Count Five Statistics Using Trimmed Mean

Chong Sun Hong[1] and Jae Woon Jun[2]

Abstract

There are many statistical methods of testing the equality of two population variances. Among them, the well-known F test is very sensitive to the normality assumption. Several other tests that do not assume normality have been proposed, but these tests usually need tables of critical values or software for hypotheses testing. McGrath and Yeh (2005) suggested a quick and compact Count Five test requiring only the calculation of the number of extreme points. Since the Count Five test uses only extreme values, this discards some information from the samples, often resulting in a degradation in power. In this paper, an alternative Count Five test using the trimmed mean is proposed and its properties are discussed for some distributions and normal mixtures.

Keywords : Dispersion; Normal mixture; Power; Size; Trimmed mean.

1. Introduction

In this paper, we discuss the procedures that can be used for testing the equality of two population variances. The well-known F test statistic is commonly used for testing this hypothesis. However, the F test has the disadvantage that it is extremely sensitive to the normality assumption. Moore and McCabe (2003) discussed the assumption of the F test and warned against its use. McClave, Benson, and Sincich (2005) mentioned the sensitivity of the F test and suggested using an alternative test proposed by Levene (1960), and modified by Brown and Forsythe (1974). Ott and Longnecker (2001) discussed the modified Levene's test and extended it for the purpose of testing the equality of three or more variances.

Most alternatives to the F test are rank-based, including those proposed by Mood (1954), Siegel and Tukey (1960), Ansari and Bradley (1960), Klotz (1962), Fligner and Killeen (1976), and Conover (1980). More recently, Hall and Padmanabhan (1997) used a bootstrapping approach to the problem. Pan (1999) provided another modification of Levene's test. Shoemaker (2003) suggested an alternative F test using adjusted degrees of freedom.

---

1) Professor, Department of Statistics, SungKyunKwan University, Seoul, Korea.
   Correspondence : cshong@skku.ac.kr
2) Research Fellow, School of Economics, SungKyunKwan University, Seoul, Korea.

McGrath and Yeh (2005) developed the Count Five statistics that use the transformations of Levene's (mean centered) and modified-Levene's (median centered) statistics. These statistics require the number of extreme points to be counted, that is, those observations that are larger than all of the observations in the other sample. The simulation results of McGrath and Yeh (2005) showed that the Count Five (mean centered) test is more reliable for attaining the nominal size reasonably well and that the power of this test is higher than that of all other tests for the uniform distribution. However, they also found that the power of the Count Five (mean centered) test is lower than that of all other tests for skewed and heavy tailed distributions. Since a trimmed mean is more robust than a mean for skewed distributions, we propose an alternative Count Five test using the trimmed mean.

The properties of the Count Five test are introduced briefly in Section 2. In Section 3, we propose a trimmed mean centered Count Five statistic, and compare the results obtained using this method with those of other statistical methods described in McGrath and Yeh (2005). And the normal mixture studied by Pan (1999) is extended and used to generate random samples. The Count Five statistic of McGrath and Yeh (2005) and the alternative statistic proposed in this paper are compared for normal mixtures and their properties are discussed. Some advantages of the Count Five tests are given in Section 4.

2. Count Five Statistic

Let $X_1, ..., X_{n_x}$ and $Y_1, ..., Y_{n_y}$ be independent random samples from the same distribution with $E(X_i) = \mu_x$, $Var(X_i) = \sigma_x^2$, $E(Y_i) = \mu_y$, and $Var(Y_i) = \sigma_y^2$ for $i = 1, ..., n_x$, $j = 1, ..., n_y$. We assume that μ_x and μ_y are known. Then, the absolute deviations $|X_i - \mu_x|$ and $|Y_j - \mu_y|$ are iid under $H_0 : \sigma_x^2 = \sigma_y^2$. Let $C_x^{(\mu)}$ be the extreme count for the X sample, that is, the number of $|X_i - \mu_x|$ that exceed the maximum among $|Y_j - \mu_y|$'s:

$$C_x^{(\mu)} = \#\{i : |X_i - \mu_x| > \max_j |Y_j - \mu_y|\}$$

Also $C_y^{(\mu)}$ can be defined analogously.

