Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

2016 - Volume: 17 Number: 3
Page: 506 - 511
DOI: 10.18038/btda.20571
Received: 26 February 2016 Revised: 20 June 2016 Accepted: 12 August 2016

COMPARISON OF TWO INDEPENDENT GROUPS BY USING THE LOWER AND UPPER QUANTILES AND PERCENTILE BOOTSTRAP

Gözde NAVRUZ [1,*] , A. Fırat Özdemir[1]

[1] Department of Statistics, Faculty of Science, Dokuz Eylül University, İzmir, Turkey

ABSTRACT

The frequently used way of comparing two independent groups is to compare in terms of some measure of location such as mean. For non-normal and heteroscedastic cases, trimmed mean, median or some other robust measures of location can be used instead. However, determination of the differences in the tails of the groups might be of interest. For this reason, comparing the lower and upper quantiles becomes an important issue. In this study, Harrell-Davis estimator and the default quantile estimator of R are compared in terms of actual Type I error rates. When quantiles close to zero or one are compared with small sample sizes Gumbel's estimator, and when quantiles close to median are compared with large sample sizes Harrell Davis estimator saved actual Type I error rate better.

Keywords: Two independent groups, Gumbel estimator, Harrell Davis estimator, Percentile bootstrap

1. INTRODUCTION

The most common idea for comparing two independent groups is to use a measure of location such as mean. For non-normal distributions the median or some other robust measure of location might be used instead. But, when the matter is to determine whether the differences occur in the tails of distributions or not, the quantiles should be considered as well.

In a wide range of working areas, there can be an interest in determining how the high scoring individuals in one group compare to the high scoring individuals in the other group. On the other side, to compare the low scoring observations can be the research interest.

Consider an experimental method and suppose it is being compared with a control group. In such a study, the experimental method can be effective for low-scoring participants, that is, the low-scoring participants in the experimental group have higher scores than low-scoring participants in the control group. On the other hand, the experimental method can be detrimental for high-scoring participants, which means the high-scoring participants in the experimental group have lower scores than high-scoring participants in the control group. In short, different subpopulations of participants respond in different ways to the experimental method.

For example, consider an experimental method for teaching reading. Students who do poorly under the standard method benefit from the experimental method, but the experimental method is detrimental to students who do well using the standard technique [1].

Furthermore, if an experimental method is expensive or invasive, knowing how different subpopulations compare might affect the policy or strategy one is willing to adopt when dealing with a particular problem [2].

When comparing two independent groups by using Student's *t*-test, the result may indicate that, there is no significant difference between these two groups. However, there can be a significant difference

*Corresponding Author: gnavruz@gmail.com

between the lower quantiles of populations. Hence, it can be useful to compare the quantiles of two groups instead. The quantile of a distribution is defined as

$$Q(q) = F^{-1}(q) = \inf\{x : F(x) \geq q\}, \quad 0 < q < 1 \tag{1}$$

where q denotes the quantile to be estimated and F denotes the cumulative distribution function.

A variety of methods for estimating population quantiles and additional comparisons of various estimators are available in the literature. Some of them have advantages in particular situations, but certainly none of them is the best. In this study Harrell Davis estimator and another quantile estimator which was studied by Gumbel are investigated [3,4].

2. QUANTILE ESTIMATORS AND PERCENTILE BOOTSTRAP

Harrell Davis estimator uses all of the order statistics by taking a weighted average. For estimating the qth quantile, consider the random variable Y that have a beta distribution with parameters a=(n+1)q and b=(n+1)(1-q) where n is the sample size. The probability distribution of Y is

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} \quad , 0 \leq y \leq 1, \tag{2}$$

where Γ is the gamma function. Let

$$W_i = P\left(\frac{i-1}{n} \leq Y \leq \frac{i}{n}\right) \tag{3}$$

Then, the Harrell Davis estimate of the qth quantile is

$$\hat{\theta}_q = \sum_{i=1}^{n} W_i X_{(i)} \tag{4}$$

where $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ denotes the order statistics of the sample $X_1, X_2, ..., X_n$.

The quantile estimator that was considered by Gumbel is the default value in R and it is also the definition 7 of Hyndman and Fan [4,5]. For estimating the qth quantile, $m = 1 - q$ and $j = \text{floor}(nq + m)$ are evaluated where $\text{floor}(x)$ means the largest integer not greater than x. Then, $\gamma = nq + m - j$ is calculated. By using the general definition

$$\hat{\theta}_q = (1-\gamma)X_{(j)} + \gamma X_{(j+1)} \tag{5}$$

quantile q is estimated by taking the weighted average of two order statistics where $X_{(j)}$ denotes the jth order statistics.

The aim is to test $H_0 : \theta_{q1} = \theta_{q2}$ where θ_{q1} and θ_{q2} are the qth quantiles of the first and second populations respectively. For the purpose of investigating the control over actual Type I error rates, The Harrell Davis estimator and the quantile estimator that is referred to Gumbel are used in conjunction with a percentile bootstrap method. The details of the mentioned method are explained.

1. Let X_{ij} be a random sample from the jth group, $j = 1, 2$ and $i = 1, ..., n_j$. A bootstrap sample from the jth group is generated by randomly sampling with replacement yielding X^*_{ij}.

