J. Appl. Math. & Informatics Vol. 26(2008), No. 1 - 2, pp. 177 - 189
Website: http://www.kcam.biz

GOODNESS OF FIT TESTS BASED ON DIVERGENCE MEASURES

EYNOLLAH PASHA, MOHSEN KOKABI AND GHOLAM REZA MOHTASHAMI*

ABSTRACT. In this paper, we have considered an investigation on goodness of fit tests based on divergence measures. In the case of categorical data, under certain regularity conditions, we obtained asymptotic distribution of these tests. Also, we have proposed a modified test that improves the rate of convergence. In continuous case, we used our modified entropy estimator [10], for Kullback-Leibler information estimation. A comparative study based on simulation results is discussed also.

AMS Mathematics Subject Classification : 62B10, 94A17.
Key words and phrases : Goodness of fit test, divergence measures, entropy, categorical data, density estimation.

1. Introduction

Suppose Y is a random variable with a density or probability mass function $f(y,\theta)$, that support of which is partitioned in "k" paired disjoint sets $A_1, \cdots, A_k$ in which $\theta = (\theta_1, \cdots, \theta_m)'$ is a vector of unknown parameters that $\theta \in \Theta \subset R^m$ and $m \leq k-1$. We define $\pi_j = P(Y \in A_j|\theta)$ for $j = 1, 2, \cdots, k$.

In n independent observations of random variable Y, suppose X_j is the number of observations belonging to the set A_j, in this case $(X_1, \cdots, X_k)$ have multinomial distribution with parameters $(\pi_1, \cdots, \pi_k)$. The goal is to testing: $H_0 : \pi_j = \pi_{j0}$ for $j = 1, 2, \cdots, k$ against any alternative H_1 where π_{j0}s are some preassigned probability with $\sum_{j=1}^{k} \pi_{j0} = 1$.

If p_j is the proportion of observations in A_j, then a natural criteria for the test, is comparison of two vectors $p = (p_1, \cdots, p_k)'$ and $\hat{\pi}_0 = (\hat{\pi}_{10}, \cdots, \hat{\pi}_{k0})'$ based on a divergence measure in which $\hat{\pi}_{j0} = P_0(Y \in A_j|\hat{\theta})$ and $\hat{\theta}$ is a suitable estimator based on X_js. From [1] and [3], for any continuous and convex function

Received January 10, 2005.

$h : [0, \infty) \to (-\infty, +\infty)$ that is zero only at point 1, a general class of divergence measures, under discrimination function "h" for two probability vectors p and q has been defined as follows:

$$I_h(p,q) = \sum_i q_i h\left(\frac{p_i}{q_i}\right), \quad I_{h^*}(p,q) = \sum_i p_i h^*\left(\frac{q_i}{p_i}\right)$$

in which $h^*(x) = xh(\frac{1}{x})$. It is well known that $I_h(p,q) \geq 0$ and $I_h(p,q) = 0$ if and only if $p = q$. Thus, the class of tests based on divergence measures I_h can be considered as:

$$I_h(p, \pi_0) = \sum_{j=1}^{k} \hat{\pi}_{j0} h(\frac{p_j}{\hat{\pi}_{j0}}).$$

Also, the Jensen difference is given in [13] as:

$$J_H(x,y) = H\left(\frac{x+y}{2}\right) - \frac{1}{2}H(x) - \frac{1}{2}H(y)$$

in which H is an entropy function, $H(p_1, \cdots, p_k) = \sum_{i=1}^{k} \Phi(p_i)$, such that $\Phi : (0, \infty) \to (-\infty, +\infty)$ is a concave and continuous function where $\Phi(1) = 0$. Therefore, the other class of tests are as follows:

$$J_\Phi(p, \hat{\pi}_0) = \sum_{i=1}^{k} \left(\Phi\left(\frac{p_i + \hat{\pi}_{i0}}{2}\right) - \frac{1}{2}\Phi(p_i) - \frac{1}{2}\Phi(\hat{\pi}_{i0})\right).$$

