Computational Statistics & Data Analysis 1 (1983) 157–165
North-Holland

On entropy-based goodness-of-fit tests

D.V. GOKHALE *

Department of Statistics, University of California, Riverside, CA 92521, USA

Received February 1983

Abstract: A general form of a goodness-of-fit test statistic for families of maximum entropy distributions is obtained. Particular cases of such families are normal, exponential, double-exponential, gamma and beta distributions. The proposed test is shown to be consistent against an appropriate class of alternatives. Simulation and Monte Carlo results are given for exponential and double exponential distributions. The proposed test compares favorably with other goodness-of-fit tests. This paper generalizes the work of Vasicek [7] regarding entropy-based test of normality and further work by Dudewicz and van der Meulen [1].

Keywords: Entropy-based goodness-of-fit tests, Exponential distributions, Families of maximum entropy distributions.

1. Introduction

A test of the composite hypothesis of normality based on sample estimate of entropy is introduced by Vasicek [7]. The test is shown to be consistent against all alternatives without a singular continuous part. Critical values of the test statistic and its power against various alternatives, obtained by simulation, are given in his paper. It is shown that the test has good power against the alternatives considered and compares well with other tests of normality.

Dudewicz and van der Meulen [1], [2] prove asymptotic (large sample) normality of Vasicek's statistic, suitably standardized, under the hypothesis of uniformity and also under a special class of (fixed) alternative densities. They show by simulation and Monte Carlo that the entropy-based test of uniformity has good power properties when compared with several other tests.

The present paper begins with a test-statistic for testing a simple hypothesis that the sample comes from a given continuous distribution F_0. Its asymptotic normality under the hypothesis follows directly from the results of [1]. It is shown here that under a fixed continuous alternative F_1 satisfying some regularity

* A part of this work was done when the author was visiting Department of Statistics, The Ohio State University, Columbus, during Spring Quarter 1980.

0167-9473/83/$3.00

conditions the test statistic converges in probability to the negative of the Kullback–Liebler information $I(f_1 : f_0)$ where f_1 and f_0 are the respective densities. The concepts of 'entropy-distinguishability' and 'entropy-uniqueness', based on *entropy power variance ratio,* due to Dudewicz and van der Meulen [1], are then generalized by using maximum entropy characterizations. Many well known families of distributions such as normal, exponential, double exponential, gamma and beta are characterized by maximization of entropy subject to constraints (Gokhale [4]). Thus they enjoy the property of entropy-uniqueness in certain families of densities. A general form of a test statistic for the composite hypothesis of a maximum entropy distribution is derived. Tests for exponential, double exponential, gamma and beta distributions are obtained as special cases. Quantiles of the test statistic are obtained by simulation for the important special case of exponential distributions and for double exponential distributions. Power comparisons in these two cases are studied in relation to available test-statistics.

2. An entropy-based goodness-of-fit test

Let $X_1, X_2, \ldots, X_n$ be a random sample from a continuous distribution function $F(x)$ with density $f(x)$. Let $Y_1, Y_2, \ldots, Y_n$ denote the order statistics of the sample. To test the hypothesis $F = F_0$ we construct

$$U_i = F_0(Y_i), \quad i = 1, 2, \ldots, n, \tag{2.1}$$

and

$$H_{mn}(f_0) = n^{-1} \sum_{i=1}^{n} \log\left\{ \frac{n}{2m} (U_{i+m} - U_{i-m}) \right\} \tag{2.2}$$

where m is a positive integer less than $\frac{1}{2}n$ and

$$U_j = U_1 \quad \text{if } j < 1, \qquad U_j = U_n \quad \text{if } j > n.$$

Since, under the hypothesis, the U_i, $i = 1, 2, \ldots, n$, are order statistics from the uniform distribution on (0, 1), it follows from [7], Theorem 1, and [1], Theorem 4, that if $m \to \infty$, $n \to \infty$, $m = O(n^{1/3-\delta})$ then

$$H_{mn}(f_0) \xrightarrow{P} 0 \tag{2.3}$$

and

$$\sqrt{(6mn)}\,[H_{mn}(f_0) + \log 2m + \gamma - R(1, 2m-1)] \tag{2.4}$$

is asymptotically standard normal. Here γ is Euler's constant and

$$R(1, m-1) = \begin{cases} \sum_{j=1}^{m-1} \dfrac{1}{m-j} & \text{if } m \geqslant 2, \\ 0 & \text{if } m = 1. \end{cases}$$

It may be noted that asymptotic normality holds under the (weaker) assumption that $m = O(n^{1/3})$ (Theorem 3, [1]) but the standardizing constants are very complicated to write down. Tables of significance points obtained by simulation are given by Dudewicz and van der Meulen [1]. They recommend normal approximation for $n \geqslant 100$ and relatively small m.