The tail probability under H_0 is obtained in order to use an application of the hypergeometric distribution such as

$$P(C_x^{(\mu)} \geq m \mid H_0) = \frac{\binom{n_x}{m}\binom{n_y}{0}}{\binom{n_x + n_y}{m}}. \tag{2.1}$$

If $n_x = n_y = n = N/2$ and $m \geq 2$, then equation (2.1) can be simplified to

$$P(C_x^{(\mu)} \geq m|H_0) = P(C_y^{(\mu)} \geq m|H_0) \qquad (2.2)$$

$$= \frac{1}{2^m} \prod_{k=1}^{m-1} \frac{N-2k}{N-k}.$$

Note that for $m < n < \infty$ in (2.2), $P(C_x^{(\mu)} \geq m|H_0) = P(C_y^{(\mu)} \geq m|H_0) < 2^{-m}$ $= .03125$ for $m = 5$. Thus, we can perform a two-sided test using a critical value of $m = 5$ and have a significance level $< .0625$ for finite n regardless of the distributional family.

In practice, both μ_x and μ_y are generally not known and must be estimated from the sample. When we substitute $\overline{X}$ and $\overline{Y}$ for μ_x and μ_y in (2.1), respectively, we obtain

$$C_x^{mean} = \# \left\{ i : |X_i - \overline{X}| > \max_j |Y_j - \overline{Y}| \right\}$$

The size of this test remains approximately the same as that of $C_x^{(\mu)}$. Hence, McGrath and Yeh (2005) suggested that if we "count five" when $n_x = n_y$, then we reject the null hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$. In other words, if $C_x^{mean} \geq 5$, we conclude that $\sigma_x^2 > \sigma_y^2$ and if $C_y^{mean} \geq 5$, we conclude that $\sigma_y^2 > \sigma_x^2$. The statistic C for equal sample sizes is referred to as *Count Five* statistic. McGrath and Yeh (2005) also proposed a median centered Count Five statistic $C_x^{med} = \# \left\{ i : |X_i - \widetilde{X}| > \max_j |Y_j - \widetilde{Y}| \right\}$, where $\widetilde{X}$ and $\widetilde{Y}$ are the sample medians.

When $n_x \neq n_y$, the critical value of $m = 5$ may not be appropriate for both C_x and C_y. We can find an appropriate critical value for C_x by substituting different values of m into equation (2.1) until we obtain a value m_0 satisfying $P(C_x \geq m_0|H_0) \approx \alpha/2$. Alternatively, equation (2.1) could be approximated as $(\frac{n_x}{n_x + n_y})^m$ using Stirling's formular and, assuming that $n_x, n_y > m \geq 2$, and one may solve this for $\alpha/2$.

McGrath and Yeh (2005) did simulation studies of the F, F_1 tests of Shoemaker (2003), the L test of Levene (1960), the modified Levene's test (*ML*) of Brown and Forsythe (1974), and the C^{mean} and C^{med} tests. The sizes and powers of these tests were obtained for equal sample sizes of $n_x = n_y = n = 20$, 30, 40, 50, 60 and the ratios of the dispersion parameters, $\delta = \sigma_y/\sigma_x = 1$, $\sqrt{2}$. Also, a variety of distributions of X_i were studied: standard normal, laplace, uniform, standard lognormal, exponential, t-distribution with $d.f. = 4$, and a normal mixture previously studied by Pan (1999). The Y_j observations were generated independently from the above mentioned distributions and then multiplied by the

dispersion ratio, so that $Var(Y_j) = \delta^2 Var(X_i)$ with $\delta = 1, \sqrt{2}$.

Based on the performance of each test under the null hypothesis of $\delta = 1$, one can see that while all attain a size close to the nominal size=0.05 for the normal distribution, the F, L, and C^{med} tests fail to attain the nominal size for the other distributions. The F test performs poorly for all nonnormal distributions, whereas the size of the L test is larger than the nominal size for the lognormal and exponential distributions, as is the size of the C^{med} test for the uniform distribution. So, of the six tests studied, it appears that the F_1, ML, and C^{mean} tests are the most reliable for attaining the nominal size.

