Communications of the Korean Statistical Society
2012, Vol. 19, No. 1, 129–134
DOI: http://dx.doi.org/10.5351/CKSS.2012.19.1.129

# A Study on the Bi-Aspect Test for the Two-Sample Problem

Seung-Man Hong[a], Hyo-Il Park[1,b]

[a]Department of Informational Statistics, Korea University
[b]Department of Statistics, Chongju University

**Abstract**

In this paper we review a bi-aspect nonparametric test for the two-sample problem under the location translation model and propose a new one to accommodate a more broad class of underlying distributions. Then we compare the performance of our proposed test with other existing ones by obtaining empirical powers through a simulation study. Then we discuss some interesting features related to the bi-aspect test with a comment on a possible expansion for the proposed test as concluding remarks.

Keywords: Bi-aspect test, combining function, nonparametric test, permutation principle.

## 1. Introduction

When one considers a comparison study for a treatment with control, one may try to apply a suitable nonparametric test when the underlying distribution cannot be assumed with any specific one. All the nonparametric test statistics are directly related with ranks that may alleviate some severe departure from the usual observations that may be called as outliers. Because of minimal assumptions for the underlying distributions, it may well be that the powers of resulting nonparametric tests are lower than those of parametric ones under some specific distributions. Then in order to enhance the power of test, one may consider to use several test procedures simultaneously (*cf.* Park, 2011a, 2011b). This procedure may be called the versatile test. In addition, one may try to reduce the scope of the null hypothesis by splitting the null hypothesis into sub-hypotheses and then intersecting the splitted sub-hypotheses. Pesarin (2001) initiated this procedure and named it as the multi-aspect test. This test procedure has been developed and expanded in various situations by many authors. Marozzi (2004a) considered a bi-aspect test procedure for the location parameter for the two-sample problem and expanded it for the multi-sample case (*cf.* Marozzi, 2004b). In addition, for testing equality of two distributions in a case-control design with treatment effects, Salmaso and Solari (2005) considered several different features of a null hypothesis that leads to the multiple-aspect test.

In order to complete the chosen test, one has to derive the null distribution of the test statistics to obtain the critical value for any given significance level or $p$-value for a more general conclusion of the test. However, the derivation of the null distribution may be difficult if not impossible since the used statistics may be correlated to each other in a complicated manner. One way out this quagmire would be to use a re-sampling method such as the bootstrap or permutation method. This approach may depend heavily on the computer ability and applicable software.

In this research, we propose a new bi-aspect nonparametric test for the two- sample problem under the location translation model. In the next section we review Marozzi's result and propose a new counterpart for broad applications to various distributions. We consider the use of permutation principle

This work was supported by Korea University Research Fund.
[1] Corresponding author: Professor, Department of Statistics, Chongju University, Chongju, Choongbook 360-764, Korea.
E-mail: hipark@cju.ac.kr

for obtaining $p$-values. Then we compare the performance of the proposed test with Marozzi's that includes some individual tests (through a simulation study) and discuss interesting features related to the bi-aspect tests as concluding remarks.

## 2. Bi-Aspect Test

Let $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$ be two independent random samples from populations with continuous but unknown distribution functions $F_1$ and $F_2$, respectively. Then we consider the following location translation model such that for some $\delta \in (-\infty, \infty)$,

$$F_2(x) = F_1(x + \delta), \quad \text{for all } x \in (-\infty, \infty).$$

Under this model, Marozzi (2004a, 2004b) proposed bi-aspect nonparametric test procedures for testing the null hypothesis

$$H_0 : \{\delta = 0\} \cap \{F_1 = F_2\}$$

based on the following two statistics $T_1$ and $T_2$ such that

$$T_1 = \sum_{i=1}^{n_1} X_{1i} \quad \text{and} \quad T_2 = \sum_{i=1}^{n_1} I\left(X_{1i} > \tilde{M}\right),$$

where $I(\cdot)$ is an indicator function and $\tilde{M}$ is a sample median from the combined sample. We note that $T_1$ is a version of two-sample $t$-statistic and $T_2$, the Mood-type median test statistic (*cf.* Mood, 1950). This means that for testing $H_{01} : \delta = 0$ and $H_{02} : F_1 = F_2$, one may use $T_1$ and $T_2$ as test statistics, respectively. For the time being, for more detailed discussion of our arguments, we now consider the one-sided alternative such that $H_{11} : \delta > 0$ or $H_{12} : F_1(x) < F_2(x)$ for some $x \in (-\infty, \infty)$. Let $\lambda_1$ and $\lambda_2$ be the respective $p$-values for testing $H_{01} : \delta = 0$ against $H_{11} : \delta > 0$ and $H_{02} : F_1 = F_2$ against $H_{12} : F_1(x) < F_2(x)$ for some $x \in (-\infty, \infty)$ with $T_1$ and $T_2$. Then by choosing a suitable combining function to obtain an overall $p$-value, one can continue this testing procedure as follows. With the Tippett combining function (*cf.* Pesarin, 2001), Marozzi (2004a, 2004b) proposed a test statistic $T_{12}$ based on $T_1$ and $T_2$ as follows.