In the case of uncategorical data, suppose random variables $Y_1, \cdots, Y_n$ are iid with a continuous distribution $F(y, \theta)$. For testing $H_0 : Y \sim F_0(y, \theta)$, let $\{b_1 < \cdots < b_{k+1}\}$ be a partition of the sample space, so that $F_0(b_{k+1}, \theta^*) = 1$ and $F_0(b_0, \theta^*) = 0$ in which θ^* is an estimator of θ. Thus, with setting X_i as the number of observations belonging to the category $[b_i, b_{i+1})$ and $\hat{\pi}_{i0} = F_0(b_{i+1}, \theta^*) - F_0(b_i, \theta^*)$, we define:

$$\tilde{f}_n(y) = \frac{1}{t}\left(\tilde{F}_n(y + \frac{t}{2}) - \tilde{F}_n(y - \frac{t}{2})\right),$$

$$\tilde{f}_0(y) = \frac{1}{t}\left(\tilde{F}_0(y + \frac{t}{2}) - \tilde{F}_0(y - \frac{t}{2})\right)$$

in which

$$\tilde{F}_n(y) = \frac{1}{n}\sum_{j=1}^{i} X_j, \qquad \tilde{F}_0(y) = \sum_{j=1}^{i} \hat{\pi}_{j0}$$

for every $y \in [b_i, b_{i+1})$ and $0 < t < 2 \min_{1 \leq i \leq k}(b_{i+1} - b_i)$. Now, we can set $p_i = t\tilde{f}_n(b_i)$ and $\hat{\pi}_{i0} = t\tilde{f}_0(b_i)$.

On the other hand, divergence measure in continues case, is defined as [3]:

$$I_{h^*}(\tilde{f}_n, \tilde{f}_0) = \int h^* \left(\frac{\tilde{f}_0(y)}{\tilde{f}_n(y)} \right) d\tilde{F}_n(y).$$

Hence, with replacement of the usual empirical distribution function and a density estimator such as f_n, we have:

$$I_{h^*}(f_n, f_0) = \frac{1}{n} \sum_{i=1}^{n} h^* \frac{f_0(y_i)}{f_n(y_i)}.$$

Therefore, we can use $I_{h^*}(f_n, f_0)$ as a test statistic with the rejection region $\{I_{h^*}(f_n, f_0) > I^*_\alpha\}$ in which $P(I_{h^*}(f_n, f_0) > I^*_\alpha) = \alpha$.

In this paper the class of goodness of fit tests is considered on the basis of divergence measures (DM-class), that is a kind of measure between empirical density function and hypothesized density function. In the sections 2 and 3, with simulation methods, some selected tests in this class were compared for the categorical and uncategorical data, respectively.

2. Goodness of fit tests and simulation results for categorical data

From [12, p. 360- 363 and p. 391], we have the following theorem:

Theorem. *Under the following conditions, the statistic* $n \sum_{i=1}^{k} \frac{(p_i - \hat{\pi}_{i0})^2}{\hat{\pi}_{i0}}$ *has asymptotic distribution* $\chi^2_{(k-m-1)}$, *in which* $\hat{\pi}_{i0} = P_0(Y \in A_i | \hat{\theta})$ *and* $\hat{\theta}$ *is a maximum likelihood estimator based on* X_i*s. (suppose k is constant).*

a) *Suppose* θ_0 *is the true value of* θ, *as interior point* Θ, *for given a* $\delta > 0$, *it is possible to find an* ε *such that:*

$$\inf_{|\theta-\theta_0|>\delta} \sum_i \pi_{i0}(\theta_0) \log \left(\frac{\pi_{i0}(\theta_0)}{\pi_{i0}(\theta)} \right) \geq \varepsilon.$$

b) *The functions* π_{i0} *admit continuous first order partial derivatives with respect to* θ_j *for* $i = 1, \cdots, k$ *and* $j = 1, \cdots, m$.

c) *Matrix* $\left(\pi_{i0}^{-\frac{1}{2}} \frac{d\pi_{i0}}{d\theta_j} \right)$ *is of rank m if* $\theta = \theta_0$.

d) *Information matrix* $\left(\sum_j \frac{1}{\pi_{j0}} \frac{d\pi_{j0}}{d\theta_r} \frac{d\pi_{j0}}{d\theta_s} \right)$ *is non-singular if* $\theta = \theta_0$.

Now, with using this theorem, we prove the following corollary:

Corollary. *Under the conditions of above theorem, suppose* $h(.)$ *is a continuous function and* $h^{(i)}$ *is a derivative of order i, such that:*

a) $h^{(1)}$ *and* $h^{(2)}$ *exists and are continuous in a closed neighborhood of 1 and* $h^{(2)}(1) = 0$.

b) $h^{(3)}$ *exist and is uniformly bounded in an open neighborhood of 1.*

Then, the statistic $\frac{2n}{h^{(2)}(1)} I_h(p, \hat{\pi}_0)$ *has asymptotic distribution* $\chi^2_{(k-m-1)}$.