2. $\hat{\theta}^*_j$ is the estimate of the qth quantile for group j based on the bootstrap sample.

3. The difference between the quantile estimators of groups which are obtained from bootstrap sample is defined as $d^* = \hat{\theta}^*_1 - \hat{\theta}^*_2$.

4. First three steps are repeated B times (for this study B=2000).

5. $d^*_1, ..., d^*_B$ values are obtained and putted in ascending order in this manner $d^*_{(1)} \leq ... \leq d^*_{(B)}$.

6. $\ell = \alpha B / 2$ is rounded to nearest integer where α is the nominal significance level (for this study $\alpha = 0.05$) and $u = B - \ell$.

7. An approximate $1-\alpha$ confidence interval for $\theta_1 - \theta_2$ is found as $(d^*_{(\ell+1)}, d^*_{(u)})$.

Furthermore, let A is the number of times that $d^* < 0$ and C is the number of times that $d^* = 0$.

$$\hat{p}^* = \frac{A + 0.5C}{B} \tag{6}$$

A generalized p-value is $2\min(\hat{p}^*, 1 - \hat{p}^*)$ and clearly, if $p - \text{value} < \alpha$, H_0 is rejected.

3. DESIGN OF SIMULATION AND RESULTS

In this part of the study, the performance of the given quantile estimators (Harrell Davis estimator and Gumbel) are compared with a simulation study. The comparison is performed in terms of actual type I error rates and the nominal significance level was set at $\alpha = 0.05$.

Normal, symmetric and heavy-tailed, asymmetric and light-tailed, asymmetric and heavy-tailed distributions are used. In particular, g-and-h distributions are used with different g and h parameters in order to generate data from those specific types of distributions. g-and-h distribution allows to observe how a distribution differs from normal distribution with the parameters g, which is about skewness and with the parameter h, which is about kurtosis; and when g=h=0 the g-and-h distribution is equivalent to the standard normal distribution [6].

Let Z be a random variable which is generated from standard normal distribution, with the transformation $X = \frac{(\exp(gZ) - 1)\exp(\frac{hZ^2}{2})}{g}$ when $g \neq 0$, and with the transformation $X = Z\exp(\frac{hZ^2}{2})$ when g=0 the data is generated from g-and-h distribution [6]. In this study, four different g-and-h distributions are used.

- Standard normal distribution, g=0 and h=0
- Symmetric and heavy tailed distribution, g=0 and h=0.5
- Asymmetric and light tailed distribution, g=0.5 and h=0
- Asymmetric and heavy tailed distribution, g=0.5 and h=0.5

Both small and large sample sizes are considered, sample sizes are chosen as n=10, n=20 and n=40. Besides median (q=0.5), lower (q=0.1) and upper (q=0.9) quantiles are also compared. In total, 36 different experimental conditions are examined which are obtained by combining the sample sizes, quantile estimators and compared quantiles. Simulations were based on 10000 replications and all simulations were done with R programming language (R version 3. 2. 2).

To interpret the simulation results, Bradley's liberal criteria of robustness is considered [7]. When the actual Type I error rates of quantile estimators fall within the interval 0.025 and 0.075, the control over the Type I error is achieved when the nominal significance level is set at 0.05.

The following four tables represent the actual significance level results for each distribution. Furthermore, the marked cells are the results that could not fall in the interval (0.025, 0.075).

Table 1. Actual Type I Error Rates, g=0, h=0

	n	Gumbel’s	HD
q = 0.1	10	0.0631	0.1208
	20	0.0341	0.0807
	40	0.0287	0.0446
q = 0.5	10	0.0351	0.0476
	20	0.0355	0.0468
	40	0.0344	0.0441
q = 0.9	10	0.0633	0.1234
	20	0.0356	0.0855
	40	0.0335	0.0467

Table 1 shows the actual significance level results for standard normal distribution. Gumbel’s quantile estimator could control the actual significance levels for all settings, whereas Harrell Davis estimator could not control the results within the bounds 0.025 and 0.075 four times. The reason of deviations from nominal significance level with Harrell Davis estimator is comparing the lower or upper quantiles with relatively small sample sizes.

Table 2. Actual Type I Error Rates, g=0, h=0.5

	n	Gumbel’s	HD
q = 0.1	10	0.0798	0.1376
	20	0.0369	0.0966
	40	0.0336	0.0537
q = 0.5	10	0.0340	0.0443
	20	0.0385	0.0462
	40	0.0332	0.0430
q = 0.9	10	0.0739	0.1381
	20	0.0383	0.0978
	40	0.0329	0.0583

Table 2 shows the actual significance level results for symmetric and heavy tailed g-and-h distribution. Gumbel’s estimator was able to keep actual significance levels in the interval (0.025, 0.075) in all cases except one. On the other side, Harrell Davis estimator could not control the actual significance levels when comparing lower and upper quantiles with small sample sizes again four times.

