Scientific Research

New Nonparametric Rank-Based Tests for Paired Data

Guogen Shan

Department of Environmental and Occupational Health, Epidemiology and Biostatistics Program, University of Nevada Las Vegas, Las Vegas, NV, USA
Email: guogen.shan@unlv.edu

Abstract

We propose a new nonparametric test based on the rank difference between the paired sample for testing the equality of the marginal distributions from a bivariate distribution. We also consider a modification of the novel nonparametric test based on the test proposed by Baumgartern, Weiβ, and Schindler (1998). An extensive numerical power comparison for various parametric and nonparametric tests was conducted under a wide range of bivariate distributions for small sample sizes. The two new nonparametric tests have comparable power to the paired t test for the data simulated from bivariate normal distributions, and are generally more powerful than the paired t test and other commonly used nonparametric tests in several important bivariate distributions.

1. Introduction

Paired data are very common in statistical and medicinal research. A typical example is a clinical trial where subjects are measured prior to a treatment, say for elevated systolic blood pressure, and then measured again after the treatment with a drug to lower the blood pressure. Another example is the use of matched cases and controls. One sample from the the case group and another matched sample from the control group may be used to form a paired sample by using additional variables that are measured in addition to the variable of interest. Paired data are often used to reduce variability and to make more precise comparisons with fewer subjects, and this has resulted in attracting many statisticians to develop more efficient tests and inferences for paired data.

Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be n random samples from a bivariate distribution with continuous end-

points. The marginal distributions of X and Y follow $F_X(x)$ and $F_Y(y)$, respectively. The null hypothesis of interest is $H_0 : F_X = F_Y$. This problem often occurs in applied research for testing the equality of the marginal distributions. For example, in a one-arm Oncolgy study, the tumor size of each patient is measured before and after treatment. If the cancer treatment is effective on patients, the tumor sizes in the majority of patients are expected to be smaller after the treatment than the baseline measurement. Therefore, an appropriate alternative hypothesis is given as $H_a : F_X \leq F_Y$ with at least one point z such that $F_X(z) < F_Y(z)$. One important typical case in the above problem is the location problem, that is, $F_Y(x) = F_X(x + \theta)$ for all x, where $\theta > 0$. The X distribution has a positive shift compared to that of the Y distribution.

The two sample paired t test is a commonly used parametric approach for comparing the means of two distributions. It computes the difference between the two measurements of each subject $D_i = X_i - Y_i, i = 1, 2, \cdots, n$, and then tests whether the average of these differences is significantly different from zero by using the test statistic

$$PT = \frac{\mu_D}{s_D / \sqrt{n}}, \tag{1.1}$$

where μ_D and s_D are the sample mean and the standard deviation, respectively. The two sample t test makes certain assumptions, such as the normality of the sample difference which needs to be checked by normality tests [1] [2] before applying the paired t test. If one or more of these assumptions can't reasonably be met, then the paired t test may be not appropriately applied.

An alternative to the two sample paired t test is the Wilcoxon signed rank (WSR) test [3], which is a commonly used nonparametric test for paired data when at least one of the assumptions is not satisfied. The Wilcoxon rank sum test (also known as the Mann-Whitney test) [3] [4] is a nonparametric statistical test for assessing whether the two independent samples are from the same distribution. It may be not be suitable for testing paired data without some modification. Later, Lam and Longnecker [5] proposed a modification of the Wilcoxon rank sum (MWRS) test by introducing a consistent variance estimator for assessing the equality of the marginal distributions of a bivariate distribution. The MWRS test was compared to other tests based on Monte Carlo simulation with small sample sizes, and was shown to be as powerful as the two sample paired t test for the bivariate normal data, and more powerful than both the two sample paired t test and the WSR test for the Farlie-Gumbel-Morgenstern distribution with exponential marginals. We propose a new rank difference (RD) test for paired data based on the rank difference between the paired sample to capture the sample difference. We also introduce the modified Baumgartern, Weiβ, and Schindler (MBWS) test proposed by Shan *et al.* [6] for paired data. A discussion on choosing between the parametric and nonparametric tests may be found in Fay and Proschan [7].

