

Abstract

The report looks to compare the size and power of three statistical tests: t test, Mann U Whitney test and a parametric- Monte Carlo test. For the purpose of this report two equal samples from either a normal or gamma distribution were used, with code implemented on R. And the results, show for increasing size in samples, the power increases for all three tests, regardless of distribution of the sample data. Furthermore, as expected the non-parametric: Mann Whitney U test performs better irrespective of distributions compared to the other two tests.

Contents

Abstract.....	1
Introduction	3
Method	3
Investigating size:.....	3
T-test:	3
Mann-Whitney U test	3
Monte- Carlo test	4
Investigating power:	4
Generating data:	4
Results:.....	5
Size:	5
Power:	6
Investigating Power under assumptions.....	6
Investigating Power under assumption violations:.....	7
Conclusion and Further Work	8

Introduction

Hypothesis testing forms the basis of all statistical tests. This report looks, to investigate how size and power varies with sample size, effect size and the distribution of the sample data.

Method

Investigating size:

In order to investigate size, a significance level of 0.05 is set two samples of data are generated from two normal distributions, with equal means (0) and standard deviation (1), the p value of each test is calculated for n such generations:

T-test:

For t-test: the test with `var.equal="T"` is used, i.e. equal variances are assumed.

The Null Hypothesis being setup as, is the difference in population means is 0 and the alternative being:

Mean of population 1 < mean of population 2.

And finally, the resulting p-value is extracted from the the output of the t-test function.

Size, as a result is incorrectly rejecting the Null hypothesis when True, and therefore, the proportion of rejections under the Null Hypothesis is calculated for N simulations

Mann-Whitney U test

Mann-Whitney U test looks it if the two samples come from the same population through their Medians. However, due to the symmetry of the two distributions being tested (normal, with equal means), the null hypothesis can be stated similar to the above i.e. the difference in population means is 0 or the two samples come from the same distribution (iid). And size, calculated similarly to the above.

Monte- Carlo test

The Monte Carlo Simulation, used here is a parametric one to look at estimating the means.

The Null Hypothesis as: difference in means of the population = 0

The Algorithm is implemented as follows:

- Obtain mean of sample 1 and sample 2, calculate the difference
- Calculated the Standard error for the two samples
- Generate from two samples of $\bar{X} \sim N(999, 0, SE1)$, $\bar{Y} \sim N(999, 0, SE2)$ and obtain the difference in means for each pair and continue as above.

However, the above assumes the data can be approximated using a Normal Distribution. In this particular scenario, we can use the Central Limit Theorem, to look at the means for large samples to follow a normal distribution. And hence, restricts this simulation to looking at large sample and the test statistic of means.

Investigating power:

In order, to calculate the power, the data generated is under the alternative hypothesis:

mean (population1) < mean(population2).

The resulting p values are calculated for the three tests and the proportion of correct rejections under the Null Hypothesis (equal means) are calculated for N simulations.

Power for each test is then investigated when the two sample distributions are from a normal and a gamma distribution.

Generating data:

For the purpose of this investigation, only two distributions have been looked at:

1: The Normal Distribution

2: The Gamma Distribution

The Normal Distribution being a distribution which satisfies all three test assumptions is used as a baseline. While, the Gamma distribution displays skewness with parameters that can be easily expressed in terms of mean and standard deviation. The power of the t-test and Monte-Carlo test are investigated under such violation of assumptions.

Results:

Size:

The figure's below, shows the results of 100 simulations for each n increasing

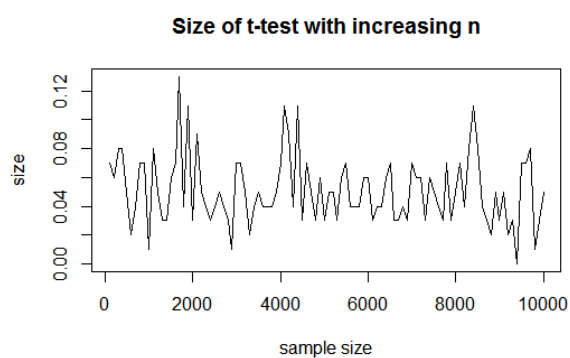


Figure 1. Size of t-test with increasing sample sizes

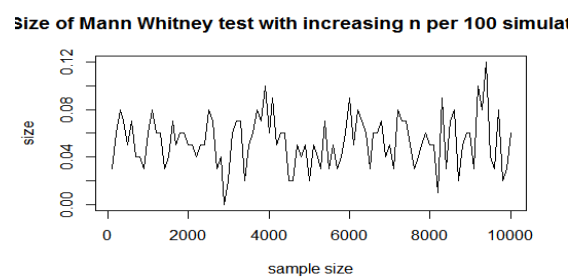


Figure 2: Size of Mann Whitney U test with increasing N

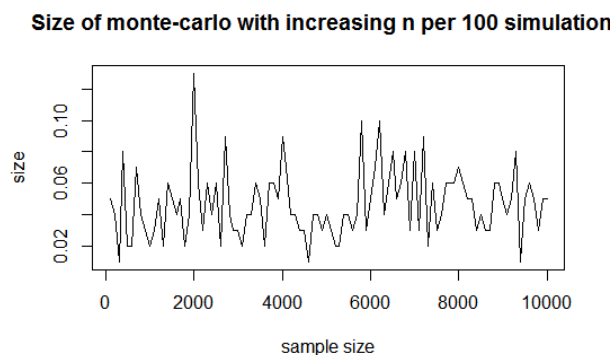


Figure 3. Size of Monte-Carlo test with increasing n

The above figures show, large variations in sizes for the samples due to the pseudo-randomness in the `rnorm` function, but with size approximately equal to the α .

Power:

Investigating Power under assumptions

To calculate the power, the data generated is under the normal distribution with a larger mean in the second sample. Variances are assumed equal set to 1, in all three test.

Power with increasing sample size (equal sample sizes):

The plot below shows the power for increasing sample sizes for the aforementioned tests:

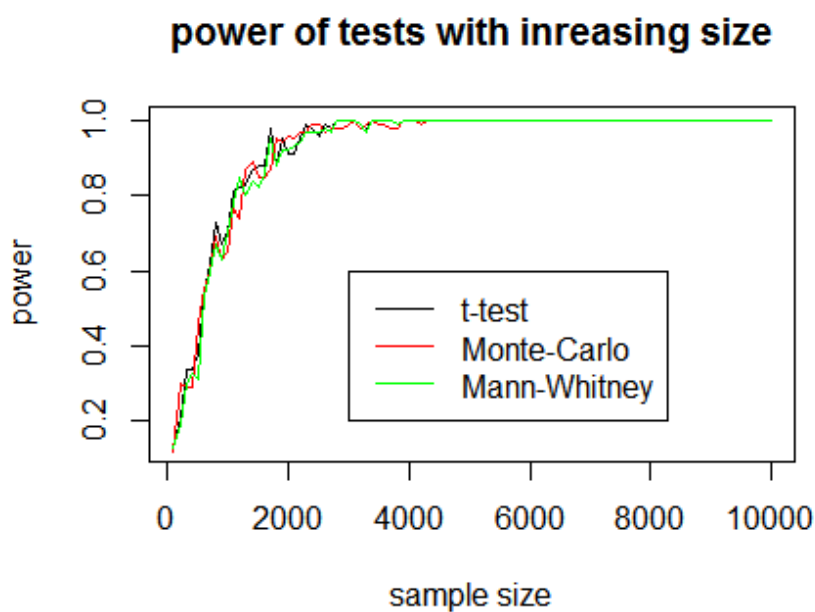


Figure 4: Power with increasing sample size

The data generated for the above calculation is with sample 1 of mean 0 and sample 2 of mean 0.1, with variance 1. At very large samples (>4000), the power approaches one for all three tests. Hence, power of a test increases with large sample sizes.

Power with increasing effect size:

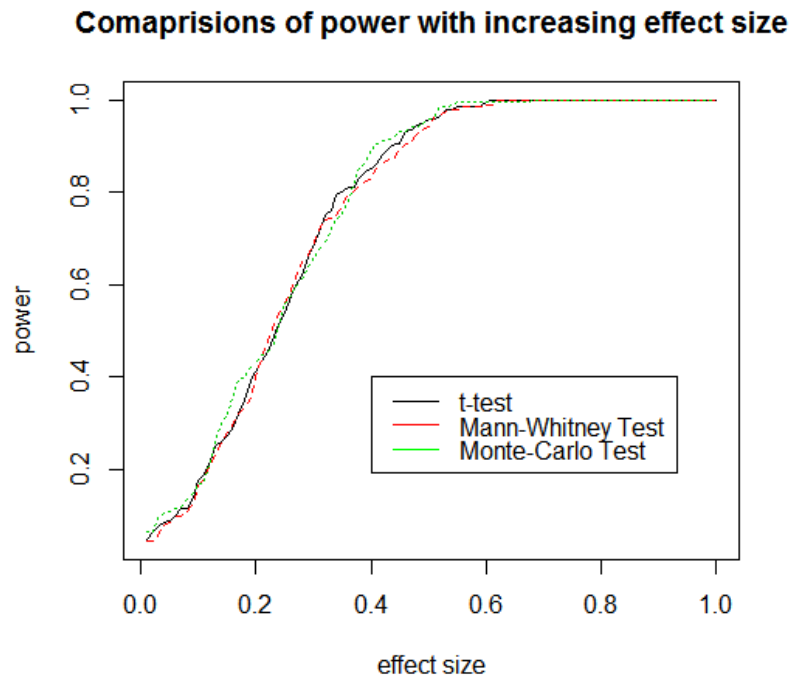


Figure 5: Power against effect size (sample size of 100 each)

The above figure, shows as effect size increases the power increases, with maximum power at effect size of 0.6. As the two samples move further apart, the test are able to detect the variation in means better, as the correct rejection rates increases ,thereby increasing the power of a test.

Investigating Power under assumption violations:

The assumptions of the t-test and parametric –Monte Carlo test (for small sample size) is violated, i.e. the data is normally distributed, by using sample data from a gamma distribution function. It also, should be noted, as the distribution is no longer symmetric, the alternative hypothesis of Mann Whitney –U -test changes to: the median of sample 1 < median of sample 2.

power with increasing sample size from a Gamma-distribution

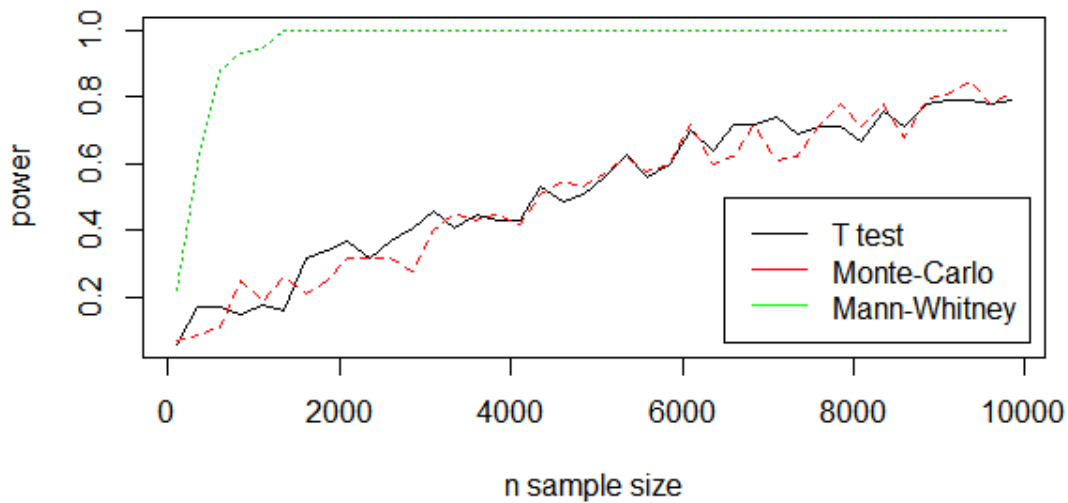


Figure 6: Power of tests with increasing sample size (Gamma distribution)

Fig 6: shows Mann-U Whitney performing much better than the Monte-Carlo and Mann-Whitney due to the Non-parametric assumptions. As expected, Monte-Carlo, performs very similar to the t-test due to the Normality assumption and power increasing linearly for large sample sizes.

Conclusion and Further Work

Due to time constraints, this report only attempts to look at power and size for equal samples for a two sample one sided test. It becomes clear, if the distribution is unknown, a non-parametric test like Mann-Whitney or Randomization test will be more appropriate in testing different hypothesis.

Furthermore, any future work should look to investigate how different variances can effect power and size.