

Test procedures for the mean and variance simultaneously under normality

Hyo-II Park^{1,a}

^aDepartment of Statistics, Cheongju University, Korea

Abstract

In this study, we propose several simultaneous tests to detect the difference between means and variances for the two-sample problem when the underlying distribution is normal. For this, we apply the likelihood ratio principle and propose a likelihood ratio test. We then consider a union-intersection test after identifying the likelihood statistic, a product of two individual likelihood statistics, to test the individual sub-null hypotheses. By noting that the union-intersection test can be considered a simultaneous test with combination function, also we propose simultaneous tests with combination functions to combine individual tests for each sub-null hypothesis. We apply the permutation principle to obtain the null distributions. We then provide an example to illustrate our proposed procedure and compare the efficiency among the proposed tests through a simulation study. We discuss some interesting features related to the simultaneous test as concluding remarks. Finally we show the expression of the likelihood ratio statistic with a product of two individual likelihood ratio statistics.

Keywords: combination function, likelihood ratio principle, Monte-Carlo method, permutation principle, union-intersection test

1. Introduction

Comparing if distributions of two populations coincide or not between two populations has been an interesting and important research topic for statistical inference. One common procedure is the two-sample t -test by assuming normality and equal variance for underlying distributions. The Wilcoxon rank-sum test is the most popular among nonparametric tests and is widely used under the location translation model. In addition, one may apply the Ansari-Bradley test when one considers a scale model. However, one may sometimes be uncertain if the location or scale model should be chosen. In that case, one may consider accommodating location and scale models simultaneously. In this approach, Lepage (1971) proposed a nonparametric test procedure by considering applying and combining the Wilcoxon and Ansari-Bradley tests for the location and scale parameters simultaneously. From then on, many nonparametric tests have been proposed with various tests for each parameter and various combining functions (Park, 2015). Under the normality assumption for the one sample problem setting, several simultaneous test procedures have been considered with the likelihood ratio (LR) principle (Choudhari *et al.*, 2001; Park and Han, 2013). For the multi-sample case with the multivariate data, Mardia *et al.* (1979) derived the LR statistic using conditional arguments and considered obtaining the null distribution asymptotically. Flury *et al.* (1995), under the common principal component model, also considered the simultaneous inferences for multivariate data by applying the LR principle for the test.

¹ Department of Statistics, Cheongju University, 298 Daeseong-ro, Cheongwon-gu, Cheongju 28503, Korea.
E-mail: hipark@cju.ac.kr

We first introduce the null hypothesis and its alternative in order to discuss simultaneous tests that more precisely compare two independent populations with respective normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$:

$$H_0 : \{\mu_1 = \mu_2\} \cap \{\sigma_1^2 = \sigma_2^2\} \quad (1.1)$$

and

$$H_0 : \{\mu_1 \neq \mu_2\} \cup \{\sigma_1^2 \neq \sigma_2^2\}.$$

Up to now, several tests have been proposed by combining two partial tests to test $H_{01} : \mu_1 = \mu_2$ and $H_{02} : \sigma_1^2 = \sigma_2^2$ with choices of suitable tests under the linear model containing covariates. In this approach, we note that in order to combine two partial tests, one should use the p -values rather than test statistics since two forms of statistics and the testing rules might be different to test $H_{01} : \mu_1 = \mu_2$ and $H_{02} : \sigma_1^2 = \sigma_2^2$. Therefore, one may overcome this difficulty using two respective p -values with combining functions. The multiple use of tests requires the use of combining functions to combine the individual tests, $H_{01} : \mu_1 = \mu_2$ and $H_{02} : \sigma_1^2 = \sigma_2^2$. There have been several widely used combining functions such as Fisher, Liptak and Tippett for nonparametric simultaneous tests. The quadratic form has also been used in the various cases especially for multivariate data in parametric and nonparametric tests. Pesarin and Salmaso (2010) have reviewed and discussed these combining functions extensively for their applications and properties.

To obtain the null distribution of test statistics or combining functions rely on two types of statistics for the location and scale parameters that are uncorrelated since the two statistics are odd and even translation invariant, respectively (Randles and Wolfe, 1979). This then implies that the two statistics are independent in the limiting sense when we consider the asymptotic normality for each statistic. Therefore, it would be easy to derive the limiting null distributions for any given combining functions. It would be difficult to obtain null distributions even when the components of combining functions are not independent, even in the limiting sense. However this difficulty may be overcome by applying the permutation principle, which is a re-sampling method initiated by Fisher (1932) that has been recently popular to perform a test since it yields the exact test even though conditional. However, the real application to the test should be based on the Monte-Carlo approach since all the configurations of permutations for the given data require an enormous burden for the calculations.

In this paper, we propose simultaneous tests to detect differences between means and variances for the two sample problem under the normality assumption. For this purpose, the rest of the paper will be organized as follow. We first begin to derive the LR statistic in the next section in order to obtain an expression with a product of two functions, each of which consists of a random quantity. We then propose a union-intersection test with other tests using various combining functions that include the LR test and applying the permutation principle to obtain the null distributions. Then we illustrate our procedure with an example and compare the efficiency among the proposed tests by obtaining empirical powers through a simulation study in the Section 3. In Section 4, we discuss some interesting features related to the simultaneous tests as concluding remarks. Finally, we show the expression of likelihood ratio statistic with a product of two individual likelihood ratio statistics in a descriptive manner.