The remainder of this article is organized as follows. In Section 2, we briefly review the two existing nonparametric tests for paired data and introduce the two new nonparametric tests. In Section 3, we compare the performance of the competing tests, studying the simulated power of the tests under a wide range of bivariate distributions. A real example is given to illustrate the application of the parametric and nonparametric tests in Section 4. Section 5 is given to discussion.

2. Nonparametric Tests

A nonparametric counterpart to the two sample paired t test is the WSR test for paired samples. The WSR test begins by transforming each difference D_i into its absolute value $|D_i|$, then the absolute differences are ranked from the lowest to the highest $R_i = Rank(|D_i|)$. For continuous endpoints, there is no tie between measurement, and all D_i's are used in the ranking precess. The WSR test statistic is then expressed as

$$\text{WSR} = \left| \sum_{i=1}^{n} sign(D_i) R_i \right|. \tag{1.2}$$

The value of the WSR test statistic is a non-negative integer between 0 and $n(n+1)/2$. The upper bound would be reached when all signed values are either positive or negative. The standardized WSR test statistic $\dfrac{\text{WSR} - 1/2}{\sqrt{n(n+1)(2n+1)/6}}$ asymptotically follows a standard normal distribution. The asymptotic distribution can be

used to calculate the p-value and to find the threshold values. But, for small sample sizes, the exact distribution of the WSR test provides accurate and reliable results. The exact sampling distribution of the WSR test can be obtained by enumerating all possible combinations of the positive and negative signs. For example, if we have n subjects in the study, then the absolute differences, $|D_i|, i = 1, 2, \cdots, n,$ produce the order of ranks $1, 2, \cdots, n$. All possible combinations of plus and minus signs that could be distributed among these ranks are 2^n. Then, the exact p-value of a given data is the proportion the combinations whose WSR test statistic is as extreme as that of the given data.

Another nonparametric test considered is the MWRS test proposed by Lam and Longnecker [5] for assessing the equality of the marginal distributions of a bivariate distribution. Let S_i and T_i denote the rank for X_i and Y_i in the combined sample, and U_i be the rank for X_i in the X sample and V_i be the rank for Y_i in the Y sample. Then the MWRS test is defined as

$$\text{MWRS} = \sqrt{2n+1}\left\{\frac{\dfrac{1}{n(2n+1)}\sum_{i=1}^{n}S_i - \dfrac{1}{2}}{\sigma_W}\right\}, \tag{1.3}$$

where $\sigma_W = \sqrt{(1-\hat{\rho}_S)}$, $\hat{\rho}_S = \dfrac{12}{n(n^2-1)}\sum_{i=1}^{n}U_iV_i - \dfrac{3(n+1)}{n-1}$ is the Spearman's coefficient of rank correlation.

The asymptotic distribution of MWRS is a standard normal distribution due to the consistency of the variance estimator [5]. The MWRS test was shown to have comparable power to paired t test and the WSR test.

Two Proposed Nonparametric Tests

Two steps are implemented in the Wilcoxon signed rank test: calculation of the absolute difference followed by the ranking of these differences. The new proposed RD test calculates the test statistic by revising the order the the two steps in the WSR test: ranking the observations followed by the difference of the ranks. Specifically, the associated test statistic of the RD test is

$$RD = \sum_{i=1}^{n}\left(S_i - T_i\right). \tag{1.4}$$

The value of the RD test statistic is an integer between $\dfrac{-(n+1)(2n+1)}{2}$ and $\dfrac{(n+1)(2n+1)}{2}$, which in-

cludes the sample space of the WSR test. A larger sample space could potentially have a less discrete type I error rate in studies with small to medium sample sizes. The sign of $S_i - T_i$ is the same as that of $sign(D_i)R_i$ in the WSR test. The new proposed RD test captures not only the difference within each subject, but also the rank of the observations within each subject.

