Y_1),...,(X_n, Y_n)$. Our focus is on the nonparametric approximation to the likelihood L given above by applying the dbEmpLike methodology. We begin by evaluating the H_0-likelihood $\prod_{i=1}^{n} f_Y(Y_{(i)})$, the denominator of the likelihood ratio L. In the empirical likelihood approach, values of

$$f_i = f_Y(Y_{(i)}), i = 1,...,n,$$

can be estimated by maximizing the likelihood $\prod_{i=1}^{n} f_i$ with respect to $f_i, i = 1,...,n$, given an empirical constraint to control the assumption $\int f_Y(u)du = 1$ under H_0. We obtain a corresponding empirical form of the constraint $\int f_Y(u)du = 1$ into which $f_i, i = 1,...,n$ are convoluted. By virtue of Proposition 2.1 in Gurevich and Vexler (2011) one can show (for details see Appendix A) that for all positive integers, $m < n/2$ and $r < n/2$,

$$1 \geq \int_{Y_{(1)}}^{Y_{(n)}} f_Y(u)du \cong \frac{1}{2m} \sum_{i=1}^{n} \frac{f_i}{f_{Y|X}(Y_{(i)} | X_{(s_i)})} \tilde{\Delta}_i(m,r), \tag{1}$$

where $X_{(1)} \leq ... \leq X_{(n)}$ denote the order statistics based on the observations $\{X_1,...,X_n\}$; s_i is an integer number such that $X_{(s_i)} = X_{t(i)}$;

$$\tilde{\Delta}_i(m,r) \equiv \frac{F_n(X_{(s_i+r)}, Y_{(i+m)}) - F_n(X_{(s_i-r)}, Y_{(i+m)}) - F_n(X_{(s_i+r)}, Y_{(i-m)}) + F_n(X_{(s_i-r)}, Y_{(i-m)}) + n^{-\beta_1}}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})}, \tag{2}$$

$X_{(s_i+r)} = X_{(n)}$, if $s_i + r > n$; $X_{(s_i-r)} = X_{(1)}$, if $s_i - r < 1$; $F_{Xn}(u) = n^{-1}\sum_{i=1}^{n} I(X_i \le u)$ denotes the corresponding empirical distribution function based on $(X_1,...,X_n)$, $F_n(x,y)$ denotes the corresponding bivariate empirical distribution function based on $(X_1,Y_1),...,(X_n,Y_n)$, and $\beta_1 \in (0,0.5)$. In this case, following Crouse (1966), $F_n(x,y)$ is defined as

$$F_n(u_1,u_2) = n^{-1}\left\{\sum_{i=1}^{n} I(X_i < u_1, Y_i < u_2) + \frac{1}{2}\left[\sum_{i=1}^{n} I(X_i = u_1, Y_i < u_2) + \sum_{i=1}^{n} I(X_i < u_1, Y_i = u_2)\right] \right.$$
$$\left. + \frac{1}{4}\sum_{i=1}^{n} I(X_i = u_1, Y_i = u_2)\right\}.$$

In order to find the values for each $f_i, i = 1,...,n,$ that maximize the likelihood $\prod_{i=1}^{n} f_i$ under the empirical constraint

$$\frac{1}{2m}\sum_{i=1}^{n} \frac{f_i}{f_{Y|X}\left(Y_{(i)}\big|X_{(s_i)}\right)} \tilde{\Delta}_i(m,r) \le 1$$

implied by Equation (1), we can apply the corresponding Lagrangian function

$$\ell = \sum_{i=1}^{n} \log f_i + \lambda\left[\frac{1}{2m}\sum_{i=1}^{n} \frac{f_i}{f_{Y|X}\left(Y_{(i)}\big|X_{(s_i)}\right)} \tilde{\Delta}_i(m,r) - 1\right],$$

where λ is a Lagrange multiplier. Calculating roots of $\partial\ell/\partial f_i = 0, i = 1,...,n,$ one can show that

$$f_i = \frac{2m}{n\tilde{\Delta}_i(m,r)} f_{Y|X}\left(Y_{(i)}\big|X_{(s_i)}\right).$$

Thus, the dbEmpLike estimator of the likelihood ratio $\prod_{i=1}^{n} f_{Y|X}(Y_{(i)}|X_{t(i)})/f_i$ has the form of

$$\tilde{V}T_n(m,r) = \prod_{i=1}^{n} \frac{n}{2m}\tilde{\Delta}_i(m,r). \tag{3}$$

Note that the term $n^{-\beta_1}$, $0 < \beta_1 < 0.5$, in definition (2) ensures the consistency of the proposed test as is shown in the proof of Proposition 1 below (see also Appendix A for details). In Equation (2), $F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})$ takes on the values of $2r/n$, $(n-s_i+r)/n$, $1-1/n$, and $(s_i+r-1)/n$, when $(s_i+r<n, s_i-r>1)$, $(s_i+r>n, s_i-r>1)$, $(s_i+r>n, s_i-r<1)$, and $(s_i+r<n, s_i-r<1)$, respectively. The form of the test-statistic at (3) depends on values of the integer parameters m and r involved in the $\tilde{\Delta}_i(m,r)$-construction. Following the density-based EL literature (e.g., Vexler and Gurevich, 2010, p. 5; Gurevich and Vexler, 2011, pp. 5, 10-11), we can estimate m and r via the EL methodology obtaining the test statistic

$$VT_n = \max_{0.5n^{\beta_2} \le m \le \gamma_n} \max_{0.5n^{\beta_2} \le r \le \gamma_n} \prod_{i=1}^{n} \frac{n\tilde{\Delta}_i(m,r)}{2m}, \tag{4}$$

where $\gamma_n = \min(n^{0.9}, n/2)$, $0.75 < \beta_2 < 0.9$ and $\tilde{\Delta}_i(m,r)$ is defined in Equation (2). The conditions on m and r at definition (4) are sufficient to guarantee the proposed test is an asymptotic power one procedure as stated in Proposition 1.

Various Monte Carlo experiments based on more than one thousand different scenarios of (X,Y)-distributions and a variety of fixed sample sizes n showed that the statistic $\tilde{V}T_n(m,r)$ reaches its maximums with respect to $m \ge 0.5n^{\beta_2}$ and $r \ge 0.5n^{\beta_2}$ at $m = 0.5n^{\beta_2}$ and $r = 0.5n^{\beta_2}$ for all the experiments. Thus, we obtain the simple test statistic

$$VT_n = \prod_{i=1}^{n} n^{1-\beta_2} \tilde{\Delta}_i\left([0.5n^{\beta_2}], [0.5n^{\beta_2}]\right), \tag{5}$$

where the function $[x]$ denotes the nearest integer to x.

Accordingly, the decision rule of the proposed test is to reject the null hypothesis if $\log(VT_n) > C_\alpha$, where C_α is an α-level test threshold. The proposed test is exact, as shown in

9

Section 2.2 below. Note that extensive Monte Carlo simulations confirmed the robustness of the proposed test with respect to the values of β_1 at (2) and β_2, i.e. one can show that the power of the new test does not depend significantly on values of $\beta_1 \in (0,0.5)$ and $\beta_2 \in (0.75,0.9)$ under various scenarios of alternative distributions applied to the hypothesis of bivariate independence. Thus, without loss of generality, we set β_1 and β_2 in the test statistic to be 0.45 and 0.8, respectively.

Remark: The test statistic defined at (5) has a dbEmpLike ratio type form. Note that one can simplify this test statistic by taking into account that product of all the denominators from Equation (2) is simply a constant dependent on n.

Let E_{H_0} denote the expectation under H_0 and E_{H_1} denote the expectation under the alternative hypothesis $H_1 : F_{XY}(x, y) \neq F_X(x)F_Y(y)$, for some x, $y \in \mathbb{R}$. Assuming that

$$E_{H_0}(\log(\frac{f_{XY}(X_1,Y_1)}{f_X(X_1)f_Y(Y_1)})) \neq E_{H_1}(\log(\frac{f_{XY}(X_1,Y_1)}{f_X(X_1)f_Y(Y_1)})),$$

the following proposition demonstrates the consistency of the proposed test.

Proposition 1. Let (X,Y) be a random bivariate vector with the absolute continuous distribution function $F_{XY}(x, y)$ defined on $\Re^2$ and with the corresponding marginal distribution functions $F_X(x)$ and $F_Y(y)$. Assume that the expectations $E(\log(f_X(X_1)))$, $E(\log(f_Y(Y_1)))$ and $E(\log(f_{XY}(X_1Y_1)))$ are finite. Then, under H_0,

$$\frac{1}{n}\log(VT_n) \xrightarrow{p} E_{H_0}(\log(\frac{f_{XY}(X_1,Y_1)}{f_X(X_1)f_Y(Y_1)})) = 0, \text{ whereas, under } H_1,$$

$$\frac{1}{n}\log(VT_n) \xrightarrow{p} E_{H_1}(\log(\frac{f_{XY}(X_1,Y_1)}{f_X(X_1)f_Y(Y_1)})) > 0, \text{ as } n \rightarrow \infty.$$

Proof. The proof of this proposition is outlined in Appendix B.