2. Derivation of likelihood ratio statistic for the simultaneous test

Suppose that we have two independent random samples, X_1, \dots, X_m and Y_1, \dots, Y_n from populations with distribution functions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. First, based on these samples, we

consider to apply the LR principle for testing H_0 against H_1 . For this, we set up some notation:

$$\begin{aligned} \bar{X} &= \frac{1}{m} \sum_{i=1}^m X_i, & \bar{Y} &= \frac{1}{n} \sum_{j=1}^n Y_j, & S_1^2 &= \frac{1}{m} \sum_{i=1}^m (X_i - \bar{X})^2, & S_2^2 &= \frac{1}{n} \sum_{j=1}^n (Y_j - \bar{Y})^2, \\ \hat{\mu} &= \frac{m\bar{X} + n\bar{Y}}{m+n}, & \text{and } S_{12}^2 &= \frac{1}{m+n} \left\{ \sum_{i=1}^m (X_i - \hat{\mu})^2 + \sum_{j=1}^n (Y_j - \hat{\mu})^2 \right\}. \end{aligned}$$

We note that the first four terms are the LR statistics under $H_0 \cup H_1$ and the last two terms are the LR ones under H_0 . Then we can derive the LR statistic, $\text{LR}(X, Y; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, to test (1.1) in the following way.

$$\begin{aligned} \text{LR}(X, Y; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= \frac{\sup \{f(X, Y; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : H_0 \cup H_1\}}{\sup \{f(X, Y; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) : H_0\}} \\ &= \frac{(2\pi S_1^2)^{-\frac{m}{2}} (2\pi S_2^2)^{-\frac{n}{2}} \exp\left[-\frac{m+n}{2}\right]}{(2\pi S_{12}^2)^{-\frac{m+n}{2}} \exp\left[-\frac{m+n}{2}\right]} \\ &= \frac{(S_{12}^2)^{\frac{m+n}{2}}}{(S_1^2)^{\frac{m}{2}} (S_2^2)^{\frac{n}{2}}}. \end{aligned} \tag{2.1}$$

Thus one may reject H_0 in favor of H_1 for some large values of LR, (2.1) in view of the LR principle. Then in order to complete the LR test, we have to derive the null distribution of (2.1). However it would be difficult to derive the null distribution of (2.1) since the components of (2.1) are not independent. We note that Mardia *et al.* (1979) also derived the statistic (2.1) using conditional arguments and concluded that the null distribution of

$$-2 \log(\text{LR}) = m \log\left(\frac{S_1^2}{S_{12}^2}\right) + n \log\left(\frac{S_2^2}{S_{12}^2}\right)$$

has a chi-square with 2 degrees of freedom (d.f.) approximately, where means the natural logarithm. However, the limiting distribution would be too crude to apply to the real problem as identified in the simulation section. For this reason, we will consider to apply the permutation principle to obtain the null distribution of (2.1) and compare the results with those of the asymptotic distributions.

In order to construct a union-intersection procedure for the null hypothesis (1.1), we investigate the LR statistic (2.1) by decomposing its components with the notation introduced above, we have that

$$\text{LR}(X, Y; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \left(1 + \frac{nS_2^2}{mS_1^2}\right)^{\frac{m}{2}} \left(1 + \frac{mS_1^2}{nS_2^2}\right)^{\frac{n}{2}} \left\{1 + \frac{mn}{m+n} \frac{(\bar{X} + \bar{Y})^2}{mS_1^2 + nS_2^2}\right\}^{\frac{m+n}{2}}. \tag{2.2}$$

The derivation of the expression (2.2) will appear in Appendix briefly. Now we note that the two random quantities in (2.2), S_2^2/S_1^2 and S_1^2/S_2^2 are reciprocals for each other. Then by denoting

$$F = \frac{S_2^2}{S_1^2} \quad \text{and} \quad T = \frac{(\bar{X} + \bar{Y})^2}{mS_1^2 + nS_2^2},$$

we see that (2.2) can be re-expressed as since $F > 0$,

$$\frac{\left[1 + \left(\frac{n}{m}\right)F\right]^{\frac{m+n}{2}}}{\left(\frac{n}{m}\right)F^{\frac{n}{2}}} \left(1 + \frac{mn}{m+n}T\right)^{\frac{m+n}{2}} = g(F)h(T), \quad \text{say.}$$

Therefore, (2.2) consists of the product of two functions, g and h , each of which is a function of a random quantity. The maximization of LR or equivalently, (2.2) can then be achieved by maximizing g and h independently. This point leads us to consider the union-intersection (UI) test (Roy, 1957) based on g and h or equivalently, F and T . We note that F and T can be considered as joint LR statistics that are not independent. We first need the following simple result for the construction of the UI test.

