A density based empirical likelihood approach for testing bivariate normality

Gregory Gurevich[a] and Albert Vexler[b,1]

[a] Department of Industrial Engineering and Management, SCE-Shamoon College of Engineering, Beer Sheva 84100, Israel

[b] Department of Biostatistics, The State University of New York, Buffalo, NY 14214, USA

Sample entropy based tests, methods of sieves and Grenander estimation type procedures are known to be very efficient tools for assessing normality of underlying data distributions, in one-dimensional nonparametric settings. Recently, it has been shown that the density based empirical likelihood (EL) concept extends and standardizes these methods, presenting a powerful approach for approximating optimal parametric likelihood ratio test statistics, in a distribution-free manner. In this paper, we discuss difficulties related to constructing density based EL ratio techniques for testing bivariate normality and propose a solution regarding this problem. Toward this end, a novel bivariate sample entropy expression is derived and shown to satisfy the known concept related to bivariate histogram density estimations. Monte Carlo results show that the new density based EL ratio tests for bivariate normality behave very well for finite sample sizes. In order to exemplify the excellent applicability of the proposed approach, we demonstrate a real data example related to a study of biomarkers associated with myocardial infarction.

Keywords: Bivariate normality; Density estimation; Empirical likelihood; Entropy; Goodness-of-fit; Histogram density estimation

1. Introduction and the statement of problem

Various statistical topics dealt with bivariate normally distributed data have broadened their appeal in recent theoretical and applied publications that provide a long cohort of new methods for multivariate statistical analysis (e.g., Balakrishnan and Lai [2]). This motivates the growing need for developing and evaluating powerful tests for bivariate normality (e.g., Balakrishnan and Lai [2]; Hawkins [7]; Kowalski [13]; Mecklin and Mundfrom [19]). Testing bivariate data for normality is much more difficult in practice than when data are univariate. Commonly techniques for detecting departures from the bivariate normal distribution are developed using modifications of conventional test procedures

---

[1] Corresponding author. Email: avexler@buffalo.edu

known in the context of assessing univariate normality. In many cases, in order to test goodness-of-fit of two-dimensional normal distribution functions, the literature proposes one-dimensional test statistics (e.g., Balakrishnan and Lai [2]). In this framework, we note that it is not sufficient to test corresponding univariate marginal distributions for normality, since scenarios, when the marginal distributions are normal but the joint distribution fails to be bivariate normal, may be in effect.

In this paper we propose and examine a bivariate extension to the one-dimensional sample entropy based concept (e.g., Vasicek [30]) using the density based empirical likelihood (EL) methodology (e.g., Vexler et al. [33]). Then, we propose the density based EL ratio tests for bivariate normality. To this end, we shall first outline the following material regarding basic sources we use in the new development.

1.1. Empirical likelihood and sample entropy

When functional forms of underlying data distributions are completely specified, the parametric likelihood approach is unarguably a powerful tool that provides optimal statistical inference. In such cases, by virtue of the Neyman-Pearson lemma, the likelihood ratio tests yield the most powerful decision making rules (e.g., Lehmann and Romano [16]; Vexler et al. [33]). However, the parametric likelihood methods cannot be applied properly if assumptions on the forms of data distributions do not hold. The distribution function based EL methods were introduced as nonparametric alternatives to parametric likelihood techniques (e.g., Lazar and Mykand [14]; Owen [22]). Commonly, the conventional EL function has the form $EL = \prod_{i=1}^{n} p_i$, where the probability weights, $p_i, i = 1,...,n$, satisfy the 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 observe independent and identically distributed (i.i.d.) data points $X_1,..., X_n$, under the null hypothesis that $E(X_1) = 0$, the corresponding empirical constraint is $\sum_{i=1}^{n} p_i X_i = 0$.

The recent statistical literature has introduced the density-based EL (dbEL) approach for creating nonparametric test statistics that approximate parametric Neyman-Pearson statistics (e.g., Vexler et al. [33]). The dbEL method proposes to consider the likelihood function in the form

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

where $f(\cdot)$ is a density function of observations $X_1,..., X_n$, and $X_{(1)} \le ... \le X_{(n)}$ are the order statistics based on $X_1,..., X_n$. Then, one can estimate values of $f_j, j = 1,...,n$, maximizing L_f given a

constraint related to the empirical version of the density property $\int f(x)dx = 1$. In this case, the following lemma (Vexler and Gurevich [31]) has a key role.

Lemma 1. Let $f(x)$ be a density function. Then

$$\frac{1}{2m}\sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx = \int_{X_{(1)}}^{X_{(n)}} f(x)dx - \frac{1}{2m}\sum_{k=1}^{m-1}(m-k)\int_{X_{(n-k)}}^{X_{(n-k+1)}} f(x)dx - \frac{1}{2m}\sum_{k=1}^{m-1}(m-k)\int_{X_{(k)}}^{X_{(k+1)}} f(x)dx,$$

where $m < n/2$, $X_{(j)} = X_{(1)}$, if $j \le 1$, and $X_{(j)} = X_{(n)}$, if $j \ge n$.

In terms of constructing the dbEL ratio test for univariate normality, Lemma 1 implies the next inference. By virtue of the inequality

$$\int_{X_{(1)}}^{X_{(n)}} f(x)dx \le \int_{-\infty}^{\infty} f(x)dx = 1,$$

we obtain the empirical constraint

$$(2m)^{-1}\sum_{j=1}^{n}\left(X_{(j+m)} - X_{(j-m)}\right)f_j \le 1,$$

where the expressions $\left(X_{(j+m)} - X_{(j-m)}\right)f_j, j = 1,..,n,$ are Mean-Value Theorem type approximations to

$\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)dx, j = 1,...,n,$ which appear in Lemma 1. Thus, the method of Lagrange multipliers

provides values of $f_1,...,f_n$, which maximize $\log(L_f)$ and satisfy

$$(2m)^{-1}\sum_{j=1}^{n}\left(X_{(j+m)} - X_{(j-m)}\right)f_j \le 1,$$

resulting in

$$f_i = 2mn^{-1}\left(X_{(i+m)} - X_{(i-m)}\right)^{-1}, i = 1,...,n.$$

Therefore, taking into account the maximum likelihood function, say L_N, under the null hypothesis

$H_0 : X_1,...,X_n$ are normally distributed, where $L_N \propto \left(2\pi e s^2\right)^{-n/2}$ with $s^2 = n^{-1}\sum_{j=1}^{n}\left(X_j - n^{-1}\sum_{k=1}^{n}X_k\right)^2$,

we obtain the dbEL ratio

$$T_{mn} = \left(2\pi e s^2\right)^{n/2}\prod_{i=1}^{n}\frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}$$

that is known to be an efficient test statistic based on sample entropy (e.g., Vasicek [30]; Arizono and Ohta [1]). In order to develop the sample entropy based test for normality, Vasicek [30] applied the property of the normal distribution that its entropy exceeds that of any other distribution with a density that has the same variance. The dbEL approach extends this sample entropy based mechanism to

3

general methods for univariate goodness-of-fit testing. The test for normality based on sample entropy is an exponential rate optimal procedure (see Tusnady [28] for details). This is in conjunction with the fact that likelihood ratio type tests oftentimes have optimal statistical properties.

In the construction process of the test statistic T_{mn} shown above, we used the approximation to the constraint

$$(2m)^{-1} \sum_{j=1}^{n} \int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx \leq 1.$$

By virtue of Lemma 1, one can expect that

$$(2m)^{-1} \sum_{j=1}^{n} \int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx \to 1 \text{ when } m/n \to 0 \text{ as } m, n \to \infty.$$

In this case, the integer m should increase when $n \to \infty$, provided that $m/n \to 0$, since, in light of Lemma 1, corresponding remainder terms related to the constraint

$$\int_{X_{(1)}}^{X_{(n)}} f(x) dx \leq \int_{-\infty}^{\infty} f(x) dx = 1$$

need to vanish asymptotically (e.g., Vexler et al. [32]). In general cases, if m has a fixed value, the approximation $\prod_{i=1}^{n} 2mn^{-1} \left(X_{(i+m)} - X_{(i-m)} \right)^{-1}$ to the parametric likelihood L_f is not consistent (see the Supplement, Appendix A, for technical details). This is an interesting point, since naively one can anticipate that fixed values of m can provide 'good' approximations to $\int_{-\infty}^{\infty} f(x) dx$ as $n \to \infty$, shortening distances between $\int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx, j = 1, ..., n,$ and their estimators $\left(X_{(j+m)} - X_{(j-m)} \right) f_j, j = 1, .., n.$ However, when m is fixed and $n \to \infty$, the number of $\int_{X_{(j-m)}}^{X_{(j+m)}} f(x) dx, j = 1, ..., n,$ is bigger than that in the case with $m \to \infty$. This enlarges a total error of the applied Mean-Value Theorem type approximations in the context of the $\int_{-\infty}^{\infty} f(x) dx$-estimation, when intervals $\left[X_{(i-m)}, X_{(i+m)} \right], i = 1, ..., n,$ overlap.

Note that the dbEL technique mentioned above can be employed in order to estimate density functions in the maximum likelihood manner, obtaining a class of histogram density estimators (e.g., Izenman [9]; Prakasa Rao [23]). In this framework, we may use fixed values of m, assuming that f is a monotonically decreasing density function. This can yield procedures related to Grenander's

estimation and the method of sieves in the context of nonparametric density functions' evaluations (Izenman [9]; Carolan and Dykstra [4]; Efromovich [5: pp. 341-343]).

The next advance of the dbEL approach based on the Lemma 1's consequence is found by attending to that we should specify values of m to apply the test statistic T_{mn}, in practice. It turns out that, since we employ the likelihood concept to derive T_{mn}, the maximum likelihood principle implies the test statistic

$$T_n = \min_{1 \le m < n^{1-\delta}} (T_{mn}),\ 0 < \delta < 1,$$

where the operator '$\min_m$' automatically sets up nearly optimal values of m at the test statistic T_{mn} (Vexler and Gurevich [31]). In general cases, the optimal values of m, which maximize the power of the test based on T_{mn}, can be calculated using information regarding the alternative distribution of the observations.

The dbEL method was successfully applied to construct various nonparametric decision-making schemes, significantly improving power as compared to the corresponding classical procedures (e.g., Vexler, Hutson and Chen, 2016).

1.2.   Bivariate extensions to the univariate density based empirical likelihood and sample entropy expressions

Let $(X, Y)^T$ denote the bivariate random vector from the joint density function $f(x, y)$. Consider a sample from $f(x, y)$ in the form $\{(X_i, Y_i), i = 1, .., n\}$. In this case, the likelihood function is

$$L_f = \prod_{i=1}^n f(X_i, Y_i) = \prod_{i=1}^n f\left(X_{(i)}, Y_{[i]}\right) = \prod_{i=1}^n f_i,\tag{1}$$

where the sample is ordered by X_i, the notation $Y_{[i]}$ is termed the concomitant of the *i*-th order statistic $X_{(i)}$, representing the *Y*-variate associated with $X_{(i)}$ and $f_i = f\left(X_{(i)}, Y_{[i]}\right), i = 1, ..., n$.

In this case, according to the dbEL technique, we aim to achieve maximization of L_f, holding an empirical constraint with respect to the requirement $\iint f(x, y) dx dy = 1$. The problem is to approximate the double integral $\iint f(x, y) dx dy$ using only n data points. That is to say although it would be desired to apply $n \times n$ points in a Riemann-type manner to approximate the double integral, we cannot employ the couples $(X_i, Y_j), i \ne j$, that are not associated with $f_i, i = 1, ..., n$, and do not

5

appear at L_f defined in (1). In this context, one can attempt to adapt multivariate entropy estimation algorithms (Berrett et al. [3]; Kozachenko and Leonenko [12]) using open circles around the points $\left(X_{(i)}, Y_{[i]}\right), i = 1,...,n$, with the radiuses

$$\rho_i = min_{j=1,...,n;\, j\neq i} \left\{ \left(X_{(i)} - X_{(j)}\right)^2 + \left(Y_{[i]} - Y_{[j]}\right)^2 \right\}^{1/2}, i = 1,...,n.$$

The methods for estimating the random vector entropy can provide consistent evaluations of $\iint f(x,y)\log\left(f(x,y)\right)dxdy$. However, in terms of the target construction of approximations to $\iint f(x,y)dxdy$ based on $f_i, i = 1,...,n,$ these approaches cannot be directly involved due to very complicated overlaps between the open circles around the points $\left(X_{(i)}, Y_{[i]}\right), i = 1,...,n.$

In order to avoid the issue above, one can reduce the dimension of the testing problem via an application of projection pursuit techniques (Zhu et al. [38]). In this framework, it can be proposed to use the fact that $(X,Y)^T$ is bivariate normally distributed, say $(X,Y)^T \sim N_2(\mu, V)$ with $\mu = E\left\{(X,Y)^T\right\}$ and variance-covariance matrix V, if and only if for every vector $a \in R^2$ such that $a^T a = 1$ and $a^T V a \neq 0$, we have

$$Z_a = \left(a^T (X,Y)^T - a^T \mu\right) / \left(a^T V a\right) \sim N_1(0,1).$$

Now, we can compute estimators $\hat{\mu}$, $\hat{V}$ of the parameters μ, V and then consider one-dimensional observations

$$Z_{a,i} = \left(a^T (X_i, Y_i)^T - a^T \hat{\mu}\right) / \left(a^T \hat{V} a\right), i = 1,...,n,$$

in order to derive the dbEL ratio depending on a via the method shown in Section 1.1. Then the obtained dbEL ratio as a function of a can be, for example, integrated over different values of a, yielding a final test-statistic. It is clear that this approach might suffer from an efficiency loss when the likelihood function L_f under the alternative hypothesis is replaced by that based on $Z_{a,i}, i = 1,...,n.$ Properties of the decision making procedure based upon projection pursuit are significantly depend on a way for summarizing the final test statistic with respect to different values of a.

In a similar manner to that shown above, one can evaluate an option to transform the original variates X and Y to independent normal variates (Balakrishnan and Lai [2: p. 509]) that is a correct

operation only under the null hypothesis, $H_0 : (X,Y)^T \sim N_2(\mu, V)$, in general cases. Note also that algorithms to transform $(X,Y)^T \sim N_2(\mu, V)$ into two independent normal variates, say Z_X and Z_Y, depend on the parameters μ, V that are unknown. If μ and V are estimated, the target distributional properties of (Z_X, Z_Y) are only approximate.