2.2 Null distribution

In this section, we show that the proposed test statistic is distribution-free under H_0. We then present the critical values for the proposed test for different sample sizes. The test statistic VT_n depends only on the empirical distribution functions F_{Xn} and F_n, which in turn depend only on certain indicator functions. For example, for fixed x and y, we have

$$I(X \le x, Y \le y) = I\big(F_X(X) \le F_X(x), F_Y(Y) \le F_Y(y)\big) = I\big(U_1 \le u_1, U_2 \le u_2\big),$$

where U_1, U_2 are *Uniform*(0, 1) distributed, $u_1 = F_X(x)$ and $u_2 = F_Y(y)$. (For details regarding distribution free test constructions based on F_n see Crouse, 1966). Hence, it follows that

$$P_{H_0}\big(\log(VT_n) > C_\alpha\big) = P_{\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n \sim Uniform[0,1]}\big(\log(VT_n) > C_\alpha\big).$$

Therefore, the proposed method is distribution-free. Moreover the critical values for the dbEmpLike test can be accurately approximated using Monte Carlo techniques. In order to tabulate the percentiles of the null distribution of the test statistic $\log(VT_n)$ with $\beta_1 = 0.45$ and $\beta_2 = 0.8$, we drew 50,000 samples of $X_1,...,X_n \sim Uniform[0,1]$ and $Y_1,...,Y_n \sim Uniform[0,1]$ calculating values of $\log(VT_n)$ at each sample size n. The generated values of the test statistic $\log(VT_n)$ were used to determine the critical values C_α of the null distribution of $\log(VT_n)$ at the significance level α. The results of this Monte Carlo study are presented in Table 1.

Table 1. *Critical Values of the Proposed Test Statistic*

Sample size	α			
n	0.2	0.1	0.05	0.01
5	3.7592	3.7982	4.0709	4.3405

7	5.7211	5.9138	6.0390	6.3010
10	7.4321	7.6529	7.8549	8.2521
15	10.8383	11.1268	11.3766	11.8930
17	11.6964	11.9860	12.2427	12.7964
20	14.0863	14.4149	14.7061	15.3051
23	15.6590	15.9941	16.2962	16.9399
25	16.6106	16.9430	17.2632	17.9485
30	19.7139	20.0724	20.4089	21.1245
35	22.7744	23.1606	23.5258	24.2849
40	25.7939	26.1989	26.5672	27.3895
45	28.7824	29.2060	29.6057	30.4312
50	32.1148	32.5648	32.9714	33.8318
60	37.9011	38.3995	38.8356	39.7367
70	43.6649	44.1765	44.6390	45.6783
80	49.4025	49.9159	50.3733	51.3542
90	55.3020	55.8526	56.3487	57.4026
100	60.9451	61.4996	62.0054	63.0367

An R function (R Development Core Team, 2012) for Monte Carlo approximations for the critical values C_α of the null distribution of $\log(VT_n)$ is available as online supplementary material. This function can be easily modified to execute the proposed test based on real data.

Note that, since the statistic $\log(VT_n)$ is distribution-free under the null hypothesis, Monte Carlo or Bootstrap type procedures can be employed to estimate different characteristics of the distribution of $\log(VT_n)$ under H_0, e.g., its variance.

3. SIMULATION STUDY

We carried out an extensive Monte Carlo study to evaluate the performance of the proposed test. Lehmann (1975) noted that "the study of the power and efficiency of tests of independence is complicated by the difficulty of defining natural classes of alternatives to the hypothesis of independence". Furthermore, Kallenberg and Ledwina (1999) pointed out that power comparisons of tests of independence are rarely performed in statistical literature. These comments indicate the difficulty of implementing a Monte Carlo study for the power of tests of independence.

In this study we attended to general forms of dependence, which have been commonly pointed out with respect to bivariate data. That is, we consider linear and nonlinear forms of dependence between *X* and *Y*, including random-effect type structures of dependence. Following Kallenberg and Ledwina (1999), we analyzed three groups of alternatives: Group 1: non-linear correlation of *X* and *Y*; Group 2: linear correlation of *X* and *Y*; and Group 3: different bivariate distributions as alternatives to independence, including Pearson Type VII, Morgenstern, Plackett, and Cauchy distributions as described in Johnson (1987). Both the Groups, 1 and 2, include random-effects models. Table 2 displays formal definitions of the designs applied in the Monte Carlo experiments. A Pearson Type VII distribution represents symmetric distributions of *X* and *Y*, where random variables *X* and *Y* are uncorrelated but dependent.

Table 2. *Distributions for (X,Y) used in the power study*

Alternative Designs			*Models/Description*
		$X_i, i = 1…n$ $\quad$ $Y_i, i = 1…n$	
Group 1 (non-linear)	*Design 1.1*	$N(0,1)$	$1 + 0.2X_i + 0.8X_i^2 + \varepsilon_i$, $iid\,\varepsilon_i \sim N(0,1)$
	Design 1.2	$N(0,1)$	$0.5 + 0.1X_i + X_i^2 + \gamma_i X_i + \varepsilon_i$, $iid\,\gamma_i, \varepsilon_i \sim N(0,1)$
	Design 1.3	$N(0,1)$	$\log(1 + |X_i|)$
	Design 1.4	$N(0,1)$	$\log(1 + |X_i|)\gamma_i$, $\gamma_i \sim N(0,1)$
	Design 1.5	$N(0,1)$	$2 + 0.1\varepsilon_i / X_i$, $\varepsilon_i \sim N(0,1)$
Group 2 (linear)	*Design 2.1*	$Lognormal(0, 1)$	$1 + \gamma_i X_i$, $\gamma_i \sim N(0,1)$
	Design 2.2	$Lognormal(0, 1)$	$1 + 0.1X_i + 4\gamma_i X_i + \varepsilon_i$, $\gamma_i, \varepsilon_i \sim N(0,1)$
	Design 2.3	$N(0, 1)$	$2 + 0.1X_i + \varepsilon_i$, $\varepsilon_i \sim N(0,1)$
	Design 2.4	$U[0,1]$	$2 + 0.5X_i + \varepsilon_i$, $\varepsilon_i \sim N(0,1)$
	Design 2.5	$U[0,1]$	$2 + 0.5X_i + \gamma_i X_i + \varepsilon_i, \varepsilon_i \sim N(0,1), \gamma_i \sim N(0,2^2)$
	Design 2.6	$U[0,1]$	$2 + X_i + \varepsilon_i, \varepsilon_i \sim N(0,1)$
	Design 2.7	$U[0,1]$	$2 + X_i + \gamma_i X_i + \varepsilon_i, \varepsilon_i \sim N(0,1), \gamma_i \sim N(0,2^2)$
Group 3 (bivariate distributions)	*Design 3.1*	*Morgenstern ($\alpha = 1$)*	*Reference Johnson (1987), pp. 180-190*
	Design 3.2	*Plackett ($\psi = 3.5$)*	*Reference Johnson (1987), pp. 191-197*
	Design 3.3	*Pearson Type VII (m=1.1) with $\mu = 0$ and $\Sigma = I$*	*Reference Johnson (1987), pp. 117-121*
	Design 3.4	*The multivariate*	*Reference Johnson (1987), p. 44*

In this section we also examine numerically the power of a test proposed by Einmahl and McKeague (2003). Those authors constructed a test statistic by localizing the empirical likelihood using one or more 'time' variables implicit in the null hypothesis and then forming integrals of the log-likelihood ratio statistic.

Table 3 shows the results of the power evaluations of the proposed test ("$\log(VT_n)$"), the classical tests ("Kendall", "Pearson", "Spearman") and the Einmahl and McKeague's test ("EMcK") via the Monte Carlo study based on 10,000 replications of $X_1,...,X_n$ and $Y_1,...,Y_n$ for the designs corresponding to Groups 1-3 given in Table 2 at each sample size *n*. This study demonstrates the dbEmpLike test is superior to the considered classical tests in most scenarios under the designs of Groups 1-3.

The power differences between the novel test and the classical tests become more substantial as the sample size increases. For example, when the sample sizes are *n=30* and *n=35* the power of the proposed test is roughly two times larger than that of the classical tests given Group 1 alternatives. Under the designs 2.1 and 2.2 it is clear that the proposed test dramatically outperforms the classical tests. In these cases, the proposed test has approximately a 10%-60% power gain as compared to the classical procedures when $n \geq 25$.