Proof. We show that $n \left| \frac{2}{h^{(2)}(1)} I_h(p, \hat{\pi}_0) - \sum_{i=1}^{k} \frac{(p_i - \hat{\pi}_{i0})^2}{\hat{\pi}_{i0}} \right| \to 0$ in distribution as $n \to \infty$. With Taylor's expansion for function "h", we have:

$$h\Big(\frac{p_i}{\hat{\pi}_{i0}}\Big) = h(1) + h^{(1)}(1)\Big(\frac{p_i}{\hat{\pi}_{i0}} - 1\Big) + \frac{h^{(2)}(1)}{2}\Big(\frac{p_i}{\hat{\pi}_{i0}} - 1\Big)^2 + \frac{h^{(3)}(c_i)}{6}\Big(\frac{p_i}{\hat{\pi}_{i0}} - 1\Big)^3$$

in which $|c_i - 1| < \left|\frac{p_i}{\hat{\pi}_{i0}} - 1\right|$, for $i = 1, 2, \cdots, k$. Therefore:

$$I_h(p, \hat{\pi}_0) = \sum_{i=1}^{k} \frac{(p_i - \hat{\pi}_{i0})^2}{\hat{\pi}_{i0}} \left\{ \frac{h^{(2)}(1)}{2} + \frac{h^{(3)}(c_i)}{6}\Big(\frac{p_i}{\hat{\pi}_{i0}} - 1\Big) \right\}$$

Considering $\left|\frac{p_i}{\hat{\pi}_{i0}} - 1\right| \to 0$ in probability as $n \to \infty$, then the value of c_i to be in the neighborhood of 1, and the corollary is proved. □

Hence, the asymptotic rejection region of this test is $\left\{ \frac{2n}{h^{(2)}(1)} I_h(p, \hat{\pi}_0) > \chi^2_\alpha \right\}$ in which χ^2_α is the $100\alpha\%$ upper point of Chi-square distribution with $k - m - 1$ degrees of freedom. This approximation is suitable provided that in the majority of $n\hat{\pi}_{j0}$ are not too small. The sensitivity to this considerably depends on function "h". Thus, the best discrimination function should be selected to have high convergence speed and less sensitivity to small $n\hat{\pi}_{j0}$.

When a model under null hypothesis is not dependent on unknown parameters, a modification statistic is presented for speed up convergence rate. Suppose "h" has continuous derivatives up to fourth order; after a fairly long algebraic work as mentioned in the appendix, the following modified statistic is proposed:

$$\frac{\frac{2n}{h^{(2)}(1)} I_h(p, \pi_0) - \mu_k}{\sigma_k}$$

where

$$\sigma_k = 1 + \frac{1}{2n(k-1)} \left(B_2 - B_1^2 - 2(k-1)B_1\right),$$

$$\mu_k = (k-1)(1 - \sigma_k) + B_1$$

and

$$\begin{aligned} B_1 &= \left(\frac{h^{(3)}(1)}{3h^{(2)}(1)}\right)(2 - 3k + t) + \left(\frac{h^{(4)}(1)}{4h^{(2)}(1)}\right)(1 - 2k + t), \\ B_2 &= (2 - 2k - k^2 + t) + \left(\frac{2h^{(3)}(1)}{3h^{(2)}(1)}\right)\left(t(k+8) - 6k^2 - 13k + 10\right) \\ &+ \frac{1}{3}\left(\frac{h^{(3)}(1)}{h^{(2)}(1)}\right)^2 (4 - 6k - 3k^2 + 5t) \end{aligned}$$

$$+ \left(\frac{h^{(4)}(1)}{2h^{(2)}(1)}\right)\left(t(3+k) - 5k - 2k^2 + 3\right),$$

in which $t = \sum_{i=1}^{k} \frac{1}{\pi_{i0}}$.

Now, the efficiency and the power of some tests related to DM-class have been studied for categorical data on using simulation with selecting some special forms for function "h" as follows:

h	*Divergence measure*	*Test statistic*
$x\log(x)$	$\sum_{i=1}^{k} p_i \log(\frac{p_i}{q_i})$	$ST1 = 2n\sum_{i=1}^{k} p_i \log(\frac{p_i}{\hat{\pi}_{i0}})$
$(x-1)^2$	$\sum_{i=1}^{k} \frac{(p_i - q_i)^2}{q_i}$	$ST2 = n\sum_{i=1}^{k} \frac{(p_i - \hat{\pi}_{i0})^2}{\hat{\pi}_{i0}}$
$x(x^{\frac{2}{3}} - 1)$	$\sum_{i=1}^{k} p_i \left((\frac{p_i}{q_i})^{\frac{2}{3}} - 1\right)$	$ST3 = \frac{9}{5}n\sum_{i=1}^{k} p_i \left(\left(\frac{p_i}{\hat{\pi}_{i0}}\right)^{\frac{2}{3}} - 1\right)$
$(\sqrt{x} - 1)^2$	$\sum_{i=1}^{k}(\sqrt{p_i} - \sqrt{q_i})^2$	$ST4 = 4n\sum_{i=1}^{k}(\sqrt{p_i} - \sqrt{\hat{\pi}_{i0}})^2$
$\frac{(x-1)^2}{x+1}$	$\sum_{i=1}^{k} \frac{(p_i - q_i)^2}{p_i + q_i}$	$ST5 = 2n\sum_{i=1}^{k} \frac{(p_i - \hat{\pi}_{i0})^2}{p_i + \hat{\pi}_{i0}}$

ST1 is likelihood ratio test based on the Kullback-Leibler information measure. ST2 is famous Pearson's test [11]. ST3 in [14] and ST4 on the basis of square of Hellinger's distance in [7] has been proposed.

