TWO-SAMPLE DENSITY-BASED EMPIRICAL LIKELIHOOD TESTS FOR STOCHASTICALLY ORDERED ALTERNATIVES

Gregory Gurevich

Abstract

The empirical likelihood method based on the empirical distribution functions is a well-accepted statistical tool for testing. Recently, the density-based empirical likelihood technique was proposed and applied successfully to construct a powerful two-sample nonparametric likelihood ratio test based on samples entropy. However, while the problem of one-sided alternatives has received considerable attention in the case of the likelihood ratio tests, the two-sample density-based empirical likelihood test was proposed only in the context of the two-sided alternative. Hence, it is not suitable for a variety of applied statistical problems where the one-sided test should be driven by the scientific question and the data analyzed. In this paper we show how one-sample density-based empirical likelihood tests can be constructed and provide a proof of their consistency. Monte Carlo simulations confirm that the proposed one-sided nonparametric tests have approximately same powers as that of the Wilcoxon test detecting a constant shift in the one-sided two-sample problem and are preferable to the Wilcoxon test detecting a nonconstant shift.

Key words: Density-based empirical likelihood, Two-sample location problem, Paired data, Wilcoxon test

JEL Code: C14, C12

Introduction

The likelihood principle is arguably the most important concept for inference in parametric models. Recently it has also been shown to be useful in nonparametric contexts (e.g., Qin and Lawless, 1994; Lazar and Mykland, 1998; Owen, 2001; Lazar, 2003; Vexler and Gurevich, 2010a, 2010b; Vexler et al., 2012a, 2012b). Let $X_1,...,X_k \sim F$ be a sample of independent identically distributed observations where F is some distribution with a density function

$f(x)$. The Empirical Likelihood (EL) function has the form of $L_p = \prod_{i=1}^{k} p_i$, where the components p_i, $i = 1,...,k$, maximize L_p and satisfy empirical constraints corresponding to hypotheses of interest. For example, if the null hypothesis is $H_0 : E(X_1) = 0$, then the values of p_i's in the H_0-empirical likelihood L_p should be chosen to maximize L_p given $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} p_i X_i = 0$, where the constraint $\sum_{i=1}^{k} p_i X_i = 0$ is an empirical version of $E(X_1) = 0$. Computation of p_i, $i = 1,...,k$, is based on a simple exercise in Lagrange multipliers. This nonparametric approach is a result of consideration of the 'distribution functions'-based likelihood $\prod_{i=1}^{k} (F(X_i) - F(X_i -))$ over all distribution functions F (see, for details, Owen, 2001). Vexler and Gurevich (2010a, 2010b) proposed to use the central idea of the EL technique to develop density-based empirical approximations to the likelihood $L_f = \prod_{i=1}^{k} f(X_i)$, where $f(x)$ is a density function. To outline this technique, we present the likelihood function L_f in the form of $L_f = \prod_{i=1}^{k} f(X_i) = \prod_{i=1}^{k} f(X_{(i)}) = \prod_{i=1}^{k} f_i$ with $f_i = f(X_{(i)})$, and $X_{(1)} \le X_{(2)} \le \ldots \le X_{(k)}$ are the order statistics derived from $X_1, \ldots, X_k$. Following the maximum EL method, we can obtain estimated values of f_i, $i = 1, \ldots, k$, that maximize L_f and satisfy empirical constraints. Obviously, the equation $\int f(u)du = 1$ constrains values of f_i, $i = 1, \ldots, k$. To formalize this constraint, Vexler and Gurevich (2010a) proposed the following result.

