Published in final edited form as:

J Stat Plan Inference. 2013 February 1; 143(2): 334–345. doi:10.1016/j.jspi.2012.07.018.

An Exact Density-Based Empirical Likelihood Ratio Test for Paired Data

Albert Vexler¹,

Department of Biostatistics, New York State University at Buffalo, Buffalo, NY 14214, USA

Gregory Gurevich, and

Department of Industrial Engineering and Management, SCE- Shamoon College of Engineering

Alan D. Hutson

Department of Biostatistics, New York State University at Buffalo, Buffalo, NY 14214, USA

Abstract

The Wilcoxon rank-sum test and its variants are historically well-known to be very powerful nonparametric decision rules for testing no location difference between two groups given paired data versus a shift alternative. In this article, we propose a new alternative empirical likelihood (EL) ratio approach for testing the equality of marginal distributions given that sampling is from a continuous bivariate population. We show that in various shift alternative scenarios the proposed exact test is superior to the classic nonparametric procedures, which may break down completely or are frequently inferior to the density-based EL ratio test. This is particularly true in the cases where there is a non-constant shift under the alternative or the data distributions are skewed. An extensive Monte Carlo study shows that the proposed test has excellent operating characteristics. We apply the density-based EL ratio test to analyze real data from two medical studies.

Keywords

Density-based empirical likelihood; Entropy; Empirical likelihood; Likelihood ratio; Paired data; Paired t-test; Skewed distributions; Test for symmetry; Two-sample location problem; Wilcoxon test

1. Introduction

In various fields of medical research one oftentimes encounters paired data based on pre- and post-experimental measurements. In the paired data setting the observations frequently come from distributions with skewed marginal densities. In addition, studies may suffer from small sample sizes, for example, due to the pilot nature of the study setting (e.g., Limpert et al., 2001). In several situations differences between the marginal distributions corresponding to the pre- and post-experimental conditions may be associated with different characteristics of the underlying distributions. For example, paired data can be presented by pre- and post-treatment measurements with unequal variances and a non-significant

© 2012 Elsevier B.V. All rights reserved.

Corresponding author. Tel.: +1 716 829 2754.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

difference between their means. In such cases classical nonparametric tests, which oftentimes rely on a strong exchangeability assumption based on the symmetry of paired differences, may break down completely (e.g., Albers *et al.*, 2001). In order to provide an efficient nonparametric comparison of distributions of pre- and post-measurements we propose and examine an exact test based on paired data. The proposed test is created by approximating an optimal Neyman-Pearson parametric decision rule in a novel manner through the use of a density-based empirical likelihood technique (e.g., Vexler and Gurevich, 2010; Vexler *et al.*, 2011; Vexler and Yu, 2011). We show that in several cases the proposed test is superior to the classic procedures.

In order to express the general framework for the statement of the test development process let us first define $(X_1, Y_1), \dots, (X_n, Y_n)$ to be a random sample from a bivariate population with absolutely continuous joint distribution function $F_{XY}(x, y)$ with marginal distributions of X_i and Y_i given as $F_X(x)$ and $F_Y(y)$, respectively. Oftentimes, given this type of data the general location testing problem consists of testing $H_0: \theta = 0$ versus $H_1: \theta \neq 0$ through the relationship $F_X(x) = F_Y(x - \theta)$. For example, in biomedical experiments θ may represent the mean or median difference between subject values measured pre- and post-treatment. Common statistical procedures for testing $H_0: \theta = 0$ versus $H_1: \theta \neq 0$ consist of the paired t-test, the sign test and the Wilcoxon signed-rank test. These tests are based on the n paired differences $Z_i = X_i - Y_i$ (e.g., Bhattacharya *et al.*, 1982; Fellingham and Stoker, 1964; Lam and Longnecker, 1983). In the paired data case, the classical nonparametric Wilcoxon signed-rank procedure for testing $H_0: \theta = 0$ versus $H_1: \theta \neq 0$ is a permutation based method under the assumption that Z is symmetric about the parameter θ under the null hypothesis. Similarly, the sign test assumes symmetry under the null hypothesis and utilizes the binomial distribution with respect to generating the null distribution of the test statistic. This article develops a novel nonparametric procedure for testing the hypothesis that the distribution of a sequence of *i.i.d.* observations, $Z_i = X_i - Y_i$, $i = 1, \dots, n$, is symmetric. The nonparametric literature tends to associate the problem of testing the hypothesis $F_X(x) = F_Y(x)$, e.g. in the context of detecting treatment effects, with the statement of testing the symmetry of $\Pr\{Z_1 < x\}$ when the forms of the distribution functions $F_X(x)$, $F_Y(y)$ and $F_{XY}(x, y)$ are completely unknown (e.g., Wilcoxon, 1945). It should also be noted that when the parametric forms of the distribution functions, F_X , F_Y and F_{XY} are assumed to be known the parametric likelihood ratio can efficiently be applied to the testing problem described above. The likelihood principle is arguably one of the most important concepts relative to designing inferential procedures within the parametric modeling framework. Theoretically it can be shown that when a density function f_Z of Z is assumed to be known under the null and alternative hypotheses an optimal parametric likelihood ratio test can be developed (e.g. see Lehmann and Romano, 2005; Vexler and Wu, 2009; Vexler *et al.* (2010a). In this article we develop a likelihood ratio type statistic in an empirical form as an alternative to the classical nonparametric test statistics for testing symmetry or the null hypothesis of the form $H_0: \theta = 0$.

Recently, the likelihood approach has also been shown to be useful in a nonparametric context. Many nonparametric and semiparametric approximations to powerful parametric likelihood procedures have been used routinely in both statistical theory and practice. Well known examples are the quasi-likelihood method (e.g., Wedderburn, 1974), approximations of parametric likelihoods via orthogonal functions (e.g., Claeskens and Hjort, 2004), techniques based on quadratic artificial likelihood functions (e.g., Wang, 2006), the local maximum likelihood methodology (e.g., Fan, Farmen and Gijbels, 1998) etc. Various studies have shown that frequently artificial or approximate likelihood-based techniques efficiently incorporate information expressed through data and have many of the same asymptotic properties as those derived from corresponding parametric likelihoods. The empirical likelihood (EL) method is one of a growing array of artificial or approximate

likelihood-based methods currently in use in statistical practice. A question of major interest focuses on the performance of this new construct relative to ordinary parametric likelihood ratio-based procedures. The EL method for testing has been dealt with extensively in the literature within a variety of settings (e.g., Owen, 2001; Qin and Lawless, 1994; Lazar, 2003; Vexler *et al.*, 2009; Yu, Vexler and Tian, 2010; Vexler *et al.*, 2010b).

As background relative to the development of our new EL test we first outline the classic approach. This will allow us to point out our extensions that delineate it from the current EL

methods. The classic EL takes the form $\prod_{i=1}^n (F(Z_i) - F(Z_{i-}))$, which is a function of the cumulative distribution function F . This EL technique is “distribution function-based”. An

empirical estimator of this likelihood may take the form of $L_p = \prod_{i=1}^n p_i$, where the components p_i , $i = 1, \dots, n$ should maximize the likelihood L_p provided that empirical

constraints based on $Z_1, \dots, Z_n$ hold, e.g. $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n p_i Z_i = 0$, where $\sum_{i=1}^n p_i Z_i$ is an empirical form of $E(Z)$ under the hypothesis $E(Z) = 0$. Computation of the EL’s components p_i , $i = 1, \dots, n$ is commonly a simple exercise in the utilization of Lagrange multipliers, e.g. see the references listed above. The EL technique is commonly applied to test for hypotheses regarding parameters (e.g., moments) of distributions (e.g., Qin and Lawless, 1994; Owen, 2001; Vexler and Yu, 2011). In this note we focus on overall distribution functions of paired observations.

