Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 4, pp. 1145 ~ 1155

Permutation tests for the multivariate data[1]

Hyo-Il Park[2] · Ju Sung Kim[3]

Abstract

In this paper, we consider the permutation tests for the multivariate data under the two-sample problem setting. We review some testing procedures, which are parametric and nonparametric and compare them with the permutation ones. Then we consider to try to apply the permutation tests to the multivariate data having the continuous and discrete components together by choosing some suitable combining function through the partial testing. Finally we discuss more aspects for the permutation tests as concluding remarks.

Keywords: Mahalanobis Distance, Multivariate Data, Nonparametric Test, Partial Test, Permutation Principle,

1. Introduction

Suppose that we have two independent d-variate samples $X_1,...,X_m$ and $Y_1,...,Y_n$ whose distribution functions F and G, respectively. Then for testing

$$H_0 : F = G \text{ versus } H_1 : F \neq G, \tag{1.1}$$

one may carry out the F-test based on the Hotelling's T^2 statistic or Mahalanobis distance with the normality assumption. We note that both the statistics Hotelling's T^2 and Mahalanobis distance take the quadratic forms. When it is difficult to assume the normality but continuous, it is customary to adopt a nonparametric testing procedure. For this case, there are two kinds of test statistics. One is some metric type of statistics such as the nearest neighbors, the

[1] This work was supported by the grant of the Chungbuk National Univ. in 2007.

[2] Professor, Department of Statistics, Chong-ju Univ, Chong-ju, 360-764 Korea
E-mail : hipark@cju.ac.kr

[3] Corresponding Author : Professor, Department of Informational Statistics, Chungbuk National Univ, 361-763 Korea E-mail : kimjs@chungbuk.ac.kr

other, the quadratic form of statistics which consist of the marginal univariate nonparametric statistics. Then the former uses the standard normal distribution as its reference one for the critical value or p-value while the latter does the chi-square distribution with d degrees of freedom. The reason for taking quadratic form even in nonparametric approach, is nothing but to use the chi-square distribution. When $H_0 : F = G$ is rejected, it might be natural to ask which component or components has or have the responsibility to be arrived at such conclusion as much as the multiple comparisons for the one-way anova case. Then we note that the forementioned testing procedures which are parametric or nonparametric, can not provide the answer directly. One way out from this quagmire might be the adoption of the partial testing or marginal testing approach(cf. Pesarin, 2001). The partial testing procedure can be proceeded as follows. First, one may take test for each component for the sub- hypothesis then cook them up together for the global hypothesis using suitable combining function. This approach requires computer-intensive computation process since the distribution of the combined statistic for the global hypothesis can be hardly derived to obtain the corresponding p-value or critical value for any given significance level. Therefore re-sampling methods have become important methodology for this problem. Especially, for this setting, the permutation principle would be desirable since under $H_0 : F = G$, the data have the exchangeability property. The permutation principle has a long history since it has been proposed by Fisher(1935). However its practical application to the real problem has been postponed until recently because of its excessive computational burden. Except the computational aspect, the tests relied on the permutation principle enjoy many positive properties such as exactness(cf. Good, 2000), which will be discussed in later chapter.

In this paper, we review some well-known test procedure based on the permutation principle for the multivariate data and show several combining functions with an example. Finally we discuss some interesting features as concluding remarks.

2. Permutation tests

The main issue in this section is to review various testing procedures based on two independent samples $X_1, ..., X_m$ and $Y_1, ..., Y_n$ of d-variate observations with d -variate location parameter θ_X and θ_Y, respectively. Then for the following problem,

$$H_0 : \theta_X = \theta_Y \quad \text{versus} \quad H_1 : \theta_X \neq \theta_Y \tag{2.1}$$

under the normality assumption with common but unknown covariance matrix Σ

for X and Y, one may carry out a test based on the following Mahalanobis distance D^2:

$$D^2 = (\overline{X} - \overline{Y})^T S_u^{-1} (\overline{X} - \overline{Y}) \tag{2.2}$$

where $(\bullet)^T$ means the transpose of a matrix or vector and

$$S_u = \frac{1}{N-2}\left[\sum_{i=1}^{m}(X_i - \overline{X})(X_i - \overline{X})^T + \sum_{j=1}^{n}(Y_j - \overline{Y})(Y_j - \overline{Y})^T\right],$$

