Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy

Albert Vexler [a,*], Gregory Gurevich [b]

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

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

ARTICLE INFO

Article history:
Received 3 February 2009
Received in revised form 14 August 2009
Accepted 18 September 2009
Available online xxxx

ABSTRACT

The likelihood approach based on the empirical distribution functions is a well-accepted statistical tool for testing. However, the proof schemes of the Neyman–Pearson type lemmas induce consideration of density-based likelihood ratios to obtain powerful test statistics. In this article, we introduce the distribution-free density-based likelihood technique, applied to test for goodness-of-fit. We focus on tests for normality and uniformity, which are common tasks in applied studies. The well-known goodness-of-fit tests based on sample entropy are shown to be a product of the proposed empirical likelihood (EL) methodology. Although the efficiency of test statistics based on classes of entropy estimators has been widely addressed in the statistical literature, estimation of the sample entropy has been not invariantly defined, and hence this estimation produces tests that are difficult to be applied to real data studies. The proposed EL approach defines clear forms of the entropy-based tests. Monte Carlo simulation results confirm the preference of the proposed method from a power perspective. Real data examples study the proposed approach in practice.

1. Introduction

Our aim in the present article is to develop empirical likelihood ratio tests for goodness-of-fit. The proposed tests are simple and very competitive with well-known tests for uniformity and composite hypothesis of normality. We demonstrate that the efficient goodness-of-fit tests, based on sample entropy and Kullback–Leibler information (an extended concept of entropy), are an intermediate product of the proposed EL methodology. Moreover, the proposed method standardizes the entropy-based tests. In contrast to the entropy-based tests, only tables of critical points are required for implementation of the density-based EL ratio tests. Thus, we suggest tests that are convenient to practical applications. The proposed density-based EL ratio technique can be applied to create test statistics that approximate nonparametrically powerful parametric likelihood ratios in various statistical problems. We begin by describing the principle ingredients that contribute in deriving the suggested approach.

Likelihood ratio tests provide a useful blueprint for various applied and theoretical statistical problems (e.g. Lehmann and Romano, 2006). Commonly, decision rules based on likelihood ratio type statistics have some optimal properties (e.g. Choi et al., 1996; Lehmann and Romano, 2006; Vexler et al., in press). This is, perhaps, partly yielded by the following simple consideration that belongs to proof schemes of the Neyman–Pearson type lemmas. Assume that we can use the likelihood ratio $f_{H_1}(X)/f_{H_0}(X)$, where likelihoods $f_{H_1}(X)$ and $f_{H_0}(X)$ correspond to density functions of data X under hypotheses H_1 and

* Corresponding author. Tel.: +1 716 829 2754.
E-mail addresses: avexler@buffalo.edu (A. Vexler), gregoryg@sce.ac.il (G. Gurevich).

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.09.025

H_0, respectively. By virtue of the trivial inequality,

$$(A - B)(I\{A > B\} - \eta) \geq 0, \quad \text{for all } A, B \text{ and } \eta \in [0, 1]$$

($I\{\cdot\}$ is the indicator function), we conclude that the test for H_0 versus H_1 based on the statistic $A = f_{H_1}(X)/f_{H_0}(X)$ is most powerful (here, when the ratio A exceeds a test-threshold B, H_0 is assumed to be rejected). To prove this classical proposition, we derive expectation, under H_0, from the above inequality, where the notation η denotes any decision rule based on X. Note that the density-based structure of the likelihood ratio has the main role in the Neyman–Pearson proof scheme. This sets up a worthwhile problem regarding estimation of the likelihood ratio, when data distribution functions are unknown. In this case, we can imagine that an asymptotic version of the Neyman–Pearson proof scheme can show an optimality of likelihood-ratio-related tests.

In a nonparametric context, the empirical likelihood is one of a growing array of artificial or approximate likelihoods currently in use in statistics. A question of major interest focuses on the performance of this new construction relative to ordinary parametric likelihood ratios. The EL method for testing has been dealt with extensively in the literature (e.g. Owen, 2001; Wang and Rao, 2002a,b; Yu et al., in press). An outline of the EL approach can be presented as follows. Given independent identically distributed observations $X_1, \ldots, X_n$, the EL function has the form of $L_p = \prod_{i=1}^n p_i$, where the components $p_i, i = 1, \ldots, n$ maximize the likelihood L_p (maximum likelihood estimation) provided that empirical constraints, based on $X_1, \ldots, X_n$, are in effect (e.g., $\sum_{i=1}^n p_i = 1$, $\sum_{i=1}^n p_i X_i = 0$, under the hypothesis $EX_1 = 0$). Computation of the EL's components $p_i, i = 1, \ldots, n$ was used to be a simple exercise in Lagrange multipliers. This nonparametric approach is a product of consideration of the 'distribution functions'-based likelihood $\prod_{i=1}^n F(X_i) - F(X_i-)$ over all distribution functions F. Results presented in this article utilize the main idea of this EL technique; however, we would like to present a density-based likelihood methodology for goodness-of-fit testing.

Many statistical procedures are, strictly speaking, only appropriate when a parametric assumption about data distribution is made, and such an assumption may need to be tested. Thus, in many cases related to biostatistics projects, reliability studies, and engineering and management sciences, it is required to detect whether the assumption of a particular distribution is statistically justified. Testing distribution assumptions, in general, and for normality and uniformity, in particular, has been one of the major areas of continuing statistical research both theoretically and practically. In most applications of statistical modeling, it is convenient to assume normality of the underlying error distribution. Before making such an assumption about the residual features of a model, a check against possible departures is often appropriate. Hence, the problem of testing composite hypotheses of normality (when the mean and variance are unknown) is, perhaps, the most common goodness-of-fit problem in current statistical practice. Historically, tests for normality have been widely addressed in the literature (e.g. Pearson, 1930). Although it is difficult to propose a test for normality competing with the highly efficient family of Shapiro–Wilk tests (e.g. Royston, 1982; Shapiro and Wilk, 1965; Shapiro et al., 1968; Verrill and Johnson, 1987), development of simple and powerful, with respect to several alternative distributions, tests for normality is still an issue of concern in the area of statistical methodology. (Note that the Shapiro–Wilk test (Shapiro and Wilk, 1965) includes components that should be tabulated and, hence, are difficult to be computed for different sample sizes, e.g., Verrill and Johnson (1987).) Various statistical phenomena, characterizing normal distributions, can be used as basic ingredients for normality testing (e.g. Hall and Welsh, 1983; Lin and Mudholkar, 1980; Oja, 1983; Zhang, 2002). Vasicek (1976) constructed a test based on the entropy characterization of normality. This entropy-based test appears to compare reasonably with the existing tests for normality in power and convenience (e.g. Arizono and Ohta, 1989; Park and Park, 2003; Prescott, 1976; Senoglu and Sürücü, 2004; Tusnady, 1977; Vasicek, 1976). The simple method proposed by Vasicek (1976) does not require transformations to uniformity as in Kolmogorov–Smirnov and Anderson–Darling type tests (e.g. Stephens, 1974) or use of tables as in the Shapiro–Wilk test. However, the entropy-based test statistic includes an integer parameter, whose optimal values are, generally speaking, unknown (in the above literature, authors considered these values asymptotically or/and conditionally on information about alternative distributions). Wrong selected values of the integer parameter strongly decrease the power of the test. This reduces the ability of the entropy-based test for being applied to real data studies. Dong and Giles (2007) proposed an EL ratio test for normality. These authors directly utilize the EL technique outlined above in this section. Dong and Giles (2007) considered the empirical H_0-likelihood function L_p, the denominator of the likelihood ratio, that was obtained under constraints corresponding to four moment equations to set up skewness and kurtosis coefficients as 0 and 3, respectively. Although the power of the Dong and Giles's test suffers from ignoring the role of f_{H_0} in constituting the likelihood ratio, Dong and Giles (2007) reported good Monte Carlo operating characteristics of the test.

Results presented in this article also deal with uniformity testing. Various tests for uniformity have been proposed in the literature (e.g. Durbin, 1973; Dudewicz and Van Der Meulen, 1981; Stephens, 1974). Frequently, practical applications of testing for uniformity follow from the fact that if the null hypothesis H_0 states $F(x) = F_0(x)$, where $F_0(x)$ is continuous and completely specified, then tests of uniformity also allow one to test H_0. Such situations commonly correspond to control problems, where the baseline in-control distribution is completely known. Dudewicz and Van Der Meulen (1981) investigated the Monte Carlo power of the entropy-based test for uniformity and showed that this test is superior to several known tests against different alternatives. The test proposed by Dudewicz and Van Der Meulen (1981) has the same drawback, with respect to the test definition, as that belongs to the entropy-based test for normality.

In the Section 2, we shall describe our technique, supported by both the empirical and theoretical arguments, in greater detail. We show that the proposed tests are asymptotically consistent and have a density-based likelihood ratio structure. We also provide Monte Carlo simulation results, real data examples and some concluding remarks.

2. Methodology

In this section, we derive the EL ratio tests for normality. The analysis of Section 2.1 is relatively clear, and has the basic ingredients for more general cases (Sections 2.2 and 2.3) in which our aim is to suggest simple EL goodness-of-fit tests corresponding to situations where distribution functions, under the null hypothesis, are known up-to-parameters or completely defined.

