AN OBSERVATION ABOUT GOODNESS-OF-FIT TESTS OF DISTRIBUTIONS

Pritibhushan Sinha
Associate Professor (Operations Management)
ICFAI Business School Hyderabad – 501504 Andhra Pradesh INDIA
Email: p_sinha@live.com

Abstract

We consider univariate goodness-of-fit tests. A result stating the non-existence of a uniformly most powerful test, for all combinations of null and alternative distributions, which are completely specified, is given. Key Words: Univariate Distribution, Goodness-of-Fit Tests, Uniformly Most Powerful Test.

In a univariate goodness-of-fit (GoF) test of distributions, one verifies if a random variable (RV) \( X \) follows a hypothesized distribution function \( F(x) = \Pr\{X \leq x\} \) or otherwise. Another distribution may be considered as an alternative for the RV. Distributions may or may not be completely specified. Some parameters may be estimated from the sampled observations. Such GoF tests are of singular importance and have many applications in management, engineering and science. For example, we may try to see the statistical distribution for the lifetime of a machine and its reliability of operation at a particular time point. We may need to estimate repair times. Such distributions then may be used arrive at optimal maintenance policies. The estimated distribution of demand of an item may be used in production planning. GoF tests also arise in a regression analysis, which again is a topic of high practical significance.

In principle, a GoF test would compare some characteristic features of the ideal distribution, with what is observed in the sampled data, corresponding to such features. There have been extensive discussions in the literature on GoF methods. Many non-parametric and parametric tests have been suggested, using different statistics. The readers may find more elaborate discussion on the topic, for example, in the text by D’Agostino and Stephens (1986).

One commonly used method for GoF test of distribution is the Chi-squared Test (see, for instance, D’Agostino and Stephens (1986), Spanos (1999)) given by Prof. Karl Pearson. The test is based on differences in expected and actual number of different cells or subsets of the population. This gives binomial distributions which are approximated with the standard normal distribution (N(0,1)). Then the fact that the sum of the squares of a few random variables which are independent and follow N(0,1) is a RV, which follows the chi-square distribution, is used. Chi-squared Test is a large sample test, because of the normal approximation used. But, it is a non-parametric test, applicable for both of continuous and discrete distributions. The test is also modified without much complication, to take into account the situations where some parameters of the distribution have to be estimated from sampled observations.

Another important type of GoF test is based on empirical distribution function (EDF). The well-known Kolmogorov-Smirnov Test (refer to, for instance, Spanos 1999) is one of such techniques. The EDF may be described in the following manner. Let \( X_{(1)}, X_{(2)}, \ldots, X_{(n)} \) be n ordered sampled observations from the distribution \( F(.) \) and let one realization be denoted as, \( x_{(1)}, x_{(2)}, \ldots, x_{(n)} \). Then EDF \( \hat{F}_n(x) \) is given as:

\[
\hat{F}_n(x) = \begin{cases} 
0, & x \leq X_{(1)}; \\
i/n, & X_{(i)} \leq x < X_{(i+1)}, i = 1, 2, ..., n-1; \\
1, & x \geq X_{(n)}. 
\end{cases}
\]

An oft-used test based on the EDF is the Anderson-Darling Test (Anderson and Darling 1952). Further discussions on some EDF tests and a comparison are found in the article by Stephens (1974).
Various other statistics as empirical characteristic function (see, for instances, the papers by Gurtler and Henze 2000, Wong and Kim 2000), empirical Melin transform of distribution function (by Meintanis 2008) etc. have been proposed by different authors.

In this article, we give a result that, there cannot be a single test which would be uniformly most powerful in all combinations of null and alternative single variable distributions, which are completely specified. As much as we are aware, such a result, although quite intuitive, has not been discussed analytically in the relevant literature. The result has some practical implications.

**A Result on Goodness-of-Fit Tests**

Denote with n the sample size and with \( a \), the level of significance. Let us consider tests as,

- \( H_0 \) (null hypothesis): \( X \) follows the distribution \( F_0 \);
- \( H_1 \) (alternative hypothesis): \( X \) follows the distribution \( F_1 \);

Where \( F_0 \) and \( F_1 \) are completely specified (\( F_0 \neq F_1 \)). We assume simple random sampling, \( X_1, X_2, \ldots, X_n \) being the independent sample observations. Tests have acceptance criteria as, \( a_i(\alpha, n) \leq t(X_1, X_2, \ldots, X_n) \leq a_2(\alpha, n) \), where \( t(X_1, X_2, \ldots, X_n) \) is the statistic used in the test. It does not depend on \( F_0 \) and \( F_1 \). The critical values \( a_1, a_2 \) are real values which possibly are dependent on \( \alpha \) and \( n \); but are not, like the test statistic, dependent on \( F_0 \) and \( F_1 \). For example, this is the case with the Chi-squared Test. A test should have level of significance \( \alpha \) or lower.

**Proposition 1:** For fixed level of significance \( \alpha < 1 \) and sample size, there cannot be a test which has higher power than any other test for all combinations of \( F_0 \) and \( F_1 \).

Proof: Let there be such a test with the statistic \( t_0(X_1, X_2, \ldots, X_n) \). Consider, \( F_0 \) as \( X = 0 \), with probability 1; \( F_1 \) as \( X = 1 \), with probability 1. Then, \( a_1 \leq t_0(0, 0, \ldots, 0) \leq a_2 \); otherwise, level of significance is 1. Power is zero if, \( a_1 \leq t_0(1, 1, \ldots, 1) \leq a_2 \). So take, \( t_0(1, 1, \ldots, 1) < a_1 \); or, \( t_0(1, 1, \ldots, 1) > a_2 \), when power is 1.

Then, take the reversed condition, \( F_0 \) as \( X = 1 \), with probability 1; \( F_1 \) as \( X = 0 \), with probability 1. In this case, level of significance is 1, power is zero. A test can be designed for this case as, \( t(X_1, X_2, \ldots, X_n) = \sum X_i/n, a_1 = a_2 = 1 \) giving level of significance of 0 and power 1.

Thus, the proposition is proved. \( \square \)

It is not difficult to see that, instead of degenerate Bernoulli distributions, Bernoulli distributions with probabilities in \((0, 1)\) could also be used in the proof, with some probabilities sufficiently close to 1. Other distributions may also be found to construct such examples.

The proposition is an analytical statement about the non-existence of a uniformly most powerful test in the situation considered and highlights the need of devising different types of tests for different combinations of hypothesized null and alternative distributions. Acceptance criteria too should take into account the combination of null and alternative distributions, for more accurate tests. This in most cases can be done with numerical experiments.

We may also note that, from the above it follows that if the tests are as,

- \( H_0 \) (null hypothesis): \( X \) follows the distribution \( F_0 \);
- \( H_1 \) (alternative hypothesis): \( X \) does not follow the distribution \( F_0 \),

In that case also there cannot be any test which is uniformly most powerful for all \( F_0 \). This is so as power has to be considered here for various alternative distributions.

**Conclusion**

Goodness-of-fit tests for distribution are of high practical importance. For a given situation, such a test needs to be selected appropriately. A suitable test can give higher accuracy, both in terms of level of significance and power, even with a smaller sample size. But selection of an appropriate test is sometimes difficult- a large number of tests being available. In this article, we have presented a result highlighting the need of having different tests for different conditions. Variety of tests is preferable. But there should be guidelines to select the most appropriate test for the problem at hand, which would help the practitioners and the researchers.

**References**


