A Cheap Trick to Improve the Power of a Conservative Hypothesis Test

Thomas J. Fisher
Michael W. Robbins
fishert4@miamioh.edu

August 6, 2020

A Cheap Trick to Improve the Power of a Conservative Hypothesis Test

DISCLAIMER!

WARNING!!!

Presentation contains hypothesis testing and p-values

DISCLAIMER!

WARNING!!!

Presentation contains hypothesis testing and p-values

In no way do the authors of this article advocate for the misuse of p-values and/or hypothesis testing. The results presented herein are primarily motivated by the pedagogical findings.

Motivation I

Some background motivation

- In multivariate analysis, the likelihood ratio test for a covariance matrix is based on the following, see Anderson [1],

$$\log|\hat{\boldsymbol{\Sigma}}|$$

where $\hat{\boldsymbol{\Sigma}}$ is the sample covariance matrix and $|\cdot|$ is the determinant.

- In time series, the goodness-of-fit test of Peña & Rodríguez [2] is

$$\frac{-3n}{2m+1}\log|\hat{\boldsymbol{R}}|$$

where $\hat{\boldsymbol{R}}$ is an $m \times m$ matrix of autocorrelations: $\hat{\boldsymbol{R}}_{i,j} = \hat{\rho}(|i-j|)$.

Motivation II

More recent background:

- Mahdi & McLeod [3] and Robbins & Fisher [4] extend ideas from Peña & Rodríguez [2] to multivariate time series.
- In Fisher & Robbins [5], we measure the lag k autocorrelation matrix in a multivariate time series with:

$$-\log|\boldsymbol{R}_k| \quad \text{where} \quad \boldsymbol{R}_k = \begin{bmatrix} \boldsymbol{I}_d & \hat{\boldsymbol{R}}_k \\ \hat{\boldsymbol{R}}_k^T & \boldsymbol{I}_d \end{bmatrix}$$

for a d-dimensional time series where $\hat{\boldsymbol{R}}_k$ is the autocorrelation matrix at lag k and $\boldsymbol{I}_d$ is a $d \times d$ identity.

Motivation III

The *math* in all these time series applications involves a bunch of linear algebra (Kronecker products, eigenvalues) but ultimately the asymptotic results depend on a few fundamental ideas:

- The determinant value, $|\cdot|$, in all these results is a value in $(0, 1)$.
- Some 1^{st} order Taylor expansions.
- Fairly basic limiting arguments.

Main Result - Framework

Let $\mathbf{X}_n = \{X_1, X_2, \ldots, X_n\}$ be a sample and $T_n = T_n(\mathbf{X}_n)$ denote a statistic for testing the competing hypotheses H_0 and H_1.

Assume the following:

(a) T_n is strictly non-negative: $P(T_n \geq 0) = 1$,

(b) When H_0 is true: $T_n = \mathcal{O}_p(1)$ (likewise, T_n has a limit distribution),

(c) When H_1 is true: $T_n = \mathcal{O}_p(n^{\kappa})$ for some $\kappa > 0$;
that is, T_n diverges to $+\infty$ at rate n^{κ}.

Main Result

For a given statistic T_n satisfying the stated assumptions, consider the modified test statistic:

$$T_n^* = -n^\kappa \log(1 - T_n/n^\kappa). \tag{1}$$

Main Result - Math results

Theorem

When H_0 is true, $T_n^* \xrightarrow{p} T_n$ as $n \to \infty$; moreover, T_n^* and T_n share the same asymptotic distribution.

Main Result - Math results

Theorem

When H_0 is true, $T_n^* \xrightarrow{p} T_n$ as $n \to \infty$; moreover, T_n^* and T_n share the same asymptotic distribution.

Theorem

When H_1 is true, T_n^* diverges from T_n and will be more powerful than T_n if decisions are based off the same critical values.

Proof of First Theorem: H_0 Theorem

Proof.

Consider

$$\begin{aligned} T_n^* &= -n^\kappa \log(1 - T_n/n^\kappa) \\ &= n^\kappa \left[\frac{T_n}{n^\kappa} + \frac{1}{2}\left(\frac{T_n}{n^\kappa}\right)^2 + \frac{1}{3}\left(\frac{T_n}{n^\kappa}\right)^3 + \dots \right] \\ &= T_n + \frac{T_n^2}{2n^\kappa} + \frac{T_n^3}{3n^{2\kappa}} + \dots \\ &= T_n + A_n. \end{aligned} \tag{2}$$

When H_0 is true, $A_n = \mathcal{O}_p(n^{-\kappa})$ by assumptions (a) and (b).

Whence $T_n^* \xrightarrow{p} T_n$, and T_n^* shares the same asymptotic distribution as T_n. □

Proof of Second Theorem: H_A Theorem

Proof.

