ANALYSIS ON GRAPHICAL PROCEDURES AND REFINEMENTS FOR NORMALITY TEST

Famus Xie, Raymond Xue, Richard Zhu
University of Washington - Department of Statistics

INTRODUCTION

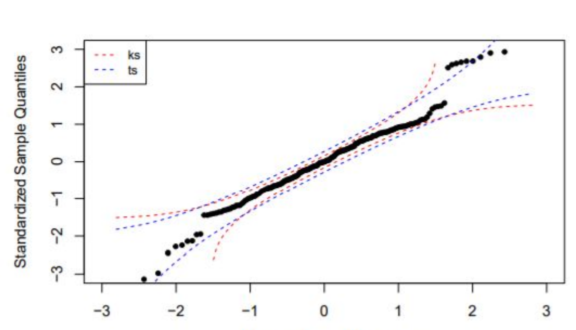
Assessing normality is an important way of summarizing the nature of the data. While many useful methods for doing so exist, we have explored two papers that adopt two different approaches for graphical analysis of normality, namely the TS test and Monte Carlo test. Both advocate for an improvement on the existing methodology.

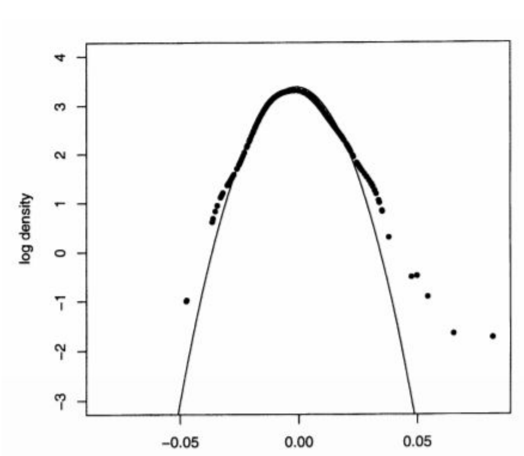
Among the common tests for testing normality, Kolmogorov-Smirnov (KS) test is the only method for which has a graphical representation (Aldor-Noiman, 2013). The KS test compares an empirical cumulative distribution of a sample with the target null distribution, which in our case, is the normal distribution. The test statistic can be used to generate confidence bands under the null hypothesis and plot them along with the QQ-plot for the test data. Then the test simply rejects the null hypothesis if any test data point falls outside of the bands.

The similar tail-sensitive (TS) approach has higher power when applied to a distribution with heavy tails or outliers. The TS test constructs confidence bands using simulation. The procedure first generates Uniform(0,1) samples and forms confidence intervals around the order statistics. Next, it adjusts the confidence level of the intervals to account for a simultaneous coverage rate as 1 - α where α is the specified type I error. The collection of confidence intervals forms the confidence bands which are then transformed back into the normal scale by inverse normal CDF. This is valid and easy to compute because the i-th order statistic of each uniform quantile follows a Beta distribution.

Monte Carlo simulation test (MC) based on kernel density estimation was the second graphical procedure that we looked at. The test compares the estimated log-density with the log of the fitted normal distribution. The test statistic is simply the sum of squared residuals (SSR) obtained from squaring the difference between the two terms. Higher values for the statistic suggest that it is more likely to reject the null hypothesis. This procedure appears to have higher power than other well-known tests such as the Shapiro-Wilk test at N = 50 (Hazelton, 2003).

KS and TS band





log density used in Monte Carlo test

INDEX FOR ABBREVIATIONS

KS - Kolmogorov-Smirnov test; TS - tail-sensitive test; MC - monte carlo simulation; KDE - kernel density estimation; SSR - sum of squared residuals; MLE - maximum likelihood estimator; MAD - mean absolute deviation;

METHODOLOGY

We performed a power analysis on simulated data from 12 different distributions. Most of these distributions are either skewed or heavily-tailed. We compared the KS test, TS test, and Monte Carlo test based on their power to determine whether these simulated data can be well modeled by a normal distribution. We prefer the test which gives larger power against these data, except for the normal data that are used for assessing type I error.