$$T_{12} = \max\{1 - \lambda_1, 1 - \lambda_2\}.$$

For each $j$, $j = 1, 2$, we note that the $p$-value $\lambda_j$ is a random variable since $\lambda_j$ is a function from $T_j$. Since at least any one of $\lambda_j$'s tends to be 0 if $H_0 : \{\delta = 0\} \cap \{F_1 = F_2\}$ is false, the test based on $T_{12}$ would be to reject $H_0$ when the value of $T_{12}$ approaches 1. Then in order to complete this test procedure, we need the null distribution of $T_{12}$. For this, Marozzi (2004a, 2004b) applied the permutation principle (*cf.* Good, 2000; Pesarin, 2001) for obtaining the overall $p$-value for $T_{12}$. Especially, Marozzi (2004a) presented a procedure by obtained the $p$-value with the Monte-Carlo approach and modified slightly the form of statistics by adding 1/2 and 1 in the numerator and denominator, respectively to ensure to obtain the $p$-values in the interval because of computational convenience. Since the statistic $T_1$ is a version of $t$-statistic that produces an optimal test when the underlying distribution is normal, one may worry about the power of the test when the underlying distribution may be skewed or heavy-tailed. Thus the test based on $T_{12}$ may be inappropriate when the underlying distribution is exponential or Cauchy. For this reason, we may propose a new test that uses the Wilcoxon rank-sum

Table 1: Normal distribution

| Test | $(n_1, n_2)$ | $\delta$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 |
| | (15, 15) | 0.0495 | 0.1361 | 0.2759 | 0.4842 | 0.6831 | 0.8442 | 0.9384 |
| $T_1$ | (20, 10) | 0.0455 | 0.1252 | 0.2560 | 0.4428 | 0.6372 | 0.8059 | 0.9129 |
| | (20, 30) | 0.0530 | 0.1748 | 0.3990 | 0.6682 | 0.8614 | 0.9626 | 0.9941 |
| | (15, 15) | 0.0129 | 0.0367 | 0.0934 | 0.1837 | 0.3164 | 0.4685 | 0.6355 |
| $T_2$ | (20, 10) | 0.0237 | 0.0579 | 0.1263 | 0.2299 | 0.3678 | 0.5225 | 0.6744 |
| | (20, 30) | 0.0221 | 0.0755 | 0.1834 | 0.3581 | 0.5665 | 0.7587 | 0.8897 |
| | (15, 15) | 0.0486 | 0.1295 | 0.2630 | 0.4527 | 0.6517 | 0.8180 | 0.9222 |
| $T_3$ | (20, 10) | 0.0438 | 0.1187 | 0.2441 | 0.4174 | 0.6053 | 0.7773 | 0.8942 |
| | (20, 30) | 0.0528 | 0.1711 | 0.3766 | 0.6427 | 0.8431 | 0.9555 | 0.9905 |
| | (15, 15) | 0.0534 | 0.1409 | 0.2878 | 0.4919 | 0.6909 | 0.8482 | 0.9396 |
| $T_{12}$ | (20, 10) | 0.0543 | 0.1407 | 0.2774 | 0.4617 | 0.6561 | 0.8171 | 0.9180 |
| | (20, 30) | 0.0595 | 0.1888 | 0.4137 | 0.6808 | 0.8883 | 0.9648 | 0.9942 |
| | (15, 15) | 0.0504 | 0.1320 | 0.2674 | 0.4574 | 0.6553 | 0.8198 | 0.9228 |
| $T_{32}$ | (20, 10) | 0.0491 | 0.1290 | 0.2565 | 0.4290 | 0.6168 | 0.7843 | 0.8968 |
| | (20, 30) | 0.0563 | 0.1783 | 0.3853 | 0.6480 | 0.8466 | 0.9563 | 0.9905 |

statistic, $T_3$,

$$T_3 = \sum_{i=1}^{n_1} R_{1i},$$

where $R_{1i}$ is the rank of $X_{1i}$ from the combined sample. Then based on $T_3$, let $\lambda_3$ be the individual $p$-value for testing $H_{01} : \delta = 0$ against $H_{11} : \delta > 0$. Then using the Tippett combining function, we may propose a test statistic $T_{32}$ as follows:

$$T_{32} = \max\{1 - \lambda_3, 1 - \lambda_2\}.$$

The testing rule also rejects $H_0$ in favor of $H_1$ for large values of $T_{32}$. Then by using the permutation principle to obtain the overall $p$-value, we can complete the test. In the next section, we compare the performance between $T_{12}$ and $T_{32}$ through a simulation study.