Lemma 1. *The function g on $(0, \infty)$ defined by*

$$g(f) = \frac{\left[1 + \left(\frac{n}{m}\right)f\right]^{\frac{m+n}{2}}}{f^{\frac{n}{2}}}$$

achieves its minimum at $f = 1$.

Proof: Using the differentiation of g with respect to f , we see the result. \square

From Lemma 1, we note that g can be maximized when the value of F becomes closer to 0 or larger than 1. Since F should be used for testing the sub-null hypothesis $H_{02} : \sigma_1^2 = \sigma_2^2$ against $H_{12} : \sigma_1^2 \neq \sigma_2^2$, the LR testing rule would be to reject $H_{02} : \sigma_1^2 = \sigma_2^2$ for the value a or b of F such that $0 < a < 1 < b$ from Lemma 1. The choices of a and b will depend on the given significance level and be discussed in detail later. We should also use T for the LR test when testing the sub-null hypothesis $H_{01} : \mu_1 = \mu_2$ against $H_{11} : \mu_1 \neq \mu_2$. The LR testing rule would be to reject $H_{01} : \mu_1 = \mu_2$ for some large values of T since the function h is strictly increasing in T . To combine the two statistics F and T for the UI test, it is difficult to use statistics themselves since they have different types of test criteria even though their null distributions are related to the F -distribution. It is then customary to use the individual p -values for the construction of a test statistic. However, we first need their null distributions for both as summarized in the following lemma.

Lemma 2. *Under H_0 , the null distributions of F and T can be summarized as*

- (I) $((n-1)/(m-1))F$ is distributed as the F -distribution with $n-1$ and $m-1$ d.f.
- (II) $(m+n-2)T$ is distributed as the F -distribution with 1 and $m+n-2$ d.f.

Let Λ_T and Λ_F be the corresponding p -values to test the sub-null hypotheses, $H_{01} : \mu_1 = \mu_2$ and $H_{02} : \sigma_1^2 = \sigma_2^2$ with T and F , respectively. Then the UI statistic, UI to test H_0 would be

$$\text{UI} = \max \{1 - \Lambda_F, 1 - \Lambda_T\}. \quad (2.3)$$

Then the testing rule of the UI test would be to reject H_0 for some large values of UI. Then we need the null distribution to complete the UI test. However, it would be difficult to obtain the null distribution of UI since Λ_F and Λ_T are not independent. This difficulty can also be resolved by applying the permutation principle.

Table 1: Kerosene heater data

Brand	Time in second														
A	69.3	56.0	22.1	47.6	53.2	48.1	23.2	13.8	52.6	34.4	60.2	43.8			
B	28.6	25.1	26.4	34.9	29.8	28.4	38.5	30.2	30.6	31.8	41.6	21.1	36.0	37.9	13.9

In view of the nonparametric simultaneous test approach, we note that one may consider the UI statistic in (2.3) as the Tippett combining function (Tippett, 1931). In this regard, one can also consider the following two more combining functions to combine the two p -values, Λ_F and Λ_T in this study. One is the Fisher combining function FI (Fisher, 1932), the other is the Liptak combining function LI (Liptak, 1958), defined by

$$FI = -2 \{ \log(1 - \Lambda_F), \log(1 - \Lambda_T) \}$$

and

$$LI = \Psi^{-1}(1 - \Lambda_F) + \Psi^{-1}(1 - \Lambda_T),$$

where \log means the natural logarithm and Ψ , a distribution function and Ψ^{-1} , its inverse. Then we note that it is customary to choose the standard normal distribution, Φ for Ψ for the distributional convenience for the real applications. The testing rule for each combining function would be to reject some large values of each combining function. We then have to derive the null distributions for FI and LI in order to obtain the overall p -value for each test. However, the null distributions of FI and LI would be difficult to derive with the same reason for UI. Therefore, we will apply the permutation principle to complete these tests.

One then may perform a simultaneous test for the mean and variance by choosing a suitable test as proposed. A comparison study for the performance among the proposed tests would be necessary to obtain some idea to choose a suitable test. We compare their efficiency in the next section by obtaining empirical powers through a simulation study by choosing Φ for Ψ for the LI test with other famous tests. We then begin the next section with an example in order to illustrate our procedure.

3. An example and simulation study

We first illustrate our procedure with a numerical example using the data in Milton and Arnold (2003) contained in Table 1 which tabulated the test results of two brands of kerosene heaters.