To compute the power, we run experiments with 1000 iterations for each distribution. That is, we draw 1000 samples each from the 12 distributions and then calculate the proportions of times that the three tests reject the null hypothesis that the data are normally distributed. We set alpha = 0.05 and sample size n = 100.

Since the Monte Carlo test requires computing the kernel density estimate for the sample data, we also compare power with different kernels and bandwidths. We pick the following commonly used kernel functions: Gaussian, Biweight, Box, Triangular, Epanechnikov, Tricube.

All the simulations are implemented in R.

RESULTS

Figure 1 shows the powers of TS and MC Test on 12 distributions that 1 of them is normal and all of the others are not normal. Estimates for the mean and standard deviation are based on each corresponding paper. Kernel functions are Gaussian kernel with bandwidth chosen by asymptotic normal reference rule. According to the plot, TS and MC Test have very similar or higher power than KS Test except for Poisson(15). TS Test has higher power than MC Test for Chi-square(df = 5), Poisson(15), uniform(1,18), binomial(10,0.2), hypergeometric(20,10,5). MC Test has higher power than TS Test for T(df = 4), logistic distribution, and Laplace(0,50). These two have very similar power for log-normal distribution, Chi-square(df = 1), Chi-square(df = 100) and normal(5,2). Both methods demonstrated to have a much high power when testing fat-tail distributions such as t(df = 4).

