A test for variability in the two-sample case

Huang, Tzee-Ming
National Chengchi University, Department of Statistics
NO.64, Sec.2, Zhinan Rd., Wenshan District
Taipei City 11605, Taiwan (R.O.C)
E-mail: tmhuang@nccu.edu.tw

1 Introduction

Suppose that $(X_{1,1}, \ldots, X_{1,n_1})$ and $(X_{2,1}, \ldots, X_{2,n_2})$ are two independent random samples. The problem of interest is to test whether the distribution of $X_{2,1}$ has a larger variability than that of $X_{1,1}$. In addition to the well-known F test for equal variance that assumes normality, other variability tests have been proposed and found robust even when the data are not normal. Overviews of variability tests can be found in Conover et al [3] and Ramsey and Ramsey [7].

A nonparametric variability test, mentioned in Moses [6], applies the Wilcoxon-Mann-Whitney rank-order test to intra absolute differences. That is, the test applies the Wilcoxon-Mann-Whitney rank-order test to the two samples of differences, where the first sample of differences is

$$(|X_{1,1} - X_{1,2}|, |X_{1,3} - X_{1,4}|, \ldots)$$

and the second sample of differences is

$$(|X_{2,1} - X_{2,2}|, |X_{2,3} - X_{2,4}|, \ldots).$$

This test is denoted by WM hereafter. It is also possible to use randomly selected pairs for the above test, as in Blair and Thompson [1]. However, this approach is criticized in [7] for the random selection and for using only half of the data.

In this paper, the proposed test is based on the Wilcoxon-Mann-Whitney rank-order test statistic of two samples of differences of all pairs, where the i-th sample of differences is consisted of

$$\{|X_{i,j} - X_{i,k}| : 1 \le j < k \le n_i\} \tag{1}$$

for $i = 1, 2$. By considering all pairs, the problem of pair selection and the problem of wasting data are taken care of. The rest of the paper is organized as follows. The details of the testing procedure are in Section 2. Simulation results are in Section 3, followed by the conclusion.

2 Model assumption and testing procedure

It is assumed that the distributions for the two independent samples are in the same location-scale family. That is, there exist constants μ and $\sigma > 0$ such that $X_{1,1}$ and $\mu + \sigma X_{2,1}$ have the same distribution. Thus the testing problem of interest becomes

$$H_0 : \sigma = 1 \text{ versus } H_1 : \sigma > 1.$$

As mentioned in Section 1, the test is based on the ranks of the two samples of differences given in (1). Specifically, the rank sum for the first sample of difference is computed as the test statistic, so the test rejects H_0 when the test statistic is small.

The test statistic is not an ancillary under H_0. Below are the steps to obtain the rejection region for the test at level α for the data $(X_{1,1}, \ldots, X_{1,n_1})$ and $(X_{2,1}, \ldots, X_{2,n_2})$.

Step 1. Specify $\mathcal{F}$: a collection of distributions supported on $[0,1]$. The distributions in $\mathcal{F}$ serve as candidate distributions for $X_{1,1}/C$, where C is a constant so that the scaled variable $X_{1,1}/C$ is in $[0,1]$ (or in $[0,1]$ with large probability).

Step 2. For each distribution F in $\mathcal{F}$, compute the critical value assuming $X_{1,1}$ is from F. The critical values are obtained via simulation. 1000 sets of two-sample data are generated from F and 1000 test statistics are computed. A $(1-\alpha)$ quantile of the 1000 test statistics is computed as the critical value, denoted by $c(F)$.

Step 3. Compute the combined sample that is consisted of

$$\frac{X_{i,j} - \min_j X_{i,j}}{\max_j X_{i,j} - \min_j X_{i,j}} : 1 \le j \le n_i, i = 1, 2.$$

Step 4. For each F in $\mathcal{F}$, apply the Kolmogrov one-sample test at 0.05 level to the combined sample to check whether the sample is from F. Let H_0^* be the hypothesis that the combined sample is from F and

$$\mathcal{F}_1 = \{F \in \mathcal{F} : \text{ the Kolmogrov test does not reject } H_0^* .\}$$

The test rejects H_0 if and only if the test statistic is less than

$$\min\{c(F) : F \in \mathcal{F}_1\}.$$

Note that the test statistic remains the same under location and scale transforms. Therefore, it is fine to consider candidate distributions for the scaled variable $X_{1,1}/C$, which is supported on $[0,1]$. For unbounded distributions, it is impossible to rescale $X_{1,1}$ so that the rescaled variable is supported on $[0,1]$. However, it is possible to find C such that $X_{1,1}/C$ is in $[0,1]$ with large probability.