The observations are the times in seconds required to raise the room temperature 10°F. Moser and Stevens (1992) raised the question on if only the two-sample t -tests would be appropriate for this data set by obtaining 0.0449 as p -value to test of the equality of two variances. With this result, they concluded that one should consider using the Satterthwaite test which is a testing procedure for the mean difference when the two variances seem to be unequal rather than applying the usual two-sample t -test for this data set in Table 1. From Figure 1, also we note that the observations of Brand A are apt to be dispersed relatively more widely than those of B. From those arguments, it would be reasonable to apply a simultaneous test rather than the two-sample t -test to the data in Table 1. We have conducted 4 different simultaneous tests (LR, UI, FI, and LI) with the two-sample t -test (T) and the Satterthwaite one (SW). In addition, we considered the following two types of LR tests. One is the exact test (LR(P)) and the other, an asymptotic one with chi-square distribution with 2 d.f. (LR(A)). For Ψ in the LI test, we chose Φ . We summarized the results in Table 2 with respective p -values. We applied the permutation principle with 10,000 iterations with the Monte-Carlo approach to obtain

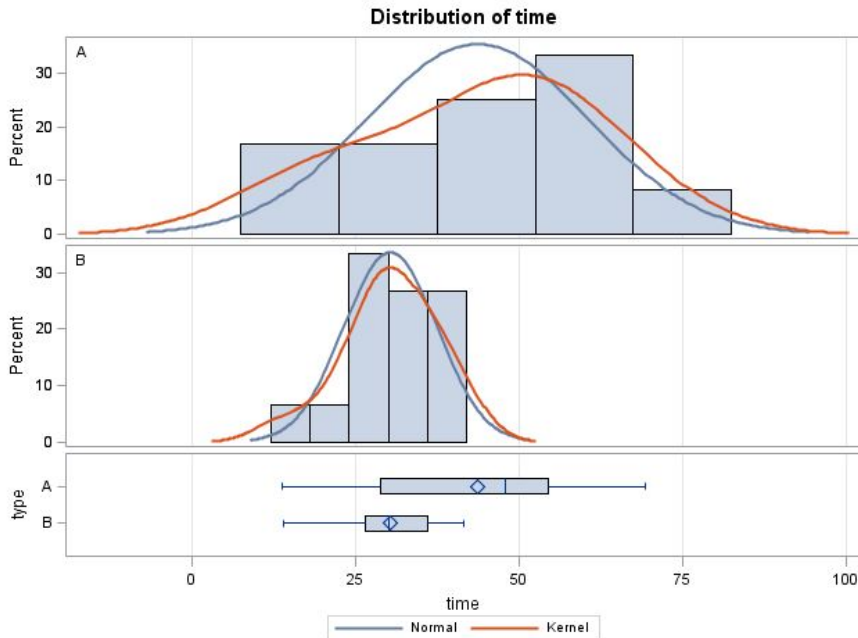


Figure 1: Plots for the data in Table 1.

Table 2: Scatter plot for the data in Table 1

Test	LR(P)	LR(A)	UI	FI	LI	T	SW
<i>p</i> -value	0.0006	0.0554	0.0029	0.0004	0.0001	0.0051	0.0112

p-values. Table 2 indicates that all the simultaneous tests proposed in the previous section provide similar results except LR(A).

It is important to note that Moser and Stevens (1992) reported 0.0449 while we obtained 0.0051 as its *p*-value of the two-sample *t*-test. The reason for this difference may be because Moser and Stevens (1992) obtained 0.0449 by the *t*-distribution with 25 d.f. while we did 0.0051 from the permutation distribution with the computer.

We now compare the efficiency among the proposed simultaneous tests by obtaining empirical powers through a simulation study. For this comparison study, we consider all simultaneous tests used for the previous example that include the two types of LR tests. For the LI test, also we choose Φ for Ψ . Furthermore, we also include the Lepage test (LE) to compare the performance with the nonparametric test. We did not include two types of *t*-tests since this study was for a comparison among simultaneous tests. We use the permutation principle to obtain the null distributions except for the LR(A) test. We consider the following three scenarios by varying the values of $\delta = \mu_1 - \mu_2$ and $\eta = \sigma_1/\sigma_2$:

- (I) Both values of δ and η vary.
- (II) Only the value of δ varies while the value of η is fixed at $\eta = 1$.
- (III) Only the value of η varies while the value of δ is fixed at $\delta = 0$.

We also consider two more cases for (I) and (III) such as (i) $\eta > 1$ with the increment 0.2 (Tables 3 and

Table 3: Both values change with $\eta > 1$

Test	(m, n)	(δ, η)					
		(0.0, 1.0)	(0.2, 1.2)	(0.4, 1.4)	(0.6, 1.6)	(0.8, 1.8)	(1.0, 2.0)
LR(P)	(10, 10)	0.0486	0.0738	0.1494	0.2516	0.3818	0.5029
	(10, 15)	0.0469	0.0784	0.1666	0.3083	0.4618	0.5932
	(15, 10)	0.0508	0.0771	0.1849	0.3295	0.4949	0.6310
LP(A)	(10, 10)	0.0131	0.0497	0.0931	0.1417	0.1869	0.2083
	(10, 15)	0.0243	0.1078	0.2184	0.3372	0.4119	0.4757
	(15, 10)	0.0260	0.0703	0.1101	0.1518	0.1698	0.1818
UI	(10, 10)	0.0483	0.0740	0.1385	0.2222	0.3446	0.4486
	(10, 15)	0.0478	0.0786	0.1608	0.2746	0.4094	0.5325
	(15, 10)	0.0527	0.0814	0.1711	0.2927	0.4328	0.5598
FI	(10, 10)	0.0499	0.0731	0.1489	0.2531	0.3848	0.5068
	(10, 15)	0.0458	0.0773	0.1658	0.3051	0.4650	0.5956
	(15, 10)	0.0510	0.0802	0.1933	0.3372	0.4985	0.6336
LI	(10, 10)	0.0498	0.0752	0.1413	0.2453	0.3714	0.4904
	(10, 15)	0.0466	0.0729	0.1604	0.2947	0.4527	0.5818
	(15, 10)	0.0487	0.0805	0.1891	0.3336	0.4923	0.6214
LE	(10, 10)	0.0447	0.0690	0.1197	0.1836	0.2523	0.3132
	(10, 15)	0.0465	0.0727	0.1310	0.2160	0.3075	0.3768
	(15, 10)	0.0488	0.0810	0.1485	0.2339	0.3115	0.3854