## 3. Simulation Study

In order to compare the performance of the proposed test with other procedures, we now carry out a simulation study. We tabulate the simulation results in Table 1 through Table 4. In this comparison study, we consider the two bi-aspect tests, $T_{12}$ and $T_{32}$ with the individual tests, $T_1$, $T_2$ and $T_3$ to obtain empirical powers through the application of the permutation principle. We consider the normal, Cauchy, exponential and uniform distributions with unit variance except the Cauchy distribution under the location translation model. We consider the standard Cauchy distribution. The value of varies from 0.0 to 1.2 with increment 0.2. The nominal significance level is 0.05 for all cases. We consider the three cases $(15, 15)$, $(20, 10)$ and $(20, 30)$ for $(n_1, n_2)$. All the results are based on 10,000 simulations with the Monte-Carlo method and within a simulation, we applied the permutation principle by 10,000 iterations also with the Monte-Carlo approach to estimate the distribution. All the computations were carried out by SAS/IML with PC-version. First, we note that the test based on $T_2$ barely achieves the nominal significance level for all cases. This may be a drawback of the median test. For the normal and uniform distributions(Table 1 and Table 4), the tests, $T_1$ and $T_{12}$ show the best performance while for the Cauchy and exponential case(Table 2 and Table 3), $T_3$ and $T_{32}$ achieve more powerful results. Especially we note that $T_{32}$ performs the best of all for the Cauchy distribution(Table 2). Therefore $T_{32}$ would be effective for the heavy-tailed underlying distribution.

Table 2: Cauchy distribution

| Test | $(n_1, n_2)$ | $\delta$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 |
| $T_1$ | (15, 15) | 0.0524 | 0.0716 | 0.0934 | 0.1199 | 0.1540 | 0.1880 | 0.2260 |
| | (20, 10) | 0.0510 | 0.0682 | 0.0903 | 0.1180 | 0.1446 | 0.1765 | 0.2093 |
| | (20, 30) | 0.0512 | 0.0702 | 0.0941 | 0.1196 | 0.1495 | 0.1809 | 0.2139 |
| $T_2$ | (15, 15) | 0.0125 | 0.0289 | 0.0601 | 0.1068 | 0.1720 | 0.2581 | 0.3506 |
| | (20, 10) | 0.0248 | 0.0521 | 0.0918 | 0.1477 | 0.2223 | 0.3094 | 0.3966 |
| | (20, 30) | 0.0218 | 0.0552 | 0.1197 | 0.2166 | 0.3480 | 0.4934 | 0.6271 |
| $T_3$ | (15, 15) | 0.0476 | 0.0862 | 0.1387 | 0.2126 | 0.2977 | 0.3939 | 0.4869 |
| | (20, 10) | 0.0483 | 0.0826 | 0.1314 | 0.1962 | 0.2734 | 0.3603 | 0.4457 |
| | (20, 30) | 0.0527 | 0.1021 | 0.1821 | 0.2874 | 0.4150 | 0.5473 | 0.6617 |
| $T_{12}$ | (15, 15) | 0.0589 | 0.0684 | 0.1262 | 0.1788 | 0.2495 | 0.3335 | 0.4244 |
| | (20, 10) | 0.0657 | 0.0997 | 0.1452 | 0.2053 | 0.2795 | 0.3646 | 0.4495 |
| | (20, 30) | 0.0642 | 0.1050 | 0.1737 | 0.2703 | 0.3945 | 0.5288 | 0.6534 |
| $T_{32}$ | (15, 15) | 0.0489 | 0.0898 | 0.1458 | 0.2234 | 0.3149 | 0.4169 | 0.5160 |
| | (20, 10) | 0.0526 | 0.0939 | 0.1494 | 0.2225 | 0.3112 | 0.4058 | 0.4959 |
| | (20, 30) | 0.0553 | 0.1097 | 0.2021 | 0.3228 | 0.4601 | 0.6041 | 0.7229 |