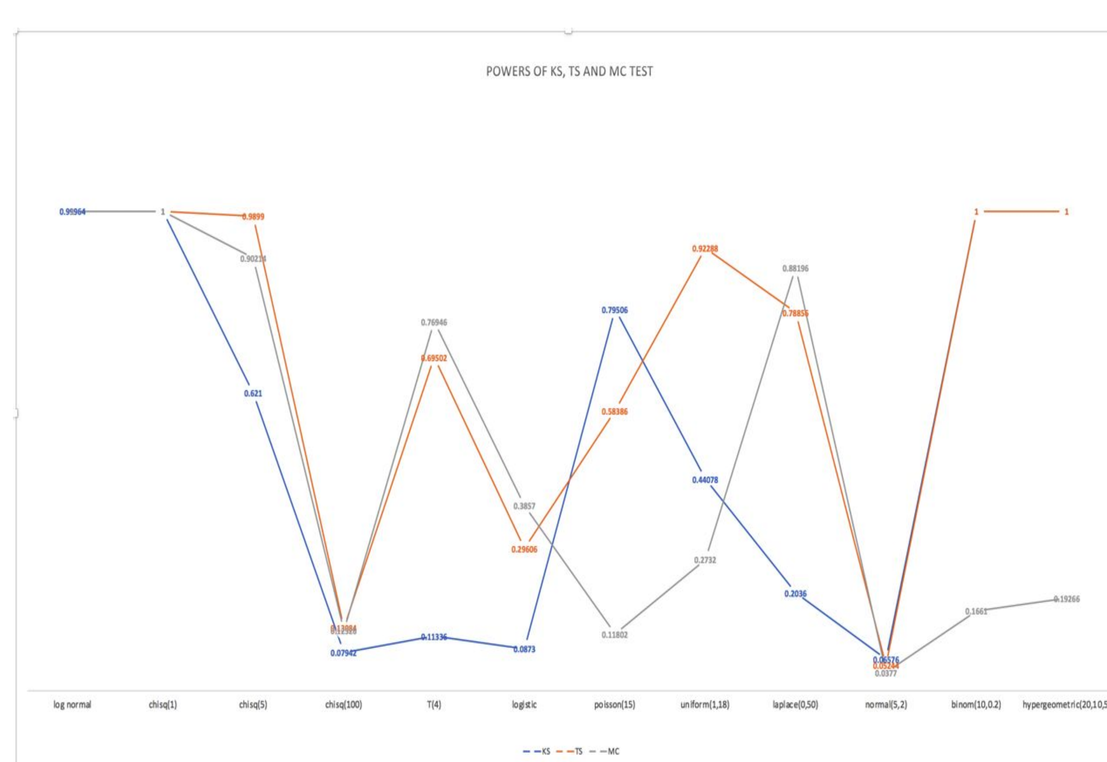


Figure 1

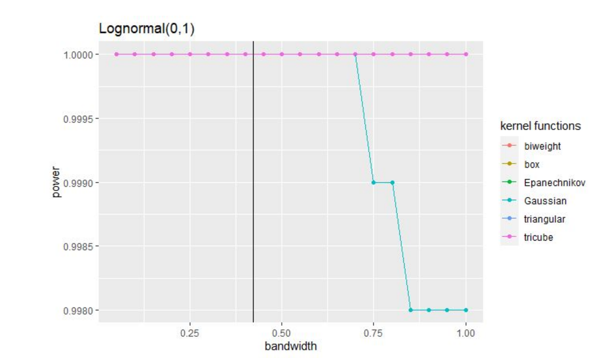


Figure 2 - Log Normal

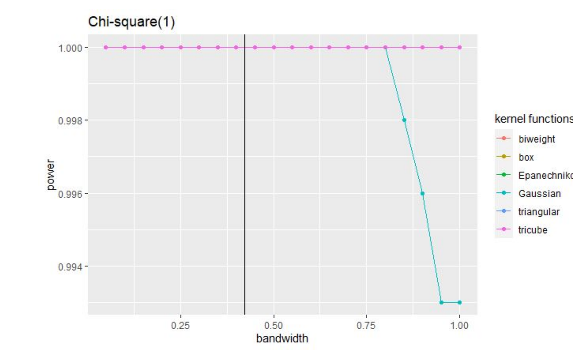


Figure 3 - Chi-square(df = 1)

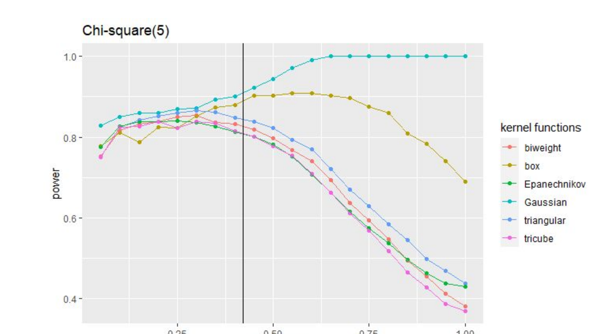


Figure 4 - Chi-square(df = 5)

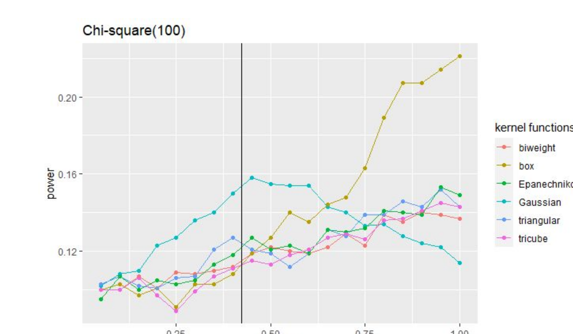


Figure 5 - Chi-square(df = 100)

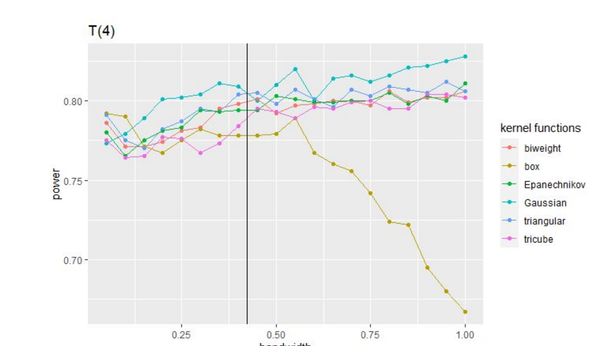


Figure 6 - T(df = 4)