The $\mathcal{F}$ in Step 1 is taken to be the collections of distributions with cumulative distribution functions of the form

$$F_{c_1,\ldots,c_{10}}(x) = \sum_{i=1}^{10} \left(\left(\sum_{1 \le j < i} \frac{c_j}{10} \right) + c_i \left(x - \frac{i-1}{10} \right) \right) I_{\left(\frac{i-1}{10}, \frac{i}{10}\right]}(x),$$

where $c_1, \ldots, c_{10}$ are non-negative integers such that $\sum_{i=1}^{10} c_i = 10$. This choice of $\mathcal{F}$ contains 92378 distributions. Due to the limitation of computing power, no other $\mathcal{F}$'s are considered in this study.

Another approach to find the rejection region for the proposed test statistic is to use

$$\min\{c(F) : F \in \mathcal{F}\}$$

from Step 2 as the critical value, so Steps 3 and 4 can be eliminated. This approach is less computational extensive but more conservative than the proposed approach. The test based on the conservative critical value is denoted by WM.AC hereafter. The proposed test, which uses the rejection region obtained from Steps 1–4, is denoted by WM.A hereafter.

3 Simulation study

In the simulation study, the Type I error control and power performance for the proposed test (WM.A) are examined and compared with those of several other tests under various data distributions. The tests to be compared are WM, WM.AC and the test BFO in [7]. Probability calculation for the test statistic of WM can be found in [5]. The data distributions (for $X_{1,1}$) considered are $N(0,1)$, $t(1)$, $t(4)$, $t(10)$ (the t distributions with degrees of freedom 1, 4 and 10 respectively) and $S_B(0, 1.1)$ and

$S_B(1.2, 5)$. Here $S_B(\beta_1, \beta_2)$ denotes Johnson's S_B distribution ([4]) described in Tadikamalla [8] with skewness $\sqrt{\beta_1}$ and kurtosis β_2 when the location and scale parameters are set to 0 and 1 respectively. For a distribution, the skewness is $\mu_3/\mu_2^{1.5}$ and the kurtosis is μ_4/μ_2^2, where μ_j denotes the j-th centered moment of the distribution. For power examination, $X_{2,1}$ is generated so that it is distributed as $\sigma X_{1,1}$ with $\sigma = \sqrt{3}$. The two sample sizes n_1 and n_2 are in $\{5, 7, 10\}$. The number of replications is 10000 for WM and BFO. For WM.AC, the number of replications is 6800 for the $n_1 = 10 = n_2$ case and 2000 for all other cases. For WM.A, the sub-collection $\mathcal{F}_1$ in Step 4 can be empty, and the replication is disregarded in such case. Therefore, the effective number of replications varies. The test level is taken to be 0.05.

The estimated Type I error probabilities are given in Table 1. All tests show good Type I error control except that the estimated Type I error probabilities for BFO are significantly larger than the nominal level 0.05 when $(n_1, n_2) = (5, 10)$ and the data distribution is $S_B(0, 1.1)$ or $S_B(1.2, 5)$.