Now suppose that the sample $X_1, X_2, \ldots, X_n$ actually comes from the distribution F_1 with density f_1. We will assume that (i) both F_0 and F_1 are mutually absolutely continuous and absolutely continuous w.r.t. the Lebesgue measure, so that

$$\frac{\mathrm{d}F_1}{\mathrm{d}F_0}(x) = \frac{\mathrm{d}F_1(x)}{\mathrm{d}x}\,\frac{\mathrm{d}x}{\mathrm{d}F_0(x)} = \frac{f_1(x)}{f_0(x)}, \tag{2.5}$$

and (ii)

$$\left| \int_{-\infty}^{\infty} f_1(x) \log \frac{f_1(x)}{f_0(x)} \mathrm{d}x \right| < \infty. \tag{2.6}$$

Notice that the integral in (2.6) is the Kullback–Liebler information $I(f_1 : f_0)$ in f_1 with respect to f_0.

Theorem 2.1. *Under assumptions* (2.5) *and* (2.6)

$$H_{mn}(f_0) \xrightarrow{P} -I(f_1 : f_0) \quad \textit{as } m \to \infty,\ n \to \infty, \frac{m}{n} \to 0.$$

Proof. Following [7], $H_{mn}(f_0)$ defined in (2.2) can be written as

$$\begin{aligned} H_{mn}(f_0) &= H_{mn}(f_1) - \frac{1}{n} \sum \log \left[\frac{F_1(Y_{i+m}) - F_1(Y_{i-m})}{F_0(Y_{i+m}) - F_0(Y_{i-m})} \right] \\ &= H_{mn}(f_1) + \frac{1}{2m} \sum_{j=1}^{2m} T_j \end{aligned} \tag{2.7}$$

where

$$T_j = -\sum_{i=1}^{n} \log \left[\frac{F_1(Y_{i+m}) - F_1(Y_{i-m})}{F_0(Y_{i+m}) - F_0(Y_{i-m})} \right] \{ F_n(Y_{i+m}) - F_n(Y_{i-m}) \},$$
$$i \equiv j \bmod (2m).$$

Then, as $m, n \to \infty$, $m/n \to 0$, $H_{mn}(f_1) \xrightarrow{P} 0$ and the second term on the right-hand side of (2.7) converges to $-I(f_1 : f_0)$, by an argument similar to that used in proving Theorem 1 of [7].

3. Tests of maximum entropy distributions

Let $\mathscr{F}$ be the class of densities f with real line as support and with a given mean μ and a given variance σ^2. It is well known that among all $f \in \mathscr{F}$ the one that maximizes the entropy $H(f) = -E(\log f(X))$ is the normal density. The maximum value max $H(f)$ in this case is $\log(\sigma\sqrt{2\pi e})$. For any other density $f \in \mathscr{F}$ the ratio

$$\exp\{H(f)\}/\sigma_f\sqrt{2\pi e} \tag{3.1}$$

can be regarded as a measure of the closeness of f to the normal density having the same variance. The ratio (3.1), apart from a constant, is termed as the *entropy power variance ratio* (EPVR) in [1]. Note that (3.1) is always less than or equal to unity and equals unity if and only if f is normal. Based on this characterization we have the following definitions from [1].

Definition 3.1. Two densities belonging to $\mathscr{F}$ are *EPVR-distinguishable* if their respective ratios (3.1) are unequal.

Definition 3.2. A density $f_0 \in \mathscr{F}$ is *EPVR-unique* if for every $f_1 \in \mathscr{F}$,

$$\frac{\exp H(f_1)}{\sigma_{f_1}\sqrt{2\pi e}} = \frac{\exp H(f_0)}{\sigma_{f_0}\sqrt{2\pi e}}$$

implies $f_1 = f_0$ a.e. Lebesgue.

A consistent test of the composite hypothesis of normality is then given by

$$K_{mn} = \exp(H_{mn})/s_n\sqrt{2\pi e} \tag{3.2}$$

where H_{mn} is the sample estimate of $H(f)$ given by (Vasicek [7])

$$H_{mn} = \frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{n}{2m}(Y_{i+m} - Y_{i-m})\right\} \tag{3.3}$$

with $Y_j = Y_1$ if $j < 1$, $Y_j = Y_n$ if $j > n$, and s_n^2 is the sample variance.

