Hypothesis testing for normal distributions

Zhou, Yuejin; Ho, Sze Yui ; Liu, Jiahua; Tong, Tiejun

Published in:
Statistics and its Interface

DOI:
10.4310/SII.2020.v13.n2.a3

Published: 30/01/2020

Document Version:
Publisher's PDF, also known as Version of record

Statistics and Its Interface Volume 13 (2020) 167–179

Hypothesis testing for normal distributions: a unified framework and new developments

Yuejin Zhou, Sze-Yui Ho, Jiahua Liu, and Tiejun Tong*

Hypothesis testing for normal distributions is one important problem in statistics and related fields including management science, engineering science and medical science. In this paper, from a very unique perspective, we propose a unified framework to comprehensively review the existing literature on the one- and two-sample testing problems of normal distributions. The unified framework has integrated the literature in a way that it includes most commonly used tests as special cases, including the one-sample mean test, the one-sample variance test, the two-sample mean test, the two-sample variance test, and the Behrens-Fisher test. The unified framework has also put forward two new hypothesis tests that are rarely studied in the literature. To complete the puzzle, we propose two likelihood ratio test statistics to solve those new testing problems. Simulation studies and real data examples are also provided to demonstrate that our proposed test statistics are appropriate for practical implementation.

Keywords and phrases: Hypothesis test, Likelihood ratio test, Normal distribution, Unified framework.

1. INTRODUCTION

The normal distribution, also known as the Gaussian distribution, is one of the most important distributions in statistics and probability. Hypothesis testing for one or two normal distributions have been extensively studied in the literature for more than 100 years. Among the available tests, testing equality of two normal distributions is one frequently encountered problem and it has wide applications in various fields, including management science, engineering science and medical science. Such a test is needed, for instance, to assess the acceptability of a new design or treatment to some standard medicine or to proceed the quality control of a new product to the existing ones.

Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ and $\boldsymbol{Y} = \{Y_1, \ldots, Y_m\}$ be two independent random samples from the normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Recall that a normal distribution is completely determined by its two parameters: the mean value and the variance. To test equality of two normal distributions, it is hence equivalent to testing the following hypothesis:

$$(1) \qquad \begin{aligned} &H_0 : \mu_1 = \mu_2 \text{ and } \sigma_1^2 = \sigma_2^2 \\ \text{versus} \quad &H_1 : \mu_1 \neq \mu_2 \text{ or } \sigma_1^2 \neq \sigma_2^2. \end{aligned}$$

In 1930, Pearson and Neyman proposed a likelihood ratio test (LRT) for hypothesis (1) with the test statistic given by

$$\Lambda_{n,m} = \frac{\left(\sum_{i=1}^n (X_i - \bar{X})^2/n\right)^{\frac{n}{2}} \left(\sum_{j=1}^m (Y_j - \bar{Y})^2/m\right)^{\frac{m}{2}}}{\left\{\left(\sum_{i=1}^n (X_i - \hat{\mu})^2 + \sum_{j=1}^m (Y_j - \hat{\mu})^2\right)/(n+m)\right\}^{\frac{n+m}{2}}},$$

where $\bar{X}$, $\bar{Y}$ and $\hat{\mu}$ are the sample means of the $\boldsymbol{X}$ sample, the $\boldsymbol{Y}$ sample and the pooled sample, respectively. For simplicity, we refer to the above LRT test as the P&N test and the corresponding hypothesis as the P&N hypothesis. Under the null hypothesis, the limiting distribution of the test statistic $\Lambda_{n,m}$ follows a uniform distribution on the interval $[0, 1]$ when n and m are both large. While for finite n and m, Pearson (1930) proposed to approximate the exact distribution of $\Lambda_{n,m}$ by a beta distribution. In addition, [45] and [4] proposed other methods for testing the P&N hypothesis.

Due to the complexity in the P&N test and its variants, hypothesis (1) is often overlooked or may not even be introduced in most introductory statistics textbooks. In contrast, one often prefer to introduce some simpler hypothesis tests in the textbooks, e.g., testing whether the two population means are equal, or whether the two population variances are equal. In this paper, according to whether (or to what extent) the prior knowledge of means and variances is available, we propose a unified framework to integrate the existing literature so that it includes most commonly used tests as special cases, including the one-sample mean test, the one-sample variance test, the two-sample mean test, and the two-sample variance test.

To present the main idea of our unified framework, we treat "Means" as the row variable and "Variances" as the column variable in Table 1, each with 4 different levels of assumptions. To start with, we make no assumption or have no prior knowledge on the population means and variances. Then to test equality of two normal distributions, we have

*Corresponding author.

Table 1. A unified framework of hypothesis testing for normal distributions. (I): two-sample mean test; (II): two-sample variance test; (III): one-sample mean test, (IV): one-sample variance test; (V): the equality test of two normal distributions with one mean known; (VI): the equality test of two normal distributions with one variance known. In addition, the P&N test denotes the equality test of two normal distributions, and the K&S test denotes the equality test of one normal distribution with a reference distribution

Means \ Variances	No assumption	Assuming equal variances	Assuming unequal variances	Assuming one variance known
No assumption	P&N test	(I)		(VI)
Assuming equal means	(II)	No test		(IV)
Assuming unequal means				
Assuming one mean known	(V)	(III)		K&S test

no simple choice but to deal with the composite hypothesis in (1). In this setting, some sophisticated test procedures, such as the P&N test, are needed.

Now we take a step forward. If the two population variances are assumed or known to be equal, then to test hypothesis (1), we need no longer to take into account the term $\sigma_1^2 = \sigma_2^2$ in the null hypothesis. As a consequence, the original test is simplified as a two-sample mean test:

$$(2) \quad H_0^{(\mathrm{I})}: \mu_1 = \mu_2 \qquad \text{versus} \qquad H_1^{(\mathrm{I})}: \mu_1 \neq \mu_2.$$

On the contrary, if the two population variances are unequal, then the null hypothesis in (1) is false and it will be rejected automatically. Given, however, that the two normal distributions are not the same, in many applications one may still be interested in testing hypothesis (2). Under such a scenario, it results in the famous Behrens-Fisher problem [5, 18, 24, 51].

On the other side, if the two population means are assumed or known to be equal, then the term $\mu_1 = \mu_2$ will be removed from the null hypothesis in (1). This leads to the two-sample variance test as follows:

$$(3) \quad H_0^{(\mathrm{II})}: \sigma_1^2 = \sigma_2^2 \qquad \text{versus} \qquad H_1^{(\mathrm{II})}: \sigma_1^2 \neq \sigma_2^2.$$

Conversely, if the two population means are unequal, then the null hypothesis of (1) will be rejected automatically. In this setting, however, one may still perform the two-sample variance test in (3).

To achieve the two-sample tests in (2) and (3), we have imposed assumptions only on the population variances or only on the population means. In what follows, we will impose assumptions on both means and variances in order to draw connections between hypothesis (1) and the one-sample tests. In the extreme case, if we assume that $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2$, then the null hypothesis in (1) is true and no further test is needed. In contrast, if either $\mu_1 \neq \mu_2$ or $\sigma_1^2 \neq \sigma_2^2$ is assumed, then the null hypothesis in (1) is false, and once again, no further test is needed. These extreme cases correspond to the four middle entries in Table 1, marked with “No test”.

To derive the one-sample mean test, we assume that one population mean is known. Without loss of generality, let the mean of the second population be known, denoted by $\mu_2 = \mu_0$. With this prior knowledge, if we further assume that the two variances are equal (or unequal), then hypothesis (1) reduces to the one-sample mean test as follows:

$$(4) \quad H_0^{(\mathrm{III})}: \mu_1 = \mu_0 \qquad \text{versus} \qquad H_1^{(\mathrm{III})}: \mu_1 \neq \mu_0.$$

Accordingly, if the variance of the second population is known (denoted by $\sigma_2^2 = \sigma_0^2$) and the two population means are equal (or unequal), then hypothesis (1) reduces to the one-sample variance test as follows:

$$(5) \quad H_0^{(\mathrm{IV})}: \sigma_1^2 = \sigma_0^2 \qquad \text{versus} \qquad H_1^{(\mathrm{IV})}: \sigma_1^2 \neq \sigma_0^2.$$

If, instead, the mean and variance of the second population are both known as μ_0 and σ_0^2, then hypothesis (1) becomes a simultaneous test of the mean and variance of a normal distribution:

$$(6) \quad \begin{aligned} &H_0^{(\mathrm{KS})}: \mu_1 = \mu_0 \text{ and } \sigma_1^2 = \sigma_0^2 \\ \text{versus} \quad &H_1^{(\mathrm{KS})}: \mu_1 \neq \mu_0 \text{ or } \sigma_1^2 \neq \sigma_0^2. \end{aligned}$$

Hypothesis (6) was first introduced in [25] and [58]. For simplicity, we refer to it as the K&S test in Table 1.

Apart from the aforementioned tests, there are still two interesting settings in Table 1 remaining unvisited: one is located in the bottom-left corner and the other is located in the top-right corner. Specifically, if we assume $\mu_2 = \mu_0$ is known but make no assumption on the population variances, then hypothesis (1) yields the hypothesis test in the bottom-left corner:

$$(7) \quad \begin{aligned} &H_0^{(\mathrm{V})}: \mu_1 = \mu_0 \ \text{ and } \ \sigma_1^2 = \sigma_2^2 \\ \text{versus} \quad &H_1^{(\mathrm{V})}: \mu_1 \neq \mu_0 \ \text{ or } \ \sigma_1^2 \neq \sigma_2^2. \end{aligned}$$

In contrast, if we assume $\sigma_2^2 = \sigma_0^2$ is known but make no assumption on the population means, then we have the hypothesis test in the top-right corner:

$$(8) \quad \begin{aligned} &H_0^{(\mathrm{VI})}: \mu_1 = \mu_2 \ \text{ and } \ \sigma_1^2 = \sigma_0^2 \\ \text{versus} \quad &H_1^{(\mathrm{VI})}: \mu_1 \neq \mu_2 \ \text{ or } \ \sigma_1^2 \neq \sigma_0^2. \end{aligned}$$

To the best of our knowledge, hypotheses (7) and (8) have rarely been studied in the literature. Although not very common in practice, we propose to derive two likelihood ratio tests for the new testing problems, and hence complete the puzzle that we have laid out in Table 1.