Recall A_n in (2),

$$A_n = \frac{T_n^2}{2n^\kappa} + \frac{T_n^3}{3n^{2\kappa}} + \dots$$

If H_0 is true, $A_n = \mathcal{O}_p(n^{-\kappa})$ (or $A_n \to 0$ for all practical purposes).

If H_1 is true, $0 \leq A_n = \mathcal{O}_p(n^\kappa)$ by assumption (c) (or $A_n \to \infty$ for all practical purposes).

It follows that for all c, $P(T_n^* > c) \geq P(T_n > c)$ and hence T_n^* can offer more power than T_n. □

Caveats

Some (obvious) cautions:

- If T_n has an exact distribution (i.e., F-stat satisfies our assumptions but is $F(\nu_1, \nu_2)$ distributed) – don't use T_n^*.
- If the asymptotic distribution of T_n produces type I error at the nominal level, or is liberal, using T_n^* will create a liberal statistic or amplify the poor type I error performance.

Caveats

Some (obvious) cautions:

- If T_n has an exact distribution (i.e., F-stat satisfies our assumptions but is $F(\nu_1, \nu_2)$ distributed) – don't use T_n^*.
- If the asymptotic distribution of T_n produces type I error at the nominal level, or is liberal, using T_n^* will create a liberal statistic or amplify the poor type I error performance.

But if T_n is conservative in practice:

- T_n^* may result in type I errors closer to the nominal level.
- T_n^* will provide more detection power than T_n.
- T_n^* will diverge from T_n.

Simulation Setup I

Consider testing for significant correlation between two sets of observations, x_i and y_i using Pearson correlation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Covered in nearly every introductory statistics course.

Simulation Setup II

If (x_i, y_i) are bivariate Normal, there is the well known result

$$F = r^2 \frac{n-2}{1-r^2} \sim F(\nu_1 = 1, \nu_2 = n-2)$$

and note $F \longrightarrow \chi_1^2$ as $n \rightarrow \infty$.

Alternatively we could use the simpler statistic

$$T = nr^2 \sim \chi_1^2, \text{ as } n \rightarrow \infty.$$

T relates to time series goodness-of-fit test an is known to be conservative since it is negatively biased compared to it asymptotic distribution [see Box & Pierce 6, for further details].

Simulation Setup III

Data is generated as:

- x_i are uniform over the interval $(1, 20)$.
- $y_i = 5 + \delta x_i + 3\varepsilon_i$ where,
  - $\varepsilon \sim t(\nu = 3)$,
  - δ acts as a perturbation parameter.

The underlying stochastic distribution (ε terms) are leptokurtic. It is known that normal theory results tend to be conservative [see 7] in such situations, but the F test can be justified asymptotically for non-normal data [8].

Simulation Setup IV

Using our *trick*, one could also consider the statistics:

$$F^* = -n\log(1 - F/n)$$

and

$$T^* = -n\log(1 - T/n) = -n\log\left(1 - r^2\right).$$

For comparison, we also include a bootstrapped version of T denoted as T_B where the y_i terms are resampled with replacement.

Simulation - Type I error rates

Table 1: Rate of rejections at $\alpha = 1\%$, out of 10,000 replications, of F statistic based on an $F(1, n-2)$ distribution, the χ^2-based test T, the transformed T^* and bootstrapped T_B (based on 1,000 resamples) under the null hypothesis at seven sample sizes n.

n	25	30	35	40	45	50	100
F	0.8	0.8	0.8	0.9	1.0	0.9	1.1
T	0.7	0.6	0.7	0.8	0.9	0.8	1.0
T^*	1.1	1.2	1.1	1.2	1.1	1.0	1.2
T_B	0.9	1.1	1.0	1.1	1.2	1.0	1.2

Simulation - Statistical Power

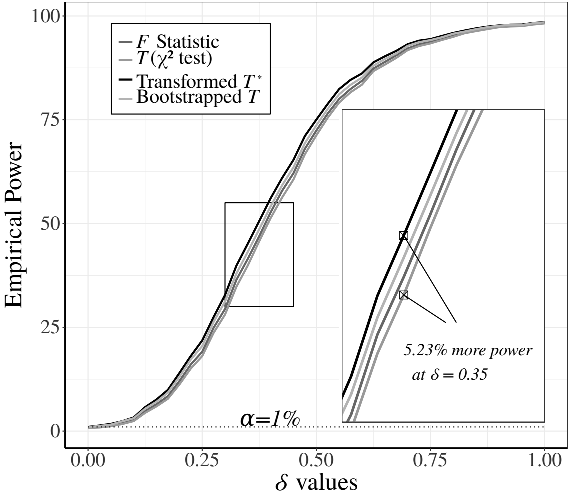

Figure 2: Power of F, T, T^* and T_B at $\alpha = 1\%$ under the alternative hypothesis as a function of the perturbation parameter δ for $n = 35$.