Recently, Baumgartern, Weiß, and Schindler (BWS) [8] proposed a novel nonparametric test for two independent sample problem, which is based on the squared value of the difference between the two empirical distribution functions weighted by the respective variance. This weighting places more emphasize on the tails of the distribution functions. This new test is not suitable for a one sided problem due the nature of the construction of the test statistic. For this reason, Neuhauser [9] proposed a modified BWS test using the sign of the difference of the rank and the mean of the rank to enable the one sided problem. It was then further modified by Shan *et al.* [6] with the exact mean and variance estimates of ranks [10] for an one sided two independent sample problem. We consider this MBWS test [6] for paired data, and the test statistic is of the form

$$\text{MBWS} = \frac{1}{2}\left(B_X - B_Y\right), \tag{1.5}$$

where

$$B_X = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(S_i - \dfrac{2n+1}{n+1}i\right)\left|S_i - \dfrac{2n+1}{n+1}i\right|}{\dfrac{i}{n+1}\left(1 - \dfrac{i}{n+1}\right)\dfrac{n(2n+1)}{n+2}}$$

and

$$B_Y = \frac{1}{n} \sum_{j=1}^{n} \frac{\left(T_j - \frac{2n+1}{n+1} j\right) \left| T_j - \frac{2n+1}{n+1} j \right|}{\frac{j}{n+1}\left(1 - \frac{j}{n+1}\right)\frac{n(2n+1)}{n+2}}.$$

Although the asymptotic distribution of the test statistic for the MBWS test may not be easily derived, an exact permutation test or a simulation based test can readily be performed in order to calculate the p-value for a given data set. It should be note that all the nonparametric procedures aforementioned can be used for data with or without ties; in the case of ties the ranks are defined to be the midranks.

3. Numerical Study

To evaluate the performance of the parametric and nonparametric test, sample size $n = 10$, significance level of $\alpha = 0.05$ and 20,000 simulated iterations were used in the Monte Carlo exact simulation. Five different tests were competed for each plot: 1) the RD test; 2) the MBWS test; 3) the MWRS test; 4) the WSR test; and 5) the two sample paired t test. The two sample paired t test is the only parametric test in this article, and all the other four tests are nonparametric approaches. Four difference bivariate distributions were examined: 1) the bivariate normal distribution; 2) the bivariate distribution with gamma marginal distributions; 3) the bivariate generalized exponential distribution; and 4) the bivariate distribution with a gamma and a exponential marginal distributions.

The first considered bivariate distribution is a bivariate distribution with mean $(\theta_1, \theta_2)'$ and variance covariance matrix $\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$, where ρ is the correlation coefficient, $0 \le \rho \le 1$. **Figure 1** shows the power plots for the bivariate normal distribution of different means with a fixed covariance matrix. Equal variances are assumed $\sigma_1 = \sigma_2 = \sigma$, and four different ρ values are considered in the figure: 0, 0.2, 0.4, and 0.7. The 95% threshold value was simulated from the bivariate normal distribution with $\theta = \theta_1 - \theta_2 = 0, \sigma = 1$, and a given ρ for each plot in **Figure 1**. As seen, the simulated power of each test is an increasing function of θ/σ. The two sample paired t test is the most powerful test as expected due to the fact that this is the uniform most powerful unbiased test for this problem when the data is from the bivariate normal distribution of different means for a given covariance matrix. The new proposed RD test and the MBWS test are compatible with regard to the power, and both are generally more powerful than the WSR test. The MBWS test has greater power than the MWRS test for a small to medium ρ, and the RD test is generally more powerful than the MWRS test. Given a large ρ, the MWRS could be more powerful than the proposed MBWS test, but less powerful than the RD test. **Figure 2** shows the power plots of the correlation coefficient ρ given equal variances $\sigma_1 = \sigma_2 = \sigma = 1$ and the ratio of mean difference and variance $\theta/\sigma = 0.3$. Similar results are observed as the results from **Figure 1**. It should be noted that the paired t test is only appropriate when the difference D_i follows a normal distribution. The other four tests considered in this article are nonparametric approaches that are applicable to any continuous distributions with fewer assumptions.