The rest of the paper is organized as follows. In Section 2, we provide a review on the commonly used one- and two-sample tests from the perspective of LRT. In Section 3, we propose two likelihood ratio tests for the new testing problems and derive their exact and approximate null distributions, respectively. In Section 4, we conduct simulation studies to assess the performance of our proposed test statistics. In Section 5, two real data examples are analyzed and they demonstrate the practical values of our proposed test statistics. Finally, we conclude the paper in Section 6 and postpone the derivations to the Appendix.

2. A REVIEW: ONE- AND TWO-SAMPLE TESTS

2.1 Two-sample tests

Following the notation in Section 1, let $\bar{X} = \sum_{i=1}^{n} X_i/n$ be the sample mean, and $S_1^2 = \sum_{i=1}^{n}(X_i - \bar{X})/(n-1)$ be the sample variance from the $\boldsymbol{X}$ sample; and let $\bar{Y} = \sum_{i=1}^{m} Y_i/m$ be the sample mean, and $S_2^2 = \sum_{i=1}^{m}(Y_i - \bar{Y})/(m-1)$ be the sample variance from the $\boldsymbol{Y}$ sample. Let also $\hat{\mu} = (n\bar{X} + m\bar{Y})/(n+m)$ be the pooled sample mean, and $S_{\text{pool}}^2 = \{(n-1)S_1^2 + (m-1)S_2^2\}/(n+m-2)$ be the pooled sample variance.

2.1.1 Two-sample mean tests

This section reviews the two-sample mean tests for hypothesis (2). We first consider the scenario of equal variances with $\sigma_1^2 = \sigma_2^2 = \sigma^2$. When σ^2 is known, the LRT statistic is given as

$$Z_1^{(\mathrm{I})} = \frac{\bar{X} - \bar{Y}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Under the null hypothesis, $Z_1^{(\mathrm{I})}$ follows a standard normal distribution. This test is well known as the Z-test. When σ^2 is unknown, the LRT statistic is

$$T_1^{(\mathrm{I})} = \frac{\bar{X} - \bar{Y}}{S_{\text{pool}}\sqrt{\frac{1}{n} + \frac{1}{m}}}, \tag{9}$$

where $S_{\text{pool}} = (S_{\text{pool}}^2)^{1/2}$ is the pooled sample standard deviation. Under the null hypothesis, $T_1^{(\mathrm{I})}$ follows a Student's t distribution with $n + m - 2$ degrees of freedom. This test is referred to as the two-sample pooled t-test.

Now we consider the scenario that the two variances are unequal. When σ_1^2 and σ_2^2 are known, the LRT statistic is given as

$$Z_2^{(\mathrm{I})} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}.$$

Under the null hypothesis, $Z_2^{(\mathrm{I})}$ follows a standard normal distribution and it results in the Z-test. When the two variances are unequal and unknown, the pooled sample variance in the test statistic (9) will be less meaningful. As an alternative, one may prefer the following "plug-in" test statistic:

$$T_2^{(\mathrm{I})} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}.$$

Under the null hypothesis, however, the test statistic $T_2^{(\mathrm{I})}$ does not follow a Student's t distribution. This is widely known as the Behrens-Fisher problem [5, 18].

[62] proposed an approximate solution for the Behrens-Fisher problem. In his method, $T_2^{(\mathrm{I})}$ is assumed to follow an approximate t distribution with ν degrees of freedom under the null hypothesis, where ν is estimated by the Welch-Satterthwaite equation [53, 54]:

$$\hat{\nu} = \left\{\frac{(S_1^2/n)^2}{n-1} + \frac{(S_2^2/m)^2}{m-1}\right\}^{-1}\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)^2.$$

This test is referred to as Welch's t-test. Other methods in this direction include but not limited to the following: [63], [19], [12], [41], [38], [27], [49], [50], [29], [59], [24], [51] and [36].

2.1.2 Two-sample variance tests

This section reviews the two-sample variance tests for hypothesis (3). The LRT statistic for this test is given as

$$F_1^{(\mathrm{II})} = \frac{S_1^2}{S_2^2}.$$

Note that $(n-1)S_1^2/\sigma_1^2$ follows a chi-square distribution with $n-1$ degrees of freedom, $(m-1)S_2^2/\sigma_2^2$ follows a chi-square distribution with $m-1$ degrees of freedom, and the two terms are independent of each other. Under the null hypothesis, $F_1^{(\mathrm{II})}$ follows an F distribution with $n-1$ and $m-1$ degrees of freedom. For more properties of this F-test, one may refer to, for example, [9], [46] and [23].

Note, however, that the classic F-test is very sensitive to the assumption that the data are normally distributed. For robust testing of hypothesis (3), [48] proposed a Wald test with the test statistic as

$$\chi_1^{(\mathrm{II})} = \frac{(S_1^2 - S_2^2)^2}{2S_1^4/(n+1) + 2S_2^4/(m+1)}.$$

Under the null hypothesis, it follows an asymptotic chi-square distribution with one degree of freedom. For more test methods including Bartlett's test, Levene test and Bootstrap test, one may refer to, for example, [3], [28], [10], [21], [13], [8] and [68].

2.1.3 Equality tests of two normal distributions

In this section, we review the equality tests of two normal distributions specified in hypothesis (1). As mentioned in Section 1, [42] constructed an LRT statistic as

$$\Lambda_{n,m} = \frac{\left(\sum_{i=1}^{n}(X_i-\bar{X})^2/n\right)^{\frac{n}{2}}\left(\sum_{j=1}^{m}(Y_j-\bar{Y})^2/m\right)^{\frac{m}{2}}}{\left\{\left(\sum_{i=1}^{n}(X_i-\hat{\mu})^2+\sum_{j=1}^{m}(Y_j-\hat{\mu})^2\right)/(n+m)\right\}^{\frac{n+m}{2}}}.$$

Under the null hypothesis, the limiting distribution of the test statistic $\Lambda_{n,m}$ follows a uniform distribution on the interval [0,1] when n and m are both large. While for finite n and m, [42] proposed to approximate the exact distribution of $\Lambda_{n,m}$ by a beta distribution with the density function

$$f(x)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1},$$

where $p=-a(a^2-a+b)/b$ and $q=(a-1)(a^2-a+b)/b$ with

$$a = E(\Lambda_{n,m}) = \left(\frac{(n+m)^{n+m}}{n^n m^m}\right)^{\frac{1}{2}}\frac{\Gamma(\frac{2n-1}{2})\Gamma(\frac{2m-1}{2})}{\Gamma(\frac{2(n+m)-1}{2})} \times\frac{\Gamma(\frac{n+m-1}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{m-1}{2})}$$

and

$$b = \mathrm{Var}(\Lambda_{n,m}) = \frac{(n+m)^{n+m}}{n^n m^m}\frac{\Gamma(\frac{3n-1}{2})\Gamma(\frac{3m-1}{2})}{\Gamma(\frac{3(n+m)-1}{2})} \times\frac{\Gamma(\frac{n+m-1}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{m-1}{2})}-a^2.$$

To name a few other popular methods for testing hypothesis (1), [45] applied Fisher's method to combine two separated hypotheses: one is for the equality of two means and the other is for the equality of two variances. Recently, [4] proposed a test statistic based on the maximum likelihood estimator of Weitzman's overlapping coefficient. Note that the Weitzman's overlapping coefficient is defined as $\Delta=\int\min\{f_1(x),f_2(x)\}dx$, where $f_1(x)$ and $f_2(x)$ are the density functions of two normal distributions.

2.2 One-sample tests

In this section, we review the one-sample tests listed in Table 1, including the one-sample mean tests, the one-sample variance tests, and the equality tests of one normal distribution with a reference distribution. Without loss of generality, we assume that the second population is known. All other notations and assumptions are the same as in Section 2.1.

2.2.1 One-sample mean tests

The one-sample mean tests for hypothesis (4) are commonly introduced in elementary statistics textbooks, see, for example, [14] and [57]. When $\sigma_1^2=\sigma_0^2$ is known, the LRT statistic is

$$Z_1^{(\mathrm{III})}=\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma_0}.$$

Under the null hypothesis, $Z_1^{(\mathrm{III})}$ follows a standard normal distribution and the test leads to a Z-test.

When σ_1^2 is unknown, the LRT statistic is given as

$$T_1^{(\mathrm{III})}=\frac{\sqrt{n}(\bar{X}-\mu_0)}{S_1}.$$

Under the null hypothesis, $T_1^{(\mathrm{III})}$ follows a Student's t distribution with $n-1$ degrees of freedom. This test is known as the one-sample t-test. [33] and [34] further studied its power and the type II error, respectively. For more tests of hypothesis (4), one may refer to [64], [30], [16], and the references therein.

2.2.2 One-sample variance tests

For the one-sample variance tests of hypothesis (5), the LRT statistic is given as

$$\chi_1^{(\mathrm{IV})}=\frac{(n-1)S_1^2}{\sigma_0^2}.$$

Under the null hypothesis, $\chi_1^{(\mathrm{IV})}$ follows a chi-square distribution with $n-1$ degrees of freedom. This test is known as the χ^2-test.

