Survey of Misuses of the Kolmogorov-Smirnov Test

Anthony Zeimbekakis[1], Jun Yan[1]

[1] Department of Statistics, University of Connecticut

1. Introduction

The Kolmogorov-Smirnov (K-S) test is one 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:

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

- $H_0 =$ The data follows a specified distribution $F(x)$
- $H_A =$ The data does not follow a specified distribution $F(x)$
- If the value of d exceeds the test's corresponding critical value, the null hypothesis is rejected

The standard one-sample KS test applies to **independent, continuous data with a hypothesized distribution that is completely specified**, though it's often used incorrectly. We aim to survey misuses of the K-S test, demonstrate their consequences through simulation, and provide remedies as needed.

- All simulations unless specified are performed with the standard normal distribution $N(0, 1)$

2. Fitted Parameters

2.1. Assumption:

- Hypothesized distribution $F(x)$ is completely specified

2.2. Misuse:

- Hypothesized distribution $F(x)$ is fitted with parameters from the sample

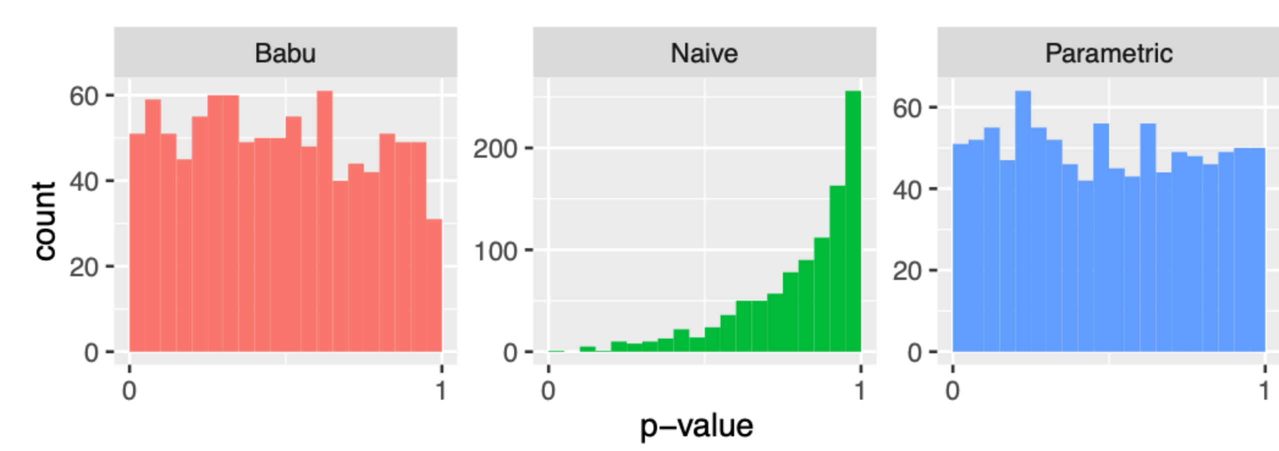
Ex: For the normal distribution, the test is performed where $F(x) \sim N(\bar{X}, s^2)$ (Lilliefors 1967).

2.3. Remedy:

A remedy for this issue is to use bootstrap, detailed by (Genest and Rémillard 2008) and (Babu and Rao 2004):

- d_0 is the observed K-S statistic
- Let $X_1^*, \ldots, X_n^*$ be i.i.d random variables from $\hat{S}_n$, an estimator of the empirical distribution function S based on the sample $X_1, \ldots, X_n$
- Let $\hat{\theta}_n^* = \theta_n(X_1^*, \ldots, X_n^*)$. θ represent the sample mean and variance of the bootstrap sample $X_1^*, \ldots, X_n^*$
- New K-S statistic: $d = max(|F(x) - S_n(x; \hat{\theta}_n^*)|)$
- Count the number of bootstrap K-S statistics greater than or equal to d_0, and divide by the number of bootstrap samples to calculate p-value

2.4. Results:



- Naive plot was performed with $F(x) \sim N(\bar{X}, s^2)$
- Parametric plot was performed with $\hat{S}_n = F(.; \hat{\theta}_n)$
- Babu plot was performed with $\hat{S}_n = S_n$ and a correction for bias from (Babu and Rao 2004)

With replicate tests, the distribution of p-values is $U(0, 1)$. However since the assumption of a specified distribution is violated, the naive plot too often does not reject the null hypothesis. Bootstrap restores the power of the test.

