A Simulation Study for the Evaluation of a Novel Class of Statistical Tests

Nicholas Chandler, Samuel Burnett, Fiona Cleary, Kimihiro Noguchi

Department of Mathematics, Western Washington University, Bellingham, WA 98225, USA

Introduction

Background:

- Often, researchers are concerned with conducting statistical tests between two independent samples.
- We developed a new class of two-sample tests which test both the equality of means and stochastic equality simultaneously. We also conducted a Monte Carlo simulation study to compare the performance of our tests, one based on the asymptotic normality and the other based on MaxT bootstrap [3], to the traditional Brunner-Munzel test [1] and Welch's t-test [2].

MaxT Bootstrap:

- Our test relies on two statistics D_b and K_d. D_b is used for comparing the means, and K_d is used for assessing the stochastic equality. The joint test statistics (D_b, K_d) are asymptotically bivariate normal, implying that a small-sample performance can be improved by applying the MaxT bootstrap.
- In the MaxT bootstrap, we randomly sample observations from the underlying distribution of the data. Then, the joint test statistics are computed from the i-th resample, $i = 1, \ldots, 1000$. Then, $\#\{\max\{|D_b|, |K_d|\} \leq \max\{|D_b^{(i)}|, |K_d^{(i)}|\}\}/1000$, where $(D_b^{(i)}, K_d^{(i)})$ represents the joint test statistics from the i-th resample, gives the p-value.

Simulation Setup: We examined combinations of different sampling distribution parameters to investigate the behavior of each test.

- Sample Sizes: 10, 20, and 50.
- Distribution: Two-parameter exponential distribution.
- The location parameter difference $(a1)$: -0.5 to 0.5 with an increment of 0.1. The scale parameter is set to 1 for both samples.

Computation:

- With many combinations of parameters to test, the computational time was on the order of days. To mitigate this, we employed multi-threaded programming within the R code via the `foreach` and `doParallel` packages. We also distributed the work using the HTCondor computing cluster. We managed to reduce the time needed to less than a day.

Monte Carlo Simulation Results

The Tests

- Brunner-Munzel: A non-parametric test for stochastic equality.
- Welch's T: A parametric test for the equality of means.
- Normal: Our test which relies on the asymptotic bivariate normality of (D_b, K_d).
- Max T: Our test which relies on the Max T bootstrap.

Robustness and Power of the Test

Here, we examine Type 1 error rates and power curves to evaluate the performance of our tests against the traditional tests on the two-parameter exponential distribution.

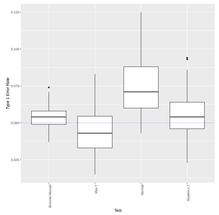

Figure 1: Box plots for Type 1 error rates of all the sample size combinations under the null hypothesis of the equality of distributions.

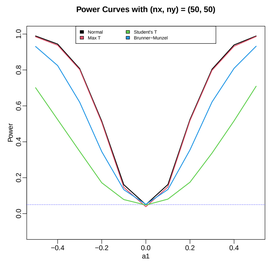

Figure 2: Power curves with balanced sample sizes $(50, 50)$ for the tests.

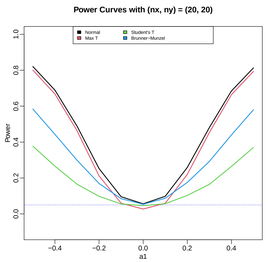

Figure 3: Power curves with balanced sample sizes $(20, 20)$ for the tests.

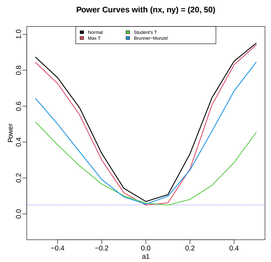

Figure 4: Power curves with unbalanced sample sizes $(20, 50)$ for the tests.

Interpretation of the Simulation Results

- The box plots imply that the Max T test is slightly conservative and the normal approximation is slightly liberal at the significance level $\alpha = 0.05$.
- The power curves for the balanced sample sizes $(20, 20)$ and $(50, 50)$ show that each curve attains a minimum at $a1 = 0$, where the null hypothesis is true.
- On the other hand, the power curves for the unbalanced sample sizes $(20, 50)$ show that some curves do not attains a minimum at $a1 = 0$. However, our tests still achieve a minimum at $a1 = 0$ as intended with only slight asymmetry.
- The Max T version of our test is slightly conservative (below 0.05) under the null hypothesis when $a1 = 0$, but is much more powerful than the traditional tests when $a1 \neq 0$.
- The normal version of our test is slightly liberal (above 0.05) under the null hypothesis when $a1 = 0$. Still, it is much more powerful than the traditional tests when $a1 \neq 0$.

Conclusions and Future Work

- Our tests perform better than the traditional tests when applied to data from the two-parameter exponential distribution. This is expected as our tests aim to simultaneously detect any difference in means and stochastic inequality where as the traditional tests are capable of doing just one of them.
- The next steps in our work concern evaluating the performance of our tests under different distributional assumptions for the sampled data as well as applying the tests to real world data to examine how they can be used.

References

1. Brunner, E. and Munzel, U. (2000). The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation, *Biometrical Journal* 42, 17-25.
2. Ruxton, G.D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test, *Behavioral Ecology*, 17(4), 688-690.
3. Westfall, P.H. (2011). On using the bootstrap for multiple comparisons, *Journal of Biopharmaceutical Statistics* 21(6), 1187-1205.