Note that the usual (equal-tail) procedure for the above χ^2-test is to reject the null hypothesis at the significance level of α if $(n-1)s_1^2/\sigma_0^2\le\chi^2_{n-1,\alpha/2}$ or $(n-1)s_1^2/\sigma_0^2\ge\chi^2_{n-1,1-\alpha/2}$, where s_1^2 is the observed value of S_1^2 and $\chi^2_{v,\beta}$ is the upper β percentile of the chi-square distribution with v degrees of freedom. Nevertheless, such a procedure may lead to a biased test ([26]). For more choices of the critical region and the unbiased tests in the literature, one may refer to, for example, [35], [17], [55], [47], [61] and [37].

2.2.3 Equality tests of one normal distribution with a reference distribution

In many practical situations, it is desired to make a decision on the mean μ and the variance σ^2 simultaneously, e.g., in constructing the joint confidence intervals for μ and σ^2. This results in the K&S test in the bottom-right corner of Table 1 ([25, 58]). For testing (6), many methods have been developed in the literature. Among them, [2] proposed an asymptotic LRT statistic as

$$\chi_1^{(\mathrm{KS})}=\left(\frac{S_1^2}{\sigma_0^2}\right)^{n/2}\exp\left\{-\frac{nS_1^2}{2\sigma_0^2}-\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{2\sigma_0^2}\right\}.$$

Under the null hypothesis, $-2\log\chi_1^{(\mathrm{KS})}$ follows an asymptotic chi-square distribution with 2 degrees of freedom.

More recently, [11] revisited the exact LRT statistic as

$$T_1^{(\mathrm{KS})} = \left(\frac{e}{n}\right)^{n/2} U_1^{n/2}\exp\{-(U_1+U_2)/2\},$$

where $U_1 = nS_1^2/\sigma_0^2$ and $U_2 = n(\bar{X}-\mu_0)^2/\sigma_0^2$. They derived the exact distribution of $T_1^{(\mathrm{KS})}$ and also reported the critical values for the first time in the literature. For other tests of the K&S hypothesis, one may refer to, for example, [66], [65], [15], [52], [20], [67], [39], [22] and [40].

3. NEW DEVELOPMENTS

In this section, we consider the hypothesis testing problems with partial information known on the population means and variances. The situation where one variance is known was first studied in [31] in the framework of two-sample mean tests. They developed a Welch type test statistic to account for the information in the known variance, and then proposed to approximate the null distribution by Student's t distribution. [43] further improved their approximate null distribution by providing an unbiased estimator for the degrees of freedom. To complete the puzzle, we propose some new LRT statistics for the unsolved testing problems presented in the bottom-left and top-right corners of Table 1.

3.1 Likelihood ratio test with one mean known

To our knowledge, there is little work in the literature on testing equality of two normal distributions with one mean known. Without loss of generality, we assume $\mu_2 = \mu_0$ is known. In what follows, we develop an LRT statistic for hypothesis (7):

$$\begin{aligned} & H_0^{(\mathrm{V})}: \mu_1 = \mu_0 \ \text{ and } \ \sigma_1^2 = \sigma_2^2 \\ \text{versus} \quad & H_1^{(\mathrm{V})}: \mu_1 \neq \mu_0 \ \text{ or } \ \sigma_1^2 \neq \sigma_2^2. \end{aligned}$$

We also derive its exact and approximate null distributions.

Under the null hypothesis $H_0^{(\mathrm{V})}$, it is evident that $\mu_1 = \mu_2 = \mu_0$ is known and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown. This yields the likelihood function as

$$\begin{aligned} L(\sigma^2|x,y) \;=\; & \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n+m}{2}} \exp\Big\{-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2} \\ & -\frac{\sum_{j=1}^m (y_j-\mu_0)^2}{2\sigma^2}\Big\}. \end{aligned}$$

Accordingly, under $H_0^{(\mathrm{V})}\bigcup H_1^{(\mathrm{V})}$, the likelihood function is

$$\begin{aligned} L(\mu_1,\sigma_1^2,\sigma_2^2|x,y) = & \frac{\sigma_1^{-n}\sigma_2^{-m}}{(2\pi)^{(n+m)/2}} \exp\Big\{-\frac{\sum_{i=1}^n (x_i-\mu_1)^2}{2\sigma_1^2} \\ & -\frac{\sum_{j=1}^m (y_j-\mu_0)^2}{2\sigma_2^2}\Big\}. \end{aligned}$$

By maximizing the above likelihood functions, we can derive their maximum likelihood estimators (MLEs), respectively. Further, we have the LRT statistic for hypothesis (7) as

$$\Lambda_1 = \frac{\Big\{\sum_{i=1}^n (X_i-\bar{X})^2/n\Big\}^{\frac{n}{2}} \Big\{\sum_{j=1}^m (Y_j-\mu_0)^2/m\Big\}^{\frac{m}{2}}}{\Big\{\Big(\sum_{i=1}^n (X_i-\mu_0)^2 + \sum_{j=1}^m (Y_j-\mu_0)^2\Big)/(n+m)\Big\}^{\frac{n+m}{2}}}.$$

For an exact test for hypothesis (7), we need to derive the exact distribution of Λ_1 under the null hypothesis. In Appendix A1, we show that for any $\lambda > 0$, the exact null distribution of the test statistic Λ_1 is

$$\begin{aligned} F(\lambda) &= P(\Lambda_1 \le \lambda) = 1 - P(\Lambda_1 > \lambda) \\ &= 1 - \int_{r_1}^{r_2} dw_1 \int_{\frac{\lambda^{n/m} n^{n/m} m}{(n+m)^{(n+m)/m} w_1^{n/m}}}^{1-w_1} f(w_1,w_2)dw_2, \end{aligned}$$

where r_1 and r_2 ($r_1 < r_2$) are the two roots of the function $g(w_1) = 1 - w_1 - \frac{\lambda^{2/m} n^{n/m} m}{(n+m)^{(n+m)/m} w_1^{n/m}}$ and $f(w_1,w_2)$ is the probability density function of the two-dimensional Dirichlet distribution.

Apart from the exact test, we also propose an approximate null distribution for the test statistic. Following Pearson and Neyman (1930), for finite sample sizes, we approximate the null distribution of Λ_1 as a beta distribution with the probability density function

$$f(x,p_1,q_1) = \frac{\Gamma(p_1+q_1)}{\Gamma(p_1)\Gamma(q_1)} x^{p_1-1}(1-x)^{q_1-1},$$

where $p_1 = -a_1(a_1^2 - a_1 + b_1)/b_1 > 0$ and $q_1 = (a_1-1)(a_1^2 - a_1 + b_1)/b_1 > 0$ are the two parameters of the beta distribution, determined by matching the first two moments with

$$a_1 = E(\Lambda_1) = \frac{(n+m)^{(n+m)/2}}{n^{n/2}m^{m/2}} \frac{\Gamma(\frac{n+m}{2})\Gamma(\frac{2n-1}{2})\Gamma(m)}{\Gamma(\frac{n-1}{2})\Gamma(\frac{m}{2})\Gamma(n+m)}$$

and

$$b_1 = \mathrm{Var}(\Lambda_1) = \frac{(n+m)^{(n+m)}}{n^n m^m} \frac{\Gamma(\frac{n+m}{2})\Gamma(\frac{3n-1}{2})\Gamma(\frac{3m}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{m}{2})\Gamma(\frac{3(n+m)}{2})} - a_1^2.$$

Note that a smaller value of Λ_1 supports the alternative hypothesis $H_1^{(\mathrm{V})}$. As a decision rule, we reject the null hypothesis $H_0^{(\mathrm{V})}$ at the significance level α if $p = F(\lambda_1) < \alpha$ or $p = \mathrm{Beta}(\lambda_1, p_1, q_1) < \alpha$, where λ_1 is the observed value of Λ_1, $F(\lambda_1)$ is the distribution function of Λ_1 evaluated at λ_1, and $\mathrm{Beta}(\lambda_1,p_1,q_1) = \int_0^{\lambda_1} f(x,p_1,q_1)dx$.

3.2 Likelihood ratio test with one variance known

As mentioned above, there is little work in the literature on testing equality of two normal distributions with one variance known. Assuming that $\sigma_2^2 = \sigma_0^2$ is known, we are interested in testing the hypothesis:

$$
\begin{aligned}
& H_0^{(\mathrm{VI})}: \mu_1 = \mu_2 \ \text{ and } \ \sigma_1^2 = \sigma_0^2 \\
\text{versus} \quad & H_1^{(\mathrm{VI})}: \mu_1 \neq \mu_2 \ \text{ or } \ \sigma_1^2 \neq \sigma_0^2.
\end{aligned}
$$

We will propose an LRT statistic for this new hypothesis, and derive its null distribution by the exact and approximate methods, respectively.