We also compare the bivariate normal distribution with equal means but different variances given the same covariance $\rho\sigma_1\sigma_2 = 0.6$. The power plots as a function of σ_2^2/σ_1^2 are shown in **Figure 3**. The threshold value is simulated from a bivariate normal distribution with equal variances. The paired t test, the WSR test, and the MWRS test appear to have less power than the two new proposed tests. The MBWS test is clearly more powerful than the other proposed RD test. The two new proposed tests are able to detect the variance change in the distribution, while others do not.

In addition to the bivariate normal distribution, we also consider other bivariate distributions. One example is the bivariate distribution with gamma marginal distributions $G(\kappa, \eta)$, where κ and η are the shape and scale parameters, respectively. The data may be generated from the function *rmvdc* in the R package *copula*. The two marginal gamma distributions with the same scale parameter but different shape parameters are considered, *i.e.*, $G(\kappa_1, 1)$ and $G(\kappa_2, 1)$. **Figure 4** shows the power plot as a function of the ratio of the shape parameters κ_1/κ_2. The two proposed tests have the highest power, followed by the MWRS test, the WSR test, and the paired t test. The two new proposed tests dominate other tests and the power gains are substantial.
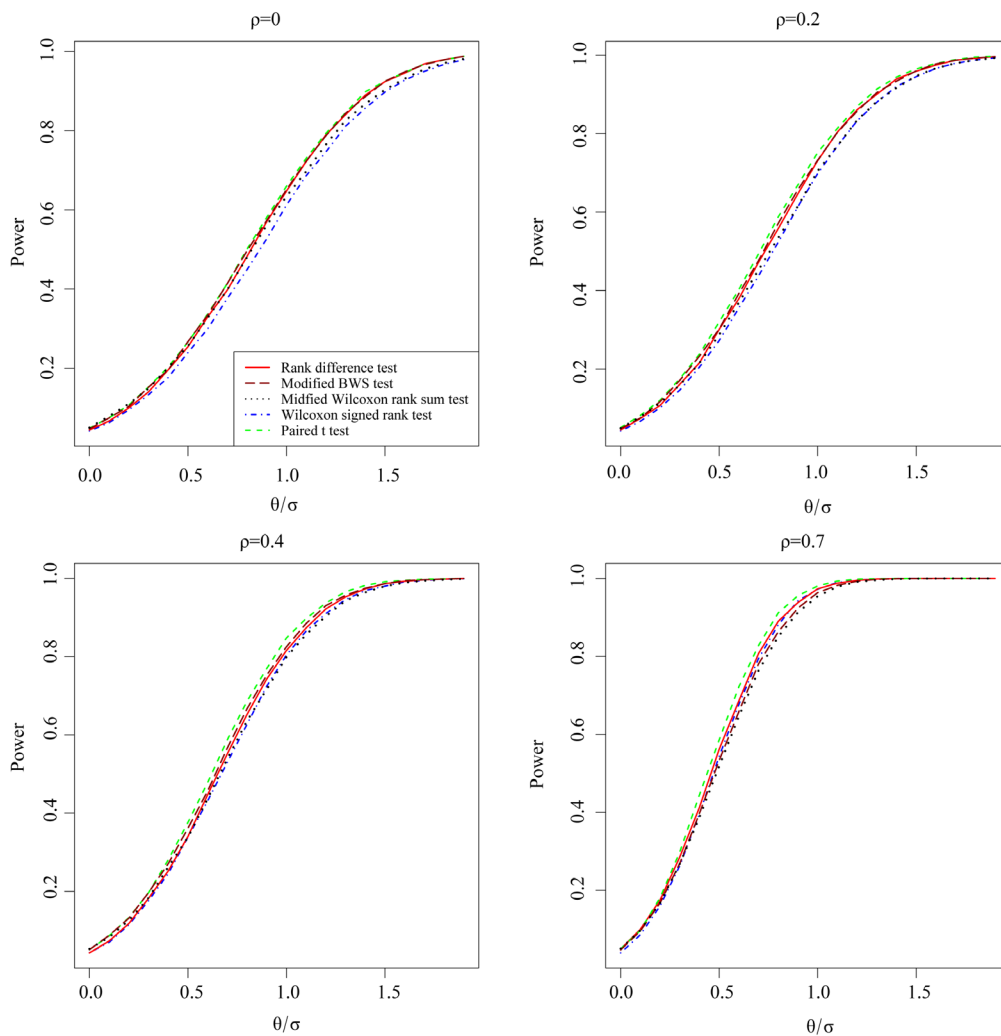
Figure 1. Power study for a bivariate normal distribution with difference mean given four different covariance matrices.
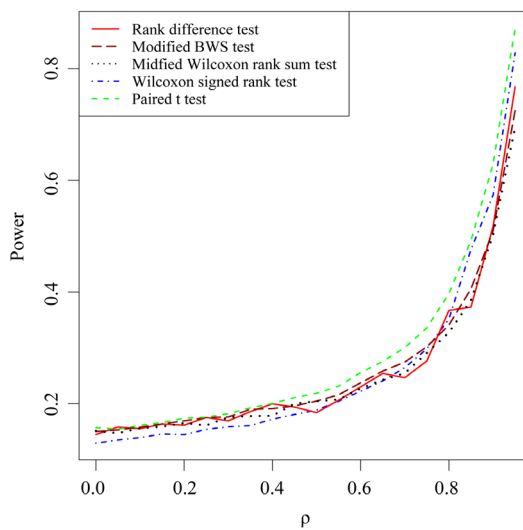