2.1. The EL ratio test for normality

Suppose that the data consist of independent observations $X_1, \ldots, X_n$. Consider the problem of testing the composite hypothesis that a sample $X_1, \ldots, X_n$ is from a normal population. Obviously, when density functions f_{H_1} and f_{H_0}, corresponding to the baseline and alternative hypotheses, are completely known, the most powerful test statistic is the likelihood ratio

$$\frac{\prod_{i=1}^{n} f_{H_1}(X_i)}{\prod_{i=1}^{n} f_{H_0}(X_i)} = \frac{\prod_{i=1}^{n} f_{H_1}(X_i)}{(2\pi\sigma^2)^{-n/2}\exp\left(-\sum_{i=1}^{n}(X_i-\mu)^2/2\sigma^2\right)}, \tag{2.1.1}$$

where under the null hypothesis $X_1, \ldots, X_n$ are normal with mean μ and variance σ^2. In the case of unknown μ and σ^2, the maximum likelihood estimation applied to (2.1.1) changes the ratio (2.1.1) to

$$\frac{\prod_{i=1}^{n} f_{H_1}(X_i)}{\left(2\pi e s^2\right)^{-n/2}}, \quad s^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \frac{1}{n}\sum_{k=1}^{n} X_k\right)^2. \tag{2.1.2}$$

Let us apply the maximum EL technique to nonparametric estimate of the numerator of the ratio (2.1.2). To this end, rewrite the likelihood function $L_f = \prod_{i=1}^{n} f(X_i)$ in the form of

$$L_f = \prod_{i=1}^{n} f(X_i) = \prod_{i=1}^{n} f(X_{(i)}) = \prod_{i=1}^{n} f_i, \quad f_i = f(X_{(i)}), \tag{2.1.3}$$

where $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ are the order statistics based on the observations $X_1, \ldots, X_n$. Following the maximum EL methodology, mentioned in Section 1, we must derive values of f_i, $i = 1, \ldots n$ that maximize L_f and satisfy empirical constraints corresponding to the alternative hypothesis. In this case, the equation $\int f(u)\mathrm{d}u = 1$ constrains values of f_i, $i = 1, \ldots n$. To formalize this constraint, we present the following general proposition.

Proposition 2.1. *Let $f(x)$ be a density function. Then*

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

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

Proof in Appendix.

Particularly, since $\int_{X_{(1)}}^{X_{(n)}} f(x)\,\mathrm{d}x \le \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$, Proposition 2.1 denotes that

$$\Delta_m \le 1, \quad \Delta_m = \frac{1}{2m}\sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)\,\mathrm{d}x. \tag{2.1.4}$$

Moreover, one can expect that $\Delta_m \approx 1$ when $m/n \to 0$ as $m, n \to \infty$.

The inequality (2.1.4) can be approximated by

$$\frac{1}{2m}\sum_{j=1}^{n} g_{jm}\left(f_{j-m}, \ldots, f_{j+m}, X_{(j-m)}, \ldots, X_{(j+m)}\right) \le 1,$$

where g_{jm} is an approximate expression of the integral $\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)\,\mathrm{d}x$. For simplicity, in this article, we approximate $\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)\,\mathrm{d}x \cong g_{jm} = \left(X_{(j+m)} - X_{(j-m)}\right) f_j$ and represent the condition (2.1.4) in the empirical form

$$\tilde{\Delta}_m \le 1, \quad \tilde{\Delta}_m = \frac{1}{2m}\sum_{j=1}^{n}\left(X_{(j+m)} - X_{(j-m)}\right) f_j. \tag{2.1.5}$$

Please cite this article in press as: Vexler, A., Gurevich, G., Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.09.025

Table 1
Monte Carlo power estimates of some tests for normality; $a = 0.05$.

	T_{1n}	T_{2n}	T_{3n}	T_{4n}	T_{5n}	T_{6n}	T_{7n}	T_{10n}	T_{nn}	MT_n	V_n^1	V_n^2	W
$n = 20$													
Exp(1)	0.6686	0.8100	0.8428[a]	0.8526	0.8454	0.8358	0.8180	0.6840	0.1310	0.6700	0.8555	0.8562	0.8330
Gamma(2, 1)	0.2820	0.4079	0.4570[a]	0.4724	0.4659	0.4546	0.4352	0.3194	0.0919	0.2817	0.4748	0.4742	0.5281
Unif(0, 1)	0.2919	0.3867	0.4206[a]	0.4373	0.4413	0.4489	0.4558	0.4671	0.5411	0.2876	0.4380	0.4379	0.1968
Betta(2, 1)	0.2752	0.3876	0.4322[a]	0.4538	0.4548	0.4544	0.4484	0.3926	0.2971	0.2744	0.4474	0.4519	0.3058
Cauchy(0, 1)	0.7259	0.7639	0.7405[a]	0.6888	0.6138	0.5350	0.4632	0.3727	0.0015	0.7228	0.6516	0.7097	0.8630
Lnorm(0, 1)	0.8100	0.9011	0.9233[a]	0.9289	0.9267	0.9211	0.9121	0.8257	0.0704	0.8126	0.9393	0.9386	0.9322
$n = 50$													
Exp(1)	0.9786	0.9975	0.9991	0.9995[a]	0.9997	0.9996	0.9994	0.9989	0.1300	0.9802	0.9996	0.9996	0.9994
Gamma(2, 1)	0.6286	0.8216	0.8840	0.9107[a]	0.9220	0.9243	0.9232	0.8914	0.1064	0.6412	0.9362	0.9299	0.9479
Unif(0, 1)	0.6402	0.8327	0.8954	0.9260[a]	0.9438	0.9554	0.9644	0.9754	0.9839	0.9305	0.9442	0.9422	0.7478
Betta(2, 1)	0.6052	0.8172	0.8886	0.9197[a]	0.9355	0.9460	0.9500	0.9474	0.7004	0.7706	0.9392	0.9398	0.8426
Cauchy(0, 1)	0.9769	0.9875	0.9893	0.9882[a]	0.9849	0.9789	0.9702	0.9006	0.0000	0.9736	0.4789	0.9774	0.9960
Lnorm(0, 1)	0.9951	0.9991	0.9996	0.9998[a]	0.9998	0.9999	0.9999	0.9996	0.0389	0.9953	0.9999	0.9998	0.9994
$n = 70$													
Exp(1)	0.9978	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.1277	0.9979	1.0000	1.0000	1.0000
Gamma(2, 1)	0.7784	0.9316	0.9648	0.9780	0.9821	0.9832	0.9832	0.9761	0.1098	0.7864	0.9824	0.9852	0.9926
Unif(0, 1)	0.7860	0.9416	0.9722	0.9854	0.9906	0.9937	0.9958	0.9976	0.9997	0.9936	0.9928	0.9947	0.9349
Betta(2, 1)	0.7610	0.9321	0.9687	0.9833	0.9882	0.9905	0.9917	0.9921	0.8611	0.9144	0.9905	0.9907	0.9661
Cauchy(0, 1)	0.9953	0.9983	0.9988	0.9987	0.9985	0.9981	0.9970	0.9885	0.0000	0.9947	0.3333	0.9966	0.9996
Lnorm(0, 1)	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0246	0.9997	1.0000	1.0000	1.0000

MT_n denotes the naive maximum EL test $\max_{m<n/2} T_{mn}$; $V_n^j, j = 1, 2$ by (2.1.11), $\delta = 0.5$; W corresponds to the Shapiro and Wilk (1965) W test.
[a] Indicates cases of the entropy-based tests T_{mn} with most favorable m's recommended by Vasicek (1976).

Thus, values of $f_1, \ldots, f_n$, which maximize $\log(L_f)$ and satisfy $\tilde{\Delta}_m \le 1$ from (2.1.5), can be derived using Lagrange multipliers:

$$f_i = \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}$$

(here $X_{(j)} = X_{(1)}$, if $j \le 1$, and $X_{(j)} = X_{(n)}$, if $j \ge n$).

Therefore, the maximum EL method applied to (2.1.2) forms the likelihood ratio test statistic

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

that is known, for $m < n/2$, to be an efficient test statistic based on sample entropy (e.g. Vasicek, 1976; Arizono and Ohta, 1989; Park and Park, 2003). The tests for normality based on sample entropy are exponential rate optimal procedures (Tusnady, 1977). The sample entropy has been most widely used as a nonparametric entropy estimator owing to its simplicity (e.g. Vasicek, 1976; Park and Park, 2003). This agrees with the fact that commonly likelihood ratio tests have optimal statistical properties and likelihood ratio type decision rules are simple in applications. Note that we can rewrite the entropy-based test statistic in the EL ratio form

$$T_{mn} = \frac{\prod_{i=1}^{n} \left(F_n\left(X_{(i+m)}\right) - F_n\left(X_{(i-m)}\right)\right) / \left(X_{(i+m)} - X_{(i-m)}\right)}{\max_{\mu,\sigma} \prod_{i=1}^{n} f_{H_0}\left(X_i; \mu, \sigma\right)},$$

where F_n is the empirical distribution function. With this, following Vasicek (1976), $n^{-1} \log \prod_{i=1}^{n} 2m\left[n\left(X_{(i+m)} - X_{(i-m)}\right)\right]^{-1}$ does not approximate, under H_1, the log-likelihood $n^{-1} \log f_{H_1}(X)$, if m is fixed as $n \to \infty$ (i.e., $m \to \infty$ should be considered).