Table 4: Both values change with $\eta < 1$

Test	(m, n)	(δ, η)					
		(0.0, 1.0)	(0.2, 0.9)	(0.4, 0.8)	(0.6, 0.7)	(0.8, 0.6)	(1.0, 0.5)
LR(P)	(10, 10)	0.0486	0.0698	0.1542	0.3298	0.5941	0.8464
	(10, 15)	0.0469	0.0703	0.1804	0.4081	0.7038	0.9248
	(15, 10)	0.0508	0.0696	0.1668	0.3905	0.6859	0.9206
LP(A)	(10, 10)	0.0131	0.0051	0.0006	0.0000	0.0000	0.0000
	(10, 15)	0.0243	0.0061	0.0009	0.0000	0.0000	0.0000
	(15, 10)	0.0260	0.0101	0.0023	0.0003	0.0000	0.0000
UI	(10, 10)	0.0483	0.0699	0.1464	0.2990	0.5289	0.7745
	(10, 15)	0.0478	0.0670	0.1540	0.3388	0.5912	0.8302
	(15, 10)	0.0527	0.0701	0.1643	0.3476	0.6024	0.8542
FI	(10, 10)	0.0499	0.0683	0.1549	0.3336	0.6012	0.8556
	(10, 15)	0.0458	0.0734	0.1753	0.3977	0.6963	0.9234
	(15, 10)	0.0510	0.0682	0.1702	0.3979	0.6927	0.9242
LI	(10, 10)	0.0498	0.0646	0.1470	0.3189	0.5840	0.8533
	(10, 15)	0.0466	0.0730	0.1758	0.3955	0.7049	0.9246
	(15, 10)	0.0487	0.0684	0.1641	0.3841	0.6882	0.9267
LE	(10, 10)	0.0447	0.0546	0.0824	0.1122	0.1383	0.1522
	(10, 15)	0.0465	0.0562	0.0802	0.0876	0.0692	0.0304
	(15, 10)	0.0488	0.0566	0.1029	0.1648	0.2245	0.2937

6) and (ii) $\eta < 1$ with the decrement 0.1 (Tables 4 and 7) when the value of η varies. However, we only consider the case of increase of δ with the increment 0.2 when the value of δ varies (Tables 3–5). For this simulation study, we generate pseudo-random numbers from the standard normal distribution with sample sizes (10, 10), (10, 15) and (15, 10) for the pair (m, n) . The simulation has been conducted with 10,000 iterations with a Monte-Carlo approach with the SAS/IML PC-version. When the permutation principle has been applied, 2,000 iterations have also been conducted with a Monte-Carlo method for each simulation to obtain the null distribution.

When both values of δ and η vary (Tables 3 and 4), the LR(P), FI and LI tests show similar results and perform relatively well. For the case that only δ changes, the UI test achieves better performance

Table 5: Only mean changes when variance is fixed at $\eta = 1$

Test	(m, n)	(δ, η)					
		(0.0, 1.0)	(0.2, 1.0)	(0.4, 1.0)	(0.6, 1.0)	(0.8, 1.0)	(1.0, 1.0)
LR(P)	(10, 10)	0.0486	0.0626	0.1114	0.1933	0.3157	0.4633
	(10, 15)	0.0469	0.0624	0.1207	0.2319	0.3792	0.5408
	(15, 10)	0.0508	0.0624	0.1207	0.2319	0.3792	0.5408
LP(A)	(10, 10)	0.0131	0.0135	0.0108	0.0083	0.0070	0.0052
	(10, 15)	0.0243	0.0242	0.0193	0.0140	0.0136	0.0094
	(15, 10)	0.0260	0.0235	0.0186	0.0151	0.0130	0.0109
UI	(10, 10)	0.0483	0.0640	0.1086	0.1926	0.3213	0.4713
	(10, 15)	0.0478	0.0561	0.0975	0.1901	0.3191	0.4664
	(15, 10)	0.0527	0.0702	0.1428	0.2653	0.4289	0.6148
FI	(10, 10)	0.0499	0.0630	0.1126	0.1920	0.3139	0.4595
	(10, 15)	0.0458	0.0590	0.1059	0.1955	0.3242	0.4685
	(15, 10)	0.0510	0.0648	0.1395	0.2526	0.4112	0.5851
LI	(10, 10)	0.0498	0.0600	0.1061	0.1677	0.2629	0.3731
	(10, 15)	0.0466	0.0583	0.1059	0.1847	0.2851	0.3997
	(15, 10)	0.0487	0.0621	0.1175	0.2079	0.3220	0.4621
LE	(10, 10)	0.0447	0.0585	0.1012	0.1692	0.2604	0.3465
	(10, 15)	0.0465	0.0627	0.1081	0.1761	0.2410	0.2911
	(15, 10)	0.0488	0.0659	0.1266	0.2309	0.3500	0.4913