n_1	n_2	Distribution	WM	BFO	WM.A	WM.AC	WM.A effective replications
5	5	$N(0,1)$	0	0.0131	0.01	0.0075	2000
5	5	$t(1)$	0	0.0178	0.0411	0.0316	1994
5	5	$t(4)$	0	0.019	0.022	0.016	2000
5	5	$t(10)$	0	0.0137	0.0135	0.0095	2000
5	5	$S_B(0,1.1)$	0	0.0213	0.0147	0.0116	1638
5	5	$S_B(1.2,5)$	0	0.0202	0.0146	0.0109	1644
5	7	$N(0,1)$	0	0.027	0.009	0.0075	2000
5	7	$t(1)$	0	0.014	0.0371	0.0351	1967
5	7	$t(4)$	0	0.0222	0.0145	0.014	2000
5	7	$t(10)$	0	0.0251	0.0145	0.0115	2000
5	7	$S_B(0,1.1)$	0	0.0272	0.0149	0.0134	1341
5	7	$S_B(1.2,5)$	0	0.0302	0.0171	0.0164	1402
5	10	$N(0,1)$	0.0456	0.0464	0.009	0.0085	2000
5	10	$t(1)$	0.0465	0.0147	0.0438	0.0376	1942
5	10	$t(4)$	0.0488	0.0397	0.023	0.0205	2000
5	10	$t(10)$	0.0442	0.0465	0.0135	0.0125	2000
5	10	$S_B(0,1.1)$	0.0463	0.0821	0.0164	0.0146	1162
5	10	$S_B(1.2,5)$	0.0523	0.0788	0.018	0.0172	1165
7	5	$N(0,1)$	0	0.016	0.0095	0.0095	2000
7	5	$t(1)$	0	0.0289	0.052	0.0515	1960
7	5	$t(4)$	0	0.0186	0.018	0.0175	2000
7	5	$t(10)$	0	0.016	0.012	0.012	2000
7	5	$S_B(0,1.1)$	0	0.0084	0.0045	0.0037	1341
7	5	$S_B(1.2,5)$	0	0.0107	0.0047	0.0039	1280
7	7	$N(0,1)$	0.0504	0.0243	0.0065	0.0055	2000
7	7	$t(1)$	0.0488	0.0236	0.0433	0.0398	1962
7	7	$t(4)$	0.051	0.025	0.0135	0.0105	2000
7	7	$t(10)$	0.0504	0.027	0.0135	0.0125	2000
7	7	$S_B(0,1.1)$	0.049	0.0186	0.006	0.0053	1327
7	7	$S_B(1.2,5)$	0.0527	0.0147	0.0061	0.0053	1320
7	10	$N(0,1)$	0.0337	0.0456	0.0125	0.0115	2000

Table 1: Estimated Type I error probabilities (continued)

n_1	n_2	Distribution	WM	BFO	WM.A	WM.AC	WM.A effective replications
7	10	$t(1)$	0.0356	0.0178	0.0438	0.0392	1941
7	10	$t(4)$	0.0383	0.0408	0.017	0.0165	2000
7	10	$t(10)$	0.0358	0.0401	0.012	0.0105	2000
7	10	$S_B(0, 1.1)$	0.0366	0.0488	0.0084	0.0074	952
7	10	$S_B(1.2, 5)$	0.0314	0.0455	0.0056	0.0045	892
10	5	$N(0, 1)$	0.0467	0.0147	0.006	0.0055	2000
10	5	$t(1)$	0.0483	0.044	0.052	0.0484	1923
10	5	$t(4)$	0.0446	0.0252	0.0295	0.0285	2000
10	5	$t(10)$	0.0486	0.0205	0.0195	0.0195	2000
10	5	$S_B(0, 1.1)$	0.0478	0.0045	0.0026	0.0017	1154
10	5	$S_B(1.2, 5)$	0.047	0.0041	0	0	1160
10	7	$N(0, 1)$	0.0344	0.024	0.013	0.01	2000
10	7	$t(1)$	0.0325	0.0355	0.0541	0.0424	1959
10	7	$t(4)$	0.0359	0.0313	0.027	0.0205	2000
10	7	$t(10)$	0.0347	0.0253	0.014	0.0105	2000
10	7	$S_B(0, 1.1)$	0.0381	0.0086	0.0021	0.001	953
10	7	$S_B(1.2, 5)$	0.0347	0.0086	0.0011	0.0011	881
10	10	$N(0, 1)$	0.0455	0.0399	0.0124	0.011	6800
10	10	$t(1)$	0.0467	0.026	0.0524	0.0472	6541
10	10	$t(4)$	0.0464	0.04	0.0226	0.0201	6800
10	10	$t(10)$	0.0464	0.0443	0.0166	0.0141	6800
10	10	$S_B(0, 1.1)$	0.0444	0.0288	0	0	2380
10	10	$S_B(1.2, 5)$	0.0471	0.0294	0.0008	0.0008	2390