Proposition 1. Assume $X_{(j)} = X_{(1)}$, if $j \le 1$, and $X_{(j)} = X_{(k)}$, if $j \ge k$. Then for all integer *m*, we have

$$\sum_{j=1}^{k} \int_{X_{(j-m)}}^{X_{(j+m)}} f(u)du = 2m \int_{X_{(1)}}^{X_{(k)}} f(u)du - \sum_{l=1}^{m-1} (m-l) \int_{X_{(k-l)}}^{X_{(k-l+1)}} f(u)du - \sum_{l=1}^{m-1} (m-l) \int_{X_{(l)}}^{X_{(l+1)}} f(u)du .$$

Denote $\mathrm{H}_m = \frac{1}{2m} \sum_{j=1}^{k} \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx$. Since $\int_{X_{(1)}}^{X_{(k)}} f(x)dx \le \int_{-\infty}^{+\infty} f(x)dx = 1$, Proposition 1 shows that $\mathrm{H}_m \le 1$, as well as, one can expect that $\mathrm{H}_m \approx 1$, when $m/k \to 0$ as $m,k \to \infty$. While approximating $\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx \cong (X_{(j+m)} - X_{(j-m)}) f_j$, we represent the condition $\mathrm{H}_m \le 1$ in the empirical form of

$$\tilde{H}_m \le 1,\ \ \tilde{H}_m = \frac{1}{2m}\sum_{j=1}^{k}\left(X_{(j+m)} - X_{(j-m)}\right)f_j\ . \tag{1}$$

Deriving $\partial/\partial f_i$, $i = 1,\ldots,k$, from the function $\log L_f + \lambda\left(1 - \tilde{H}_m\right)$ with the Lagrange multiplier λ, and then solving the resulting equation $\frac{\partial}{\partial f_i}\left(\log L_f + \lambda\left(1 - \tilde{H}_m\right)\right) = \frac{1}{f_i} - \lambda\frac{1}{2m}\left(X_{(i+m)} - X_{(i-m)}\right) = 0$, we obtain that the values $f_i = 2m\left(k\left(X_{(i+m)} - X_{(i-m)}\right)\right)^{-1}$, $i = 1,\ldots,k$, maximize $\log L_f$, satisfying the constraint (1) (here $X_{(j)} = X_{(1)}$, if $j \le 1$, and $X_{(j)} = X_{(k)}$, if $j \ge k$). Finally, the EL estimate of the likelihood has the form of $\prod_{i=1}^{k} 2m\left(k\left(X_{(i+m)} - X_{(i-m)}\right)\right)^{-1}$. Gurevich and Vexler (2011) and Vexler et al. (2012a, 2012b) utilized the above EL estimate to construct efficient two-sample density-based empirical likelihood tests for two-sided alternatives. This paper proceeds as follows. In Section 1 we shortly outline the way of constructing the existing EL two-sample test for the two-sided alternative (Gurevich and Vexler, 2011). We also extend this test to deal with one-sided alternatives and prove the consistency of a novel proposed test. In Section 2 we present the existing EL two-sample test for paired data (Vexler et al., 2012a, 2012b) and propose its modificated forms for one-sided alternatives. The obtained theoretical results and partially presented Monte Carlo study confirm the high efficiency of the proposed modified test.

1 A two-sample density-based empirical likelihood test for the classical two-sample problem

Let $X_1,...,X_n$ and $Y_1,...,Y_k$ be independent samples that consist of independent identically distributed observations from distributions F_X and F_Y with density functions $f_X(x)$ and $f_Y(y)$, respectively. We are interested to verify if the both samples are from the same distribution.

1.1 A two-sided alternative

Formally, we want to test for

$$H_0 : F_Y = F_X = F_Z \ \text{ versus } \ H_1 : F_Y \ne F_X, \tag{2}$$

where distributions F_Z, F_X and F_Y are unknown. In this case, the likelihood ratio statistic based on all $n + k$ observations has the form of

$$\frac{\prod_{i=1}^{n} f_X(X_i)\prod_{j=1}^{k} f_Y(Y_j)}{\prod_{i=1}^{n} f_Z(X_i)\prod_{j=1}^{k} f_Z(Y_j)} = \frac{\prod_{i=1}^{n} f_{X,i}\prod_{j=1}^{k} f_{Y,j}}{\prod_{i=1}^{n} f_{ZX,i}\prod_{j=1}^{k} f_{ZY,j}}, \tag{3}$$