In the case of design 2.3, where the linear model of dependence is dominant, the proposed test is comparable to the classical tests. In particular, note that for the sample size $n \geq$ 35 the proposed test provides the same Monte Carlo power as that of the classical procedures. In these scenarios, it is anticipated that the Pearson correlation test has higher power than the other considered tests since the Pearson correlation test is known to be developed to detect linear

forms of dependence between two random variables. Note that in the situations where the random-effect type structures of the dependence is present (e.g., the designs 1.2 and 2.2) the proposed test clearly outperforms the classical tests. Designs 2.4-2.7 consider scenarios based on mixed linear models as a function of an increasing linear fixed effect. In these cases, when the models do not include the random component γ the Pearson correlation test can be recommended as a very efficient test. However, when the random-effect γ is incorporated the dbEmpLike test has higher power than the other considered tests.

The conclusions regarding the power behavior of the tests under the designs 3.1 and 3.2 are similar to those related to the Monte Carlo simulations under the design 2.3. Under the designs given at 3.3 and 3.4 the proposed test has a considerable power gain in comparison to the Spearman and Kendall correlation tests (design 3.3) and the Spearman, Kendall correlation tests and the EMcK test (design 3.4), respectively.

Although Einmahl and McKeague's test demonstrated very powerful characteristics capturing various non-linear and linear dependence structures, the proposed dbEmpLike test outperformed the Einmahl and McKeague's test in many cases.

We conducted a brief evaluation of the data-driven rank tests for independence proposed in Kallenberg and Ledwina (1999). In order to develop these tests it was assumed that the observed samples are distributed following the joint density function given as

$$h\big(F_X(x),F_Y(y)\big)=c(\theta)\exp\left\{\sum_{j=1}^{}\theta_j b_j(F_X(x))b_j(F_y(y))\right\},$$

where b_j denotes the jth orthonormal Legendre polynomial, $\theta=(\theta_1,\theta_2,\ldots)^T$ and $c(\theta)$ is a normalizing constant. Kallenberg and Ledwina proposed the score tests for testing $\theta=0$ against $\theta\neq 0$. They assumed exponential families of data distributions corresponding to a very wide class of alternatives under which it is most probable that model-free tests for independence have less power than the data-driven rank tests. Note that one can demonstrate that many familiar

classes of distributions are not exponential families (e.g., Klauer, 1986). For example, we generated $(X_i, Y_i), i = 1,\ldots,50$ from the reciprocal-normal type distribution, using the R-command (R Development Core Team, 2012): 1/mvrnorm(n, rep(0, 2), Sigma), where

$\text{Sigma} = \begin{bmatrix} 0.1 & 0.05 \\ 0.05 & 0.1 \end{bmatrix}$. In this case the data-driven rank tests in the forms mentioned in Section 3

of Kallenberg and Ledwina (1999), gave the Monte Carlo powers 0.15 and 0.12 as compared to the Monte Carlo powers of 0.31, 0.15, 0.06, 0.12 and 0.13 obtained by using the tests "$\log(VT_n)$", "Kendall", "Pearson", "Spearman" and "EMcK", respectively. Note also, by virtue of the tests' structures, Kallenberg and Ledwina's tests require $E_{H_1} b_j(F_X(X))b_s(F_y(Y)) \neq 0$, for some j and s, to be consistent. Our Monte Carlo study showed that the data-driven rank tests had relatively low power when we generated, e.g., $X_i \sim Unif[0,1]$ and $Y_i = \arg\min_z \sum_{s,j} a_{sj} \left(b_j(X_i)b_s(z)\right)^2$ (or

$Y_i = \arg\min_z \left(\sum_{s,j} a_{sj} b_j(X_i)b_s(z)\right)^2$), where $a_{11}, a_{12},\ldots$ are constant. For example, when

$X_i \sim Unif[0,1]$, $Y_i = \arg\min_z \left(\sum_{j=1}^2 b_j(X_i)b_j(z)\right)^2, i = 1,\ldots,50$, the Monte Carlo powers had values of 0.69, 0.83 (the data-driven rank tests) and 1, 0.07, 0.37, 0.60, 1 ("$\log(VT_n)$", "Kendall", "Pearson", "Spearman" and "EMcK" tests). Similarly, when $X_i \sim Unif[0,1]$,

$Y_i = \arg\min_z \left(b_1(X_i)b_1(z) + 0.3b_1(X_i)b_2(z) + 0.5b_2(X_i)b_1(z) + b_2(X_i)b_2(z)\right)^2$, $1 \leq i \leq 50$, the results were 0.72, 0.66 (the data-driven rank tests) versus 1, 0.08, 0.26, 0.40, 1 ("$\log(VT_n)$", "Kendall", "Pearson", "Spearman" and "EMcK" tests). When $X_i \sim Unif[0,1]$,

$Y_i = \arg\min_z \sum_{j=1,2,4} \left(b_j(X_i)b_j(z)\right)^2, i = 1,\ldots,50$, the results were 0.58, 0.54 (the data-driven rank tests) versus 0.83, 0.02, 0.011, 0.03, 0.78 ("$\log(VT_n)$", "Kendall", "Pearson", "Spearman" and "EMcK" tests).

Based on the Monte Carlo results, we conclude that the proposed test exhibits high and stable power characteristics in comparison to the well-known classical procedures. Specifically, the proposed test performs reasonably well, and is generally competitive with the classical tests in the cases of linear forms of dependence. On the other hand, the proposed test significantly

outperforms the classical tests in terms of the power properties when detecting the nonlinear forms of bivariate dependence including random-effect type dependencies.

Table 3. *The Monte Carlo power of the tests.*

Tests	Design 1.1						Design 1.2					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.36	0.48	0.57	0.66	0.85	0.99	0.35	0.47	0.57	0.67	0.85	0.98
Kendall	0.13	0.14	0.15	0.16	0.18	0.22	0.11	0.11	0.12	0.13	0.13	0.13
Pearson	0.26	0.28	0.29	0.31	0.32	0.36	0.26	0.26	0.27	0.28	0.28	0.30
Spearman	0.12	0.13	0.14	0.15	0.17	0.19	0.09	0.10	0.10	0.10	0.11	0.11
EMcK	0.26	0.31	0.47	0.56	0.81	0.95	0.19	0.29	0.37	0.48	0.74	0.90
Tests	Design 1.3						Design 1.4					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.99	0.99	1.00	1.00	1.00	1.00	0.33	0.43	0.54	0.64	0.86	0.97
Kendall	0.23	0.23	0.23	0.24	0.24	0.24	0.12	0.12	0.13	0.13	0.13	0.13
Pearson	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.20	0.20	0.20	0.20
Spearman	0.14	0.14	0.14	0.15	0.15	0.15	0.10	0.10	0.10	0.10	0.10	0.10
EMcK	0.78	0.89	0.97	0.99	1.00	1.00	0.10	0.13	0.14	0.14	0.18	0.25
Tests	Design 1.5						Design 2.1					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70

Tests	Design (left)						Design (right)					
$\log(VT_n)$	0.12	0.15	0.23	0.36	0.76	0.94	0.51	0.65	0.76	0.87	0.97	1.00
Kendall	L[1]	L	L	L	L	L	0.09	0.09	0.10	0.10	0.11	0.11
Pearson	L	L	L	L	L	L	0.46	0.48	0.50	0.51	0.55	0.57
Spearman	L	L	L	L	L	L	0.09	0.09	0.09	0.10	0.10	0.10
EMcK	0.03	0.04	0.07	0.08	0.19	0.61	0.33	0.52	0.70	0.77	0.91	0.97

Tests	Design 2.2						Design 2.3					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.44	0.57	0.68	0.80	0.93	1.00	0.05	0.06	0.07	0.08	0.08	0.12
Kendall	0.09	0.10	0.10	0.10	0.10	0.10	0.05	0.06	0.07	0.08	0.09	0.12
Pearson	0.44	0.49	0.50	0.50	0.55	0.56	0.06	0.07	0.08	0.08	0.10	0.13
Spearman	0.08	0.09	0.09	0.09	0.09	0.09	0.06	0.07	0.07	0.08	0.09	0.12
EMcK	0.28	0.45	0.58	0.70	0.90	0.98	0.06	0.07	0.07	0.07	0.08	0.11

Tests	Design 2.4						Design 2.5					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.08	0.09	0.10	0.11	0.12	0.14	0.14	0.15	0.17	0.20	0.24	0.35
Kendall	0.08	0.09	0.11	0.11	0.16	0.20	0.06	0.07	0.08	0.09	0.11	0.12
Pearson	0.09	0.10	0.11	0.12	0.17	0.21	0.07	0.08	0.09	0.10	0.12	0.14
Spearman	0.09	0.10	0.11	0.11	0.16	0.21	0.07	0.07	0.08	0.09	0.11	0.12
EMcK	0.12	0.14	0.14	0.15	0.16	0.17	0.10	0.11	0.11	0.14	0.20	0.30