To generalize these ideas consider a family $P(\Theta) = \{f_0(x, \theta)\colon \theta \in \Theta,\ \theta = (\theta_1, \theta_2, \ldots, \theta_m)\}$ of densities of the form

$$f_0(x, \theta) = C(\theta)\exp\left[u_1(\theta)T_1(x) + \sum_{j=2}^{m} u_j(\theta)T_j\big(x; \eta_1(\theta), \ldots, n_{j-1}(\theta)\big)\right] \tag{3.4}$$

where

$$E_\theta T_1(X) = \eta_1(\theta),$$
$$E_\theta T_j\big(X; \eta_1(\theta), \ldots, \eta_{j-1}(\theta)\big) = \eta_j(\theta), \quad j = 2, \ldots, m.$$

The norming constant $C(\theta)$ and functions $u_j(\theta)$ are assumed to be continuous and depend on θ through functions $\eta_j(\theta)$, $j = 1, 2, \ldots, m$. Let $\eta(\theta) = (\eta_1(\theta), \ldots, \eta_m(\theta))$ and $\eta(\Theta) = \{\eta(\theta)\colon \theta \in \Theta\}$. For a fixed $\eta = \eta(\theta) \in \eta(\Theta)$, let $\mathscr{F}^*$ denote the class of densities $f(x)$ such that

$$E_f T_1(X) = \eta_1,$$
$$E_f T_j(X; \eta_1, \eta_2, \ldots, \eta_{j-1}) = \eta_j, \quad j = 2, 3, \ldots, m,$$

and every $f \in \mathscr{F}^*$ has the same support as $f_0(x, \xi) = \eta$.

Now, it is shown in Gokhale [4] that $f_0(x, \theta)$ maximizes the entropy among all $f \in \mathscr{F}^*$ and

$$\max_{\mathscr{F}^*} H(f) = H(f_0) = -\log C(\xi) - \sum_{j=1}^{m} u_j(\xi)\eta_j \tag{3.5}$$

where $\eta(\xi) = \eta$.

Defining as *entropy power fraction* (EPF) the ratio

$$\mathrm{EPF}(f) = \frac{\exp\{H(f)\}}{[C(\xi)]^{-1}\exp\{-\sum_{j=1}^{m} u_j(\xi)\eta_j\}} \tag{3.6}$$

for any density f for which θ is finite and $\eta \in \eta(\Theta)$ it can be seen that Definitions 3.1 and 3.2 generalize as follows:

Definition 3.3. Two densities are *EPF-distinguishable* if their respective ratios (3.6) are unequal.

Definition 3.4. A density f_0 in a family $\mathscr{F}$ is *EPF-unique* if for every $f_1 \in \mathscr{F}$

$$\mathrm{EPF}(f_1) = \mathrm{EPF}(f_0)$$

implies $f_1 = f_0$ a.e. Lebesgue.

Note that for the family $\mathscr{F}^*$ defined above $f_0(x, \xi)$ with $\eta(\xi) = \eta$ is EPF-unique.

Let $\hat{\eta}_n$ be a consistent estimator of η based on the sample $X_1, \ldots, X_n$, and let $\hat{\xi}_n$ be such that $\eta(\hat{\xi}_n) = \hat{\eta}_n$. Then as a generalization of the consistent test statistic K_{mn} of (3.2) for testing normality, one can define

$$K_{mn}^* = \frac{\exp\{H_{mn}\}}{\left[C(\hat{\xi}_n)\right]^{-1}\exp\left\{-\sum_{j=1}^{m} u_j(\hat{\xi}_n)\hat{\eta}_n\right\}} \tag{3.7}$$

where H_{mn} is the sample estimate (3.3) of $H(f)$ as before. The K_{mn}^* test is consistent for testing the form of the density to be (3.4) against any alternative density f for which η is finite and $\eta \in \eta(\Theta)$. In the examples given below a number of special cases of families of densities are considered. Expressions for η_j, $j = 1, \ldots, m$, and the quantity $H(f_0) = \max_{\mathscr{F}^*} H(f)$ as given by (3.5) in terms of η_j, are derived wherever possible. The specific form of K_{mn}^* is also provided.

Example 3.1. Exponential densities:

$$f_0(x, \theta) = \frac{1}{\theta}\exp(-x/\theta), \quad 0 < x < \infty, \theta > 0, \tag{3.8}$$

$$T(X) = X, \quad \eta = \theta, \quad H(f_0) = \log \mathrm{e}\theta,$$

$$K_{mn}^* = \exp\{H_{mn}\}/\mathrm{e}\bar{X}, \tag{3.9}$$

$\bar{X}$ being the sample mean.