According to the well-known Neyman-Pearson Lemma, density-based likelihood ratios can provide uniformly most powerful tests. Using this as a starting point, Vexler and Gurevich (2010) proposed an alternative to the “distribution function-based” EL methodology. They utilized the approximate density-based likelihood, which has the following form:

$$L_f = \prod_{i=1}^n f(Z_i) = \prod_{i=1}^n f_i, \quad f_i = f(Z_{(i)}),$$

where $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$ are the order statistics based on, $Z_1, \dots, Z_n$ and $f_1, \dots, f_n$ take on the values that maximize L_f given the empirical constraint corresponding to $\int f(u) du = 1$. This density-based EL approach was utilized successfully in order to construct efficient entropy-based goodness-of-fit test procedures. The entropy-based tests had improved power as compared to the classical tests (e.g., Vasicek, 1976; Dudewicz and Van Der Meulen, 1981; Arizono and Ohta, 1989; Mudholkar and Tian, 2002; Vexler and Gurevich, 2010; Vexler *et al.*, 2011).

In terms of constructing the nonparametric likelihood test for the general location testing problem there are two critical issues in the context of a direct use of the technique proposed by Vexler and Gurevich (2010). First, in contrast to the density-based EL approach for goodness-of-fit testing, parametric forms of relevant density functions cannot be assumed to be known, even under the null hypothesis. That is, ratios such as “(density-based EL)/(parametric likelihood)” cannot serve as test statistics in the setting being examined in this note. The second issue is that, using the Vexler and Gurevich approach, the approximate likelihood ratio might be presented in a form of “(H_I -density-based EL)/(H_0 -density-based EL)”, where the numerator and denominator consist of density-based empirical likelihoods derived under empirical constraints with respect to the alternative and null hypotheses, respectively. In this case, one can easily show the relevant constraints are very complicated and the constructed test statistic has an H_0 -distribution function that depends on data distributions. These factors bring difficulties to provide the corresponding Type I error control of the test, whereas it is reasonable to derive a density-based EL ratio test statistic developing an exact test.

In this article, we develop a new density-based EL approach for constructing a novel exact test for the general two-sample problem of detecting shift alternatives in the bivariate data setting. The proposed distribution-free test statistic approximates a parametric likelihood ratio test using an extension of the one-sample density-based EL methodology described above. Given that it was noted by Albers *et al.* (2001) that for many applications a simple constant location-shift alternative is far from realistic we provide a more general framework in terms of various types of alternatives. For example, in the case when the variance of X differs from that of Y and the marginal distributions $F_X(x)$, $F_Y(y)$ are symmetric roughly about the same point the standard procedures for testing $F_X(x) = F_Y(x)$ may break down completely. This is not the case for the new EL-based method. In addition, we show that the new EL-based test is sensitive with respect to detecting non-constant shift alternatives occurring in the two-sample problem.

The newly developed density-based EL decision rule is found to be very powerful test for the challenging two-sample paired data problem. We show that under a variety of scenarios the proposed test is oftentimes superior to the classical procedures. We examine the property of superior power of the proposed procedure over classic tests with a focus on skewed data (e.g., data from lognormal distributions). In various fields of medical research measurements biomarker data are commonly represented by skewed data. Such skewed distributions frequently are accommodated by the lognormal distribution (e.g., Limpert *et al.*, 2001). We utilize this notion of lognormality with respect to developing our simulation study.

The article is organized as follows: In Section 2 we describe the proposed method. The development of the main test statistic is introduced and theoretical propositions are presented; An extensive simulation study is performed and shown in Section 3; In Section 4 we illustrate the proposed method with data from the two real studies: 1) A study of 56 assay pairs of cyclosporine from blood samples of organ transplant recipients previously analyzed by Hawkins (2002) using parametric methods and 2) An evaluation of a group-based therapy for children with attention-deficit/hyperactivity disorder and severe mood dysregulation. We show that the new test clearly outperforms the standard procedures, while analyzing data related to the second study. Finally, we conclude this article with a discussion in Section 5.

2. Constructing the density-based EL ratio

Let $Z_1, \dots, Z_n$ be independent identically distributed (i.i.d.) random variables with corresponding distribution function F , where $Z_i = X_i - Y_i$ as described in Section 1. We begin by developing the basic methodology from which to construct the density-based EL ratio test statistic given the following null and alternative hypotheses:

$$H_0: F = F_{H_0}, F_{H_0}(u) = 1 - F_{H_0}(-u), \text{ for all } -\infty < u < \infty \quad \text{versus} \quad (1)$$

$$H_1: F = F_{H_1}, F_{H_1}(u) \neq 1 - F_{H_1}(-u), \text{ for some } -\infty < u < \infty.$$

The parametric likelihood ratio statistic based on $Z_1, \dots, Z_n$ takes the form of

$$LR = \frac{\prod_{i=1}^n f_{H_1}(Z_i)}{\prod_{i=1}^n f_{H_0}(Z_i)} = \frac{\prod_{j=1}^n f_{H_1}(Z_{(j)})}{\prod_{j=1}^n f_{H_0}(Z_{(j)})} = \frac{\prod_{j=1}^n f_{H_1,j}}{\prod_{j=1}^n f_{H_0,j}}, \quad (2)$$

where $f_{H_1}(u)$, $f_{H_0}(u)$ denote the density functions of Z under H_1 and H_0 , respectively, $Z_{(1)} \dots Z_{(n)}$ is the order statistic and $f_{H_{kj}} = f_{H_k}(Z_{(j)})$, $k = 0, 1, j = 1, \dots, n$. In order to generate the nonparametric test for (1) we first estimate the values of $f_{H_{1,j}}$, $j = 1, \dots, n$. This

is accomplished via maximizing the log-likelihood $\sum_{j=1}^n \log(f_{H_{1,j}})$ provided that $f_{H_{1,j}}$, $j = 1, \dots, n$ satisfy an empirical constraint that is the empirical equivalent of $\int f_{H_1}(u) du = 1$, e.g. see Vexler and Gurevich (2010) for details. To formalize this constraint, we specify $Z_{(j)} = Z_{(1)}$, if $j = 1$, and $Z_{(j)} = Z_{(n)}$, if $j = n$, and utilize the relationship

$$\frac{1}{2m} \sum_{j=1}^n \int_{Z_{(j-m)}}^{Z_{(j+m)}} f_{H_1}(u) du = \int_{Z_{(1)}}^{Z_{(n)}} f_{H_1}(u) du - \sum_{l=1}^{m-1} \frac{(m-l)}{2m} \left(\int_{Z_{(n-l)}}^{Z_{(n-l+1)}} f_{H_1}(u) du + \int_{Z_{(l)}}^{Z_{(l+1)}} f_{H_1}(u) du \right), \text{ for all integer } m \leq \frac{n}{2}, \quad (3)$$

which appears in Proposition 2.1 of Vexler and Gurevich (2010). Since

$$\int_{Z_{(1)}}^{Z_{(n)}} f_{H_1}(u) du \leq \int_{-\infty}^{\infty} f_{H_1}(u) du = 1, \text{ equation (3) implies the inequality that}$$

$$(1/2m) \sum_{j=1}^n \int_{Z_{(j-m)}}^{Z_{(j+m)}} f_{H_1}(u) du \leq 1. \text{ In addition, based on previous work one can expect that}$$

$$(1/2m) \sum_{j=1}^n \int_{Z_{(j-m)}}^{Z_{(j+m)}} f_{H_1}(u) du \approx 1 \text{ when } m/n \rightarrow 0, \text{ as } m, n \rightarrow \infty \text{ (e.g., Vexler et al., 2011).}$$