The next statistic is the selection of Shannon entropy in the class of $J_\Phi(p, \pi_0)$ tests with $\Phi(x) = -x\log(x)$. Therefore, the test statistic is as follows:

$$ST6 = -4n\sum_{i=1}^{k}\left\{(\frac{p_i + \hat{\pi}_{i0}}{2})\log(\frac{p_i + \hat{\pi}_{i0}}{2}) - \frac{p_i \log p_i + \hat{\pi}_{i0}\log\hat{\pi}_{i0}}{2}\right\}.$$

In the table 1, the amount of asymptotic significant level

$$\beta(k) = p(ST > \chi^2_{1-\alpha}(k-1)|H_0)$$

for homogeneity test of a population with 6 category and different sample sizes, and in the table 2, for a sample size 30 and different categories "k", has been simulated on the basis of 10,000 samples and at the significant level of 5%. According to this tables the asymptotic significant level of tests ST1, ST4 and ST6 are increasing function of "k", that is, these tests are sensitive to small expected values in the categories, so that, the status for ST4 is much critical. When "k" is small, the proposed modification is effective and the asymptotic significant level will tend to the true value as n increases. Moreover, the best test according to these tables are ST2 and ST3 which preserve the significant level when "k" increases.

In the table 3, for $k = 6$ and $n = 30$, the true and asymptotic powers of these tests have been simulated on the basis of modified and non-modified statistics based on 10,000 samples and at the significant level 5%. As one can see, the proposed modification will make the true and asymptotic powers closer; moreover, those are the same in the test ST2, and in the test ST3 are almost the same. No specific rule can be proposed whether which test is more powerful.

In the table 4, the amount of asymptotic significant level

$$\beta(k) = P(ST > \chi^2_{1-\alpha}(k-3)|H_0)$$

for normality test of data has been simulated based on 5000 samples and at the significant level 5%. Here, we have categorized our data to have $\hat{\pi}_{10} = \cdots = \hat{\pi}_{k0}$ asymptotically, under normal hypothesis. In this case, the categories should be selected as follows:

$$(-\infty, \bar{x} + z_{\frac{1}{k}} s), [\bar{x} + z_{\frac{1}{k}} s, \bar{x} + z_{\frac{2}{k}} s), \cdots, [\bar{x} + z_{\frac{k-1}{k}}, \infty),$$

in which $\bar{x}$ and s are mean and standard deviation of sample, respectively, and Z_α is $100\alpha\%$ upper point of standard normal distribution. This selection of the categories will make the power function to approach to 1. As one can see in the table 4, the asymptotic significant level of ST1, ST4 and ST6 tests are increasing function of "k" and in the other tests are decreasing function of "k" so that for middle "k" the best tests from the view of significant level are ST2 and ST3.

3. Goodness of fit tests and simulation results for uncategorical data

In this section on using simulation, we have compared the power of some tests in DM-class with two tests based on the empirical distribution function (EDF-class), such as Cramer-von Mises and Anderson-Darling, for normality test of data. The values of simulations have presented in the table 5, that are based on 5000 samples and at the significant level 5%. Here, with selecting some special forms for function "h", we obtain the following tests:

h^*	*Divergence measure*	*Test statistic*
$(x-1)^2$	$\int \left(\frac{g}{f} - 1\right)^2 dF$	$ST1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f_0(x_i)}{f_n(x_i)} - 1\right)^2$
$(\sqrt{x}-1)^2$	$\int \left(\sqrt{\frac{g}{f}} - 1\right)^2 dF$	$ST2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{\frac{f_0(x_i)}{f_n(x_i)}} - 1\right)^2$

h^*	*Divergence measure*	*Test statistic*
$\lvert x-1\rvert$	$\int \left\lvert \frac{g}{f}-1\right\rvert dF$	$ST3=\frac{1}{n}\sum_{i=1}^{n}\left\lvert \frac{f_0(x_i)}{f_n(x_i)}-1\right\rvert$
$(x-1)\log x$	$\int \left(\frac{g}{f}-1\right)\log\left(\frac{g}{f}\right)dF$	$ST4=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{f_0(x_i)}{f_n(x_i)}-1\right)\log\left(\frac{f_0(x_i)}{f_n(x_i)}\right)$
$-\log(x)$	$\int \log\left(\frac{f}{g}\right)dF$	$ST5=\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{f_n(x_i)}{f_0(x_i)}\right)$

ST1 is the continuous version of Neyman statistic [6] and the statistics ST2, ST3, ST4 and ST5 are based on divergence measures: square of Hellinger, Total variations, Jeffreys and Kullback-Leibler, respectively.