an unbiased estimate of Σ with $N = m+n$. Then it is well-known that under $H_0 : \theta_X = \theta_Y$

$$H = \frac{mn(N-d-1)}{N(N-1)d} D^2 \sim F_{d,N-d-1},$$

where $F_{d,k}$ means the F-distribution with d and k degrees of freedoms. We note that θ_X and θ_Y are the mean vectors of X and Y, respectively in this case. When it is difficult to assume the normality even though the continuity is certain, one can apply a nonparametric test based on ranks. In this way, we may consider the componentwise approach, which is extensively dealt with by Puri and Sen(1971). For this purpose, let $\phi_i(R)$ be a test statistic, which is a function of d-variate rank matrix R for the ith component from the combined sample. Also let $E(\phi_i)$ and $V(\phi)$ be the expectation and covariance matrix for $\phi_i(R)$ and $(\phi_1(R),\ldots,\phi_d(R))^T$ under $H_0 : \theta_X = \theta_Y$, respectively. We note that in this case, θ_X and θ_Y may be median vectors of X and Y, respectively. Then the testing statistic for testing $H_0 : \theta_X = \theta_Y$ is of the form by assuming that $V(\phi)$ is full-rank,

$$Q = (\phi_1(R) - E(\phi_1),\ldots,\phi_d(R) - E(\phi_d))\, V(\phi)^{-1} (\phi_1(R) - E(\phi_1),\ldots,\phi_d(R) - E(\phi_d))^T.$$

Then it is well known that the limiting distribution of Q is a chi-square with d degrees of freedom.

The data in Table 1, are from the geological problem(cf. Mardia et al., 1979). The first components(X_1 and Y_1) represent the distances between the shoulders of the larger left value and the second components(X_2 and Y_2) represent the lengths of specimens of Bairda oklahomaensis from two geological levels(Levels 1 and 2). For the data in Table 1, $H = 7.528$ and the corresponding p-value becomes 0.0050 from the F-distribution with 2 and 16 degrees of freedom. Now for the nonparametric procedure, we provide the corresponding ranks in Table 2. We used the mid-ranks for the tied observations in both components. If one uses the Wilcoxon statistic for ϕ_i for each i, $i = 1,2$, one may obtain $Q = 10.1376$ with 0.0063 as its p-value from the chi-square distribution with 2 degrees of freedom.

<Table 1> Geological data

Level 1		Level 2	
X_1	X_2	Y_1	Y_2
631	1167	682	1257
606	1222	631	1227
682	1278	631	1237
480	1045	707	1368
606	1151	631	1227
556	1172	682	1262
429	970	707	1313
454	1166	656	1283
		682	1298
		656	1283
		672	1278

<Table 2> Ranks for geological data

Level 1		Level 2	
X_1	X_2	Y_1	Y_2
8.5	5	15.5	11
5.5	8	8.5	8
13.5	14.5	8.5	10
3	2	18.5	13
5.5	3	8.5	8
4	6	13.5	12
1	1	18.5	19
2	4	11.5	16.5
		15.5	18
		11.5	16.5
		13	14.5

From the example, we note that both the test procedures rely completely on the normal distribution theory in any sense even the nonparametric case to obtain the null distribution for the test statistics. Here is one way out for this normal theory even when the data are from the population with normal distribution. This can be done by applying the permutation principle. Since D^2 and H produce the same result, one may use D^2 instead of H for the computational consideration. The permutation tests can be performed in the following order for the two-sample problem setting:

(a) Set $COUNT_{D^2}=0$(and $COUNT_Q=0$).

(b) Combine two samples $X_1,\dots,X_m$ and $Y_1,\dots,Y_n$ into one sample.

(c) Re-sample $X_1^*,\dots,X_m^*$ and $Y_1^*,\dots,Y_n^*$ without replacement.

(d) From $X_1^*,\dots,X_m^*$ and $Y_1^*,\dots,Y_n^*$, compute D^{*2} and Q^* and compare them with D^2

and Q, respectively.

(e) If $D^{*2} > D^2$(and $Q^* > Q$) then

$$COUNT_{D^2} \leftarrow COUNT_{D^2}+1(\text{and } COUNT_Q \leftarrow COUNT_Q+1)$$

(f) Repeat (c)-(e) B times. Usually the number B should be large enough.

(g) Obtain the permutational p-value by $COUNT_{D^2}/B$(and $COUNT_Q/B$).