Let $\hat{F}_{H_0}$ and $\hat{F}_{H_1}$ denote estimators of F_{H_0} and F_{H_1} , respectively. Utilizing the approximate analog to the mean-value integration theorem, the empirical approximations

$$\begin{aligned} & \sum_{j=1}^n \int_{Z_{(j-m)}}^{Z_{(j+m)}} f_{H_1}(u) du \\ &= \sum_{j=1}^n \int_{Z_{(j-m)}}^{Z_{(j+m)}} \frac{f_{H_1}(u)}{f_{H_0}(u)} f_{H_0}(u) du \approx \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} \int_{Z_{(j-m)}}^{Z_{(j+m)}} f_{H_0}(u) du \approx \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} (F_{H_0}(Z_{(j+m)}) - F_{H_0}(Z_{(j-m)})) \approx \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} (\hat{F}_{H_0}(Z_{(j+m)}) \\ & \quad - \hat{F}_{H_0}(Z_{(j-m)})), \end{aligned}$$

$$\int_{Z_{(1)}}^{Z_{(n)}} f_{H_1}(u) du \approx \hat{F}_{H_1}(Z_{(n)}) - \hat{F}_{H_1}(Z_{(1)}),$$

$$\left(\int_{Z_{(n-l)}}^{Z_{(n-l+1)}} f_{H_1}(u) du + \int_{Z_{(l)}}^{Z_{(l+1)}} f_{H_1}(u) du \right) \approx (\hat{F}_{H_1}(Z_{(n-l+1)}) - \hat{F}_{H_1}(Z_{(n-l)}) + \hat{F}_{H_1}(Z_{(l+1)}) - \hat{F}_{H_1}(Z_{(l)}))$$

can be derived and applied directly to functions defined at (3). If we let $\hat{F}_{H_1}$ be the empirical distribution function in these approximations we then can record the empirical version of (3) having the following form:

$$\frac{1}{2m} \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} (\hat{F}_{H_0}(Z_{(j+m)}) - \hat{F}_{H_0}(Z_{(j-m)})) = \left(1 - \frac{1}{n}\right) - \frac{(m-1)}{2n}.$$

To define $\hat{F}_{H_0}$, the distribution free estimation for a symmetric distribution proposed by Schuster (1975) is applied. Thus, we formulate the empirical constraint

$$\begin{aligned} \frac{1}{2m} \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} \Delta_{jm} &= 1 - \frac{(m+1)}{2n}, \\ \Delta_{jm} &:= \frac{1}{2n} \sum_{i=1}^n \left(I(Z_i \leq Z_{(j+m)}) + I(-Z_i \leq Z_{(j+m)}) - I(Z_i \leq Z_{(j-m)}) - I(-Z_i \leq Z_{(j-m)}) \right) \end{aligned} \quad (4)$$

on $f_{H_{1,j}}, j = 1, \dots, n$, where $I(\cdot)$ denotes the standard indicator function. In order to find values of $f_{H_{1,j}}$ that maximize the log-likelihood $\sum_{j=1}^n \log(f_{H_{1,j}})$ subject to the constraints at (4), we derive $f_{H_{1,j}}, j = 1, \dots, n$, from the Lagrange function given by

$$\Lambda_\lambda = \sum_{j=1}^n \log(f_{H_{1,j}}) + \lambda \left(1 - \frac{m+1}{2n} - \frac{1}{2m} \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} \Delta_{jm} \right)$$

with a Lagrange multiplier λ . It is straightforward to obtain that

$$f_{H_{1,j}} = f_{H_{0,j}} \frac{2m(1 - (m+1)(2n)^{-1})}{n\Delta_{jm}}, \quad j = 1, \dots, n,$$

are the roots of the equation $\Lambda_\lambda / f_{H_{0,j}} = 0$ such that $f_{H_{1,j}}, j = 1, \dots, n$, satisfy the constraints at (4). This implies that the empirical maximum likelihood approximation to the H_I -

likelihood $\prod_{j=1}^n f_{H_{1,j}}$ can be presented as

$$\prod_{j=1}^n f_{H_{0,j}} \frac{2m(1 - (m+1)(2n)^{-1})}{n\Delta_{jm}}.$$

Therefore, the likelihood ratio given at (2) has the empirical form

$$V_{nm} = \prod_{j=1}^n \frac{2m(1 - (m+1)(2n)^{-1})}{n\Delta_{jm}}. \quad (5)$$

This test statistic is the density-based empirical likelihood ratio having a structure similar to those of statistics based on sample entropy used to test for goodness-of-fit (e.g., Vasicek, 1976; Dudewicz and Van Der Meulen, 1981; Arizono and Ohta, 1989; Mudholkar and Tian, 2002; Vexler and Gurevich, 2010). Although entropy-based tests are known to be very powerful there are difficulties when applying these tests in practice. The actual power of the entropy-based tests strongly depends on integer parameters involved in the related structure of the test statistic. For example, the distribution function of V_{nm} at (5) depends on m . We employ the empirical likelihood concept to specify the test statistic for (1). In Appendix A

we outline the technical details that are necessary in order to modify the test statistic (5) to be of the slightly modified form

$$V_n = \min_m \prod_{j=1}^n \frac{2m(1 - (m+1)(2n)^{-1})}{n\Delta_{jm}}. \quad (6)$$

Now to complete the construction of the test statistic, we note that $m/n \rightarrow 0$ as $m, n \rightarrow \infty$ is a necessary assumption. In accordance with the constraints at (3) and (4), the requirement $m/n \rightarrow 0$ as $m, n \rightarrow \infty$ reduces asymptotically the remainder terms

$$\sum_{l=1}^{m-1} \frac{(m-l)}{2m} \left(\int_{Z_{(n-l)}}^{Z_{(n-l+1)}} f_{H_1}(u) du + \int_{Z_{(l)}}^{Z_{(l+1)}} f_{H_1}(u) du \right) \text{ and } \frac{(m+1)}{2n},$$

respectively. Here, as mentioned above,

$$\sum_{l=1}^{m-1} \frac{(m-l)}{2m} \left(\int_{Z_{(n-l)}}^{Z_{(n-l+1)}} f_{H_1}(u) du + \int_{Z_{(l)}}^{Z_{(l+1)}} f_{H_1}(u) du \right) \approx \frac{(m+1)}{2n} \xrightarrow{m, n \rightarrow \infty} 0, \text{ if } m/n \rightarrow 0 \text{ as } m, n \rightarrow \infty.$$

The condition on m as stated in Proposition 1 of this article is necessary in order to facilitate the EL ratio test as an asymptotic power one procedure. Investigating asymptotic properties of entropy based tests for goodness-of-fit, the published research often assumes conditions similar to the requirement $m/n \rightarrow 0$ as $m, n \rightarrow \infty$ (e.g., Vasicek, 1976; Tusnady, 1977; Vexler and Gurevich, 2010). We also note that to preserve the consistency of the new test statistic (Proposition 1 below), we require that $m \rightarrow \infty$ as $n \rightarrow \infty$, which is also in line with related restrictions mentioned in the entropy-based literature.

Thus, based on combining all of the results provided above we propose the new test statistic taking the form

$$V_n = \min_{a(n) \leq m \leq b(n)} \prod_{j=1}^n \frac{2m(1 - (m+1)(2n)^{-1})}{n\Delta_{jm}}, a(n) = n^{0.5+\delta}, b(n) = \min(n^{1-\delta}, n/2), \delta \in (0, 1/4) \quad (7)$$

(here, in a similar manner to Canner (1975), we will arbitrarily define $\Delta_{jm} = 1/n$, if $\Delta_{jm} = 0$).

The V_n -test of the hypothesis (1) is a test with critical region

$$\log(V_n) > C, \quad (8)$$

where C is a test-threshold. More specially, in this article we arrive at the practical value of $\delta = 0.1$. It is shown in Section 3 within our extensive Monte-Carlo study that the power of the test statistic at (7) does not differ substantially for values of $\delta \in (0, 1/4)$ given various alternatives.

Remark. Since the test statistic at (7) approximates the optimal parametric likelihood ratio defined at (2) the test based on V_n can be anticipated to be very efficient. This is illustrated in the context of our simulation study.