Table 6: Only variance changes for $\eta > 1$ when mean is fixed at $\delta = 0$

Test	(m, n)	(δ, η)					
		(0.0, 1.0)	(0.0, 1.2)	(0.0, 1.4)	(0.0, 1.6)	(0.0, 1.8)	(0.0, 2.0)
LR(P)	(10, 10)	0.0486	0.0611	0.1031	0.1493	0.2283	0.3135
	(10, 15)	0.0469	0.0658	0.1121	0.1924	0.2828	0.3808
	(15, 10)	0.0508	0.0612	0.1247	0.2095	0.3194	0.4241
LP(A)	(10, 10)	0.0131	0.0520	0.0995	0.1629	0.2242	0.2660
	(10, 15)	0.0243	0.1116	0.2320	0.3715	0.4786	0.5627
	(15, 10)	0.0260	0.0729	0.1198	0.1765	0.2146	0.2446
UI	(10, 10)	0.0483	0.0633	0.1031	0.1510	0.2324	0.3212
	(10, 15)	0.0478	0.0699	0.1314	0.2222	0.3330	0.4359
	(15, 10)	0.0527	0.0621	0.1133	0.1810	0.2751	0.3724
FI	(10, 10)	0.0499	0.0619	0.1013	0.1458	0.2198	0.3024
	(10, 15)	0.0458	0.0646	0.1174	0.2033	0.2913	0.3902
	(15, 10)	0.0510	0.0629	0.1164	0.1904	0.2860	0.3812
LI	(10, 10)	0.0498	0.0606	0.0959	0.1303	0.1839	0.2470
	(10, 15)	0.0466	0.0595	0.0990	0.1559	0.2144	0.2705
	(15, 10)	0.0487	0.0625	0.1151	0.1840	0.2588	0.3281
LE	(10, 10)	0.0447	0.0555	0.0679	0.0692	0.0800	0.0909
	(10, 15)	0.0465	0.0582	0.0661	0.0766	0.0795	0.0791
	(15, 10)	0.0488	0.0624	0.0781	0.0862	0.1023	0.1063

than other ones. We note that the case that $\eta = 1$ seems to have the lowest power while the case that $\eta > 1$ achieves the highest one when we compare the three cases that $\eta > 1$, $\eta < 1$ and $\eta = 1$ in Tables 3–5. This phenomenon may be because we deal with simultaneous test procedures for mean and variance. We also note that the performances of the UI, FI and LI tests are often highly dependent upon the order of sample sizes such as (10, 15) and (15, 10) when only η varies (Tables 6 and 7). We observe that the empirical powers of these tests for (15, 10) are lower than the tests for (10, 10) in Table 7. In general, the LR(P) shows relatively consistent performance for all cases. Especially, it is noticeable that the performances of LE are poor and even indicate a test bias when $\eta < 1$. Also, the LR(A) is very conservative for H_0 and even shows absurd results related to the test bias for H_1 when

Table 7: Only variance changes for $\eta < 1$ when mean is fixed at $\delta = 0$

Test	(m, n)	(δ, η)					
		(0.0, 1.0)	(0.0, 1.2)	(0.0, 1.4)	(0.0, 1.6)	(0.0, 1.8)	(0.0, 2.0)
LR(P)	(10, 10)	0.0486	0.0530	0.0746	0.1059	0.1852	0.3205
	(10, 15)	0.0469	0.0522	0.0742	0.1297	0.2417	0.4239
	(15, 10)	0.0508	0.0538	0.0765	0.1184	0.2207	0.3884
LP(A)	(10, 10)	0.0131	0.0052	0.0009	0.0000	0.0000	0.0000
	(10, 15)	0.0243	0.0063	0.0012	0.0000	0.0000	0.0000
	(15, 10)	0.0260	0.0103	0.0026	0.0004	0.0000	0.0000
UI	(10, 10)	0.0483	0.0532	0.0771	0.1056	0.1854	0.3303
	(10, 15)	0.0478	0.0565	0.0858	0.1566	0.2827	0.4884
	(15, 10)	0.0527	0.0463	0.0561	0.0836	0.1519	0.2916
FI	(10, 10)	0.0499	0.0533	0.0731	0.1044	0.1796	0.3101
	(10, 15)	0.0458	0.0525	0.0830	0.1524	0.2800	0.4739
	(15, 10)	0.0510	0.0491	0.0659	0.0880	0.1524	0.2674
LI	(10, 10)	0.0498	0.0541	0.0683	0.0953	0.1529	0.2511
	(10, 15)	0.0466	0.0516	0.0746	0.1334	0.2279	0.3732
	(15, 10)	0.0487	0.0492	0.0677	0.0919	0.1488	0.2207
LE	(10, 10)	0.0447	0.0395	0.0337	0.0236	0.0206	0.0148
	(10, 15)	0.0465	0.0403	0.0325	0.0267	0.0202	0.0089
	(15, 10)	0.0488	0.0380	0.0308	0.0232	0.0179	0.0114

