Journal of Mathematical Extension
Vol. 3, No. 2 (2009), 43-59

# Goodness of Fit Test and Test of Independence by Entropy

**M. Sharifdoost**
Islamic Azad University Science & Research Branch, Tehran

**N. Nematollahi**
Allameh Tabatabaie University

**E. Pasha**
Teacher Training University

**Abstract.** To test whether a set of data has a specific distribution or not, we can use the goodness of fit test. This test can be done by one of Pearson $X^2$-statistic or the likelihood ratio statistic $G^2$, which are asymptotically equal, and also by using the Kolmogorov-Smirnov statistic in continuous distributions. In this paper, we introduce a new test statistic for goodness of fit test which is based on entropy distance, and which can be applied for large sample sizes. We compare this new statistic with the classical test statistics $X^2$, $G^2$, and $T_n$ by some simulation studies. We conclude that the new statistic is more sensitive than the usual statistics to the rejection of distributions which are almost closed to the desired distribution. Also for testing independence, a new test statistic based on mutual information is introduced.

**AMS Subject Classification:** 62G99; 62H10.
**Keywords and Phrases:** Chi-squared test, goodness of fit test, test of independence, Kolmogorov-Smirnov test, likelihood ratio test, mutual information, relative entropy.

## 1. Introduction

One of the interesting problems in statistics is finding a distribution which fits to a given set of data. In other words, we want to test whether a specific distribution coincides with given data or not. In a goodness of fit test, we compare an unknown distribution q of a random variable, or a random sample, with a given known distribution p.

There are many ways to test goodness of fit. Karl Pearson [18] introduced a statistic for goodness of fit test that is asymptotically equal to the likelihood ratio statistic for large sample sizes, and it has the Chi-squared distribution. Moreover, for testing independence of two random variables, we can use this statistic and compare it to quantiles of Chi-squared distribution.

Sometimes we can compare two distributions by measuring the distance between them with a suitable criterion. One such comparison is Kolmogorov-Smirnov distance $d(F, G) = \sup_t |F(t) - G(t)|$ for continuous distributions F and G. The Kolmogorov-Smirnov distance is zero if and only if F=G (Lehmann and Romano, [14], p. 584). Another way of measuring the distance between two distributions is relative entropy. Also this measure is zero if and only if the two specified distributions are equal (Cover and Thomas, [5], p. 26). By using the concept of relative entropy, we want to construct a test statistic for testing goodness of fit.

The concept of entropy was first introduced in thermodynamics, where it was used to provide a statement of the second law of thermodynamics. Later, statistical mechanics provided a connection between macroscopic property of entropy and microscopic state of the system. This work was the crowning achievement of Boltzman [4]. In ([12]) Hartley introduced a logarithmic measure of the alphabet size. Shanon [20] was the first who defined entropy and mutual information as defined in this paper.

Entropy of a random variable is the measure of uncertainty of that random variable, i.e., measure of the amount of information required on the average to describe random variable. Entropy has many applications in statistical science and engineering. One of the subjects in information theory is mutual information, i.e., the amount of information that one random variable has from other random variables. On the other hand, it is the uncertainty of one random variable with knowledge of other random variable. If the mutual information is zero, then the two random variables are independent. Taking this into consideration, we introduce a measure based on mutual information for testing independence of two random variables in contingency tables.

There are ideas about the construction of the test of goodness of fit based on entropy but they are based on maximum entropy principal; considering a class of densities satisfying criterion restriction and finding a member of this class that maximizes entropy and finding its parametric consistent estimators. Based on this principal, one can find the test statistic of some densities that maximizes the entropy of special class, including uniform, normal, exponential and inverse Gaussian.

Vasicek [21] proposed an entropy-based test for composite hypothesis of normality and provided the critical values and power with Monte Carlo simulation. Dudewicz and Meulen [7,8,9] extended Vasicek's work and proposed an entropy-based test for uniformity and provided the critical values and power by Monte Carlo simulation. Gokhale [10] proposed the entropy-based test construction for a broad class of distributions. Mudholkar and Lin [16] introduced an entropy-based test for exponential hypothesis and prepared the critical values and power by Monte Carlo simulation. Parzen [17] introduced entropy-based test statistic based on difference of order statistic to test goodness of fit of parametric model $\{\mathrm{F}(x, \theta)\}$. Mergel [15] found a test statistic based on maximum entropy for null hypothesis inverse Gaussian with different estimator.