It should be noted that Kullback-Leibler information measure can also be written as follows:

$$KL(f,g)=-H(f)-\int log(g)dF$$

in which $H(f)=-\int\log(f)dF$ is called entropy of f. In this case, we can use the following test statistic:

$$-H(f_n)-\frac{1}{n}\sum_{i=1}^{n}\log f_0(x_i,\hat{\theta})$$

in which $H(f_n)$ is an estimator of entropy and $\hat{\theta}$ is a maximum likelihood estimator of θ.

Many papers such as [2], [4], [5], [9], [15] have been written on goodness of fit tests for various distributions through the above-mentioned statistic with using Vasicek entropy estimator [16]. In [10], a modification of entropy estimator has been introduced which has an smaller bias and MSE, also the best selection of window parameter in various distributions and different sample sizes have been obtained. Here, we will use this entropy estimator as follows:

$$H(f_n)=-\frac{1}{n}\sum_{i=1}^{n}\log\left\{\frac{Z_{i+m}-Z_{i-m}}{d_i\frac{m}{n}}\right\}$$

where

$$\begin{cases} Z_{i-m}=a+\frac{i-1}{m}(X_{(1)}-a) & 1\le i\le m,\\ Z_i=X_{(i)} & m+1\le i\le n-m,\\ Z_{i+m}=b-\frac{n-i}{m}(b-X_{(n)}) & n-m+1\le i\le n\end{cases}$$

and

$$d_i=\begin{cases} 1+\frac{i}{i-1+m} & 1\le i\le m,\\ 2 & m+1\le i\le n-m,\\ 1+\frac{n-i+1}{n-i+m} & n-m+1\le i\le n\end{cases}$$

where $b=X_{(n)}+\frac{X_{(n)}-X_{(1)}}{n-1}$ and $a=X_{(1)}-\frac{X_{(n)}-X_{(1)}}{n-1}$.

Moreover, $X_{(1)} < ... < X_{(n)}$ are ordered statistics and $m \leq \frac{n}{2}$ is window parameter. Here, the entropy estimation method has been used for estimation of Kullback-Leibler information measure (ST6). Considering that in practice, Kullback-Leibler information estimation may not be positive, so window parameter used to have positive estimation. Based on our simulation results in Normal distribution, suitable window parameter is closest integer value of $\sqrt[3]{n}$.

The density estimator of data has been calculated with Kernel method with Gaussian Kernel. With regard to our simulation results, to have positive Kullback-Leibler information estimation, the smoothing amount in calculation of the density estimator should be low. Therefore, the band width of $\frac{X_{(n)} - X_{(1)}}{2(1+\log_2^n)}$ is proposed. Also, forms of Cramer-von Mises and Anderson-Darling statistics related to EDF-class is as follows, respectively:

$$CM = \sum_{i=1}^{n} \left(y_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n},$$

$$AD = -n - \frac{1}{n} \sum_{i=1}^{n} \left\{ (2i-1)(\log(1 - y_{n-i+1}) + \log(y_i)) \right\}$$

in which $y_i = F_0(x_{(i)}, \hat{\theta})$ and $x_{(1)} < \cdots < x_{(n)}$ are ordered observations. Taking into account that we intend to perform the hypothesis test $H_0 : X_i \sim N(\mu, \sigma^2)$ with unknown parameters μ and σ^2, to omit unknown parameters, we consider the test problem $H_0 : \frac{X_i - \bar{X}}{S} \sim N(0,1)$ for $i = 1, \cdots, n$ and $n \geq 20$, in which $\bar{X}$ and S are mean and standard deviation of sample respectively, therefore, f_0 and F_0 are density and standard normal distribution functions. Here, the family of distributions namely t-student (t), Chi-square (χ^2), Weibull (W), Beta (β), Skew Normal (SN), Laplace, Logistic and Lognormal (LN) are considered as alternative hypothesis H_1.

On noting to the table 5, when the distribution under H_1 is fairly similar to Normal, the power of the DM-class in compared with EDF-class is lower, and amid tests of DM-class, ST4 is better than others. The important point is that alternative hypothesis is the symmetric Beta distribution, then the power of tests of DM-class is considerably higher.