$\eta < 1$ (Tables 4 and 7). Finally, we notice that all simultaneous tests, except LR(P), show the test bias for case III when $\eta < 1$ (Table 7). The following tables highlight all biased results with boldfaced figures.

4. Discussions for some related topics

From the simulation results, we noted that the performance of LR(A) is inferior to any tests considered in this study. Even the LR(A) shows the bias of test when the ratio of variances is less than 1 or the value of mean only changes. Therefore one should be careful to apply LR(A) test to real data. However, one may expect that the performance of LR(A) would improve if the sample sizes increase.

We have used the permutation principle to obtain the distributions of the proposed tests since it would be difficult to derive distributions of the combining functions and LR statistic in a theoretical manner. The reason for this difficulty may be due to statistics that consist in the LR statistic and combining functions that are not independent. One may consider to use the bootstrap (Efron, 1979) method, which is also a re-sampling one, to obtain the null distributions of combining functions. The difference between the two methods is as follows. The permutation method re-samples without replacement versus the bootstrap one with replacement. However, it is known that sometimes the results between the two methods are quite different in some cases (Good, 2000). If one can use either one for the testing problem, it is recommended to choose the permutation method since the permutation method estimates the distribution function while the bootstrap one does the parameter.

The choices of a and b from the distribution of F or more precisely, of $(m - 1)F/(n - 1)$ should satisfy that

$$G(b) - G(a) = 1 - \alpha$$

and

$$\left(\frac{a}{b}\right)^{\frac{(n-3)}{2}} = \left(\frac{1 + (n - 1)a/(m - 1)}{1 + (n - 1)b/(m - 1)}\right)^{\frac{-(m+n-2)}{2}} \tag{4.1}$$

for the size α LR test, where G is the cumulative F -distribution function with $n - 1$ and $m - 1$ d.f. However, it would be difficult to find exact values of a and b . Fortunately, the traditionally convenient choices such that $a = g_{\alpha/2}(n - 1, m - 1)$ and $b = g_{1-\alpha/2}(n - 1, m - 1)$ approximately satisfy (4.1) (Bickel and Doksum, 1977), where $g_{\alpha}(n - 1, m - 1)$ is an α^{th} quantile point of the F -distribution with $n - 1$ and $m - 1$ d.f.

The UI test was initiated by Roy (1957) and applied mainly to multivariate data. As we might see during the construction of the test statistics, the UI test may indicate which component or components might be responsible for the rejection of H_0 when H_0 is rejected. Unfortunately the LR test does not have this property. Therefore, one cannot ask more details about the reason for the rejection of H_0 . In addition, the UI test procedure can lead to the construction of simultaneous confidence intervals for parameters (Mardia *et al.*, 1979). Berger (1996) showed that the intersection-union test which is the reverse of the UI test for the null hypothesis and its alternative, is uniformly more powerful than the LR ones in some cases for the multivariate data.

One may also note that the combining functions FI and LI used in the Section 2 weigh equally for the two components T and F in their importance. The considerations of the mean and variance are equally important when one uses FI or UI for the statistical inference on the mean and variance simultaneously. Therefore, when one considers that one is more important than the other, one may use the following weighted combining functions such as

$$FI_w = -2 \{w_F \log(1 - \Lambda_F), w_T \log(1 - \Lambda_T)\}$$

and

$$LI_w = w_F \Psi^{-1}(1 - \Lambda_F) + w_T \Psi^{-1}(1 - \Lambda_T),$$

where w_F and w_T mean weights and satisfy that $w_F + w_T = 1$. Won *et al.* (2009) discussed how to determine an optimal choice of the weights for the Liptak function under some conditions. Pesarin and Salmaso (2010) reviewed and investigated extensively combining functions with several desirable properties and useful selection guidelines based on the permutation principle. In passing, we note that the UI test approach takes naturally the Tippett combining function for the construction of a test statistic from the individual partial tests to test H_0 .

Acknowledgements

This work was supported by a 2-year research grant of Cheongju University (2016.03.01–2018.02.28).