Tests	Design 2.6	Design 2.7
	Sample size (n)	Sample size (n)

	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.16	0.17	0.20	0.21	0.26	0.38	0.19	0.21	0.23	0.26	0.28	0.45
Kendall	0.18	0.23	0.30	0.33	0.46	0.62	0.12	0.13	0.16	0.18	0.25	0.33
Pearson	0.22	0.26	0.32	0.36	0.50	0.65	0.13	0.16	0.18	0.20	0.26	0.34
Spearman	0.20	0.24	0.31	0.34	0.47	0.62	0.13	0.14	0.16	0.18	0.24	0.32
EMcK	0.18	0.23	0.24	0.30	0.44	0.57	0.14	0.17	0.19	0.20	0.24	0.40

Tests	Design 3.1 ($\alpha = 1$)						Design 3.2 ($\psi = 3.5$)					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.24	0.31	0.34	0.43	0.63	0.75	0.58	0.68	0.79	0.88	0.98	1.00
Kendall	0.26	0.33	0.41	0.45	0.64	0.80	0.59	0.70	0.81	0.87	0.96	0.99
Pearson	0.30	0.38	0.45	0.50	0.68	0.82	0.54	0.63	0.72	0.79	0.91	0.98
Spearman	0.28	0.35	0.42	0.49	0.66	0.81	0.58	0.69	0.77	0.84	0.94	0.99
EMcK	0.23	0.30	0.34	0.45	0.62	0.73	0.59	0.68	0.80	0.89	0.97	0.98

Tests	Design 3.3						Design 3.4					
	Sample size (n)						Sample size (n)					
	20	25	30	35	50	70	20	25	30	35	50	70
$\log(VT_n)$	0.99	0.99	1.00	1.00	1.00	1.00	0.23	0.30	0.38	0.44	0.65	0.82
Kendall	0.19	0.20	0.20	0.20	0.21	0.22	0.11	0.12	0.12	0.13	0.13	0.13
Pearson	0.91	0.92	0.93	0.93	0.94	0.94	0.57	0.60	0.64	0.67	0.71	0.75
Spearman	0.13	0.13	0.14	0.14	0.14	0.14	0.09	0.09	0.09	0.10	0.11	0.11
EMcK	0.23	0.27	0.32	0.36	0.62	0.93	0.12	0.12	0.13	0.15	0.18	0.25

[1] The notation L indicates a number that is less than 0.007

4. DATA ANALYSIS

In this section, we present a data example to illustrate the practical application of the proposed test.

The use of thiobarbituric acid-reactive substances (TBARS) as a value to summarize total circulating oxidative stress in individuals is common in laboratory research (Armstrong, 1994), but its use as a discriminant factor between individuals with and without myocardial infarction (MI) disease is still controversial (e.g., Schisterman *et al*., 2001). Some authors have found a positive association between TBARS and MI disease (e.g., Jayakumari *et al*., 1992; Miwa *et al*., 1995), while others did not find corresponding significant associations (e.g., Karmansky *et al*., 1996). The aim of this study is to investigate the discriminative properties of TBARS with regard to MI disease by indirectly evaluating an association between TBARS and MI disease. Towards this end, we implemented separately two groups of tests of independence: one is in order to test for independence between TBARS and high-density lipoprotein (HDL)-cholesterol, and the other is to test for the independence between TBARS and vitamin E, where both the biomarkers, HDL-cholesterol and vitamin E, are historically known to be significantly associated with MI disease (e.g., Schisterman *et al*., 2001). Therefore, the results of these two groups of tests can be beneficial to entail indirect evidence on the discriminative ability of TBARS as a single biomarker in individuals with MI disease versus healthy individuals.

A sample of randomly selected residents of Erie and Niagara counties, 35 to 79 years of age, was employed in this investigation. The New York State department of Motor Vehicles drivers' license rolls was used as the sampling frame for adults between the age of 35 and 65, while the elderly sample (age 65 to 79) was randomly selected from the Health Care Financing Administration database. The study evaluated 230 measurements of TBARS, HDL-cholesterol

and vitamin E biomarkers. Half of them were collected on cases, who recently survived on MI disease, and the other half on controls, who had no previous MI disease. Table 4 depicts the *p*-values obtained via the dbEmpLike, Pearson, Spearman and Kendall procedures for the case and control groups, respectively.

Table 4*: The p-values obtained via the proposed test and the classical procedures*

Test of independence	Groups	n	$\log(VT_n)$	Pearson	Spearman	Kendall
TBARS *versus* HDL-cholesterol	Control	115	0.0147	0.0514	0.0709	0.0773
	Case	115	0.0228	0.1619	0.05671	0.0730
TBARS *versus* Vitamin E	Control	115	0.0440	0.0508	0.1168	0.0911
	Case	115	0.0019	0.0977	0.0503	0.0844

The classical tests provide *p*-values that are slightly larger than a significance level of 5%. As a result, the dependence between TBARS and HDL-cholesterol biomarkers, as well as the association between TBARS and vitamin E biomarkers, are not detected by the classical procedures. The proposed dbEmpLike test reveals a strong evidence of an association between TBARS and HDL-cholesterol biomarkers and a significant dependence between TBARS and vitamin E biomarkers. That is, the dbEmpLike test is more sensitive as compared with the classical methods to rejecting the null hypothesis of independence between TBARS and HDL-cholesterol biomarkers as well as for the test of independence between TBARS and vitamin E biomarkers.

5. CONCLUDING REMARKS

In this article we proposed and developed a novel density-based empirical likelihood ratio test for independence of two random variables. The proposed test is distribution-free, simple and can be easily applied in practice. The new procedure has very favorable and robust power properties against linear and non-monotone forms of dependence, with or without random effect type structures. Through extensive Monte Carlo simulation studies, we showed that the proposed test has significantly higher power as compared with the classical Pearson, Spearman, and Kendall test across a variety of scenarios. This study demonstrated that the proposed test can efficiently detect a broader class of dependence structures than can the classical techniques.

SUPPLEMENTARY MATERIALS

R Code: Code for Monte Carlo computing the critical values of the null distribution of the proposed test statistic.

REFERENCES

1. Armstrong, D. (1994), *Free radicals in diagnostic medicine: a systems approach to laboratory, technology, clinical correlations, and antioxidant therapy*, New York: Plenum Press.

2. Christensen, R. (2002), *Advanced Linear Modeling*, New York: Springer-Verlag New York, LLC.

3. Crouse, C. F. (1966), "Distribution Free Tests Based on the Sample Distribution Function," *Biometrika*, 53, 99-108.

4. David, H. A., and Nagaraja, H. N. (2003), *Order Statistics.* New York: Wiley

5. Einmahl, J. H. J. and McKeague, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli*, 9, 267-290.

6. Embrechts, P., McNeil, A., and Straumann, D. (2002), "Correlation and dependence in risk management: properties and pitfalls," In *Risk Management: Value at Risk and Beyond*, ed. M.A.H. Dempster, Cambridge University Press, Cambridge, 176-223.

7. Gu, M., Dong, X., Zhang, X., Wang, X., Qi, Y., Yu, J., and Niu, W. (2012), "Strong Association between Two Polymorphisms on 15q25.1 and Lung Cancer Risks: A Meta-Analysis," *PLoS ONE*, 7, e37970. DOI: 10.1371/journal.pone.0037970

8. Gurevich G., and Vexler, A. (2011), "A two-sample empirical likelihood ratio test based on samples entropy," *Statistics and Computing*, 21, 657-670.

9. Hauke, J., and Kossowksi, T. (2011), "Comparison of values of Pearson's and Spearman's correlation coeffcienets on the same sets of data," Quaestiones Geographicae, 3, 87-93.

10. Jayakumari, N., Ambikakumari, V., Balakrishnan, K. G., Subramonia Lyer, K. (1992), "Antioxidant status in relation to free radical production during stable and unstable angina syndromes," *Atherosclerosis*, 94, 183-190.

11. Johnson, M. E. (1987), *Multivariate Statistical Simulation*, New York: Wiley.

12. Kallenberg, W. C. M., and Ledwina, T. (1999), "Data-Driven Rank Tests for Independence," *Journal of the American Statistical Association*, 94, 285-301.

13. Karagrigoriou, A. (2012), *Goodness-of-Fit Tests for Reliability Modeling*, Springer: New York, pp. 253-267.

14. Karmansky, I., Shnaider, H., Palant, A., Gruener, N. (1996), "Plasma lipid oxidation and susceptibility of low-density lipoproteins to oxidation in male patients with stable coronary artery disease," *Clin Biochem*, 29, 573-579.

