

Simultaneous Tests with Combining Functions under Normality

Hyo-II Park^{1,a}

^aDepartment of Statistics, Chongju University, Korea

Abstract

We propose simultaneous tests for mean and variance under the normality assumption. After formulating the null hypothesis and its alternative, we construct test statistics based on the individual p -values for the partial tests with combining functions and derive the null distributions for the combining functions. We then illustrate our procedure with industrial data and compare the efficiency among the combining functions with individual partial ones by obtaining empirical powers through a simulation study. A discussion then follows on the intersection-union test with a combining function and simultaneous confidence region as a simultaneous inference; in addition, we discuss weighted functions and applications to the statistical quality control. Finally we comment on nonparametric simultaneous tests.

Keywords: combining function, likelihood ratio statistic, multiple test, normal distribution, partial test, simultaneous test

1. Introduction

There is a long history on the inference about mean and variance under the normality assumption but the inference always has been concentrated on one of both parameters while neglecting the other. However, sometimes one may be interested in searching for both values since both parameters can provide equally important information. As an example, suppose that a production process should be maintained and controlled with the variance of the process as well as the mean for the quality control. This can be achieved by using a control chart such as the $\bar{X} - S$ chart. Then what is the overall probability limit of the $\bar{X} - S$ chart when the $3 - \sigma$ control limits are used for both charts? Or what is the overall average run length of the $\bar{X} - S$ chart until the chart can detect the shift in the mean or change in the variance? Those answers should be provided to maintain and enhance the quality of the process on the floor. Consequently there is a need to provide a statistical methodology to make inferences about the mean and variance simultaneously under the normality assumption.

Simultaneous tests for the location and scale parameters have been proposed for more than 40 years in nonparametric statistics (Lepage, 1971, 1973) that are still reported by many authors (Park, 2012; Park and Kim, 2012). Choudhari *et al.* (2001) proposed a likelihood ratio simultaneous test under the normality assumption. They derived the null distribution of the likelihood ratio statistic. However the form of the distribution is too awkward to apply to obtain the probability and so Choudhari *et al.* (2001) provided approximate critical values for selected sample sizes and significance levels. Park and Han (2013) also considered union-intersection tests (Roy, 1957) with the partially

¹ Department of Statistics, Chongju University, 298 Daesung-Ro, Chungwon-Gu, Chongju 28503, Korea.
E-mail: hipark@cju.ac.kr

likelihood ratio statistics. However, simultaneous inferences have already been applied with an ad hoc approach in application branches of statistics such as statistical process control (Hawkins and Deng, 2009). Here again, it would be imperatively necessary to establish more simultaneous test procedures for the mean and variance under the normality assumption for the underlying distribution.

In multiple testing, one chooses a suitable combining function (Pesarin and Salmaso, 2010) to combine the individual test statistics or individual testing results. Three widely used combining functions are Fisher (1932), Tippett (1931) and Liptak (1958); in addition, it is also possible to use the quadratic form. We note that the quadratic form is suitable to combine statistics and not p -values.

We consider simultaneous tests to jointly test hypothesis on the mean and variance under normality assumption. The rest of the paper will be organized in the following order. In the next section, we begin our discussion by formulating the null hypothesis and its alternative. For the construction of the simultaneous tests, we use optimal statistics derived from the likelihood ratio principle. We then obtain respective p -values by testing the individual partial null hypothesis and combine the two individual p -values with the combining functions to complete the simultaneous test by obtaining an overall p -value. Then we illustrate our test procedure with industrial data and compare our proposed tests by obtaining empirical powers through a simulation study. We discuss some features of the test procedure.

2. Simultaneous Tests with Combining Functions

Suppose that we have a sample X_1, \dots, X_n from a population with the normal distribution whose mean and variance are μ and σ^2 , respectively. Our main interest is to test

$$H_0 : \{\mu = \mu_0\} \cap \{\sigma^2 = \sigma_0^2\}, \quad (2.1)$$

where μ_0 and σ_0^2 are some pre-specified quantities. The corresponding alternative hypothesis is

$$H_1 : \{\mu \neq \mu_0\} \cup \{\sigma^2 \neq \sigma_0^2\}. \quad (2.2)$$

Choudhari *et al.* (2001) proposed a likelihood ratio test based on the following likelihood ratio statistic $\text{LR}(\mathbf{X}; \mu, \sigma^2)$.