For testing independence, Robinson [19] introduced a test based on an estimator of Kullback-Leibler divergence and studied consistency on testing serial independence for time series. Zheng [22] claimed: " Robinson's test does not have good power against a broad range of alternatives. Morever, the regularity assumptions imposed by the test are so strong that is rules out even some commonly used distribution such as normal."

Goria et al. [11] constructed goodness of fit tests for normal, Laplace, exponential, Gamma, Beta based on maximum entropy principal and introduced a consistent test of independent.

Our test statistics is based on property of relative entropy that is zero if and only if two distributions are equal. This test can be used for every distribution in null and alternative hypotheses.

In Section 2 we review the definition of relative entropy and mutual information and their properties. In Section 3, we review the goodness of fit test and three well-known statistics. In Section 4, we introduce the new statistic for testing goodness of fit based on relative entropy and derive its asymptotic distribution by using limit theorems. To compare the new test procedure with the usual goodness of fit tests, we provide some examples in Section 5. In Section 6, it is shown by some simulation studies that in contrast to the usual tests, the new test is sensitive to the rejection of distributions which are almost close to the desired distribution. In Section 7, a new test statistic based on mutual information for testing independent is derived, and some examples are given. Finally, in Section 8, some conclusions are given.

## 2. Elementary Concepts

Relative entropy was first defined by Kullback and Leibler (1951). It is known under a variety of names including the Kullback-Leibler distance, cross entropy, information diverges and information for discrimination, and has been studied in detail by Csiszar (1967) and Amari (1985).
In this section, we review two related concepts; namely, relative entropy and mutual information.

### 2.1. Relative Entropy

The relative entropy is a measure of the distance between two distributions. It arises on an expected logarithm of the likelihood ratio. The relative entropy $D(q \| p)$ is a measure of the inefficiency of assuming that the distribution is p when the true distribution is $q$.

**Definition 1.** *The relative entropy or Kullback-Leibler distance between two probability mass functions $p(x)$ and $q(x)$ is defined as:*

$$D(q \| p) = \sum_{x \in S(x)} q(x) \log_2 \frac{q(x)}{p(x)} = \mathrm{E}_q \left[ \log_2 \frac{q(X)}{p(X)} \right], \tag{1}$$

*where $S(x)$ is the support of random variable $X$.*
*In the above definition, we use the convention (based on continuity argument) that $0 \log \frac{0}{p} = 0$ and $q \log \frac{q}{0} = \infty$ .*

*The relative entropy is always non-negative and is zero if and only if $p = q$. However, it is not a true distance between two distributions because it is not symmetric and does not satisfy the triangle inequality (Cover and Thomas, 1991, p. 26).*

### 2.2. Mutual Information

The mutual information is a measure of the amount of information that one random variable contains about another random variable. It is the reduction in the uncertainty of one random variable due to knowledge of the other.

**Definition 2.** *Let $X$ and $Y$ be two random variables with joint probability mass functions $p(x,y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The mutual information $I(X,Y)$ is the relative entropy between the joint distribution and the product $p(x)p(y)$, i.e.,*

$$\begin{aligned} I(X,Y) &= \sum_x \sum_y p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \\ &= D(p(x,y)||p(x)p(y)) \\ &= E_{p(X,Y)} \left[ \log_2 \frac{p(X,Y)}{p(X)p(Y)} \right] \end{aligned} \tag{2}$$

*In discrete case, the property of the mutual information is the same as that of relative entropy, i.e., it is always non-negative. This criterion is the measure of dependence of the two random variables and is zero if and only if $X$ and $Y$ are independent. It is symmetric with respect to $X$ and $Y$.*

## 3. Goodness of Fit Test

Suppose the result of a random experiment belongs to one of k disjoint mutual categories $A_1, A_2, \ldots, A_k$, where the probability of belonging to category $A_i$ is $q_i$, $(0 \leqslant q_i \leqslant 1)$ , $\sum_{i=1}^{k} q_i = 1$ , this unknown probability distribution is denoted by $q(x)$. We want to test whether this random experiment has a known probability distribution $p(x)$ that is $p(A_i) =$

$p_i$ , $(0 \leqslant p_i \leqslant 1)$ , $\sum_{i=1}^{k} p_i = 1$. In other words, we want to test $H_0 : q(x) = p(x)$ for all $x$ versus $H_1 : q(x) \neq p(x)$ for at least one $x$. To do this test, we repeat the random experiment n times and let $O_i$ be the frequency of results that belong to $A_i$. If $H_0$ is true, then the expected frequency of results that belong to category $A_i$ $(O_i)$ is $np_i$, that is, $e_i = E(O_i) = np_i$ in n experiments.