15. Kendall, M. G. (1938), "A new measure of rank correlation," *Biometrika*, 30, 81-89.

16. Kendall, M. G. (1948), *Rank Correlation Methods*, London: Griffin.

17. Kiefer, J. (1961), "On large deviations of the empiric d.f. of vector chance variables and a law of the iterated logarithm," *Pacific J. Math.*, 11, 649-660.

18. Klauer, K. C. (1986), "Non-exponential families of distributions," *Metrika*, 33, 299-305.

19. Lazar, N. A. (2003), "Bayesian Empirical Likelihood," *Biometrika*, 90, 319–326.

20. Lazar, N, and Mykland, P. A. (1998), "An evaluation of the power and conditionality properties of empirical likelihood," *Biometrika*, 85, 523-534.

21. Lehmann, E. L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, Oakland, CA: Holden-Day.

22. Lehmann, E. L., and Romano, J. P. (2005), *Testing Statistical Hypotheses*, Springer: New York.

23. Miwa, K., Miyagi, U., and Fujita, M. (1995), "Susceptibility of plasman low density liporprotein to cupric ion-induced peroxidation in patients with variant angina," *J Am Coll Cardiol*, 26, 632-638.

24. Miecznikowski, J. C., Vexler, A., and Shepherd, L. A. (2013), "dbEmpLikeGOF : An R package for nonparametric likelihood ratio tests for goodness-of-fit and two sample comparisons based on sample entropy." *Journal of Statistical Software*. In press.

25. Mudholkar, G. S., and Wilding, G. E. (2003), "On the conventional wisdom regarding two consistent tests of bivariate independence," *Journal of the Royal Statistical Society, Series D*, 52, 41-57.

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

27. Owen A. B. (1990), "Empirical Likelihood Ratio Confidence Regions," *The Annals of Statistics*, 18, 90-120.

28. Owen A. B. (2001), *Empirical Likelihood*, Chapman and Hall/CRC, New York.

29. Pearson, K. (1920), "Notes on the history of correlation," *Biometrika*, 13, 25-45.

30. Qin, J., and Lawless J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300-325.

31. R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0 <http://www.R-project.org>.

32. Schisterman, E. F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, N., and Trevisan, M. (2001), "TBARS and cardiovascular disease in a population-based sample," *Journal of Cardiovascular Risk*, 8, 219-225.

33. Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*. Wiley: New York, 1980.

34. Spearman, C. E. (1904), "The proof and measurements of association between two things," *American Journal of Psychology*, 15, 72-101.

35. Vasicek, O. (1976). "A test for normality based on sample entropy," Journal of the Royal Statistical Society, Ser. B, 38, 54-59.

36. Vexler A, and Gurevich G. (2010), "Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy," *Computational Statistics & Data Analysis*, 54, 531-545.

37. Vexler, A., Liu, S., Kang, L., and Hutson, A. D. (2009), "Modifications of the Empirical Likelihood Interval Estimation with Improved Coverage Probabilities," *Communications in Statistics (Simulation and Computation)*, 38, 2171-2183.

38. Vexler, A., Shan, G., Kim, S., Tsai, W.-M., Tian, L., and Hutson, A.D. (2011), "An empirical likelihood ratio based goodness-of-fit test for Inverse Gaussian distributions," *Journal of Statistical Planning and Inference*, 141, 2128-2140.

39. Vexler, A., Tsai, W.-M., Gurevich G., and Yu, J. (2012a), "Two-sample density-based empirical likelihood ratio tests based on paired data with an application to a treatment study of Attention-Deficit/Hyperactivity Disorder and Severe Mood Dysregulation," *Statistics in Medicine*, 31, 1821-1837.

40. Vexler, A., Tsai, W.-M., and Malinovsky, Y. (2012b), "Estimation and testing based on data subject to measurement errors: from parametric to non-parametric likelihood methods," *Statistics in Medicine*, 31, 2498-2512.

41. Vexler, A., and Wu, C. (2010), "An Optimal Retrospective Change Point Detection Policy," *Scandinavian Journal of Statistics*, 36, 542-558.

42. Vexler, A., and Yu, J. (2011), "Two-sample density-based empirical likelihood tests for incomplete data in application to a pneumonia study," *Biometrical Journal*, 53, 628-651.

43. Vexler, A., Wu, C., and Yu, K. F. (2010a), "Optimal hypothesis testing: from semi to fully Bayes factors," *Metrika*, 71, 125-138.

44. Vexler, A., Yu, J., Tian, L., and Liu, S. (2010b), "Two-sample nonparametric likelihood inference based on incomplete data with an application to a pneumonia study," *Biometrical Journal*, 52, 348–361.

45. Yu, J., Vexler, A., and Tian L. (2011), "Analyzing Incomplete Data Subject to a Threshold Using Empirical Likelihood Methods: An Application to a Pneumonia Risk Study in an ICU Setting," *Biometrics*, 66, 123-130.

46. Yu, J., Vexler, A., Kim, S., and Hutson, A. D. (2011), "Two-sample Empirical likelihood ratio tests for medians application to biomarker evaluations," *The Canadian Journal of Statistics*, 39, 671-689.

APPENDIX A: **The empirical constraint (1)**

Proposition 2.1 in Gurevich and Vexler (2011) shows that for all positive integers, $m < n/2$,

$$\frac{1}{2m}\sum_{i=1}^{n}\int_{Y_{(i-m)}}^{Y_{(i+m)}} f_Y(y)dy = \int_{Y_{(1)}}^{Y_{(n)}} f_Y(y)dy - \frac{1}{2m}\sum_{l=1}^{m-1}(m-l)\left[\int_{Y_{(n-l)}}^{Y_{(n-l+1)}} f_Y(y)dy + \int_{Y_{(l)}}^{Y_{(l+1)}} f_Y(y)dy\right],$$

where $Y_{(i+m)} = Y_{(n)}$, if $i+m > n$; $Y_{(i-m)} = Y_{(1)}$, if $i-m < 1$. Since $\int_{Y_{(1)}}^{Y_{(n)}} f_Y(y)dy \leq \int_{-\infty}^{+\infty} f_Y(y)dy = 1$,

$$\frac{1}{2m}\sum_{i=1}^{n}\int_{Y_{(i-m)}}^{Y_{(i+m)}} f_Y(y)dy \leq 1. \tag{A.1}$$

It is clear that when $m/n \to 0$ as $m,n \to \infty$, we have $(2m)^{-1}\sum_{i=1}^{n}\int_{Y_{(i-m)}}^{Y_{(i+m)}} f_Y(y)dy \cong 1$. In the interest of economy of space we refer the reader to Vexler and Gurevich (2010, p.533-534), Gurevich and Vexler (2011), Vexler and Yu (2011), Vexler *et al.*, (2011) and the "sample entropy"-literature cited in these papers for more details regarding the equations above and the integer parameter m at Equation (A.1).

Consider (A.1) in the form of

$$1 \geq \frac{1}{2m} \sum_{i=1}^{n} \int_{Y_{(i-m)}}^{Y_{(i+m)}} f_Y(y)dy = \frac{1}{2m} \sum_{i=1}^{n} \int_{Y_{(i-m)}}^{Y_{(i+m)}} \frac{f_{Y|X}\left(y|X_{t(i)}\right)}{f_{Y|X}\left(y|X_{t(i)}\right)} f_Y(y)dy \,, \tag{A.2}$$

where the conditional density $f_{Y|X}$ is denoted as

$$f_{Y|X}\left(y|X_{t(i)}\right) = \frac{f_{XY}(X_{t(i)}, y)}{f_X(X_{t(i)})} = \left(\frac{d}{dx} P\{X \leq x\}\Big|_{x=X_{t(i)}}\right)^{-1} \frac{d}{dy}\frac{d}{dx} P\{X \leq x, Y \leq y\}\Big|_{x=X_{t(i)}} \,.$$

By virtue of the Mean Value Theorem we have

$$\int_{Y_{(i-m)}}^{Y_{(i+m)}} \frac{f_{Y|X}\left(y|X_{t(i)}\right)}{f_{Y|X}\left(y|X_{t(i)}\right)} f_Y(y)dy \cong \frac{f_i}{f_{Y|X}\left(Y_{(i)}|X_{t(i)}\right)} \int_{Y_{(i-m)}}^{Y_{(i+m)}} f_{Y|X}\left(y|X_{t(i)}\right)dy \,.$$