Under the alternative hypothesis of $\delta = \sqrt{2}$, the powers of the F_1, ML, and C^{mean} tests did not differ substantially for six of the seven distributions studied. (The exception is the uniform distribution.) In fact, the Count Five test is more powerful than any of the other tests in the case of a uniform distribution, and its power lags behind that of the other two-sample dispersion tests for the other distributions when $n > 20$. As the sample size increases, the power increases slowly for most distributions, but decreases slightly for a normal mixture, where $X_i \sim N(0,1)$ with probability 0.9 and $X_i \sim N(0,4)$ with probability 0.1.

In the next section, we will explore these properties of the Count Five test for certain distributions, including various normal mixtures, and propose an alternative Count Five statistic using the trimmed mean.

3. Count Five with Trimmed Means and Simulation Results

The critical value of the Count Five test is always 5 for equal sample sizes, so that the advantage of these quick and compact test for comparing the dispersions of two samples does not require the calculation of ranks or reference tables. However, since the Count Five test uses only extreme values, this statistic discards some information from the samples, which often results in a degradation in power. For large samples, the power of the Count Five test may be relatively low and other tests may be preferred (McGrath and Yeh 2005).

In order to overcome these disadvantages, we propose an alternative Count Five statistic using the trimmed mean as the follows:

$$C_x^{\alpha\,mean} = \# \left\{ i : |X_i - \overline{X^\alpha}| > \max_j |Y_j - \overline{Y^\alpha}| \right\}$$

where $\overline{X^\alpha}$ and $\overline{Y^\alpha}$ are the $\alpha\%$ trimmed sample means.

Adapting the simulation methods of McGrath and Yeh (2005), 10,000 sets of random samples of X_i and Y_i are generated with the above mentioned distributions, the sample sizes of $n = 20, 30, 40, 50, 60$, and dispersion ratios of

$\delta = 1, \sqrt{2}$. The F, F_1, L, ML, C^{mean}, and C^{med} tests are considered, as well as the $\alpha\%$ trimmed mean centered Count Five $C^{\alpha\,mean}$ test.

Since the sizes and powers of the F, F_1, L, ML, C^{mean}, and C^{med} tests for some distributions were shown and discussed in <Figures 2 and 3> of McGrath and Yeh (2005), As would be expected, the sizes and powers of the $\alpha\%$ trimmed mean centered Count Five $C^{\alpha\,mean}$ ($\alpha = 5, 10, 20$) tests have similar behaviors to those of the C^{mean} test. <Figure 3.1> presents the sizes and powers of all of the Count Five tests (C^{mean}, C^{med}, $C^{5\,mean}$, $C^{10\,mean}$, and $C^{20\,mean}$) for some distributions in order to compare the three kinds of trimmed mean centered Count Five tests with the C^{mean} and C^{med} tests.

From the plots on the left-hand side of <Figure 3.1>, we could find that the C^{mean}, $C^{5\,mean}$, and $C^{10\,mean}$ tests (but not the $C^{20\,mean}$ test with a large sample size and C^{med} test) attain a size of close to the nominal size=0.05 for the uniform distribution. For the normal and laplace distributions, all of the tests obtain the nominal size. The C^{med} test comes closer to the nominal size than the others for the exponential distribution. For the lognormal distribution, the C^{med} and $C^{20\,mean}$ tests perform better than the other tests when the sample size is greater than 40.

From the plots on the right-hand side of <Figure 3.1>, it can be seen that for the uniform, normal, and laplace distributions, the powers of all of the Count Five tests exhibit similar behaviors. However the powers of all of the Count Five tests come close to 1 for the uniform distribution, whereas the powers for the other distributions are very low. In this situations, we can see that the performances of the tests for the exponential and lognormal distributions are in the order C^{med} test $> C^{20\,mean}$ test $> C^{10\,mean}$ test $> C^{5\,mean}$ test $> C^{mean}$ test. In other words, for the exponential and lognormal distributions, the power of the C^{med} test is the highest and that of C^{mean} test is the lowest. As the value of $\alpha\%$ increases, the power of the $C^{\alpha\,mean}$ test comes closer to that of the C^{med} test, which is the most powerful test for skewed distributions.