Significance level of the proposed test. It turns out that the null distribution of the test statistic V_n is independent of the distribution of the observations $Z_1, \dots, Z_n$. In order to substantiate this claim we note that under H_0

$$I(Z_i \leq Z_j) = I(\Phi^{-1}(F_{H_0}(Z_i)) \leq \Phi^{-1}(F_{H_0}(Z_j))), \text{ and}$$

$$\begin{aligned} I(-Z_i \leq Z_j) &= I(\Phi^{-1}(F_{H_0}(-Z_i)) \leq \Phi^{-1}(F_{H_0}(Z_j))) \\ &= I(\Phi^{-1}(1 - F_{H_0}(Z_i)) \leq \Phi^{-1}(F_{H_0}(Z_j))) \\ &= I(-\Phi^{-1}(F_{H_0}(Z_i)) \leq \Phi^{-1}(F_{H_0}(Z_j))), \text{ for } i \neq j \in [1, n], \end{aligned}$$

where $\Phi^{-1}(x)$ denotes the inverse function of the standard normal cumulative distribution function $\Phi(x)$. This fact implies, via the construction of the test statistic V_n that

$$P_{H_0} \{\log(V_n) > C\} = P_{X_1, \dots, X_n \sim \text{Norm}(0,1)} \{\log(V_n) > C\}.$$

From this result, it is clear that the proposed test is exact and the corresponding critical values can be tabulated for fixed sample sizes. Certain lines of research have developed around the asymptotic distribution problems involving Vasicek's entropy estimator and different entropy-based statistics (e.g., Cressie, 1976; Dudewicz and van der Meulen, 1981; Hall, 1984). It is generally recognized that the asymptotic distribution of the test statistic $\log(V_n)$, which includes the estimates of distribution functions, is analytically difficult. Practical applications motivated us to consider critical values for fixed sample sizes. Thus, in this article and following the recent literature related to nonparametric tests (e.g., Canner, 1975; Mudholkar and Tian, 2002, 2004), we will not attempt to provide an analytical solution for the critical values for the test (8), tabulating the critical values for fixed sample sizes using an alternative Monte Carlo approach. The definition at (7) includes $\delta \in (0, 0.25)$. One can show that the power of the new test does not depend significantly on values of δ under various scenarios of alternative distributions applied to the hypotheses at (1). The computations involved in such conclusions were executed via an extensive Monte Carlo study. Section 3 displays a part of results related to the relevant Monte Carlo simulations. In this article we focus on $\delta = 0.1$ in the definition (7). In order to tabulate the percentiles of the null distribution of $\log(V_n)$ with $\delta = 0.1$, we drew 75,000 replicate samples of the test statistic $\log(V_n)$ at each sample size n . In this experiment, data were generated from the $N(0, 1)$ distribution. The generated values of the test statistic $\log(V_n)$ were used to determine the critical values C_α : $P_{H_0}\{\log(V_n) > C_\alpha\} = \alpha$ of the null distribution of $\log(V_n)$ at the significance level α . The results of the experiment are presented in Table 1.

Consistency. The consistency of the proposed test defined at (8) is stated in the following proposition:

Proposition 1. Let $f(x)$ define a density function of the observations $Z_1, \dots, Z_n$ with the finite expectations $E(\log f(Z_1))$ and $E(\log(0.5 f(Z_1) + 0.5 f(-Z)))$. Then

$$\frac{1}{n} \log(V_n) \xrightarrow{p} -E \left(\log \left(0.5 + 0.5 \frac{f(-Z_1)}{f(Z_1)} \right) \right), \text{ as } n \rightarrow \infty,$$

for all $0 < \delta < 1/4$ in the definition (7) of the statistic V_n .

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

It is obvious that the limiting value of $n^{-1} \log(V_n)$, the expectation

$$-E \left(\log \left(0.5 + 0.5 f(-Z_1) f(Z_1)^{-1} \right) \right)$$

stated in Proposition 1, has the forms of $-E_{H_0} \log(0.5 + 0.5 f_{H_0}(-Z_1) (f_{H_0}(Z_1))^{-1}) = 0$ and $-E_{H_1} (\log(0.5 + 0.5 f_{H_1}(-Z_1) (f_{H_1}(Z_1))^{-1})) - \log(0.5 + 0.5 E_{H_1} (f_{H_1}(-Z_1) (f_{H_1}(Z_1))^{-1})) = 0$, under H_0 and H_1 respectively. This implies that the test statistic $\log(V_n)$ is consistent.

3. Simulation Study

An extensive simulation study was carried out in order to evaluate the performance of the new test as compared to the classical procedures. The Monte Carlo experiments for investigating the power properties of the tests were repeated 25,000 times. Critical values of the tests were set at the 5% level of significance. The alternatives used in the Monte Carlo evaluations were chosen to represent the following situations: 1) constant shifts in location, e. g., $(X \sim N(0, 1), Y \sim N(0.5, 1))$, $(X - Y \sim U[-0.3, 0.7])$; 2) constant and nonconstant shifts, e. g., $(X \sim N(0, 1), Y \sim N(0.5, 2^2))$; 3) nonconstant shifts, e. g., $(X \sim \text{Beta}(0.7, 1), Y \sim \text{Exp}(2))$; 4) skewed alternatives, e. g., $(X \sim \text{LogN}(1, 1), Y \sim \text{LogN}(1, 0.5^2))$; 5) tail alternatives (Albers, Kallenberg and Martini, 2001). Note that in the case of situation (1) the literature anticipates that the two-sided Wilcoxon signed-rank test and the two-sided paired t-test will be very efficient procedures. As noted earlier, a constant shift in location does not necessarily correspond to cases of practical applications (e. g., Albers *et al.*, 2001.)

It can be seen by examination of Tables 2 and 3 that the power of the new test is robust to the choice of δ from definition (7). Thus, for practical use we recommend a value of $\delta = 0.1$.

When a constant shift is in effect, the new test demonstrates a relatively small power loss in comparison to Wilcoxon's signed-rank test and the paired t-test based given normally distributed observations. Generally speaking these cases are standard with respect to demonstrating the efficiency properties of classical tests (here the t-test is expected to be most powerful). However, when a constant shift is in effect and samples do not follow normal distributions, the proposed test can be superior to the classical procedures, e. g., the situations with uniformly distributed alternatives as shown in Table 3. The new test clearly shows high and stable power as compared to the classical procedures, e. g., values of the power of the proposed test with $\delta = 0.1$ at (7) vs those of Wilcoxon's test are 0.999 vs 0.263 (the case of $X \sim \chi^2_{df=6}, Y \sim \text{LogN}(0.3^2), n=50$); 0.293 vs 0.078 (the case of $X \sim \text{Beta}(0.7, 1), Y \sim \text{Exp}(2), n=50$); 0.464 vs 0.128 (the case of $X \sim \text{LogN}(1, 1), Y \sim \text{LogN}(1, 0.5^2), n=50$); 0.966 vs 0.257 (the case of $X \sim U(1, 2), Y \sim \text{LogN}(0, 1), n=50$). Thus, the proposed method is a powerful way to carry out testing by using paired data.

4. Practical Data Examples

We exemplify the proposed method with data related to the two biomedical studies.

4.1 A Quantitative Assay Problem

Hawkins (2002) presented measurements of 56 assay pairs of cyclosporine. Data points were derived by utilizing 56 blood samples from organ transplant recipients and assaying an aliquot of each sample by a standard method called “High Performance Liquid Chromatography” (HPLC) and an alternative technique, “Radio-Immunoassay” (RIA). The research focus of Hawkins (2002) was to show that the new clinical assay via RIA (the Y) is “substantially equivalent” to the approved assay via HPLC (the X). Utilizing values of X and Y depicted in Hawkins (2002), we obtain p-values corresponding to two-sided Wilcoxon signed-rank test and the two-sided paired t-test of p-value=0.6566 and p-value=0.6364, respectively. In this case, the proposed test statistic $\log(V_n)$ with $\delta=0.1$ has a value of 1.7941, which corresponds to a p-value>0.4. This supports the conclusion originally obtained by Hawkins (2002) indicating that the HPLC and RIA methods are not statistically significantly different.