The power of the test based on statistic T_{mn} strongly depends on values of m. Assuming information regarding the distribution functions of the alternative hypothesis, Monte Carlo simulation results, published in the relevant literature, point out values of m (subject to n) that provide high levels of the power of the test based on T_{mn}. However, for example when $n = 50$ the test statistic T_{mn} with $m = 1$ was used to compare unfavorably with other tests for normality (see, for example, Table 1 in Section 3). This lack of balance restricts fields of real data applications of tests based on the sample entropy. The proposed EL point of view can attend to this issue.

First reconsideration of the test statistic. To derive the EL ratio T_{mn}, the constraint (2.1.4) was taken into account. Assume that $\Delta_{m_0} = (2m_0)^{-1} \sum_{j=1}^{n} \int_{X_{(j-m_0)}}^{X_{(j+m_0)}} f(x)\,dx > 1$, for some m_0. In this case, Proposition 2.1 concludes $\int_{X_{(1)}}^{X_{(n)}} f(x)\,dx > 1$, which is

Please cite this article in press as: Vexler, A., Gurevich, G., Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.09.025

unacceptable. Thus, it is reasonable to force f_i, $i = 1, \dots n$ to satisfy (2.1.5) for all $m < n/2$. It is clear that for a fix integer $m_1 < n/2$,

$$\max_{f_1,\dots,f_n:\tilde{\Delta}_r \le 1,\ \text{for all } r<n/2} \prod_{i=1}^{n} f_i \le \max_{f_1,\dots,f_n:\tilde{\Delta}_{m_1} \le 1} \prod_{i=1}^{n} f_i.$$

Therefore,

$$\max_{f_1,\dots,f_n:\tilde{\Delta}_r \le 1,\ \text{for all } r<n/2} \prod_{i=1}^{n} f_i \le \min_{1\le m<n/2} \max_{f_1,\dots,f_n:\tilde{\Delta}_m \le 1} \prod_{i=1}^{n} f_i. \tag{2.1.6}$$

On the other hand, the empirical interpretation of Proposition 2.1 can be formulated as

$$\begin{aligned}\sum_{j=1}^{n} \int_{X_{(j-m)}}^{X_{(j+m)}} f(x)\,dx &= 2m \int_{X_{(1)}}^{X_{(n)}} f(x)\,dx - \sum_{k=1}^{m-1} (m-k)\left(F\left(X_{(n-k+1)}\right) - F\left(X_{(n-k)}\right)\right)\\ &\quad - \sum_{k=1}^{m-1} (m-k)\left(F\left(X_{(k+1)}\right) - F\left(X_{(k)}\right)\right)\\ &\cong 2m \int_{X_{(1)}}^{X_{(n)}} f(x)\,dx - \sum_{k=1}^{m-1} (m-k)\left(F_n\left(X_{(n-k+1)}\right) - F_n\left(X_{(n-k)}\right)\right)\\ &\quad - \sum_{k=1}^{m-1} (m-k)\left(F_n\left(X_{(k+1)}\right) - F_n\left(X_{(k)}\right)\right)\\ &= 2m \int_{X_{(1)}}^{X_{(n)}} f(x)\,dx - \frac{m(m-1)}{n},\end{aligned} \tag{2.1.7}$$

where F and F_n are the theoretical and empirical distribution functions, respectively. Therefore, by virtue of the definition (2.1.4), $\Delta_{\mathrm{m}} \cong \int_{X_{(1)}}^{X_{(n)}} f(x)\,dx - (m-1)/(2n)$ and hence $\Delta_k \tilde{\gtrsim} \Delta_r$, for $k < r$. This leads us to consider that $\tilde{\Delta}_1 \tilde{\gtrsim} \tilde{\Delta}_2 \tilde{\gtrsim} \tilde{\Delta}_3 \tilde{\gtrsim} \cdots$ (here $\tilde{\Delta}_m$, from (2.1.5), approximate Δ_m). Thus, if f_i, $i = 1, \dots n$ satisfy $\tilde{\Delta}_1 \le 1$, taking into account $\Delta_k \tilde{\gtrsim} \Delta_r (k < r)$, than we would like to claim that these f_i, $i = 1, \dots n$ satisfy $\tilde{\Delta}_m \le 1$, for all $m < n/2$, i.e., the maximum density-based EL is

$$\max_{\substack{f_1,\dots,f_n:\tilde{\Delta}_r \le 1,\ \text{for all } r<n/2,\\ \text{taking into account } \Delta_k \tilde{\gtrsim} \Delta_r (k<r),}} \prod_{i=1}^{n} f_i \cong \max_{f_1,\dots,f_n:\tilde{\Delta}_1 \le 1} \prod_{i=1}^{n} f_i \ge \min_{1\le m<n/2} \max_{f_1,\dots,f_n:\tilde{\Delta}_m \le 1} \prod_{i=1}^{n} f_i. \tag{2.1.8}$$

That is, on the one hand, $f_1, \dots, f_n$ need to satisfy $\tilde{\Delta}_m \le 1$, for all $m < n/2$, and, on the other hand, if $\tilde{\Delta}_1, \tilde{\Delta}_2, \dots$ are close to $\Delta_1, \Delta_2, \dots$ then $f_1, \dots, f_n$ can comply with $\tilde{\Delta}_1 \le 1$ only. By virtue of (2.1.6) and (2.1.8), this implies that the maximum EL should be defined as

$$\min_{1\le m<n/2} \max_{f_1,\dots,f_n:\tilde{\Delta}_m \le 1} \prod_{i=1}^{n} f_i.$$

Finally, we propose the empirical modification of the test statistic (2.1.2) in the form of

$$V_n^1 = \min_{1\le m<n/2} \left(2\pi e s^2\right)^{n/2} \prod_{i=1}^{n} \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}. \tag{2.1.9}$$

Second reconsideration of the test statistic. The empirical form (2.1.7) of Proposition 2.1 concludes that

$$\Delta_{\mathrm{m}} \cong \int_{X_{(1)}}^{X_{(n)}} f(x)\,dx - \frac{(m-1)}{2n}.$$

Thus, to let the constraints $\Delta_m \le 1$ to be in effect, we must require $m/n \to 0$ as $m, n \to \infty$. This obviously displays the condition on m mentioned in asymptotic theorems of Vasicek (1976) and Tusnady (1977). In this article, we propose considering m in the range $\left(1, n^{1-\delta}\right)$, $0 < \delta < 1$, and hence the empirical modification of the test statistic (2.1.2) has the form of

$$V_n^2 = \min_{1\le m<n^{1-\delta}} \left(2\pi e s^2\right)^{n/2} \prod_{i=1}^{n} \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}, \tag{2.1.10}$$

where $0 < \delta < 1$.

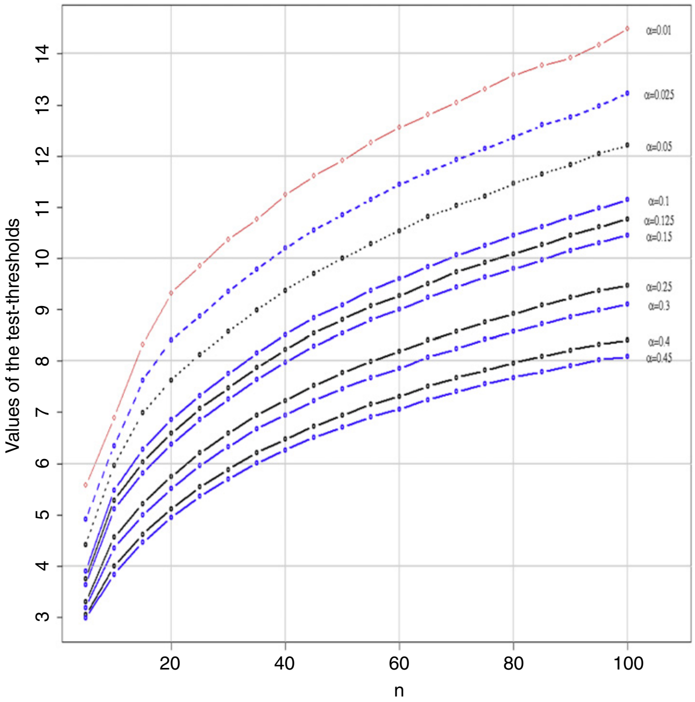

Fig. 1a. The curves present the values of thresholds C_α for the test $\log\left(V_n^1\right)$ by (2.1.11), corresponding to the significance levels $\alpha = 0.01, 0.025, 0.05, 0.1, 0.125, 0.15, 0.25, 0.3, 0.4, 0.45$ that are plotted against the sample sizes $n = 5, 10, 15, \dots, 100$.

(Note that, we minimize the entropy-based test statistic T_{mn} with respect to m. The remark following Proposition 2.2 of the next section asymptotically argues this suggested minimization.)

Suggested tests. Define the following decision rules. We reject the null hypothesis iff

$$\log\left(V_n^j\right) > C, \tag{2.1.11}$$

where C is a test-threshold, $j = 1, 2$ and V_n^j are the test statistics defined in (2.1.9) and (2.1.10), respectively.

In this article, Proposition 2.2 of the next section evaluates asymptotically the operating characteristics of the test (2.1.11) with $j = 2$. (We do not prove the asymptotic consistency of the test (2.1.11) with $j = 1$, but owing to high levels of the power of this test against alternatives, which are considered in Section 3 in several situations, we can recommend the test (2.1.11) with $j = 1$ to be applied.)