where a density function f_Z corresponds to the null hypothesis, $f_{X,i} = f_X(X_{(i)})$, $f_{Y,j} = f_Y(Y_{(j)})$, and $f_{ZX,i} = f_Z(X_{(i)})$, $f_{ZY,j} = f_Z(Y_{(j)})$, $i=1,...,n$, $j=1,...,k$; $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, $Y_{(1)} \le Y_{(2)} \le \ldots \le Y_{(k)}$ are the order statistics based on the observations $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_k$, respectively. Gurevich and Vexler (2011) applied the method of the density-based EL mentioned in Introduction to estimate $f_{X,i}$, $i=1,...,n$, and $f_{ZY,j}$, $j=1,...,k$. Thus, they derived values of $f_{X,i}$, $i=1,\ldots,n$ that maximize the likelihood $\prod_{i=1}^{n} f_{X,i}$, satisfying an empirical constraint. Here the equation $\int f_X(u)du = 1$ constrains values of $f_{X,i}$, $i=1,\ldots,n$. By virtue of Proposition 1,

$$\sum_{i=1}^{n}\int_{X_{(i-m)}}^{X_{(i+m)}} \frac{f_X(u)}{f_Z(u)} f_Z(u)du = 2m\int_{X_{(1)}}^{X_{(n)}} \frac{f_X(u)}{f_Z(u)} f_Z(u)du - \sum_{l=1}^{m-1}(m-l)\int_{X_{(n-l)}}^{X_{(n-l+1)}} \frac{f_X(u)}{f_Z(u)} f_Z(u)du$$
$$-\sum_{l=1}^{m-1}(m-l)\int_{X_{(l)}}^{X_{(l+1)}} \frac{f_X(u)}{f_Z(u)} f_Z(u)du .$$

That is, since $\int_{X_{(1)}}^{X_{(n)}} f_X(u)du \le \int_{-\infty}^{+\infty} f_X(u)du = 1$, one can conclude

$$\Delta_m \le 1, \quad \Delta_m = \frac{1}{2m}\sum_{i=1}^{n}\int_{X_{(i-m)}}^{X_{(i+m)}} \frac{f_X(u)}{f_Z(u)} f_Z(u)du, \tag{4}$$

and $\Delta_m \approx 1$ when $m/n \to 0$ as $m,n \to \infty$. In a similar manner to deriving the constraint (1), by applying the approximate analog to the mean-value integration theorem, Gurevich and Vexler (2011) approximated Δ_m as

$$\Delta_m \cong \frac{1}{2m}\sum_{i=1}^{n}\frac{f_{X,i}}{f_{ZX,i}}\int_{X_{(i-m)}}^{X_{(i+m)}} f_Z(u)du = \frac{1}{2m}\sum_{i=1}^{n}\left(F_Z\left(X_{(i+m)}\right) - F_Z\left(X_{(i-m)}\right)\right)\frac{f_{X,i}}{f_{ZX,i}} \tag{5}$$
$$\cong \frac{1}{2m}\sum_{i=1}^{n}\left(F_{Z(n+k)}\left(X_{(i+m)}\right) - F_{Z(n+k)}\left(X_{(i-m)}\right)\right)\frac{f_{X,i}}{f_{ZX,i}},$$

where

$$F_{Z(n+k)}(u)=\frac{1}{n+k}\left(\sum_{i=1}^{n}I(X_i\le u)+\sum_{j=1}^{k}I(Y_j\le u)\right) \quad (6)$$

is the empirical distribution function that estimates the distribution $F_Z(u)$ ($I(\cdot)$ is the indicator function). Futher considerations presented in Gurevich and Vexler (2011) lead to the EL ratio test-statistic $V_{nk}=ELR_{X,n}ELR_{Y,k}$, where

$$ELR_{X,n}=\min_{a_n\le m\le b_n}\prod_{i=1}^{n}\frac{2m}{n\left(F_{Z(n+k)}\left(X_{(i+m)}\right)-F_{Z(n+k)}\left(X_{(i-m)}\right)\right)}, \quad (7)$$