Table 3: Exponential distribution

| Test | $(n_1, n_2)$ | $\delta$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 |
| $T_1$ | (15, 15) | 0.0505 | 0.1458 | 0.3167 | 0.5234 | 0.7156 | 0.8473 | 0.9251 |
| | (20, 10) | 0.0503 | 0.1563 | 0.3261 | 0.5250 | 0.7066 | 0.8311 | 0.9150 |
| | (20, 30) | 0.0508 | 0.1735 | 0.4083 | 0.6634 | 0.8576 | 0.9535 | 0.9873 |
| $T_2$ | (15, 15) | 0.0122 | 0.0506 | 0.1419 | 0.3028 | 0.4825 | 0.6676 | 0.8061 |
| | (20, 10) | 0.0245 | 0.0749 | 0.1677 | 0.2976 | 0.4447 | 0.5900 | 0.7142 |
| | (20, 30) | 0.0238 | 0.1006 | 0.2891 | 0.5693 | 0.8107 | 0.9412 | 0.9864 |
| $T_3$ | (15, 15) | 0.0497 | 0.2080 | 0.4593 | 0.6981 | 0.8522 | 0.9370 | 0.9762 |
| | (20, 10) | 0.0476 | 0.1991 | 0.4253 | 0.6376 | 0.7941 | 0.8911 | 0.9438 |
| | (20, 30) | 0.0508 | 0.2779 | 0.6446 | 0.8870 | 0.9758 | 0.9957 | 0.9995 |
| $T_{12}$ | (15, 15) | 0.0545 | 0.1619 | 0.3449 | 0.5633 | 0.7541 | 0.8796 | 0.9461 |
| | (20, 10) | 0.0593 | 0.1747 | 0.3520 | 0.5541 | 0.7305 | 0.8494 | 0.9272 |
| | (20, 30) | 0.0609 | 0.2091 | 0.4741 | 0.7503 | 0.9205 | 0.9812 | 0.9969 |
| $T_{32}$ | (15, 15) | 0.0506 | 0.2096 | 0.4599 | 0.6983 | 0.8522 | 0.9370 | 0.9762 |
| | (20, 10) | 0.0543 | 0.2038 | 0.4267 | 0.6383 | 0.7943 | 0.8911 | 0.9438 |
| | (20, 30) | 0.0553 | 0.2813 | 0.6455 | 0.8873 | 0.9759 | 0.9957 | 0.9995 |

## 4. Some Concluding Remarks

The bi-aspect test has some flexibility since one may consider various existing tests that combine them together upon the situations. Thus when no information for the underlying distributions is available, one can try to combine various tests and increase the significance of the test.

In Section 2, we stated that the statistic $T_1$ is a version of $t$-statistic without the expression of the pooled sample variance $S_p^2$. Since we have chosen the permutation principle to obtain $p$-value, the divisor $\sqrt{S_p^2}$ at $T_1$ would be redundant since $\sqrt{S_p^2}$ is common for all permutations. If we take the approach of invoking the $t$-distribution table, then the complete form for the $t$-statistic should be used.

For the construction of $T_{12}$ and $T_{32}$ in Section 2, we have used the respective $p$-values instead of the statistics themselves since the form of $T_{12}$ and $T_{32}$ is maximal. If $T'_{12} = \max\{T_1, T_2\}$ or $T'_{32} = \max\{T_3, T_2\}$, then both $T'_{12}$ and $T'_{32}$ would be meaningless. Then in order to make them meaningful, one has to obtain the null means and variances of the respective statistics in order to normalize them. This is why we use the respective $p$-values rather than the original statistics.

