Survey of Misuses of the KS-Test

Anthony Zeimbekakis and Jun Yan
anthony.zeimbekakis@uconn.edu
Department of Statistics, University of Connecticut

November 8, 2021

Abstract

The Kolmogorov–Smirnov (KS) test is one the most popular goodness-of-fit tests for comparing a sample with a hypothesized parametric distribution. Nevertheless, it has often been misused. The standard one-sample KS test applies to independent, continuous data with a hypothesized distribution that is completely specified. It is not uncommon, however, to see in the literature that it was applied to dependent, discrete, or rounded data, with hypothesized distributions containing estimated parameters. For example, it has been "discovered" multiple times that the test is too conservative when the parameters are estimated [e.g. Steinskog et al. (2007)]. This paper aims to survey the misuses of the KS test, demonstrate their consequences through simulation, and provide remedies as needed.

1 Introduction

The Kolmogorov-Smirnov (K-S) statistic is one of the most popular goodness-of-fit tests for comparing a sample with a hypothesized parametric distribution. Given a sample of n observations, let $S_n(x)$ be the empirical cumulative distribution and $F(x)$ be the population cumulative distribution. The K-S statistic is defined by:

$$d = max(|F(x) - S_n(x)|)$$

If the value of d exceeds the critical value given by the table, the null hypothesis that the observations are from a specified distribution is rejected.

However, the test is often misused. The standard one-sample K-S test applies to independent, continuous data with a hypothesized distribution that is completely satisfied. Often in literature, the test is applied to dependent, discrete, or rounded data, with hypothesized distributions containing estimated parameters. As shown by Steinskog et al. (2007) and later in this paper, the test is too conservative when the parameters are estimated. Throughout this paper, the cumulative distribution $F(x)$ is standard normal.

To begin, a simple demonstration is performed to show the impact of sample size and replicate tests. For n in $[10, 100, 1000]$, a random standard normal distributions were generated and K-S tests were performed. With the hypothesized distribution of standard normal, the histograms in Figure 1 show that the sample size has little effect on the distribution of P-values.

The standard K-S test also shows misuse when the data has been rounded. For this simulation, standard normal distributions are rounded up to 4 decimal places. The histograms in Figure 2 display the distribution of P-values for the tests.

As the level of rounding increases, the resulting P-values stray from the expected uniform distribution.

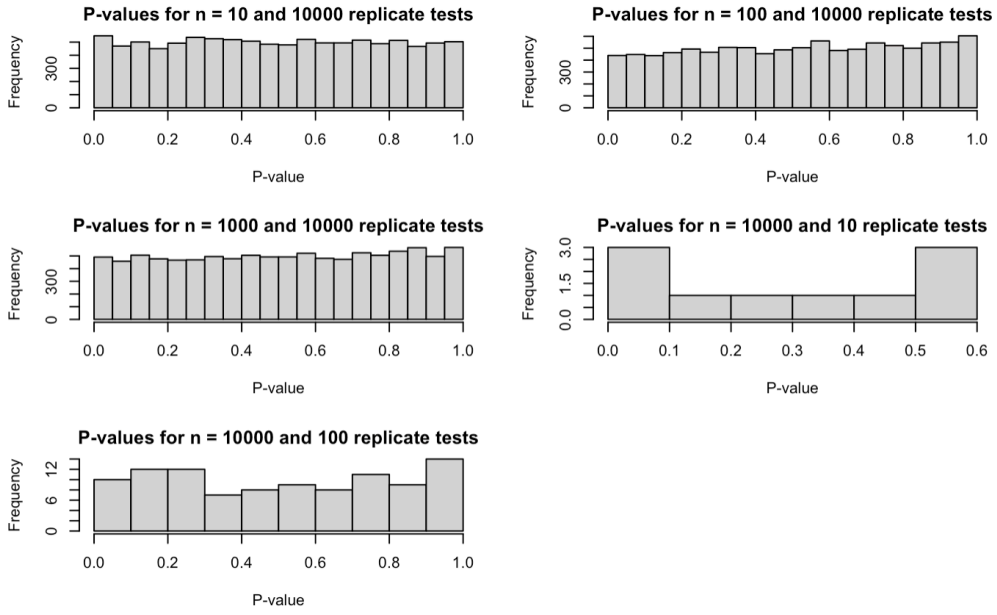

Figure 1: Histograms of impact of sample size and number of tests on P-value.

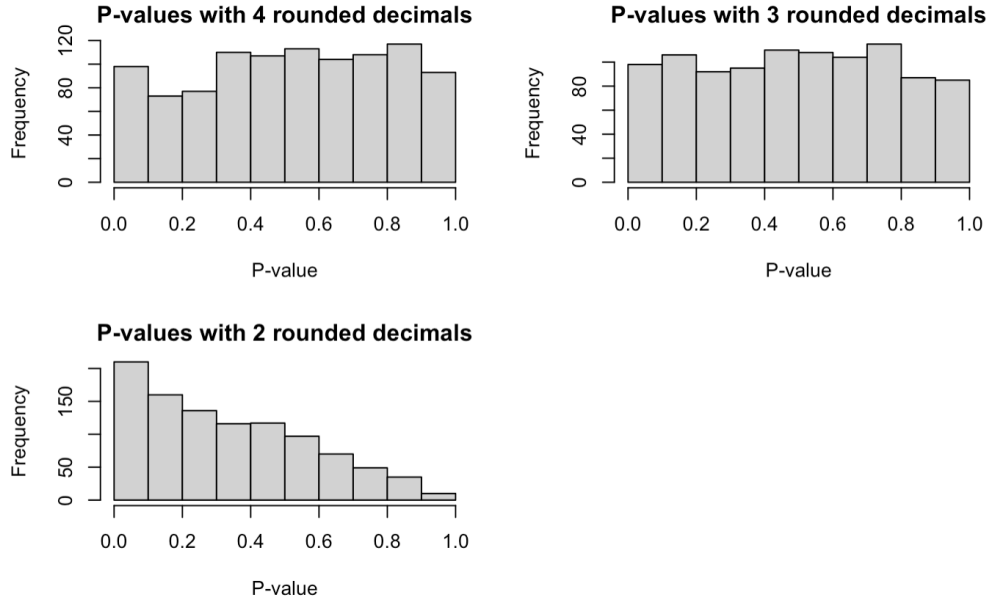

Figure 2: Histograms of P-values with rounded data

2 Problem Under Serial Dependence

The K-S test shows significant issues in the case of serial dependence.

	Tests, Samples	PowerLevel
1	10000, 10	0.95
2	10000, 100	0.96
3	10000, 1000	0.95
4	10, 10000	0.90
5	100, 10000	0.92
6	1000, 10000	0.94

3 Simulation

Simulation here.

4 Real Data Analysis

Real Data Analysis here.

5 Conclusion

Conclusion here.

Adding these to see the full bibliography:

Steinskog et al. (2007) Weiss (1978) Massey (1951) Lilliefors (1967)

References

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association 62*(318), 399–402.

Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association 46*(253), 68–78.

Steinskog, D. J., D. B. Tjøstheim, and N. G. Kvamstø (2007). A cautionary note on the use of the kolmogorov-smirnov test for normality. *Monthly Weather Review 135*(3), 1151–1157.

Weiss, M. S. (1978). Modification of the Kolmogorov-Smirnov statistic for use with correlated data. *Journal of the American Statistical Association 73*(364), 872–875.