$a_n=n^{0.5+\delta}$, $b_n=\min\left(n^{1-\delta},\frac{n}{2}\right)$, $\delta\in(0,0.25)$, $X_{(j)}=X_{(1)}$, if $j\le 1$, and $X_{(j)}=X_{(n)}$, if $j\ge n$;

$$ELR_{Y,k}=\min_{a_k\le r\le b_k}\prod_{i=1}^{k}\frac{2r}{k\left(F_{Z(n+k)}\left(Y_{(i+r)}\right)-F_{Z(n+k)}\left(Y_{(i-r)}\right)\right)}, \quad (8)$$

$a_k=k^{0.5+\delta}$, $b_k=\min\left(k^{1-\delta},\frac{k}{2}\right)$, $\delta\in(0,0.25)$, $Y_{(j)}=Y_{(1)}$, if $j\le 1$; $Y_{(j)}=Y_{(k)}$, if $j\ge k$. This EL ratio test-statistic V_{nk} approximates the optimal likelihood ratio test statistic (3). Finally, the EL density-based test rejects the null hypothesis of (2) if $\log(V_{nk})>C$, where C is a test-threshold.

1.2 One-sided alternatives

Without loss of generality, we consider the following one-sided version of the two-sided problem (2)

$$H_0: F_Y=F_X=F_Z \quad \text{versus} \quad (9)$$
$$H_1: F_X(u)\le F_Y(u) \text{ for all } -\infty<u<\infty,\ F_X(u)<F_Y(u) \text{ for some } -\infty<u<\infty.$$

This formulation means that under H_1, $X\overset{st}{\succ}Y$. In this case we propose to apply to (5) the definition $F_Z^*(u)=\max\left(F_{X,n}(u),F_{Y,k}(u)\right)=\max\left(\frac{1}{n}\sum_{i=1}^{n}I(X_i\le u),\frac{1}{k}\sum_{j=1}^{k}I(Y_j\le u)\right)$ instead of the definition (6). That is, we define $ELR^*_{X,n}=\min_{a_n\le m\le b_n}\prod_{i=1}^{n}\frac{2m}{n\left(F_Z^*\left(X_{(i+m)}\right)-F_Z^*\left(X_{(i-m)}\right)\right)}$ as the modification of (7). By the similar way, we define $ELR^*_{Y,k}=\min_{a_k\le r\le b_k}\prod_{i=1}^{k}\frac{2r}{k\left(F_Z^{**}\left(Y_{(i+r)}\right)-F_Z^{**}\left(Y_{(i-r)}\right)\right)}$, where $F_Z^{**}(u)=\min\left(F_{X,n}(u),F_{Y,k}(u)\right)$, as the modification of (8). Thus, we propose to reject the null hypothesis of (9) if

$$\log\left(V_{nk}^{*}\right) > C, \tag{11}$$

where $V_{nk}^{*} = ELR_{X,n}^{*} ELR_{Y,k}^{*}$, $a_n = n^{0.5+\delta}$, $b_n = \min\left(n^{1-\delta}, \frac{n}{2}\right)$, $a_k = k^{0.5+\delta}$, $b_k = \min\left(k^{1-\delta}, \frac{k}{2}\right)$, $\delta \in (0, 0.25)$, and C is a test-threshold.

The next proposition indicates that the test (11) is consistent as $n,k \to \infty$, $n/k \to \eta$, where a constant $\eta > 0$. To formulate the following result, we assume that F_X and F_Y, mentioned in the statement (9), are the continuous cumulative distribution functions with density functions f_X and f_Y, respectively.

Proposition 2. If the expectations $E(\log f_X(X_1))$, $E(\log f_X(Y_1))$, $E(\log f_Y(Y_1))$ and $E(\log f_Y(X_1))$ are finite, then

$$\frac{1}{n+k}\log\left(V_{nk}^{*}\right) \xrightarrow{P} \gamma, \text{ as } n,k \to \infty,\ n/k \to \eta,\ \eta > 0 \text{ is a constant,}$$

where

$$\gamma = -\frac{\eta}{1+\eta} E\left(\log\left(\frac{f_Y(X_1)}{f_X(X_1)}\right)\right) - \frac{1}{1+\eta} E\left(\log\left(\frac{f_X(Y_1)}{f_Y(Y_1)}\right)\right).$$