We can test the above hypothesis by the following well-known test statistic:

1. Pearson statistic $X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}$. For large sample sizes, $X^2$ has approximately Chi-squared distribution with $k-1$ degrees of freedom. Therefore, we reject $H_0$ if $X^2 > \chi^2(k-1,\alpha)$ where $\chi^2(k-1,\alpha)$ is the $1-\alpha$ quantile of Chi-squared distribution. This test is asymptotic maximin at level of $\alpha$ (Lehmann and Romano, 2005, p.593).

2. Likelihood ratio statistic $G^2 = 2 \sum_{i=1}^{k} O_i \log_2 \left(\frac{O_i}{e_i}\right)$. For large sample sizes, $G^2$ has Chi-squared distribution with k-1 degrees of freedom. Therefore, reject $H_0$ if $G^2 > \chi^2(k-1,\alpha)$. The statistic $G^2$ is asymptotically equivalent to Pearson statistic $X^2$ (Cover and Thomas, [5], p. 333).

3. The Kolmogrov-Smirnov statistic $T_n = \sup_{x \in S(x)} n^{\frac{1}{2}} |F_n(x) - F(x)|$ where $F_n(x)$ is the empirical distribution of sample and F is the distribution function of continuous random variable $X$. Reject $H_0$ if $T_n$ is more than critical value in related tables (Lehmann and Romano, [14], p.584).

In the next section, we introduce another statistic based on relative entropy for testing goodness of fit.

## 4. The Goodness of Fit Test Based on Relative Entropy

Consider testing the hypothesis $H_0 : q = p$ versus $H_1 : q \neq p$. Using the relative entropy given in definition 1. The above testing problem is equivalent to testing the hypothesis $H_0 : D(q \| p) = 0$ versus $H_1$ :

$D(q\,\|p) > 0$. For testing these hypotheses, we know that $D(q\,\|p) = \sum_{i=1}^{k} q_i \log_2 \frac{q_i}{p_i} = E_q\left[\log_2 \frac{q}{p}\right]$.
Let $O_i$ and $e_i$ , $i = 1,\ldots,k$, be the values that are given in Section 3, then the maximum likelihood (ML) estimator of $q_i$ is $\hat{q}_i = \frac{O_i}{n}$. Since $p_i = \frac{e_i}{n}$, so the ML estimator of $D(q\,\|p)$ is given by:

$$\hat{D}\;(\,q\;\|\;p\,) = \frac{1}{n}\sum_{i=1}^{k} O_i \log_2 \frac{O_i}{e_i}. \tag{1}$$

We know that $O = (O_1, O_2, \ldots, O_k)$ has multinomial distribution, i.e.,

$$O = (O_1, O_2, \ldots, O_k) \sim M_k(n, (q_1, q_2, \ldots, q_k)).$$

For a large sample size, $O$ has an asymptotic multivariate normal distribution $N_k(nq, n(D_q - qq'))$, where $D_q$ is a diagonal matrix with diagonal elements $q_i$, $i = 1, 2, \ldots, k$ , and $q = (q_1, q_2, \ldots, q_k)$. Therefore,

$$\sqrt{n}\left(\frac{1}{n}O - q\right) \xrightarrow{d} N_k\left(0, D_q - qq'\right),$$

where $\xrightarrow{d}$ denotes the convergence in distribution (Agresti, [1], p. 580). With simple algebra and using limit theorems and equations (1) and (3), we can show:

$$Z = \sqrt{n}\left(\frac{\hat{D}(q||p) - D(q||p)}{\hat{\sigma}}\right) \xrightarrow{d} N(0,1), \tag{2}$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\left[\sum_i O_i\left(\log_2 \frac{O_i}{e_i}\right)^2 - \left(\sum_i O_i \log_2 \frac{O_i}{e_i}\right)^2\right]. \tag{3}$$

So, from the above argument and asymptotic distribution of Z given by (4), in testing the hypothesis $H_0 : D(q\,\|p) = 0$ versus $H_1 : D(q\,\|p) > 0$, we can reject $H_0$ if $Z_0 > Z_\alpha$ where

$$Z_0 = \frac{\sqrt{n}\hat{D}(q||p)}{\hat{\sigma}}, \tag{4}$$

and $Z_\alpha$ is the 1- $\alpha$ quantile of standard normal distribution.