Significance level of the tests. Since

$$\sup_{\mu,\sigma} P_{H_0}\left\{\log\left(V_n^j\right) > C\right\} = P_{X_1,\dots,X_n\sim N(0,1)}\left\{\log\left(V_n^j\right) > C\right\}, \quad j = 1, 2,$$

the type I error of the tests (2.1.11) can be calculated exactly. Note that, a very substantial body of literature has now grown around the asymptotic distributional problems involving the Vasicek's entropy estimator and the analogous statistics (e.g. Dudewicz and Van Der Meulen, 1981; van Es, 1992). However, it is generally recognized that even the asymptotic distribution of the statistic T_{mn}, which includes the estimates of nuisance parameter of f_{H_0}, is analytically difficult. (When the sample size is relatively large, we can also assume various tests provide powerful inference.) Thus, following the recent literature related to goodness-of-fit tests (e.g. Hall and Welsh, 1983; Mudholkar and Tian, 2002, 2004), we will not attempt to provide here an analytical solution for the critical values for the proposed tests. Set up $\delta = 0.5$ in the definition of the statistic V_n^2. Fig. 1 plots Monte Carlo roots C_α of the equations $P_{X_1,\dots,X_n\sim N(0,1)}\left\{\log\left(V_n^j\right) > C_\alpha\right\} = \alpha$, for different values of α and n. (For each value of α and n, the type I error results were derived from 75,000 samples of size n.)

Remark. The empirical conclusion (2.1.7) can adjust the constraint (2.1.5), i.e., $\tilde{\Delta}_m \le 1 - (m-1)/(2n)$ can be required. This defines

$$f_i = \frac{2m - m(m-1)/n}{n\left(X_{(i+m)} - X_{(i-m)}\right)}, \quad i = 1, \dots, n$$

that maximize L_f under condition $\tilde{\Delta}_m \le 1 - (m-1)/(2n)$. Thus, forms of the test statistics T_{mn}, V_n^1, V_n^2 can be adjusted. Moreover, one can naively consider m as an unknown parameter and estimate it based on the maximum EL. In this case,

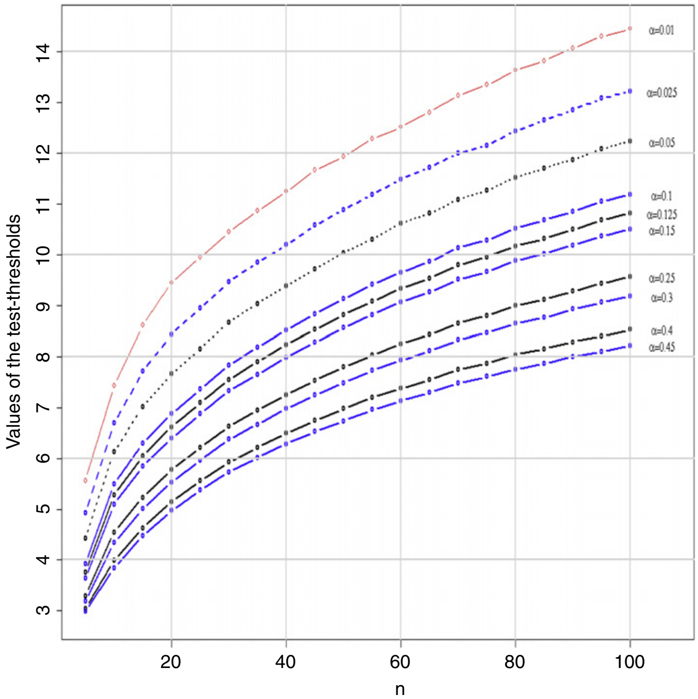

Fig. 1b. The curves present the values of thresholds C_α for the test $\log\left(V_n^2\right)$ by (2.1.11) with $\delta = 0.5$, corresponding to the significance levels $\alpha = 0.01, 0.025, 0.05, 0.1, 0.125, 0.15, 0.25, 0.3, 0.4, 0.45$ that are plotted against the sample sizes $n = 5, 10, 15, \ldots, 100$.

the test statistics are

$$VG_n^1 = \max_{1\le m<n/2} \left(2\pi es^2\right)^{n/2} \prod_{i=1}^{n} \frac{2m - m(m-1)/n}{n\left(X_{(i+m)} - X_{(i-m)}\right)}, \qquad VG_n^2 = \max_{1\le m<n^{1-\delta}} \left(2\pi es^2\right)^{n/2} \prod_{i=1}^{n} \frac{2m - m(m-1)/n}{n\left(X_{(i+m)} - X_{(i-m)}\right)},$$

where $0 < \delta < 1$. However, our broad Monte Carlo investigations compared unfavorably the powers of these tests (utilizing VG_n^1, VG_n^2, and the modifications of T_{mn}, V_n^1, V_n^2) with the proposed tests for normality.

2.2. The EL ratio goodness-of-fit test for composite hypotheses

Given a random sample $X_1, \ldots, X_n$ from a population with a density function f and a finite variance, consider the problem of testing for

$$H_0 : f = f_{H_0} \quad \text{vs} \quad H_1 : f = f_{H_1}, \tag{2.2.1}$$

where, under the alternative hypothesis, f_{H_1} is completely unknown, whereas, under the null hypothesis, $f_{H_0}(x) = f_{H_0}(x; \boldsymbol{\theta})$ is known up to the vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ (here, $d \ge 1$ defines a dimension of the vector $\boldsymbol{\theta}$). In accordance with the technique mentioned in Section 2.1, in the context of the statement (2.2.1), we suggest the test statistic

$$G_n = \min_{1\le m<n^{1-\delta}} \frac{\prod_{i=1}^{n} \frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)}}{\prod_{i=1}^{n} f_{H_0}\left(X_i, \hat{\boldsymbol{\theta}}\right)}, \tag{2.2.2}$$

where $0 < \delta < 1$ and $\hat{\boldsymbol{\theta}}$ estimates $\boldsymbol{\theta}$ (e.g., $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator of $\boldsymbol{\theta}$).

To examine asymptotic properties of the statistic (2.2.2), we denote

$$g_i(x, \boldsymbol{\theta}) = \frac{\partial}{\partial\theta_i} \log f_{H_0}(x; \boldsymbol{\theta}), \quad i = 1, \ldots, d,$$

as well as $|\mathbf{y}| = \max_{1\le i\le d} |y_i|$ (where $\mathbf{y} = (y_1, \ldots, y_d)$), and assume the following conditions hold:

(C1) $E\left(\log f(X_1)\right)^2 < \infty$;
(C2) under the null hypothesis H_0, $\left|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right| \xrightarrow{P} 0$ as $n \to \infty$;
(C3) under the alternative H_1, $\hat{\boldsymbol{\theta}} \xrightarrow{P} \mathbf{a}$ as $n \to \infty$, where $\mathbf{a} = (a_1, \ldots, a_d)$ is a vector with finite components;
(C4) there are open intervals $\Theta_0 \subseteq R^d$ and $\Theta_{\mathbf{a}} \subseteq R^d$ containing $\boldsymbol{\theta}$ and $\mathbf{a}$, respectively, as well as there exists a function $t(x)$ such that $|g_i(x, \boldsymbol{\eta})| \le t(x)$, for all $x \in R$, $\boldsymbol{\eta} \in \Theta_0 \cup \Theta_{\mathbf{a}}$, $1 \le i \le d$, and $Et(X_1) < \infty$.

Proposition 2.2. *Assume* (C1)–(C4). *Then, under* H_0,

$$n^{-1}\log(G_n) \xrightarrow{P} 0,$$

while, under H_1,

$$n^{-1}\log(G_n) \xrightarrow{P} E\log\left(\frac{f_{H_1}(X_1)}{f_{H_0}(X_1;\mathbf{a})}\right) > 0,$$

as $n \to \infty$.

Proof in Appendix.

Given conditions (C1)–(C4), Proposition 2.2 shows that, with a test-threshold C_α related to the type I error $\alpha = \sup_\theta P_{H_0}(\log(G_n) > C_\alpha)$ in mind,

$$P_{H_1}(\log(G_n) > C_\alpha) \xrightarrow[n\to\infty]{} 1.$$

Therefore, we suggest the asymptotic power one (i.e., consistent) tests.

Remarks. The formal proof of this proposition is presented in Appendix. We just note that in accordance with the proof scheme of Proposition 2.2, the proposed test statistic $n^{-1}\log(G_n)$ approximates the most powerful log-likelihood ratio $n^{-1}\log\left(f_{H_1}(X)/f_{H_0}(X)\right)$. This demonstrates that the suggested testing has a density-based likelihood ratio structure. (Moreover, utilizing the proof scheme of Proposition 2.2, one can show that the naive "maximum EL estimation" of m applied to the EL ratio

$$R_{mn} = \prod_{i=1}^{n}\frac{2m}{n\left(X_{(i+m)} - X_{(i-m)}\right)} \Big/ \prod_{i=1}^{n} f_{H_0}\left(X_i;\hat{\boldsymbol{\theta}}\right)$$

leads to the test statistic $\max_m R_{mn}$ that is not consistent.)

Remark that while the density function $f_{H_0}(x;\boldsymbol{\theta})$ obeys some regularity conditions, the maximum likelihood estimator of $\boldsymbol{\theta}$ satisfies the condition $(C2)$ (e.g. Serfling, 1980, pp. 144–149).