Therefore, for the uniform distribution, the mean centered Count Five statistic, and the 5%, 10%, and possibly 20%, trimmed mean centered Count Five statistics perform better than the median centered Count Five statistic with respect to the size and power of the test. We can say that, in comparison with the other two-sample dispersion test statistics such as the F, F_1, L, ML tests, both of the mean centered Count Five test and the trimmed mean centered Count Five test perform very well for the uniform distribution. Hence we would like to generate random samples following analogous uniform distributions, such as the following

normal mixtures with three normal distributions:

$$① \quad (N(0,1)+N(2,1)+N(4,1))/3$$
$$② \quad (N(0,1)+N(2.5,1)+N(5,1))/3$$
$$③ \quad (N(0,1)+N(3,1)+N(6,1))/3$$

From the sizes and powers for these normal mixtures in <Figure 3.2>, it can be seen that the C^{mean}, $C^{5\,mean}$, $C^{10\,mean}$, and $C^{20\,mean}$ tests behave similarly and attain the nominal size, whereas the C^{med} test exceeds the nominal size. Moreover, the powers of all of these Count Five tests are very much identical and have low values. This phenomenon becomes even more remarkable as the differences in the means in the normal mixtures increase from ① to ③.

$P(\text{Type I error})$, nominal=0.05   Power under H_1
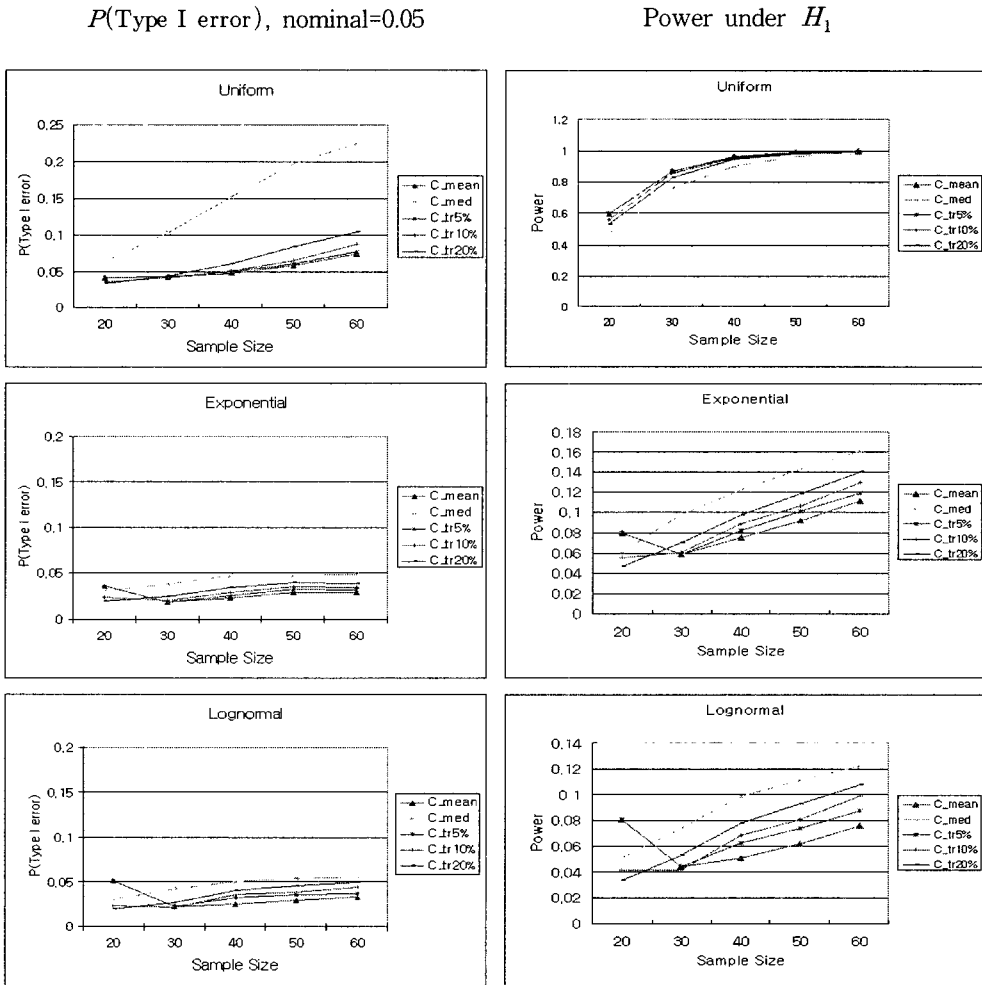
<Figure 3.1> $P(\text{Type I error})$ and Power