3. Temporal Dependence

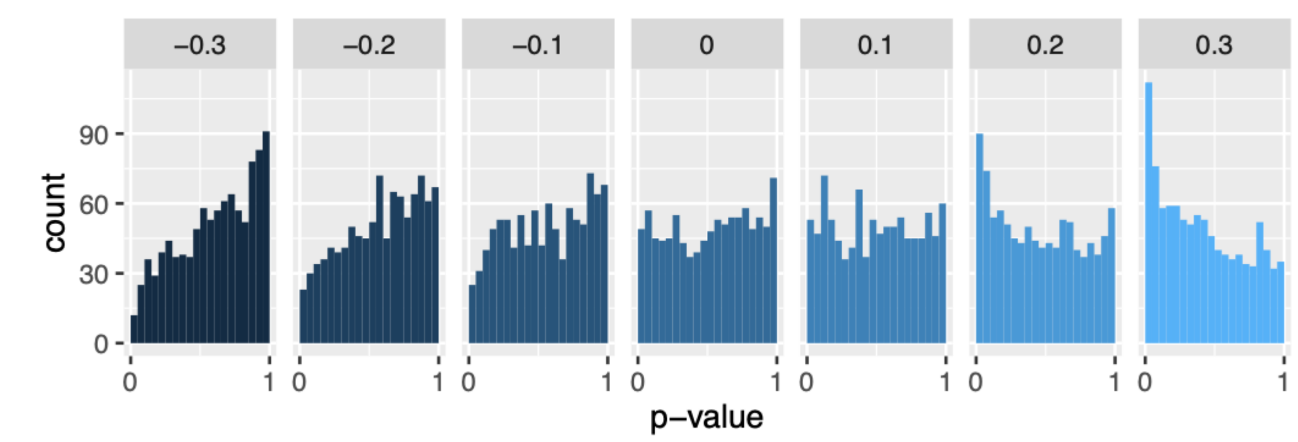
Assumption:

- The empirical distribution $S_n(x)$ contains independent data

Misues:

- The empirical distribution $S_n(x)$ contains dependent data

To demonstrate this, data is simulated from an $AR(1)$ process with $AR = (-3, 3)$. There is an inverse relationship between negative and positive AR values (Durilleul and Legendre 1992), and the distributions are not uniform when temporal dependence is present.

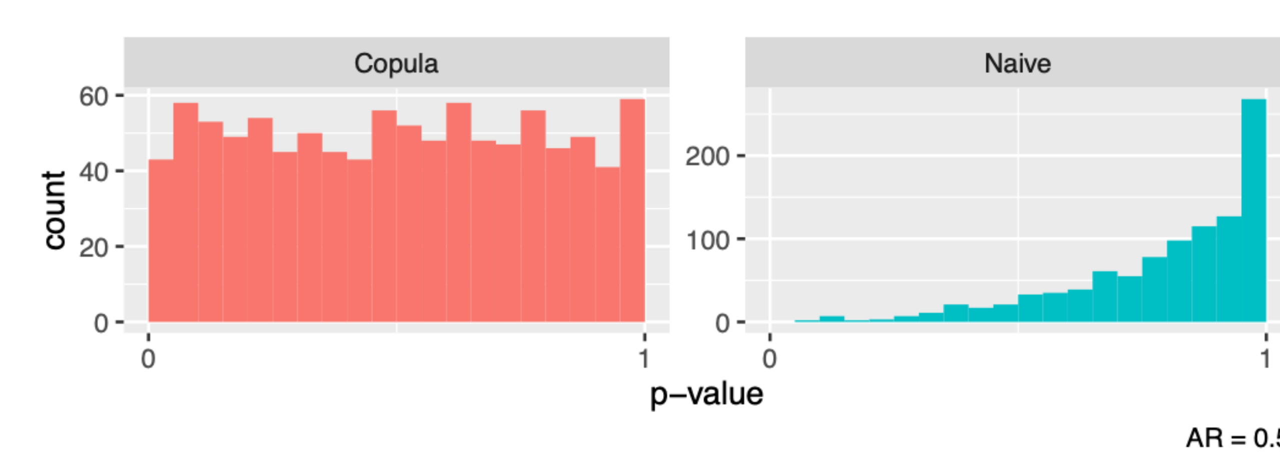


Remedy:

- Use copulas combined with bootstrap, for two reasons

1. Ensure that the marginal distribution of our bootstrap data is the same as the fitted distribution
2. Introduce temporal dependence at the same level as the fitted distribution The procedure used is the same as bootstrap, except the re-sampling comes from correlated data generated by copulas

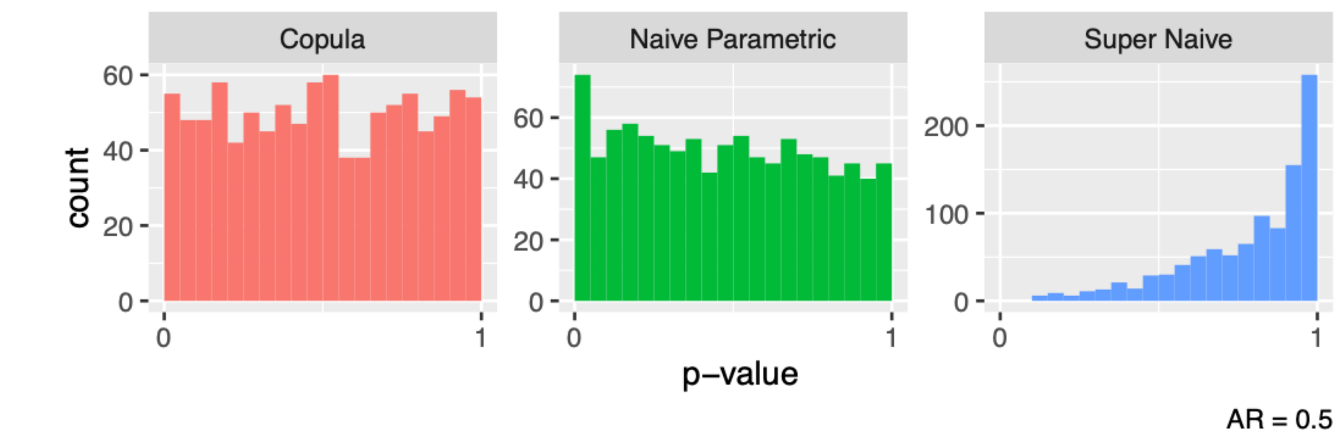
Results:



- The copula remedy restores the $U(0, 1)$ nature of the p-values

4. Fitted Parameters + Temporal Dependence

If both assumptions are violated, the copula fix works as well.



- First plot shows the distribution of p-values when performing the copula remedy
- Second plot shows the distribution of p-values when performing basic parametric bootstrap, which naively ignores dependence
- Third plot shows the violation of both assumptions, where fitted parameters are used and dependence is not considered

5. Conclusion

- When the K-S test is used with violated assumptions, it no longer holds it's power and corrections must be made
- In the case of fitted parameters, parametric and non-parametric bootstrap are recommended solutions
- When there is dependence in the data, performing bootstrap with copulas is recommended.
- When both assumptions are violated, meaning there is dependent data and parameters must be estimated, perform bootstrap with copulas
- If the guess isn't far from the true distribution it may still be reasonably accurate for other distributions.

6. References

Babu, Gutti Jogesh, and C. R. Rao. 2004. "Goodness-of-Fit Tests When Parameters Are Estimated." *Sankhya: The Indian Journal of Statistics* 66: 63–74.

Durilleul, Pierre, and Pierre Legendre. 1992. "Lack of Robustness in Two Tests of Normality Against Autocorrelation in Sample Data." *Journal of Statistical Computation and Simulation* 42 (1-2): 79–91.

Genest, Christian, and Bruno Rémillard. 2008. "Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models." *Annales de l'Institut Henri Poincaré, Probabilités Et Statistiques* 44 (6): 1096–1127.

Lilliefors, Hubert W. 1967. "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown." *Journal of the American Statistical Association* 62 (318): 399–402.