Simulation - Divergence of Statistics

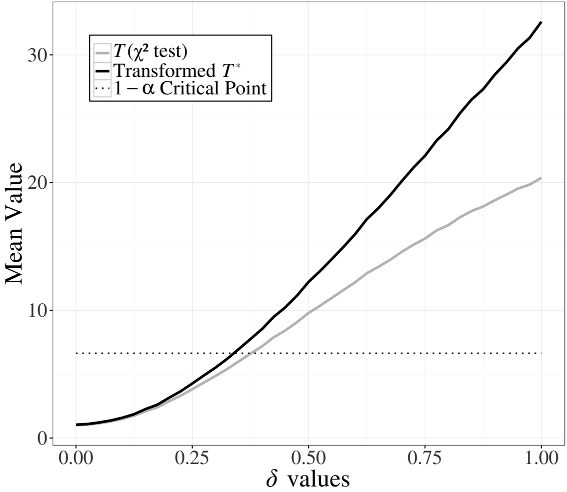

Figure 3: Mean value of T and T^* under H_1 under the alternative hypothesis as a function of perturbation parameter δ for $n = 35$.

Additional results I

How to determine κ?

- The two theorems hold for any $\kappa > 0$.
- Value of T_n^* increases as κ approaches 0.
- Sensitivity study on κ is provided in the article.
- We found setting κ equal to the H_1 rate of divergence yields a test that performs well under both hypotheses.

Comparison to other correction methods

- Compared to multiplicative corrected statistics $T_n^\dagger = b_n T_n$.

Additional results II

Discussion on Exact level and UMP tests

- Clearly do not want to use the transformation for exact level tests.
- The transformation cannot improve power on a UMP test.
- Details provided on how the transformation can correct conservative test.

Connections to education

- Results rest on Taylor expansions and basic convergence results.
  - Two topics I found students struggle with historically.
- Mechanism to introduce what it means to be a UMP test, the limits of hypothesis testing and using asymptotic results.
- Article could be used in an undergraduate Mathematical Statistics class.

Additional Implementations

- Correlation Example presented here.
- In main article:
  - Change point testing using a CUSUM statistic.
  - Wald statistic in Logistic Regression.
    (including an application on the Challenger O-ring data)
- Additional examples in supplemental code:
  - A likelihood ratio test on Gamma distributed data.
  - A t-test and ANOVA F-test.
  - Kolmogorov-Smirnov test.

Thanks for the memories JSM!

Thanks for the memories JSM!

Unapologetic Self-references

Presentation is based on the article

- T. J. Fisher and M. W. Robbins, “A cheap trick to improve the power of a conservative hypothesis test,” *The American Statistician*, vol. 73, no. 3, pp. 232–242, 2019. DOI: 10.1080/00031305.2017.1395364

Slides and other work can be found

- https://tjfisher19.github.io/

Bibliography I

[1] T. W. Anderson, *An introduction to multivariate statistical analysis*, Third, ser. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience [John Wiley & Sons], 2003, pp. xx+721, ISBN: 0-471-36091-0.

[2] D. Peña and J. Rodrıguez, "The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series," *J. Statist. Plann. Inference*, vol. 136, no. 8, pp. 2706–2718, 2006, ISSN: 0378-3758. DOI: 10.1016/j.jspi.2004.10.026.

[3] E. Mahdi and I. A. McLeod, "Improved multivariate portmanteau test," *J. Time Ser. Anal.*, vol. 33, no. 2, pp. 211–222, 2012, ISSN: 1467-9892. DOI: 10.1111/j.1467-9892.2011.00752.x.

[4] M. W. Robbins and T. J. Fisher, "Cross-correlation matrices for tests of independence and causality between two multivariate time series," *J. Bus. Econom. Statist.*, vol. 33, no. 4, pp. 459–473, 2015. DOI: 10.1080/07350015.2014.962699.

Bibliography II

[5] T. J. Fisher and M. W. Robbins, "An improved measure for lack of fit in time series models," *Statist. Sinica*, vol. 10, no. 28, pp. 1285–1305, 2018. DOI: 10.5705/ss.202016.0286.

[6] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive integrated moving average time series models," *J. Amer. Statist. Assoc.*, vol. 65, pp. 1509–1526, 1970, ISSN: 0162-1459.

[7] E. S. Pearson, "The analysis of variance in cases of non-normal variation," *Biometrika*, vol. 23, no. 1/2, pp. 114–133, 1931, ISSN: 00063444.

[8] M. Kendall and A. Stuart, *The Advanced Theory of Statistics: Inference and relationship*, ser. Kendall's Advanced Theory of Statistics: Classical Inference and the Linear Model. C. Griffin, 1977, ISBN: 9780852642559.

[9] T. J. Fisher and M. W. Robbins, "A cheap trick to improve the power of a conservative hypothesis test," *The American Statistician*, vol. 73, no. 3, pp. 232–242, 2019. DOI: 10.1080/00031305.2017.1395364.