Under the null hypothesis $H_0^{(\mathrm{VI})}$, $\sigma_1^2 = \sigma_2^2 = \sigma_0^2$ is known and $\mu_1 = \mu_2 = \mu$ is unknown. This yields the likelihood function as

$$
\begin{aligned}
L(\mu|x,y) &= \Big(\frac{1}{2\pi\sigma_0^2}\Big)^{\frac{n+m}{2}} \exp\Big\{ -\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma_0^2} \\
&\quad -\frac{\sum_{j=1}^m (y_j-\mu)^2}{2\sigma_0^2}\Big\}.
\end{aligned}
$$

Correspondingly, under $H_0^{(\mathrm{VI})} \cup H_1^{(\mathrm{VI})}$, the likelihood function is

$$
\begin{aligned}
L(\mu_1,\mu_2,\sigma_1^2|x,y) = \frac{\sigma_1^{-n}\sigma_0^{-m}}{(2\pi)^{(n+m)/2}} \exp\Big\{ -\frac{\sum_{i=1}^n (x_i-\mu_1)^2}{2\sigma_1^2} \\
-\frac{\sum_{j=1}^m (y_j-\mu_2)^2}{2\sigma_0^2}\Big\}.
\end{aligned}
$$

By maximizing the above likelihood functions, we can obtain the MLEs for the unknown parameters. Then by the MLEs, the LRT statistic is given as

$$
\begin{aligned}
\Lambda_2 &= \Big\{\frac{\sum_{i=1}^n (X_i-\bar{X})^2}{n\sigma_0^2}\Big\}^{\frac{n}{2}} \exp\Big\{\frac{n}{2} \\
&\quad -\frac{\sum_{i=1}^n (X_i-\bar{X})^2 + n(\bar{X}-\hat{\mu})^2 + m(\bar{Y}-\hat{\mu})^2}{2\sigma_0^2}\Big\}.
\end{aligned}
$$

In Appendix A2, we show that, for any $\lambda > 0$, the exact null distribution of Λ_2 is

$$
\begin{aligned}
F(\lambda) &= P(\Lambda_2 \le \lambda) = 1 - P(\Lambda_2 > \lambda) \\
&= 1 - \int_{r_1}^{r_2} dt_1 \int_0^{n(1-\log n)+n\log t_1 - t_1 - 2\log\lambda} f(t_1,t_2)dt_2,
\end{aligned}
$$

where r_1 and r_2 ($r_1 < r_2$) are the two roots of $n(1-\log n) + n\log t_1 - t_1 - 2\log\lambda = 0$ and $f(t_1,t_2) = t_1^{(n-3)/2} t_2^{-1/2} \exp\{-(t_1+t_2)/2\}/\{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})2^{n/2}\}$.

Apart from the exact null distribution, we also develop an approximate null distribution for finite sample sizes. As in Section 2.1, we approximate the null distribution of Λ_2 as a beta distribution with the probability density function

$$
f(x,p_2,q_2) = \frac{\Gamma(p_2+q_2)}{\Gamma(p_2)\Gamma(q_2)} x^{p_2-1}(1-x)^{q_2-1},
$$

where $p_2 = -a_2(a_2^2 - a_2 + b_2)/b_2 > 0$ and $q_2 = (a_2-1)(a_2^2 - a_2 + b_2)/b_2 > 0$ are the two parameters of the beta distribution with

$$
a_2 = \Big(\frac{e}{2n}\Big)^{\frac{n}{2}} \frac{\Gamma(n-\frac{1}{2})}{\Gamma(\frac{n-1}{2})}
$$

and

$$
b_2 = 3^{-\frac{3n}{2}} \Big(\frac{2e}{n}\Big)^n \frac{\Gamma(\frac{3n-1}{2})}{\Gamma(\frac{n-1}{2})} - a_2^2.
$$

The null hypothesis $H_0^{(\mathrm{VI})}$ is rejected at the significance level α if $p = F(\lambda_2) < \alpha$ or $p = \mathrm{Beta}(\lambda_2, p_2, q_2) < \alpha$, where λ_2 is the observed value of Λ_2, $F(\lambda_2)$ is the distribution function of Λ_2 evaluated at λ_2, and $\mathrm{Beta}(\lambda_2,p_2,q_2) = \int_0^{\lambda_2} f(x,p_2,q_2)dx$.

4. SIMULATION STUDIES

4.1 The test with one mean known

To evaluate the performance of our proposed tests and compare them with the P&N test, we consider two simulation studies. One is to assess whether the proposed tests are valid so that the type I errors are well controlled under the significance level, and the other is to assess whether our proposed tests are more powerful than the existing method.

To assess the type I error under various settings, without loss of generality, we assume that the samples $\boldsymbol{X} = \{X_1,\ldots,X_n\}$ and $\boldsymbol{Y} = \{Y_1,\ldots,Y_m\}$ are both independently generated from $N(0,1)$. We consider three different combinations of the sample sizes: $(n,m) = (30,10), (100,100)$ and $(300,100)$, and three different significance levels at $\alpha = 0.01, 0.05$ and 0.1, respectively.

We repeat the simulation 10,000 times for each setting and report their average type I errors in Table 2. From the results, we note that our proposed test with the exact null distribution provides the smallest type I errors in most settings, no matter whether the sample sizes of two populations are small or not. Overall, our proposed test with the exact null distribution provides a more accurate control for the type I error.

To assess the statistical power of the tests, we simulate the samples $\boldsymbol{X} = \{X_1,\ldots,X_n\}$ independently from $N(0,1)$, and $\boldsymbol{Y} = \{Y_1,\ldots,Y_m\}$ independently from $N(1,1.5^2)$. Other settings are kept the same as before. For each setting, we repeat the simulation 10,000 times and report the average power in Table 3. Simulation results indicate that both of our tests are more powerful than the P&N test in most settings. In particular, the proposed test with the exact null distribution performs the best.

4.2 The test with one variance known

To evaluate the performance of the proposed test, we also conduct two simulation studies where the first one is to assess the type I errors and the second one is to assess the

Table 2. Average type I errors of our proposed test with one mean known under the exact null distribution (EX) and the approximate null distribution (AP) and the P&N test.

Method	Significance level α		
	0.1	0.05	0.01
		$n = 30, m = 10$	
EX	0.1009	0.0490	0.0104
AP	0.1031	0.0525	0.0107
P&N	0.1008	0.0499	0.0104
		$n = m = 100$	
EX	0.0985	0.0467	0.0094
AP	0.1014	0.0502	0.0115
P&N	0.1011	0.0504	0.0109
		$n = 300, m = 100$	
EX	0.0967	0.0490	0.0089
AP	0.0987	0.0492	0.0108
P&N	0.1033	0.0482	0.0092

Table 3. Average power of our proposed test with one mean known under the exact null distribution (EX) and the approximate null distribution (AP) and the P&N test.

Method	Significance level α		
	0.1	0.05	0.01
		$n = 30, m = 10$	
EX	0.9994	0.9981	0.9848
AP	0.9982	0.9950	0.9732
P&N	0.6655	0.5687	0.3592
		$n = m = 100$	
EX	0.8230	0.6899	0.3455
AP	0.6959	0.5501	0.2660
P&N	0.4378	0.3228	0.1258
		$n = 300, m = 100$	
EX	0.9994	0.9981	0.9848
AP	0.9982	0.9950	0.9732
P&N	0.6655	0.5687	0.3592

Table 4. Average type I errors of our proposed test with one variance known under the exact null distribution (EX) and the approximate null distribution (AP) and the P&N test.

Method	Significance level α		
	0.1	0.05	0.01
		$n = 30, m = 10$	
EX	0.0925	0.0413	0.0059
AP	0.1027	0.0532	0.0109
P&N	0.1008	0.0499	0.0104
		$n = m = 100$	
EX	0.0964	0.0446	0.0064
AP	0.1007	0.0486	0.0099
P&N	0.1011	0.0504	0.0109
		$n = 300, m = 100$	
EX	0.0430	0.0410	0.0395
AP	0.0699	0.0312	0.0060
P&N	0.1043	0.0498	0.0116

Table 5. Average power of our proposed test with one variance known under the exact null distribution (EX) and the approximate null distribution (AP) and the P&N test.

Method	Significance level α		
	0.1	0.05	0.01
		$n = 30, m = 10$	
EX	0.9560	0.8993	0.6555
AP	0.9141	0.8449	0.6286
P&N	0.6655	0.5687	0.3592
		$n = m = 100$	
EX	0.9998	0.9997	0.9941
AP	0.9998	0.9996	0.9931
P&N	0.9563	0.9237	0.8087
		$n = 300, m = 100$	
EX	0.9222	0.9790	0.9934
AP	1.0000	1.0000	1.0000
P&N	0.9943	0.9907	0.9694

statistical power. We also compare them with the P&N test. For the purpose of transparency, we follow the same simulation settings as in Section 3.1. We repeat the simulation 10,000 times for each setting and report their average type I errors and average power in Tables 4 and 5, respectively.

In comparison with the P&N test, our proposed test with the exact null distribution provides the best control for the type I errors in most settings. Meanwhile, our proposed test method is more powerful than the P&N test in most settings, no matter whether the exact null distribution or the approximate null distribution is applied. In conclusion, our proposed test with the exact null distribution provides the best test for hypothesis (8).

5. CASE STUDIES

In this section, we consider two real data examples to illustrate the applications of our proposed tests. One is to determine whether a nutrient is effective in increasing the height of five-year-old boys, which is from the website: www.real-statistics.com. The other is to test whether there is the difference in the batch viscosity resulting from the process change, which is from [32]. The data has also been used by [31] to interpret the application of two sample t-test with one variance unknown.

Example 1. The data is from a survey to determine whether some nutrient is effective in increasing the height of five-year-old boys. The average height of five-year-old boys in a certain country is known to be normally distributed with mean 95 cm. A firm is selling a nutrient, which it claims will significantly increase the height of children. In order to demonstrate its claim, it selects 60 random samples of four-year-old boys, half of whom are given the nutrient for one year and half of whom are not. Given that the height of the boys at the age of five is given as follows, determine whether the nutrient is effective in increasing the height.