P(Type I error), nominal=0.05   Power under H_1
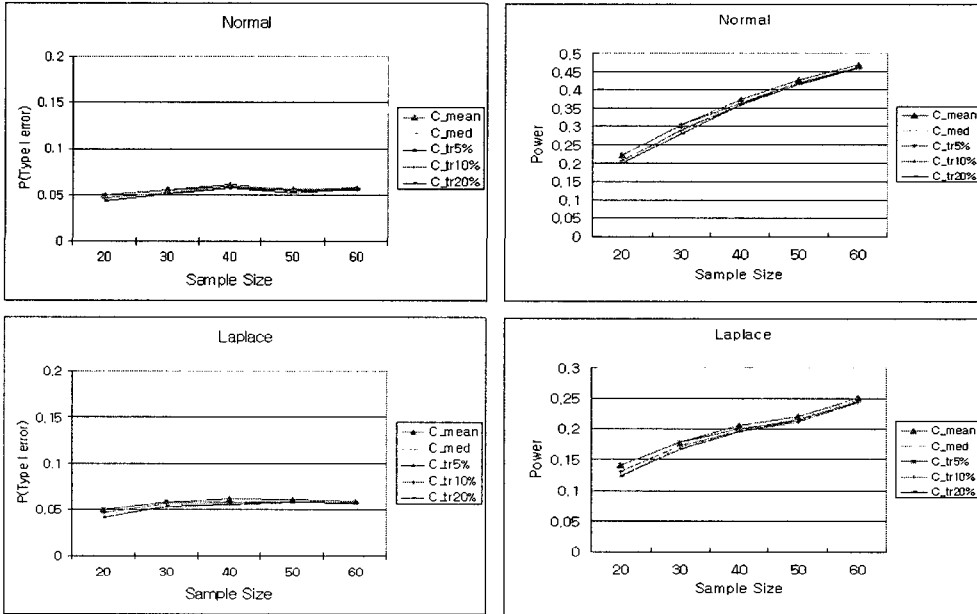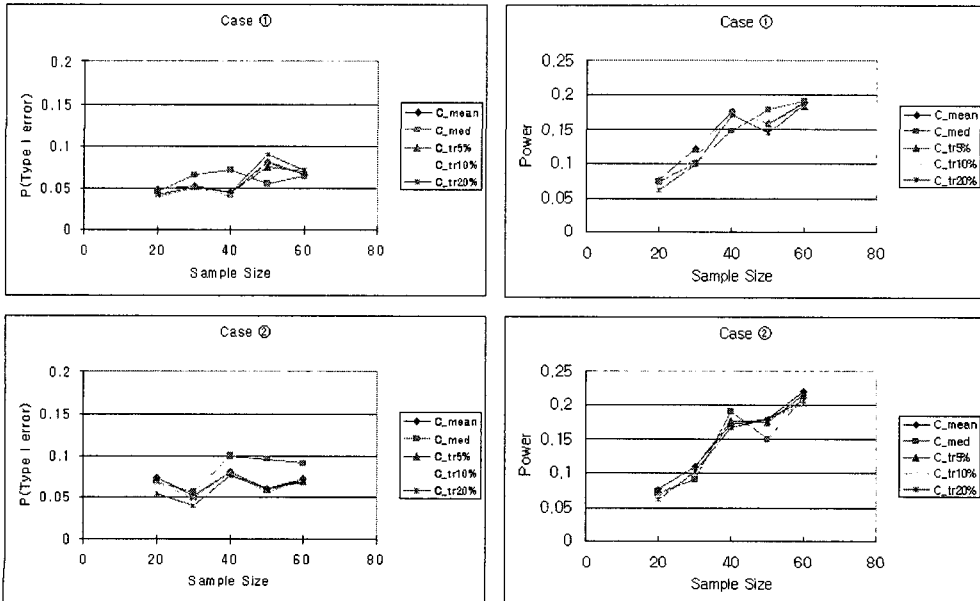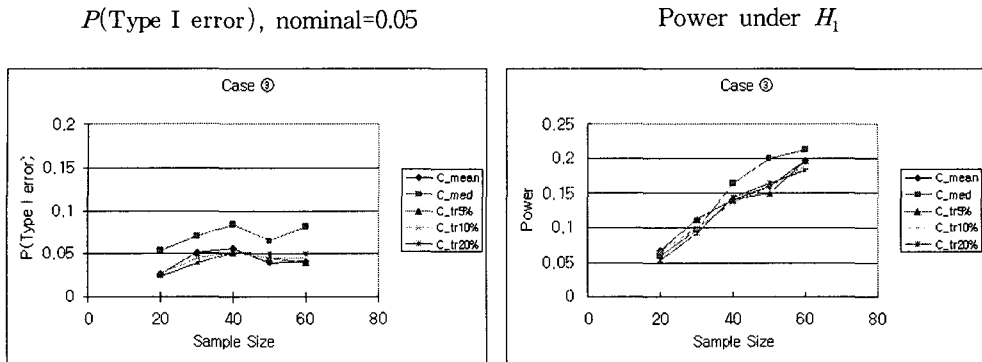