Figure 2. Power study for a bivariate normal distribution with the same equal variances and the ratio of mean difference and variance but different ρ.
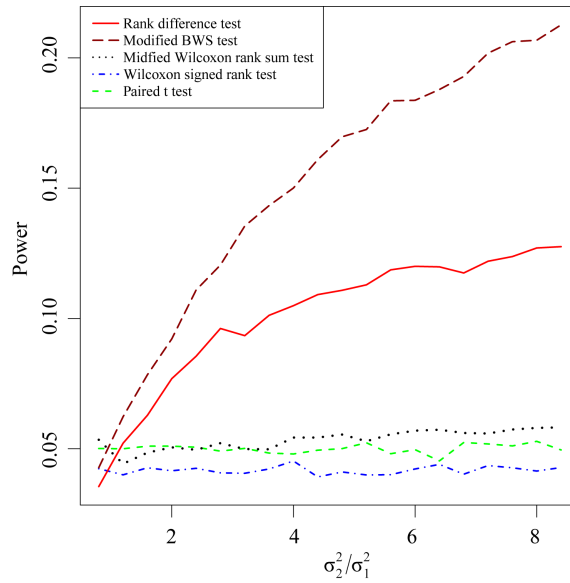
Figure 3. Power study for a bivariate normal distribution with the same mean but different variances σ_1, σ_2 given the covariance 0.6.
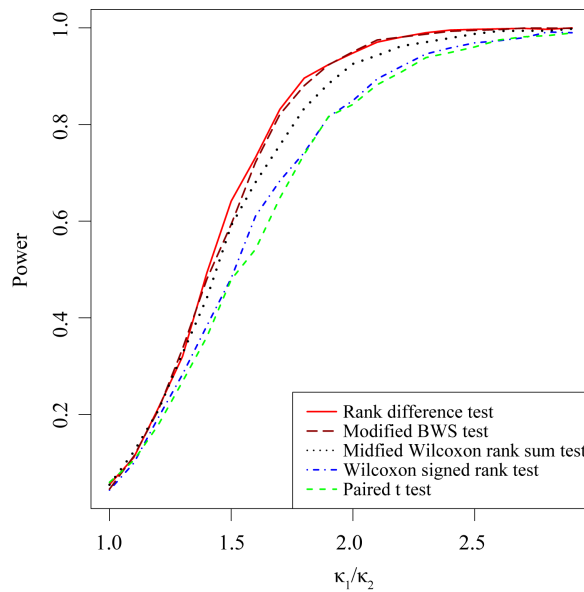

Figure 4. Power study for a bivariate distribution with gamma marginal distributions, $G(\kappa_1, 1)$ and $G(\kappa_2, 1)$.