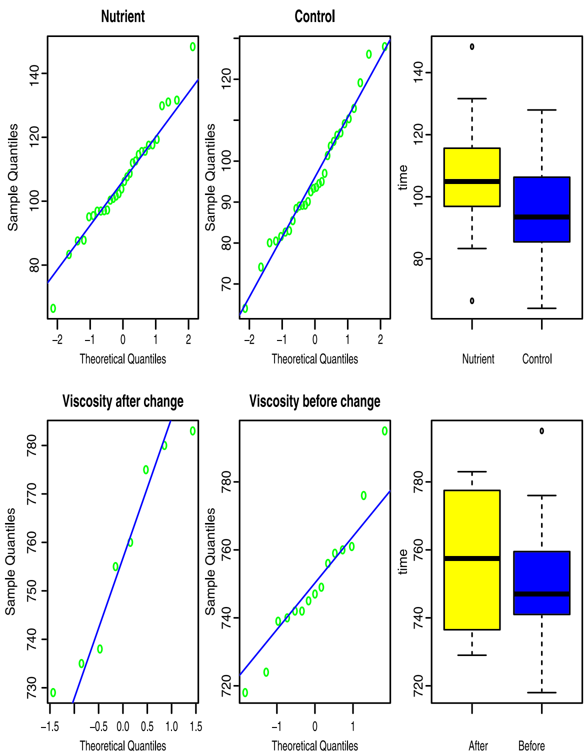

Figure 1. The normal Q-Q plots and boxplots for the data used in Example 1 and Example 2, respectively.

Nutrient: 106.18, 95.47, 117.51, 100.86, 108.66, 115.64, 129.85, 83.32, 97.22, 100.30, 117.64, 131.04, 87.56, 96.90, 101.58, 96.87, 66.46, 103.80, 112.57, 87.80, 111.99, 148.36, 115.52, 119.34, 131.62, 102.34, 95.10, 114.60, 97.01, 107.62;

Control: 82.67, 89.03, 109.13, 90.11, 94.51, 81.59, 89.20, 93.32, 94.99, 119.15, 89.26, 101.34, 83.01, 110.36, 104.82, 93.61, 92.52, 106.92, 88.42, 112.87, 80.50, 97.02, 64.05, 106.31, 126.11, 80.06, 85.46, 127.96, 74.13, 103.69.

We treat 'Nutrient' and 'Control' as the $\boldsymbol{X}$ and $\boldsymbol{Y}$ samples, respectively. From the Q-Q plots in the top panel of Figure 1, the normality assumptions for both samples are verified. As the data were randomly collected, we treat the two samples are independent of each other. From the data, we know that the true mean of the $\boldsymbol{Y}$ sample be known at $\mu_2 = \mu_0 = 95$. Then by the test statistic Λ_1, the observed value of the test statistic is $\lambda_1 = 0.000446$. This yields the p-value of the exact test as $F(\lambda_1) = 0.00289$, and the p-value of the approximate test as $\text{Beta}(\lambda_1, 0.9666, 1.0000) = 0.00057$, respectively. For both tests, we reject the null hypothesis $H_0^{(\text{V})}$ at the significance level $\alpha = 0.05$, that is, there is significantly effective in the nutrient creasing the height of five-year-old boys. Finally, we plot the boxplots of the two samples in Figure 1 and that also supports our decision.

Example 2. A polymer is manufactured in a batch chemical process. Viscosity measurements are normally made on each batch, and long experience with the process has indicated that the variability in the process is fairly stable with $\sigma = 20$. Fifteen batch viscosity measurements are given as follows (called 'Viscosity before change'):

724, 718, 776, 760, 745, 759, 795, 756, 742, 740, 761, 749, 739, 747, 742.

A process change is made which involves switching the type of catalyst used in the process. Following the process change, eight batch viscosity measurements are taken (called 'Viscosity after change'):

735, 775, 729, 755, 783, 760, 738, 780.

Based on the data, test whether there is the difference in the batch viscosity resulting from the process change.

We treat 'Viscosity after change' and 'Viscosity before change' as the $\boldsymbol{X}$ and $\boldsymbol{Y}$ samples, respectively. From the Q-Q plots in the bottom panel of Figure 1, we do not see any severe departure from the normality assumption for both samples. Following the design, we can treat the two samples as independence of each other. From the data example, we know that the true variance of the sample $\boldsymbol{Y}$ is known at $\sigma_2^2 = \sigma^2 = 20^2$. Then by the test statistic Λ_2, the observed value of the test statistic is $\lambda_2 = 0.7477$. This yields the p-value of the exact test as $F(\lambda_2) = 0.7668$, and the p-value of the approximate test as $\text{Beta}(\lambda_2, 0.8830, 0.9996) = 0.7734$, respectively. For both tests, we accept the null hypothesis $H_0^{(\text{VI})}$ at the significance level $\alpha = 0.05$. That is, we conclude that there is no significant difference in the batch viscosity resulting from the process change.

6. CONCLUSION

In this paper, we proposed a unified framework of hypothesis testing for two normal distributions from a very unique perspective. The unified framework has integrated the existing literature by including most commonly used tests as special cases. Following the unified framework, we comprehensively reviewed the one- and two-sample tests from the likelihood ratio test perspective, including the one-sample mean test, the one-sample variance test, the two-sample mean test, the two-sample variance test, and the equality test of one or two normal distributions. And more importantly, the unified framework has also put forward two new hypothesis tests that are rarely studied in the literature.

To solve the new testing problems in the unified framework, we proposed two likelihood ratio test statistics and derived their exact and approximate null distributions. To evaluate the finite sample performance of the proposed tests, we also conducted two simulation studies to assess the type I errors and the statistical power, and compared them with the benchmark P&N test. Simulation results indicated that our proposed tests perform better than the P&N test in a wide range of settings, especially for the proposed likelihood ratio tests with the exact null distribution. Finally, we also applied our new tests to two real data examples to demonstrate their usefulness in practice.

Note that hypothesis (1) is only for testing equality of two univariate normal populations. Statistically, our unified framework can also be extended to the testing problem for multivariate normal data. Assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two independent random vectors from the multivariate normal distributions $N_p(\boldsymbol{\mu}_1, \Sigma_1)$ and $N_p(\boldsymbol{\mu}_2, \Sigma_2)$ respectively, where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are two p-dimensional mean vectors, and Σ_1 and Σ_2 are two $p \times p$ covariance matrices. To test equality of two multivariate normal distributions, it is then equivalent to testing the hypothesis:

$$\begin{aligned} &H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \ \text{ and } \ \Sigma_1 = \Sigma_2 \\ \text{versus} \quad &H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \ \text{ or } \ \Sigma_1 \neq \Sigma_2. \end{aligned} \tag{10}$$

The likelihood ratio tests for this hypothesis have been proposed by [60] and [44]. If the two covariance matrices are assumed or known to be equal, then this leads to test equality of two mean vectors [6, 1]. If the two mean vectors are assumed or known to be equal, it leads to test equality of two covariance matrices [7, 56, 1]. If we further assume that one mean vector is known or one covariance matrix is known, then it yields two new testing problems for multivariate normal data as in Section 3, and which may deserve further research.

APPENDIX A

The appendix is organized as follows. In Appendix A1, we derive the LRT statistic and its exact null distribution for hypothesis (7). In Appendix A2, we derive the LRT statistic and its exact null distribution for hypothesis (8).

A.1 Likelihood ratio test with one mean known

In this section, we consider an LRT for hypothesis (7). The main result includes the following theorem.

Theorem A.1. *Assume that the mean $\mu_2 = \mu_0$ is known. The likelihood ratio test statistic for testing $H_0^{(\mathrm{V})}$ versus $H_1^{(\mathrm{V})}$ is given as*

$$\Lambda_1 = \frac{\Big\{\sum_{i=1}^n (X_i - \bar{X})^2/n\Big\}^{\frac{n}{2}} \Big\{\sum_{j=1}^m (Y_j - \mu_0)^2/m\Big\}^{\frac{m}{2}}}{\Big\{\Big(\sum_{i=1}^n (X_i - \mu_0)^2 + \sum_{j=1}^m (Y_j - \mu_0)^2\Big)/(n+m)\Big\}^{\frac{n+m}{2}}}.$$

Furthermore, for any $\lambda > 0$, the exact null distribution of Λ_1 is denoted as

$$\begin{aligned} F(\lambda) &= P(\Lambda_1 \le \lambda) = 1 - P(\Lambda_1 > \lambda) \\ &= 1 - \int_{r_1}^{r_2} dw_1 \int_{\frac{\lambda^{n/m} n^{n/m} m}{(n+m)^{(n+m)/m} w_1^{n/m}}}^{1-w_1} f(w_1, w_2) dw_2, \end{aligned}$$

where r_1 and r_2 ($r_1 < r_2$) are the two roots of the function $g(w_1) = 1 - w_1 - \frac{\lambda^{2/m} n^{n/m} m}{(n+m)^{(n+m)/m} w_1^{n/m}}$ and $f(w_1, w_2)$ is the probability density function of the two-dimensional Dirichlet distribution.