$$\begin{aligned} \text{LR}(\mathbf{X}; \mu, \sigma^2) &= \frac{\sup\{f(\mathbf{X}; \mu, \sigma^2) : H_0 \cup H_1\}}{\sup\{f(\mathbf{X}; \mu, \sigma^2) : H_0\}} \\ &= \frac{(2\pi S^2)^{-\frac{n}{2}} \exp\left[-(2S^2)^{-1} nS^2\right]}{(2\pi\sigma_0^2)^{-\frac{n}{2}} \exp\left[-(2\sigma_0^2)^{-1} nS_0^2\right]} \\ &\propto \left(\frac{S^2}{\sigma_0^2}\right)^{-\frac{n}{2}} \exp\left[\frac{n S^2}{2 \sigma_0^2}\right] \exp\left[\frac{n (\bar{X} - \mu_0)^2}{2 \sigma_0^2}\right], \end{aligned} \quad (2.3)$$

where

$$S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2 \quad \text{and} \quad S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

From (2.3), Park and Han (2013) identified that

$$\exp\left[\frac{n(\bar{X} - \mu_0)^2}{2\sigma_0^2}\right] \quad \text{or} \quad M = \frac{n n(\bar{X} - \mu_0)^2}{2\sigma_0^2}$$

can be used for testing the partial null hypothesis $H_0^1 : \mu = \mu_0$ with the condition that $\sigma^2 = \sigma_0^2$ and

$$\left(\frac{S^2}{\sigma_0^2}\right)^{-\frac{n}{2}} \exp\left[\frac{n S^2}{2\sigma_0^2}\right] \quad \text{or} \quad V = n \frac{S^2}{\sigma_0^2}$$

can be used for testing the partial hypothesis $H_0^2 : \sigma^2 = \sigma_0^2$. Then the partial testing rule would be that for large values of M , one will reject $H_0^1 : \mu = \mu_0$ in favor of $H_1^1 : \mu \neq \mu_0$ and for smaller or larger values than 1 of V , one will reject $H_0^2 : \sigma^2 = \sigma_0^2$ in favor of $H_1^2 : \sigma^2 \neq \sigma_0^2$. We note that M and V are independent. For the testing rule for $H_0^2 : \sigma^2 = \sigma_0^2$ based on V , you may refer to Park and Han (2013). Since we take the multiple testing approach for the simultaneous test, for the construction of test statistics for testing (2.1) with these partial individual tests, let Λ_i be the corresponding p -value for testing the partial hypothesis H_0^i , $i = 1, 2$. Then we may consider the following combining functions:

(1) Fisher combining function

$$C_F = -2 \log(\Lambda_1) - 2 \log(\Lambda_2).$$

(2) Tippett combining function

$$C_T = \Psi^{-1}(\Lambda_1) + \Psi^{-1}(\Lambda_2).$$

(3) Liptak combining function

$$C_L = \min\{\Lambda_1, \Lambda_2\}.$$

The log in C_F stands for the natural logarithm and Ψ in C_T is an arbitrary distribution function and Ψ^{-1} , its inverse. Usually, one chooses Φ , the standard normal distribution for Ψ for some theoretic and application reasons. Then the overall testing rule would be to reject (2.1) in favor of (2.2) for some small values of any chosen combining function. The distributions for C_F , C_T and C_L are required to decide the critical value for any given significance level or more generally the p -value. We now summarize the null distributions of the three combining functions. The derivations for those results are contained in the intermediate mathematical statistics textbooks (Mood *et al.*, 1974).

Lemma 1. Under (2.1), the distribution of C_F is a chi-square with 4 degrees of freedom.

Lemma 2. Under (2.1) with Φ for Ψ , the distribution of C_T is the normal distribution with mean 0 and variance 2.

Lemma 3. Under (2.1), the distribution of C_L is as follows: for some real number c_l , we have that

$$\begin{aligned} \Pr\{C_L \leq c_l\} &= \Pr\{\min\{\Lambda_1, \Lambda_2\} \leq c_l\} \\ &= 1 - (1 - c_l)^2. \end{aligned}$$