Another bivariate distribution examined here is the bivariate generalized exponential distribution [11] with the joint cumulative distribution function

$$P(X < x, Y < y) = \left(1 - e^{-x}\right)^{\tau_1} \left(1 - e^{-y}\right)^{\tau_2} \left(1 - e^{-\min(x,y)}\right)^{\tau_3},$$

where $\tau_1, \tau_2,$ and τ_3 are the three parameters in the distribution. The marginal distributions for X and Y are generalized exponential distributions with parameters $(\tau_1 + \tau_3, 1)$ and $(\tau_2 + \tau_3, 1)$, respectively. The third parameter in the generalized exponential distribution is given as $\tau_3 = 1$ in the simulation study. The null distribution is simulated with equal τ_1 and τ_2, *i.e.*, $\tau_1 = \tau_2 = 1$. The power plot is drawn as a function of τ_1/τ_2, see **Figure 5**. The signed rank test is very lower in power as compared to other procedures; the two new proposed tests are not as powerful as the paired t test and the MWRS test.
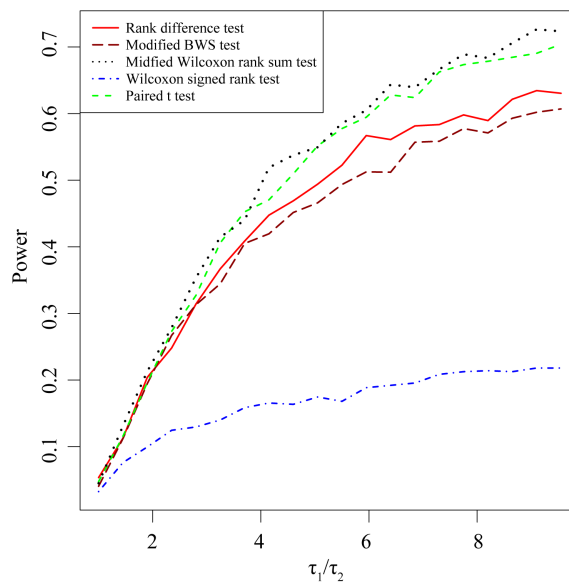
Figure 5. Power study for a bivariate generalized exponential distribution with parameters $(\tau_1, \tau_2, \tau_3 = 1)$.

For further comparison, we examined the bivariate distribution with different types of marginal distributions, for example, one marginal distribution follows a gamma distribution $G(\kappa, 1)$ and the other is an exponential distribution $\mathrm{Exp}(1)$. Equal mean is assumed under the null hypothesis with $\kappa = 1$ in the gamma distribution. The power plots as a function of κ are displayed in **Figure 6**. The paired t test and the WSR test are less powerful than the other three tests. The proposed MBWS test is generally more powerful than the MWRS test under large κ alternatives.

4. Example

We consider an example and apply the five different tests discussed in this article: 1) the paired t test; 2) the WSR test; 3) the MWRS test; 4) the RD test; and 5) the MBWS test. Suppose a pharmaceutical company wants to assess the efficacy of a drug in lowering systolic blood pressure. The systolic blood pressure reading in mmHg for 10 subjects were measured before and after the administration of the drug, and the associated data can be found in Antonisamy *et al.* [12]. The systolic blood pressure is expected to be lower after the drug treatment, therefore a one sided alternative is appropriate for this study. The p-value of the WSR test was calculated based on the exact permutation approach, the p-value of the paired t test was computed using the asymptotic approach, and the p-values of all the other three nonparametric tests were calculated based on the 100,000 Monte Carlo exact simulation. The p-values are reported in **Table 1**. All five tests conclude that the drug is effective in lowering the systolic blood pressure at the significance level of 0.05.