4. Conclusions

In this paper, an investigation has been performed on some selected goodness of fit tests from the class of divergence measures. Under some regularity conditions, asymptotic distribution of the tests is obtained and a modification is proposed. According to our simulation studies, for normal testing, when the distribution under H_1 is fairly similar to Normal, the power of the DM-class in compared with EDF-class is lower, and when the alternative hypothesis is symmetric Beta distribution, the power of DM-class is considerably higher. Also in the case in which the data are categorized, the best test is still Pearson [11] so that its convergence rate to Chi-square distribution is high and has less sensitivity to small expected values in the categories. In any case, this class of tests can also be studied for Log-linear models and dependence in cross tables.

Remark. The values of tables 1 to 5 are multiply by 1000.

Appendix:

Suppose function "h" has continuous derivatives to fourth order, and $h^{(i)}$ be a derivative of order i; we have the following equations via Taylor's expansion :

$$
\begin{aligned}
E\left(\frac{2n}{h^{(2)}(1)}I(p,\pi_0)\right) &= \sum_{i=1}^{k}\frac{E(W_i^2)}{\pi_{i0}} + \frac{h^{(3)}(1)}{3h^{(2)}(1)}\sum_{i=1}^{k}\frac{E(W_i^3)}{\sqrt{n}\pi_{i0}^2} \\
&+ \frac{h^{(4)}(1)}{12h^{(2)}(1)}\sum_{i=1}^{k}\frac{E(W_i^4)}{n\pi_{i0}^3} + o(n^{-\frac{3}{2}}) \\
E\left(\frac{2n}{h^{(2)}(1)}I(p,\pi_0)\right)^2 &= \sum_{i=1}^{k}\frac{E(W_i^4)}{\pi_{i0}^2} + \sum_{i\neq j}^{k}\frac{E(W_i^2W_j^2)}{\pi_{i0}\pi_{j0}} \\
&+ \left(\frac{h^{(3)}(1)}{3h^{(2)}(1)}\right)^2\left\{\sum_{i=1}^{k}\frac{E(W_i^6)}{n\pi_{i0}^4} + \sum_{i\neq j}^{k}\frac{E(W_i^3W_j^3)}{n\pi_{i0}^2\pi_{j0}^3}\right\} \\
&+ \left(\frac{2h^{(3)}(1)}{3h^{(2)}(1)}\right)\left\{\sum_{i=1}^{k}\frac{E(W_i^5)}{\sqrt{n}\pi_{i0}^3} + \sum_{i\neq j}^{k}\frac{E(W_i^2W_j^3)}{\sqrt{n}\pi_{i0}\pi_{j0}^2}\right\} \\
&+ \left(\frac{h^{(4)}(1)}{6h^{(2)}(1)}\right)\left\{\sum_{i=1}^{k}\frac{E(W_i^6)}{n\pi_{i0}^4} + \sum_{i\neq j}^{k}\frac{E(W_i^2W_j^4)}{n\pi_{i0}\pi_{j0}^3}\right\} + o(n^{-\frac{3}{2}})
\end{aligned}
$$

where $W_i = \sqrt{n}(p_i - \pi_{i0})$, and with respect to results in [14, p. 177-178] ,we have:

$$
\begin{aligned}
E\left(\frac{2n}{h^{(2)}(1)}I(p,\pi_0)\right) &= k - 1 + \frac{1}{n}\{\frac{h^{(3)}(1)}{3h^{(2)}(1)}(2-3k+t) \\
&+ \frac{h^{(4)}(1)}{4h^{(2)}(1)}(2-2k+t)\} + o(n^{-\frac{3}{2}}) \\
E\left(\frac{2n}{h^{(2)}(1)}I(p,\pi_0)\right)^2 &= k^2 - 1 + \frac{1}{n}\{(2-2k-k^2+t) \\
&+ \frac{2h^{(3)}(1)}{3h^{(2)}(1)}(t(k+8) - 6k^2 - 13k + 10) \\
&+ \frac{1}{3}\left(\frac{h^{(3)}(1)}{h^{(2)}(1)}\right)^2(4-6k-3k^2+5t) \\
&+ \left(\frac{h^{(4)}(1)}{2h^{(2)}(1)}\right)(t(3+k) - 5k - 2k^2 + 3)\} \\
&+ o(n^{-\frac{3}{2}})
\end{aligned}
$$

TABLE 1. Monte Carlo Estimations of the asymptotic significant level for homogeneity test for $k = 6$ and $\alpha = 0.05$.