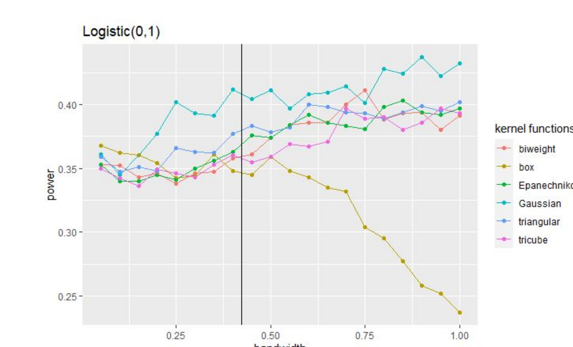


Figure 7 - Logistic

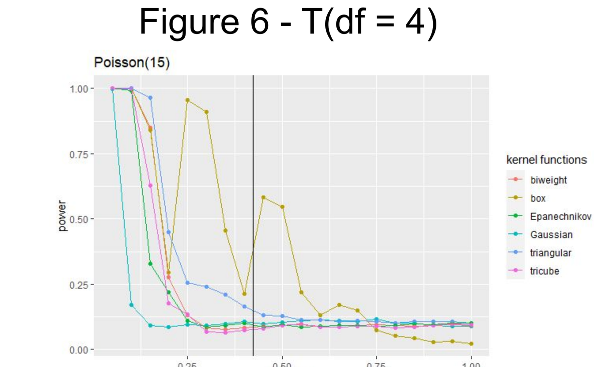


Figure 8 - Poisson(lambda = 15)

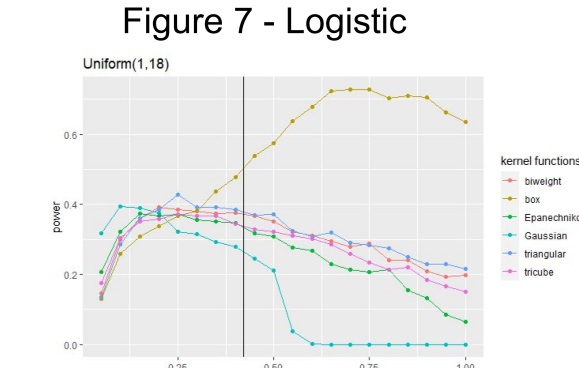


Figure 9 - Uniform(1, 18)

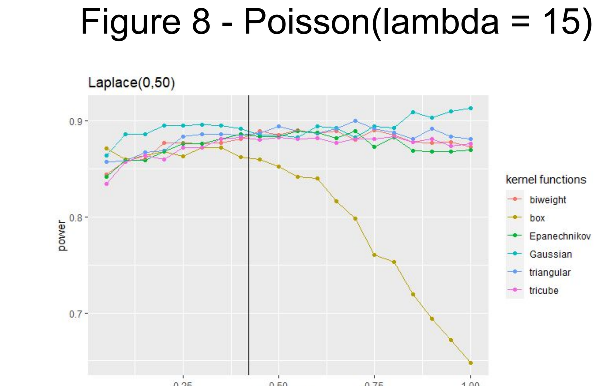


Figure 10 - Laplace(mu = 0, b = 50)

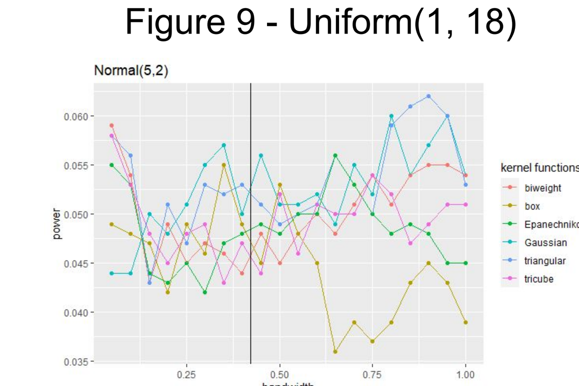


Figure 11 - Normal(mean = 5, sd = 2)

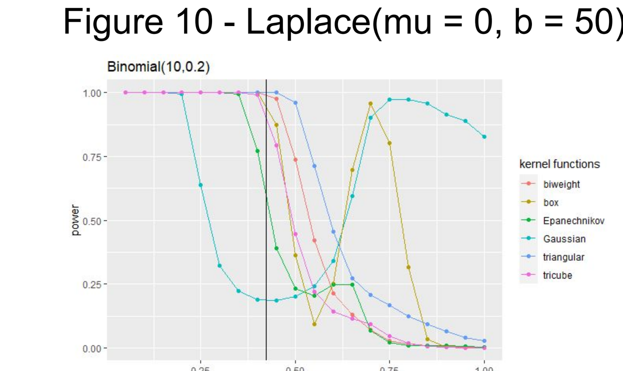


Figure 12 - Binomial(n = 10, p = 0.2)