5. Conclusion

In this article, we introduce two new nonparametric tests for testing whether paired samples come from the same population. The two new proposed nonparametric tests are comparable to the paired t test for testing the mean difference for the bivariate normal distribution given a covariance matrix, and much more powerful than the paired t test and another two nonparametric tests for the difference in variances for the bivariate normal distribution. Extensive numerical power comparison was conducted for various other important bivariate distributions. The proposed RD test and the MBWS test have greater power than other tests in several important scenarios, and the power gains are substantial. These two proposed tests are recommended for use in practice due the power gains as compared to other competitors. One limitation of the MBWS test is the difficulty to find the asymptotic distribution. However, permutation-based or simulation-based tests can always be used for the p-value calculation. We consider exact testing procedures as future work [13]-[20]. The extension of the RD test
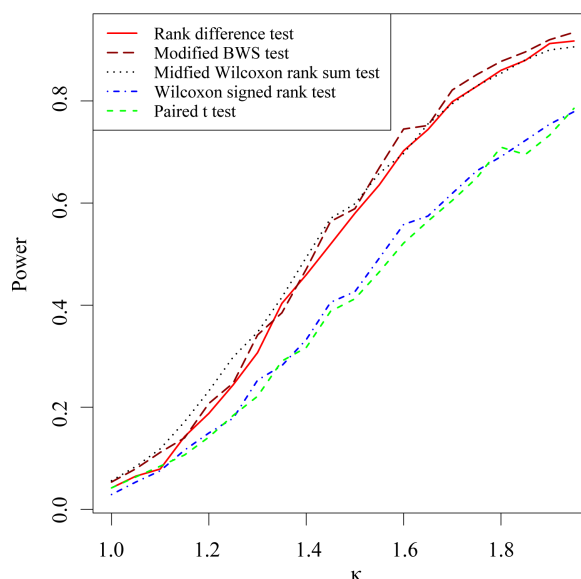
Figure 6. Power study for a bivariate distribution with $G(\kappa,1)$ and Exp(1) as marginal distributions.

Table 1. p-values for the example from the systolic blood pressure study.

Tests				
Paired t test	WSR	MWRS	RD	MBWS
0.0104	0.0049	0.0005	0.0149	0.0145

and the MBWS test to the *k*-sample independent and dependent problems [21]-[24] is currently underway.

Acknowledgements

References

[1] Shapiro, S.S. and Wilk, M.B. (1965) An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, **52**, 591-611. http://dx.doi.org/10.2307/2333709

[2] Shan, G.G., Vexler, A., Wilding, G. and Hutson, A. (2011) Simple and Exact Empirical Likelihood Ratio Tests for Normality Based on Moment Relations. *Communications in Statistics*: *Simulation and Computation*, **40**, 129-146. http://dx.doi.org/10.1080/03610918.2010.532896

[3] Wilcoxon, F. (1945) Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, **1**, 80-83. http://dx.doi.org/10.2307/3001968

[4] Mann, H.B. and Whitney, D.R. (1947) On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, **18**, 50-60. http://dx.doi.org/10.1214/aoms/1177730491

[5] Lam, F.C. and Longnecker, M.T. (1983) A Modified Wilcoxon Rank Sum Test for Paired Data. *Biometrika*, **70**, 510-513. http://dx.doi.org/10.1093/biomet/70.2.510

[6] Shan, G.G., Ma, C.X., Hutson, A.D. and Wilding, G.E. (2013) Some Tests for Detecting Trends Based on the Modified Baumgartner Weiß Schindler Statistics. *Computational Statistics & Data Analysis*, **57**, 246-261. http://dx.doi.org/10.1016/j.csda.2012.04.021

[7] Fay, M.P. and Proschan, M.A. (2010) Wilcoxon-Mann-Whitney or t-Test? On Assumptions for Hypothesis Tests and Multiple Interpretations of Decision Rules. *Statistics Surveys*, **4**, 1-39.