It is clear that, under the null hypothesis, the ratio $f_Y / f_X = 1$ that implies $\gamma = 0$. Under the alternative H_1, we have $E(f_Y(X_1)/f_X(X_1)) = E(f_X(Y_1)/f_Y(Y_1)) = 1$ that implies

$$\gamma \geq -\frac{\eta}{1+\eta}\left(\log\left(E\frac{f_Y(X_1)}{f_X(X_1)}\right)\right) - \frac{1}{1+\eta}\left(\log\left(E\frac{f_X(Y_1)}{f_Y(Y_1)}\right)\right) = 0.$$

Thus, the consistency of the proposed density-based EL test (11) is given by Proposition 2.

2 A two-sample density-based empirical likelihood test for the paired data

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a bivariate population with continuous joint distribution function $F_{XY}(x, y)$ and the marginal distributions $F_X(x)$, $F_Y(y)$ of X_i and Y_i, respectively. Common statistical procedures for testing the problem (2) consist of the paired t-test, the sign test and the Wilcoxon signed-rank test. These tests are based on the *n* paired differences $Z_i = X_i - Y_i$. In the paired data case, the classical nonparametric Wilcoxon signed-rank procedure is a permutation based method under the assumption that Z is symmetric about zero under the null hypothesis. Similarly, the sign test assumes symmetry about zero under the null hypothesis and utilizes the binomial distribution with respect to generating the test procedure. Note that, when parametric forms of the distribution functions F_X, F_Y and F_{XY} can be assumed to be known, the parametric likelihood ratio can be efficiently applied to the problem (2). Alternatively to the classical nonparametric tests, Vexler et al. (2012a) developed the procedure presenting the corresponding likelihood ratio statistic in an empirical form. This procedure was developed for two-sided alternatives and is based on the *n* paired differences $Z_i = X_i - Y_i$. Here we outline shortly this method.

2.1 A two-sided alternative

Let $Z_1, \ldots, Z_n$ be independent identically distributed (i.i.d.) random variables with the distribution function F. The considered hypotheses testing problem is presented as:

$$H_0 : F = F_{H_0},\ F_{H_0}(u) = 1 - F_{H_0}(-u), \text{ for all } -\infty < u < \infty \text{ versus} \qquad (12)$$

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

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

$$LR = \frac{\prod_{i=1}^{n} f_{H_1}(Z_i)}{\prod_{i=1}^{n} f_{H_0}(Z_i)} = \frac{\prod_{j=1}^{n} f_{H_1}\left(Z_{(j)}\right)}{\prod_{j=1}^{n} f_{H_0}\left(Z_{(j)}\right)} = \frac{\prod_{j=1}^{n} f_{H_1,j}}{\prod_{j=1}^{n} f_{H_0,j}}, \qquad (13)$$

where $f_{H_1}(u)$ and $f_{H_0}(u)$ denote the density functions of Z under H_1 and H_0, respectively, $Z_{(1)} \le Z_{(2)} \le \ldots \le Z_{(n)}$ is the order statistic and $f_{H_k,j} = f_{H_k}\left(Z_{(j)}\right)$, $k = 0,1$, $j = 1,\ldots,n$. The nonparametric approximation for the likelihood ratio statistic (13) is obtained by estimating the values of $f_{H_1,j}, j = 1,\ldots,n$. This is accomplished via maximizing the log-likelihood $\sum_{j=1}^{n} \log\left(f_{H_1,j}\right)$ provided that $f_{H_1,j}$, $j = 1,\ldots,n$, satisfy an empirical constraint that is an empirical form of $\int f_{H_1}(u)du = 1$. Based on the Proposition 1, by the similar way as described in Section 1 and presented in Vexler et al. (2012a) we formulate this constraint in form of

$$\frac{1}{2m}\sum_{j=1}^{n} \frac{f_{H_1,j}}{f_{H_0,j}} \Delta_{jm} = 1, \qquad (14)$$