Thus, taking into account (A.2), we can constrain values of $f_i, i = 1, \ldots, n,$ to satisfy

$$\frac{1}{2m} \sum_{i=1}^{n} \frac{f_i}{f_{Y|X}\left(Y_{(i)}|X_{t(i)}\right)} \int_{Y_{(i-m)}}^{Y_{(i+m)}} f_{Y|X}\left(y|X_{t(i)}\right)dy \equiv \frac{1}{2m} \sum_{i=1}^{n} \frac{f_i \, \Delta_i}{f_{Y|X}\left(Y_{(i)}|X_{t(i)}\right)} \leq 1, \tag{A.3}$$

where $\Delta_i = \int_{Y_{(i-m)}}^{Y_{(i+m)}} f_{Y|X}\left(y|X_{t(i)}\right)dy$. It follows straightforward that

$$\Delta_i = \frac{\dfrac{d}{dx} P\{X \leq x, Y \leq Y_{(i+m)}\}\Big|_{x=X_{t(i)}} - \dfrac{d}{dx} P\{X \leq x, Y \leq Y_{(i-m)}\}\Big|_{x=X_{t(i)}}}{\dfrac{d}{dx} P\{X \leq x\}\Big|_{x=X_{t(i)}}} \,.$$

The constraint at Equation (A.3) depends on the unknown theoretical distributions of the underlying observations. To obtain an empirical constraint corresponding to Equation (A.2), we need to estimate $\Delta_i, i = 1, \ldots, n$. Towards this end let $X_{(1)} \leq \ldots \leq X_{(n)}$ denote the order statistics based on the observations $\{X_1, \ldots, X_n\}$ and s_i be an integer number such that $X_{(s_i)} = X_{t(i)}$. Then it can be shown through the implementation of the dbEmpLike methodology that Δ_i can be approximated by

$$\Delta_i \cong \frac{P\{X \in (X_{(s_i-r)}, X_{(s_i+r)}), Y \le Y_{(i+m)}\} - P\{X \in (X_{(s_i-r)}, X_{(s_i+r)}), Y \le Y_{(i-m)}\}}{P\{X \le X_{(s_i+r)}\} - P\{X \le X_{(s_i-r)}\}}$$

$$= \frac{F_{XY}(X_{(s_i+r)}, Y_{(i+m)}) - F_{XY}(X_{(s_i-r)}, Y_{(i+m)}) - F_{XY}(X_{(s_i+r)}, Y_{(i-m)}) + F_{XY}(X_{(s_i-r)}, Y_{(i-m)})}{F_X(X_{(s_i+r)}) - F_X(X_{(s_i-r)})} \cong \tilde{\Delta}_i(m,r), \text{ (A.4)}$$

where $\tilde{\Delta}_i(m,r)$, an empirical estimator of Δ_i, is defined in (2). This empirical estimation is

based on sample entropy considerations, e.g. see Vexler and Gurevich (2010) as well as

Gurevich and Vexler (2011) for details. These papers extend the concept for estimating density

functions' values using sample entropy (e.g., Vasicek, 1976), which are presented as a

consequence of the dbEmpLike approach. For example, $dP\{X \le x\}/dx\big|_{x=X_{t(i)}}$ in the definition of

Δ_i can be approximated by $\left(F_X(X_{(s_i+r)}) - F_X(X_{(s_i-r)})\right)/(2r)$. The sample entropy and dbEmpLike

literature show that such approximations applied to construct test statistics are very efficient even

when the observed samples have relatively small sizes.

Regarding the term $n^{-\beta_1}$, $0 < \beta_1 < 0.5$, in the definition (2) of $\tilde{\Delta}_i(m,r)$, we note that

theoretically, across several situations, the values of the numerator of Δ_i can be of an order that

is comparable with that of $n^{-0.5}$. In these cases it can lead to a bias of the estimation of F_{XY} of an

order that is greater than that of F_{XY}. Additionally, in such scenarios the $n^{-\beta_1}$ term helps distance

the numerator of Δ_i from zero. When F_{XY} is close to zero, the term $n^{-\beta_1}$ makes $\tilde{\Delta}_i(m,r)$

estimable.

Thus, Equations (A.3) and (A.4) imply that the empirical constraint on the values of

$f_{Yi}, i = 1,...,n,$ is

$$\frac{1}{2m} \sum_{i=1}^{n} \frac{f_i}{f_{Y|X}\left(Y_{(i)}\big|X_{(s_i)}\right)} \tilde{\Delta}_i(m,r) \le 1.$$

APPENDIX B: PROOFS

This appendix comprises a proof scheme to establish Proposition 1.

Proof of Proposition 1.

We first study the elements of $\tilde{\Delta}_i(m,r)$ defined in (2). Towards this end we denote

$$Q_{XY,i} = F_{XY}(X_{(s_i+r)}, Y_{(i+m)}) - F_{XY}(X_{(s_i-r)}, Y_{(i+m)}) - F_{XY}(X_{(s_i+r)}, Y_{(i-m)}) + F_{XY}(X_{(s_i-r)}, Y_{(i-m)}) + n^{-\beta_1},$$

$$Q_{n,i} = F_n(X_{(s_i+r)}, Y_{(i+m)}) - F_n(X_{(s_i-r)}, Y_{(i+m)}) - F_n(X_{(s_i+r)}, Y_{(i-m)}) + F_n(X_{(s_i-r)}, Y_{(i-m)}) + n^{-\beta_1} \text{ and}$$

$$F_{Yn}(u) = n^{-1}\sum_{i=1}^{n} I(Y_i < u), \text{ where } F_n \text{ is defined in (2). It is clear that } F_{Yn}(Y_{(i+m)}) - F_{Yn}(Y_{(i-m)})$$

$= 2m/n$, when $i+m < n$, and $i-m > 1$. Then, since the definitions (2), (4) and $m/n \to 0$, as

$n \to \infty$, the statistic at Equation (3) can be reformulated as

$$n^{-1}\log(\tilde{V}T_n) = n^{-1}\sum_{i=1}^{n}\log\frac{\tilde{\Delta}_i(m,r)}{(2m/n)} \cong n^{-1}\sum_{i=1}^{n}\log\frac{Q_{n,i}}{\left(F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})\right)\left(F_{Yn}(Y_{(i+m)}) - F_{Yn}(Y_{(i-m)})\right)}$$

$$= -n^{-1}\sum_{i=1}^{n}\log\left(F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})\right) - n^{-1}\sum_{i=1}^{n}\log\left(F_{Yn}(Y_{(i+m)}) - F_{Yn}(Y_{(i-m)})\right) + n^{-1}\sum_{i=1}^{n}\log Q_{n,i}, \quad \text{(B.1)}$$

where $m, r \in (0.5n^{\beta_2}, n^{0.9})$ and $0.75 < \beta_2 < 0.9$. Regarding this equation, we show that

$$n^{-1}\sum_{i=1}^{n}\log\frac{F_X(X_{(s_i+r)}) - F_X(X_{(s_i-r)})}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})} \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Towards this end, we apply the theorem of Kolmogorov (e.g., Serfling, 1980): for all $\varepsilon \in (0, 1/2)$,

$$P\left\{\sup_{-\infty < u < \infty}\left|(F_X(u) - F_{Xn}(u))\right| > n^{-1/2+\varepsilon}\right\} \to 0, \text{ as } n \to \infty.$$

Now we consider the case of $\sup_{-\infty < u < \infty}\left|(F_X(u) - F_{Xn}(u))\right| \le n^{-1/2+\varepsilon}$. Using the fact that

$F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)}) \ge 2r/n$, through the definition of F_{Xn}, and $0.5n^{\beta_2} \le r$, we obtain that

for each $\varepsilon \in (0, 1/4)$,

$$n^{-1} \sum_{i=1}^{n} \log \frac{F_X(X_{(s_i+r)}) - F_X(X_{(s_i-r)})}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})}$$

$$= n^{-1} \sum_{i=1}^{n} \log \frac{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)}) + \left(F_X(X_{(s_i+r)}) - F_{Xn}(X_{(s_i+r)})\right) + \left(F_X(X_{(s_i-r)}) - F_{Xn}(X_{(s_i-r)})\right)}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})}$$

$$\leq n^{-1} \sum_{i=1}^{n} \log \frac{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)}) + 2n^{-1/2+\varepsilon}}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})}$$

$$\leq n^{-1} \sum_{i=1}^{n} \log\left(1 + \frac{n^{-1/2+\varepsilon}}{r/n}\right) \sim n^{-1} \sum_{i=1}^{n} \frac{n^{-1/2+\varepsilon}}{n^{\beta_2}/n} = n^{1/2-\beta_2+\varepsilon} \to 0 \text{, as } n \to \infty.$$

In this case, we also have

$$n^{-1} \sum_{i=1}^{n} \log \frac{F_X(X_{(s_i+r)}) - F_X(X_{(s_i-r)})}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})} \geq n^{-1} \sum_{i=1}^{n} \log \frac{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)}) - 2n^{-1/2+\varepsilon}}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})}$$

$$\geq n^{-1} \sum_{i=1}^{n} \log\left(1 - \frac{n^{-1/2+\varepsilon}}{2r/n}\right) \sim -n^{-1} \sum_{i=1}^{n} \frac{n^{-1/2+\varepsilon}}{n^{\beta_2}/n} = -n^{1/2-\beta_2+\varepsilon} \to 0 \text{, as } n \to \infty.$$

This implies

$$n^{-1} \sum_{i=1}^{n} \log \frac{F_X(X_{(s_i+r)}) - F_X(X_{(s_i-r)})}{F_{Xn}(X_{(s_i+r)}) - F_{Xn}(X_{(s_i-r)})} \xrightarrow{p} 0, \tag{B.2}$$

uniformly over $m, r \in (0.5n^{\beta_2}, n^{0.9})$, as $n \to \infty$.