[8] Baumgartner, W., Weiß, P. and Schindler, H. (1998) A Nonparametric Test for the General Two-Sample Problem. *Biometrics*, **54**, 1129-1135. http://dx.doi.org/10.2307/2533862

[9] Neuhäuser, M. (2001) One-Sided Two-Sample and Trend Tests Based on a Modified Baumgartner-Weiß Schindler

Statistic. *Journal of Nonparametric Statistics*, **13**, 729-739. http://dx.doi.org/10.1080/10485250108832874

[10] Murakami, H. (2006) A *k*-Sample Rank Test Based on Modified Baumgartner Statistic and Its Power Comparison. *Journal of the Japanese Society of Computational Statistics*, **19**, 1-13. http://dx.doi.org/10.5183/jjscs1988.19.1

[11] Kundu, D. and Gupta, R.D. (2009) Bivariate Generalized Exponential Distribution. *Journal of Multivariate Analysis*, **100**, 581-593. http://dx.doi.org/10.1016/j.jmva.2008.06.012

[12] Antonisamy, B., Christopher, S. and Samuelson, P. (2010) Biostatistics: Principles and Practice. McGraw-Hill Education, New York.

[13] Wilding, G.E., Shan, G. and Hutson, A.D. (2012) Exact Two-Stage Designs for Phase II Activity Trials with Rank-Based Endpoints. *Contemporary Clinical Trials*, **33**, 332-341. http://dx.doi.org/10.1016/j.cct.2011.10.008

[14] Shan, G., Ma, C., Hutson, A.D. and Wilding, G.E. (2012) An Efficient and Exact Approach for Detecting Trends with Binary Endpoints. *Statistics in Medicine*, **31**, 155-164. http://dx.doi.org/10.1002/sim.4411

[15] Shan, G. and Ma, C. (2012) Unconditional Tests for Comparing Two Ordered Multinomials. *Statistical Methods in Medical Research*, Published Online. http://dx.doi.org/10.1177/0962280212450957

[16] Shan, G. (2013) More Efficient Unconditional Tests for Exchangeable Binary Data with Equal Cluster Sizes. *Statistics & Probability Letters*, **83**, 644-649. http://dx.doi.org/10.1016/j.spl.2012.11.014

[17] Shan, G. (2013) A Note on Exact Conditional and Unconditional Tests for Hardy-Weinberg Equilibrium. *Human Heredity*, **76**, 10-17. http://dx.doi.org/10.1159/000353205

[18] Shan, G. and Ma, C. (2014) Exact Methods for Testing the Equality of Proportions for Binary Clustered Data from Otolaryngologic Studies. *Statistics in Biopharmaceutical Research*, **6**, 115-122.

[19] Shan, G., Ma, C., Hutson, A.D. and Wilding, G.E. (2013) Randomized Two-Stage Phase II Clinical Trial Designs Based on Barnard's Exact Test. *Journal of Biopharmaceutical Statistics*, **23**, 1081-1090. http://dx.doi.org/10.1080/10543406.2013.813525

[20] Shan, G. (2014) Exact Approaches for Testing Non-Inferiority or Superiority of Two Incidence Rates. *Statistics & Probability Letters*, **85**, 129-134. http://dx.doi.org/10.1016/j.spl.2013.11.010

[21] Jonckheere, A.R. (1954) A Distribution-Free *k*-Sample Test against Ordered Alternatives. *Biometrika*, **41**, 133-145. http://dx.doi.org/10.2307/2333011

[22] Terpstra, T.J. (1952) The Asymptotic Normality and Consistency of Kendall's Test against Trend, When Ties Are Present in One Ranking. *Indigationes Mathematicae*, **14**, 327-333.

[23] Shan, G., Hutson, A.D. and Wilding, G.E. (2012) Two-Stage *k*-Sample Designs for the Ordered Alternative Problem. *Pharmaceutical Statistics* **11**, 287-294. Http://Dx.Doi.Org/10.1002/Pst.1499

[24] Page, E.B. (1963) Ordered Hypotheses for Multiple Treatments: A Significance Test for Linear Ranks. *Journal of the American Statistical Association*, **58**, 216-230. http://dx.doi.org/10.1080/01621459.1963.10500843

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.