Table 1: Estimated Type I error probabilities

The power estimates are given in Table 2. For small n_1 and n_2, WM does not reject H_0 at all and has no power. WM.AC is less powerful than WM.A for all cases. BFO appears to be the most powerful except for the $t(1)$ case, in which WM.A is the most powerful. When WM can reject H_0 with positive probability, its power is comparable with that of WM.A except for the $t(1)$ case.

n_1	n_2	Distribution	WM	BFO	WM.A	WM.AC	WM.A effective replications
5	5	$N(0, 1)$	0	0.076	0.06	0.047	2000
5	5	$t(1)$	0	0.0458	0.0958	0.0787	1994
5	5	$t(4)$	0	0.0704	0.078	0.0655	2000
5	5	$t(10)$	0	0.0726	0.0705	0.0525	2000
5	5	$S_B(0, 1.1)$	0	0.2758	0.0201	0.0153	1638
5	5	$S_B(1.2, 5)$	0	0.2712	0.0182	0.0128	1644
5	7	$N(0, 1)$	0	0.123	0.0650	0.0610	2000
5	7	$t(1)$	0	0.0416	0.0961	0.0946	1967
5	7	$t(4)$	0	0.1004	0.0865	0.0825	2000
5	7	$t(10)$	0	0.1119	0.0740	0.0705	2000
5	7	$S_B(0, 1.1)$	0	0.3167	0.0194	0.0164	1341

Table 2: Power estimates (continued)

n_1	n_2	Distribution	WM	BFO	WM.A	WM.AC	WM.A effective replications
5	7	$S_B(1.2, 5)$	0	0.3246	0.0243	0.0221	1402
5	10	$N(0, 1)$	0.1172	0.2081	0.0735	0.0620	2000
5	10	$t(1)$	0.1037	0.0429	0.1344	0.1189	1942
5	10	$t(4)$	0.1192	0.1521	0.0990	0.0895	2000
5	10	$t(10)$	0.1161	0.1887	0.0825	0.0755	2000
5	10	$S_B(0, 1.1)$	0.0691	0.4945	0.0215	0.0164	1162
5	10	$S_B(1.2, 5)$	0.0774	0.5041	0.0275	0.0249	1165
7	5	$N(0, 1)$	0	0.1217	0.0955	0.0955	2000
7	5	$t(1)$	0	0.0752	0.1337	0.1332	1960
7	5	$t(4)$	0	0.112	0.1175	0.1165	2000
7	5	$t(10)$	0	0.1132	0.093	0.093	2000
7	5	$S_B(0, 1.1)$	0	0.3719	0.0075	0.0045	1341
7	5	$S_B(1.2, 5)$	0	0.3688	0.0156	0.0063	1280
7	7	$N(0, 1)$	0.1544	0.1821	0.1140	0.1025	2000
7	7	$t(1)$	0.1095	0.0693	0.1320	0.1208	1962
7	7	$t(4)$	0.1420	0.1487	0.1090	0.0990	2000
7	7	$t(10)$	0.1486	0.1639	0.1035	0.0890	2000
7	7	$S_B(0, 1.1)$	0.1308	0.4216	0.0128	0.0083	1327
7	7	$S_B(1.2, 5)$	0.1451	0.4267	0.0121	0.0076	1320
7	10	$N(0, 1)$	0.1201	0.2694	0.1260	0.1125	2000
7	10	$t(1)$	0.0919	0.063	0.1535	0.1417	1941
7	10	$t(4)$	0.1153	0.2046	0.1435	0.1330	2000
7	10	$t(10)$	0.1216	0.2409	0.1250	0.1160	2000
7	10	$S_B(0, 1.1)$	0.0827	0.5875	0.0231	0.0084	952
7	10	$S_B(1.2, 5)$	0.0763	0.5858	0.0067	0.0045	892
10	5	$N(0, 1)$	0.1750	0.1633	0.1390	0.1325	2000
10	5	$t(1)$	0.0997	0.1135	0.1550	0.1472	1923
10	5	$t(4)$	0.1528	0.1602	0.1600	0.1555	2000
10	5	$t(10)$	0.1703	0.1645	0.1515	0.1480	2000
10	5	$S_B(0, 1.1)$	0.2302	0.4467	0.0113	0.0043	1154
10	5	$S_B(1.2, 5)$	0.2287	0.4457	0.0078	0.0009	1160
10	7	$N(0, 1)$	0.1563	0.2272	0.1575	0.1320	2000
10	7	$t(1)$	0.0929	0.1086	0.1644	0.1440	1959
10	7	$t(4)$	0.1413	0.2036	0.1695	0.1440	2000
10	7	$t(10)$	0.1479	0.2208	0.1670	0.1435	2000
10	7	$S_B(0, 1.1)$	0.1390	0.5715	0.0168	0.0021	953
10	7	$S_B(1.2, 5)$	0.1361	0.5772	0.0216	0.0023	881
10	10	$N(0, 1)$	0.2148	0.3317	0.1966	0.1793	6800
10	10	$t(1)$	0.1325	0.095	0.1801	0.1666	6541
10	10	$t(4)$	0.1943	0.2725	0.2097	0.1966	6800
10	10	$t(10)$	0.1972	0.3143	0.2029	0.1894	6800
10	10	$S_B(0, 1.1)$	0.1701	0.7095	0.0210	0.0038	2380
10	10	$S_B(1.2, 5)$	0.1622	0.7018	0.0268	0.0025	2390