where $\Delta_{jm} := \frac{1}{2n}\sum_{i=1}^{n}\Big(I\big(Z_i \le Z_{(j+m)}\big) + I\big(-Z_i \le Z_{(j+m)}\big) - I\big(Z_i \le Z_{(j-m)}\big) - I\big(-Z_i \le Z_{(j-m)}\big)\Big)$ is an estimator for the difference $F_{H_0}(Z_{(j+m)}) - F_{H_0}(Z_{(j-m)})$ based on the distribution free estimation $\hat{F}_{H_0}$ for a symmetric distribution F_{H_0} proposed by Schuster (1975):

$$\hat{F}_{H_0}(u) = \frac{1}{2n}\left(\sum_{i=1}^{n} I\big(Z_i \le u\big) + \sum_{i=1}^{n} I\big(-Z_i \le u\big)\right). \tag{15}$$

Futher considerations similar to that presented in Vexler et al. (2012a) lead to the following approximation of the parametric likelihood ratio statistic (13):

$$V_n = \min_{a(n)\le m\le b(n)} \prod_{j=1}^{n} \frac{2m}{n\Delta_{jm}}, \tag{16}$$

where $a(n) = n^{0.5+\delta}$, $b(n) = \min(n^{1-\delta}, n/2)$, $\delta \in (0, 0.25)$. Thus, the EL density-based test rejects the null hypothesis of (12) if $\log(V_n) > C$, where C is a test-threshold.

2.2 One-sided alternatives

Without loss of generality, we consider the following one-sided version of the two-sided problem (12)

$$H_0 : F = F_{H_0},\ F_{H_0}(u) = 1 - F_{H_0}(-u), \text{ for all } -\infty < u < \infty \text{ versus} \tag{17}$$

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

This formulation means that under H_1, $Z \overset{st}{\succ} -Z$ (that corresponds to $X \overset{st}{\succ} Y$). In this case we propose to apply to (14) the definition

$\hat{\hat{F}}_{H_0}(u) = \max\frac{1}{n}\left(\sum_{i=1}^{n} I\big(Z_i \le u\big), \sum_{i=1}^{n} I\big(-Z_i \le u\big)\right)$ instead of the definition (15). That is, we define

$\Delta^{*}_{jm} := \hat{\hat{F}}_{H_0}\big(Z_{(j+m)}\big) - \hat{\hat{F}}_{H_0}\big(Z_{(j-m)}\big)$ as the modification of $\Delta_{jm} := \hat{F}_{H_0}\big(Z_{(j+m)}\big) - \hat{F}_{H_0}\big(Z_{(j-m)}\big)$. Thus, the proposed test statistic for the problem (16) has the form

$$V^{*}_n = \min_{a(n)\le m\le b(n)} \prod_{j=1}^{n} \frac{2m}{n\Delta^{*}_{jm}}, \tag{18}$$

where

$$\Delta^{*}_{jm} := \max\frac{1}{n}\left(\sum_{i=1}^{n} I\big(Z_i \le Z_{(j+m)}\big), \sum_{i=1}^{n} I\big(-Z_i \le Z_{(j+m)}\big)\right) - \max\frac{1}{n}\left(\sum_{i=1}^{n} I\big(Z_i \le Z_{(j-m)}\big), \sum_{i=1}^{n} I\big(-Z_i \le Z_{(j-m)}\big)\right),$$

if $\Delta_{jm}^* < \frac{1}{n}$ then $\Delta_{jm}^* = \frac{1}{n}$, $a(n) = n^{0.5+\delta}$, $b(n) = \min(n^{1-\delta}, n/2)$, $\delta \in (0, 0.25)$. The V_n^*-test rejects the null hypothesis of (17) if

$$\log\left(V_n^*\right) > C, \tag{19}$$

where C is a test-threshold.

The consistency of the proposed test (19) is stated in the following proposition.

Proposition 3. Let $f(x)$ define a density function of the observations $Z_1,...,Z_n$ with the finite expectations $E(\log f(Z_1))$ and $E(\log f(-Z_1))$. Then

$$\frac{1}{n}\log\left(V_n^*\right) \overset{p}{\to} -E\left(\log\left(\frac{f(-Z_1)}{f(Z_1)}\right)\right), \text{ as } n \to \infty,$$

for all $0 < \delta < 0.25$ in the definition (18) of the statistic V_n^*.

