Journal of the Korean Data & Information Science Society
2010, **21**(2), 291–296

한국데이터정보과학회지

Some versatile tests based on percentile tests†

Hyo-Il Park[1] · Ju Sung Kim[2]

[1]Department of Statistics, Chong-ju University
[2]Department of Informational Statistics, Chungbuk National University

Received 20 October 2009, revised 7 January 2010, accepted 8 February 2010

Abstract

In this paper, we consider a versatile test based on percentile tests. The versatile test may be useful when the underlying distributions are unknown or quite different types. We consider two kinds of combining functions for the percentile statistics, the quadratic and summing forms and obtain the limiting distributions under the null hypothesis. Then we illustrate our procedure with an example. Finally we discuss some interesting features of the test as concluding remarks.

Keywords: Nonparametric test, percentile test, permutation principle, two-sample problem, versatile test.

1. Introduction

When taking comparison study between two kinds of treatments, one may use two- sample t-test under the normality assumption or a nonparametric test only by assuming continuity of the underlying distributions under some specific model. There have been proposed several nonparametric tests for the two-sample problem (Randles and Wolfe, 1979) and for the consideration of the power of test, one may choose suitable one according as the underlying distribution. For this direction, one may apply the Wilcoxon rank sum test for the logistic and the median test, for the double exponential under the location translation model. Park and Kim (2009) considered testing problem for the multi- sample case under the location translation model. Also Hong and Kim (2009) applied the nonparametric test for the credit rating model validation. The location translation implies that the difference between quantile points of the two distribution functions remains same all the time. However this assumption may be violated when the effect of some new treatment may appear initial stage of treatment but disappears as time goes or may appear after some time passes by. Therefore in this case, the nonparametric test procedure based on the location translation model may incur some loss of power.

Several authors have been considered some simultaneous use of several nonparametric test statistics when the underlying distributions are not certain. For example, as we have already

† This work was supported by the grant of the Chungbuk National University in 2009.
[1] Professor, Department of Statistics, Chong-ju University, Chong-ju 360-764, Korea.
[2] Corresponding author: Professor, Department of Informational Statistics, Chungbuk National University, Chong-ju 361-763, Korea. E-mail: kimjs@chungbuk.ac.kr

mentioned, Wilcoxon rank sum and median tests are locally most powerful for the logistic and double exponential distributions, respectively. Therefore when one is not sure about the underlying distribution, one may apply both the statistics together using some combining function. In this direction, Lee (1996), Chi and Tsai (2001) and Wu and Gilbert (2002) proposed several nonparametric testing procedures for the survival data. Also Fleming and Harrington (1991) considered extensively the use and efficiency of these test procedures. They termed this test procedure as the versatile test.

In this study, we consider a test procedure based on the percentile test statistics in the spirit of the versatile test. In the next section, we introduce the test statistics with quadratic and summing forms for combining functions for the quantile functions. Also we derive the limiting distribution under the null hypothesis for obtaining the critical value or p -value. For obtaining the p -value, we also consider the permutation principle (Good, 2001). Then we illustrate our procedure with an example and discuss some interesting features of our test as concluding remarks.

2. Some versatile test

Suppose that we have two independent random samples, $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ with continuous distribution functions F and G, respectively. For any p, $0 < p < 1$, let $r_p = [Np] + 1$, where $N = m + n$ and $[x]$ means the largest integer which does not exceed the real number x. Also let $R_1, \ldots, R_m$ be the ranks of $X_1, \ldots, X_m$ from the combined sample of $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ together. Then a p th percentile test statistic $T_N(p)$ for testing $H_0 : F = G$ can be defined as

$$T_N(p) = \sum_{i=1}^{m} I(R_i \leq r_p),$$

where $I(\bullet)$ is an indicator function. We note that when $p = 1/2$, $T_N\,(1/2)$ is the well- known Mood type median test statistic (Mood, 1950). Let ξ_p be the p th quantile of F. Then roughly speaking, since the statistic $T_N(p)/m$ can be considered as a consistent estimate of $F(\xi_p)$ under $H_0 : F = G$, the test based on $T_N(p)$ may be effective for detectection of the difference between the two p th quantiles of F and G. Therefore when the difference in quantiles of F and G seems to appear in the early stage of experiment, one may choose p such as $0 < p < 1/2$ and $1/2 < p < 1$ when it does to appear in the later stage of experiment. If one may not be sure when the difference happens, one may try to use several p's simultaneously to ensure to detect any difference. With this in mind, for a versatile test, we may choose d number of p_j's such that $0 < p_1 < \ldots < p_d < 1$ and make use of them simultaneously for testing $H_0 : F = G$. For this purpose, let $T_N(p_j)$ be the p_j th quantile test statistic, $j = 1, \ldots, d$. Since $T_N(p_j)$ is a linear rank statistic (Randles and Wolfe, 1979) for each j, $j = 1, \ldots, d$, one may obtain easily the expectation and variance of $T_N(p)$ under $H_0 : F = G$, as