4.2. A Group-based Therapy for Children with Attention-Deficit/Hyperactivity Disorder (ADHD) and Severe Mood Dysregulation (SMD)

The following study was carried out at the Center for Children and Families, University at Buffalo, The State University of New York. To examine the feasibility and efficacy of a group-based therapy program for children with ADHD and SMD, 17 children ages 7 to 12 with ADHD and SMD were assigned to receive the experimental 11 week group therapy program. The group based therapy program consisted of eleven 90-minute sessions with concurrent parent and child groups. Assessments were completed at 2 time points: Baseline (Week 0) and Endpoint (Week 11). Clinician ratings included the Children’s Depression Rating Scale – Revised total score (CDRS-Rts), see, e.g., Leibenluft et al. (2003). Write $X_1, \dots, X_{n=17}$ and $Y_1, \dots, Y_{n=17}$ for the CDRS-Rts observations obtained with respect to Baseline and Endpoint, respectively. The empirical histograms of $X_1, \dots, X_n$, $Y_1, \dots, Y_n$ and $Z_1 = X_1 - Y_1, \dots, Z_n = X_n - Y_n$ corresponding to this dataset are depicted in Figure 1. The differences in location between the two groups, as indicated by the empirical density estimates seen in Figure 1, was confirmed by the two-sided Wilcoxon signed-rank test (p-value=0.00071). The corresponding value for the test statistic at (8) with $\delta=0.1$ had a value 10.828 with corresponding p-value=0.000055. These results show that the proposed testing procedure is coherent with the classic Wilcoxon signed-rank test in terms of detecting location differences, thus demonstrating that the density-based EL test can be utilized in this type of inference.

In addition to our illustration, a bootstrap type study was conducted to examine the proposed test. In accordance with ideas introduced by Stigler (1977), since the considered tests based on the full dataset reject the null hypothesis, indicating a strong evidence of the treatment effects, the rejection rate of the null hypothesis for a test based on samples (with relatively small sizes) from the data can be examined in the context of the efficiency (i.e., power) of the test. Thus, the following procedures were executed. From the dataset, samples with the sizes 5, 10, 11, 12 and 15 were randomly selected, in order to be tested for (1) at a 5% level of significance. We repeated this strategy 30,000 times calculating the frequencies of the events $\log(V_5) > C_{0.05} = 3.514$, $\log(V_{10}) > 2.601$, $\log(V_{11}) > 2.535$, $\log(V_{12}) > 2.592$, and $\log(V_{15}) > 2.541$. The new test did not reject the null hypothesis at (1) in 11145 (the case of $n=5$), 0 ($n=10$), 0 ($n=11$), 0 ($n=12$) and 0 ($n=15$) events, respectively, while the Wilcoxon signed-rank test did not reject the null hypothesis in 18573, 1696, 362, 3 and 0 events, respectively. For example, consider a subsample ($n=11$) displayed in Figure 2. In this case, the Wilcoxon signed-rank test breaks down completely (p-value= 0.1050), detecting a change in distributions of CDRS scores. Thus, this indicates that our proposed method is more sensitive and reliable as compared to the classical test in this real data study. The

density-based EL ratio test can be recommended to be applied as a primary statistical tool in ADHD and SMD investigations.

5. Concluding Remarks

The newly proposed method described in this article employs the EL concept in a nonparametric fashion in order to approximate a Neyman-Pearson type parametric likelihood ratio test-statistic given paired data. The benefit of using this approach is threefold. First, we are able to construct an exact and robust nonparametric test. Second, the proposed technique is highly efficient given that it approximates well the optimal parametric likelihood ratio. Third, the new test has consistently high power over a variety of alternative distributions, resulting in a large power gain in comparison to the classical procedures. We illustrated that the new method is easily applied to real-life data and provides meaningful results. Finally, it is important to note that for future work the method developed in this article can be modified for a variety of other nonparametric problems. For example, a one-sided version of the two-sided problem (1) may be a point of interest, e.. say $X \stackrel{st}{>} Y$ under H_1 , here the notation $X \stackrel{st}{>} Y$ can be formalized admitting nonconstant shifts in location (e.g., Albers *et al.* (2001)). In this case, it turns out that the definition

$$\hat{F}_{H_0}(u) = \max \frac{1}{n} \left[\sum_{i=1}^n I(Z_i \leq u), \sum_{i=1}^n I(-Z_i \leq u) \right] \text{ applied instead of}$$

$\hat{F}_{H_0}(u) = \frac{1}{2n} \left[\sum_{i=1}^n I(Z_i \leq u) + \sum_{i=1}^n I(-Z_i \leq u) \right]$ to (4) (where $\Delta_{jm} := \hat{F}_{H_0}(Z_{(j+m)}) - \hat{F}_{H_0}(Z_{(j-m)})$) significantly improves the power of the test. For example, the Monte Carlo powers of the modified test and the one-sided Wilcoxon rank-sum were calculated as 0.92168 and 0.81252 at 5% level of significance and based on 25,000 samples of $Z_1, \dots, Z_{n=25} \sim U[-0.5, 1]$. This is one of topic that is subject to further investigation.

Acknowledgments

The authors are indebted to Professor Govind S. Mudholkar for many helpful discussions and comments. We also would like to acknowledge Professor Bradley Efron for providing valuable commentaries. This research is supported by NIH grant 1R03DE020851 - 01A1 (the National Institute of Dental and Craniofacial Research). Dr. Gurevich was supported by the Internal Funding Program of the Shamoon College of Engineering (SCE). The authors are grateful to the Editor and the referee for suggestions that led to a substantial improvement in this paper.

Appendix A: Constructing the test statistic (6)

Let M define a set of appropriate values of the integer parameter m in (5). Assume there

exists an integer $r \in M$ such that $(1/2r) \sum_{j=1}^n \int_{Z_{(j-r)}}^{Z_{(j+r)}} f_{H_1}(u) du \geq 1$. In this case, by virtue of

(3), we conclude $\int_{Z_{(1)}}^{Z_{(n)}} f_{H_1}(u) du \geq 1$, which is unacceptable. Then, it is natural to require that $f_{H_1,j}$, $j = 1, \dots, n$ are subject to (4), for all possible $m \in M$. Therefore, for a fixed integer $m_0 \in M$, the approximation to the H_f -likelihood $\prod_{j=1}^n f_{H_1,j}$ complies with

$$\max_{f_{H_1,1}, \dots, f_{H_1,n} \text{ subject to (4), for all } m \in M} \prod_{j=1}^n f_{H_1,j} \leq \max_{f_{H_1,1}, \dots, f_{H_1,n} \text{ subject to (4) with } m=m_0 \in M} \prod_{j=1}^n f_{H_1,j}.$$

Thus,

$$\min_{m_0} \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4), for all } m \in M} \prod_{j=1}^n f_{H_{1,j}} \leq \min_{m_0} \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4) with } m=m_0 \in M} \prod_{j=1}^n f_{H_{1,j}},$$

and hence we can write

$$\max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4), for all } m \in M} \prod_{j=1}^n f_{H_{1,j}} \leq \min_{m_0} \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4) with } m=m_0 \in M} \prod_{j=1}^n f_{H_{1,j}}. \quad (\text{A.1})$$

(Here, if $a > b$, then, $\min(a) > \min(b)$.)