Proof. Let x_i and y_j be the observed values of the X_i and Y_j samples for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, respectively. Let $\bar{x}$ and $\bar{y}$ be observed values of the sample means $\bar{X}$ and $\bar{Y}$, respectively. For the two independent samples, the likelihood function is denoted as
$L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 | x, y)$

$$\begin{aligned} &= \prod_{i=1}^n f_X(x_i) \prod_{j=1}^m f_Y(y_j) \\ &= \Big(\frac{1}{2\pi\sigma_1^2}\Big)^{\frac{n}{2}} \Big(\frac{1}{2\pi\sigma_2^2}\Big)^{\frac{m}{2}} \exp\Big\{ - \frac{\sum_{i=1}^n (x_i - \mu_1)^2}{2\sigma_1^2} \\ &\quad - \frac{\sum_{j=1}^m (y_j - \mu_2)^2}{2\sigma_2^2} \Big\}, \end{aligned}$$

where $f_X(x)$ and $f_Y(y)$ are the density functions of the X and Y samples, respectively. Note that

$$\sum_{i=1}^n (x_i - \mu_1)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_1)^2,$$

and

$$\sum_{j=1}^m (y_j - \mu_2)^2 = \sum_{j=1}^m (y_j - \bar{y})^2 + m(\bar{y} - \mu_2)^2.$$

Under $H_0^{(\mathrm{V})}$, $\mu_1 = \mu_2 = \mu_0$ is known and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ is unknown. This yields the likelihood function as

$$\begin{aligned} L(\sigma^2 | x, y) &= \Big(\frac{1}{2\pi\sigma^2}\Big)^{\frac{n+m}{2}} \exp\Big\{ - \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2} \\ &\quad - \frac{\sum_{j=1}^m (y_j - \mu_0)^2}{2\sigma^2} \Big\}. \end{aligned}$$

Taking the derivative of the log-likelihood function with respect to σ^2 and setting it to zero, we have

$$\begin{aligned} \frac{d}{d\sigma^2} \log L(\sigma^2 | x, y) &= -\frac{m+n}{2} \frac{1}{\sigma^2} + \Big\{ \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2} \\ &\quad + \frac{\sum_{j=1}^m (y_j - \mu_0)^2}{2} \Big\} \frac{1}{\sigma^4} = 0 \end{aligned}$$

Then, we get

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \mu_0)^2 + \sum_{j=1}^m (y_j - \mu_0)^2}{n+m}.$$

Note that $\hat{\sigma}^2$ makes the likelihood function achieve the maximum. Thus, we have

$$\begin{aligned} &\sup_{H_0^{(\mathrm{V})}} L(\hat{\sigma}^2 | x, y) \\ &= \Big(\frac{1}{2\pi}\Big)^{\frac{n+m}{2}} \exp\Big(-\frac{n+m}{2}\Big) \end{aligned} \tag{11}$$

$$\times \left\{\frac{n+m}{\sum_{i=1}^{n}(x_i-\mu_0)^2+\sum_{j=1}^{m}(y_j-\mu_0)^2}\right\}^{\frac{n+m}{2}}.$$

Under $H_0^{(\mathrm{V})} \cup H_1^{(\mathrm{V})}$, $\mu_2 = \mu_0$ is known, but μ_1, σ_1^2 and σ_2^2 are unknown. This yields the likelihood function as

$$L(\mu_1,\sigma_1^2,\sigma_2^2|x,y) = \frac{\sigma_1^{-n}\sigma_2^{-m}}{(2\pi)^{(n+m)/2}} \exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{2\sigma_1^2} - \frac{\sum_{j=1}^{m}(y_j-\mu_0)^2}{2\sigma_2^2}\right\}.$$

Taking the partial derivative of the log-likelihood function with respect to each unknown parameter and setting it to zero, we have

$$\begin{aligned}
\frac{\partial \log L(\mu_1,\sigma_1^2,\sigma_2^2|x,y)}{\partial \mu_1} &= \frac{\sum_{i=1}^{n}(x_i-\mu_1)}{\sigma_1^2} = 0,\\
\frac{\partial \log L(\mu_1,\sigma_1^2,\sigma_2^2|x,y)}{\partial \sigma_1^2} &= -\frac{n}{2\sigma_1^2} + \frac{\sum_{i=1}^{n}(x_i-\mu_1)^2}{2\sigma_1^4} = 0,\\
\frac{\partial \log L(\mu_1,\sigma_1^2,\sigma_2^2|x,y)}{\partial \sigma_2^2} &= -\frac{m}{2\sigma_2^2} + \frac{\sum_{j=1}^{m}(y_j-\mu_0)^2}{2\sigma_2^4} = 0.
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\hat{\mu}_1 &= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},\\
\hat{\sigma}_1^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2, \hat{\sigma}_2^2 = \frac{1}{m}\sum_{j=1}^{m}(y_j-\mu_0)^2.
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\sup_{H_0^{(\mathrm{V})} \cup H_1^{(\mathrm{V})}} L(\hat{\mu}_1,\hat{\sigma}_1^2,\hat{\sigma}_2^2|x,y)\\
(12) \quad = \quad &\left\{\frac{n}{2\pi\sum_{i=1}^{n}(x_i-\bar{x})^2}\right\}^{\frac{n}{2}} \left\{\frac{m}{2\pi\sum_{j=1}^{m}(y_j-\mu_0)^2}\right\}^{\frac{m}{2}}\\
&\times \exp\left(-\frac{n+m}{2}\right).
\end{aligned}$$

Combined with (11) and (12), the LRT statistic is given as

$$\begin{aligned}
\Lambda_1 &= \frac{\sup_{H_0^{(\mathrm{V})}} L(\hat{\sigma}^2|X,Y)}{\sup_{H_0^{(\mathrm{V})} \cup H_1^{(\mathrm{V})}} L(\hat{\mu}_1,\hat{\sigma}_1^2,\hat{\sigma}_2^2|X,Y)}\\
&= \frac{\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right\}^{\frac{n}{2}}\left\{\frac{1}{m}\sum_{j=1}^{m}(Y_j-\mu_0)^2\right\}^{\frac{m}{2}}}{\left\{\frac{1}{n+m}\left(\sum_{i=1}^{n}(X_i-\mu_0)^2+\sum_{j=1}^{m}(Y_j-\mu_0)^2\right)\right\}^{\frac{n+m}{2}}}.
\end{aligned}$$

In what follows, we study the exact null distribution of Λ_1. Note that $\sum_{i=1}^{n}(X_i-\mu_0)^2 + \sum_{j=1}^{m}(Y_j-\mu_0)^2 = \sum_{i=1}^{n}(X_i-\bar{X})^2 + \sum_{j=1}^{m}(Y_j-\mu_0)^2 + n(\bar{X}-\mu_0)^2$. Let $U_1 = \sum_{i=1}^{n}(X_i-\bar{X})^2/\sigma^2$, $U_2 = \sum_{j=1}^{m}(Y_j-\mu_0)^2/\sigma^2$ and $U_3 = n(\bar{X}-\mu_0)^2/\sigma^2$. Then, $U_1 \sim \chi^2_{n-1}$, $U_2 \sim \chi^2_m$, $U_3 \sim \chi^2_1$, and U_1, U_2 and U_3 are mutually independent. Thus, the LRT statistic Λ_1 is rewritten as

$$\Lambda_1 = \frac{(U_1/n)^{\frac{n}{2}}(U_2/m)^{\frac{m}{2}}}{\{(U_1+U_2+U_3)/(n+m)\}^{\frac{n+m}{2}}}.$$

Let $W_1 = \frac{U_1}{U_1+U_2+U_3}$, $W_2 = \frac{U_2}{U_1+U_2+U_3}$ and $W_3 = U_1+U_2+U_3$. Then, the test statistic Λ_1 is further written as

$$\Lambda_1 = \frac{(n+m)^{\frac{n+m}{2}}}{n^{\frac{n}{2}}m^{\frac{m}{2}}} W_1^{\frac{n}{2}} W_2^{\frac{m}{2}}.$$

Following the mentioned above, we derive the joint density function of $U = (U_1, U_2, U_3)$ as

$$\begin{aligned}
f_U(u_1,u_2,u_3) \quad = \quad &\frac{1}{2^{\frac{n+m}{2}}\Gamma(\frac{1}{2})\Gamma(\frac{n-1}{2})\Gamma(\frac{m}{2})} u_1^{\frac{n-3}{2}} u_2^{\frac{m}{2}-1} u_3^{-\frac{1}{2}}\\
&\times \exp(-\frac{u_1+u_2+u_3}{2}).
\end{aligned}$$

Note that $U_1 = W_1W_3$, $U_2 = W_2W_3$ and $U_3 = (1 - W_1 - W_2)W_3$. Then, the joint density function of $W = (W_1, W_2, W_3)$ is

$$\begin{aligned}
&f_W(w_1,w_2,w_3)\\
&= f_U(u_1,u_2,u_3) \mid J \mid\\
&= f_U(w_1w_3, w_2w_3, (1-w_1-w_2)w_3) \mid J \mid\\
&= \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{m}{2})\Gamma(\frac{1}{2})} w_1^{\frac{n-3}{2}} w_2^{\frac{m}{2}-1}(1-w_1-w_2)^{-\frac{1}{2}}\\
&\quad \times \frac{1}{2^{\frac{n+m}{2}}\Gamma(\frac{n+m}{2})} w_3^{\frac{n+m}{2}-1}\exp(-\frac{w_3}{2}),
\end{aligned}$$

where $w_1 > 0, w_2 > 0, w_3 > 0$ and $w_1 + w_2 < 1$. Therefore, (W_1, W_2) and W_3 are independent, and (W_1, W_2) is distributed as the two-dimensional Dirichlet distribution with the density function

$$f(w_1,w_2) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n-1}{2})\Gamma(\frac{m}{2})\Gamma(\frac{1}{2})} w_1^{\frac{n-3}{2}} w_2^{\frac{m}{2}-1}(1-w_1-w_2)^{-\frac{1}{2}}.$$

Then, for any $\lambda > 0$, the exact null distribution of Λ_1 is denoted as

$$\begin{aligned}
F(\lambda) &= P(\Lambda_1 \le \lambda) = 1 - P(\Lambda_1 > \lambda)\\
&= 1 - \int\int_{D_1} f(w_1,w_2)dw_1dw_2\\
&= 1 - \int_{r_1}^{r_2} dw_1 \int_{\frac{\lambda^{n/m}n^{n/m}m}{(n+m)^{(n+m)/m}w_1^{n/m}}}^{1-w_1} f(w_1,w_2)dw_2,
\end{aligned}$$

where $D_1 = \{(w_1,w_2) : w_1 > 0, w_2 > 0, w_1 + w_2 < 1, \frac{(n+m)^{\frac{n+m}{2}}}{n^{\frac{n}{2}}m^{\frac{m}{2}}} w_1^{\frac{n}{2}} w_2^{\frac{m}{2}} > \lambda\}$, and r_1 and r_2 ($r_1 < r_2$) are the two roots of the function $g(w_1) = 1 - w_1 - \frac{\lambda^{2/m}n^{n/m}m}{(n+m)^{(n+m)/m}w_1^{n/m}}$. This completes the proof of theorem. $\square$

A.2 Likelihood ratio test with one variance known

In the section, we derive the LRT statistic and its exact null distribution for hypothesis (8). The result is presented in the following theorem.