**Remark:** The above statistic is very close to $G^2$, but we will see by simulation study, it is more sensitive than $G^2$.
Note that we cannot compare the power of the new test procedure to the usual goodness of fit tests based on $X^2$, $G^2$ and $T_n$ statistics, since the class of alternatives typically is enormously large and can no longer be described by a parametric model (Lehmann and Romano, 2005, p.583). Instead, in Section 6 we use some simulation studies to compare the p-values of these test procedures and compare the sensitivity of these test statistics. Before doing this, we will take a look at some examples in the next section.

## 5. Examples

In this section, some examples for using and comparing the test procedures that considered in previous sections are given (Bhattacharyya and Johnson,1977).

**Example 1**. From a large population, a sample of 200 is selected and the number of times that they go to an insurance company in a period of 4-years is recorded in the following table. We want to test that the distribution of data follows a Poisson distribution.

| Number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | sum |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 22 | 53 | 58 | 39 | 20 | 5 | 2 | 1 | 200 |

Let X be the number of times that a person goes to the insurance company. We want to test $H_0 : X \sim P(\lambda)$ versus $H_1 : X \nsim P(\lambda)$. Using test statistic (6), we have $Z_0 = 0.8012$ and $Z_{0.05} = 1.645$. So, $H_0$ is accepted. The Pearson statistic is $X^2 = 2.33$, which is not greater than $\chi^2(4, 0.05) = 9.487$. So, again the hypothesis $H_0$ is accepted.

**Example 2**. In a general election in a country about a subject, the percentage of answers is:

| opinion | very agreeable | agreeable | abstention | disagreeable | very disagreeable |
|---|---|---|---|---|---|
| percentage | 20 | 30 | 20 | 20 | 10 |

A sample of 100 is selected and the results are collected in the following table.

| opinion | very agreeable | agreeable | abstention | disagreeable | very disagreeable |
|---|---|---|---|---|---|
| frequency | 14 | 18 | 18 | 26 | 24 |

Do these data coincide with general election?
By considering the above tables and using (6),we have $Z_0 = 2.3156$, and by comparing it to $Z_{0.05} = 1.645$, we conclude that the hypothesis is rejected. Also $X^2 = 46.5$ and $\chi^2(4, 0.05) = 9.487$, so at the level $\alpha = 0.05$, we conclude that $H_0$ is rejected and the two methods have the same conclusion.

**Example 3**. The following table shows the results of tossing a dice 120 times. We want to test whether the dice is biased or not.

| spot | 1 | 2 | 3 | 4 | 5 | 6 | sum |
|---|---|---|---|---|---|---|---|
| frequency | 18 | 23 | 16 | 21 | 18 | 24 | 120 |

By considering $p_i = \frac{1}{6}$ and $e_i = 20$, we have $Z_0 = 0.7914$. Comparing it with $Z_{0.05} = 1.645$, we conclude that $H_0$ is accepted. Also, $X^2 = 2.5$ and comparing it with $\chi^2(5, 0.05) = 11.1$, we conclude that $H_0$ is also not rejected.

## 6. Simulation Study

As we told in Section 4, the class of alternative hypothesis is very large and contains all distributions except the distribution given in $H_0$. Thus, we cannot compute and compare the power function of the new test procedure with the usual ones. In this section, we generate random numbers from Poisson and exponential distributions and carry out the goodness of fit test by the test statistics which are discussed in this paper. To compare these test functions, we use their p-values.