2.3. The EL ratio test for uniformity

Consider (2.2.1), where f_{H_0} is a completely known continuous density function that is related to a distribution function F_{H_0}. At this rate, the transformation $Y_i = F_{H_0}(X_i)$, $i = 1,\dots,n$, causes observations $Y_1,\dots,Y_n$ to be uniformly distributed under the null hypothesis. Therefore, we aim to test for

$$H_0 : Y_1, Y_2, \dots, Y_n \sim Uni(0,1) \tag{2.3.1}$$

versus the alternative that $Y_1,\dots,Y_n$ are from a nonuniformly distribution $F(y)$ with a finite variance and continuous density function $f(y)$ concentrated on the interval $[0, 1]$. In accordance with (2.2.2), we suggest the following EL ratio test statistic

$$U_n = \min_{1\le m<n^{1-\delta}} \prod_{i=1}^{n}\frac{2m}{n\left(Y_{(i+m)} - Y_{(i-m)}\right)}, \tag{2.3.2}$$

where $0 < \delta < 1$. The event

$$\log(U_n) > C \tag{2.3.3}$$

implies that H_0 is rejected, C here is a test-threshold.

Note that the statistic

$$U_{mn} = \prod_{i=1}^{n}\frac{2m}{n\left(Y_{(i+m)} - Y_{(i-m)}\right)}$$

with a fixed $m < n/2$ is an intermediate product of the EL technique presented in Section 2.1. Dudewicz and Van Der Meulen (1981) considered U_{mn} as a test statistic of the entropy-based test for uniformity. This test is a very efficient decision rule provided that optimal values of m, subject to f_{H_1} and n, are applied to the statistic U_{mn} (Dudewicz and Van Der Meulen, 1981). In practice, since f_{H_1} is completely unknown, we risk choosing m that leads to a U_{mn}-based test having the power that is lower than that of other known tests for uniformity (e.g. Zhang, 2002). In contrast to this, a Monte Carlo study, presented in the next section, demonstrates that, in many cases, the test (2.3.3) provides the power that is close to the power of U_{mn}-based tests with optimal ms, calculated empirically (with respect to the type of f_{H_1}).

Since Proposition 2.2 is in effect, the test (2.3.3) is consistent as $n \to \infty$.

Significance level of the test. Fig. 2 depicts Monte Carlo roots C_α of the equations

$$P_{X_1,\dots,X_n\sim Uni(0,1)}\{\log(U_n) > C_\alpha\} = \alpha,$$

for $\delta = 0.5$ and different values of α and n. (For each value of α and n, the type I error result was derived from 75,000 samples of size n.)

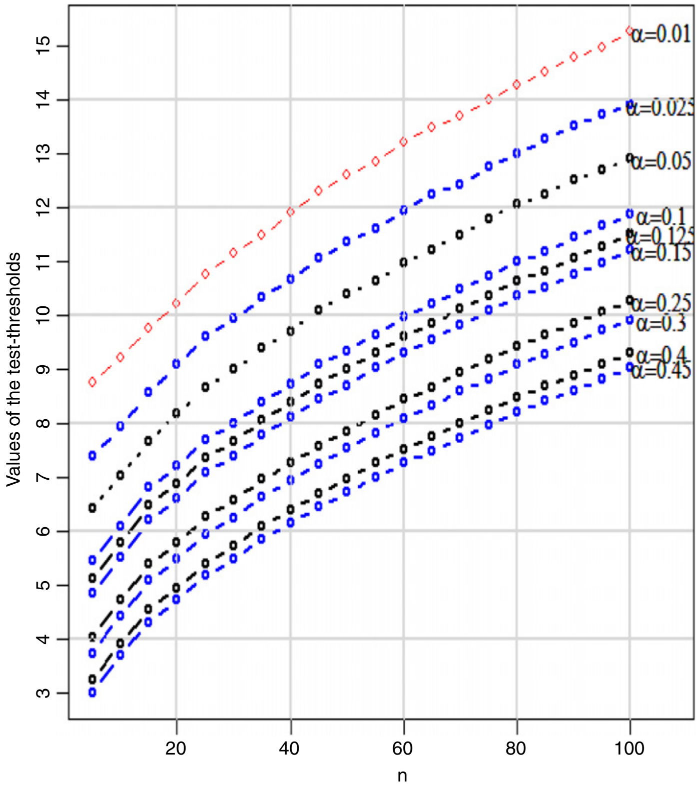

Fig. 2. The curves present the values of threshold C_α for the test (2.3.3) with $\delta = 0.5$, corresponding to $\alpha = 0.01, 0.025, 0.05, 0.1, 0.125, 0.15, 0.25, 0.3, 0.4, 0.45$ that are plotted against the sample sizes $n = 5, 10, 15, \ldots, 100$.

3. Monte Carlo study

3.1. The EL ratio test for normality

To study properties of the tests (2.1.11), Monte Carlo simulations were employed. For each $n = 20$, 50 and 70, 25000 samples for size $n$from Exponential, Gamma, Uniform, Beta, Cauchy, and Log-Normal distributions were produced. The entropy-based test statistic T_{mn} for several values of m was calculated from each sample. The number of times H_0 was rejected by each test for each sample size is shown in Table 1, at 5% level of significance.

Besides the Monte Carlo powers of the entropy-based tests and tests (2.1.11) with $\delta = 0.5$, Table 1 gives estimates of power of the Shapiro and Wilk (1965) W test to judge the proposed tests. The literature cited in Section 1 can be used to compare Table 1 with simulation results related to known tests for normality (e.g., proposed by Hall and Welsh, 1983; Lin and Mudholkar, 1980; Oja, 1983). We mark the power points of T_{mn}-based tests that correspond to m s recommended by Vasicek (1976). (The entropy-based test T_{mn} was Monte Carlo compared with different tests for normality by Arizono and Ohta (1989), Prescott (1976), Senoglu and Sürücü (2004) and Vasicek (1976); etc.) The naive consideration of m as an unknown parameter defines the "maximum" EL test $MT_n = \max_{m<n/2} T_{mn}$. Table 1 confirms the lack of this definition with respect to the power perspective. The proposed tests tend to be superior to other tests for normality against the considered alternatives. However, when an H_1-distribution is a Cauchy distribution, the Shapiro–Wilk test is more powerful than the entropy-based tests. Note that, in this article, we did not prove consistency of the test (2.1.11) with $j = 1$. Table 1 displays problems with this test against the Cauchy alternative when the sample size increases. This confirms our analysis mentioned in Section 2.1 (paragraph "Second reconsideration of the test statistic"), but because of high levels of the demonstrated power of the test (2.1.11) with $j = 1$, we are afraid to reject the practical meaning of this decision rule. It is interesting to remark, under the non-Cauchy alternative distributions, that observed histograms of the random values $m_{\min} = \arg\min_{m\leq n} T_{mn}$ were concentrated around the values of ms that were empirically recommended by Vasicek (1976) for the entropy-based test statistics. Fig. 3 plots the histograms of $m_{\min}$ based on Cauchy distributed data.

In these cases, the second mode of the Monte Carlo distributions of $m_{\min}$ is close to n. In accordance with asymptotic results of Vasicek (1976) as well as with the consideration in Section 2.1 confirmed by Table 1, T_{nn} is not an asymptotic power one test statistic. Therefore, the definition (2.1.11) with $j = 2$ reasonably corrects the test (2.1.11) with $j = 1$ eliminating the concentration of $m_{\min}$ near n.

Please cite this article in press as: Vexler, A., Gurevich, G., Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.09.025

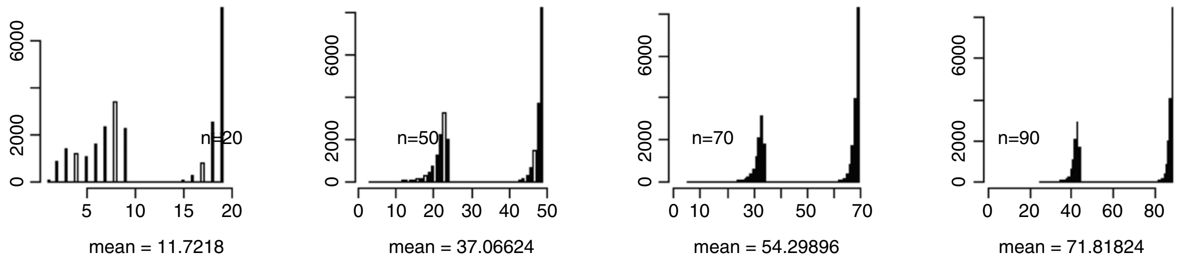

Fig. 3. The histograms of the random value $m_{\min}$, for different n under the Cauchy (location = 0, scale = 1) alternative.

3.2. The EL ratio test for uniformity

Dudewicz and Van Der Meulen (1981) compared the entropy-based test for uniformity with different procedures that include the Kolmogorov–Smirnov, Anderson–Darling, log-statistic, χ^2, and other tests. The authors considered the cases of distribution functions, under the alternative hypothesis H_1, in the forms of

$$A_k: \quad F(y) = 1-(1-y)^k, \quad 0 \le y \le 1; \qquad B_k: \quad F(y) = \begin{cases} 2^{k-1}y^k & 0 \le y \le 0.5 \\ 1-2^{k-1}(1-y)^k & 0.5 \le y \le 1, \end{cases}$$

$$C_k: \quad F(y) = \begin{cases} 0.5-2^{k-1}(0.5-y)^k, & 0 \le y \le 0.5 \\ 0.5+2^{k-1}(y-0.5)^k, & 0.5 \le y \le 1 \end{cases} \quad ; \quad k = 1.5, 2.0.$$

(Alternatives A, B, C were used by Stephens (1974) to compare the power of tests for uniformity. According to Stephens, the alternatives A give data points closer to zero than expected under the null hypothesis of uniformity, whereas B_k, $k = 1.5, 2$ give points near 0.5 and C_k, $k = 1.5, 2$ give two clusters (close to 0 and 1). Stephens (1974) interpreted alternatives A as a change in mean, alternatives B as a change toward a smaller variance, and alternatives C as a change toward a larger variance.) The Monte Carlo power comparisons confirmed that the entropy-based test is a very efficient procedure provided that optimal values of m in U_{mn}, depending on the f_{H_1} and n, are found (Dudewicz and Van Der Meulen, 1981, Table 3). The entropy-based test tends to be superior to the considered tests against alternatives B. However, the entropy-based test tends to be inferior when alternatives C are in effect and n is relatively small ($n = 10, 20$). When $n = 40$, under alternatives C, the Monte Carlo power of the entropy-based test was shown to be close or superior to the considered tests. Given alternatives A, the entropy-based test performs moderately well, providing power levels that are approximately similar to those of the other tests.