$$E_0\left(T_N(p_j)\right) = m\frac{r_{p_j}}{N} \text{ and } V_0\left(T_N(p_j)\right) = \frac{mn}{N-1}\frac{r_{p_j}(N - r_{p_j})}{N^2}.$$

Also for $j < k$, we need the covariance between $T_N(p_j)$ and $T_N(p_k)$ under $H_0 : F = G$, which can be obtained as follows:

Lemma 1. Under $H_0 : F = G$, we have the covariance between $T_N(p_j)$ and $T_N(p_k)$ as

$$COV_0\left(T_N(p_j), T_N(p_k)\right) = \frac{mn}{N-1}\frac{r_{p_j}(N - r_{p_k})}{N^2}.$$

Proof. First of all, we note that for any pair j and k such that $1 \leq j < k \leq d$, we have

$$\left\{\sum_{i=1}^{m} I(R_i \leq r_{p_j})\right\}\left\{\sum_{i=1}^{m} I(R_i \leq r_{p_k})\right\} = \sum_{i=1}^{m} I(R_i \leq r_{p_j}) + \sum\sum_{i \neq l} I(R_i \leq r_{p_j}) I(R_l \leq r_{p_k}).$$

Also we note that

$$E_0\left(I(R_1 \leq r_{p_j})\right) = \Pr_0\left\{R_1 \leq r_{p_j}\right\} = \frac{r_{p_j}}{N}$$

and $E_0\left(I(R_1 \leq r_{p_j}) I(R_2 \leq r_{p_k})\right) = \Pr_0\left\{R_1 \leq r_{p_j}, R_2 \leq r_{p_k}\right\} = \dfrac{r_{p_j}\left(r_{p_k} - 1\right)}{N(N-1)}$.

Then from the fact

$$COV_0\left(T_N(p_j), T_N(p_k)\right) = E_0\left(T_N(p_j) T_N(p_k)\right) - E_0\left(T_N(p_j)\right) E_0\left(T_N(p_k)\right)$$

we obtain the result by manipulating algebraically.

Now we consider the standardized form for each j as follows:

$$Z_N(p_j) = \frac{T_N(p_j) - E_0\left(T_N(p_j)\right)}{\sqrt{V_0\left(T_N(p_j)\right)}}.$$

It is well-known that for each j, the distribution of $Z_N(p_j)$ converges in distribution to the standard normal distribution. Then there are several ways to combine the d number of standardized form of the univariate percentile statistics. If there is no more information about F and G in advance, then it would be appropriate to consider the general or two-sided alternative $H_1 : F \neq G$. This in turn makes us to consider a quadratic form. Then we note that the covariance between $Z_N(p_j)$ and $Z_N(p_k)$ under $H_0 : F = G$ for $j < k$ is of the form

$$COV_0\left(Z_N(p_j), Z_N(p_k)\right) = \frac{r_{p_j}(N - r_{p_k})}{\sqrt{r_{p_j} r_{p_k}(N - r_{p_j})(N - r_{p_k})}} = \rho_{jkN}.$$

Also let P_N be the covariance matrix whose component is 1 if $j = k$ and ρ_{jkN}, $1 \leq j < k \leq d$ and $\boldsymbol{Z}_N = \left(Z_N(p_1), \ldots, Z_N(p_d)\right)^T$, where $()^T$ means the transpose of a vector or matrix. Then we can propose a versatile test statistic for testing $H_0 : F = G$ against $H_1 : F \neq G$ with the assumption that P_N is positive definite as

$$Q_N(\boldsymbol{p}) = \boldsymbol{Z}_N^T \mathbf{P}_N^{-1} \boldsymbol{Z}_N,$$

where P_N^{-1} is the inverse of P_N. Then for large value of $Q_N(\boldsymbol{p})$, we may reject $H_0 : F = G$ in favor of $H_1 : F \neq G$. For any given significance level α, in order to have the critical value $C_N(\alpha)$ or more generally, p -value, we should have the null distribution of $Q_N(\boldsymbol{p})$. This can be done by applying the permutation principle for small or reasonable sample sizes.