Table 2: Power estimates

4 Conclusion

It is recommended to use the proposed test WM.A when the WM test has no power at all. When WM can reject H_0, its power is comparable with (and sometimes better than) that of WM.A for most cases. Since it is computationally expansive to apply WM.A, WM is more recommended. WM.AC is found to be inferior to WM.A and not recommended. BFO is the most powerful for most cases, but has low power for the $t(1)$ case, and the Type I error can be out of control for the S_B distributions. If one looks for a test that has stable and acceptable performance for all cases, WM is recommended.

Acknowledgement

Cheng [2] also considers a variability test that is similar to WM.A but applies the Wilcoxon-Mann-Whitney rank-order test to the following two samples of differences:

$$(|X_{1,1} - X_{1,2}|, |X_{1,2} - X_{1,3}|, \ldots, |X_{1,n_1-1} - X_{1,n_1}|)$$

and

$$(|X_{2,1} - X_{2,2}|, |X_{2,2} - X_{2,3}|, \ldots, |X_{2,n_2-1} - X_{2,n_2}|).$$

The author would like to thank Cheng for helpful discussion.

REFERENCES (RÉFERENCES)

[1] R. C. Blair and G. L. Thompson, A distribution-free rank-like test for scale with unequal population locations, Communications in Statistics: Simulation and Computation, 21 (1992), pp. 353–371.

[2] Y. Cheng, A nonparametric test for detecting increasing variability, master's thesis, National Chengchi University, Taiwan, R.O.C., 2011.

[3] W. J. Conover, M. E. Johnson, and M. M. Johnson, A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data (Corr: V26 p302), Technometrics, 23 (1981), pp. 351–361.

[4] N. L. Johnson, Systems of frequency curves generated by methods of translation, Biometrika, 36 (1949), pp. 149–176.

[5] H. B. Mann and D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Annals of Mathematical Statistics, 18 (1947), pp. 50–60.

[6] L. E. Moses, Rank tests of dispersion, The Annals of Mathematical Statistics, 34 (1963), pp. 973–983.

[7] P. H. Ramsey and P. P. Ramsey, Testing variability in the two-sample case, Communications in Statistics–Simulation and Computation, 36 (2007), pp. 233–248.

[8] P. R. Tadikamalla, On simulating non-normal distributions, Psychometrika, 45 (1980), pp. 273–279.