n		ST1	ST2	ST3	ST4	ST5	ST6
10	a	26	40	40	74	26	12
	b	77	40	40	254	51	185
15	a	35	53	44	80	48	33
	b	74	39	44	359	56	126
20	a	50	48	46	148	61	45
	b	85	48	46	165	75	144
25	a	46	47	47	72	57	52
	b	77	47	45	104	64	100
30	a	48	47	47	46	61	42
	b	66	47	47	87	65	79
35	a	43	48	47	41	55	32
	b	59	48	47	78	61	73
40	a	43	50	48	40	55	33
	b	57	45	48	77	61	64
50	a	43	44	45	40	51	35
	b	53	44	45	66	55	62

a: modified statistic and b: non-modified statistic

TABLE 2. Monte Carlo Estimations of the asymptotic significant level for homogeneity test for $n = 30$ and $\alpha = 0.05$.

k		ST1	ST2	ST3	ST4	ST5	ST6
8	a	39	45	45	115	55	32
	b	72	45	44	159	66	123
10	a	35	44	45	76	51	21
	b	83	44	44	297	70	141
12	a	24	49	41	77	44	10
	b	92	39	39	306	69	176
14	a	15	57	42	72	37	4
	b	102	42	41	436	70	203
16	a	9	45	42	47	29	1
	b	97	45	40	473	62	215
18	a	5	48	44	40	21	0
	b	105	48	45	567	61	249
20	a	3	60	40	18	15	0
	b	100	45	39	684	53	275
22	a	9	50	35	9	8	0
	b	108	50	34	732	46	285

a: modified statistic and b: non-modified statistic

TABLE 3. Monte Carlo Estimations of the power for homogeneity test for $n = 30$, $k = 6$ and $\alpha = 0.05$.

$H_1 : (\pi_1, ..., \pi_6)$		ST1	ST2	ST3	ST4	ST5	ST6
1/6,1/6,1/6,1/6,1/6,1/6	A	50	50	50	50	50	50
	B	45	48	47	41	59	37
	C	65	48	47	83	64	74
1/8,1/8,1/8,1/8,2/8,2/8	A	239	265	277	244	238	223
	B	229	265	259	199	256	184
	C	283	265	259	323	275	301
1/9,1/9,1/9,1/9,2/9,3/9	A	468	523	533	431	454	430
	B	454	523	509	399	477	387
	C	517	523	509	522	503	528
1/10,1/10,2/10,2/10,2/10,2/10	A	190	175	190	187	192	186
	B	180	174	176	168	212	156
	C	224	174	176	274	222	254
1/12,1/12,1/12,3/12,3/12,3/12	A	542	549	570	532	548	529
	B	528	548	547	499	570	473
	C	591	548	547	643	590	623
2/15,2/15,2/15,3/15,3/15,3/15	A	107	107	117	103	106	101
	B	100	107	105	90	119	83
	C	131	107	105	163	128	149

A: true power and B(C): modified (non-modified) asymptotic power

TABLE 4. Monte Carlo Estimations of the asymptotic significant level for Normality test for $\alpha = 0.05$.

n	k	ST1	ST2	ST3	ST4	ST5	ST6
30	6	77	66	68	107	85	99
	8	92	53	60	124	87	120
	10	90	56	55	273	81	159
	12	92	54	51	307	73	160
	14	109	53	51	376	84	210
	16	116	50	45	485	77	241
	18	114	48	43	577	74	272
	20	122	54	46	626	72	299
50	6	72	66	68	85	74	82
	10	76	54	54	98	75	92
	14	85	52	52	230	80	141
	18	101	53	50	326	81	202
	22	118	52	50	482	84	244
	26	130	52	46	615	83	302
	30	135	50	40	711	76	362
	34	147	51	37	816	68	413

TABLE 5. Monte Carlo Estimations of the power for Normality test for $\alpha = 0.05$.