Table 1: Overall p -values

Test	p -value
C_F	0.0031
C_T	0.0137
C_L	0.0012

Table 2: Empirical powers for the location change only

Test	n	μ					
		0.0	0.2	0.4	0.6	0.8	1.0
C_F	6	0.0491	0.0608	0.0952	0.1499	0.1990	0.2412
	10	0.0516	0.0702	0.1258	0.1903	0.2449	0.2753
	15	0.0500	0.0839	0.1565	0.2319	0.2744	0.2915
C_T	6	0.0501	0.0607	0.0957	0.1509	0.2045	0.2490
	10	0.0516	0.0708	0.1271	0.1965	0.2551	0.2895
	15	0.0500	0.0836	0.1593	0.2413	0.2884	0.3070
C_L	6	0.0495	0.0654	0.1281	0.2404	0.4043	0.5942
	10	0.0499	0.0806	0.1887	0.3857	0.6229	0.8216
	15	0.0523	0.0943	0.2671	0.5491	0.8060	0.9483
T^2	6	0.0502	0.0665	0.1247	0.2235	0.3524	0.5023
	10	0.0503	0.0872	0.2036	0.3942	0.6149	0.7998
	15	0.0510	0.1139	0.3046	0.5759	0.8165	0.9456

Then we can complete the test procedure to test (2.1) by obtaining the overall p -value by choosing a suitable combining function using the null distribution summarized in Lemmas.

We note that

$$T^2 = (n - 1) \frac{(\bar{X} - \mu_0)^2}{S^2}$$

is the well-known likelihood ratio statistic for testing $H_0^1 : \mu = \mu_0$ without any further assumption about σ^2 (Bickel and Doksum, 1977). Therefore it would be interesting if we compare the performance with the proposed simultaneous tests when only the mean value changes. This comparison will be performed in the next section through a simulation study.

3. An Example and Simulation Study

In this section, first of all, we provide an example to illustrate our simultaneous test procedure using a part of the industrial data (DeVor *et al.*, 1992). The data set consists of the measurements of the cylinder bores in an engine block. The measurements are made to 1/10,000 of an inch, refer to DeVor *et al.* (1992) for further information. Chen and Cheng (1998) used the whole data consisted with 35 samples for the construction of the max chart which is a combined control chart for the mean and variance under the normality. For our purpose, only we use the first sample in the following.

205 202 204 207 205

Chen and Cheng (1998) considered 200.25 and 2.85 for the values of the central line for each \bar{X} - and S -charts, respectively. Therefore we will take $\mu_0 = 200.25$ and $\sigma_0^2 = 2.85^2$ for the null hypothesis (2.1). Then we have the respective statistics and p -values for each partial tests $M = 11.65$ and $\Lambda_1 = 0.000642$ and $V = 2.03139$ and $\Lambda_2 = 0.540$. We summarized the simultaneous testing results in Table 1 below.

Table 3: Empirical powers for the scale change only

Test	<i>n</i>	δ					
		1.0	1.2	1.4	1.6	1.8	2.0
C_F	6	0.0491	0.1076	0.2090	0.3110	0.4128	0.4971
	10	0.0516	0.1328	0.2694	0.4009	0.4950	0.5626
	15	0.0500	0.1533	0.3159	0.4489	0.5368	0.5979
C_T	6	0.0501	0.1087	0.2115	0.3175	0.4196	0.5057
	10	0.0516	0.1353	0.2732	0.4074	0.5053	0.5725
	15	0.0500	0.1554	0.3218	0.4575	0.5494	0.6094
C_L	6	0.0495	0.1369	0.2813	0.4511	0.5956	0.7129
	10	0.0499	0.1621	0.3750	0.5925	0.7560	0.8660
	15	0.0523	0.1939	0.4765	0.7331	0.8786	0.9471
V	6	0.0502	0.1240	0.2611	0.4139	0.5539	0.6697
	10	0.0508	0.1598	0.3742	0.5980	0.7538	0.8590
	15	0.0530	0.1994	0.4988	0.7492	0.8846	0.9484

In this example, we see that C_T is less significant among the three combining functions. In passing, we note when the probability of control limit is 0.0053927 for a combined chart that the combined charts based on C_F and C_L can issue alarms for the out-of-control state in the process while the one based on C_T cannot. We also note that one can easily understand which one characteristic between the two causes the out-of-control state in the process.

In the following, we carry on a simulation study to compare the efficiency among the proposed tests by obtaining empirical powers for some small sample sizes. For this comparison study, we consider five different tests, C_F , C_T , C_L , T^2 and Λ_2 along with the three different sample sizes, $n=6$, 10 and 15. In this study, we also included individual likelihood ratio tests T^2 and Λ_2 to compare the performance with C 's when only one parameter value is changed (Table 2 and Table 3). We generate pseudo-random numbers from the standard normal distribution and consider the following three different cases:

- (1) Location change only (Table 2).
- (2) Scale change only (Table 3).
- (3) Both location and scale change (Table 4).