We, in this paper, carry out an accurate scheme to apply the Lemma 1's result to approximate $\iint f(x, y)dxdy$, basing on $f_i, i = 1,...,n$. Then, in Section 2, maximizing L_f defined in (1), we obtain estimators of $f_i, i = 1,...,n,$ in forms that can be directly associated with the bivariate histogram density estimation (Kim and van Ryzin [11]; van Ryzin [29]; Prakasa Rao [23: pp. 234-235]; Izenman [9: pp. 209-210, 212-213]). This displays a natural linkage between the univariate sample entropy based approach shown in Section 1.1 and the proposed methodology (see Section 3 for details). We study the asymptotic consistency of the bivariate dbEL technique in Section 4. In this context, due to the complexity of a structure of the dbEL, we assume a condition on $f(x, y)$ to show rigorously that the bivariate dbEL is a consistent approximation of the parametric likelihood L_f. Various Monte Carlo experiments based on more than one hundred different scenarios of (X,Y)-distributions and a variety of sample sizes *n* showed that the theoretical condition applied in Section 4 is not critical. Then, we believe the bivariate dbEL approach is consistent in more general cases.

In Section 5, we develop the dbEL ratio tests for bivariate normality. Consider the simple example assuming that $X_1,...,X_{50}$ are i.i.d. observations from a standard normal distribution and $Y_1,...,Y_{50}$ are defined as $Y_i = \tau_i X_i$, *i=1,...50,* where random variables $\tau_i = -1$ or 1, $i = 1,...,50$, are i.i.d. and independent of $X_1,...,X_{50}$ with $Pr(\tau_1 = -1) = 0.5$. This is a conventional scenario when $X_1,...,X_{50} \sim N_1(0,1)$ and $Y_1,...,Y_{50} \sim N_1(0,1)$, but $(X_1,Y_1),...,(X_{50},Y_{50})$ are not bivariate normal. In this case, the Shapiro-Wilk test (R procedure "*mvShapiro.Test*", R Development Core Team [24]) and classical Mardia's test for bivariate normality show power of 0.06 and 0.38 at the significance level of 5%, respectively, whereas the new tests we propose provide power of 0.83 and 0.92 (see Section 6 for details). One advantage of the proposed technique is that by applying the dbEL approach we can powerfully detect failures of underlying data to be bivariate normal. In Section 6, an extensive Monte Carlo study is employed to support this conclusion.

In Section 7 the proposed tests are applied to a biomarker study associated with myocardial infarction (MI) disease. The epidemiological literature indicates significant associations between the

biomarkers "vitamin E", "cholesterol" and the MI disease. We demonstrate that the new tests based on measurements related to "vitamin E" and "cholesterol" biomarkers exhibit high and stable power characteristics in comparison to the well-known decision making procedures. We conclude with remarks in Section 8. Finally the technical proofs of the theoretical results shown in this paper are given in the Supplement. The online supplementary material of this paper also presents R code to implement the proposed method.

2. The bivariate density based empirical likelihood

In this section, we introduce the algorithm for developing the bivariate dbEL approximation to the likelihood function L_f defined in (1). Towards this end, we begin with outlining the following scheme to construct an empirical estimation of the constraint $\iint f(x,y)dxdy = 1$ based on the observations $\left(X_{(i)}, Y_{[i]} \right), i = 1,...,n$. The proposed algorithm consists of two stages: (A) we use Lemma 1 with respect to the density function $f(x)$ of X, employing $X_{(1)},...,X_{(n)}$, and then (B) we use Lemma 1 regarding the conditional density function $f(y|x)$, employing Y's that link to X's involved in corresponding procedures related to Stage (A).

Lemma 1 applied to the density function

$$\int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} f(y|x)f(x)dy = f(x)\int_{-\infty}^{\infty} f(y|x)dy$$

provides the inequality

$$\frac{1}{2m}\sum_{i=1}^{n}\int_{X_{(i-m)}}^{X_{(i+m)}} f(x)\int_{-\infty}^{\infty} f(y|x)dydx \leq \int_{-\infty}^{\infty} f(x)\int_{-\infty}^{\infty} f(y|x)dydx = 1, \tag{2}$$

where $m < n/2$.

Next, we will use Lemma 1 to approximate $\int f(y|x)dy$. To this end, we rewrite the left term of (2) as

$$\frac{1}{2m}\sum_{i=1}^{n}\int_{X_{(i-m)}}^{X_{(i+m)}} f(x)\int_{-\infty}^{\infty} f(y|x)dydx = \frac{1}{2m}\sum_{i=1}^{m}\int_{X_{(i-m)}}^{X_{(i+m)}} f(x)\int_{-\infty}^{\infty} f(y|x)dydx \tag{3}$$

$$+\frac{1}{2m}\sum_{i=m+1}^{n-m}\int_{X_{(i-m)}}^{X_{(i+m)}}f(x)\int_{-\infty}^{\infty}f(y|x)dydx+\frac{1}{2m}\sum_{i=n-m+1}^{n}\int_{X_{(i-m)}}^{X_{(i+m)}}f(x)\int_{-\infty}^{\infty}f(y|x)dydx.$$

In order to consider the summands in (3) with $1\le i\le m$, $m+1\le i\le n-m$ and $n-m+1\le i\le n$, corresponding to the sums related to the right side of (3), we define the order statistics

$$Y_{(1:k,j)}\le Y_{(2:k,j)}\le...\le Y_{(j-k+1:k,j)}$$

based on $Y_{[k]},...,Y_{[j]}$ that are the concomitants of $X_{(k)},...,X_{(j)}$, $j\ge k$, respectively. We specify that

$Y_{(r:k,j)}=Y_{(1:k,j)}$, if $r\le 1$ and $Y_{(r:k,j)}=Y_{(j-k+1:k,j)}$, if $r\ge j-k+1$.

In the case of $1\le i\le m$, we have $i+m$ observations $Y_{(1:1,i+m)}\le...\le Y_{(i+m:1,i+m)}$, since it is

defined that $X_{(j)}=X_{(1)}$, if $j\le 1$. Then, by virtue of Lemma 1, for $k<m/2$ and a fixed x, we obtain

$$h_{1,i,m,k}(x)\equiv\frac{1}{2k}\sum_{j=1}^{i+m}\int_{Y_{(j-k:1,i+m)}}^{Y_{(j+k:1,i+m)}}f(y|x)dy\le\int_{-\infty}^{\infty}f(y|x)dy=1.\tag{4}$$

In the case of $m+1\le i\le n-m$, we observe $2m+1$ data points $Y_{(1:i-m,i+m)}\le..\le Y_{(2m+1:i-m,i+m)}$. Then, we have

$$h_{2,i,m,k}(x)\equiv\frac{1}{2k}\sum_{j=1}^{2m+1}\int_{Y_{(j-k:i-m,i+m)}}^{Y_{(j+k:i-m,i+m)}}f(y|x)dy\le\int_{-\infty}^{\infty}f(y|x)dy=1.\tag{5}$$

In the case of $n-m+1\le i\le n$, we observe $n-i+m+1$ data points $Y_{(1:i-m,n)}\le..\le Y_{(n-i+m+1:i-m,n)}$, since it is defined that $X_{(j)}=X_{(n)}$, if $j\ge n$. Then, we have

$$h_{3,i,m,k}(x)\equiv\frac{1}{2k}\sum_{j=1}^{n-i+m+1}\int_{Y_{(j-k:i-m,n)}}^{Y_{(j+k:i-m,n)}}f(y|x)dy\le\int_{-\infty}^{\infty}f(y|x)dy=1.\tag{6}$$

Note that, for any fixed x, $h_{r,i,m,k}(x)\to1, r=1,2,3,$ when $k/m\to0$ as $k,m\to\infty$.

It is clear that Equations (2)-(6) provide the constraint

$$H_{m,k}\le1,\tag{7}$$

where $H_{m,k}$ it defined as

$$H_{m,k} = \frac{1}{2m} \sum_{i=1}^{m} \int_{X_{(i-m)}}^{X_{(i+m)}} f(x) h_{1,i,m,k}(x) dx + \frac{1}{2m} \sum_{i=m+1}^{n-m} \int_{X_{(i-m)}}^{X_{(i+m)}} f(x) h_{2,i,m,k}(x) dx$$

$$+ \frac{1}{2m} \sum_{i=n-m+1}^{n} \int_{X_{(i-m)}}^{X_{(i+m)}} f(x) h_{3,i,m,k}(x) dx.$$

A simple algebra shows that the statistic $H_{m,k}$ can be presented in the form

$$H_{m,k} = \frac{1}{4km} \sum_{i=1}^{m} \sum_{j=1}^{i+m} \int_{X_{(i-m)}}^{X_{(i+m)}} \int_{Y_{(j-k:1,i+m)}}^{Y_{(j+k:1,i+m)}} f(x,y) dxdy$$

$$+ \frac{1}{4km} \sum_{i=m+1}^{n-m} \sum_{j=1}^{2m+1} \int_{X_{(i-m)}}^{X_{(i+m)}} \int_{Y_{(j-k:i-m,i+m)}}^{Y_{(j+k:i-m,i+m)}} f(x,y) dxdy$$

$$+ \frac{1}{4km} \sum_{i=n-m+1}^{n} \sum_{j=1}^{n-i+m+1} \int_{X_{(i-m)}}^{X_{(i+m)}} \int_{Y_{(j-k:i-m,n)}}^{Y_{(j+k:i-m,n)}} f(x,y) dxdy.$$

Now, in accordance with the dbEL-developing procedure introduced in Section 1.1, we will use the mean-value approximations to the integrals in $H_{m,k}$. Let the notation $X_{[r:k,j]}$ state the concomitant of $Y_{(r:k,j)}$. The couple $\left(X_{[r:k,j]}, Y_{(r:k,j)} \right)$ belongs to the data points $\{ (X_i, Y_i), i = 1,..,n \}$ and $f\left(X_{[r:k,j]}, Y_{(r:k,j)} \right)$ has a place in L_f at (1). Consider the following situations with respect to the summands that appear in the definition of $H_{m,k}$.

In the cases with $1 \le i \le m$, $1 \le j \le i + m$, we obtain

$$\int_{X_{(i-m)}}^{X_{(i+m)}} \int_{Y_{(j-k:1,i+m)}}^{Y_{(j+k:1,i+m)}} f(x,y) dxdy \tag{8}$$

$$\simeq f\left(X_{[j:1,i+m]}, Y_{(j:1,i+m)} \right) \left(X_{(i+m)} - X_{(i-m)} \right) \left(Y_{(j+k:1,i+m)} - Y_{(j-k:1,i+m)} \right).$$

Note that, in this framework, it is held that $Y_{(j-k:1,i+m)} \le Y_{(j:1,i+m)} \le Y_{(j+k:1,i+m)}$ and $X_{(i-m)} \le X_{[j:1,i+m]} \le X_{(i+m)}$.

10

In the cases with $m+1 \le i \le n-m$, $1 \le j \le 2m+1$, we have

$$\int_{X_{(i-m)}}^{X_{(i+m)}} \int_{Y_{(j-k:i-m,i+m)}}^{Y_{(j+k:i-m,i+m)}} f(x,y)\,dxdy \tag{9}$$

$$\simeq f\left(X_{[j:i-m,i+m]}, Y_{(j:i-m,i+m)}\right)\left(X_{(i+m)} - X_{(i-m)}\right)\left(Y_{(j+k:i-m,i+m)} - Y_{(j-k:i-m,i+m)}\right).$$

In the cases with $n-m+1 \le i \le n$, $1 \le j \le n-i+m+1$, we have

$$\int_{X_{(i-m)}}^{X_{(i+m)}} \int_{Y_{(j-k:i-m,n)}}^{Y_{(j+k:i-m,n)}} f(x,y)\,dxdy \tag{10}$$

$$\simeq f\left(X_{[j:i-m,n]}, Y_{(j:i-m,n)}\right)\left(X_{(i+m)} - X_{(i-m)}\right)\left(Y_{(j+k:i-m,n)} - Y_{(j-k:i-m,n)}\right).$$

Thus, using (8)-(10), we represent constraint (7) in its empirical form

$$\tilde{H}_{m,k} \le 1, \tag{11}$$

defining

$$\tilde{H}_{m,k} = \frac{1}{4km}\sum_{i=1}^{m}\sum_{j=1}^{i+m} f\left(X_{[j:1,i+m]}, Y_{(j:1,i+m)}\right)\left(X_{(i+m)} - X_{(i-m)}\right)\left(Y_{(j+k:1,i+m)} - Y_{(j-k:1,i+m)}\right)$$

$$+ \frac{1}{4km}\sum_{i=m+1}^{n-m}\sum_{j=1}^{2m+1} f\left(X_{[j:i-m,i+m]}, Y_{(j:i-m,i+m)}\right)\left(X_{(i+m)} - X_{(i-m)}\right)\left(Y_{(j+k:i-m,i+m)} - Y_{(j-k:i-m,i+m)}\right)$$

$$+ \frac{1}{4km}\sum_{i=n-m+1}^{n}\sum_{j=1}^{n-i+m+1} f\left(X_{[j:i-m,n]}, Y_{(j:i-m,n)}\right)\left(X_{(i+m)} - X_{(i-m)}\right)\left(Y_{(j+k:i-m,n)} - Y_{(j-k:i-m,n)}\right).$$

Note that $\tilde{H}_{m,k}$ in (11) is a sum of

$$w = (m+1) + (m+2) + \ldots + 2m + (n-2m)(2m+1) + (m+1) + (m+2) + \ldots + 2m = n + m(2n-m-1) > n$$

different summands and each summand involves one multiplier $f(X_l, Y_l)$ for some $l \in [1, n]$.