### 6.1. Poisson Test

In this section, we simulate 10000 random numbers from $P(\lambda_1)$ for $\lambda_1 = 1, 2$ and consider testing hypothesis $H_0 : X \sim P(\lambda_2)$ versus $H_1 : X \nsim P(\lambda_2)$, where values of $\lambda_2$ are given in Tables 1 and 2, and compute values of $\hat{d} = \hat{D}(q \| p)$, $\hat{\sigma}^2 = S_d^2$, $Z, X^2, G^2$ and p-values corresponding to Z, $X^2$ and $G^2$ statistics. Since the data are generated from Poisson distribution with $\lambda_1 = 1$, from Table 1, we see that for values of $\lambda_2$ near 1(rows I), all procedures accept $H_0$ and for values of $\lambda_2$ far from 1(rows III), all procedures reject $H_0$. But, for values of $\lambda_2$ between (1.0196 , 1.02) the new method based on Z statistic reject $H_0$ (which is correct) but the usual methods based on $X^2$ and $G^2$ accept $H_0$ (which is not correct). This shows that the new statistic Z is more sensitive than the usual statistic to the rejection of distributions which are almost close to the desired distribution.

**Table 1.** Results of testing $H_0 : X \sim P(\lambda_2)$ versus $H_1 : X \nsim P(\lambda_2)$ for 10000 random numbers generated from P(1).

| | $\lambda_2$ | $\hat{d}$ | $S_d^2$ | Z | $X^2$ | $G^2$ | p-value (z) | p-value ($X^2$) | p-value ($G^2$) |
|---|---|---|---|---|---|---|---|---|---|
| I | 1 | 0.0008 | 0.0048 | 1.15 | 2.32 | 2.36 | 0.1251 | 0.89 | 0.88 |
| | 1.005 | 0.0008 | 0.0049 | 1.24 | 3.21 | 3.29 | 0.1075 | 0.7788 | 0.774 |
| | 1.01 | 0.0009 | 0.005 | 1.35 | 4.58 | 4.71 | 0.0855 | 0.6 | 0.584 |
| | 1.015 | 0.0011 | 0.005 | 1.49 | 6.43 | 6.61 | 0.068 | 0.4 | 0.37 |
| | 1.019 | 0.0012 | 0.006 | 1.63 | 8.25 | 8.48 | 0.0526 | 0.22 | 0.215 |
| | 1.0195 | 0.0013 | 0.006 | 1.646 | 8.49 | 8.74 | 0.0505 | 0.21 | 0.201 |
| II | 1.0196 | .0013 | 0.006 | 1.65 | 8.54 | 8.79 | 0.0495 | 0.21 | 0.198 |
| | 1.0197 | 0.0013 | 0.006 | 1.654 | 8.59 | 8.84 | 0.0495 | 0.209 | 0.196 |
| | 1.02 | .0013 | 0.006 | 1.66 | 8.75 | 8.99 | 0.0485 | 0.2 | 0.188 |
| III | 1.03 | 0.0017 | 0.007 | 2.05 | 14.77 | 15.2 | 0.0202 | 0.023 | 0.02 |
| | 1.04 | 0.0023 | 0.0085 | 2.5 | 22.6 | 23.27 | 0.0062 | 0 | 0 |
| | 1.05 | 0.003 | 0.001 | 2.96 | 32.18 | 33.18 | 0.0015 | 0 | 0 |

**Table 2.** Results of testing $H_0 : X \sim P(\lambda_2)$ versus $H_1 : X \nsim P(\lambda_2)$ for 10000 random numbers generated from $P(2)$.

| | $\lambda_2$ | $\hat{d}$ | $S_d^2$ | Z | $X^2$ | $G^2$ | p-value (z) | p-value ($X^2$) | p-value ($G^2$) |
|---|---|---|---|---|---|---|---|---|---|
| I | 2 | 0.001 | 0.005 | 1.58 | 3.72 | 3.81 | 0.0571 | 0.88 | 0.86 |
| | 2.001 | 0.001 | 0.005 | 1.6 | 3.87 | 3.96 | 0.0548 | 0.86 | 0.855 |
| | 2.005 | 0.0017 | 0.005 | 1.65 | 4.53 | 4.64 | 0.0495 | 0.8 | 0.79 |
| II | 2.01 | 0.0012 | 0.005 | 1.73 | 5.58 | 5.72 | 0.0418 | 0.69 | 0.68 |
| | 2.03 | 0.002 | 0.006 | 2.17 | 12.17 | 12.47 | 0.0150 | 0.16 | 0.14 |
| | 2.04 | 0.002 | 0.008 | 2.45 | 16.89 | 17.3 | 0.0071 | 0.033 | 0.028 |
| III | 2.05 | 0.003 | 0.008 | 2.75 | 22.52 | 23.08 | 0.003 | 0 | 0 |
| | 2.06 | 0.003 | 0.0095 | 3.06 | 29.07 | 29.82 | 0.0011 | 0 | 0 |
| | 2.07 | 0.0035 | 0.011 | 3.38 | 36.52 | 37.48 | 0 | 0 | 0 |