Then our null hypothesis in this simulation study would be

$$H_0 : \{\mu = 0\} \cap \{\sigma^2 = 1\}.$$

We have considered varying the value of μ from 0 up to 1 with the increment 0.2 while fixing σ with 1 for the case (1) and varying the value of σ from 1 up to 2 with the increment 0.2 while fixing μ with 0 for the case (2). For the case (3), we considered varying the values of (μ, σ) from (0.2, 1.2) up to (1.0, 2.0) with the increment 0.2 for each component. Therefore the rate of increments for μ and σ are 20% each. For example, $\mu = 0.4$ or $\sigma = 1.4$ implies that the mean or standard deviation increased 40% from 0 or 1 which is the null value of μ or σ . The nominal significance level considered in this study is 0.05 for each case. The simulation has been conducted with 10,000 iterations with the Monte-Carlo approach with SAS/IML PC-version.

From the tables we note that the performance by C_L is superior to those of C_F and C_T among the proposed simultaneous tests. In Table 2, we did not include the simulation results for Λ_2 since the application of V to the location shift only case would be at least meaningless. We also excluded the results for T^2 in Table 3 since the statistic should be applied to the case of location shift only. In Table

Table 4: Empirical powers for both location and scale changes

Test	n	(μ, δ)					
		(0.0, 1.0)	(0.2, 1.2)	(0.4, 1.4)	(0.6, 1.6)	(0.8, 1.8)	(1.0, 2.0)
C_F	6	0.0491	0.1221	0.2688	0.4273	0.5605	0.6660
	10	0.0516	0.1585	0.3790	0.5838	0.7245	0.8257
	15	0.0500	0.1987	0.4910	0.7168	0.8477	0.9162
C_T	6	0.0501	0.1233	0.2736	0.4342	0.5695	0.6742
	10	0.0516	0.1604	0.3857	0.5903	0.7313	0.8207
	15	0.0500	0.2018	0.4962	0.7235	0.8521	0.9135
C_L	6	0.0495	0.1545	0.3482	0.5551	0.7127	0.8210
	10	0.0499	0.1932	0.4714	0.7177	0.8631	0.9372
	15	0.0523	0.2399	0.5993	0.8479	0.9485	0.9853
T^2	6	0.0502	0.0608	0.0876	0.1155	0.1444	0.1695
	10	0.0503	0.0742	0.1274	0.1851	0.2410	0.2926
	15	0.0510	0.0918	0.1816	0.2753	0.3624	0.4352
V	6	0.0502	0.1240	0.2611	0.4139	0.5539	0.6697
	10	0.0508	0.1598	0.3742	0.5980	0.7538	0.8590
	15	0.0530	0.1994	0.4988	0.7492	0.8846	0.9484

2 and Table 3, one may note that C_L achieves almost the same performance (or better) than T^2 and Λ_2 which are optimal for the location shift and scale change only models, respectively. Therefore C_L can also be applied to the case of the location shift only or scale change only in order to enhance the power and preserve the stability of the test.

4. Concluding Remarks

The test C_L also can be called the intersection-union one (Roy, 1957) since the combining function C_L consists of the likelihood ratio statistic. For the intersection-union test which is the reverse procedure of the union-intersection test for the null hypothesis and its alternative, Berger (1996) showed that intersection-union tests are uniformly more powerful than the likelihood ratio tests in some cases. This may a reason that the test C_L achieves most powerful results in the simulation study.

As another simultaneous inference for the mean and variance, Mood *et al.* (1974) considered obtaining a simultaneous confidence region for (μ, σ^2) using jointly the statistics M and V on an ad hoc basis; however, it is difficult to obtain the confidence coefficient for the simultaneous confidence region if one uses T^2 and V since T^2 and V are not independent. We may also expect that apart from obtaining the confidence coefficient, the confidence region for (μ, σ^2) based on M and V would be more accurate than that based on T^2 and V since M and V are jointly likelihood ratio statistics.

When we construct the test statistics for testing (2.1), we have used the p -values rather than the statistics themselves since there is no way to combine $(\bar{X} - \mu_0)^2/\sigma_0^2$ and S^2/σ_0^2 into a statistic and the respective testing criteria are different. We also note that the two combining functions, C_F and C_T deal with the mean and variance equally according to their importance. If one considers that one is more important than the other, one can use the following weighted combining function such as

$$C_F^w = -2w_1 \log(\Lambda_1) - 2w_2 \log(\Lambda_2)$$

and

$$C_T^w = w_1 \Psi^{-1}(\Lambda_1) + w_2 \Psi^{-1}(\Lambda_2),$$

where w_i means weight for the i th component, $i = 1, 2$ with $w_1 + w_2 = 1$. Pesarin and Salmaso (2010) reviewed and introduced extensively the combining functions based on the permutation prin-

ciple (Good, 2000) for the nonparametric testing problems. There is also one additional advantage when we use the p -values for the construction of the test statistics. When one applies our tests to the one-sided alternatives, there is no need to modify or standardize the corresponding individual test statistics.