Therefore, there are several summands in $\tilde{H}_{m,k}$ with equivalent multipliers $f\left(X_{(l)}, Y_{[l]}\right) = f_l$,

$l = 1, 2, \ldots, n$. This can complicate the use of the Lagrange method for deriving values of $f_i, i = 1, \ldots, n$,

that maximize L_f defined in (1), satisfying (11). Taking into account this issue, we will rewrite $\tilde{H}_{m,k}$

via a sum of n variables with coefficients $f_i, i = 1, \ldots, n$. To this end, we define the rank of the

observation $Y_{[r]}$ with respect to $Y_{[c]}, Y_{[c+1]}, \ldots, Y_{[d]}$ as

11

$$\rho\left(Y_{[r]},c,d\right) = \sum_{l=c}^{d} I\left(Y_{[r]} \le Y_{[l]}\right),$$

where $I(\cdot)$ is the indicator function and $1 \le c \le d \le n$. Then, using some reorganization (see the Supplement, Appendix B, for details), we have

$$\tilde{H}_{m,k} = \frac{1}{4km} \sum_{i=1}^{n} f_i G_{i,m,k}, \tag{12}$$

where $G_{i,m,k}$ are defined in accordance with the following scheme:

(a)   for $1 \le i \le m$,

$$G_{i,m,k} = \sum_{j=1}^{m} \left[Y_{\left(\rho\left(Y_{[i]},1,m+j\right)+k:1,m+j\right)} - Y_{\left(\rho\left(Y_{[i]},1,m+j\right)-k:1,m+j\right)} \right]\left(X_{(m+j)} - X_{(1)}\right)$$

$$+ \sum_{j=1}^{i} \left[Y_{\left(\rho\left(Y_{[i]},j,j+2m\right)+k:j,j+2m\right)} - Y_{\left(\rho\left(Y_{[i]},j,j+2m\right)-k:j,j+2m\right)} \right]\left(X_{(j+2m)} - X_{(j)}\right);$$

(b)   for $m+1 \le i \le 2m$,

$$G_{i,m,k} = \sum_{j=0}^{2m-i} \left[Y_{\left(\rho\left(Y_{[i]},1,i+j\right)+k:1,i+j\right)} - Y_{\left(\rho\left(Y_{[i]},1,i+j\right)-k:1,i+j\right)} \right]\left(X_{(i+j)} - X_{(1)}\right)$$

$$+ \sum_{j=1}^{i} \left[Y_{\left(\rho\left(Y_{[i]},j,j+2m\right)+k:j,j+2m\right)} - Y_{\left(\rho\left(Y_{[i]},j,j+2m\right)-k:j,j+2m\right)} \right]\left(X_{(j+2m)} - X_{(j)}\right);$$

(c)   for $2m+1 \le i \le n-2m$,

$$G_{i,m,k} = \sum_{j=0}^{2m} \left[Y_{\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)+k:i-j,i-j+2m\right)} - Y_{\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)-k:i-j,i-j+2m\right)} \right]\left(X_{(i-j+2m)} - X_{(i-j)}\right);$$

(d)   for $n-2m+1 \le i \le n-m$,

$$G_{i,m,k} = \sum_{j=0}^{i-n+2m} \left[Y_{\left(\rho\left(Y_{[i]},i-j,n\right)+k:i-j,n\right)} - Y_{\left(\rho\left(Y_{[i]},i-j,n\right)-k:i-j,n\right)} \right]\left(X_{(n)} - X_{(i-j)}\right)$$

$$+\sum_{j=1}^{n-i}\left\{\left[Y_{\left(\rho\left(Y_{[i]},n-j-2m,n-j\right)+k:n-j-2m,n-j\right)}-Y_{\left(\rho\left(Y_{[i]},n-j-2m,n-j\right)-k:n-j-2m,n-j\right)}\right)\right.$$

$$\left.\left(X_{(n-j)}-X_{(n-j-2m)}\right)\right\};$$

(e)   for $n-m+1\le i\le n$,

$$G_{i,m,k}=\sum_{j=0}^{m-1}\left[Y_{\left(\rho\left(Y_{[i]},n-m-j,n\right)+k:n-m-j,n\right)}-Y_{\left(\rho\left(Y_{[i]},n-m-j,n\right)-k:n-m-j,n\right)}\right]\left(X_{(n)}-X_{(n-m-j)}\right)$$

$$+\sum_{j=0}^{n-i}\left\{\left[Y_{\left(\rho\left(Y_{[i]},n-j-2m,n-j\right)+k:n-j-2m,n-j\right)}-Y_{\left(\rho\left(Y_{[i]},n-j-2m,n-j\right)-k:n-j-2m,n-j\right)}\right)\right.$$

$$\left.\left(X_{(n-j)}-X_{(n-j-2m)}\right)\right\}.$$

Note that, corresponding to Scenarios (a)-(e) above, the statistic $G_{i,m,k}$ consists of $m+i$, $2m+1$, $2m+1$, $2m+1$ and $m+n-i+1$ different summands, respectively. Then $\tilde{H}_{m,k}$ at (12) includes w different summands that is consistent with definition (11).

According to the dbEL concept, we derive values of f_i, $i=1,...,n$, that maximize the logarithm of the likelihood $L_f=\prod_{i=1}^{n}f_i$ defined in (1), subject to the constraint

$$(4km)^{-1}\sum_{i=1}^{n}f_iG_{i,m,k}\le 1$$

obtained with respect to (11), where $\tilde{H}_{m,k}$ has the form (12). This procedure provides the values

$$f_i=\frac{4mk}{nG_{i,m,k}},\quad i=1,...,n,\tag{13}$$

yielding the approximation

$$\Delta_{n,m,k}=\prod_{i=1}^{n}\frac{4mk}{nG_{i,m,k}}$$

to the likelihood function

$$L_f=\prod_{i=1}^{n}f\left(X_i,Y_i\right),$$

where $n^\delta \le m \le n^{1-\delta}$ and $m^\delta \le k \le m^{1-\delta}$ with $0 < \delta < 0.5$. Employing the maximum likelihood technique described in Section 1.1, we conclude that the dbEL approximation to L_f is

$$\Delta_n = \min_{n^\delta \le m \le n^{1-\delta}} \; \min_{m^\delta \le k \le m^{1-\delta}} \Delta_{n,m,k}, \tag{14}$$

where $0 < \delta < 0.5$. Note that, in contrast to the univariate dbEL approach shown in Section 1.1, it is required that $m \ge n^\delta$ and $k \ge m^\delta$. Explanations regarding these restrictions are provided in the sections below.

3.   A bivariate version of the variable partition histogram

In this section, we demonstrate that the proposed method satisfies the principle of the bivariate histogram construction in the context of a maximum likelihood estimation of the density function $f(X,Y)$ (Kim and van Ryzin [11]; van Ryzin [29]; Prakasa Rao [23: pp. 234-235]; Izenman [9: pp. 209-210, 212-213]).

Consider, for example, Scenario (c) in the definition (12), where, for $i \in [2m+1, n-2m]$,

$$G_{i,m,k} = \sum_{j=0}^{2m} \left(Y_{\left(\rho\left(Y_{[i]}, i-j, i-j+2m\right)+k:i-j, i-j+2m\right)} - Y_{\left(\rho\left(Y_{[i]}, i-j, i-j+2m\right)-k:i-j, i-j+2m\right)} \right) \left(X_{(i-j+2m)} - X_{(i-j)} \right)$$

consists of $2m+1$ summands. Taking into account the formal notations used in Kim and van Ryzin [11], we denote the statistics

$$\hat{f}_{n,j} = \frac{\left(D_{n,j} - B_{n,j} \right)}{n \left(Y_{\left(D_{n,j}: i-j, i-j+2m\right)} - Y_{\left(B_{n,j}: i-j, i-j+2m\right)} \right) \left(X_{\left(C_{n,j} \right)} - X_{\left(A_{n,j} \right)} \right)}, \quad j = 0,...,2m,$$

where $A_{n,j} = i-j$, $C_{n,j} = i-j+2m$, $B_{n,j} = \rho\left(Y_{[i]}, i-j, i-j+2m\right)-k$ $D_{n,j} = \rho\left(Y_{[i]}, i-j, i-j+2m\right)+k.$, The

statistics $\hat{f}_{n,j}$, $j=0,...,2m$, are consistent approximations to $f\left(X_{(i)}, Y_{[i]} \right)$, $2m+1 \le i \le n-2m$,

uniformly for $j=0,...,2m$, if $A_{n,j}$, $B_{n,j}$, $C_{n,j}$ and $D_{n,j}$ satisfy the conditions presented in Kim and

van Ryzin [11]. In this context, we note that, for $n^\delta \le m \le n^{1-\delta}$ and $n \to \infty$,

$$Pr\left(0 \le A_{n,j} < C_{n,j} \le n, \; X_{\left(A_{n,j} \right)} \le X_{(i)} < X_{\left(C_{n,j} \right)} \right) = 1; \; C_{n,j} - A_{n,j} = 2m \to \infty;$$

$\left(C_{n,j} - A_{n,j} \right)/n \to 0$; $A_{n,j}$ and $C_{n,j}$ are invariant under permutations of $\left(X_r, Y_r \right), r=1,...,n$, for $X_{(i)}$.

Regarding the positive integer-valued and indexing random variables $B_{n,j}$ and $D_{n,j}$, we have

14

$$Pr\left(0 \le B'_{n,j} < D'_{n,j} \le C_{n,j} - A_{n,j} + 1 = 2m+1, Y_{\left(B_{n,j}:i-j,i-j+2m\right)} \le Y_{[i]} < Y_{\left(D_{n,j}:i-j,i-j+2m\right)}\right) = 1,$$

where $D'_{n,j} = D_{n,j}I\left(D_{n,j} \le 2m+1\right) + (2m+1)I\left(D_{n,j} > 2m+1\right)$ and $B'_{n,j} = B_{n,j}I\left(B_{n,j} \ge 1\right) + I\left(B_{n,j} < 1\right)$

corresponding to the definition of subscripts $\left(D_{n,j}:i-j,i-j+2m\right)$ and $\left(B_{n,j}:i-j,i-j+2m\right)$ of Y's, since the

order statistics $Y_{\left(1:i-1,i-j+m\right)} \le Y_{\left(2:i-1,i-j+m\right)} \le ... \le Y_{\left(2m+1:i-1,i-j+m\right)}$ are based on

$Y_{[i-j]}, Y_{[i-j+1]}, ..., Y_{[i-j+2m]}$. It is clear that, for $n^\delta \le m \le n^{1-\delta}$ and $m^\delta \le k \le m^{1-\delta}$, $D_{n,j} - B_{n,j} = 2k \to \infty$,

$\left(B_{n,j} - D_{n,j}\right)/\left(C_{n,j} - A_{n,j}\right) \to 0$ as well as $D_{n,j}$ and $B_{n,j}$ are invariant under permutation of

$\left(X_r, Y_r\right), r = 1, ..., n,$, $i = 1, 2, ..., n$ for $\left(X_{(i)}, Y_{[i]}\right)$.

Now, requiring $m \ge n^\delta$ and $k \ge m^\delta$, we obtain that

$$log(n)/(D_{n,j} - B_{n,j}) = log(n)/(2k) \le log(n)/(2m^\delta) \le log(n)/(2n^{2\delta}) \to 0 \text{ as } n \to \infty.$$

Thus, the theoretical arguments shown in Kim and van Ryzin [11] provide that, for all $j \in [0, 2m]$,

$1/\hat{f}_{n,j} \to 1/f\left(X_{(i)}, Y_{[i]}\right)$. This implies

$$\left(2m\right)^{-1} \sum_{j=0}^{2m}\left(1/\hat{f}_{n,j}\right) \to 1/f\left(X_{(i)}, Y_{[i]}\right),$$

concluding that

$$4mk\left(nG_{i,m,k}\right)^{-1} \to f\left(X_{(i)}, Y_{[i]}\right), \; i \in [2m+1, n-2m], \text{ as } n \to \infty.$$

4. An asymptotic consistency of the bivariate density based empirical likelihood

Here we confine the density function $f(x, y)$ to be continuous and bounded on its support,

$a_1 < f(x, y) < a_2$, where $0 < a_1 < a_2 < \infty$ are fixed constants. Then,

$$n^{-1} log\left(\varDelta_n\right) \xrightarrow{p} E\left\{log\left(f\left(X_1, Y_1\right)\right)\right\} \text{ as } n \to \infty,$$

where the dbEL, $\varDelta_n$, is defined by (14).

The proof of the result above is based on a formal scheme in a manner, which can be associated with the explanations mentioned in Section 3. We use the theoretical arguments shown in [11], [29] and [37] to present the rigorous proof of the dbEL consistency in the Supplement (Appendix C).

We employed various Monte Carlo evaluations based on more than one hundred different scenarios of (X,Y)-distributions and a variety of sample sizes n in order to examine a critical necessity of the condition $a_1 < f(x,y) < a_2$ for the asymptotic result shown in this section. These studies demonstrated that the bivariate dbEL approach is consistent in more general cases of $f(x,y)$-forms.

5. The dbEL ratio tests for bivariate normality

Should the underlying data follow the density function

$$f(x,y) = \frac{0.5}{\pi\sigma_x\sigma_y\left(1-\rho^2\right)^{1/2}} exp\left[-\frac{0.5}{\left(1-\rho^2\right)}\left\{\frac{\left(x-\mu_x\right)^2}{\sigma_x^2} - 2\rho\frac{\left(x-\mu_x\right)\left(y-\mu_y\right)}{\sigma_x\sigma_y} + \frac{\left(y-\mu_y\right)^2}{\sigma_x^2}\right\}\right],$$

where the parameters $\mu_x = E(X), \mu_y = E(Y), \sigma_x^2 = var(X), \sigma_y^2 = var(Y)$ and

$\rho = E\left(\left(X-\mu_x\right)(Y-\mu_y)\right)/\left(\sigma_x\sigma_y\right) \in (0,1)$, the maximum log likelihood function is

$-n\,log\left\{2\pi\hat{\sigma}_x\hat{\sigma}_y\left(1-\hat{\rho}^2\right)\right\} - n$ with $\left(\hat{\sigma}_x\right)^2 = n^{-1}\sum_{i=1}^{n}\left(X_i - n^{-1}\sum_{j=1}^{n}X_j\right)^2$, $\left(\hat{\sigma}_y\right)^2 =$

$n^{-1}\sum_{i=1}^{n}\left(Y_i - n^{-1}\sum_{j=1}^{n}Y_j\right)^2$ and $\hat{\rho} = \left(n\hat{\sigma}_x\hat{\sigma}_y\right)^{-1}\sum_{i=1}^{n}\left(X_i - n^{-1}\sum_{j=1}^{n}X_j\right)\left(Y_i - n^{-1}\sum_{j=1}^{n}Y_j\right)$

(Balakrishnan and Lai [2: p. 490]). Then we state the log dbEL ratio test that rejects the null hypothesis iff

$$V_n^1 \equiv log\left(\Delta_n\right) + n\,log\left\{2\pi\hat{\sigma}_x\hat{\sigma}_y\left(1-\hat{\rho}^2\right)\right\} + n > C, \tag{15}$$

where the statistic Δ_n is defined in (14) and C is a test-threshold.

It is clear that the transformation $\left(\left(X_i - \mu_x\right)/\sigma_x, \left(Y_i - \mu_y\right)/\sigma_y\right)^T, i = 1,...,n,$ of the data does not change a value of the statistic V_n^1. That is to say the null distribution of V_n^1 is unaltered with respect to the parameters $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2$. However, the H_0 – distribution of V_n^1 depends on ρ and one can easily show that the probability

$$Pr\left\{V_n^1 > C / \left(X_i, Y_i\right)^T \sim N_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right), i = 1,...,n\right\}$$

decreases when $|\rho| \in (0,1)$ increases. We will evaluate this fact in the next section.

Various efficient statistics applied for testing univariate normality have forms that distributed independently of the mean and variance of observations under the null hypothesis. For example, the

distribution of the statistic T_n defined in Section 1.1 does not depend on μ_x and σ_x when

$X_1 \sim N_1(\mu_x, \sigma_x)$. Then, in order to calculate the critical values of the T_n based test for $X_i \sim$

$N_1(\mu_x, \sigma_x)$, $i = 1,...,n$, one can use pre-tabulated critical points or/and Monte Carlo simulations

without restricting the sample size n to be relatively large.