Example 3.2. Laplace densities:

$$f_0(x, \theta_1, \theta_2) = \frac{1}{2\theta_2} \exp\{-\theta_2^{-1}|x - \theta_1|\}, \quad -\infty < x < \infty, \theta_1 \in \mathbb{R}, \theta_2 > 0, \tag{3.10}$$

$$T_1(X) = X, \quad \eta_1 = \theta_1, \quad T_2(X, \theta_1) = |X - \theta_1|, \quad \eta_2 = \theta_2,$$
$$H(f_0) = \log 2 + \log \theta_2 + 1 = \log 2\mathrm{e}\theta_2,$$
$$K^*_{mn} = \exp\{H_{mn}\}/2\mathrm{e}S_m, \tag{3.11}$$

where S_m is the mean deviation of the sample.

Example 3.3. Gamma densities:

$$f_0(x, \theta_1, \theta_2) = \frac{\theta_1^{\theta_2}}{\Gamma(\theta_2)} \mathrm{e}^{-\theta_1 x} x^{\theta_2 - 1}, \quad 0 < x < \infty, \theta_1 > 0, \theta_2 > 0, \tag{3.12}$$

$$T_1(X) = X, \quad \eta_1 = \theta_2/\theta_1, \quad T_2(X_j, \eta_1) = \log X,$$
$$\eta_2 = \Psi(\theta_2) - \log \theta_1, \quad H(f_0) = \log \Gamma(\theta_2) + \theta_2(1 - \Psi(\theta_2)) + \Psi(\theta_2) - \log \theta_1,$$

where $\Psi(\cdot)$ is the Digamma function. In this case

$$K^*_{mn} = \exp\{H_{mn}\}/\left[\Gamma(\hat{\theta}_{2n}) \exp\{\hat{\theta}_{2n}(1 - \Psi(\hat{\theta}_{2n})) + \Psi(\hat{\theta}_{2n})\}/\hat{\theta}_{1n}\right], \tag{3.13}$$

where $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are consistent estimators of θ_1 and θ_2. They cannot be explicitly obtained in terms of η_1 and η_2, but can be obtained iteratively (see Johnson and Kotz [5], p. 187).

Example 3.4. Beta densities:

$$f_0(x, \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1 - 1}(1 - x)^{\theta_2 - 1}, \quad 0 < x < 1, \theta_1 > 0, \theta_2 > 0, \tag{3.14}$$

$$T_1(X) = \log X, \quad \eta_1 = \Psi(\theta_1) - \Psi(\theta_1 + \theta_2), \quad T_2(X) = \log(1 - X),$$
$$\eta_2 = \Psi(\theta_2) - \Psi(\theta_1 + \theta_2),$$
$$H(f_0) = \log \Gamma(\theta_1) + \log \Gamma(\theta_2) - \log \Gamma(\theta_1 + \theta_2)$$
$$\qquad - (\theta_1 - 1)\Psi(\theta_1) - (\theta_2 - 1)\Psi(\theta_2) + (\theta_1 + \theta_2 - 2)\Psi(\theta_1 + \theta_2),$$
$$K^*_{mn} = \exp\{H_{mn}\}/\exp \hat{H}_n(f_0), \tag{3.15}$$

where $\hat{H}_n(f_0)$ is obtained by using consistent estimators $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ for θ_1 and θ_2 in $H(f_0)$. (See Johnson and Kotz [6], p. 44.) Again it is not possible to express θ_1 and θ_2 explicitly in terms of η_1 and η_2.

Table 1
Quantiles of K^*_{mn} for exponential distributions

n	m	Left-tail probability			
		0.05	0.10	0.90	0.95
10	1	0.41	0.46	0.78	0.81
10	2	0.49	0.55	0.81	0.83
10	3	0.50	0.56	0.80	0.82
10	4	0.50	0.55	0.80	0.83
20	1	0.53	0.56	0.79	0.82
20	2	0.64	0.68	0.86	0.87
20	3	0.67	0.70	0.87	0.88
20	4	0.68	0.71	0.88	0.89
30	1	0.59	0.62	0.79	0.81
30	2	0.70	0.73	0.87	0.89
30	3	0.74	0.76	0.89	0.90
30	4	0.75	0.77	0.90	0.92

4. Examples and power comparisons

In this section the foregoing methodology is applied to goodness-of-fit tests for (i) exponential and (ii) Laplace distributions. The K^*_{mn} statistic is given in (3.9) and (3.11) respectively. The exact distribution of K^*_{mn} statistics (3.7) is very difficult to obtain in any special case. Hence computer simulation and Monte Carlo were used to obtain critical values of K^*_{mn} and to make power comparisons for the two examples. Five thousand samples were taken for each null hypothesis for various values of (m, n), to determine selected quantiles of K^*_{mn}. Power of the test was determined from one thousand samples each for different alternatives and a few values of (m, n). The results are summarized in Tables 1–4. In case of the exponential distribution alternative distributions were taken from Gail and Gastwirth [3].