However for large sample case, one can obtain the asymptotic distribution using the large sample approximation. We begin this matter by showing that ρ_{jkN} converges to ρ_{jk}, where

$$\rho_{jk} = \sqrt{\frac{p_j}{p_k}\frac{1-p_k}{1-p_j}}.$$

□

Lemma 2. The covariance ρ_{jkN} converges to ρ_{jk}.
Proof. First of all, we note that

$$\rho_{jkN} = \frac{r_{p_j}(N - r_{p_k})}{\sqrt{r_{p_j} r_{p_k}(N - r_{p_j})(N - r_{p_k})}} = \sqrt{\frac{r_{p_j}(N - r_{p_k})}{r_{p_k}(N - r_{p_j})}}.$$

Then the result follows easily by noting that $r_{p_j}/N \to p_j$ and $r_{p_k}/N \to p_k$.
Now we have our main result in the following theorem. □

Theorem 1. With the assumption that $m/N \to \lambda$ as $\min\{m, n\} \to \infty$ and assuming that P_N is positive definite, $Q_N(\boldsymbol{p})$ converges in distribution to a chi-square distribution with d degrees of freedom under $H_0 : F = G$.

Proof. For each component j, it is well-known that under $H_0 : F = G$, $Z_N(p_j)$ converges in distribution to a standard normal distribution with the assumption that $m/N \to \lambda$ as $\min\{m, n\} \to \infty$. Therefore from the Cramer-Wold device (Mardia *et al.*, 1979) and Slutsky's theorem with Lemma 2 and the assumption that P_N is positive definite, we have the result.

If someone believes that $F(x) \geq G(x)$ for all $x \in \boldsymbol{R}^1$, then one may consider different type of statistics rather than the quadratic form itself. Also the corresponding alternative would be

$$H_1 : F(x) \geq G(x) \ \ with \ \ strict \ \ inequality \ \ for \ \ some \ \ x \ \ \in \boldsymbol{R}^1. \tag{2.1}$$

The alternative (2.1) is one-sided. Therefore it would be better to consider the following form $S_N(\boldsymbol{p})$ for the test statistic rather than the quadratic form.

$$S_N(\boldsymbol{p}) = \sum_{j=1}^{d} Z_N(p_j).$$

Since if (2.1) is true, then $Z_N(p_j)$ tends to have large positive value for all j, for large values of $S_N(\boldsymbol{p})$ $H_0 : F = G$ would be rejected in favor of the alternative (2.1). For the completion of test procedure, one needs the null distribution of $S_N(\boldsymbol{p})$. For small or reasonable sample sizes, one may easily obtain the exact null distribution by applying the permutation principle. Park and Kim (2008) applied the permutation principle for testing multivariate data. For the large sample case, we may obatin the limiting distribution from Theorem 1 as follows.
□

Theorem 2. With the assumption that $m/N \to \lambda$ as $\min\{m, n\} \to \infty$ $S_N(\boldsymbol{p})$ converges in distribution to a normal random variable with mean 0 and variance σ_S^2 under $H_0 : F = G$,where

$$\sigma_S^2 = d + 2\sum\sum_{1 \leq j < k \leq d} \rho_{jk}.$$

Proof. This result follows easily from the proof of Theorem 1 by taking a vector $\boldsymbol{l} = (1, \ldots, 1)^T$ during application of Cramer-Wold device. □

3. An example and concluding remarks

The following data set in Table 1 is the perceived degree of job satisfaction measured by a proper psychological index consisting of a sum of a finite sub-responses each related to a specific sub-aspect among individuals classified as extroverted (X) and introverted (Y) in Pesarin (2001). Here $m = 12$, $n = 8$ and $N = 20$. Then it is of interest to test whether there exists any difference between the two groups, X and Y .