Unfortunately, in the bivariate case, when we construct appropriate test statistics, the property

mentioned above cannot be held in many scenarios. In this framework, the conventional approach is to

standardize the bivariate data $(X_i, Y_i)^T$, $i = 1,...,n$, obtaining the transformed data

$$\left(X'_i, Y'_i\right)^T = S_n^{-0.5}\left(X_i - n^{-1}\sum_j^n X_j, Y_i - n^{-1}\sum_j^n Y_j\right), i = 1,...,n,$$

where $S_n^{-0.5}$ is the square root of the inverse of the estimated covariance matrix. Then, under the null

hypothesis, $\left(X'_1, Y'_1\right)^T$ has approximately the bivariate standard normal distribution

$N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$. In this context, in addition to the relevant literature mentioned in Section 1, we

would like to refer the reader to Looney [18], Henze and Zirkler [8], Lee et al. [15], Villaseñor-Alva

and González-Estrada [36]. For example, the widely-applied R procedure (R Development Core Team

[24]) "*mvShapiro.Test*" for Shapiro-Wilk type testing bivariate normality is based on the principle

shown above. Note that applications of the code "*mvShapiro.Test*" are restricted by the requirement

$n \in [12,5000]$ related to the use of tabulated values of the theoretical expectations of order statistics in

Shapiro-Wilk's manner for testing normality.

We may then propose the dbEL ratio test that rejects the null hypothesis iff

$$V_n^2 \equiv log\left(\Delta_n\right) - n\,log(1/(2\pi)) + n > C, \tag{16}$$

where the statistic Δ_n by (14) is calculated employing the data $\left(X'_i, Y'_i\right)^T$, $i = 1,...,n$, instead of

$(X_i, Y_i)^T$, $i = 1,...,n$, the notation $n\,log(1/(2\pi)) - n$ corresponds to the approximate likelihood function

of $\left(X'_i, Y'_i\right)^T$, $i = 1,...,n$, under the null hypothesis, and C is a test-threshold.

Remark. The dbEL literature shows that the power of dbEL tests does not depend significantly on

values of parameters that have roles similar to that of δ at definition (14) (e.g., Tsai et al. [27]; Vexler

et al. [33, 35]). In this context, we note that extensive Monte Carlo simulations confirmed the

robustness of the proposed tests with respect to the values of δ at (14), i.e. one can show that the

power of the new tests does not depend significantly on values of $\delta \in (0,0.5)$ under various scenarios of alternative distributions. Thus, without loss of generality, we set δ in the proposed test statistics to be 0.4.

5.1. Null distributions of the dbEL ratio tests

In the one-dimensional setting, a very substantial body of literature has now grown around the asymptotic distribution problems involving the Vasicek entropy type statistics. One can generally recognize that proofs regarding the asymptotic distribution of the statistic T_{mn} defined in Section 1.1 are analytically very complicated. Note that, when the sample size is relatively large, we can also anticipate that various tests provide very powerful inference. Thus, following the recent literature related to goodness-of-fit tests (e.g., Hall and Welsh [6]; Mudholkar and Tian [20, 21]), we will focus on finite sample sizes without attempting to provide here an asymptotic solution for the critical values for the proposed tests in the two-dimensional setting.

The critical values for the dbEL ratio tests can be accurately approximated using Monte Carlo techniques. In order to tabulate the percentiles of the null distributions of the test statistics V_n^1 and V_n^2 with $\delta = 0.4$ in definition (14), we drew 50,000 samples of $(X_i, Y_i)^T \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), i = 1,...,n,$ calculating values of $V_n^j, j = 1,2,$ at each sample size n. The generated values of the test statistics $V_n^j, j = 1,2,$ were used to determine the critical values C_α of the corresponding null distributions of $V_n^j, j = 1,2,$ at the significance level α. The results of this Monte Carlo study are presented in Tables 1 and 2.

Table 1. *Critical Values, C_α, of the Test Statistic V_n^1*

Sample size	α				
n	0.1	0.05	0.04	0.025	0.01
15	13.78831	14.45392	14.67231	15.10617	15.89147
20	16.01282	16.82075	17.06102	17.56599	18.44998
25	17.79228	18.67010	18.93496	19.48494	20.54437
30	19.34573	20.38065	20.67776	21.28747	22.37941
35	20.88404	21.99563	22.33991	23.01375	24.19156
40	22.07930	23.26584	23.64351	24.36966	25.67781

45	23.30695	24.54427	24.90769	25.70937	27.06523
50	24.67166	26.00287	26.34771	27.08118	28.43204
55	25.53667	26.91676	27.33220	28.19032	29.63046
60	26.44042	27.77653	28.22558	29.12152	30.66377
70	28.04247	29.53837	29.96516	30.84174	32.44158
80	29.36747	31.02797	31.51445	32.44036	34.09259
90	30.75603	32.51081	33.05028	34.05678	35.57193
100	32.07309	33.76989	34.37907	35.40567	37.31336
120	34.02814	36.02057	36.52973	37.70851	39.86061

Table 2. *Critical Values, C_{α}, of the Test Statistic V_n^2*

Sample size	α				
n	0.1	0.05	0.04	0.025	0.01
15	13.96789	14.68150	14.85996	15.30876	16.03354
20	16.22554	17.06107	17.31668	17.77547	18.72344
25	17.94104	18.90434	19.19077	19.72435	20.72166
30	19.47291	20.53942	20.78474	21.38950	22.44172
35	21.31563	22.44544	22.77901	23.46965	24.70225
40	22.29129	23.48504	23.82955	24.50040	25.67661
45	23.52354	24.73849	25.09407	25.84452	27.13543
50	24.77544	26.05601	26.43748	27.21586	28.60631
55	25.65505	26.96317	27.35342	28.14286	29.63217
60	26.52463	27.98900	28.42992	29.29108	30.66241
70	28.15409	29.68741	30.12672	31.04799	32.67359
80	29.46820	31.10201	31.56947	32.42746	34.07674
90	30.92357	32.62724	33.13130	34.23576	36.06089
100	32.00653	33.72502	34.26569	35.32767	37.10607
120	34.01007	35.96420	36.54560	37.82717	39.62449

In order to verify the results shown in Tables 1 and 2, for different values of $\rho \in (-1,1)$ and n, we calculated the Monte Carlo approximations to

$$Pr\left\{V_n^j > C_{0.05} / \left(X_i, Y_i\right)^T \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), i = 1,...,n\right\}, j = 1, 2,$$

where $C_{\alpha=0.05}$'s are shown in Tables 1 and 2. In this study, we also analyzed the Shapiro-Wilk test (SW), using the R procedure "*mvShapiro.Test*". For each value of ρ and n, the Type I error rates were derived using 25,000 samples of $\left(X_i, Y_i\right)^T \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), i = 1,...,n$. Table 3 presents the results of this Monte Carlo evaluation.

Table 3. *The Monte Carlo Type I error probabilities of the proposed tests (15), (16) and the Shapiro-Wilk test (SW), when $\left(X_i, Y_i\right)^T \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), i = 1,...,n$ and the anticipated significance level is $\alpha = 0.05$.*

ρ	$n = 35$			$n = 50$			$n = 70$		
	V_n^1	V_n^2	SW	V_n^1	V_n^2	SW	V_n^1	V_n^2	SW
-0.9	<0.0001	0.0475	0.0521	<0.0001	0.0496	0.0519	<0.0001	0.0513	0.0506
-0.8	0.0015	0.0504	0.0501	0.0014	0.0495	0.0492	0.0014	0.0477	0.0482
-0.7	0.0073	0.0501	0.0501	0.0069	0.0498	0.0480	0.0067	0.0499	0.0466
-0.6	0.0143	0.0505	0.0489	0.0139	0.0499	0.0521	0.0143	0.0479	0.0495
-0.5	0.0235	0.0511	0.0520	0.0233	0.0495	0.0521	0.0228	0.0498	0.0481
-0.4	0.0325	0.0513	0.0507	0.0319	0.0500	0.0519	0.0322	0.0495	0.0498
-0.3	0.0388	0.0509	0.0521	0.0407	0.0500	0.0493	0.0394	0.0500	0.0506
-0.2	0.0454	0.0509	0.0517	0.0465	0.0501	0.0516	0.0475	0.0495	0.0497
-0.1	0.0506	0.0461	0.0526	0.0499	0.0499	0.0489	0.0507	0.0503	0.0496
0	0.0488	0.0501	0.0500	0.0508	0.0499	0.0493	0.0504	0.0495	0.0501
0.1	0.0501	0.0473	0.0505	0.0500	0.0454	0.0510	0.0503	0.0475	0.0466
0.2	0.0466	0.0508	0.0521	0.0470	0.0499	0.0493	0.0469	0.0490	0.0508
0.3	0.0418	0.0507	0.0519	0.0400	0.0499	0.0519	0.0380	0.0493	0.0470
0.4	0.0335	0.0512	0.0509	0.0317	0.0501	0.0497	0.0323	0.0492	0.0506
0.5	0.0237	0.0510	0.0515	0.0228	0.0496	0.0481	0.0231	0.0499	0.0514
0.6	0.0143	0.0503	0.0488	0.0138	0.0498	0.0521	0.0142	0.0498	0.0497
0.7	0.0072	0.0498	0.0495	0.0071	0.0498	0.0492	0.0068	0.0489	0.0492
0.8	0.0016	0.0502	0.0498	0.0015	0.0496	0.0519	0.0012	0.0483	0.0489
0.9	<0.0001	0.0500	0.0518	<0.0001	0.0497	0.0519	<0.0001	0.0508	0.0488

According to Table 3, the validity of the critical values related to the test statistic V_n^2 is experimentally confirmed. Although the test based on V_n^1 is very conservative when $|\rho| > 0.7$, we can recommend the test (15) to be applied in practice, owing to high levels of the power of this test against alternatives considered in Section 6.

Remark. In practice, in order to implement the proposed approach in a simple and rapid manner, one can suggest applying a hybrid method for computing the p-values of the tests (15), (16), by combining Monte Carlo simulations and the critical values displayed in Tables 1 and 2. In this framework, employing Bayesian type procedures, we can derive relevant information from the Monte Carlo experiments via likelihood type functions, whereas the tabulated critical values can be used to reflect prior distributions (Vexler et al. [34]). The hybrid technique for computing the p-values has been employed in STATA and R statistical packages (Vexler et al. [33]).

6.   Power of the tests

It is clear that in the nonparametric setting of testing bivariate normality, there are no most powerful decision making procedures. In this section, we only exemplify several scenarios where the power of the proposed tests is compared with that of the Shapiro-Wilk (SW) test and classical Mardia's test (MT) for bivariate normality at the significance level of 5%. The following scenarios of source distributions were treated:

(A) $X_1,\ldots,X_n \sim N_1(0,1)$ and $Y_i = \tau_i X_i$, $i=1,\ldots n$, where random variables $\tau_i = -1$ or 1, $i=1,\ldots,n$, are i.i.d. and independent of $X_1,\ldots,X_n$ with $Pr(\tau_1 = -1) = 0.5$. In this case, $X_1,\ldots,X_n \sim N_1(0,1)$ and $Y_1,\ldots,Y_n \sim N_1(0,1)$, but $(X_1,Y_1),\ldots,(X_n,Y_n)$ are not bivariate normal.

(B) $X_1,\ldots,X_n$ are uniformly distributed over (-5,5) and independent of $Y_1,\ldots,Y_n$ that are uniformly distributed over (-5,5). This case presents a light tailed alternative distribution.

(C) Let $\xi_1,\ldots,\xi_n$ and $\eta_1,\ldots,\eta_n$ be i.i.d. random variables from $N_1(0,1)$. Define ($X_i = \xi_i + \xi_i^4 / 25$, $Y_i = \eta_i X_i$), $i = 1,\ldots,n$, to examine a case of heavy tailed alternative distributions.

(D) Define ($X_i = \sum_{j=1}^{7} \xi_{ij} / 7^{1/2}$, $Y_i = \sum_{j=1}^{7} \eta_{ij} / 7^{1/2}$), $i = 1,\ldots,n$, with independent and identically $U(0,1)$-distributed random variables ξ's and η's to evaluate a case in which the central limit theorem can be applied to approximate the data distribution.

(E) (X,Y)'s follow Morgenstern's distribution with parameter $\alpha = 0.5$ (Johnson [10: p. 185]).

(F) (X,Y)'s follow Plackett's distribution with parameter $\psi = 2$ (Johnson [10: p. 193]).

(G) $(X_1,Y_1),\ldots,(X_n,Y_n)$ follow Gumbel's Type I logistic distribution (Johnson [10: p. 199]).

(H) Assume random vectors $(Z_1,W_1)^T,\ldots,(Z_n,W_n)^T$ are from Gumbel's bivariate exponential distribution with parameter $\theta = 0.9$ (Johnson [10: p. 197]). In order to reduce the power of the considered tests, we define $(X_i = Z_i + \xi_i, Y_i = W_i + \eta_i)$, where ξ's and η's are independent and identically $U(-6,3)$-distributed random variables.

(I) Define $(X_i = \xi_i,\ Y_i = \eta_i), i = 1,\ldots,n,$ where ξ's and η's are independent and identically $Gamma(2,1)$-distributed random variables. Note that, commonly, sample entropy based tests for univariate normality do not outperform the corresponding Shapiro-Wilk test when underlying data are from a gamma distribution (e.g., Table 2 in Vasicek [30], where the case with $X \sim Gamma(2,1)$ is evaluated).

Table 4 shows the results of the power evaluations of the proposed tests (V_n^1 and V_n^2 with $\delta = 0.4$ in definition (14)), the Shapiro-Wilk (SW) test and classical Mardia's test (MT) for bivariate normality via the Monte Carlo study based on 15,000 replications of $(X_1,Y_1),\ldots,(X_n,Y_n)$ for the designs (A-I) given above at each sample size $n = 35, 50, 70$.