However, because the constraint (4) is an approximation to (3), assuming $f_{H_{1,j}} j = 1, \dots, n$ satisfy (4) with $m=r$, we have

$$\begin{aligned} 1 - \frac{1}{n} &= \frac{1}{2r} \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} \Delta_{jr} + \frac{(r-1)}{2n} \approx \frac{1}{2r} \sum_{j=1}^n \int_{Z_{(j-r)}}^{Z_{(j+r)}} f_{H_1}(u) du + \sum_{l=1}^{r-1} \frac{(r-l)}{2r} \left(\int_{Z_{(n-l)}}^{Z_{(n-l+1)}} f_{H_1}(u) du + \int_{Z_{(l)}}^{Z_{(l+1)}} f_{H_1}(u) du \right) \\ &= \int_{Z_{(1)}}^{Z_{(n)}} f_{H_1}(u) du \\ &= \frac{1}{2k} \sum_{j=1}^n \int_{Z_{(k-r)}}^{Z_{(k+r)}} f_{H_1}(u) du + \sum_{l=1}^{k-1} \frac{(k-l)}{2k} \left(\int_{Z_{(n-l)}}^{Z_{(n-l+1)}} f_{H_1}(u) du + \int_{Z_{(l)}}^{Z_{(l+1)}} f_{H_1}(u) du \right) \\ &\approx \frac{1}{2k} \sum_{j=1}^n \frac{f_{H_{1,j}}}{f_{H_{0,j}}} \Delta_{jk} + \frac{(k-1)}{2n}, \text{ for } k \leq \frac{n}{2}, \end{aligned}$$

i.e., $f_{H_{1,j}} j = 1, \dots, n$ are expected to satisfy (4), for $m=k$ too. This leads us to expect that, for some m_0 ,

$$\begin{aligned} \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4), for all } m \in M} \prod_{j=1}^n f_{H_{1,j}} &\approx \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4) with } m=m_0 \in M} \prod_{j=1}^n f_{H_{1,j}} \\ &\geq \min_{m_0} \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4) with } m=m_0 \in M} \prod_{j=1}^n f_{H_{1,j}}. \end{aligned} \quad (\text{A.2})$$

The considerations (A.1) and (A.2) imply that the maximum EL should be defined as

$$\min_{m_0 \in M} \max_{f_{H_{1,1}}, \dots, f_{H_{1,n}} \text{ subject to (4) with } m=m_0} \prod_{j=1}^n f_{H_{1,j}}.$$

Appendix B: Proof of Proposition 1

In order to prove Proposition 1, we applied some reorganization of definitions to follow proof schemes utilized in Vasicek (1976). By virtue of the definition (5), we have

$$\frac{1}{n} \log(V_{nm}) = \frac{1}{n} \log \prod_{j=1}^n \frac{2m(1 - (m+1)(2n)^{-1})}{n\Delta_{jm}} = \frac{1}{n} \log \prod_{j=1}^n \frac{2m}{n\Delta_{jm}} + \log \left(1 - \frac{m+1}{2n}\right). \quad (\text{B.1})$$

Defining

$G_n(x) = (F_n(x) + F_n^*(x)) / 2$, where $F_n(x) = n^{-1} \sum_{i=1}^n I(Z_i \leq x)$ and $F_n^*(x) = n^{-1} \sum_{i=1}^n I(-Z_i \leq x)$ are the empirical distribution functions based on $Z_1, \dots, Z_n$ and $-Z_1, \dots, -Z_n$ respectively, we use the definition of Δ_{jm} in (4) to present

$$\frac{1}{n} \log \prod_{j=1}^n \frac{2m}{n\Delta_{jm}} = -\frac{1}{n} \sum_{i=1}^n \log \frac{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})}{(2m/n)}. \quad (\text{B.2})$$

Therefore,

$$\frac{1}{n} \log \prod_{j=1}^n \frac{2m}{n\Delta_{jm}} = -\frac{1}{n} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{F(Z_{(i+m)}) - F(Z_{(i-m)})} + \frac{1}{n} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})} - \frac{1}{n} \sum_{i=1}^n \log \frac{F(Z_{(i+m)}) - F(Z_{(i-m)})}{(2m/n)}, \quad (\text{B.3})$$

where the distribution functions $G(x) = 0.5F(x) + 0.5F^*(x)$, $F^*(x) = 1 - F(-x)$ and

$$\sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{F(Z_{(i+m)}) - F(Z_{(i-m)})} = \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{Z_{(i+m)} - Z_{(i-m)}} - \sum_{i=1}^n \log \frac{F(Z_{(i+m)}) - F(Z_{(i-m)})}{Z_{(i+m)} - Z_{(i-m)}}. \quad (\text{B.4})$$

In a similar manner to the reorganization applied to prove Theorem 1 of Vasicek (1976), we record

$$n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{Z_{(i+m)} - Z_{(i-m)}} = (2m)^{-1} \sum_{j=1}^{2m} S_j,$$

$$S_j = \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{Z_{(i+m)} - Z_{(i-m)}} \{F_n(Z_{(i+m)}) - F_n(Z_{(i-m)})\}, i \equiv j \pmod{2m},$$

and note that when $Z_{(i-m)}, Z_{(i+m)}$ belong to an interval in which $g(x) = dG(x)/dx = 0.5f(x) + 0.5f(-x)$ is positive and continuous, then

$(G(Z_{(i+m)}) - G(Z_{(i-m)}))(Z_{(i+m)} - Z_{(i-m)})^{-1} = g(Z_i^*)$, for some existing value

$Z_i^* \in (Z_{(i-m)}, Z_{(i+m)})$. Therefore, S_j is a Stieltjes sum of the function $\log(g(x))$ with respect to the measure F_n over the sum of intervals of continuity of g in which $g(x) > 0$. The contribution of terms in S_j that corresponds to interval in which $g(x) = 0$ approaches zero with $n \rightarrow \infty$. Since in any interval in which $g(x)$ is positive $Z_{(i+m)} - Z_{(i-m)} \rightarrow 0$ as $n \rightarrow \infty$ uniformly over $m \in [n^{0.5+\delta}, n^{1-\delta}]$ and uniformly over x , S_j converges in probability to

$\int_{-\infty}^{\infty} \log(g(x)) dF(x) = E(\log(g(Z_1)))$. Moreover, this convergence is uniform over $n^{0.5+\delta}$ and $n^{1-\delta}$ and j . Consequently,

$$(2m)^{-1} \sum_{j=1}^{2m} S_j \xrightarrow{P} \int_{-\infty}^{\infty} \log(g(x)) dF(x) = E(\log(g(Z_1))),$$

as $n \rightarrow \infty$, uniformly over $n^{0.5+\delta} \leq m \leq n^{1-\delta}$. This implies

$$n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{Z_{(i+m)} - Z_{(i-m)}} \xrightarrow{P} E(\log(g(Z_1))) \quad (\text{B.5})$$

uniformly over $n^{0.5+\delta} \leq m \leq n^{1-\delta}$. Now, it is also clear that

$$n^{-1} \sum_{i=1}^n \log \frac{F(Z_{(i+m)}) - F(Z_{(i-m)})}{Z_{(i+m)} - Z_{(i-m)}} \xrightarrow{P} E(\log(f(Z_1))) \quad (\text{B.6})$$

uniformly over $n^{0.5+\delta} \leq m \leq n^{1-\delta}$. By (B.4), (B.5) and (B.6),

$$\frac{1}{n} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{F(Z_{(i+m)}) - F(Z_{(i-m)})} \xrightarrow{P} E \log \frac{g(Z_1)}{f(Z_1)} = E \log \left(\frac{1}{2} + \frac{1}{2} \frac{f(-Z_1)}{f(Z_1)} \right), \quad (\text{B.7})$$

as $n \rightarrow \infty$, uniformly over $n^{0.5+\delta} \leq m \leq n^{1-\delta}$.