H_1	n	ST1	ST2	ST3	ST4	ST5	ST6	CM	AD
$t(1)$	20	648	812	662	821	835	767	872	875
	50	941	993	918	993	992	990	997	997
$t(3)$	20	135	254	121	279	284	168	301	330
	50	248	505	156	583	557	398	579	606
$t(5)$	20	67	131	54	142	128	83	156	74
	50	97	233	70	300	277	141	276	311
$t(10)$	20	48	67	48	71	73	54	84	96
	50	55	88	42	120	109	57	89	105
$\chi^2(1)$	20	921	976	962	975	967	990	947	968
	50	999	1000	999	1000	1000	1000	1000	1000
$\chi^2(3)$	20	458	610	468	608	567	603	538	596
	50	911	969	897	973	977	981	929	962
$\chi^2(5)$	20	284	396	291	400	346	347	340	380
	50	703	812	665	837	828	805	732	806
$W(1,1)$	20	659	799	674	797	769	831	726	777
	50	987	999	978	999	998	999	987	995
$W(2,1)$	20	144	171	149	163	144	130	116	133
	50	283	319	276	340	335	336	217	256
$W(3,1)$	20	79	73	76	71	62	63	50	51
	50	89	68	88	71	74	75	50	54
$\beta(1,1)$	20	366	368	385	354	308	410	133	167
	50	831	832	835	829	851	924	436	566
$\beta(2,2)$	20	134	133	146	126	107	132	59	61
	50	303	268	321	259	253	308	107	130
$\beta(3,3)$	20	80	78	94	72	71	83	44	46
	50	163	141	186	137	140	156	64	70
Laplace	20	58	60	55	55	60	42	61	61
	50	50	52	53	57	52	50	63	60
Logistic	20	121	153	112	152	131	71	88	95
	50	269	319	253	341	314	122	153	169
$SN(1)$	20	224	280	220	281	225	218	226	255
	50	499	582	495	598	577	513	515	573
$SN(2)$	20	182	276	166	290	301	234	426	438
	50	460	630	402	669	642	562	749	753
$SN(3)$	20	56	77	55	85	79	53	97	106
	50	65	108	42	143	131	68	139	159
$LN(0,0.1)$	20	71	77	67	77	66	51	67	70
	50	77	82	79	96	91	73	84	92
$LN(0,0.2)$	20	98	120	100	118	117	86	109	120
	50	167	222	142	254	235	185	208	236
$LN(0,0.5)$	20	347	487	343	491	451	416	439	458
	50	738	870	707	886	880	837	821	868

in which $t = \sum_{i=1}^{k} \frac{1}{\pi_{i0}}$.

REFERENCES

1. S.M. Ali and S.D. Silvey, *A general class of coefficients of divergence of one distributin*, J. R. Statist. Soc. Ser. B, **28**(1966), 131-142.
2. I. Arizono and H. Ohta, *A Test of Normality Based on Kullback-Leibler Information*, JASA, **43**(1989), 20-22.
3. I. Csiszar, *Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizitat von Markoffschen Ketten*, Publ. Nath. Inst. Hungar. Acad. Sci., **8**(1963), 85-107.
4. E.J. Dudewics and E.C. VanderMulen, *Entropy Based Tests of Uniformity*, JASA, **76**(1981), 967-974.
5. N. Ebrahimi, M. Habibullah and E. Soofi, *Testing Exponentiality Based on Kullback-Leibler Information*, J. R. statist. Soc. B, **54**(1992), 739-748.
6. N. Ebrahimi, K. Pflughoeft and E. Soofi, *Two mearures of Sample Entropy*, Statist. prob. Lett., **20**(1994), 225-234.
7. D.S. Moore, *Tests of chi-squared type. In goodness of fit techniques*, New York: Macel Dekker, 1986.
8. J. Neyman, *Contribution to the theory of the* χ^2 *test*, Proceeding of the First Berkeley Symposium on MATHEMATICAL Statistics and Probability, (1949), 239-273.
9. S. Park and D. Park, *Correcting Moments for Goodness of fit Tests based on Two Entropy Estimetes*, J. Statist. Comput. Simul., **73**(2003), 684-694.
10. E. Pasha, M. Kokabi and G.R. Mohtashami, *A Version of the Entropy Estimator via Spacing*, Iranian Inter. J. Sci., **1**(2005), 113-122.
11. K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variable is such that it can reasonably be supposed to have risen from random sampling*, Philosophical Magazine, **5**(1900), 157-175.
12. C.R. Rao, *Linear statistical inference and its applications*, New York: John Wily and Sons, 1973.
13. C.R. Rao, *Diversity and dissimilarity coefficients: a unified approach*, Journal of Theorical Pop. Biology, **21**(1982), 24-43.
14. T.R. Read and N.A Cressie ,*Goodness of fit statistics for discrete multivariate data*, New York: Springar-Varlag, 1988.
15. K.S. Song, *Goodness of fit Tests Based on Kullback-Leibler Discriminiation Information*, IEEE Trans. Inform. Theory, **48**(2002), 1103-1117.
16. O. Vasicek, *A test for normality based on sample entropy*, J. R. Statist. Soc. Ser. B, **38**(1976), 54-59.

Eynollah Pasha: Ph.D in Mathematics,
Department of Mathematics, Teacher Training University, Tehran, Iran.
Mohsen Kokabi Nezhad: Ph.D in Statistics,
Department of Mathematics, Central Tehran Branch, Islamic Azad University,Tehran, Iran.
Gholam Reza Mohtashami Borzadaran: Ph.D in Statistics,
Department of Mathematics, Birjand University, Birjand, Iran.
e-mail:gmohtashami@Birjand.ac.ir and gmb1334@Yahoo.com