For the data in Table 1 and the corresponding ranks in Table 2, we carried out the permutation tests based on both the statistics D^2 and Q and obtained 0.00094 and 0.00234 as their respective p-values by using SAS/IML on PC with 100000 repetition. As a matter of fact, the procedure based on Q is a permutation test and the derivation of the limiting distribution has been relied on the permutation principle. However it has not been possible to use the permutational distribution even for any reasonable sample sizes until very recently since the procedure heavily depend upon the computer ability.

3. Various nonparametric statistics for the permutation tests

However there may be some other situations where the quadratic form of the test statistics is not appropriate. Suppose that a laboratory has developed a medicine which may have effects on two symptoms simultaneously. One may draw a conclusion that this medicine is acceptable if it is effective for any one of two symptoms or for both. In this problem, the null and alternative hypotheses can be expressed as follows using the notation introduced in section 2:

$$H_0 : \theta_X \leq \theta_Y \text{ versus } H_1 : \text{at least one component is not true.} \tag{3.1}$$

This is the so-called multivariate one-sided test problem(cf. Wei and Knuiman, 1987). Also Bhattacharyya and Johnson(1970) and Johnson and Mehrotra(1972) proposed nonparametric tests based on some metric under the name of the ordered alternatives for the bivariate data. All the mentioned testing procedures for (3.1) use the standard normal distribution as their limiting distributions. Therefore it is still difficult to distinguish which component has the responsibility for the rejection of the null hypothesis directly. We note that they also do not apply the permutation principle for the null distribution. There is another type of statistics for testing (3.1), which is called as the maximal type of statistics. Boyett and Shuster(1977) considered a nonparametric test procedure based on the following statistics:

$$MT = \max\{T_1, \ldots, T_d\},$$

where for each i, $i=1,\ldots,d$

$$T_i = \frac{\overline{Y}_i - \overline{X}_i}{\sqrt{S_{pi}^2}},$$

where S_{pi}^2 is the pooled sample variance of the ith component. Also Park et al.(2001) considered a nonparamertic test procedure based on the following statistics

$$MR = \max\{NS_1, \ldots, NS_d\},$$

where for each i, $i = 1, \ldots, d$, $NS_i = \{\phi_i(R) - E(\phi_i)\}/\sqrt{V(\phi_i)}$ by varying the type of ϕ_i with component by component fashion, where $V(\phi_i)$ is the null variance of ϕ_i. Park et al.(2001) allowed that the score functions ϕ_i's can be varied component by component. For example, one component may be Wilcoxon rank sum statistic and the other, median type of statistic for the bivariate case. Then one can extend these ideas further in the following way for the testing problem (1.1). For this purpose, for each i, $i = 1, \ldots, d$, let F_i and G_i be the ith marginal distribution of X and Y, respectively. Also let $H_{0i}: F_i = G_i$ and $H_{1i}: F_i \neq G_i$. Then it is interesting to observe that

$$H_0 = \bigcap_{i=1}^{d} H_{0i} \text{ and } H_1 = \bigcup_{i=1}^{d} H_{1i}. \tag{3.2}$$

We note that $|T_i|$ or $|NS_i|$ is an appropriate statistic for testing $H_{0i}: F_i = G_i$. Also we note that the maximal function is an appropriate function for the intersection. Therefore in this vein, one may use the following maximal statistic

$$AMT = \max\{|T_1|, \ldots, |T_d|\} \text{ or } AMR = \max\{|NS_1|, \ldots, |NS_d|\} \tag{3.3}$$

for testing (1.1). We note that test procedures based on (3.3) should be appropriate for the continuous components only. However some data may contain the continuous and discrete components together. In this case, we have to consider some different approach. This may be solved by introducing the marginal or partial testing approach. In this way, Pesarin(2001) considered several test procedures based on partial tests using various combining functions. He used the p-value approach instead of directly using the test statistics directly. For this, let λ_i be the corresponding p-value for each H_{0i} based on some appropriate statistic for testing $H_{0i}: F_i = G_i$ for the continuous or discrete data. Then Pesarin(2001) considered the following combining functions:

(a) The Fisher omnibus combining function is based on the statistic

$$T_F = -2\sum_{i=1}^{d} \log(\lambda_i).$$

It is easy to show that under H_0, T_F is distributed as a chi-square distribution with $2d$ degrees of freedom with the variable transformation technique when all the components are independent.