Similar results can be seen from Table 2, where we generate 10000 random numbers from P(2) and then test the hypothesis $H_0 : X \sim P(\lambda_2)$ versus $H_1 : X \nsim P(\lambda_2)$ for some values of $\lambda_2$. Again we see that for values $\lambda_2 \in (2.005, 2.03)$ the new method based on Z statistic rejects $H_0$ but the other methods accept it.

### 6.2. Exponential Test

Similar to the previous section, we simulate 10000 random numbers from exp(1) and consider testing hypothesis $H_0 : X \sim exp(\lambda_2)$ versus $H_1 : X \nsim exp(\lambda_2)$, where values of $\lambda_2$ are given in Table 3 and computed $\hat{d}$, $\hat{\sigma}^2$, Z, $X^2$, $G^2$, $T_n$ and p-values corresponding to Z, $X^2$, and $G^2$ statistics where the critical value of $T_n$ is 1.36. We see that for values of $\lambda_2$ near 1 (rows IV), all procedures accept $H_0$ and for $\lambda_2 = 1.06$ that is far from 1 (row I), all procedures reject $H_0$. For values of $\lambda_2$ in (1.039,1.041) the new method based on Z statistic rejects $H_0$ but the usual methods based on $X^2$, $G^2$, and $T_n$ accept $H_0$. This shows the new method is more sensitive than the others to the rejection of distribution which are almost close to the desired distribution. Although we see in row III for $\lambda_2 = 1.05$, the methods based on $X^2$ and $G^2$.

**Table 3.** Results of testing $H_0 : X \sim \exp(\lambda_2)$ versus $H_1 : X \nsim \exp(\lambda_2)$ for 10000 random numbers generated from exp(1).

| | $\lambda_2$ | $\hat{d}$ | $S_{\hat{d}}^2$ | Z | $X^2$ | $G^2$ | $T_n$ | p-value (z) | p-value ($X^2$) | p-value ($G^2$) |
|---|---|---|---|---|---|---|---|---|---|---|
| I | 1 | 0.0004 | 0.0012 | 1.15 | 4.09 | 4.1 | 0.76 | 0.1251 | 0.767 | 0.766 |
| | 1.02 | 0.0004 | 0.0014 | 1.17 | 4.9 | 4.66 | 0.40 | 0.121 | 0.673 | 0.7 |
| | 1.03 | 0.0007 | 0.0021 | 1.42 | 8.1 | 7.62 | 0.64 | 0.0778 | 0.337 | 0.38 |
| II | 1.039 | 0.0009 | 0.003 | 1.72 | 12.58 | 11.75 | 0.87 | 0.0427 | 0.0862 | 0.11 |
| | 1.04 | 0.001 | 0.0032 | 1.75 | 13.18 | 12.3 | 0.90 | 0.0401 | 0.072 | 0.09 |
| | 1.041 | 0.001 | 0.0033 | 1.79 | 13.79 | 12.86 | 0.92 | 0.0367 | 0.0573 | 0.083 |
| III | 1.05 | 0.0014 | 0.0046 | 2.13 | 20.13 | 18.66 | 1.10 | 0.0166 | 0.0055 | 0.009 |
| IV | 1.06 | 0.002 | 0.0065 | 2.52 | 28.97 | 26.67 | 1.38 | 0.0059 | 0 | 0 |

## 7. Test of Independence Based on Mutual Information

Suppose a random sample of size n is drawn from a population. The observations in the random sample are classified according to the two criteria. Using the first criterion, each observation is associated with one of the R rows, and using the second criterion, each observation is associated with one of the C columns. Consider $O_{ij}$ as the number of observations associated with row i and column j simultaneously. These $O_{ij}$ s are arranged in a R×C contingency table.