Table 4. *The Monte Carlo power of the tests at the significance level of 5%.*

Tests	Design (A)			Design (B)			Design (C)		
	Sample size (n)			Sample size (n)			Sample size (n)		
	35	50	70	35	50	70	35	50	70
V_n^1	0.5645	0.8327	0.9427	0.9295	0.9923	1	0.8228	0.9635	0.9945
V_n^2	0.6666	0.9249	0.9796	0.8907	0.9888	1	0.7771	0.9458	0.9914
SW	0.0570	0.0604	0.0613	0.6975	0.9316	0.9952	0.7744	0.9282	0.9878
MT	0.3711	0.3881	0.4108	0.0004	0.0002	0.0002	0.7392	0.8350	0.9156
	Design (D)			Design (E)			Design (F)		
V_n^1	0.0639	0.0643	0.0753	0.9226	0.9928	0.9998	0.9908	0.9998	1
V_n^2	0.0533	0.0639	0.0693	0.8748	0.9876	0.9994	0.9739	0.9990	1
SW	0.0375	0.0378	0.0381	0.6544	0.9098	0.9906	0.8987	0.9910	0.9999
MT	0.0230	0.0251	0.0257	0.0000	0.0001	0.0001	0.0033	0.0045	0.0049

	Design (G)			Design (H)			Design (I)		
V_n^1	1	1	1	0.4581	0.6234	0.7752	0.8959	0.9766	0.9969
V_n^2	0.9994	1	1	0.4160	0.6184	0.7796	0.8521	0.9745	0.9964
SW	0.8812	0.9801	0.9980	0.3783	0.5582	0.7535	0.9615	0.9972	0.9999
MT	0.9950	0.9997	0.9998	0.2043	0.3004	0.4178	0.8062	0.9527	0.9959

This study demonstrates that the dbEL ratio tests are superior to the considered classical tests under the designs (A-H). The new tests have significantly improved powers as compared to the corresponding classical procedures. For example, the power of the proposed tests is roughly two times larger than that of the classical tests given scenarios (A) and (D). The proposed tests have approximately 10%-30% power gains as compared to the classical procedures when $n = 35$ in scenarios (E) and (H). It seems that the Shapiro Wilk test is not efficient under the designs of (A) and (D). In scenario (D), the Shapiro Wilk test is biased. Mardia's test is biased under the designs of (B), (D), (E), and (F). In these scenarios, the new tests exhibit high and stable power characteristics. The proposed tests perform reasonably well, and are generally competitive with the classical tests in case (I). In this scenario, it is anticipated that the Shapiro Wilk test has higher power than the other considered tests. In parallel with studies regarding properties of tests for univariate normality, the shown Monte Carlo results are consistent with those related to one-dimensional sample entropy based tests (e.g., Vasicek [30]]).

7. Data analysis

Myocardial infarction is commonly caused by blood clots blocking the blood flow of the heart leading heart muscle injury. The heart disease is leading cause of death affecting about or higher than 20% of populations regardless of different ethnicities according to the Centers for Disease Control and Prevention (e.g., Schisterman et al. [25, 26]).

We illustrate the application of the proposed approach based on a sample from a study that evaluates biomarkers associated with myocardial infarction (MI). The study was focused on the residents of Erie and Niagara counties, 35-79 years of age. 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 years, while the elderly sample (age 65-79) was randomly chosen from the Health Care Financing Administration database. The biomarkers "high density lipoprotein (HDL)-cholesterol" and "vitamin E" are often used as a discriminant factor between individuals with (MI=1) and without (MI=0) myocardial infarction disease (e.g., Schisterman et al. [25, 26]). The HDL-cholesterol levels were

examined from a 12-hour fasting blood specimen for biochemical analysis at baseline. A total of 240 measurements of the biomarkers were evaluated by the study. The sample of 120 biomarkers values was collected on cases who survived on MI and the sample of 120 measurements on controls who had no previous MI.

Oftentimes, measurements related to biological processes follow a log-normal distribution (e.g., Limpert et al. [17]). The aim of this study is to investigate the joint distribution of log-transformed vitamin E measurements, say X, and log-transformed (HDL)-cholesterol measurements, say Y, with regard to MI disease. Towards this end, we implemented the new tests, the Shapiro-Wilk test (SW) and Mardia's test (MT) for bivariate normality using the data described above. Figure S1 in the Supplement depicts the histograms based on values of X, Y and the scatter plots based on (X,Y) for the case (MI=1) and control (MI=0) groups, respectively.

In this study, the considered four tests provided p-values<0.045, rejecting the hypotheses that the observations $(X_1,Y_1),...,(X_{120},Y_{120})$ are bivariate normally distributed (H_0) for the case (MI=1) and control (MI=0) groups, respectively. Then, we organized a bootstrap/Jackknife type study to examine the power performances of the test statistics. The conducted strategy was that samples with sizes 35, 50 and 70 were randomly selected from the "vitamin E/ HDL-cholesterol" data to be tested for bivariate normality at 5% level of significance. We repeated this strategy 5000 times calculating the frequencies of the events $\{V_n^1$ rejects $H_0\}$, $\{V_n^2$ rejects $H_0\}$, $\{SW$ rejects $H_0\}$ and $\{MT$ rejects $H_0\}$. The test statistics V_n^1 and V_n^2 were performed with $\delta = 0.4$ in definition (14). The obtained experimental powers of the four tests are shown in Table 5.

Table 5. *The experimental powers of the tests at the significance level of 5%.*

Tests	MI=1			MI=0		
	Sample size (*n*)			Sample size (*n*)		
	35	50	70	35	50	70
V_n^1	0.2780	0.3838	0.4716	0.3904	0.4504	0.6442
V_n^2	0.1990	0.3108	0.4277	0.2799	0.4178	0.6189
SW	0.1152	0.1896	0.3119	0.2602	0.3812	0.5910
MT	0.0940	0.1580	0.2369	0.0510	0.0786	0.1544

In this study, the proposed tests significantly outperform the SW and MT tests in terms of the power properties when detecting that the log-transformed biomarkers' values are not jointly distributed as bivariate normal random variables. For example, when $n = 35$ and MI=1, the dbEL ratio tests reveal the experimental powers that are approximately two times larger than those of the SW and MT tests. That is, the dbEL ratio tests are more sensitive as compared with the known methods to rejecting the null hypothesis of bivariate normality regarding joint distributions of the log-transformed values of the "vitamin E" and "HDL-cholesterol" biomarkers.

8. Concluding remarks

In this paper, we extended the density based empirical likelihood approach to construct new goodness of fit tests for bivariate normality. The main idea of our method was to propose a consistent technique that employs histogram/sample entropies density based estimations in the bivariate framework. We compared the performance of the dbEL ratio tests to the known decision making procedures, the Shapiro-Wilk test and Mardia's test. The conducted simulation study displayed that the proposed tests outperformed the known tests in many important scenarios of alternative distributions as well as the new tests provided power levels in a similar manner to their univariate sample entropy based analogs. Finally, we applied our tests on a real data set, where the proposed technique exhibited high and stable power characteristics.

Certainly, the proposed testing strategy is computational intensive. In this context, we would like to note that the known principles regarding bivariate histogram developments deal with strong computational requirements in general. In the modern age we are generally no longer constrained by computational issues and have a greater flexibility in terms of the statistical approaches that we may employ to data analysis problems. Advances in computation and the fast and cheap computational facilities now are available to statisticians. This can support that the dbEL methodology can be suggested to be modified and extended in order to be applied to various multivariate problems encountered in statistical studies.

Our main objective of this paper is twofold: (1) to show that the density based empirical likelihood technique can be a valuable tool in multivariate statistical analysis and (2) to convince readers of the usefulness of the sample entropy based approach that should be more widely investigated in multivariate frameworks.

Acknowledgments

Appendix. Supplementary data

Supplementary material related to this article can be found online

References

[1]   I. Arizono and H. Ohta, A test for normality based on Kullback-Leibler information. The American Statistician 43 (1989) 20-22.

[2]   N. Balakrishnan and C.-D. Lai, Continuous Bivariate Distributions. Springer, New York, 2009.

[3]   T. B. Berrett, R. J. Samworth and M. Yuan. Efficient multivariate entropy estimation via k-nearest neighbour distances. arXiv:1606.00304 (https://arxiv.org/abs/1606.00304), 2017.

[4]   C. Carolan and R. Dykstra. Asymptotic behavior of the Grenander estimator at density flat regions. The Canadian Journal of Statistics 27 (1999) 557-566.

[5]   S. Efromovich. Nonparametric Curve Estimation: Methods, Theory, and Applications. Springer-Verlag, New York, 1999.

[6]   P. Hall and A.H. Welsh. A test for normality based on the empirical characteristic function. Biometrika 70 (1983) 485-489.

[7]   D. M. Hawkins. A new test for multivariate normality and homoscedasticity. Technometrics 23:1 (1981) 105-110.

[8]   N. Henze and B. Zirkler. A Class of Invariant Consistent Tests for Multivariate Normality. Communications in Statistics-Theory and Method 19 (1990) 3595-3618.

[9]   A. J. Izenman. Resent developments in nonparametric density estimation. Journal of the American Statistical Association 86 (1991) 205-224.

[10]   M. E. Johnson. Multivariate Statistical Simulationa. John Wiley & Sons, New York, 1987.

[11]   B. K. Kim, J. Van Ryzin. A bivariate histogram density estimator: consistency and asymptotic normality. Statistics & Probability Letters 3 (1985) 167-173.

[12]   L. F. Kozachenko and N. N. Leonenko. Sample estimate of the entropy of a random vector. Probl. Inform. Transm. 23 (1987) 95–101.

[13]   C. J. Kowalski. The performance of some rough tests for bivariate normality before and after coordinate transformations to normality. Technometrics 12:3 (1970) 517-544.

[14]   N. Lazar and P. A. Mykland. An Evaluation of the Power and Conditionality Properties of Empirical Likelihood. Biometrika 85 (998) 523-534.

[15]   R. Lee, M. Qian and Y. Z. Shao. On Rotational Robustness of Shapiro-Wilk Type Tests for Multivariate Normality. Open Journal of Statistics 4 (2014) 964-969.

[16]   E. L. Lehmann and J. P. Romano. Testing Statistical Hypotheses. Springer-Verlag, New York, 2005.

[17]   E. Limpert, W. A. Stahel and M. Abbt. Log-Normal Distributions across the Sciences: Keys and Clues on the Charms of Statistics, and How Mechanical Models Resembling Gambling Machines Offer a Link to a Handy Way to Characterize Log-Normal Distributions, Which Can Provide Deeper Insight into Variability and Probability—Normal or Log-Normal: That Is the Question. BioScience. 51 (2001) 341-352.

[18]   S. W. Looney. How to Use Tests for Univariate Normality to Assess Multivariate Normality. The American Statistician 49 (1995) 64-70.

[19]   C. J. Mecklin and D. J. Mundfrom. An appraisal and bibliography of tests for multivariate normality. International Statistical Review 72 (2004) 123–138.

[20]   G. S. Mudholkar and L. Tian. An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test. Journal of Statistical Planning and Inference 102 (2002) 211–221.

[21]   G. S. Mudholkar and L. Tian. A test for homogeneity of ordered means of inverse Gaussian populations. Journal of Statistical Planning and Inference. 118 (2004) 37–49.

[22]   A. B. Owen. Empirical Likelihood. CRC press, Florida, 2001.

[23]   B. L. S. Prakasa Rao. Nonparametric functional estimation. Accademic Press, Inc., New York, 1983.

[24]   R Development Core Team. R: A language and environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. http://www.R-project.org

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

[26]   E. F. Schisterman, D. Faraggi, R. Browne, J. Freudenheim, J. Dorn, P. Muti, D. Armstrong, B. Reiser and M. Trevisan. Minimal and best linear combination of oxidative stress and antioxidant biomarkers to discriminate cardiovascular disease. Nutrition, Metabolism, and Cardiovascular Disease 12 (2002) 259-266.

[27]   W-M Tsai, A. Vexler and G. Gurevich. An extensive power evaluation of a novel two-sample density-based empirical likelihood ratio test for paired data with an application to a treatment study of attention-deficit/hyperactivity disorder and severe mood dysregulation. Journal of Applied Statistics. 40 (2013) 1189-1208.

[28]   G. Tusnady. On asymptotically optimal tests. Ann. Statist. 5 (1977) 385-393.

[29]   J. Van Ryzin. A histogram method of density estimation, Communications in Statistics 2:64 (1973) 93-506.

[30]   O. Vasicek. A Test for Normality Based on Sample Entropy. Journal of the Royal Statistical Society. Series B (Methodological) 139 (1976) 54-59.

[31]   A. Vexler and G. Gurevich, G. Empirical Likelihood Ratios Applied to Goodness-of-Fit Tests Based on Sample Entropy. Computational Statistics & Data Analysis 54 (2010) 531-545.

[32]   A. Vexler, G. Gurevich and A. D. Hutson. An Exact Density-Based Empirical Likelihood Ratio Test for Paired Data. Journal of Statistical Planning and Inference 143(2) (2013) 334-345.

[33]   A. Vexler, A. D. Hutson and X. Chen. Statistical Testing Strategies in the Health Sciences. Chapman & Hall/CRC, New York, 2016.

[34]   A. Vexler, Y. M. Kim, J. Yu, N. A. Lazar and A. D. Hutson. Computing Critical Values of Exact Tests by Incorporating Monte Carlo Simulations Combined with Statistical Tables. Scandinavian Journal of Statistics. 41 (2014) 1013–1030.

[35]   A. Vexler, W-M.Tsai and A. D. Hutson. A Simple Density-Based Empirical Likelihood Ratio Test for Independence. The American Statistician. 68 (2014) 158-169.

[36]   J. A. Villaseñor-Alva and G. González-Estrada. A Generalization of Shapiro-Wilk's Test for Multivariate Normality. Communications in Statistics-Theory and Methods 38 (2009) 1870-1883.

[37]   S. S. Wilks. Mathematical Statistics. John Wiley and Sons, New York, 1962.

[38]   L-X. Zhu, H. L. Wong and K-T. Fang. A test for multivariate normality based on sample entropy and projection pursuit. Journal of Statistical Planning and Inference 45 (1995) 373-385.

On-line Supplement to "**A density based empirical likelihood approach for testing bivariate normality**"

Gregory Gurevich[a] and Albert Vexler[b,1]

[a] Department of Industrial Engineering and Management, SCE-Shamoon College of Engineering, Beer Sheva 84100, Israel

[b] Department of Biostatistics, The State University of New York, Buffalo, NY 14214, USA

Abstract: This on-line supplement to "A density based empirical likelihood approach for testing bivariate normality" contains technical proofs of the theoretical propositions shown in the article. In this Supplement Material, in order to describe the dataset used in the article, we also provide Figure S1 that depicts histograms and scatter plots of the log-transformed biomarkers values related to vitamin E and HDL-cholesterol. This Supplement presents R code to implement the proposed method.