<Figure 3.1> Continued

P(Type I error), nominal=0.05   Power under H_1

<Figure 3.2> P(Type I error) and Power for Normal Mixtures

P(Type I error), nominal=0.05　　　　　　　　　Power under H_1

<Figure 3.2> Continued

4. Conclusion

Although most two-sample dispersion tests require tables of critical values or software to calculate the p-value, the Count Five test can be performed simply by calculating the number of extreme points. The critical value of the Count Five test is always 5 for equal sample sizes and, even in the case of unequal sample sizes, it can be obtained with ease.

In this paper, an alternative Count Five statistic using the trimmed mean, $C^{\alpha\,mean}$ ($\alpha = 5, 10, 20$), is proposed and compared with other two-sample dispersion tests, viz. the F, F_1, L, ML, C^{mean}, and C^{med} statistics, for the many distributions studied by McGrath and Yeh (2005), as well as for normal mixtures. We found that the performance of the $C^{\alpha\,mean}$ test is similar to those of the C^{mean} and C^{med} tests, but that the $C^{\alpha\,mean}$ test is more powerful than the other dispersion tests, including the F, F_1, L, ML tests, in the case of the uniform distribution (see <Figure 4.1>). Secondly, the $C^{\alpha\,mean}$ test is better than the C^{mean} test with respect to the size and power for the lognormal and exponential distributions (see <Figure 3.1>). As the value of $\alpha\%$ increases, the power of the trimmed mean centered Count Five statistic increases for skewed distributions, such as the lognormal and exponential distributions. Finally, for normal mixtures corresponding to case ③ above, the powers of the Count Five tests, including the trimmed mean centered Count Five test, are not as high as that exhibited for the uniform distribution, but are higher than those of the other tests, such as the F, F_1, L, ML tests (see <Figure 4.1>). Therefore, we can conclude that the trimmed mean centered Count Five test proposed in this paper attains the nominal size and is more powerful than the other dispersion tests for heavy tailed and skewed distributions, such as the uniform, lognormal, exponential distributions, and normal

mixtures (for example, case ③).