Compare the test (2.3.3) with the entropy-based tests proposed by Dudewicz and Van Der Meulen (1981). To this end, we conduct the Monte Carlo study based on a scheme (taking into account the alternative distributions A, B and C) that is suggested by Dudewicz and Van Der Meulen (1981). In Table 2, we report a Monte Carlo comparison between the powers of the test (2.3.3) (with $\delta = 0.5, 0.3, 0.1$) and the U_{mn}-based test, for two different values of m, corresponding to optimal and worse powers at the 10% level of significance. (Here, we fixed the 10% level of significance of the tests to utilize results presented by Dudewicz and Van Der Meulen (1981).) Table 2 contains the estimated powers of the widely used Kolmogorov–Smirnov test to judge the proposed tests.

The estimated powers of the entropy-based tests were taken from Table 2 of Dudewicz and Van Der Meulen (1981). The Monte Carlo powers of the proposed tests (2.3.3) and the Kolmogorov–Smirnov test were obtained utilizing 25 000 samples for size $n = 10, 30, 50, 100$. The critical values of the tests (2.3.3) and the Kolmogorov–Smirnov test were calculated using Monte Carlo techniques with 75 000 replications to guarantee the level of significance at 0.1. From Table 2, we conclude that under alternatives A and $B_{2.0}$ with $n = 10, 30, 50, 100$, the Monte Carlo power of the test (2.3.3), which is inessential depending on δ, is close to that of the entropy-based test with optimal m. While $B_{1.5}$ is in effect, in the cases of $n = 30, 50$ and $n = 10, 100$, the power of the test (2.3.3), which depends weakly on δ, is about 20% less than or close to that of the optimal (only theoretically existent) entropy-based test, respectively. Under alternatives C with $n = 10, 30, 50$, the power of the test (2.3.3) depends on δ. The value $\delta = 0.5$ provides the best power results. Moreover, when $n > 10$, the power of the test (2.3.3) with $\delta = 0.5$ approximates that of the entropy-based test with optimal m.

Thus, taking into account the simplicity for practical applications of the proposed test (2.3.3), we conclude that in many situations (e.g., testing for a possible change toward a smaller variance) the test (2.3.3) may be a preferable alternative to existing tests of uniformity. (For simplicity and by virtue of the presented Monte Carlo results, we can recommend applying $\delta = 0.5$ to the definition (2.3.3).)

4. Data examples

4.1. A TBARS data example

In this section, we use the proposed tests for normality based on data from a study that evaluates biomarkers related to atherosclerotic coronary heart disease. We consider the biomarker TBARS (Thiobarbituric Acid Reactive Substances),

Please cite this article in press as: Vexler, A., Gurevich, G., Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.09.025

Table 2
The Monte Carlo powers of the tests: U_{mn} (Dudewicz and Van Der Meulen, 1981), Kolmogorov–Smirnov and (2.3.3). For each $n = 10, 30, 50100$, two values of the U_{mn}-test's power are presented corresponding to ms that support optimal or worse U_{mn}-test statistics, respectively, in accordance with Table 2 of Dudewicz and Van Der Meulen (1981).

	Alternative $A_{1.5}$					Alternative $A_{2.0}$			
Sample size n	10	30	50	100	Sample size n	10	30	50	100
Entropy-based test U_{mn}									
	0.1725	0.2757	0.3851	0.5343		0.3124	0.6359	0.8183	0.9785
	0.2316	0.5106	0.7165	0.9383		0.4539	0.9086	0.9920	1.0000
Kolmogorov–Smirnov test									
	0.2509	0.5444	0.7414	0.9574		0.5387	0.9356	0.9949	1.0000
Proposed test									
$\delta = 0.5$	0.2328	0.4646	0.6506	0.9050	$\delta = 0.5$	0.4480	0.8910	0.9858	1.0000
$\delta = 0.3$	0.2396	0.4807	0.6612	0.9084	$\delta = 0.3$	0.4606	0.9005	0.9868	1.0000
$\delta = 0.1$	0.2394	0.4830	0.6616	0.9084	$\delta = 0.1$	0.4609	0.9013	0.9868	1.0000
	Alternative $B_{1.5}$					Alternative $B_{2.0}$			
n	10	30	50	100	n	10	30	50	100
Entropy-based test U_{mn}									
	0.1931	0.2919	0.3741	0.5532		0.3487	0.6597	0.8403	0.9792
	0.3318	0.6722	0.8884	0.9911		0.6217	0.9720	0.9994	1.0000
Kolmogorov–Smirnov test									
	0.0873	0.1778	0.2822	0.5578		0.1221	0.4422	0.7316	0.9868
Proposed test									
$\delta = 0.5$	0.2981	0.5374	0.7250	0.9350	$\delta = 0.5$	0.5494	0.9277	0.9934	1.0000
$\delta = 0.3$	0.3145	0.5605	0.7346	0.9378	$\delta = 0.3$	0.5778	0.9352	0.9939	1.0000
$\delta = 0.1$	0.3162	0.5620	0.7348	0.9378	$\delta = 0.1$	0.5793	0.9362	0.9940	1.0000
	Alternative $C_{1.5}$					Alternative $C_{2.0}$			
n	10	30	50	100	n	10	30	50	100
Entropy-based test U_{mn}									
	0.0452	0.0315	0.0091	0.0012		0.0411	0.0331	0.0081	0.0008
	0.1301	0.2784	0.4481	0.7777		0.2044	0.6622	0.9153	0.9992
Kolmogorov–Smirnov test									
	0.1909	0.3053	0.4145	0.6651		0.3133	0.6176	0.8243	0.9932
Proposed test									
$\delta = 0.5$	0.0673	0.2427	0.4178	0.7758	$\delta = 0.5$	0.0756	0.5893	0.8973	0.9994
$\delta = 0.3$	0.0536	0.1530	0.3972	0.7738	$\delta = 0.3$	0.0540	0.2769	0.8746	0.9993
$\delta = 0.1$	0.0507	0.1302	0.3804	0.7738	$\delta = 0.1$	0.0454	0.1596	0.6987	0.9993

that can quantify different phases of oxidative stress and antioxidant status process of an individual (e.g. Schisterman et al., 2001). A population-based sample of randomly selected residents of Erie and Niagara counties, 35 to 79 years of age, was the focus of this investigation (Schisterman et al., 2001). The sampling frame for adults between the ages of 35 and 65 was from the New York State Department of Motor Vehicle drivers' license rolls, while the elderly sample (age 65–79) was randomly chosen from the Health Care Financing Administration database. A cohort of 70 men and women were randomly selected for the analyses to present a population without myocardial infarction. Participants provided a 12-h fasting food specimen for biochemical analysis at baseline, and a number of parameters were examined from fresh blood samples.

Different studies based on independent data sets reported that measurements of TBARS follow normal distributions (e.g. Brown et al., 1998; Schisterman et al., 2001; Turoli et al., 2004; Yarman et al., 2003). This conclusion is important for analyzing properties of the TBARS biomarker, e.g., using the receiver operating characteristic curve (e.g. Schisterman et al., 2001). Fig. 4 depicts the histogram of TBARS biomarker measurements from a population without myocardial infarction.

In this case, the commonly used Shapiro–Wilk test for normality provides p-value $= 0.317$, the consistent test statistic $\log\left(V_n^2\right)$ from (2.1.11) with $\delta = 0.5$ has value 7.75 and hence, corresponding to Fig. 1b, the test's p-$value > 0.3$. In accordance with ideas introduced by Stigler (1977), we organized a bootstrap type study to examine the test based on the statistic $\log\left(V_n^2\right)$. The strategy was that a sample with size 50 was randomly selected from the TBARS data to be tested for normality at 5% level of significance. We repeated this strategy 10 000 times calculating the frequencies of the event $\log\left(V_n^2\right) > C_{0.05} = 10.035724$. This bootstrap type procedure showed that the proposed test rejected the normality assumption in 365 cases. The Shapiro–Wilk test rejected the normality assumption in 1846 cases ($\alpha = 0.05$). In this study, 269 times these tests rejected the null hypothesis together. Thus, the proposed test can be recommended to be applied for analyzing the TBARS biomarker based on data from different biomedical studies. (In this study, the test statistic $\log\left(V_n^1\right)$ by (2.1.11) rejected the null hypothesis in 401 cases.)