Data
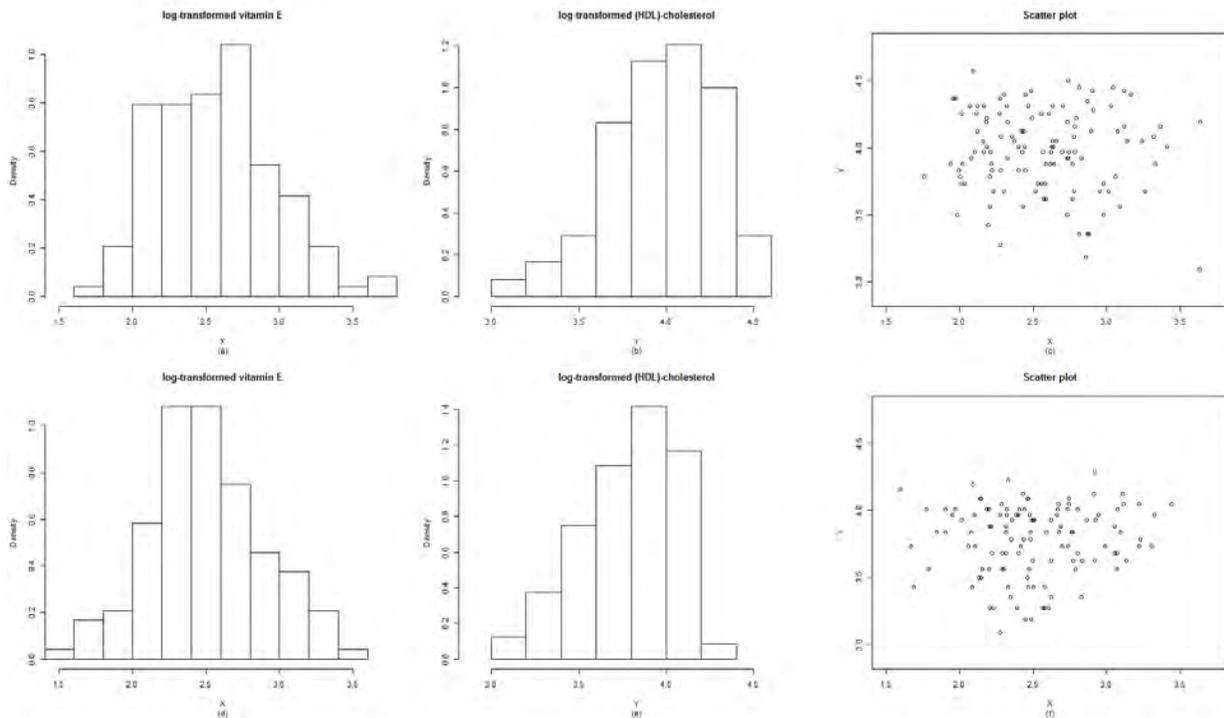
[1] Corresponding author. Email: avexler@buffalo.edu

Figure S1. Histograms and scatter plots of the log-transformed biomarkers values related to vitamin E and HDL-cholesterol. Panels (a), (b) and (c) correspond to MI=1 and panels (d), (f) and (e) correspond to MI=0.

Appendix A.

Define

$$H_{mn} = \frac{1}{n}\sum_{i=1}^{n} log\left\{\frac{n}{2m}\left(X_{(i+m)} - X_{(i-m)}\right)\right\}.$$

Using some reorganization, H_{mn} can be presented in the form $H_{mn} = (2m)^{-1}\sum_{j=1}^{2m} S_j + U_{mn}$, where

$$S_j = -\sum_{i=1}^{n} log\left\{\frac{F\left(X_{(i+m)}\right) - F\left(X_{(i-m)}\right)}{X_{(i+m)} - X_{(i-m)}}\right\}\left\{F_n\left(X_{(i+m)}\right) - F_n\left(X_{(i-m)}\right)\right\}, \quad i \equiv j \ (\mathrm{mod}\ 2m),$$

$F_n(u) = (n)^{-1}\sum_{i=1}^{n} I(X_i \le u)$ is the empirical distribution function and

$$U_{mn} = \frac{1}{n}\sum_{i=1}^{n} log\left[\frac{n}{2m}\left\{F\left(X_{(i+m)}\right) - F\left(X_{(i-m)}\right)\right\}\right].$$

Vasicek [30] showed the following results: $(2m)^{-1}\sum_{j=1}^{2m} S_j \to H(f) \equiv E\left[log\left\{f(X_1)\right\}\right]$ almost sure as $m/n \to 0$, $n \to \infty$; U_{mn} is a non-positive variable with the expectation

$$E(U_{mn}) = log(n) - log(2m) + \left(1 - \frac{2m}{n}\right)\psi(2m) - \psi(n+1) + \frac{2}{n}\sum_{i=1}^{m}\psi(i+m-1),$$

where $\psi(u) = \Gamma'(u)/\Gamma(u)$ is the digamma function; $E(U_{mn}) \to 0$, if $n \to \infty$, $m \to \infty$; and

$U_{mn} \xrightarrow{P} 0$. These results provide the conclusion $H_{mn} \xrightarrow{P} E\left[log\left\{f(X_1)\right\}\right]$. It is clear that this property is not held, if m has a fixed value (e.g., $m=1$). For example, for $m=1$, we have

$$E(U_{1n}) = log(n) - log(2) + \left(1 - \frac{2}{n}\right)\psi(2) - \psi(n+1) + \frac{2}{n}\sum_{i=1}^{1}\psi(i+1-1)$$

$$= log(n) - log(2) + \psi(2) - \frac{2}{n}\psi(2) - \psi(n+1) + \frac{2}{n}\psi(1)$$

$$= log(n) - \psi(n+1) - log(2) + \psi(2) - \frac{2}{n}(\psi(1)+1) + \frac{2}{n}\psi(1)$$

2

$$= log(n) - \psi(n+1) - log(2) + \psi(2) - \frac{2}{n} \rightarrow -log(2) + \psi(2) \neq 0 \text{ as } n \rightarrow \infty.$$

Appendix B. Proof of Equation (12).

The summands presented in the definition of $\tilde{H}_{m,k}$ at (11) consist of the items

$$A_{j,l_1,l_2} \equiv (4km)^{-1} f\left(X_{[j:l_1,l_2]}, Y_{(j:l_1,l_2)}\right)\left(X_{(l_2)} - X_{(l_1)}\right)\left(Y_{(j+k:l_1,l_2)} - Y_{(j-k:l_1,l_2)}\right),$$

where $l_1 = r - m$, $l_2 = r + m$ ($l_1 = 1$, if $r - m \leq 1$ and $l_2 = n$, if $r + m \geq n$) with $j = 1, 2, ..., l_2 - l_1 + 1$ and $r = 1, 2, ..., n$. It is clear that $m \leq l_2 - l_1 \leq 2m$.

In order to show that Definitions (11) and (12) are identical, we provide the following scheme for detecting summands in (11), which consist of equivalent multipliers $f_i = f\left(X_{(i)}, Y_{[i]}\right)$, $i = 1, 2, ..., n$.

Consider Case (a) in (12) with $1 \leq i \leq m$. According to the process of obtaining (11), A_{j,l_1,l_2} includes f_i, if $X_{(i)} \in \left\{X_{(l_1)}, X_{(l_1+1)}, ..., X_{(l_2)}\right\}$. In this case there are $m + i$ different pairs (l_1, l_2) such that $X_{(i)} \in \left\{X_{(l_1)}, X_{(l_1+1)}, ..., X_{(l_2)}\right\}$ and these scenarios can be divided into two groups:

(i)   ($l_1 = 1, l_2 = m + s$) with $s = 1, 2, ..., m$, when $X_{(i)} \in \left\{X_{(1)}, X_{(2)}, ..., X_{(m+s)}\right\}$;

(ii)   ($l_1 = h, l_2 = h + 2m$) with $h = 1, ..., i$, when $X_{(i)} \in \left\{X_{(h)}, X_{(h+1)}, ..., X_{(h+2m)}\right\}$.

Step 1.1: Let $s = 1$ in Group (i). Then, $A_{j,l_1,l_2} = A_{j,1,m+1}$, $j = 1, 2, ..., m + 1$, where $Y_{(j+k:1,m+1)}$ and $Y_{(j-k:1,m+1)}$ are the ($j + k$)th and ($j - k$)th order statistics based on $Y_{[1]}, Y_{[2]}, ..., Y_{[m+1]}$, which are the concomitants of $X_{(1)}, X_{(2)}, ..., X_{(m+1)}$, respectively. The rank of the observation $Y_{[r]}$ with respect to $Y_{[c]}, Y_{[c+1]}, ..., Y_{[d]}$ was defined as $\rho\left(Y_{[r]}, c, d\right) = \sum_{l=c}^{d} I\left(Y_{[r]} \leq Y_{[l]}\right)$. Then, $Y_{\left(\rho\left(Y_{[r]}, c, d\right):c,d\right)} = Y_{[r]}$. This implies that $Y_{(j:1,m+1)} = Y_{[i]}$, if and only if $1 \leq j = \rho\left(Y_{[i]}, 1, m + 1\right) \leq m + 1$. Furthermore, for $1 \leq j = \rho\left(Y_{[i]}, 1, m + 1\right) \leq m + 1$, we have

$$Y_{(j:1,m+1)} = Y_{\left(\rho\left(Y_{[i]}, 1, m+1\right):1,m+1\right)} = Y_{[i]} \text{ and } X_{[j:1,m+1]} = X_{\left[\rho\left(Y_{[i]}, 1, m+1\right):1,m+1\right]} = X_{(i)}.$$

3

Thus,

$$A_{j,1,m+1} = (4km)^{-1} f_i\left(X_{(m+1)} - X_{(1)}\right)\left(Y_{\left(\rho\left(Y_{[i]},1,m+1\right)+k:1,m+1\right)} - Y_{\left(\rho\left(Y_{[i]},1,m+1\right)-k:1,m+1\right)}\right).$$

Step 1.2: In a similar manner to Step 1.1, we consider Group *(i)* with $s = 2,...,m$. Then, $A_{j,l_1,l_2} = A_{j,1,m+s}$ and one can show that the corresponding summand, which has the multiplier f_i and is presented in $\tilde{H}_{m,k}$, is

$$(4km)^{-1} f\left(X_{(i)}, Y_{[i]}\right)\left(X_{(m+s)} - X_{(1)}\right)\left(Y_{\left(\rho\left(Y_{[i]},1,m+s\right)+k:1,m+s\right)} - Y_{\left(\rho\left(Y_{[i]},1,m+s\right)-k:1,m+s\right)}\right).$$

The Steps 1.1-1.2 based algorithm can be easily modified to evaluate Group *(ii)*, obtaining that the corresponding summand, which has the multiplier f_i and is presented in $\tilde{H}_{m,k}$ at (11), is

$$(4km)^{-1} f_i\left(X_{(h+2m)} - X_{(h)}\right)\left(Y_{\left(\rho\left(Y_{[i]},h,h+2m\right)+k:h,h+2m\right)} - Y_{\left(\rho\left(Y_{[i]},h,h+2m\right)-k:h,h+2m\right)}\right).$$

Thus, the part (a) of Equation (12) is justified.

Now, one can consider separately Cases (b)-(e) at Definition (12), applying directly adapted Steps 1.1-1.2. That is, each summand presented in the statistic $(4km)^{-1}\sum_{i=1}^{n} f_i G_{i,m,k}$ via the schematic form "$(4km)^{-1} f_i\left(X - X\right)\left(Y - Y\right)$" corresponds to only one summand presented in statistics $\tilde{H}_{m,k}$ defined in (11). The number of these summands in (11) and (12) is $w = n + m(2n - m - 1)$. The equation (12) is verified.

Appendix C. Proof of the asymptotic consistency of the bivariate dbEL

Assume $A_{n,i} > 0$, $B_{n,i} > 0$, $C_{n,i} > 0$, $D_{n,i} > 0$ define positive-integer valued sequences of random variables, satisfying with probability one the condition

$$n^{\delta} < B_{n,i} - A_{n,i} < 2n^{1-\delta}, \quad X_{(A_{n,i})} < X_{(i)} < X_{(B_{n,i})}, \quad \left(B_{n,i} - A_{n,i}\right)^{\delta} < D_{n,i} - C_{n,i} < 2\left(B_{n,i} - A_{n,i}\right)^{1-\delta}, \quad (C1)$$

$$Y_{(C_{n,i}:A_{n,i},B_{n,i})} < Y_{[i]} < Y_{(D_{n,i}:A_{n,i},B_{n,i})}, \quad i = 1,2,...,n,$$

where $0 < \delta < 0.5$ is defined in Equation (14). Then, we have the following lemma.

4

Lemma C. Let $f(x,y)$ be continuous and bounded on its support, $a_1 < f(x,y) < a_2$, for some constants $0 < a_1 < a_2 < \infty$. Then, for any $\varepsilon > 0$, all $i = 1, 2, ..., n$, and $A_{n,i}, B_{n,i}, C_{n,i}, D_{n,i}$ that satisfy condition (C1), we have

$$Pr\left(\left|\frac{D_{n,i} - C_{n,i}}{n\left(Y_{(D_{n,i}:A_{n,i},B_{n,i})} - Y_{(C_{n,i}:a_{n,i},b_{n,i})}\right)\left(X_{(B_{n,i})} - X_{(A_n,i)}\right)} - f_i\right| > \varepsilon\right) \to 0$$

uniformly as $n \to \infty$, where $f_i = f\left(X_{(i)}, Y_{[i]}\right)$.

Proof.

Define

$$f_{n,i} = \frac{D_{n,i} - C_{n,i}}{n\left(Y_{(D_{n,i}:A_{n,i},B_{n,i})} - Y_{(C_{n,i}:A_{n,i},B_{n,i})}\right)\left(X_{(B_{n,i})} - X_{(A_n,i)}\right)}.$$

Taking into account formal notations used in Equation (2.1) of Kim and van Ryzin [11], we present $f_{n,i}$ in the form

$$f_{n,i} = V_{n,i} W_{n,i},\tag{C2}$$

where

$$V_{n,i} = \frac{D_{n,i} - C_{n,i}}{n \Delta_{i,D-C} F\left(X_{(i)}, Y_{[i]}\right)}, \quad W_{n,i} = \frac{\Delta_{i,D-C} F\left(X_{(i)}, Y_{[i]}\right)}{\left(Y_{(D_{n,i}:A_{n,i},B_{n,i})} - Y_{(C_{n,i}:A_{n,i},B_{n,i})}\right)\left(X_{(B_{n,i})} - X_{(A_{n,i})}\right)},$$

$$\Delta_{i,D-C} F\left(X_{(i)}, Y_{[i]}\right) = Pr\left(X_{(A_{n,i})} \le X_1 \le X_{(B_{n,i})}, Y_{(C_{n,i}:A_{n,i},B_{n,i})} \le Y_1 \le Y_{(D_{n,i}:A_{n,i},B_{n,i})}\right),$$

The mean value theorem for double integrals provides the result

$$\Delta_{i,D-C} F\left(X_{(i)}, Y_{[i]}\right) = f\left(H_{1,i}, H_{2,i}\right)\left(Y_{(D_{n,i}:A_{n,i},B_{n,i})} - Y_{(C_{n,i}:A_{n,i},B_{n,i})}\right)\left(X_{(B_{n,i})} - X_{(A_{n,i})}\right),\tag{C3}$$

where $X_{(A_{n,i})} \le H_{1,i} \le X_{(B_{n,i})}$, $Y_{(C_{n,i}:A_{n,i},B_{n,i})} \le H_{2,i} \le Y_{(D_{n,i}:A_{n,i},B_{n,i})}$.