Table 2
Power of K^*_{mn} test for exponentiality, $\alpha = 0.05$

Alternative [a]	m, n		
	(1, 20)	(1, 30)	(4, 30)
Weibull ($k = 0.8$)	0.061	0.078	0.068
Weibull ($k = 1.5$)	0.236	0.268	0.617
Uniform (0, 2)	0.613	0.801	0.969
Pareto ($k = 3$)	1.000	1.000	1.000
Shifted Pareto ($k = 3$)	0.167	0.285	0.249
Shifted exponential	0.338	0.469	0.850
Gamma (2)	0.243	0.316	0.619

[a] See Gail and Gastwirth [3] for specific forms of the alternatives.

Table 3
Quantiles of K^*_{mn} for Laplace distributions

n	m	Left-tail probability			
		0.05	0.10	0.90	0.95
10	1	0.40	0.45	0.75	0.78
10	2	0.47	0.52	0.77	0.79
10	3	0.48	0.51	0.75	0.78
10	4	0.47	0.50	0.71	0.74
20	1	0.52	0.56	0.78	0.80
20	2	0.63	0.66	0.84	0.86
20	3	0.65	0.68	0.86	0.87
20	4	0.64	0.68	0.87	0.89
30	1	0.58	0.61	0.79	0.81
30	2	0.69	0.72	0.86	0.88
30	3	0.72	0.75	0.89	0.90
30	4	0.73	0.76	0.90	0.92

Table 4
Power of K^*_{mn} test for Laplace distributions, $\alpha = 0.05$

Alternative	(m, n)		
	(1, 20)	(1, 30)	(4, 30)
Normal (0, 1)	0.094	0.131	0.229
Cauchy (0, 1)	0.470	0.659	0.562
Logistic $f(x) = (1 + e^{-x})^{-1}$ $-\infty < x < \infty$	0.066	0.063	0.104
Triangular $(-1, 1)$	0.120	0.197	0.382
Uniform $(-1, 1)$	0.502	0.762	0.936
Double Weibull $f(x) = c\|x\|^{c-1}e^{-\|x\|^c}$ $c = 0.5$	0.761	0.900	0.710

5. Conclusions

The K^*_{mn} test for exponentiality compares reasonably well with its competitors against alternatives mentioned in Gail and Gastwirth [3]. A similar conclusion holds for the K^*_{mn} test for Laplace distributions. The power of the K^*_{mn} test is poor against alternatives with the same support and similar shape as the null distributions. There is a markedly substantial improvement in power when (i) the alternatives have the same support but heavier tails, or (ii) the alternatives have a singular part with respect to the null distributions. For case (i), power is larger for smaller values of m, while for (ii) it is larger when m is close to $[n/2]$.

Frequency histograms were constructed for the 5000 values of K^*_{mn} under the null hypothesis and also for the 1000 values under various alternative hypotheses, for each combination of m and n reported here. Even for $n = 10$, the histograms are unimodal and almost symmetric. They are not reproduced here for space considerations. They are available with the author.

References

[1] E.J. Dudewicz and E.C. van der Meulen, Entropy based statistical inference, I: Testing hypothesis on continuous probability densities, with special reference to uniformity, *Mededelingen uit het Wiskundig Instituut,* Katholieke Universiteit, Leuven, Belgium, No. 120 (1979).

[2] E.J. Dudewicz and E.C. van der Meulen, Entropy-based tests of uniformity, *Journal of the American Statistical Association* **76** (376) (1981) 967–974.

[3] N.H. Gail and J.L. Gastwirth. A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve, *Journal of the American Statistical Association* **73** (1978) 787–793.

[4] D.V. Gokhale, Maximum entropy characterizations of some distributions, in: G.P. Patil (Ed.), *Statistical Distributions in Scientific Work,* Vol. III (Reidel, Dordrecht–Boston, 1975) pp. 299–304.

[5] N.L. Johnson and S. Kotz, *Continuous Univariate Distributions-1, Distributions in Statistics* (John Wiley and Sons, New York, 1970).

[6] N.L. Johnson and S. Kotz, *Continuous Univariate Distributions-2, Distributions in Statistics* (John Wiley and Sons, New York, 1970).

[7] O. Vasicek, A test for normality based on sample entropy, *Journal of the Royal Statistical Society Series B* **38** (1976) 54–59.