Table 3. Actual Type I Error Rates, g=0.5, h=0

	n	Gumbel's	HD
q = 0.1	10	0.0608	0.1152
	20	0.0365	0.0785
	40	0.0293	0.0423
q = 0.5	10	0.0366	0.0494
	20	0.0361	0.0463
	40	0.0376	0.0447
q = 0.9	10	0.0698	0.1317
	20	0.0343	0.0928
	40	0.0299	0.0502

Table 3 shows the actual significance level results for asymmetric and light tailed g-and-h distribution. It is seen that, the actual significance levels fell in the stated interval (0.025, 0.075) in all settings with Gumbel's estimator. Harrell Davis estimator had four actual significance levels outside of the Bradley's interval (0.025, 0.075).

Table 4. Actual Type I Error Rates, g=0.5, h=0.5

	n	Gumbel's	HD
q = 0.1	10	0.0717	0.1307
	20	0.0371	0.0937
	40	0.0312	0.0510
q = 0.5	10	0.0331	0.0419
	20	0.0358	0.0464
	40	0.0341	0.0421
q = 0.9	10	0.0719	0.1299
	20	0.0373	0.1000
	40	0.0310	0.0549

Finally, Table 4 shows the actual significance level results for asymmetric and heavy tailed g-and-h distribution. Gumbel's estimator kept actual significance levels within the bounds 0.025 and 0.075. The results of Harrell Davis estimator were similar with the other tables, that is there are four uncontrolled results.

4. CONCLUSION

In general, 36 different conditions were examined. These conditions were obtained by combining different population distributions with different sample sizes, and changing the compared quantiles. Briefly, the actual significance levels fell in the interval (0.025, 0.075) with Gumbel's estimator 35 times and with Harrell Davis estimator 20 times.

The quantile estimator which is referred to Gumbel could control the actual significance levels in all conditions except one of them. Even when comparing quantiles that are close to zero or one with relatively small sample sizes, its results were not affected.

On the other side, Harrell Davis estimator could not control the actual significance levels when comparing the lower and upper quantiles with small sample sizes. However, when the median was considered the actual significance levels were very close to nominal level. Also note that, when sample size was relatively large, good results were obtained with all quantiles.

When population quantiles of two independent groups are compared in terms of controlling actual significance levels, none of the quantile estimators is convenient in all cases. The most appropriate quantile estimator should be preferred according to how large is the sample size and how quantiles are close to zero or one.

In particular, when quantiles that are close to zero or one are compared with small sample sizes, control over the actual Type I error rate is achieved by using the Gumbel's estimator. Hence, Gumbel's quantile estimator can be suggested since its actual significance levels fall in the interval (0.025, 0.075) in almost all cases.

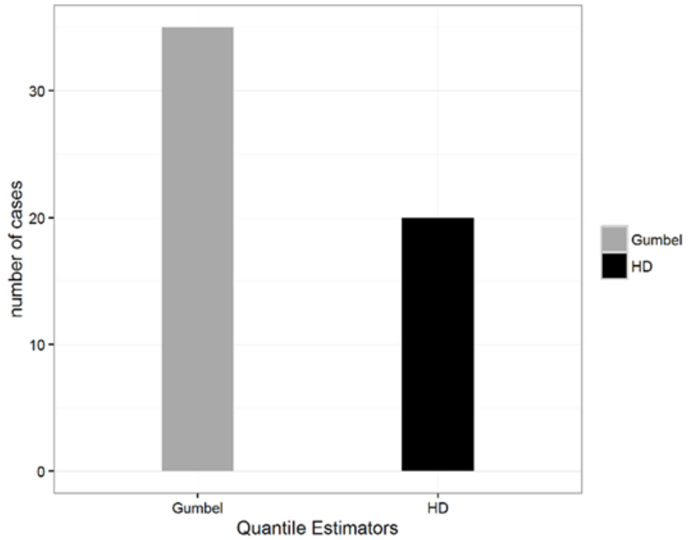

Figure 1. Number of cases that quantile estimators were able to control nominal significance level $\alpha=0.05$ within Bradley's limits

Furthermore, Figure 1 also summarizes the performance of quantile estimators. The number of cases that quantile estimators could control the actual significance levels is shown, and clearly it is seen that Gumbel's estimator is better.

REFERENCES

[1] Wilcox RR. Comparing two independent groups via multiple quantiles The Statistician, 1995; 44(1): 91-99.

[2] Wilcox RR. Introduction to robust estimation and hypothesis testing. Amsterdam: Elsevier/ Academic Press; 2012.

[3] Harrell FE, Davis CE. A new distribution-free quantile estimator Biometrika, 1982; 69 : 635-640,

[4] Gumbel EJ. La Probabilité des Hypothèses Comptes Rendus de l'Académie des Sciences (Paris), 1939; 209: 645-647.

[5] Hyndman RJ, Fan Y. Sample quantiles in statistical packages The American Statistician, 1996; 50(4): 361-365,

[6] Hoaglin DC. Summarizing shape numerically: The g-and-h distribution.In: Haoglin D, Mostseller F, Tukey J editors. Exploring data tables trends and shapes, New York: Wiley, 1985; p.461-515.

[7] Bradley JV. Robustness British Journal of Mathematical and Statistical Psychology, 1978; 31: 144-152.