Then,

$$W_{n,i} = f\left(H_{1,i}, H_{2,i}\right)$$

and

$$V_{n,i} = \frac{D_{n,i} - C_{n,i}}{n} \frac{1}{Pr\left(X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}, Y_{(C_{n,i}:A_{n,i},B_{n,i})} \leq Y_1 \leq Y_{(D_{n,i}:A_{n,i},B_{n,i})}\right)}, \quad (C4)$$

where

$$Pr\left(X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}, Y_{(C_{n,i}:A_{n,i},B_{n,i})} \leq Y_1 \leq Y_{(D_{n,i}:A_{n,i},B_{n,i})}\right) \qquad (C5)$$

$$= Pr\left(Y_{(C_{n,i}:A_{n,i},B_{n,i})} \leq Y_1 \leq Y_{(D_{n,i}:A_{n,i},B_{n,i})} \middle| X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}\right) Pr\left(X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}\right)$$

$$= \left(F_{Y_1 | X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}}\left(Y_{(D_{n,i}:A_{n,i},B_{n,i})}\right) - F_{Y_1 | X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}}\left(Y_{(C_{n,i}:A_{n,i},B_{n,i})}\right)\right)$$

$$\left(F_{X_1}\left(X_{(B_{n,i})}\right) - F_{X_1}\left(X_{(A_{n,i})}\right)\right).$$

Here the distribution functions $F_{Y_1 | X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}}(y)$ and $F_{X_1}(x)$ mean

$$F_{Y_1 | X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}}(y) = Pr\left(Y_1 \leq y \middle| X_{(A_{n,i})} \leq X_1 \leq X_{(B_{n,i})}\right) \text{ and } F_{X_1}(x) = Pr(X_1 \leq x).$$

Note that $X_{(B_{n,i})} - X_{(A_{n,i})} \xrightarrow{p} 0$ and $Y_{(D_{n,i}:A_{n,i},B_{n,i})} - Y_{(C_{n,i}:A_{n,i},B_{n,i})} \xrightarrow{p} 0$ uniformly for all $A_{n,i}, B_{n,i}, C_{n,i}, D_{n,i}$ that satisfy condition (C1) as $n \to \infty$ (see Lemma 1 of Kim and van Ryzin [11]). In addition, the density function $f(x,y)$ is uniformly continuous on its support. This implies

$$W_{n,i} \xrightarrow{p} f_i \qquad (C6)$$

uniformly for all $i = 1, 2, ..., n$ and $A_{n,i}, B_{n,i}, C_{n,i}, D_{n,i}$ that satisfy condition (C1).

Note also that the random variables $F_{X_1}\left(X_{(1)}\right)$, $F_{X_1}\left(X_{(2)}\right), ..., F_{X_1}\left(X_{(n)}\right)$ are distributed as an ordered sample of size n from a uniform [0,1] distribution and then the random variable $F_{X_1}\left(X_{(j+m)}\right) - F_{X_1}\left(X_{(j)}\right)$ has the beta distribution with parameters $(m, n-m+1)$ for all $j = 1, 2, ..., n-m$ (van Ryzin [29], Wilks [37]). Thus,

$$F_{X_1}\left(X_{(j+m)}\right) - F_{X_1}\left(X_{(j)}\right) \sim Beta(m, n-m+1), \ j = 1, 2, ..., n-m. \qquad (C7)$$

In a similar manner to the analysis shown above, one can obtain that

6

$$F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{(1:j,j+m)}\right), \ F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{(2:j,j+m)}\right),...,$$

$$F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{(m+1:j,j+m)}\right)$$

are distributed as an ordered sample of size $m+1$ from a uniform $[0,1]$ distribution, since

$Y_{(1:j,j+m)},Y_{(2:j,j+m)},...,Y_{(m+1:j,j+m)}$ are the order statistics based on $Y_{[j]},Y_{[j+1]},...,Y_{[j+m]}$ that are

the concomitants of $X_{(j)},X_{(j+1)},...,X_{(j+m)}$, and then $Y_{[r]}\sim F_{Y_1|X_{(r)}\leq X_1\leq X_{(r+m)}}$, $r=j,...,j+m$.

Therefore,

$$F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{(l+k:j,j+m)}\right)-F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{Y_{(l:j,j+m)}}\right)\sim Beta\left(k,m-k+2\right), \quad \text{(C8)}$$

$$l=1,2,...,m-k+1.$$

According to Lemma 4 presented in van Ryzin [29], if a random variable

$Z_n\sim Beta\left(k_n,n-k_n+1\right)$, where $k_n\to\infty$ as $n\to\infty$, then, for each $0<\varepsilon_1<1$, there exist constants

$A>0$, $\alpha_1>0$ and $\alpha_2>0$ such that

$$Pr\left(Z_n\geq\frac{k_n}{n(1-\varepsilon_1)}\right)\leq A\,exp\left(-\alpha_1 k_n\right) \text{ and } Pr\left(Z_n\leq\frac{k_n}{n(1+\varepsilon_1)}\right)\leq A\,exp\left(-\alpha_2 k_n\right). \qquad \text{(C9)}$$

Now, taking into account (C4) and (C5), we obtain that, for any $0<\varepsilon_2<1$ and a large enough n (n:

$n^\delta<2n^{1-\delta}$ and $m^\delta<2m^{1-\delta}$ for $n^\delta\leq m\leq 2n^{1-\delta}$),

$Pr\{\text{there exist sequences } A_{n,i},B_{n,i},C_{n,i},D_{n,i} \text{ that satisfy (C1) such that } |V_{n,i}-1|>\varepsilon_2, \text{ for some } 1\leq i\leq n\}$

$$\leq\sum_{m=n^\delta}^{2n^{1-\delta}}\sum_{k=m^\delta}^{2m^{1-\delta}}\sum_{j=1}^{n-m}\sum_{l=1}^{m-k+1}Pr\left(\begin{array}{l}F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{(l+k:j,j+m)}\right)\\[4mm]-F_{Y_1|X_{(j)}\leq X_1\leq X_{(j+m)}}\left(Y_{(l:j,j+m)}\right)\geq\dfrac{k}{(m+1)(1-\varepsilon_2)}\end{array}\right)$$

$$....\text{(C10)}$$

$$+\sum_{m=n^\delta}^{2n^{1-\delta}}\sum_{j=1}^{n-m}Pr\left(F_{X_1}\left(X_{(j+m)}\right)-F_{X_1}\left(X_{(j)}\right)\geq\frac{m}{n(1-\varepsilon_2)}\right)$$

$$+ \sum_{m=n^\delta}^{2n^{1-\delta}} \sum_{k=m^\delta}^{2m^{1-\delta}} \sum_{j=1}^{n-m} \sum_{l=1}^{m-k+1} Pr\left(F_{Y_1 \mid X_{(j)} \le X_1 \le X_{(j+m)}} \left(Y_{(l+k:j,j+m)} \right) \right.$$

$$\left. - F_{Y_1 \mid X_{(j)} \le X_1 \le X_{(j+m)}} \left(Y_{(l:j,j+m)} \right) \le \frac{k}{(m+1)(1+\varepsilon_2)} \right)$$

$$+ \sum_{m=n^\delta}^{2n^{1-\delta}} \sum_{j=1}^{n-m} Pr\left(F_{X_1}\left(X_{(j+m)} \right) - F_{X_1}\left(X_{(j)} \right) \le \frac{m}{n(1+\varepsilon_2)} \right),$$

where $0 < \delta < 0.5$ is defined in Equation (14).

Thus, by (C7)-(C10), there exist constants $A > 0$, $\alpha_1 > 0$ and $\alpha_2 > 0$ such that

$$Pr\left\{ \text{there exist sequences } A_{n,i}, B_{n,i}, C_{n,i}, D_{n,i} \text{ that satisfy (C1) such that } \left| V_{n,i} - 1 \right| > \varepsilon_2, \text{ for some } 1 \le i \le n \right\}$$

$$\le A\left(n^4 \, exp\left(-\alpha_1 n^{\delta^2} \right) + n^2 \, exp\left(-\alpha_1 n^\delta \right) + n^4 \, exp\left(-\alpha_2 n^{\delta^2} \right) + n^2 \, exp\left(-\alpha_2 n^\delta \right) \right) \to 0 \text{ as } n \to \infty. \quad \text{(C11)}$$

Rewrite (C2) in the form

$$f_{n,i} - f_i = V_{n,i}\left(W_{n,i} - f_i \right) + \left(V_{n,i} - 1 \right) f_i. \tag{C12}$$

Results (C6), (C11) applied to (C12) complete the proof of Lemma C.

In order to prove that $n^{-1} log\left(\Delta_n \right) \xrightarrow{p} E\left\{ log\left(f\left(X_1, Y_1 \right) \right) \right\}$ as $n \to \infty$, we will show that

$n^{-1} log\left(\Delta_{n,m,k} \right) \xrightarrow{p} E\left\{ log\left(f\left(X_1, Y_1 \right) \right) \right\}$ uniformly with respect to all $m \in \left[n^\delta, n^{1-\delta} \right]$ and

$k \in \left[m^\delta, m^{1-\delta} \right]$, where $0 < \delta < 0.5$. We have

$$\frac{1}{n} log\left(\Delta_{n,m,k} \right) = \frac{1}{n} log\left(\prod_{i=1}^n \frac{4mk}{nG_{i,m,k}} \right) = -\frac{1}{n} \sum_{i=1}^n log\left(\frac{n}{4mk} G_{i,m,k} \right) \tag{C13}$$

$$= -\frac{1}{n} \sum_{i=1}^{2m} log\left(\frac{n}{4mk} G_{i,m,k} \right) - \frac{1}{n} \sum_{i=2m+1}^{n-2m} log\left(\frac{n}{4mk} G_{i,m,k} \right) - \frac{1}{n} \sum_{i=n-2m+1}^n log\left(\frac{n}{4mk} G_{i,m,k} \right).$$

Consider the case *(i)*: $2m+1 \le i \le n-2m$ and the case *(ii)*: $1 \le i \le n$, $i \notin \left[2m+1, n-2m \right]$.

(i) $2m+1 \le i \le n-2m$: in this case, using (12), one can write summands $log\left(\frac{n}{4mk} G_{i,m,k} \right)$ in the form

$$log\left(\frac{n}{4mk}G_{i,m,k}\right)$$

$$=log\left(\frac{n}{4mk}\sum_{j=0}^{2m}\left[Y_{\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)+k:i-j,i-j+2m\right)}-Y_{\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)-k:i-j,i-j+2m\right)}\right]\right.$$

$$\left.\left(X_{(i-j+2m)}-X_{(i-j)}\right)\right).$$

Lemma C provides

$$\frac{n\left[Y_{\left(\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)+k:i-j,i-j+2m\right)}-Y_{\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)-k:i-j,i-j+2m\right)}\right)\right]\left(X_{(i-j+2m)}-X_{(i-j)}\right)}{min\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)+k,i-j+2m\right)-max\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)-k,i-j\right)} \quad (C14)$$

$$\xrightarrow{p}\frac{1}{f_i}\quad \text{as } n\to\infty,$$

uniformly for all $i=1,2,...,n$, $m\in\left[n^{\delta},n^{1-\delta}\right]$ and $k\in\left[m^{\delta},m^{1-\delta}\right]$.

It is clear that

$$k\le min\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)+k,i-j+2m\right)-max\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)-k,i-j\right)\le 2k. \quad (C15)$$

Following the material presented in Section 2, the number of items

$$A_{j,l_1,l_2}\equiv(4km)^{-1}f\left(X_{[j:l_1,l_2]},Y_{(j:l_1,l_2)}\right)\left(X_{(l_2)}-X_{(l_1)}\right)\left(Y_{(j+k:l_1,l_2)}-Y_{(j-k:l_1,l_2)}\right)$$

(see Appendix B), presented in statistic $\tilde{H}_{m,k}$ at (11) and in $\frac{1}{4km}\sum_{i=1}^{n}f_iG_{i,m,k}$ at (12) (among

$w=n+2nm-m(m-1)$, a total number of the summands), when

$$min(j+k,l_2)-max(j-k,l_1)<2k,$$

is less or equal to $2(k+1)n$.

Since, for any fixed $\varepsilon>0$ and a large enough m, $m\varepsilon>2(k+1)$, we have that there are at least $(n-4m)(2m+1)-nm\varepsilon$ items

$$\left[Y_{\left(\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)+k:i-j,i-j+2m\right)}-Y_{\left(\rho\left(Y_{[i]},i-j,i-j+2m\right)-k:i-j,i-j+2m\right)}\right)\right]\left(X_{(i-j+2m)}-X_{(i-j)}\right)$$

9

in

$$\sum_{i=2m+1}^{n-2m} \frac{n}{4mk} G_{i,m,k}$$

$$= \sum_{i=2m+1}^{n-2m} \left(\frac{n}{4mk} \sum_{j=0}^{2m} Y_{\left(\rho\left(Y_{[i]}, i-j, i-j+2m\right)+k : i-j, i-j+2m \right)} - Y_{\left(\rho\left(Y_{[i]}, i-j, i-j+2m\right)-k : i-j, i-j+2m \right)} \right)$$

$$\left(X_{(i-j+2m)} - X_{(i-j)} \right),$$

when the subscripts of these items satisfy

$$min\left(\rho\left(Y_{[i]}, i-j, i-j+2m\right)+k, i-j+2m \right) - max\left(\rho\left(Y_{[i]}, i-j, i-j+2m\right)-k, i-j \right) = 2k.$$

Since, for any fixed $\varepsilon > 0$, $(n-4m-n\varepsilon)(2m+1) < (n-4m)(2m+1) - nm\varepsilon$, we have that there are different integer values $i_1, i_2, ..., i_{n-4m-n\varepsilon}$, which satisfy

$$min\left(\rho\left(Y_{[i_r]}, i_r-j, i_r-j+2m\right)+k, i_r-j+2m \right) - max\left(\rho\left(Y_{[i_r]}, i_r-j, i_r-j+2m\right)-k, i_r-j \right) = 2k,$$

where $r = 1, 2, ..., n-2m-n\varepsilon$, $2m+1 \le i_r \le n-2m$, $j = 0, 1, ..., 2m$.

Then, one can represent

$$\sum_{i=2m+1}^{n-2m} \frac{n}{4mk} G_{i,m,k} = \sum_{r=1}^{n-4m-n\varepsilon} \frac{n}{4mk} G_{i_r,m,k} + \sum_{\substack{2m+1\le i \le n-2m \\ i\ne i_r, r=1,2,...,n-4m-n\varepsilon}} \frac{n}{4mk} G_{i,m,k} ,$$

where $min\left(\rho\left(Y_{[i_r]}, i_r-j, i_r-j+2m\right)+k, i_r-j+2m \right) - max\left(\rho\left(Y_{[i_r]}, i_r-j, i_r-j+2m\right)-k, i_r-j \right) = 2k$.