Theorem A.2. *Assume that the variance $\sigma_2^2 = \sigma_0^2$ is known. The likelihood ratio test statistic for testing $H_0^{(\mathrm{VI})}$ versus $H_1^{(\mathrm{VI})}$ is given as*

$$\Lambda_2 = \Big\{\frac{\sum_{i=1}^n (X_i-\bar X)^2}{n\sigma_0^2}\Big\}^{\frac{n}{2}} \exp\Big\{\frac{n}{2} - \frac{\sum_{i=1}^n (X_i-\bar X)^2 + n(\bar X-\hat\mu)^2 + m(\bar Y - \hat\mu)^2}{2\sigma_0^2}\Big\}.$$

Furthermore, for any $\lambda > 0$, the exact null distribution of Λ_2 is denoted as

$$F(\lambda) = P(\Lambda_2 \le \lambda) = 1 - P(\Lambda_2 > \lambda) = 1 - \int_{r_1}^{r_2} dt_1 \int_0^{n(1-\log n)+n\log t_1 - t_1 - 2\log\lambda} f(t_1,t_2)dt_2,$$

where r_1 and r_2 $(r_1 < r_2)$ are the two roots of $n(1-\log n) + n\log t_1 - t_1 - 2\log\lambda = 0$ and $f(t_1,t_2) = t_1^{(n-3)/2} t_2^{-1/2}\exp\{-(t_1+t_2)/2\}/\{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})2^{n/2}\}$.

Proof. Under $H_0^{(\mathrm{VI})}$, $\sigma_1^2 = \sigma_2^2 = \sigma_0^2$ is known and $\mu_1 = \mu_2 = u$ is unknown. This yields the likelihood function as

$$L(\mu|x,y) = \Big(\frac{1}{2\pi\sigma_0^2}\Big)^{\frac{n+m}{2}} \exp\Big\{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma_0^2} - \frac{\sum_{j=1}^m (y_j-\mu)^2}{2\sigma_0^2}\Big\}.$$

Let $\frac{d}{d\mu}\log L(\mu|x,y) = 0$. Then, we get

$$\hat\mu = \frac{1}{n+m}\Big(\sum_{i=1}^n x_i + \sum_{j=1}^m y_j\Big).$$

Note that $\hat\mu$ maximizes the likelihood function. Therefore, we have

$$\sup_{H_0^{(\mathrm{VI})}} L(\hat\mu|x,y) = \exp\Big\{-\frac{\sum_{i=1}^n (x_i-\hat\mu)^2}{2\sigma_0^2} - \frac{\sum_{j=1}^m (y_j-\hat\mu)^2}{2\sigma_0^2}\Big\} \times \Big(\frac{1}{2\pi\sigma_0^2}\Big)^{\frac{n+m}{2}}. \quad (13)$$

Corresponding, under $H_0^{(\mathrm{VI})} \cup H_1^{(\mathrm{VI})}$, the likelihood function is

$$L(\mu_1,\mu_2,\sigma_1^2|x,y) = \frac{\sigma_1^{-n}\sigma_0^{-m}}{(2\pi)^{(n+m)/2}} \exp\Big\{-\frac{\sum_{i=1}^n (x_i-\mu_1)^2}{2\sigma_1^2} - \frac{\sum_{j=1}^m (y_j-\mu_2)^2}{2\sigma_0^2}\Big\}.$$

Taking the partial derivative of the log-likelihood function with respect to each unknown parameter and setting it to be zero, we have

$$\hat\mu_1 = \frac{1}{n}\sum_{i=1}^n x_i = \bar x, \quad \hat\mu_2 = \frac{1}{m}\sum_{j=1}^m y_j = \bar y,$$
$$\hat\sigma_1^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar x)^2.$$

Note that $\hat\mu_1, \hat\mu_2$ and $\hat\sigma_1^2$ make the likelihood function reach the maximum. Hence, we have

$$\sup_{H_0^{(\mathrm{VI})}\cup H_1^{(\mathrm{VI})}} L(\hat\mu_1,\hat\mu_2,\hat\sigma_1^2|x,y) = \Big(\frac{1}{2\pi}\Big)^{\frac{n+m}{2}}\Big(\frac{1}{\sigma_0^2}\Big)^{\frac{m}{2}}\Big\{\frac{\sum_{i=1}^n (x_i-\bar x)^2}{n}\Big\}^{-\frac{n}{2}} \exp\Big\{-\frac{n}{2} - \frac{\sum_{j=1}^m (y_j - \bar y)^2}{2\sigma_0^2}\Big\}. \quad (14)$$

Combined with (13) and (14), the LRT statistic is denoted as

$$\Lambda_2 = \frac{\sup_{H_0^{(\mathrm{VI})}} L(\hat\mu|X,Y)}{\sup_{H_0^{(\mathrm{VI})}\cup H_1^{(\mathrm{VI})}} L(\hat\mu_1,\hat\mu_2,\hat\sigma_1^2|X,Y)} = \exp\Big\{\frac{n}{2} - \frac{\sum_{i=1}^n (X_i-\bar X)^2 + n(\bar X-\hat\mu)^2 + m(\bar Y-\hat\mu)^2}{2\sigma_0^2}\Big\} \times \Big\{\frac{\sum_{i=1}^n (X_i-\bar X)^2}{n\sigma_0^2}\Big\}^{\frac{n}{2}}.$$

Let $T_1 = \sum_{i=1}^n (X_i-\bar X)^2/\sigma_0^2$, $T_2 = (n(\bar X-\hat\mu)^2 + m(\bar Y - \hat\mu)^2)/\sigma_0^2$. Then, it is easy to confirm that $T_1 \sim \chi^2_{n-1}, T_2 \sim \chi^2_1$ and T_1, T_2 are mutually independent. Therefore, the joint density function of (T_1, T_2) is denoted as

$$f(t_1,t_2) = \frac{1}{\Gamma(\frac{n-1}{2})\Gamma(\frac{1}{2})2^{\frac{n}{2}}} t_1^{\frac{n-3}{2}} t_2^{-\frac{1}{2}} \exp\Big(-\frac{t_1+t_2}{2}\Big).$$

Meanwhile, the test statistic Λ_2 is represented as

$$\Lambda_2 = \Big(\frac{e}{n}\Big)^{\frac{n}{2}} T_1^{\frac{n}{2}} \exp\Big(-\frac{T_1+T_2}{2}\Big).$$

Then, for any $\lambda > 0$, the exact null distribution of Λ_2 is

$$F(\lambda) = P(\Lambda_2 \le \lambda) = 1 - P(\Lambda_2 > \lambda) = 1 - \iint_{D_2} f(t_1,t_2)dt_1dt_2 = 1 - \int_{r_1}^{r_2} dt_1 \int_0^{n(1-\log n)+n\log t_1 - t_1 - 2\log\lambda} f(t_1,t_2)dt_2,$$

where $D_2 = \{(t_1,t_2) : t_1 > 0, t_2 > 0, (\frac{e}{n})^{n/2} t_1^{n/2} \times \exp(-\frac{t_1+t_2}{2}) > \lambda\}$, and r_1 and r_2 ($r_1 < r_2$) are the two roots of $n(1-\log n) + n\log t_1 - t_1 - 2\log\lambda = 0$. This completes the proof of theorem. □

ACKNOWLEDGMENTS

The authors thank the editor, the associate editor, and three reviewers for their constructive comments that have led to a substantial improvement of the paper. Yuejin Zhou's research was supported in part by Natural Science Foundation of Anhui (No. KJ2017A087), Doctoral Foundation of Anhui University of Science and Technology (No. ZY514), and National Natural Science Foundation of China (No. 61703005). Tiejun Tong's research was supported in part by Hong Kong Baptist University grants (FRG1/17-18/045, FRG2/17-18/020, Century Club Sponsorship Scheme, and Initiation Grant for Faculty Niche Research Areas), General Research Fund (No. HKBU12303918), and National Natural Science Foundation of China (No. 11671338).

Received 14 May 2019

REFERENCES

[1] Anderson, T. W. (2003). *An introduction to multivariate statistical analysis, 3rd Edition*. John Wiley and Sons, Hoboken, New Jersey. MR1990662

[2] Arnold, B. C. and Shavelle, R. M. (1998). Joint confidence sets for the mean and variance of a normal distribution. *The American Statistican* **52** 133-140. MR1628439

[3] Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society, Series A* **160** 268-282.

[4] Bayoud, H. A. (2016). Testing the similarity of two normal populations with application to the bioequivalence problem. *Journal of Applied Statistics* **43** 1322-1334. MR3473773

[5] Behrens, W. U. (1929). A contribution to error estimation with few observations. *Landwirtschaftliche Jahrbücher* **68** 807-837.