4.2. Published data examples

Folks and Chhikara (1978) as well as Mudholkar et al. (2001) presented and discussed data sets to illustrate the appropriateness of the inverse Gaussian distribution in applications. The authors proved the data sets "D1: shelf life of a

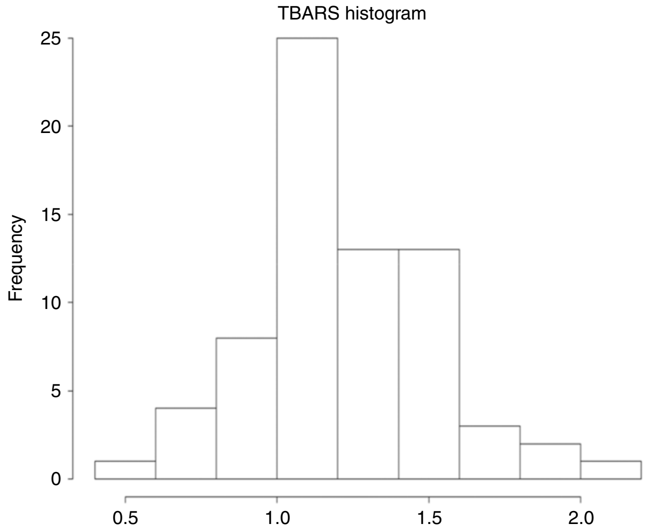

Fig. 4. The histogram of TBARS measurements.

Table 3
Comparison of *p-values* of the tests for normality.

Data set	W	V_n^1	V_n^2
D1	0.0518	<0.025 ($\log V_n^1 = 9.44784$)	<0.025 ($\log V_n^2 = 9.44784$)
D2	0.4473	<0.25 ($\log V_n^1 = 5.75710$)	<0.25 ($\log V_n^2 = 5.75710$)

$V_n^j, j = 1, 2$ by (2.1.11); W corresponds to the Shapiro and Wilk (1965) W test.

food product in days" and "D2: toughness of MIG welds" follow inverse Gaussian distributions. Thus, we expect tests for normality should provide relatively small p-values, when these data sets are utilized. Table 3 presents outputs of the tests (2.1.11) and the commonly used Shapiro–Wilk test based on the data D1 and D2.

Section 4.1 demonstrated an efficiency of the proposed tests in the context of the type I error control. In Section 4.2, we show the tests (2.1.11) are reasonable from a power perspective to be applied in practice.

4.3. Concluding remarks

In this article, we have presented a methodology for developing density-based EL tests. We have proved that the entropy-based tests for goodness-of-fit have the EL ratio structure. The proposed maximum likelihood approach also has enabled us to define steady efficient entropy-based test statistics. This technique has been supported by both empirical and asymptotic arguments. Commonly, the proof schemes of the Neyman–Pearson type lemmas can be applied to show optimal properties of likelihood ratio tests (e.g., being locally most powerful decision rules). Since, in practice, sample sizes are limited and there are a lot of asymptotic power one tests, we would like to focus on the empirical reasoning evaluated by the Monte Carlo study. (For example, Tusnady (1977) proposed optimal tests for normality based on sample entropy with m in T_{mn} that satisfies $m \log(n)/n \to 0$, as $m, n \to \infty$. When $n = 20, 30, 50$, the practical meaning of this rule is unclear.)

We conducted broad Monte Carlo analysis (partly presented in this article) for judging our conclusions and the suggested tests. Thus, we can infer that the proposed goodness-of-fit tests are simple and powerful procedures that can be applied to real data studies.

We focused on the tests for normality and uniformity. The test for exponentiality based on Kullback–Leibler information (Ebrahami et al., 1992) can be also considered paying attention to Section 2.2.

The proposed density-based EL ratio technique can be applied to create test statistics that approximate nonparametrically powerful parametric likelihood ratios in various statistical problems. For example, in a subsequent paper we plan to address two-sample EL tests based on samples entropy.

Acknowledgements

We are grateful to the co-editor, associate editor, and referee for their helpful comments that clearly improved this article. This work was partially supported by the Internal Funding Program of the Shamoon College of Engineering (SCE).

Please cite this article in press as: Vexler, A., Gurevich, G., Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.09.025

Appendix

Proof of Proposition 2.1. Since $f(x)$ is a density function, we have

$$\begin{aligned}\sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)\,\mathrm{d}x &= \sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j-m+1)}} f(x)\,\mathrm{d}x + \sum_{j=1}^{n}\int_{X_{(j-m+1)}}^{X_{(j+m-1)}} f(x)\,\mathrm{d}x + \sum_{j=1}^{n}\int_{X_{(j+m-1)}}^{X_{(j+m)}} f(x)\,\mathrm{d}x \\ &= \sum_{j=1}^{n}\int_{X_{(j)}}^{X_{(j+1)}} f(x)\,\mathrm{d}x - \int_{X_{(n-m+1)}}^{X_{(n)}} f(x)\,\mathrm{d}x + \sum_{j=1}^{n}\int_{X_{(j-m+1)}}^{X_{(j+m-1)}} f(x)\,\mathrm{d}x \\ &\quad + \sum_{j=1}^{n}\int_{X_{(j)}}^{X_{(j+1)}} f(x)\,\mathrm{d}x - \int_{X_{(1)}}^{X_{(m)}} f(x)\,\mathrm{d}x \\ &= \sum_{j=1}^{n}\int_{X_{(j-(m-1))}}^{X_{(j+(m-1))}} f(x)\,\mathrm{d}x + 2\sum_{j=1}^{n}\int_{X_{(j)}}^{X_{(j+1)}} f(x)\,\mathrm{d}x - \int_{X_{(n-(m-1))}}^{X_{(n)}} f(x)\,\mathrm{d}x - \int_{X_{(1)}}^{X_{(m)}} f(x)\,\mathrm{d}x. \end{aligned} \tag{A.1}$$

By virtue of (A.1) applied, by induction, to $\sum_{j=1}^{n}\int_{X_{(j-(m-1))}}^{X_{(j+(m-1))}} f(x)\,\mathrm{d}x$, we conclude that

$$\sum_{j=1}^{n}\int_{X_{(j-m)}}^{X_{(j+m)}} f(x)\,\mathrm{d}x = 2m\sum_{j=1}^{n}\int_{X_{(j)}}^{X_{(j+1)}} f(x)\,\mathrm{d}x - \sum_{k=1}^{m}\int_{X_{(n-(k-1))}}^{X_{(n)}} f(x)\,\mathrm{d}x - \sum_{k=1}^{m}\int_{X_{(1)}}^{X_{(k)}} f(x)\,\mathrm{d}x. \tag{A.2}$$

Since

$$\begin{aligned}&\sum_{j=1}^{n}\int_{X_{(j)}}^{X_{(j+1)}} f(x)\,\mathrm{d}x = \int_{X_{(1)}}^{X_{(n)}} f(x)\,\mathrm{d}x, \\ &\sum_{k=1}^{m}\int_{X_{(n-(k-1))}}^{X_{(n)}} f(x)\,\mathrm{d}x = \int_{X_{(n-1)}}^{X_{(n)}} f(x)\,\mathrm{d}x + \int_{X_{(n-2)}}^{X_{(n)}} f(x)\,\mathrm{d}x + \cdots + \int_{X_{(n-(m-1))}}^{X_{(n)}} f(x)\,\mathrm{d}x = (m-1)\int_{X_{(n-1)}}^{X_{(n)}} f(x)\,\mathrm{d}x \\ &\qquad + (m-2)\int_{X_{(n-2)}}^{X_{(n-1)}} f(x)\,\mathrm{d}x + \cdots + \int_{X_{(n-(m-1))}}^{X_{(n-(m-2))}} f(x)\,\mathrm{d}x = \sum_{k=1}^{m-1}(m-k)\int_{X_{(n-k)}}^{X_{(n-k+1)}} f(x)\,\mathrm{d}x\end{aligned}$$

and

$$\begin{aligned}\sum_{k=1}^{m}\int_{X_{(1)}}^{X_{(k)}} f(x)\,\mathrm{d}x &= \int_{X_{(1)}}^{X_{(2)}} f(x)\,\mathrm{d}x + \int_{X_{(1)}}^{X_{(3)}} f(x)\,\mathrm{d}x + \cdots + \int_{X_{(1)}}^{X_{(m)}} f(x)\,\mathrm{d}x \\ &= (m-1)\int_{X_{(1)}}^{X_{(2)}} f(x)\,\mathrm{d}x + (m-2)\int_{X_{(2)}}^{X_{(3)}} f(x)\,\mathrm{d}x + \cdots + \int_{X_{(m-1)}}^{X_{(m)}} f(x)\,\mathrm{d}x \\ &= \sum_{k=1}^{m-1}(m-k)\int_{X_{(k)}}^{X_{(k+1)}} f(x)\,\mathrm{d}x,\end{aligned}$$

the Eq. (A.2) completes the proof of Proposition 2.1. □

Proof of Proposition 2.2. Consider the statistic

$$Q_n = \frac{1}{n}\log \min_{1\le m<n^{1-\delta}} \prod_{i=1}^{n}\frac{2m}{n\left(X_{(i+m)}-X_{(i-m)}\right)} = -\max_{1\le m<n^{1-\delta}} q_{mn}, \quad q_{mn} = \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{n}{2m}\left(X_{(i+m)}-X_{(i-m)}\right)\right)$$

that is a component of the test statistic $\log(G_n)/n$ belonged to our interest. Following Vasicek (1976), after some reorganization, we can write