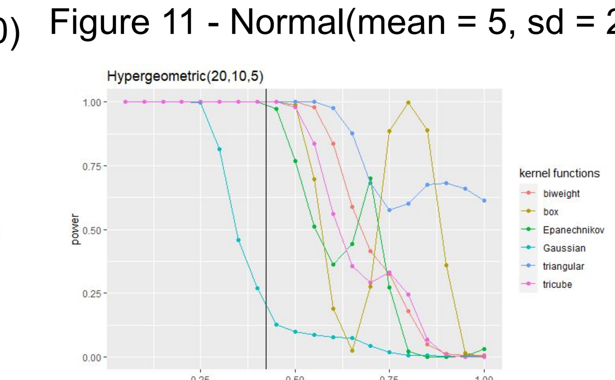


Figure 13 - Hypergeometric (m = 20, n = 10, k = 5)

From Figure 2 to Figure 13, we show powers for these 12 distributions with different choices of bandwidth and kernel functions. The bandwidths shown in the plot range from 0 to 1, but the results also include bandwidths up to 5. The black vertical line indicates the normal scale bandwidth. According to the plots, all kernel functions, except the box kernel, lead to similar powers for most distributions. The box kernel generally seems to be more sensitive to bandwidth change than the other kernels. The choice of normal scale bandwidth seems to be too large for discrete data such as Poisson(15), binomial(10,0.2), hypergeometric(20,10,5) as the maximum power is achieved at a smaller bandwidth.

DISCUSSION

We observed that most kernel functions have a similar trend of change in power as the bandwidth h increases for the MC test. Counterintuitively, a very high bandwidth h will cause the MC test power to eventually become 0 for some distribution. This may be explained by the fact that both the observed statistic and the simulated statistic use the same h, thus inflating both test statistics when h gets large. In fact, this also suggests that the MC test requires a good estimated density for the simulated normal sample to be comparable to test normality. Without a good estimated density, the Monte-Carlo statistics even for the normal sample could be large when we expect it to be small.

The paper considered the use of the normal scale bandwidth h to be optimal with the assumption that the underlying distribution of the data is not obviously nonnormal. This might explain why the MC test lacks power in the uniform case. Further research could attempt to use k-fold cross-validation to find the optimal bandwidth h at least for the test data.

In general, we obtained similar results for the powers of the two described graphical methods, the TS test, and the MC test when tested on the continuous distributions. Both methods demonstrated high power when testing fat-tail distributions such as t(df = 4) compared to the KS test. However, the MC test fell short when tested on uniform data and discrete data. We discussed the reason for the uniform case above, and, for the discrete data, we hypothesize that this phenomenon may be due to the unsuitable bandwidth choice since KDE is intended to estimate the probability density function from a continuous variable. As the figure 8,12,13 show, the powers drop quickly as the bandwidth gets over about 0.25.

It should also be noted that the papers, which we had examined, did not supply the exact algorithms used to compute the powers. Hence, it should be expected that there are slight variations in our testing algorithms although they closely follow the procedures articulated by the authors of the papers.

REFERENCES

- M. C. Jones, & Martin L. Hazelton. (2004). Hazelton, M. L. (2003), "A Graphical Tool for Assessing Normality," The American Statistician," 57, 285-288: Comment by Jones and Reply. *The American Statistician, 58*(2), 176-177.
- Aldor-Noiman, S. et al. (2013). The Power to See: A New Graphical Test of Normality. The American Statistician. 67:4.

ACKNOWLEDGEMENT

We are very much thankful to Steven Wilkins-Reeves and Sarah Teichman of the Department of Statistics for their valuable guidance and patient assistance especially under the unfortunate circumstances due to the pandemic.

We also want to acknowledge Professor Ranjini Grove of the Department of Statistics for suggesting to us this opportunity and providing us with further constructive criticisms.