| Y / X | 1 | 2 | $\cdots$ | $C$ | Sum |
|---|---|---|---|---|---|
| 1 | $O_{11}$ | $O_{12}$ | $\cdots$ | $O_{1C}$ | $O_{1.}$ |
| 2 | $O_{21}$ | $O_{22}$ | $\cdots$ | $O_{2C}$ | $O_{2.}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $R$ | $O_{R1}$ | $O_{R2}$ | $\cdots$ | $O_{RC}$ | $O_{R.}$ |
| Sum | $O_{.1}$ | $O_{.2}$ | $\cdots$ | $O_{.C}$ | $n$ |

The total number in $i$-th row is $O_i.$ and $j$-th column is $O_{.j}$ and the sum of numbers in all the cells is $n$. For testing $H_0$: (the event "an observation is in row $i$" is independent of the event "that some observation is in column $j$" for all $i$ and $j$.), by the definition of independence of events,

we can state as follows:
$H_0$: p(row $i$, column $j$) = p(row $i$) ×p(column $j$) for all $i, j$
Versus
$H_1$: p(row $i$, column $j$)≠ p(row $i$) ×p(column $j$) for some $i, j$
The test statistic is given by

$$X^2 = \sum_j \sum_i \frac{(O_{ij} - e_{ij})^2}{e_{ij}},$$

where $e_{ij} = \frac{O_{i.} \times O_{.j}}{n}$.This statistic has Chi-squared distribution with $(R-1)(C-1)$ degrees of freedom. So, reject $H_0$ at level $\alpha$ if $X^2 > \chi^2((R-1)(C-1), \alpha)$.
Now using the mutual information given in definition 2, the above testing problem is equivalent to testing the hypothesis $H_0 : I(X, Y) = 0$ versus $H_1 : I(X, Y) \neq 0$, i.e., to test $H_0 : D(p(x, y)||p(x)p(y)) = 0$ versus $H_1 : D(p(x, y)||p(x)p(y)) > 0$.
Testing the above hypothesis is similar to that of Section 4. In this case, the ML estimator of $I(X, Y)$ is given by

$$\hat{I}\,(X, Y) = \frac{1}{n} \sum_i \sum_j O_{ij} \log_2 \frac{O_{ij}}{e_{ij}},$$

where $O = (O_{11}, O_{12}, \ldots, O_{RC})$ is the observed values and has a multinomial distribution. For large sample size it has an asymptotic normal distribution. By using the same argument as in Section 4, we have.

$$Z = \frac{\sqrt{n}\left(\hat{I}\,(X, Y) - I\,(X, Y)\right)}{\hat{\sigma}} \xrightarrow{d} N\,(0, 1)\,,$$

where

$$\hat{\sigma}^{\ 2} = \frac{1}{n}\left[\sum_i \sum_j O_{ij}\left(\log_2 \frac{O_{ij}}{e_{ij}}\right)^2 - \left(\sum_i \sum_j O_{ij} \log_2 \frac{O_{ij}}{e_{ij}}\right)^2\right].$$

So, we reject $H_0$ if $Z_0 = \frac{\sqrt{n}\left(\hat{I}(X,Y)\right)}{\hat{\sigma}} > Z_\alpha$.

The following examples introduce the use of this model of testing in comparison to the usual test of independence based on $X^2$-statistic (Bhattacharyya and Johnson,1977).

**Example 4**. 1200 persons are classified into four business groups and two types of opinion. The results are classified in the following table. Are these two types of opinion independent from each other?

| opinion / group | I | II | Sum |
|---|---|---|---|
| A | 32 | 269 | 300 |
| B | 51 | 199 | 250 |
| C | 67 | 233 | 300 |
| D | 83 | 267 | 350 |
| Sum | 233 | 967 | 1200 |

From this table, we have $X^2 = 20.59$ and by comparing it with $\chi^2(3, 0.05)$ $= 7.815$, we reject the independent hypothesis. Based on mutual information, we have $Z_0 = 2.514$ and by comparing it with $Z_{0.05} = 1.645$ we also reject the independent hypothesis. Therefore, the two methods have the same result.

**Example 5**. To distinguish the efficiency of a chemical treatment on seeds, we choose 100 seeds with treatment and 150 seeds without treatment and gain the following results:

| | positive | negative |
|---|---|---|
| with treatment | 84 | 16 |
| without treatment | 132 | 18 |

for testing independence, we have $X^2 = 0.817$ and $\chi^2(1, 0.05) = 3.841$, which imply independence. Moreover, with the new method, we have $Z_0 = 0.4456$, which is not greater than $Z_{0.05} = 1.645$. Thus, we conclude that the independent hypothesis is not rejected.