Similarly, one can show that

$$n^{-1} \sum_{i=1}^{n} \log \frac{F_Y(Y_{(i+m)}) - F_Y(Y_{(i-m)})}{F_{Yn}(Y_{(i+m)}) - F_{Yn}(Y_{(i-m)})} \xrightarrow{p} 0, \tag{B.3}$$

uniformly over $m, r \in (0.5n^{\beta_2}, n^{0.9})$, as $n \to \infty$.

Now we consider the last term of (B.1) to show that $n^{-1} \sum_{i=1}^{n} \log Q_{XY,i} / Q_{n,i} \xrightarrow{p} 0$, as $n \to \infty$. To this

end, we apply Theorem 1 of Kiefer (1961); that is, for all $\varepsilon \in (0, 1/4)$,

$$P\left\{\sup_{u,v}\left|F_{XY}(u,v)-F_n(u,v)\right|>n^{-1/2+\varepsilon}\right\}\to 0,\text{ as }n\to\infty,$$

which clearly implies that

$$P\left\{\sup_i\left|Q_{XY,i}-Q_{n,i}\right|>2n^{-1/2+\varepsilon}\right\}\to 0,\text{ as }n\to\infty.$$

In the case of $\sup_i\left|Q_{XY,i}-Q_{n,i}\right|\le 2n^{-1/2+\varepsilon}$, applying the trivial inequality $Q_{n,i}\ge n^{-\beta_1}$, we have

$$n^{-1}\sum_{i=1}^n\log\frac{Q_{XY,i}}{Q_{n,i}}\le n^{-1}\sum_{i=1}^n\log\frac{Q_{n,i}+2n^{-1/2+\varepsilon}}{Q_{n,i}}\le n^{-1}\sum_{i=1}^n\log\left(1+\frac{2n^{-1/2+\varepsilon}}{n^{-\beta_1}}\right)\sim n^{-1}\sum_{i=1}^n\frac{n^{-1/2+\varepsilon}}{n^{-\beta_1}}$$
$$=n^{-0.5+\beta_1+\varepsilon}\to 0\text{ and}$$

$$n^{-1}\sum_{i=1}^n\log\frac{Q_{XY,i}}{Q_{n,i}}\ge n^{-1}\sum_{i=1}^n\log\frac{Q_{n,i}-2n^{-1/2+\varepsilon}}{Q_{n,i}}\ge n^{-1}\sum_{i=1}^n\log\left(1-\frac{2n^{-1/2+\varepsilon}}{n^{-\beta_1}}\right)\sim -n^{-1}\sum_{i=1}^n\frac{n^{-1/2+\varepsilon}}{n^{-\beta_1}}$$

$$=n^{-0.5+\beta_1+\varepsilon}\to 0,\text{ as }n\to\infty.$$

Thus,

$$n^{-1}\sum_{i=1}^n\log\frac{Q_{XY,i}}{Q_{n,i}}\xrightarrow{p}0,\tag{B.4}$$

uniformly over $m,r\in(0.5n^{\beta_2},n^{0.9})$, as $n\to\infty$.

Taking into account (B.1)-(B.4), we conclude that

$$n^{-1}\log(VT_n)=n^{-1}\sum_{i=1}^n\log\frac{Q_{XY,i}}{\left(F_X(X_{(s_i+r)})-F_X(X_{(s_i-r)})\right)\left(F_Y(Y_{(i+m)})-F_Y(Y_{(i-m)})\right)}+o_p(1),\tag{B.5}$$

uniformly over $m,r\in(0.5n^{\beta_2},n^{0.9})$.

Now we note that the test statistic at Equation (5) can be presented in the form

$$\frac{\log(VT_n)}{n}\cong n^{-1}\sum_{i=1}^n\log\frac{Q_{n,i}}{\left(F_{Xn}(X_{(s_i+r)})-F_{Xn}(X_{(s_i-r)})\right)\left(F_{Yn}(Y_{(i+m)})-F_{Yn}(Y_{(i-m)})\right)}=n^{-1}\sum_{i=1}^n\log\frac{Q_{n,i}^*}{A_iB_i},\tag{B.6}$$

where $m/n\to 0$,

$$A_i = \frac{1}{n}\left(\sum_{i=1}^{n} I\big(F_X(X_j) < F_X(X_{(s_i+r)})\big) - \sum_{j=1}^{n} I\big(F_X(X_j) < F_X(X_{(s_i-r)})\big)\right),$$

$$B_i = \frac{1}{n}\left(\sum_{i=1}^{n} I\big(F_Y(Y_j) < F_Y(Y_{(i+m)})\big) - \sum_{j=1}^{n} I\big(F_Y(Y_j) < F_Y(Y_{(i-m)})\big)\right), \text{ and}$$

$$Q_{n,i}^* = n^{-1}\Bigg\{ \sum_{j=1}^{n} I\big(F_X(X_j) < F_X(X_{(s_i+r)}), F_Y(Y_j) < F_Y(Y_{(i+m)})\big) - \sum_{j=1}^{n} I\big(F_X(X_j) < F_X(X_{(s_i-r)}), F_Y(Y_j) < F_Y(Y_{(i+m)})\big)$$

$$- \sum_{j=1}^{n} I\big(F_X(X_j) < F_X(X_{(s_i+r)}), F_Y(Y_j) < F_Y(Y_{(i-m)})\big) + \sum_{j=1}^{n} I\big(F_X(X_j) < F_X(X_{(s_i-r)}), F_Y(Y_j) < F_Y(Y_{(i-m)})\big)\Bigg\}.$$

(Here, for the sake of clarity and without loss of generality, we represent

$$F_n(u_1, u_2) = n^{-1}\sum_{i=1}^{n} I(X_i < u_1, Y_i < u_2) \text{ instead of the long notation shown in (2))}.$$

Thus, we can consider $n^{-1}\log(VT_n)$ to be based on uniformly distributed random variables

$U_i = F_X(X_i) \sim Uniform[0,1]$, and $W_i = F_Y(Y_i) \sim Uniform[0,1], i = 1,..,n$, i.e.

$$n^{-1}\log(VT_n) = n^{-1}\sum_{i=1}^{n}\log\frac{\tilde{Q}_{n,i}}{\tilde{A}_i\tilde{B}_i}, \tag{B.7}$$

where

$$\tilde{A}_i = n^{-1}\left(\sum_{j=1}^{n} I\big(U_j < U_{(s_i+r)}\big) - \sum_{j=1}^{n} I\big(U_j < U_{(s_i-r)}\big)\right), \quad \tilde{B}_i = n^{-1}\left(\sum_{j=1}^{n} I\big(W_j < W_{(i+m)}\big) - \sum_{j=1}^{n} I\big(W_j < W_{(i-m)}\big)\right),$$

and

$$\tilde{Q}_{n,i} = n^{-1}\Bigg\{ \sum_{j=1}^{n} I\big(U_j < U_{(s_i+r)}, W_j < W_{(i+m)}\big) - \sum_{j=1}^{n} I\big(U_j < U_{(s_i-r)}, W_j < W_{(i+m)}\big) -$$

$$\sum_{j=1}^{n} I\big(U_j < U_{(s_i+r)}, W_j < W_{(i-m)}\big) + \sum_{j=1}^{n} I\big(U_j < U_{(s_i-r)}, W_j < W_{(i-m)}\big)\Bigg\}.$$

Let $F_{UW}(u,w)$, $F_U(u)$, and $F_W(w)$ denote a distribution function and corresponding marginal

distribution functions of the uniformly distributed random variables $(U_i, W_i), i = 1,...,n$,

respectively, and let $f_{UW}(u,w)$, $f_U(u)$, and $f_W(w)$ be the respective density functions.