It is obvious that the limiting value of $n^{-1}\log\left(V_n^*\right)$, the expectation $-E\left(\log\left(f(-Z_1)\, f(Z_1)^{-1}\right)\right)$ stated in Proposition 3, has the forms of $-E_{H_0}\log\left(f_{H_0}(-Z_1)\left(f_{H_0}(Z_1)\right)^{-1}\right) = 0$ and $-E_{H_1}\left(\log\left(f_{H_1}(-Z_1)\left(f_{H_1}(Z_1)\right)^{-1}\right)\right) \geq -\log\left(E_{H_1}\left(f_{H_1}(-Z_1)\left(f_{H_1}(Z_1)\right)^{-1}\right)\right) = 0$, under H_0 and H_1, respectively. This implies that the test (19) is consistent.

To compare powers of the proposed test (19) with that of the classical Wilcoxon signed-rank test we conducted a simulation study. We focus on $\delta = 0.1$ applied to the definition (18). The Monte Carlo experiments for investigating the power properties of the tests were repeated 25,000 times. Critical values of the tests were set up to preserve the 5% level of significance of the decision rules. The following Table 1 presents the simulated powers of the tests for several typical alternatives.

Table 1: Monte Carlo powers of the test (19) with $\delta = 0.1$ and and the Wilcoxon signed-rank test; at the significance level α=0.05, for the sample size $n = 25$.

F_Z	Proposed test (19)	Wilcoxon signed-rank test
$Norm(0,1)$	0.999	0.999
$Norm(0.5,1)$	0.764	0.775
$Norm(0.25,1)$	0.324	0.334
$Unif(-0.3,-0.7)$	0.977	0.924
$Unif(-0.5,1)$	0.927	0.826
$Unif(-1.5,2.5)$	0.757	0.619
$F_X = LogN(2,4^2)$, $F_Y = LogN(1,1)$, $Z = X - Y$	0.968	0.776

$F_X = LogN(2,3^2)$, $F_Y = LogN(1,1)$, $Z = X - Y$	0.952	0.812
$F_X = LogN(2,1)$, $F_Y = LogN(1,1)$, $Z = X - Y$	0.929	0.937

Source: own research.

Thus, the proposed new test clearly shows high and stable power as compared to the classical Wilcoxon signed-rank test.

Conclusion

The presented method employs the EL concept in a nonparametric fashion in order to approximate Neyman-Pearson type parametric likelihood ratio test-statistics for two sample problems. The benefit of using this approach is threefold. First, we are able to construct exact and robust nonparametric tests. Second, the proposed technique is highly efficient given that it approximates well the optimal parametric likelihood ratio. Third, the obtained tests have high and stable powers over a variety of alternative distributions, resulting in a large power gain in comparison to the classical procedures. We illustrated that the EL method can be easily applied to one-sided alternatives, and provided meaningful results.

References

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

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

Lazar, N.A. (2003). Bayesian empirical likelihood. *Biometrika,* 90, 319-326.

Owen, A.B. (2001). *Empirical likelihood*. Chapman and Hall/CRC.

Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics, 22,* &300-325.

Schuster, E. F. (1975). Estimating the distribution function of a symmetric distribution, *Biometrika, 62*, 631-635.

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

Vexler, A & Gurevich, G. (2010b). Density-based empirical likelihood ratio change point detection policies. *Communications in Statistics-Simulation and Computation, 39,* 1709-1725.

Vexler, A., Gurevich, G., & Hutson, A.D. (2012a). An exact density-based empirical likelihood ratio test for paired data. *Journal of Statistical Planning and Inference*, under revision.

Vexler, A., Tsai, W.M., Gurevich, G., & Yu, J. (2012b). Two-sample density-based empirical likelihood ratio tests based on paired data, with application to a treatment study of attention-deficit/hyperactivity disorder and severe mood dysregulation. *Statistics in Medicine*, in press (DOI:10.1002/sim.4467).

Contact

Gregory Gurevich
SCE-Shamoon College of Engineering
Jabotinsky St., Ashdod 77245, Israel
gregoryg@sce.ac.il