Consider the term $n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})}$ of (B.3). The Kolmogorov Theorem (e.g., Serfling, 1980, p 62) provides

$$\Pr \left(\sup_{-\infty < x < \infty} |F(x) - F_n(x)| > n^{-0.5+\varepsilon} \right) \rightarrow 0, \quad \Pr \left(\sup_{-\infty < x < \infty} |F^*(x) - F_n^*(x)| > n^{-0.5+\varepsilon} \right) \rightarrow 0, \quad \varepsilon \in (0, \delta/2),$$

and hence $\Pr \left(\sup_{-\infty < x < \infty} |G(x) - G_n(x)| > n^{-0.5+\varepsilon} \right) \rightarrow 0$, as $n \rightarrow \infty$, $\varepsilon \in (0, \delta/2)$. Therefore, we can consider the term

$n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})}$ when $\sup_{-\infty < x < \infty} |G(x) - G_n(x)| \leq n^{-0.5+\varepsilon}$. By virtue of the definition of $G_n(x)$, we have the inequality $G_n(Z_{(i+m)}) - G_n(Z_{(i-m)}) \geq m/n$. This, taking into account the condition $\sup_{-\infty < x < \infty} |G(x) - G_n(x)| \leq n^{-0.5+\varepsilon}$, implies that, for sufficiently large n and $m \in [n^{0.5+\delta}, n^{1-\delta}]$,

$$\begin{aligned} n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})} &\leq n^{-1} \sum_{i=1}^n \log \frac{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)}) + 2n^{-0.5+\delta/2}}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})} \leq n^{-1} \sum_{i=1}^n \log \left(1 + \frac{2n^{-0.5+\delta/2}}{m/n} \right) \leq n^{-1} \sum_{i=1}^n \frac{2n^{-0.5+\delta/2}}{m/n} \\ &= \frac{2}{n^{\delta/2}} \xrightarrow{n \rightarrow \infty} 0; \end{aligned}$$

$$\begin{aligned}
n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})} &\geq n^{-1} \sum_{i=1}^n \log \frac{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)}) + 2n^{-0.5+\delta/2}}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})} \geq n^{-1} \sum_{i=1}^n \log \left(1 - \frac{2n^{-0.5+\delta/2}}{m/n} \right) \geq \\
&- n^{-1} \sum_{i=1}^n \frac{2n^{0.5+\delta/2}}{n^{0.5+\delta}} \\
&= - \frac{2}{n^{\delta/2}} \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

Thus, we conclude

$$n^{-1} \sum_{i=1}^n \log \frac{G(Z_{(i+m)}) - G(Z_{(i-m)})}{G_n(Z_{(i+m)}) - G_n(Z_{(i-m)})} \xrightarrow[n \rightarrow \infty]{p} 0 \quad (\text{B.8})$$

uniformly over $n^{0.5+\delta} \leq m \leq n^{1-\delta}$.

We complete the proof of Proposition 1 applying directly (B.7), (B.8) and the result

$$0 \leq -\frac{1}{n} \sum_{i=1}^n \log \frac{F(Z_{(i+m)}) - F(Z_{(i-m)})}{(2m/n)} \xrightarrow[n \rightarrow \infty]{p} 0 \text{ uniformly, for all } m \in [n^{0.5+\delta}, n^{1-\delta}]$$

of Vasicek (1976) to the equation (B.3).

References

- Albers W, Kallenberg WCM, Martini F. Data-Driven Rank Tests for Classes of Tail Alternatives. *Journal of the American Statistical Association*. 2001; 96:685–696.
- Arizono I, Ohta H. A Test for Normality Based on Kullback-Leibler Information. *The American Statistician*. 1989; 43:20–22.
- Bhattacharya PK, Gastwirth JL, Wright AL. Two Modified Wilcoxon Tests for Symmetry about an Unknown Location Parameter. *Biometrika*. 1982; 69:377–382.
- Canner PL. A simulation study of one-and two-sample Kolmogorov-Smirnov statistics with a particular weight function. *Journal of the American Statistical Association*. 1975; 70:209–211.
- Claeskens G, Hjort NL. Goodness of fit via non-parametric likelihood ratios. *Scandinavian Journal of Statistics*. 2004; 31:487–513.
- Cressie N. On the Logarithms of High-Order Spacings. *Biometrika*. 1976; 63:343–355.
- Dudewicz EJ, Van Der Meulen EC. Entropy-Based Tests of Uniformity. *Journal of the American Statistical Association*. 1981; 76:967–974.
- Fan J, Farnen M, Gijbels I. Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B*. 1998; 60:591–608.
- Fellingham SA, Stoker DJ. An Approximation for the Exact Distribution of the Wilcoxon Test for Symmetry. *Journal of the American Statistical Association*. 1964; 59:899–905.
- Hall P. Limit theorems for sums of general functions of m-spacings. *Mathematical Proceedings of the Cambridge Philosophical Society*. 1984; 96:517–532.
- Hawkins DM. Diagnostics for Conformity of Paired Quantitative Measurements. *Statistics in Medicine*. 2002; 21:1913–1935. [PubMed: 12111897]
- Lam FC, Longnecker MT. A Modified Wilcoxon rank-sum test for paired data. *Biometrika*. 1983; 70:510–513.
- Lazar NA. Bayesian Empirical Likelihood. *Biometrika*. 2003; 90:319–326.
- Lehmann, EL.; Romano, JP. Testing statistical hypotheses. New York: Springer; 2005.

- Leibenluft E, Chaney D, Towbin KE, Bhangoo RK, Pine DS. Defining clinical phenotypes of juvenile mania. *American Journal of Psychiatry*. 2003; 160(3):430–437. [PubMed: 12611821]
- Limpert E, Stahel W, Abbt M. Log-Normal Distributions Across the Sciences: Keys and Clues. *BioScience*. 2001; 51:341–352.
- Mudholkar GS, Tian L. An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test. *Journal of Statistical Planning and Inference*. 2002; 102:211–221.
- Mudholkar GS, Tian L. A test for homogeneity of ordered means of inverse Gaussian population. *Journal of Statistical Planning and Inference*. 2004; 118:37–49.
- Owen, AB. Empirical likelihood. New York: Chapman and Hall/CRC; 2001.
- Qin J, Lawless J. Empirical Likelihood and General Estimating Equations. *The Annals of Statistics*. 1994; 22:300–325.
- Schuster EF. Estimating the distribution function of a symmetric distribution. *Biometrika*. 1975; 62:631–635.
- Serfling, RJ. Approximation Theorems of Mathematical Statistics. Wiley: 1980.
- Stigler SM. Do robust estimators work with real data? *The Annals of Statistics*. 1977; 5:1055–1098.
- Tusnady G. On Asymptotically Optimal Tests. *The Annals of Statistics*. 1977; 5:385–393.
- Vasicek O. A Test for Normality Based on Sample Entropy. *Journal of Royal Statistical Society, B*. 1976; 38:54–59.
- Vexler A, Gurevich G. Empirical Likelihood Ratios Applied to Goodness-Of-Fit Tests Based on Sample Entropy. *Computational Statistics and Data Analysis*. 2010; 54:531–545.
- Vexler A, Liu S, Kang L, Hutson AD. Modifications of the Empirical Likelihood Interval Estimation with Improved Coverage Probabilities. *Communications in Statistics (Simulation and Computation)*. 2009; 38:2171–2183.
- Vexler A, Wu C. An optimal retrospective change point detection policy. *Scandinavian Journal of Statistics*. 2009; 36:542–558.
- Vexler A, Wu C, Yu KF. Optimal Hypothesis Testing: From Semi to Fully Bayes Factors. *Metrika*. 2010a; 71:125–138. [PubMed: 20198107]
- Vexler A, Yu J, Tian L, Liu S. Two-sample nonparametric likelihood inference based on incomplete data with an application to a pneumonia study. *Biometrical Journal*. 2010b; 52:348–361. [PubMed: 20533413]
- Vexler A, Shan G, Kim S, Tsai WM, Tian L, Hutson AD. An empirical likelihood ratio based goodness-of-fit test for Inverse Gaussian distributions. *Journal of Statistical Planning and Inference*. 2011; 141:2128–2140.
- Vexler A, Yu J. Two-sample density-based empirical likelihood tests for incomplete data in application to a pneumonia study. *Biometrical Journal*. 2011; 53:628–651. [PubMed: 21647935]
- Yu J, Vexler A, Tian L. Analyzing Incomplete Data Subject to a Threshold using Empirical Likelihood Methods: An Application to a Pneumonia Risk Study in an ICU Setting. *Biometrics*. 2010; 66:123–130. [PubMed: 19432776]
- Wang J. Quadratic Artificial Likelihood Functions Using Estimating Functions. *Scandinavian Journal of statistics*. 2006; 33:379–380.
- Wedderburn RWM. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*. 1974; 61:439–447.
- Wilcoxon F. Individual comparisons by ranking methods. *Biometrics*. 1945; 1:80–83.

Figure 1. Histograms of Baseline (the X) and Endpoint (the Y) values of the CDRS-Rts and histogram of $Z=X-Y$ based on the data set ($n=17$).