Table 3.1 Job-satisfaction

X	66 57 81 62 61 60 73 59 80 55 67 70
Y	64 58 45 43 37 56 44 42

For this we consider the quartile points such as the first (p_1), second (or median) (p_2) and third (p_3). In other words, let $p_1 = 0.25$, $p_2 = 0.5$ and $p_3 = 0.75$ be chosen. Then $r_{0.25} = 5$, $r_{05} = 10$ and $r_{.75} = 15$. Thus we have that $T_{20}(.25) = 0$, $T_{20}(.5) = 2$ and $T_{20}(.75) = 7$. Also we have that

$$E_0(T_{20}(.25)) = 3, E_0(T_{20}(.5)) = 6 and E_0(T_{20}(.75)) = 9$$

$$V_0\left(T_{20}(.25)\right) = \frac{18}{19}, V_0\left(T_{20}(.5)\right) = \frac{24}{19} and V_0\left(T_{20}(.75)\right) = \frac{18}{19}$$

$$COV_0\left(T_{20}(.25), T_{20}(.5)\right) = \frac{12}{19}, COV_0\left(T_{20}(.5), T_{20}(.75)\right) = \frac{12}{19}\ and$$

$$COV_0\left(T_{20}(.25), T_{20}(.75)\right) = \frac{6}{19}.$$

Thus we have that

$$Z_{20}(.25) = \frac{-3}{\sqrt{18/19}} = -\sqrt{\frac{19}{2}}, Z_{20}(.50) = \frac{-4}{\sqrt{24/19}} = -\sqrt{\frac{38}{3}} Z_{20}(.75) = \frac{-2}{\sqrt{18/19}} = -\sqrt{\frac{38}{9}}$$

and $Q_{20}(\boldsymbol{p}) = 14.25$ and $S_{20}(\boldsymbol{p}) = -3.557$. Then the respective p -values are 0.0026 and 0.0004, which show the strong evidence of difference between two groups. We note that we considered two-sided test for the test procedure based on $S_{20}(\boldsymbol{p})$.

For the null distribution of test statistics, mainly we have considered the limiting distribution with large sample approximation theory. Also one may consider applying the permutation principle (Good, 2000). In this case, it is customary that one may take the Monte-Carlo approach for obtaining p-values because of the excessive computational burdens when one tries to obtain the exact probabilities. Therefore the permutation principle relies heavily on the computer capability. For the one-sided test, one may use the maximal type of statistic

$$M_N(\boldsymbol{p}) = \max\left\{Z_N(p_1), \ldots, Z_N(p_d)\right\}$$

or Reiny type of statistic (Shorack and Wellner, 1986)

$$R_N(\boldsymbol{p}) = sup_{0<p<1} \left\{ itp : Z_N(p) \right\}.$$

However the limiting distributions for these statistics are not easily available since that of $M_N(\boldsymbol{p})$ is d -variate normal and that of $R_N(\boldsymbol{p})$, non-normal. Then in this case, it is recommendable to apply the permutation principle to obtain the p -value. We have obtained an approximate p -value for $M_N(\boldsymbol{p})$ as $\Pr\left\{ M_{20}(\boldsymbol{p}) \leq -\sqrt{38/3} \right\} = 0.00000079$ by applying the permutation principle.

For the re-sampling method for the hypothesis test, one may also consider the bootstrap method. While the permutation method re-samples without replacement, the bootstrap method does with replacement. However the results between two methods are known to be considerably different (Good, 2000).

References

Chi, Y. and Tsay, M. H. (2001). Some versatile tests based on the simultaneous use of weighted logrank and weighted Kaplan-Meier statistics. *Communications in Statistics: Simulation and Computation*, **30**, 743-759.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*, Wiley & Sons, Inc., New York.

Good, P. (2000). *Permutation tests-A practical guide to resampling methods for testing hypotheses*, Wiley & Sons, Inc., New York.

Hong, J. S. and Kim, J. H. (2009). Nonparametric homogeneity tests of two distributions for credit rating model validation. *Journal of the Korean Data & Information Science Society*, **20**, 261-272.

Lee, J. W. (1996). Some versatile tests based on the simultaneous use of weighted log- rank statistics. *Biometrics*, **52**, 721-725.

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

Park, H. I. and Kim, J. S. (2008). Permutation tests for multivariate data. *Journal of the Korean Data & Information Science Society*, **18**, 1145-1155.

Park, H. I. and Kim, J. S. (2009). Some nonparametric test procedure for the multi- sample case. *Journal of the Korean Data & Information Science Society*, **20**, 237-250.

Pesarin, F. (2001). *Multivariate permutation tests*, Wiley & Sons, Inc., New York.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics*, Wiley & Sons, Inc., New York.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical processes with applications to statistics*, Wiley & Sons, Inc., New York.

Wu, L. and Gilbert, P. B. (2002). Flexible weighted log-rank tests optimal for detecting early and/or late survival differences. *Biometrics*, **58**, 997-1004.