P(Type I error), nominal=0.05   Power under H_1
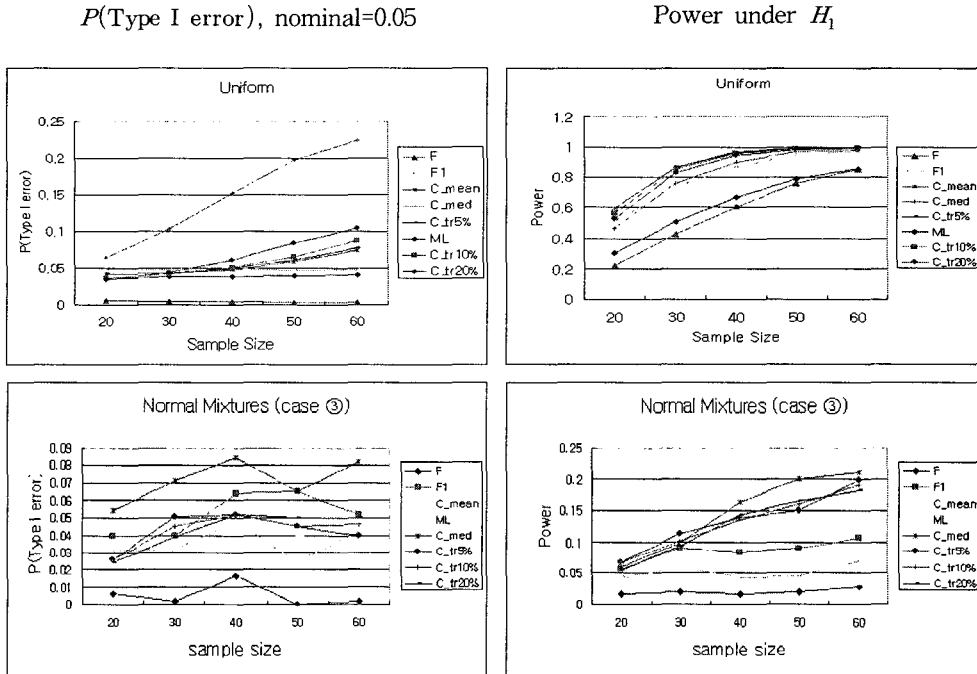
<Figure 4.1> Performance of the Count Five for Uniform and Normal Mixtures

References

[1] Ansari, A.R. and Bradley, R.A. (1960). Rank Sum Tests for Dispersions. *Annals of Mathematical statistics*, Vol. 31, 1174-1189.

[2] Brown, M.B. and Forsythe, A.B. (1974). Robust Test for the Equality of Variances. *Journal of the American Statistical Association*, Vol. 69, 364-367.

[3] Conover, W.J. (1980). *Practical Nonparametric Statistics* (2nd ed). New York: Wiley.

[4] Fligner, M.A. and Killeen, T.J. (1976). Distribution-Free Two-sample Tests for Scale. *Journal of the American Statistical Association*, Vol. 71, 210-213.

[5] Hall, P. and Padmanabhan, A.R. (1997). Adaptive Inference for the Two-sample Scales Problem. *Technometrics*, Vol. 39, 412-422.

[6] Klotz, J. (1962). Nonparametric tests for Scale. *Annals of Mathematical Statistics*, Vol. 33, 498-512.

[7] Levene, H. (1960). Robust Tests for Equality of Variances. in *Contributions to*

Probability and Statistics. Ed. I. Olkin, Palo Alto, CA: Stanford University Press.

[8] McClave, J.T., Benson, P.G., and Sincich, T. (2005). *Statistics for Business and Economics* (9th ed.). Englewood Cliffs, NJ: Pearson Prentice Hall.

[9] McGrath, R.N. and Yeh, A.B. (2005). A Quick, Compact, Two-Sample Dispersion Test : Count Five. *The American Statistician.* Vol. 59, 47-53.

[10] Mood, A.M. (1954). On the Asymptotic Efficiency of Certain Nonparametric Two-Sample Tests. *Annals of Mathematical Statistics*, Vol. 25, 514-522.

[11] Moore, D.S. and McCabe, G.P. (2003). *Introduction to the Practice of Statistics* (4th ed.). New York: W. H. Freeman.

[12] Ott, R.L. and Longnecker, M. (2001). *An Introduction to Statistical Methods and Data Analysis* (5th ed.). Belmont, CA: Duxbury.

[13] Pan, G. (1999). On a Levene Type Test for the Equality of Two Variances. *The American Statistician*, Vol. 57, 105-114.

[14] Shoemaker, L.H. (2003). Fixing the F test for Equal Variances. *The American Statistician*, Vol. 57, 105-114.

[15] Siegel, S. and Tukey, J.W. (1960). A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples. *Journal of the American Statistical Association*, Vol. 55, 429-455.