(b) The Liptak combining function is based on the statistic

$$T_L=\sum_{i=1}^{d}\Phi^{-1}(1-\lambda_i),$$

where Φ^{-1} is the quantile function of the standard normal distribution function. A version of the Liptak function considers logistic transformation of p-values such as

$$T_P=\sum_{i=1}^{d}\log\left[\frac{1-\lambda_i}{\lambda_i}\right].$$

(c) The Tippett combining function is given by

$$T_T=\max_{1\le i\le d}\{1-\lambda_i\}.$$

In this case we note that AMT and T_T are equivalent procedures but they do not produce the exactly same calculation results.

We note that the quadratic form is also a combining function among the univariate statistics. In this case, one cannot represent the quadratic form as a function of p- alues but should use the test statistics themselves. In this approach, the amount of computation should be assessed. The order for the applications of the permutation principle becomes as follows:

(i) Obtain the p-values for each component using the procedure (a)-(g) in section 2 for the original data set.
(ii) Compute the basic statistic by choosing a combination function (a)-(c) in this section.
(iii) For each permutational configuration used in (i), do the same procedure (i) and (ii).
(iv) Count the number from (iii) whose values of combined function exceed the basic statistic.

The data for the salary of the computer experts in a company in Table 3, were analyzed originally with a multiple regression model in Chatterjee and Price(1991). The data have four variables such as duration of work experience and education and status as manager or not and the amount of salary in year. In that analysis, the salary has been used as response variable and the others, as the explanatory variables. Therefore the original purpose of data analysis was to identify the explanatory variables which may influence the amount of the salary. However in this study, we are interested in comparing the three variables, the durations of experience and education and the amount of salary according as one is manager or not. Then the analysis based on the traditional parametric or nonparametric methods with the assumption of the continuity of the underlying distribution becomes impossible. We note that the first variable, work experience, has been reduced as the grouped data. Therefore this can be considered as the ordered categorical data(cf. Pesarin, 2001). Puri and Sen(1985) proposed a test procedure based on the statistics which were derived by using the likelihood ratio principle. Therefore one may use the Puri and Sen's procedure for this partial test. For the second component, the status of the duration of education, the variable can take values 1, 2 and 3 according as the graduate of highschool, college and educational experience beyond college level. Also this can be analyzed by treating them as grouped data. However since the number of categories is too small, it would be more appropriate to compare them by Anderson-Darling test statistic for the two sample case. Finally the third variable, the amount of salary, is continuous and can be distributed as normal. Therefore the famous two sample t-test can be applied to this case as the third partial test. Then one may carry out a test procedure by choosing a combining function, which is listed as (a)-(c).

<Table 3> Salary data due to status as manger

Non-Manager			Manager		
Experience	Education	Salary	Experience	Education	Salary
1	3	11608			
1	2	11283			
1	3	11767			
2	2	11772	1	1	13876
2	1	10535	1	3	18701
2	3	12195	2	2	20872
3	2	12313	3	1	14975
4	1	11417	3	2	21371
4	3	13231	3	3	19800
4	2	12884	4	3	20263
5	2	13245	5	1	15965
5	3	13677	6	3	21352
6	1	12336	6	2	22884
6	2	13839	7	1	16978
8	2	14803	8	1	17404
8	1	13548	8	3	22184
10	1	14467	10	3	23174
10	2	15942	10	2	23780
11	1	14861	11	2	25410
12	2	16882	12	3	24170
13	1	15990	13	2	26330
14	2	17949	15	3	25685
16	2	18838	16	2	27837
16	1	17483			
17	2	19207			
20	1	19346			

<Table 4> p-values for each component

Test	Wilcoxon	Anderson -Darling	T
p-value	0.9818	0.1433	0.0000

<Table 5> p-values for selected combining functions

Combining function	Omnibus	Liptak	Tippet
Global p-value	0.0038	0.0514	0.0000

We considered three types of combining functions and obtained the permutational p- alues in the Tables 4 and 5. In Table 4, we enlisted p-values for the partial tests for the original data and in Table 5, the global p-values for the three types of combining functions are listed. From Table 5, we note that the Fisher's omnibus and Tippet combination functions yield the significant results but the Liptak procedure does not under the significance level 0.05. Then from Table 4, we see the reason for the significance is due to the salary earned in year. Especially we note that since the Tippet combining function considers the largest studentized value or the smallest p-value among the components, it is useful to detect the extreme component.