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

where F_n is the empirical distribution function and X's are from a distribution F. In this case, by Vasicek (1976), $S_j \to H(f)$ (as $n \to \infty$, $m/n \to 0$) uniformly for all $1 \le m \le n^{1-\delta}$, $0 < \delta < 1$, where $H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx = -E_f(\log f(X_1))$ is the entropy of the distribution F with a density function f and E_f denotes expectation given that $X \sim F$. The statistic U_{mn} is a non-positive variable distributed independently of F and $U_{mn} \xrightarrow{P} 0$ as $n \to \infty$, $m \to \infty$. Thus,

$$Q_n \le -q_{n^{1-\delta}n} \xrightarrow{p} E_f(\log(X_1)), \qquad Q_n \ge -\max_{1\le m<n^{1-\delta}} (2m)^{-1} \sum_{j=1}^{2m} S_j \xrightarrow{p} E_f(\log(X_1)), \quad n \to \infty.$$

This implies

$$Q_n \xrightarrow{P} E_f(\log f(X_1)), \quad n \to \infty. \tag{A.3}$$

Now, we represent the statistic (2.2.2) in the form of

$$\frac{1}{n}\log(G_n) = Q_n - \frac{1}{n}\sum_{i=1}^{n} \log\left(f_{H_0}(X_i; \boldsymbol{\theta})\right) + \frac{1}{n}\left(\sum_{i=1}^{n} \log\left(f_{H_0}(X_i; \boldsymbol{\theta})\right) - \sum_{i=1}^{n} \log\left(f_{H_0}\left(X_i; \hat{\boldsymbol{\theta}}_n\right)\right)\right). \tag{A.4}$$

Since (A.3), under H_0,

$$Q_n \xrightarrow{P} E_{f_{H_0}} \log\left(f_{H_0}(X_1)\right), \tag{A.5}$$

as $n \to \infty$.

The condition (C1) leads to

$$\frac{1}{n}\sum_{i=1}^{n} \log\left(f_{H_0}(X_i; \theta)\right) \xrightarrow{P} E_{f_{H_0}} \log\left(f_{H_0}(X_1; \boldsymbol{\theta})\right). \tag{A.6}$$

By virtue of the conditions (C2), (C4) and one-term Taylor expansion, we obtain

$$\frac{1}{n}\left(\sum_{i=1}^{n} \log f_{H_0}(X_i; \theta) - \sum_{i=1}^{n} \log f_{H_0}\left(X_i; \hat{\boldsymbol{\theta}}_n\right)\right) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} g_j\left(X_i, \boldsymbol{\eta}_i\right)\left(\theta_j - \hat{\theta}_{nj}\right) \to^{P} 0 \tag{A.7}$$

as $n \to \infty$, where $\left|\boldsymbol{\eta}_i - \boldsymbol{\theta}_0\right| \le \left|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_n\right|$. Therefore, under H_0, (A.4)–(A.7) provide

$$\frac{1}{n}\log(G_n) \xrightarrow{P} 0, \quad \text{as } n \to \infty. \tag{A.8}$$

Consider, under H_1,

$$\frac{1}{n}\log(G_n) = Q_n - \frac{1}{n}\sum_{i=1}^{n} \log f_{H_1}(X_i) + \frac{1}{n}\sum_{i=1}^{n} \log\frac{f_{H_1}(X_i)}{f_{H_0}(X_i; \mathbf{a})} + \frac{1}{n}\sum_{i=1}^{n} \log\frac{f_{H_0}(X_i; \mathbf{a})}{f_{H_0}\left(X_i; \hat{\boldsymbol{\theta}}_n\right)}. \tag{A.9}$$

Similarly to the proof of the results (A.5)–(A.7), assuming conditions (C1)–(C4), we conclude that

$$\frac{1}{n}\log(G_n) \xrightarrow{P} E_f \log\left(\frac{f_{H_1}(X_1)}{f_{H_0}(X_1; \mathbf{a})}\right) > 0 \quad \text{as } n \to \infty. \tag{A.10}$$

This and (A.8) complete the proof of Proposition 2.2. □

References

Arizono, I., Ohta, H., 1989. A test for normality based on Kullback–Leibler information. The American Statistician 43, 20–22.
Brown, K., Reid, A., White, T., Henderson, T., Hukin, S., Johnstone, C., Glen, A., 1998. Vitamin E, lipids, and lipid peroxidation products in tardive dyskinesia. Biological Psychiatry 43, 863–867.
Choi, S., Hall, W.L., Schick, A., 1996. Asymptotically uniformly most powerful tests in parametric and semiparametric models. Annals of Statistics 24, 841–861.
Dudewicz, E.J., Van Der Meulen, E.C., 1981. Entropy-based tests of uniformity. Journal of the American Statistical Association 76, 967–974.
Dong, L.B., Giles, D.E.A., 2007. An empirical likelihood ratio test for normality. Communications in Statistics. Simulation and Computation 36, 197–215.
Durbin, J., 1973. Distribution Theory for Tests Based on The Sample Distribution Function (Vol. 9). In: Conference Board of the Mathematical Sciences' Regional Conference Series in Applied Mathematics, SIAM, Philadelfia.
Ebrahami, N., Habibullah, M., Soofi, E.S., 1992. Testing exponentiality based on Kullback–Leibler information. Journal of Royal Statistical Society B 54, 739–748.
Folks, J.L., Chhikara, R.S., 1978. The inverse Gaussian distribution and its statistical application — A Review. Journal of Royal Statistical Society B 40, 263–289.

Hall, P., Welsh, A.H., 1983. A test for normality based on the empirical characteristic function. Biometrika 70, 485–489.
Lin, C.-C., Mudholkar, G.S., 1980. A simple test for normality against asymmetric alternatives. Biometrika 67, 455–461.
Lehmann, E.L., Romano, J.P., 2006. Testing Statistical Hypotheses. Springer, New York.
Mudholkar, G.S., Natarajan, R., Chaubey, Y.P., 2001. A goodness-of-fit test for the inverse Gaussian distribution using its independence characterization. Sankhyā, B 63, 362–374.
Mudholkar, G.S., Tian, L., 2002. An entropy characterization of the inverse Gaussian distribution and related goodness-of-fit test. Journal of Statistical Planning and Inference 102, 211–221.
Mudholkar, G.S., Tian, L., 2004. A test for homogeneity of ordered means of inverse Gaussian populations. Journal of Statistical Planning and Inference 118, 37–49.
Oja, H., 1983. New tests for normality. Biometrika 70, 297–299.
Owen, A.B., 2001. Empirical Likelihood. Chapman and Hall/CRC.
Park, S., Park, D., 2003. Correcting moments for goodness of fit tests based on two entropy estimates. Journal of Statistical Computation and Simulation 73, 685–694.
Pearson, E.S., 1930. A further development of tests for normality. Biometika 22, 239–249.
Prescott, P., 1976. On a test for normality based on sample entropy. Journal of Royal Statistical Society B 38, 254–256.
Royston, J.P., 1982. An extension of Shapiro and Wilk's W test for normality to large samples. Applied Statistics 31, 115–124.
Schisterman, E.F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, B., Trevisan, M., 2001. TBARS and cardiovascular disease in a population-based sample. Journal of Cardiovascular Risk 8, 1–7.
Senoglu, B., Sürücü, B., 2004. Goodness-of-fit tests based on Kullback–Leibler information. IEEE Transactions on Reliability 53, 357–361.
Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.
Shapiro, S.S., Wilk, M.B., Chen, H.J., 1968. A comparative study of various tests for normality. Journal of the American Statistical Association 63, 1343–1372.
Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. Journal of the American Statistical Association 69, 730–737.
Stigler, S.M., 1977. Do robust estimators work with real data?. The Annals of Statistics 5, 1055–1098.
Turoli, D., Testolin, G., Zanini, R., Bellù, R., 2004. Determination of oxidative status in breast and formula milk. Acta Pædiatrica 93, 1569–1574.
Tusnady, G., 1977. On asymptotically optimal tests. The American Statistician 5, 385–393.
van Es, B., 1992. Estimating functionals related to a density by a class of statistics based on spacings. Scandinavian Journal of Statistics 19, 61–72.
Vasicek, O., 1976. A test for normality based on sample entropy. Journal of Royal Statistical Society B 38, 54–59.
Verrill, S., Johnson, R.A., 1987. The asymptotic equivalence of some modified Shapiro–Wilk statistics — complete and censored sample cases. The American Statistician 15, 413–419.
Vexler, A., Wu, C., Yu, K.F., 2009. Optimal hypothesis testing: From semi to fully Bayes factors, Metrika, in press (doi:10.1007/s00184-008-0205-4).
Wang, Q., Rao, J.N.K., 2002a. Empirical likelihood-based inference under imputation for missing response data. The American Statistician 30, 896–924.
Wang, Q., Rao, J.N.K., 2002b. Empirical likelihood-based inference in linear models with missing data. Scandinavian Journal of Statistics 29, 563–576.
Yarman, S., Özden, T.A., Gökkuş, C., 2003. The evaluation of lipid peroxidation and acute effect of octreotide on lipid peroxidation in patients with active acromegaly. Clinica Chimica Acta 336, 45–48.
Yu, J., Vexler, A., Tian, L., 2009. Analyzing incomplete data subject to a threshold using empirical likelihood methods: An application to a pneumonia risk study in an ICU setting. Biometrics, in press (doi:10.1111/j.1541-0420.2009.01128.x).
Zhang, J., 2002. Powerful goodness-of-fit tests based on the likelihood ratio. Journal of Royal Statistical Society B 64, 281–294.

Please cite this article in press as: Vexler, A., Gurevich, G., Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.09.025