Appendix: Derivation of equation

First we note that

$$\begin{aligned} \sum_{i=1}^m (X_i - \hat{\mu})^2 &= \sum_{i=1}^m (X_i - \bar{X})^2 + m(\bar{X} - \hat{\mu})^2 \\ &= \sum_{i=1}^m (X_i - \bar{X})^2 + \frac{mn^2}{(m+n)^2} (\bar{X} - \bar{Y})^2. \end{aligned}$$

In addition, we have

$$\sum_{j=1}^n (Y_j - \hat{\mu})^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2 + \frac{m^2 n}{(m+n)^2} (\bar{X} - \bar{Y})^2.$$

Since,

$$\begin{aligned} S_{12}^2 &= \frac{1}{m+n} \left\{ \sum_{i=1}^m (X_i - \hat{\mu})^2 + \sum_{j=1}^n (Y_j - \hat{\mu})^2 \right\} \\ &= \frac{1}{m+n} \left\{ \sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2 + \frac{mn}{(m+n)} (\bar{X} - \bar{Y})^2 \right\}. \end{aligned}$$

We have,

$$\begin{aligned} \left(\frac{S_{12}^2}{S_1^2} \right)^{\frac{m}{2}} &= \left(\frac{m}{m+n} \right)^{\frac{m}{2}} \left\{ 1 + \frac{nS_2^2}{mS_1^2} + \frac{mn}{m+n} \frac{(\bar{X} - \bar{Y})^2}{mS_1^2} \right\}^{\frac{m}{2}} \\ &= \left(\frac{m}{m+n} \right)^{\frac{m}{2}} \left\{ 1 + \frac{nS_2^2}{mS_1^2} + \frac{mn}{m+n} \frac{(\bar{X} - \bar{Y})^2}{mS_1^2 + nS_2^2} \frac{mS_1^2 + nS_2^2}{mS_1^2} \right\}^{\frac{m}{2}} \\ &\propto \left(1 + \frac{nS_2^2}{mS_1^2} \right)^{\frac{m}{2}} \left\{ 1 + \frac{mn}{m+n} \frac{(\bar{X} - \bar{Y})^2}{mS_1^2 + nS_2^2} \right\}^{\frac{m}{2}}. \end{aligned} \tag{A.1}$$

Also with similar argument used for $(S_{12}^2/S_1^2)^{m/2}$, we obtain

$$\left(\frac{S_{12}^2}{S_2^2} \right)^{\frac{n}{2}} \propto \left(1 + \frac{mS_1^2}{nS_2^2} \right)^{\frac{n}{2}} \left\{ 1 + \frac{mn}{m+n} \frac{(\bar{X} - \bar{Y})^2}{mS_1^2 + nS_2^2} \right\}^{\frac{n}{2}}. \tag{A.2}$$

Therefore, we obtain the expression (2.2) by multiplying (A.1) and (A.2).

References

- Berger RL (1996). Likelihood Ratio Tests and Intersection-Union Tests (Institute of Statistics Mimeo Series No. 2288), North Carolina State University, Department of Statistics, Raleigh, NC.
- Bickel PJ and Doksum KA (1977). *Mathematical Statistics-Basic Ideas and Selected Topics*, Holden-Day, San Francisco.
- Choudhari P, Kundu D, and Misra N (2001). Likelihood ratio test for simultaneous testing of the mean and the variance of a normal distribution, *Journal of Statistical Computation and Simulation*, **71**, 313–333.
- Efron B (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics*, **7**, 1–26.
- Fisher RA (1932). *Statistical Methods for Research Workers* (4th ed), Oliver & Boyd, Edinburgh.
- Flury BD, Nel DG, and Pienaar I (1995). Simultaneous detection of shift in means and variances, *Journal of the American Statistical Association*, **90**, 1474–1481.

- Good PI (2000). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (2nd ed), Springer, New York.
- Lepage Y (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics, *Biometrika*, **58**, 213–217.
- Liptak T (1958). On the combination of independent tests, *Magyar Tudományos Akadémia Matematikai Kutató Intézetének közleményei*, **3**, 127–141.
- Mardia KV, Kent JT, and Bibby JM (1979). *Multivariate Analysis*, Academic Press, London.
- Milton JS and Arnold JC (2003). *Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences* (4th ed), McGraw-Hill, Boston.
- Moser BK and Stevens GR (1992). Homogeneity of variance in the two-sample means test, *The American Statistician*, **46**, 19–21.
- Park HI (2015). Nonparametric simultaneous test procedures, *Revista Colombiana de Estadística*, **38**, 107–121.
- Park HI and Han KJ (2013). A simultaneous test for mean and variance based on the likelihood ratio principle, *Journal of Korean Data Analysis Society*, **15**, 1733–1742.
- Pesarin F and Salmaso L (2010). *Permutation Tests for Complex Data: Theory, Applications, and Software*, John Wiley & Sons, New York.
- Randles RH and Wolfe DA (1979). *Introduction to the Theory of Nonparametric Statistics*, John Wiley & Sons, New York.
- Roy SN (1957). *Some Aspects of Multivariate Analysis*, John Wiley & Sons, New York.
- Tippett LHC (1931). *The Methods of Statistics*, Williams & Norgate Ltd., London.
- Won S, Morris N, Lu Q, and Elston RC (2009). Choosing an optimal method to combine p -values, *Statistics in Medicine*, **28**, 1537–1553.

Received August 11, 2016; Revised November 7, 2016; Accepted November 8, 2016