4. Concluding remarks

When one applies the permutation principle to the multivariate data, the permutation should be applied objectwisely not componentwisely. Let me explain this more clearly with an example. Suppose that we have a sample $X_1,\dots,X_n$ with d-variate observations. Then the corresponding $n \times d$ data matrix X becomes

$$X=\begin{pmatrix} X_{11},\dots,X_{1d} \\ \dots \\ X_{n1},\dots,X_{nd} \end{pmatrix}.$$

We note that data matrix X consists of n rows and d columns. If we consider to apply the permutation principle, then we must exchange the rows. For example, if we want to get a permutation by exchanging X_{11} and X_{n1}, then we have to consider the following permuted data matrix X' such that

$$X' = \begin{pmatrix} X_{n1},\ldots,X_{nd} \\ \ldots \\ X_{11},\ldots,X_{1d} \end{pmatrix}.$$

Only for the case that all the components are independent, one may consider the componentwisely permutations, whose numbers are $(n!)^d$. For more discussion you may refer to Good(2000) and/or Pesarin(2001).

The permutation principle is one of the re-sampling methods. There is another famous re-sampling method-the bootstrap method. The simple distinction between the two re-sampling methods is as follows: The permutation principle re-samples without replacement while the bootstrap method, with replacement. However it is known that the difference for the results is considerable(cf. Good, 2000). The application of the bootstrap method to the testing problem has been extensively dealt with by Westfall and Young(1993) with various situations.

There are a lot of methodologies for the multivariate data(cf. Jung, 2005, Um, 2005, Kim and Jung, 2005 and Park, 2007). However almost all cases, the data are continuous or discrete for all components. For the data used in section 3, which has continuous and discrete components together, there is little result for the analysis. Therefore it would be worth to take research in this way in the future.

References

1. Bhattacharyya, G. K. and Johnson, R. A. (1970) A layer rank test for ordered bivariate alternatives. *Annals of Mathematical Statistics* 41, 1296-1310.
2. Boyett, J. J. and Shuster, J. M. (1977) Nonparemetric one-sided tests in multivariate analysis with medical applications. *Journal of American Statistical Association* 72, 177-187.
3. Chatterjee, S. and Price, B. (1991) *Regression analysis by example*, second Edition. Wiley, New York.
4. Fisher, R. A. (1935) *The design of experiments*. Oliver and Boyd, Edinburgh.
5. Good, P. (2000) *Permutation tests-A practical guide to resampling methods for testing hypothesis*, second Edition. Springer, New York.
6. Johnson, R. A. and Mehrotra, K. G. (1972) Nonparametric tests for ordered alternatives in the bivariate case. *Journal of Multivariate Analysis* 2, 219-229.
7. Jung, K. M. (2005) A Detection method of multivariate outliers using decompositions of the squared mahalanobis distance. *Journal of the Korean Data Analysis Society* 7, 1935-1953.
8. Kim M. G. and Jung, K. M. (2005) Detection of outliers in multivariate

regression using plug-in method. *Journal of the Korean Data Analysis Society* 7, 1117-1124.

9 Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate analysis*. Academic Press, New York.

10. Park, H. I. (2007) A nonparametric test procedure for multivariate censored data. *Journal of the Korean Data Analysis Society* 9, 1589-1599.

11. Park, H. I., Na, J. H. and Desu, M. M. (2001) Non-parametric one-sided tests for multivariate data. *Sankhya* Ser. B 63, 286-297.

12. Pesarin, F. (2001) *Multivariate permutation tests with applications in Biostatistics*. Wiley, New York.

13. Puri, M. L. and Sen, P. K. (1971) *Nonparametric methods in multivariate analysis*. Wiley, New York.

14. Puri, M. L. and Sen, P. K. (1985) *Nonparametric methods in general linear model*. Wiley, New York.

15. Um, Y. (2005) A stratified agreement measure among multiple raters for multivariate interval data. *Journal of the Korean Data Analysis Society* 7, 1125-1132.

16. Wei, L. J. and Knuiman, M. W. (1987) A one-sided rank test for multivariate censored data. *Australian Journal of Statistics* 29, 214-219.

17 Westfall, P. H. and Young, S. S. (1993) *Resampling-based multiple testing*. Wiley, New York.

[received date : Oct. 2007, accepted date : Nov. 2007]