Figure 2.
Histograms based on the sample ($n=11$) from the data set.

Table 1

The critical values, $\log(V_n)$, of the proposed test (8) with $\delta = 0.1$ at the significance level α .

n/α	0.010	0.015	0.020	0.025	0.040	0.050	0.060	0.100	0.150	0.200	0.300
5	3.514	3.514	3.514	3.514	3.514	3.514	3.514	2.557	1.499	1.276	0.855
10	4.682	4.342	3.728	3.507	2.924	2.601	2.469	1.957	1.638	1.403	1.153
15	4.087	3.706	3.390	3.174	2.744	2.541	2.401	1.979	1.676	1.473	1.159
20	4.107	3.729	3.456	3.254	2.867	2.681	2.531	2.149	1.834	1.608	1.291
25	4.234	3.850	3.625	3.421	3.017	2.833	2.689	2.278	1.976	1.737	1.428
30	4.160	3.790	3.558	3.376	2.981	2.808	2.666	2.293	1.997	1.777	1.471
35	4.137	3.797	3.550	3.377	3.005	2.850	2.713	2.338	2.054	1.835	1.535
40	4.249	3.915	3.704	3.539	3.172	2.992	2.849	2.462	2.178	1.956	1.656
45	4.404	4.068	3.834	3.652	3.266	3.096	2.963	2.580	2.302	2.072	1.768
50	4.401	4.084	3.855	3.692	3.329	3.161	3.013	2.627	2.357	2.139	1.840
55	4.468	4.158	3.923	3.742	3.365	3.196	3.057	2.686	2.415	2.205	1.916
60	4.547	4.228	4.007	3.834	3.477	3.319	3.187	2.816	2.543	2.328	2.031
65	4.688	4.355	4.144	3.974	3.597	3.436	3.302	2.934	2.650	2.430	2.135
70	4.611	4.355	4.155	3.984	3.639	3.491	3.347	2.992	2.722	2.508	2.219
80	4.897	4.576	4.360	4.187	3.842	3.686	3.552	3.178	2.906	2.691	2.402
90	5.045	4.711	4.524	4.365	4.013	3.861	3.722	3.353	3.073	2.857	2.566
100	5.239	4.914	4.692	4.521	4.177	4.020	3.876	3.516	3.233	3.028	2.738
110	5.384	5.073	4.870	4.700	4.339	4.186	4.049	3.679	3.397	3.196	2.899
120	5.573	5.264	5.037	4.852	4.495	4.330	4.195	3.845	3.565	3.365	3.069
135	5.792	5.478	5.266	5.086	4.738	4.573	4.433	4.062	3.769	3.574	3.281
150	6.033	5.694	5.469	5.314	4.972	4.803	4.661	4.297	4.000	3.795	3.501
175	6.336	6.041	5.830	5.671	5.329	5.165	5.033	4.649	4.359	4.140	3.836
200	6.763	6.447	6.237	6.058	5.708	5.539	5.394	5.008	4.703	4.496	4.185
250	7.414	7.071	6.850	6.671	6.311	6.139	6.000	5.620	5.312	5.075	4.731

Table 2

Monte Carlo Powers of the test (8) with different values of δ , two-sided Wilcoxon signed rank test and paired Student's t-test; at $\alpha=0.05$.

F_X	F_Y	n	δ				Proposed test (8)	Wilcoxon test	t-test
			0.025	0.05	0.10	0.15			
N(0,1)	N(0.5,1)	15	0.239	0.239	0.238	0.238	0.236	0.249	
		25	0.359	0.359	0.359	0.357	0.376	0.394	
		50	0.605	0.605	0.605	0.605	0.659	0.687	
N(0,1)	N(0.5,2 ²)	15	0.124	0.124	0.124	0.124	0.121	0.127	
		25	0.169	0.169	0.169	0.169	0.178	0.187	
		50	0.285	0.286	0.287	0.287	0.319	0.338	
N(0,1)	Cauchy(1,1)	15	0.297	0.297	0.297	0.297	0.310	0.165	
		25	0.469	0.469	0.468	0.464	0.481	0.184	
		50	0.788	0.788	0.790	0.790	0.769	0.193	
Exp(1)	LogN(0,2 ²)	15	0.285	0.285	0.281	0.281	0.221	0.121	
		25	0.552	0.552	0.548	0.535	0.367	0.307	
		50	0.938	0.937	0.935	0.933	0.655	0.668	
$\chi^2_{df=6}$	LogN(0,3 ²)	15	0.339	0.339	0.327	0.327	0.129	0.059	
		25	0.818	0.818	0.813	0.816	0.174	0.035	
		50	0.999	0.999	0.999	0.999	0.263	0.096	
Gamma(5,1)	Gamma(1,5)	15	0.090	0.090	0.089	0.089	0.080	0.068	
		25	0.165	0.165	0.164	0.160	0.104	0.064	
		50	0.449	0.448	0.445	0.441	0.150	0.060	
Beta(0.7,1)	Exp(2)	15	0.061	0.061	0.060	0.060	0.053	0.055	
		25	0.089	0.089	0.088	0.089	0.063	0.079	
		50	0.297	0.296	0.293	0.287	0.078	0.150	
LogN(1,1)	LogN(1,0.5 ²)	15	0.085	0.085	0.083	0.083	0.062	0.059	
		25	0.175	0.175	0.172	0.164	0.084	0.127	
		50	0.471	0.469	0.464	0.457	0.128	0.350	
N(0,1)	U[-1,2]	15	0.261	0.261	0.260	0.260	0.252	0.273	
		25	0.402	0.402	0.402	0.400	0.414	0.442	
		50	0.656	0.656	0.657	0.656	0.704	0.748	
U[1,2]	LogN(0,1)	15	0.179	0.179	0.173	0.173	0.117	0.086	
		25	0.500	0.500	0.496	0.485	0.164	0.068	
		50	0.967	0.966	0.966	0.965	0.257	0.053	

Table 3

Monte Carlo Powers of the test (8) with different values of δ , two-sided Wilcoxon signed rank test and paired Student's t-test; at $\alpha=0.05$.

	<i>n</i>	δ					Wilcoxon test	t-test
		0.025	0.05	0.10	0.15			
Increasing tail alternative A	15	0.218	0.218	0.216	0.216		0.167	0.222
	25	0.459	0.459	0.454	0.442		0.274	0.495
	50	0.898	0.897	0.895	0.891		0.517	0.883
Increasing tail alternative B	15	0.582	0.582	0.577	0.577		0.500	0.611
	25	0.862	0.862	0.860	0.854		0.765	0.892
	50	0.995	0.995	0.995	0.995		0.969	0.997
$Z=X \cdot Y \sim U[-0.3,0.7]$	15	0.714	0.714	0.710	0.710		0.618	0.696
	25	0.963	0.963	0.961	0.959		0.866	0.923
	50	1.000	1.000	1.000	1.000		0.992	0.999
$Z=X \cdot Y \sim U[-0.5,1]$	15	0.559	0.559	0.555	0.555		0.465	0.526
	25	0.885	0.885	0.882	0.877		0.726	0.794
	50	1.000	1.000	1.000	1.000		0.956	0.982
$Z=X \cdot Y \sim N(-0.5,1)$	15	0.418	0.418	0.418	0.418		0.420	0.439
	25	0.625	0.625	0.625	0.622		0.646	0.668
	50	0.884	0.884	0.884	0.883		0.917	0.933

$$X \sim F(x) = \left(1 + e^{-x \sqrt{3}}\right)^{-1}, Y \sim R(x - \theta B_r(F(x))), \text{ where } B_r(t) = ((4t^2 - 1)/(3t))^{1/2}, t^{r-1}, r = 8, \theta = 4;$$

Increasing tail alternative A:

Increasing tail alternative B: $X \sim R(x), Y \sim R(x - \theta B_r(F(x))), \text{ where } r = 4, \theta = 4;$

$$Z = X - Y \sim N(-0.5,1): \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0.5 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right) \text{ (the two-dimensional normal distribution)}$$