Table 4: Uniform distribution

| Test | $(n_1, n_2)$ | $\delta$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 | 1.2 |
| $T_1$ | (15, 15) | 0.0503 | 0.1258 | 0.2747 | 0.4763 | 0.6808 | 0.8481 | 0.9439 |
| | (20, 10) | 0.0470 | 0.1240 | 0.2574 | 0.4391 | 0.6371 | 0.8105 | 0.9195 |
| | (20, 30) | 0.0482 | 0.1656 | 0.2795 | 0.6475 | 0.8596 | 0.9616 | 0.9929 |
| $T_2$ | (15, 15) | 0.0140 | 0.0320 | 0.0634 | 0.1062 | 0.1807 | 0.2844 | 0.4148 |
| | (20, 10) | 0.0232 | 0.0498 | 0.0905 | 0.1541 | 0.2422 | 0.3517 | 0.4806 |
| | (20, 30) | 0.0199 | 0.0517 | 0.1111 | 0.2131 | 0.3411 | 0.5036 | 0.6776 |
| $T_3$ | (15, 15) | 0.0482 | 0.1237 | 0.2615 | 0.4419 | 0.6240 | 0.7861 | 0.8978 |
| | (20, 10) | 0.0464 | 0.1168 | 0.2405 | 0.4024 | 0.5797 | 0.7446 | 0.8625 |
| | (20, 30) | 0.0469 | 0.1609 | 0.3597 | 0.6054 | 0.8088 | 0.9283 | 0.9780 |
| $T_{12}$ | (15, 15) | 0.0525 | 0.1302 | 0.2758 | 0.4769 | 0.6811 | 0.8482 | 0.9440 |
| | (20, 10) | 0.0529 | 0.1308 | 0.2630 | 0.4428 | 0.6386 | 0.8109 | 0.9201 |
| | (20, 30) | 0.0522 | 0.1699 | 0.3817 | 0.6483 | 0.8596 | 0.9616 | 0.9929 |
| $T_{32}$ | (15, 15) | 0.0500 | 0.1250 | 0.2622 | 0.4424 | 0.6242 | 0.7861 | 0.8978 |
| | (20, 10) | 0.0516 | 0.1237 | 0.2455 | 0.4056 | 0.5816 | 0.7451 | 0.8634 |
| | (20, 30) | 0.0510 | 0.1652 | 0.3616 | 0.6062 | 0.8088 | 0.9284 | 0.9780 |

In order to express $T_2$ using the ranks, let $n = n_1 + n_2$. Then a rank for a median, $n_M$, can be defined as $n_M = [n/2] + 1$, where [x] is the largest integer of the real number $x$. Then $X_{(n_M)}$ be a median and so can be $\tilde{M}$, where $X_{(n_M)}$ is the $n_M$th order statistic from the combined sample. Since the ranks preserve the original order of the observations, we see that

$$T_2 = \sum_{i=1}^{n_1} \left(X_{1i} > \tilde{M}\right) = \sum_{i=1}^{n_1} I\left(R_{1i} > n_M\right).$$

Even though Marozzi (2004a) did not derive the limiting distribution for $T_{12}$, one may obtain the limiting distribution of $T_{12}$ using the large sample approximation theorem with applications of Slutsky's theorem and Cramer-Wold device (*cf.* Serfling, 1980). However, the resulting distribution will be a bivariate normal whose distributional tables are rare or not at all (*cf.* Owen, 1962). Also the software for the computation of probability hardly provide the multivariate normal distributions or any multivariate distribution. For example, SAS can only provide bivariate normal distribution. These points might make Marozzi (2004a) to reconsider the limiting distribution for $T_{12}$.

The simultaneous use of several statistics in the nonparametric test procedures has been developed for the purpose to obtain high power of test; however, the increase of the use of the statistics does not guarantee the increase of power (*cf.* Park, 2011a). Therefore before analyzing the data with intensity, it would be necessary to take some preliminary or explanatory analysis to choose some suitable statistics that can be included in a multi-aspect test.

For the combining functions of several statistics or $p$-values, one may consider using summing or quadratic types for the consideration of enhancing power of test; however, we note that the quadratic form can only be used for the two-sided test. Therefore, the summing and maximal type can be used in order to accommodate more general cases in our study. In addition, a comparison study for both types will appear in the near future in the suitable medium.

Finally we note that there is one more re-sampling method such as the bootstrap method. The distinction between the bootstrap and permutation methods are as follows. The bootstrap method re-samples with replacement but the permutation principle, without replacement from the original sample, however the difference can be significant in some cases (*cf.* Good, 2000).

## Acknowledgment

The authors wish to express their since appreciation to the referees for pointing out errors to enhance the quality of this paper.

## References

Good, P. (2000). *Permutation Tests-A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer, New York.

Marozzi, M. (2004a). A bi-aspect nonparametric test for the two-sample location problem, *Computational Statistics & Data Analysis*, **44**, 639–648.

Marozzi, M. (2004b). A bi-aspect nonparametric test for the multi-sample location problem, *Computational Statistics & Data Analysis*, **46**, 81–92.

Mood, A. M. (1950). *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Owen, D. B. (1962). *Handbook of Statistical Tables*, Addison-Wesley, Boston.

Park, H. I. (2011a). A nonparametric test procedure based on a group of quantile tests, *Communications in Statistics-Simulation and Computation*, **40**, 759–783.

Park, H. I. (2011b). A nonparametric test procedure based on quantiles for the right censored data, Submitted for publication.

Pesarin, F. (2001). *Multivariate Permutation Tests with Applications in Biostatistics*, Wiley, New York.

Salmaso, L. and Solari, A. (2005). Multiple-aspect testing for case-control designs, *Metrika*, **62**, 331–340.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

*Received August 2011; Revised September 2011; Accepted November 2011*