[6] Bennett, B. M. (1951). Note on a solution of the generalized Behrens-Fisher problem. *Annals of the Institute of Statistical Mathematics* **2** 87-90. MR0041399

[7] Bock, R. D. and Petersen, A. C. (1975). A multivariate correction for attenuation. *Biometrika* **62** 673-678. MR0397996

[8] Boos, D. D. and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances. *Technometrics* **31** 69-82. MR0997671

[9] Brown, G. W. (1939). On the power of the L_1 test for equality of several variances. *The Annals of Mathematical Statistics* **10** 119-128.

[10] Chao, M. and Glaser, R. E. (1978). The exact distribution of Bartlett's test statistic for homogeneity of variances with unequal sample sizes. *Journal of the American Statistical Association* **73** 422-426. MR0497423

[11] Choudhari, P., Kundu, D. and Misra, N. (2001). Likelihood ratio test for simultanous testing of the mean and the variance of a normal distribution. *Journal of Statistical Computation and Simulation* **71** 313-333. MR1888828

[12] Cochran, W. G. (1964). Approximate significance levels of the Behrens-Fisher test. *Biometrics* **20** 191-195. MR0162323

[13] Conover, W. J., Johnson, M. E. and Johnson, M. M. (1981). A comparative study of tests of homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **23** 351-361.

[14] DeGroot, M. H. and Schervish, M. J. (2002). *Probability and Statistics, 3rd Edition*. Addison Wesley, Boston. MR0373075

[15] Duran, B. S., Tsai, W. S. and Lewis, T. O. (1976). A class of location-scale tests. *Biometrika* **63** 173-176. MR0408102

[16] 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. MR2595125

[17] Fertig, J. W. and Proehl, E. A. (1937). A test of a sample variance based on both tail ends of the distribution. *The Annals of Mathematical Statistics* **8** 193-205.

[18] Fisher, R. A. (1935). The fiducial argument in statistical inference. *The Annals of Eugenics* **6** 391-398.

[19] Fisher, R. A. and Healy, M. J. (1956). New tables of Behrens's test of significance. *Journal of the Royal Statistical Society, Series B* **18** 212-216. MR0082242

[20] Frey, J., Marrero, O. and Norton, D. (2009). Minimum-area confidence sets for a normal distribution. *Journal of Statistical Planning and Inference* **139** 1023-1032. MR2479846

[21] Glaser, R. E. (1980). A characterization of Bartlett's statistic involving incomplete beta functions. *Biometrika* **67** 53-58. MR0570504

[22] Hayter, A. (2014). Identifying common normal distributions. *TEST* **23** 135-152. MR3179620

[23] Keating, J. P. and Tripathi, R. C. (1984). Comparison op some two sample tests for equality of variances. *Communications in Statistics-Theory and Methods* **13** 2691-2706. MR0759245

[24] Kim, S. H. and Cohen, A. S. (1998). On the Behrens-Fisher problem: a review. *Journal of Eduational and Behavioral Statistics* **23** 356-377.

[25] Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* **4** 83-91.

[26] Kulinskaya, E. (2008). On two-sided p-values for non-symmetric distributions. *arXiv:0810.2124*.

[27] Lee, A. and Gurland, J. (1975). Size and power of tests for equality of means of two normal populations with unequal variances. *Journal of the American Statistical Association* **70** 933-941. MR0388621

[28] Levene, H. (1960). *Robust Tests for Equality of Variances, in I.Olkin, ed., Contributions to Probability and Statistics*. Palo Alto, Calif., Stanford University Press, 278–292. MR0120709

[29] Linssen, H. N. (1991). A table for solving the Behrens-Fisher problem. *Statistics and Probability Letters* **11** 359-363.

[30] Ludbrook, J. and Dudley, H. (1998). Why permutation tests are superior to t and F tests in biomedical research. *The American Statistician* **52** 127-132.

[31] Maity, A. and Sherman, M. (2006). The two-sample t test with one variance unknown. *The American Statistician* **60** 163-166. MR2224216

[32] Montgomery, D. C. and Runger, G. C. (2003). *Applied Statistics and Probability for Engineers, 3rd Edition*. John Wiley and Sons, Hoboken, New Jersey.

[33] Neyman, J., Iwaszkiewicz, K. and Kolodziejczyk, S. (1935). Statistical problems in agricultural experimentation. *Supplement to the Journal of the Royal Statistical Society,* **2** 107-180.

[34] Neyman, J. and Pearson, E. S. (1936). Contributions to the theory of testing statistical hypothesis. I. Unbiased critical regions of Type A and Type A_1. *Statistical Research Memoirs* **1** 1-37.

[35] Neyman, J. and Tokarsk, B. (1936). Errors of the second kind in testing Student's hypothesis. *Journal of the American Statistical Association* **31** 318-326.

[36] Niwitpong, S. and Niwitpong, S. A. (2017). Confidence intervals for the difference between normal means with known coefficients of variation. *Annals of Operations Research* **256** 237-251. MR3697209

[37] Pachares, J. (1960). Tables for unbiased tests on the variance of a normal population. *The Annals of Mathematical Statistics* **31** 84-87. MR0119298

[38] Pagurova, V. I. (1968). On a comparison of means of two normal samples. *Theory of Probability and Its Application* **13** 527-534. MR0237031
[39] Park, A. I. (2013). On the study for the simultaneous test. *Communications for Statistical Applications and Methods* **20** 241-246.
[40] Park, A. I. (2015). Simultaneous test for the mean and variance with an application to the statistical process control. *Journal of Statistical Theory and Practice* **9** 868-881. MR3378119
[41] Patil, V. H. (1965). Approximation to the Behrens-Fisher distributions. *Biometrika* **52** 267-271. MR0207080
[42] Pearson, E. S. and Neyman, J. (1930). *On the Problem of Two Samples, in Joint Statistical Papers (1967)*. Eds. J. Neyman and E. S. Pearson, Cambridge: Cambridge University Press, 99–115. MR0222982
[43] Peng, L. and Tong, T. (2011). A note on a two-sample t test with one variance unknown. *Statistical Methodology* **8** 528–534. MR2834038
[44] Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *The Annals of Statistics* **8** 247–263. MR0560727
[45] Perng, S. K. and Littell, R. C. (1976). A test of equality of two normal population means and variances. *Journal of the American Statistician Association* **71** 968–971. MR0420945
[46] Pitman, E. J. G. (1939). Tests of hypotheses concerning location and scale parameters. *Biometrika* **31** 200–215. MR0000382
[47] Ramachandran, K. V. (1958). A test of variances. *Journal of the American Statistical Association* **53** 741–747. MR0101587
[48] Rayner, J. C. (1997). The asymptotically optimal tests. *Journal of the Royal Statistical Society, Series D* **46** 337–346.
[49] Robinson, G. K. (1976). Properties of Student's t and of the Behrens-Fisher solution to the two means problem. *The Annals of Statistics* **4** 963–971. MR0415868
[50] Robinson, G. K. (1982). The Behrens-Fisher problem. *Encyclopedia of Statistical Sciences* **1** 205–208.
[51] Ruben, H. (2002). A simple conservative and robust solution of the Behrens-Fisher problem. *Sankhyā* **64** 139–155. MR1968379
[52] Rublik, F. (2009). Critical values for testing location-scale hypothesis. *Measurement Science Review* **9** 9–15.
[53] Satterthwaite, F. E. (1941). Synthesis of Variance. *Psychometrika* **6** 309–316. MR0005564
[54] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2** 110–114.
[55] Scheffe, H. (1942). On the ratio of the variances of two normal populations. *The Annals of Mathematical Statistics* **13** 371–388. MR0007586
[56] Schott, J. R. (2001). Some tests for the equality of covariance matrices. *Journal of Statistical Planning and Inference* **94** 25–36. MR1820169
[57] Sheldon, M. R. (2010). *Introductory Statistics*. Academic Press, London.
[58] Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics* **19** 279–281. MR0025109
[59] Sprott, D. A. and Farewell, V. T. (1993). The difference between two normal means. *The American Statistician* **47** 126–128.
[60] Szatrowski, T. H. (1979). Asymptotic nonnull distributions for likelihood ratio statistics in the multivariate normal patterned mean and covariance matrix testing problem. *The Annals of Statistics* **7** 823–837. MR0532246
[61] Tate, R. F. and Klett, G. W. (1959). Optimal confidence intervals for the variance of a normal distribution. *Journal of the American Statistician Association* **54** 674–682. MR0107926
[62] Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29** 350–362.
[63] Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika* **34** 28–35. MR0019277
[64] Welch, B. L. (1958). 'Student' and small sample theory. *Journal of the American Statistical Association* **53** 777–788. MR0099734
[65] Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley and Sons, New York. MR0144404
[66] Wilks, S. S. and Daly, J. F. (1939). An optimum property of confidence regions associated with the likelihood function. *The Annals of Mathematics Statistics* **10** 225–235. MR0000386
[67] Won, S., Morris, N., Lu, Q. and Elston, R. C. (2009). Choosing an optimal method to combine p-values. *Statistics in Medicine* **28** 1537–1553. MR2668754
[68] Zimmerman, D. W. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology* **57** 173–181. MR2087822

Yuejin Zhou
School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, China
E-mail address: yjchow@163.com

Sze-Yui Ho
Department of Statistics, Chinese University of Hong Kong, Hong Kong, China
E-mail address: 1155087611@link.cuhk.edu.hk

Jiahua Liu
Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada
E-mail address: j664liu@uwaterloo.ca

Tiejun Tong
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
E-mail address: tongt@hkbu.edu.hk