## 8. Conclusion

In this paper, we introduce a new statistic for goodness of fit test based on relative entropy and compare it with the classical statistics $X^2$, $G^2$ and $T_n$ (in the continuous case) by simulation studies. It is seen that goodness of fit test based on relative entropy is more sensitive than the usual ones to the rejection of distributions which are almost close to the desired distribution. Also, to test the independence,we derive a new test based on mutual information.

## References

[1] A. Agresti, *Categorical data analysis.* 2rd edition. John Wiley and Sons, 2002.

[2] S. Amari, *Differential-Geometrical Methods in statistics.* Springer-Verlag, New York, 1985.

[3] K. G. Bhattacharyya and A. R. Johnson, *Statistical concept and methods.* John Wiley and Sons, 1977.

[4] L. Boltzman, *Lectures on Gas theory.* University of California Press, Berkeley, 1964.

[5] Cover, M. Thomas and Thomas, A. Joy, *Elements of information theory.* John Wiley and Sons, 1991.

[6] I. Csiszar, Information type measures of difference of probability distribution and indirect observation. *Studia Sci. Math. Hungar*, 2 (1967), 229-318.

[7] E. J. Dudewicz and E. C. Van der Meulen, Entropy- based test of uniformity, *J. Amer. Statist. Assoc.* 76 (1981), 967-974.

[8] E. J. Dudewicz and E. C. Van der Meulen, Entropy-based random evalution, *American J. Mathematical and Management Sciences*, 15 (1995), 115-153.

[9] E. J. Dudewicz and E. C. Van der Meulen, On the power of entropy-based tests against bump-type altermatives. In Contributions to Stochastics, *Ed. W. Sendler, Physica-Verlag*, (1987), 84-88.

[10] D. V. Gokhale, On entropy-based goodness-of-fit test. *Computational Statistics and Data Analysis*, 1 (1983), 157-165.

[11] M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. Novi inverardi, A new class of random vector entropy estimators and its applications in testing statistical hypotheses. *Journal of Nonparametric Statistics*. 17: 3 (2005), 227-297.

[12] R. V. Hartley, Transmission of information. *Bell Sys. Tech. Journal*, 7 (1928), 535.

[13] S. Kullback and R. A. Leibler, On information and sufficiency. *Ann. Math. Stat.*, 22 (1951), 79-86.

[14] E. L. Lehmann and J. P. Romano, *Testing statistical hypothesis*. 3rd edittion. John Wiley and Sons, 2005.

[15] V. Mergel, Test of goodness of fit for the inverse-Gaussian distribution. *Mathematical Communications*, 4 (1999), 191-195.

[16] G. S. Mudholkar and C. T. Lin, On two application of characterization theorems to goodness-of-fit. Colloquia *Mathematica Soc. Janos Bolyai*, (1984), 395-414.

[17] E. Parzen, Goodness of fit tests and entropy. *J. Combinatorics, Information and System Sciences*, 16 (1991), 129-136.

[18] K. Pearson, On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling, *Philosophical Magazine* (5), 50 (1900), 157-175.

[19] P. M. Robinson, Consistent nonparametric entropy-based testing, *Review of Economic studies*, 58 (1991), 437-453.

[20] C. E. Shanon, A Mathematical Theory of Communication, *Bell Sys. Tech. Journal*, 27 (1948), 379-423 , 623-656.

[21] O. Vasicek, A test of normality based on sample entropy, *J. Roy. Statist. Soc. Ser. B*, 38 (1976), 54-59.

[22] J. X. Zheng, A consistent specification test of independence, *Journal of Nonparametric statistics*, 7 (1997), 297-306.

**Maryam Sharifdoost**
Ph.D student, Islamic Azad University
Science & Research Branch
Tehran, Iran.
E-mail: sharifdoost@gmail.com

**Nader Nematollahi**
Department of Statistics
Allameh Tabatabaie University
Tehran, Iran.
E-mail: na_nemat@yahoo.com

**Einollah Pasha**
Department of Statistics
Teacher Training University
Tehran, Iran.
E-mail: pasha@saba.tmu.ac.ir