We note that the individual \bar{X} - and S -charts are popular to control the quality of product along with the mean and variance for the production process under the normality assumption. However one may be interested in considering a statistical procedure that controls simultaneously the location and scale parameters under the normal process assumption. Then it would be useful to consider using our proposed test based on (2.3) if one can take the hypothesis testing approach instead of using control charts.

In nonparametric statistics, the simultaneous test procedure has been initiated by Lepage (1971) for the location and scale parameters by combining Wilcoxon and Ansari-Bradley statistics. Also Lepage (1973) tabulated the exact critical values and significance levels for the combined statistic for some selected small sample sizes. Duran *et al.* (1976) obtained asymptotic distribution and investigated asymptotic relative efficiency. Statisticians have also considered the applications and modifications of the simultaneous test (Murakami, 2007; Neuhäuser *et al.*, 2011); in addition, Park (2012) also proposed the simultaneous tests with the combining functions.

Acknowledgements

The author expresses his sincere appreciation to the anonymous referees and associate editor who provided invaluable advice and suggestions to enhance the quality of this paper.

References

- Berger, R. L. (1996). Likelihood ratio tests and intersection-union tests, *Institute of Statistics Mimeo Series Number 2288*, Department of Statistics, North Carolina State University, Raleigh, North Carolina.
- Bickel, P. J. and Doksum, K. A. (1977). *Mathematical Statistics-Basic Ideas and Selected Topics*, Holden-Day, San Francisco.
- Chen, G. and Cheng, S. W. (1998). Max chart: combining X-bar chart and S chart, *Statistica Sinica*, **8**, 263–271.
- 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.
- DeVor, R. E., Chang, T. and Sutherland, J. W. (1992). *Statistical Quality Design and Control*, Macmillan, New York.
- Duran, B. S., Tsai, W. S. and Lewis, T. O. (1976). A class of location-scale tests, *Biometrika*, **63**, 173–176.
- Fisher, R. A. (1932). *Statistical Methods for Research Workers*, Fourth Edition, Oliver and Boyd, Edinburgh.
- Good, L. J. (2000). *Permutation Tests-A Practical Guide to Resampling Methods for Testing Hypotheses*, Wiley, New York.
- Hawkins, D. M. and Deng, Q. (2009). Combined charts for mean and variance information, *Journal of Quality Technology*, **41**, 415–425.
- Lepage, Y. (1971). A combination of Wilcoxon's and Ansari-Bradley's statistics, *Biometrika*, **58**, 213–217.

- Lepage, Y. (1973). A table for a combined Wilcoxon Ansari-Bradley statistic, *Biometrika*, **60**, 113–116.
- Liptak, I. (1958). On the combination of independent tests, *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, **3**, 127–141.
- Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). *Introduction to the Theory of Statistics, Third Edition (International Student Edition)*, McGraw-Hill, Tokyo.
- Murakami, H. (2007). Lepage type statistic based on the modified Baumgartner statistic, *Computational Statistics & Data Analysis*, **51**, 5061–5067.
- Neuhäuser, M., Leuchs, A.-K. and Ball, D. (2011). A new location-scale test based on a combination of the ideas of Levene and Lepage, *Biometrical Journal*, **53**, 525–534.
- Park, H. I. (2012). On the study for the simultaneous test, *Communications for Statistical Applications and Methods*, **20**, 241–246.
- Park, H. I. and Kim, J. S. (2012). A study on the bi-aspect procedure with location and scale parameters, *Journal of The Korean Official Statistics*, **17**, 19–26.
- Park, H. I. and Han, K. J. (2013). A simultaneous test for mean and variance based on the likelihood ratio principle, *Journal of the Korean Data Analysis Society*, **15**, 1733–1742.
- Pesarin, F. and Salmaso (2010). *Permutation Tests for Complex Data*, Wiley, New York.
- Roy, S. N. (1957). *Some Aspects of Multivariate Analysis*, Wiley, New York.
- Tippett, L. H. C. (1931). *The Methods of Statistics*, Williams and Norgate, London.

Received August 17, 2015; Revised September 30, 2015; Accepted October 6, 2015