

HypothesisTestforNonnormal Vignette

Qi Liu

December 16, 2011

Abstract

According to Central Limit theorem, the sampling distribution of mean converges to Normal with large sample size. T-test is widely used based on the assumption of normality of test statistics in the hypothesis test about one-sample mean or difference of two-sample mean. However, if the assumption of normality is violated, nonparametric tests, such as Wilcoxon rank sum test, is recommended. The package HypothesisTestforNonnormal provides functions to evaluate the assumption of the normality, compare t-test and Wilcoxon rank sum test on type I error and power for data from common nonnormal distributions, including Exponential, Lognormal, Gamma, Beta, Weibull, Cauchy, Bernoulli and Poisson distributions. It also contains a function to conduct rank transformed t-test for situation when assumptions of normality and homogeneity of variance are simultaneously violated.

1 Introduction

In the hypothesis test about one-sample mean or difference of two-sample mean, the exact distribution of test statistics is difficult to obtain if the underlying distribution is not normal. A common strategy is to make the assumption that the sample size is big enough so that Central Limit theorem holds. Then, the test statistic has an approximately normal distribution. However, the validity of this assumption depends on both specific distribution parameters and sample size. The functions "how_norm.distribution" (eg. how_norm.cauchy is for data from Cauchy distribution) simulate data with desired sample size and parameters, produce QQ plot of sample mean (or difference of two-sample mean) vs. Normal distribution, and therefore can be used to assess the assumption of normality of test statistics.

Nonparametric tests are recommended if the assumption of normality is violated. Wilcoxon rank sum test is usually considered as nonparametric

counterpart of t-test. The functions "t_vs_wilcox.distribution"(eg.t_vs_wilcox.exp is for Exponential distribution) simulated data with desired sample size and parameters, conduct t-test and Wilcoxon rank sum test, and calculate the simulated probability of type I error (if input parameters under null hypothesis) or power (if input parameters under alternative hypothesis) for both tests.

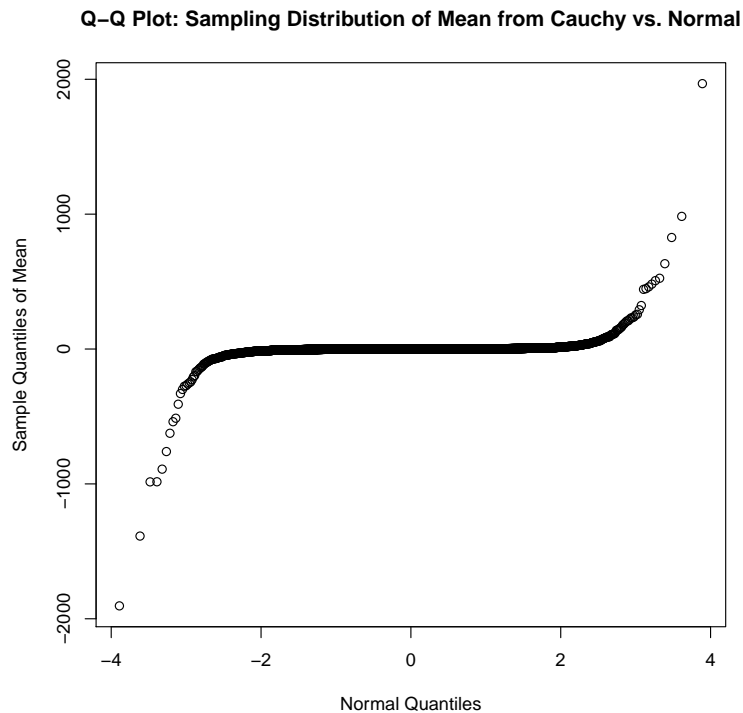
Nonparametric tests, however, can not handle the problem of unequal variances in combination with unequal sample size. Zimmerman and Zumbo (1993) proposed that when assumption of normality and homogeneity of variance were simultaneously violated, transforming the actual data to rank and then conduction unequal variance t-test (the Welch t-test) could solve the problem. The function "rank_t.test" is coded for this rank transformed t-test.

2 Examples

2.1 Assess the normality of sample mean

The functions "how_norm.distribution" simulate data with desired sample size and parameters, produce QQ plot of sample mean (or difference of two-sample mean) vs. Normal distribution, and therefore can be used to assess the assumption of normality of test statistics. For example, how_norm.cauchy is for Cauchy distribution. The user could specific the sample size, location, and scale parameter of Cauchy distribution. This function works for both one-sample and two-sample tests. Since the mean and variance doesn't exist for Cauchy distribution, we expect the sample mean doesn't converge to Normal even when large sample size. The QQ plot could show the non-normality of sample mean of Cauchy distribution even with sample size of 1000.

```
> how_norm.cauchy(n=1000,location=0, scale=1, sim=10^4)
```



2.2 Compare t-test and Wilcoxon rank sum test

The functions "t_vs_wilcox.distribution" can be used to compare t-test and Wilcoxon rank sum test. For example "t_vs_wilcox.exp" is for Exponential distribution. If user inputs desired sample size and parameters under null hypothesis, this function compares type I error for both test.

```
> t_vs_wilcox.exp(n=5,rate=1,n_2=7,rate_2=1, sim=1000, conf.level=0.95)
```

```
$t.test
[1] 0.027
```

```
$wilcox.test
[1] 0.046
```

If user inputs parameters under alternative hypothesis, this function compares power for both tests.

```
> t_vs_wilcox.exp(n=10,rate=1,n_2=10,rate_2=0.2, sim=1000, conf.level=0.95)
```

```
$t.test  
[1] 0.738
```

```
$wilcox.test  
[1] 0.791
```

2.3 Rank transformed t-test

The function "rank.t.test" conducts the rank transformed t-test proposed by Zimmerman and Zumbo (1993). It transform the actual data to their rank and then conduct the unequal variance t-test(Welch t-test) on the rank.

```
> x <- rexp(8)+8  
> y <- 5*rexp(5)+3  
> rank_t.test(x,y)
```

Welch Two Sample t-test

```
data: rank.x and rank.y  
t = 0.7516, df = 5.408, p-value = 0.4837  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
 -4.570534  8.470534  
sample estimates:  
mean of x mean of y  
    7.75    5.80
```

3 Limitation and Future Work

There are some limitations with the "rank.t.test" function. First, it can not deal with large portion of ties very well. Second, it can not transform the μ (the true sample mean or the true difference of two-sample mean in null hypothesis) to its rank. Therefore, it only works when the null hypothesis is no difference with 0. Future work needs to be done for these limitations. After that, the comparison with this test could also be added to the functions "t_vs_wilcox.distribution".

4 References

Zimmerman, D. W. and Zumbo, B. D. (1993) Rank transformations and the power of the Student t Test and Welch t's test for Non-Normal populations with unequal variances. *Canadian Journal of Experimental Psychology*, 47:3, 523-539.