By virtue of (B.5), we have

$$n^{-1}\sum_{i=1}^{n}\log\frac{\tilde{Q}_{n,i}}{\tilde{A}_i\tilde{B}_i} = n^{-1}\sum_{i=1}^{n}\log\frac{Q_{UW,i}}{(U_{(s_i+r)}-U_{(s_i-r)})(W_{(i+m)}-W_{(i-m)})}+o_p(1),\qquad\text{(B.8)}$$

where

$$Q_{UW,i} = F_{UW}(U_{(s_i+r)},W_{(i+m)}) - F_{UW}(U_{(s_i-r)},W_{(i+m)}) - F_{UW}(U_{(s_i+r)},W_{(i-m)}) + F_{UW}(U_{(s_i-r)},W_{(i-m)}) + n^{-\beta_1}.$$

Consider the numerator term in the right-hand side of (B.8). We apply the Taylor argument to obtain the following result:

$$Q_{UW,i} = \left(U_{(s_i+r)}-U_{(s_i-r)}\right)\left(W_{(i+m)}-W_{(i-m)}\right)f_{UW}(U_{(s_i-r)},W_{(i-m)}) + \left(U_{(s_i+r)}-U_{(s_i-r)}\right)^2\frac{\partial^2 F_{UW}(u,W_{(i+m)})}{\partial^2 u}\bigg|_{u=U_i^*} +$$

$$\left(U_{(s_i+r)}-U_{(s_i-r)}\right)^2\frac{\partial^2 F_{UW}(u,W_{(i-m)})}{\partial^2 u}\bigg|_{u=U_i'} + \left(U_{(s_i+r)}-U_{(s_i-r)}\right)\left(W_{(i+m)}-W_{(i-m)}\right)^2\frac{\partial^2}{\partial^2 z}\frac{\partial F_{UW}(u,z)}{\partial u}\bigg|_{u=U_{(s_i-r)}}\bigg|_{z=W_i^*},$$

(B.9)

where $U_i^* \in \left(U_{(s_i-r)},U_{(s_i+r)}\right)$, $U_i' \in \left(U_{(s_i-r)},U_{(s_i+r)}\right)$ and $W_i^* \in \left(W_{(i-m)},W_{(i+m)}\right)$.

To analyze the second and third terms of the right-hand side of (B.9), we apply Chebyshev's inequality. It follows that for $\eta < 1-(2k)^{-1}$, $m,r \in (0.5n^{\beta_2},n^{0.9})$, $0.75 < \beta_2 < 0.9$, $k=1,2...$,

$$\sum_{i=1}^{n}P\left((U_{(s_i+r)}-U_{(s_i-r)})^{2k}\geq\varepsilon\right)\leq\sum_{i=1}^{n}E(U_{(s_i+r)}-U_{(s_i-r)})^{2k}\bigg/\varepsilon^k\leq\frac{r^{2k}}{\left(n^{(2k-1)}\varepsilon^k\right)}\leq\frac{n^{2\eta k}}{\left(n^{(2k-1)}\varepsilon^k\right)}\to 0,\qquad\text{(B.10)}$$

as $n\to\infty$, where $U_{(q)}$ is the qth order statistic based on standard uniformly distributed random variables (see for details David and Nagaraja, 2003). Then, this leads to the asymptotic result

$$n^{-1}\sum_{i=1}^{n}\log\frac{Q_{UW,i}}{(U_{(s_i+r)}-U_{(s_i-r)})(W_{(i+m)}-W_{(i-m)})} = n^{-1}\sum_{i=1}^{n}\log\left(f_{UW}(U_{(s_i-r)},W_{(i-m)})\right)+o_p(1),\text{ as }n\to\infty.\text{ (B.11)}$$

Using the Taylor series expansion, for $s_i > r, i > m$, we have

$$\log\left(f_{UW}(U_{(s_i-r)},W_{(i-m)})\right) = \log\left(f_{UW}(U_{(s_i)},W_{(i)}) + (U_{(s_i-r)}-U_{(s_i)})\frac{\partial f_{UW}(u,W_{(i-m)})}{\partial u}\bigg|_{u=U^*_{(s_i)}}\right.$$

$$+ \left(W_{(i-m)} - W_{(i)}\right) \frac{\partial f_{UW}\left(U_{(s_i-r)}, u\right)}{\partial u}\Bigg|_{u=W^*_{(i)}}\right), \tag{B.12}$$

where $U^*_{(s_i)} \in \left(U_{(s_i-r)}, U_{(s_i)}\right)$ and $W^*_{(i)} \in \left(W_{(i-m)}, W_{(i)}\right)$.

Noting that that $f_U(U_{(s_i)}) = f_W(W_{(i)}) = 1$ and by the results of (B.11) and (B.12), we have

$$n^{-1} \sum_{i=1}^{n} \log \frac{Q_{UW,i}}{(U_{(s_i+r)} - U_{(s_i-r)})(W_{(i+m)} - W_{(i-m)})} =$$

$$n^{-1} \sum_{i=1}^{n} \log \frac{f_{UW}(U_{(s_i)}, W_{(i)})}{f_U(U_{(s_i)}) f_W(W_{(i)})} + o_p(1) = n^{-1} \sum_{j=1}^{n} \log \frac{f_{UW}(U_j, W_j)}{f_U(U_j) f_W(W_j)} + o_p(1) \overset{p}{\to} E(\log(\frac{f_{UW}(U_1, W_1)}{f_U(U_1) f_W(W_1)})), \tag{B.13}$$

Combining (B.6), (B.8) and (B.13), we conclude

$$\frac{1}{n} \log(VT_n) \overset{p}{\to} E(\log(\frac{f_{UW}(U_1, W_1)}{f_U(U_1) f_W(W_1)})), \text{ as } n \to \infty. \tag{B.14}$$

It is clear that (B.14) completes the proof of Proposition 1.

A Simple Density-Based Empirical Likelihood Ratio Test for Independence

Supplementary Material

Albert Vexler, Wan-Min Tsai, Alan D. Hutson

Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, U.S.A

```
##########################################################
####### R code to calculate the critical values of the proposed test
##########################################################

# The sample size
n <- 14

# The number of the Monte Carlo simulations
mc <- 50000

# Set the values of m and r in Equation (4) of the paper to be m=r=0.5*n^0.8 (round to the
#nearest integer)
m <- r <- rep(round(0.5*n^0.8), n)

# Generate a vector that has values from 1 to the sample size
i <- 1:n

# Generate normal random values (the proposed test is exact)
x <- matrix(rnorm(n*mc), nrow =n)
y <- matrix(rnorm(n*mc), nrow =n)

# Sort generated x and y values
sx <- as.matrix(apply(x, 2, sort))
sy <- as.matrix(apply(y, 2, sort))

# Calculate the values of (i-m) and (i+m) in Equation (2)
yD <- 1*((i-m) < 1) + (i-m)*((i-m) >= 1)
yU <- n*((i+m) > n) + (i+m)*((i+m) <= n)

# Obtain X's paired with the ordered Y's
xt <- sapply(seq_len(mc), function(j) {x[order(y[, j]), j]})

# Obtain the rank of xt values obtained above (si is defined at Eq. (1))
si <- apply(xt, 2, rank)

# Calculate the values of (s(i)-r) and (s(i)+r) in Eq.(2)
```

```
xD <- sapply(seq_len(mc), function(j) {1*((si[, j] - r) < 1) + (si[, j] - r)*((si[, j] - r) >= 1)})
xU <- sapply(seq_len(mc), function(j) {n*((si[, j] + r) > n) + (si[, j] + r)*((si[, j] + r) <= n)})

# Store the x and y values corresponding to their lower and upper bounds
sxxU <- sapply(seq_len(mc), function(j) sx[xU[, j], j])
sxxD <- sapply(seq_len(mc), function(j) sx[xD[, j], j])
syyU <- sapply(seq_len(mc), function(j) sy[yU, j])
syyD <- sapply(seq_len(mc), function(j) sy[yD, j])

# Obtain the denominator of Eq (2)
denom <- xU - xD
denom[denom==0] <- 1/n

# Write a function for calculating the bivariate empirical function based on (Xi, Yi), i=1,...,n
#(without n^(-1))
fn <- function(u, v, j){
        sum(1*(x[, j] < u & y[, j] < v) + 0.5*(x[, j] == u & y[, j] < v) + 0.5*(x[, j]  < u & y[, j] ==
v)  + 0.25*(x[, j] == u & y[, j] == v))
}

# Write a function for the calculation of Equation (2)
fnn <- function(s, j, n){
        delta <- ( fn(sxxU[s, j] , syyU[s, j], j) - fn(sxxD[s, j] , syyU[s, j], j) - fn(sxxU[s, j] ,
syyD[s, j], j) + fn(sxxD[s, j] , syyD[s, j], j) +  n^0.55)/denom[s, j]
        if (delta==0) delta <- 1/n^2
        return(delta)
                        }

# Obtain the test statistic value following Eq. (5) (after taking log)
vt <- sapply(seq_len(mc), function(j) {sum(log((n^0.2)*sapply(i, function(s) {fnn(s, j, n)})))})

# Obtain  C_alpha of the proposed test
alpha <- c(0.8, 0.9, 0.95, 0.99)
( critvalue <- sapply(1:4, function(t) quantile(vt, alpha[t]))  )
```