That is, applying Lemma C, we have

$$\frac{n}{2k} \left(Y_{\left(\rho\left(Y_{[i_r]}, i_r-j, i_r-j+2m\right)+k : i_r-j, i_r-j+2m \right)} - Y_{\left(\rho\left(Y_{[i_r]}, i_r-j, i_r-j+2m\right)-k : i_r-j, i_r-j+2m \right)} \right) \left(X_{(i_r-j+2m)} - X_{(i_r-j)} \right) \xrightarrow{p} \frac{1}{f_{i_r}}$$

as $n \to \infty$, uniformly for all i_r, $n^\delta \le m \le n^{1-\delta}$, $m^\delta \le k \le m^{1-\delta}$. Then

$$\frac{n}{4mk} G_{i_r,m,k} \xrightarrow{p} \frac{2m+1}{2m} \frac{1}{f_{i_r}} \xrightarrow{p} \frac{1}{f_{i_r}}.$$

By virtue of (C14) and (C15), using a proof scheme similar to that shown above, one can derive the result

$$Pr\left(\frac{1}{2f_i} \leq \frac{n}{4mk} G_{i,m,k} \leq \frac{1}{f_i}\right) > 1 - \varepsilon,$$

for large values of n and all $i \in [2m+1, n-2m]$, $i \neq i_r$, where $r = 1, 2, ..., n - 4m - n\varepsilon$.

Case (ii): $1 \leq i \leq 2m$, $n - 2m + 1 \leq i \leq n$. In this case, the evaluations shown in Case *(i)* above can be easily modified to obtain that, for large values of n and any fixed $\varepsilon > 0$, we have

$$Pr\left(\frac{1}{4f_i} \leq \frac{n}{4mk} G_{i,m,k} \leq \frac{1}{f_i}\right) > 1 - \varepsilon.$$

Thus, when $n \gg 1$, we obtain

$$Pr\left(\sum_{r=1}^{n-4m-n\varepsilon} log\left(\frac{1}{f_{i_r}}\right) + \sum_{\substack{2m+1 \leq i \leq n-2m \\ i \neq i_r, r=1,2,...,n-4m-n\varepsilon}} log\left(\frac{1}{2f_i}\right) + \sum_{\substack{i=1,...,2m, \\ i=n-2m+1,...n}} log\left(\frac{1}{4f_i}\right)\right.$$

$$\left. \leq \sum_{i=1}^{n} log\left(\frac{n}{4mk} G_{i,m,k}\right) \leq \sum_{i=1}^{n} log\left(\frac{1}{f_i}\right)\right) > 1 - \varepsilon.$$

Taking into account that $log\left(\frac{1}{a_2}\right) < log\left(\frac{1}{f_i}\right) < log\left(\frac{1}{a_1}\right)$, $i = 1,...,n$, we conclude that, when $n \gg 1$,

$$Pr\left(\sum_{i=1}^{n} log\left(\frac{1}{f_i}\right) - (4m + n\varepsilon) log\left(\frac{1}{a_1}\right) + (4m + n\varepsilon) log\left(\frac{1}{4a_2}\right)\right.$$

$$\left. \leq \sum_{i=1}^{n} log\left(\frac{n}{4mk} G_{i,m,k}\right) \leq \sum_{i=1}^{n} log\left(\frac{1}{f_i}\right)\right) > 1 - \varepsilon,$$

where $\frac{1}{n}\left(-(4m+n\varepsilon) log\left(\frac{1}{a_1}\right) + (4m+n\varepsilon) log\left(\frac{1}{4a_2}\right)\right) \rightarrow \varepsilon\left(log\left(\frac{1}{4a_2}\right) - log\left(\frac{1}{a_1}\right)\right)$ and the law of

large numbers provides that $\frac{1}{n} \sum_{i=1}^{n} log\left(\frac{1}{f_i}\right) \xrightarrow{p} E\{log(f(X_1, Y_1))\}$. This result applied to (C13)

implies that $\frac{1}{n} \sum_{i=1}^{n} log\left(\frac{n}{4mk} G_{i,m,k}\right) \rightarrow E\{log(f(X_1, Y_1))\}$, completing the proof of the asymptotic

consistency of the bivariate dbEL.

R-Code

```r
library(MASS)
library(goft)
library(strucchange)
library(mvtnorm)
library(MVN)

p <-0
Sigma <- matrix(c(1,p,p,1),2,2)

n<-50            #number of observations
MC<-25000        #number of Monte Carlo generations

EL<-array()   #dbEL Test Statistic $V_n^1$ or $V_n^2$

SW<-array() #Shapiro-Wilk (SW) test statistic
CT<-array()  # classical Mardia's test statistic

for(mc in 1:MC){
                              #(X,Y) under H0
Xo<-mvrnorm(n, rep(0, 2), Sigma)
xo<-Xo[,1]
yo<-Xo[,2]

######## Scenario A #########################################
#for(i in 1:n){
#u<-runif(1,0,1)
#xo[i]<-rnorm(1)
#yo[i]<-xo[i]*(u<=1/2)-xo[i]*(u>1/2)
#}
###############################################

############ Scenario B#####################################
#yo<-runif(n,-5,5)
#xo<-runif(n,-5,5)
#############################################################

### Scenario E: Morgenstern Alternative#######
#alpha<-0.5
#x<-runif(n,0,1)
#v2<-runif(n,0,1)
#A<-alpha*(2*x-1)-1
#B<-1-2*alpha*(2*x-1)+alpha^2*(2*x-1)^2+4*alpha*v2*(2*x-1)
#BB<-sqrt(B)
#y<-2*v2/(BB-A)
#xo<-x
#yo<-y
###############################################
```

```
#### Scenario F: Plackett Alternative######d
#sii<-2
#x<-runif(n,0,1)
#u2<-runif(n,0,1)
#a<-u2*(1-u2)
#A<-sii + a*(sii-1)^2
#B<-2*a*(x*sii^2 + 1 - x) + sii*(1-2*a)
#D<-sii*(sii + 4*a*x*(1-x)*(1-sii)^2)
#DD<-sqrt(D)
#y<-(B-(1-2*u2)*DD)/2*A
#xo<-x
#yo<-y
##################################################

################# Scenario H: Gumbel
#U1<-runif(n,0,1)
#U2<-runif(n,0,1)
#U3<-runif(n,0,1)
#U4<-runif(n,0,1)
#x1o<-(-log(U1))
#et<-0.9
#beta<-1+et*x1o
#pp<-(beta-et)/beta
#y<-(-log(U2))
#y1o<-array()
#for (i in 1:n) {
#            if(U3[i]<pp[i]) y1o[i]<-beta[i]*y[i]
#            if(U3[i]>=pp[i]) y1o[i]<-beta[i]*(y[i]-log(U4[i]))
#            }
#Sigma <- matrix(c(1/2,1/4,1/4,1/2),2,2)
#Xo<-mvrnorm(n, rep(0, 2), Sigma)
#xo<-Xo[,1]
#yo<-Xo[,2]
#yo<-runif(n,-6,3)+y1o
#xo<-runif(n,-6,3)+x1o
#############################################################

################# Scenario C
#xo<-xo+xo^4/25
#yo<-yo*xo
##################################################

#################### Scenario D
#for(i in 1:n){
#xo[i]<-sum(runif(7,0,1))/sqrt(7)
#yo[i]<-sum(runif(7,0,1))/sqrt(7)
```

13

```
#}

######################### Scenario I

xo<-rgamma(n,2,1)
yo<-rgamma(n,2,1)
############################################################

XN<-array(0,c(n,2))
XN[,1]<-xo
XN[,2]<-yo
SW[mc]<-mvShapiro.Test(XN)$'p'

a=0.4

beg_m<-round(max(c(1,n^a)))
end_m<-round(min(c(n^(1-a),n/2)))

beg_k<-function(m) round(max(c(1,m^a)))
end_k<-function(m) round(min(c(m^(1-a),m/2)))

LogL<-array(100000009,c(n,n))
```
######################################Use these (X,Y) to calculate V_n^1 or see (*) below
```
                              XY<-data.frame(XN)
                              XYoX<-XY[order(XY$X1),]
                              X<-XYoX[,1,]
                              Y<-XYoX[,2,]


#############################################################################
S2X<-mean((xo-mean(xo))^2)
S2Y<-mean((yo-mean(yo))^2)
PHO<-mean((xo-mean(xo))*(yo-mean(yo)))/sqrt(S2X*S2Y)
###################################################################
Sn<-matrix(c(S2X,mean((xo-mean(xo))*(yo-mean(yo))),mean((xo-mean(xo))*(yo-
mean(yo))),S2Y),2,2)
InvSn<-ginv(Sn)
Mult<-root.matrix(InvSn)
xn<-array()
yn<-array()
for(u in 1:n){
        PP<-Mult%*%t(matrix(c(xo[u]-mean(xo),yo[u]-mean(yo)),1,2))
        xn[u]<-PP[1]
        yn[u]<-PP[2]
        }
```
######### (*) Use these (X,Y) to calculate V_n^2 (do not use the following 4 lines if you need V_n^1)
```
                                     XY<-data.frame(xn=xn,yn=yn)
```

```
                                        XYoX<-XY[order(XY$xn),]
                                        X<-XYoX[,1,]
                                        Y<-XYoX[,2,]
NL<-2*n*log(1/sqrt(2*pi))-sum(X^2)/2-sum(Y^2)/2
###################################

#############################################Classical test######################
HatSig<-0
for(u in 1:n){
        HatSig<-HatSig+t(matrix(c(xo[u]-mean(xo),yo[u]-mean(yo)),1,2))%*%matrix(c(xo[u]-
mean(xo),yo[u]-mean(yo)),1,2)
        }
HatSig<-HatSig/n
A<-0
for(u in 1:n) {
        for(v in 1:n)
         {
          A<-A+(matrix(c(xo[u]-mean(xo),yo[u]-
mean(yo)),1,2)%*%ginv(HatSig)%*%t(matrix(c(xo[v]-mean(xo),yo[v]-mean(yo)),1,2)))^3
         }
        }
A<-A/(6*n)
df<-2*(2+1)*(2+2)/6
TS<-qchisq(0.95, df)
CT[mc]<-1*(A>TS)
##############################################################################

for(mm in (beg_m):(end_m)){
for(mk in (beg_k(mm)):(end_k(mm))){

G<-array(0,n)
for(i in 1:(max(mm,1))) {
        S<-0
        for(j in 1:(mm))
          {
          Z<-Y[1:(mm+j)]
          nz<-length(Z)
          Zs<-sort(Z)
          roj<-length(Z[Z<=Y[i]])
          Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
          Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
          S<-S+(Zs[Uy]-Zs[Dy])*(X[mm+j]-X[1])
          }
        for(j in 1:i)
          {
          Ux<-(j+2*mm)*(j+2*mm<=n)+n*(j+2*mm>n)
          Z<-Y[j:Ux]
```

```
              nz<-length(Z)
              Zs<-sort(Z)
              roj<-length(Z[Z<=Y[i]])
              Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
              Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
              S<-S+(Zs[Uy]-Zs[Dy])*(X[Ux]-X[j])
              }
            G[i]<-S
             }


for(i in (mm+1):(2*mm)) {
            S<-0
           for(jj in 1:(2*mm-i+1))
             {
             j<-jj-1
             Z<-Y[1:(i+j)]
             nz<-length(Z)
             Zs<-sort(Z)
             roj<-length(Z[Z<=Y[i]])
             Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
             Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
             S<-S+(Zs[Uy]-Zs[Dy])*(X[i+j]-X[1])
             }
           for(j in 1:i)
             {
             Ux<-(j+2*mm)*(j+2*mm<=n)+n*(j+2*mm>n)
             Z<-Y[j:Ux]
             nz<-length(Z)
             Zs<-sort(Z)
             roj<-length(Z[Z<=Y[i]])
             Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
             Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
             S<-S+(Zs[Uy]-Zs[Dy])*(X[Ux]-X[j])
             }
           G[i]<-S
            }


if((n-2*mm)>=(2*mm+1)) for(i in (2*mm+1):(n-2*mm)) {
           S<-0
          for(jj in 1:(2*mm+1))
            {
            j<-jj-1
            Dx<-1*(i-j<=1)+(i-j)*(i-j>1)
            Ux<-(i-j+2*mm)*(i-j+2*mm<=n)+n*(i-j+2*mm>n)
            Z<-Y[Dx:Ux]
            nz<-length(Z)
```

16

```
            Zs<-sort(Z)
            roj<-length(Z[Z<=Y[i]])
            Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
            Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
            S<-S+(Zs[Uy]-Zs[Dy])*(X[Ux]-X[Dx])
            }
         G[i]<-S
         }

if ((n-2*mm+1)<=(n-mm)) for(i in (n-2*mm+1):(n-mm)) {
         S<-0
         for(jj in 1:(i-n+2*mm+1))
           {
           j<-jj-1
           Z<-Y[(i-j):n]
           nz<-length(Z)
           Zs<-sort(Z)
           roj<-length(Z[Z<=Y[i]])
           Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
           Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
           S<-S+(Zs[Uy]-Zs[Dy])*(X[n]-X[i-j])
           }
         for(j in 1:(n-i))
           {
           Dx<-1*(n-j-2*mm<=1)+(n-j-2*mm)*(n-j-2*mm>1)
           Z<-Y[Dx:(n-j)]
           nz<-length(Z)
           Zs<-sort(Z)
           roj<-length(Z[Z<=Y[i]])
           Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
           Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
           S<-S+(Zs[Uy]-Zs[Dy])*(X[n-j]-X[Dx])
           }
         G[i]<-S
         }

if((n-mm+1)<=(n)) for(i in (n-mm+1):(n)) {
         S<-0
         for(jj in 1:(max(mm,1)))
           {
           j<-jj-1
           Z<-Y[(n-mm-j):n]
           nz<-length(Z)
           Zs<-sort(Z)
           roj<-length(Z[Z<=Y[i]])
           Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
           Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
```

```
                    S<-S+(Zs[Uy]-Zs[Dy])*(X[n]-X[n-mm-j])
                    }
            for(jj in 1:(n-i+1))
                 {
                 j<-jj-1
                 Dx<-(n-j-2*mm)*((n-j-2*mm)>=1)+1*(n-j-2*mm<1)
                 Z<-Y[Dx:(n-j)]
                 nz<-length(Z)
                 Zs<-sort(Z)
                 roj<-length(Z[Z<=Y[i]])
                 Dy<-1*(roj-mk<1)+(roj-mk)*(roj-mk>=1)
                 Uy<-nz*(roj+mk>=nz)+(roj+mk)*(roj+mk<nz)
                 S<-S+(Zs[Uy]-Zs[Dy])*(X[n-j]-X[Dx])
                 }
             G[i]<-S
              }
G<-G[G!=0]
LogL[mm-beg_m+1,mk-beg_k(mm)+1]<-n*log(4*mm*mk)-sum(log(n*G))
}
}
LogL<-LogL[LogL!=100000009]
EL[mc]<-(min(LogL)-NL)
###########################################################
print('---------------------')
print(c(mc,EL[mc],mean(1*(EL>= 26.00287))))    # Critical Value, $C_{0.05}$, of the Test Statistic $V_n^1$
print('--')
print(c("SW",mean(1*(SW<0.05 ))))
print(c("CT",mean(CT)))
print('---------------------')
}

quantile(EL,0.9)
quantile(EL,0.95)
quantile(EL,0.96)
quantile(EL,0.975)
quantile(EL,0.99)
```