Journal of the KSQC
Vol.21, No.1, June 1993

1993년 6월 韓國品質管理學會誌 제21권 제1호

# Tests for Asymmetry Associated with the Linear Signed Rank Statistics

Soon-Ock Kim *

## ABSTRACT

Tests associated with the linear signed rank statistics are considered for testing the symmetry of a continuous distribution about an unknown median. The results of Monte Carlo study show that the proposed tests are reasonably good in level control and powers.

## 1. Introduction

Let $F(x-\theta)$ be a distribution function of a continuous population with unknown median $\theta$. Denote a random sample from $F(x-\theta)$ by $X_1, X_2, \cdots, X_n$. It is desired to test the null hypothesis

$$H_0 : F(\theta - x) = 1 - F(\theta - x) \quad \text{for all} \quad x \qquad (1.1)$$

* Statistical Research Institute, College of Natural Science, Seoul National University

against the alternative that $F$ is not symmetric. In one-sided test, the alternative may indicate the underlying distribution which is more spread out to the right(left) of $\theta$ than it is to the left(right). To test (1.1) Gupta(1967) modified the Wilcoxon signed rank statistic. He estimated the center by the sample median and applied the Wilcoxon test. He also derived the asymptotic distribution of the proposed test statistic. When the center is estimated, linear rank statistics lose their distribution-free property. For the purpose of getting level robustness, Bhattacharya, Gastwirth and Wright(1982) and Antille, Kersting and Zucchini(1982) studied the trimmed version of Wilcoxon statistic. Here we consider the tests associated with the linear signed rank statistics proposed by S.O.Kim(1992). The tests are presented in Section 2 and their asymptotic properties are discussed. Section 3 contains the results of Monte Carlo study comparing the powers of the procedures.

## 2. Tests and Their Asymptotic Properties

For testing the symmetry of the underlying distribution, the general linear rank-like statistics can be expressed as follows :

$$S^* = \sum_{i=1}^{n} \Psi_i a_n(R_i^*) \qquad (2.1)$$

where

$$\Psi_i = \begin{cases} 1, & \text{if} \quad X_i - \hat{\theta} \geq 0 \\ 0, & \text{if} \quad X_i - \hat{\theta} < 0, \end{cases}$$

$R_i^*$ is the rank of $|X_i - \hat{\theta}|$ among $|X_1 - \hat{\theta}|, \cdots, |X_n - \hat{\theta}|$, where $\hat{\theta}$ is the sample median, and $a_n(i)$'s are scores. When $a_n(i) = i$, (2.1) is exactly Gupta's statistic, say $S_W{}^*$.

$$S_W{}^* = \sum_{i=1}^{n} \Psi_i R_i^*$$

Kim(1992) proposed to use the scale relevant scores such as

$$a_M(i) = i^2$$

and

$$a_N(i) = \left\{ \Phi^{-1}\left( \frac{1}{2} + \frac{i}{2(n+1)} \right) \right\}^2$$

which lead to the statistics

$$S_M{}^* = \sum_{i=1}^{n} \Psi_i R_i^{*2} \qquad (2.2)$$

and

$$S_N{}^* = \sum_{i=1}^{n} \Psi_i \left\{ \Phi^{-1}\left( \frac{1}{2} + \frac{R_i^*}{2(n+1)} \right) \right\}^2 \qquad (2.3)$$

The asymptotic null distribution of any linear rank-like test statistic of the form (2.1) is obtained by Bhattacharya *et al.*(1982) and given in Theorem 2.1.

**Theorem 2.1** If $f$ is a symmetric density function of $F$, differentiable at its median, and $S^*$ is a linear signed rank-like statistic of the form (2.1), then $\frac{1}{\sqrt{n}}(S^* - E_0(S^*)) = \frac{1}{2}\sum_{i=1}^{n} a_n(i)$, is asymptotically normal with mean 0 and variance

$$\int_0^\infty \phi^2\{H(x)\}dH(x) + b^{+2}\{4f^2(0)\}^{-1} - b^+\{f(0)\}^{-1}\int_0^\infty \phi\{H(x)\}dH(x), \qquad (2.4)$$

where $\phi(\cdot)$ is a square integrable score function satifying $a_n(i) = \phi(i/(n+1))$, and $H(x)$ and $b^+$ are defined by

$$H(x) = F(x) - F(-x), \qquad b^+ = \int_0^1 \phi(u)\phi^+(u, f)du,$$

respectively, with

$$\phi^+(u, f) = -f'[F^{-1}\{(u+1)/2\}] / f[F^{-1}\{(u+1)/2\}]$$

The scores defining $S_M{}^*$ and $S_N{}^*$ in (2.2) and (2.3) are related to the score generating functions

$$\phi(u) = u^2, \quad \phi(u) = \left\{ \Phi^{-1}\left( \frac{1}{2} + \frac{u}{2} \right) \right\}^2,$$

〈 Table 2.1 Asymptotic Variances of Statistics $S_M^*$, $S_N^*$ 〉

| | Normal | Logistic | Double Exponential | Cauchy |
|---|---|---|---|---|
| $\sigma_M^2$ | 0.0344 | 0.0292 | 0.0222 | 0.0226 |
| $\sigma_N^2$ | 0.75 | 0.6013 | 0.5 | 0.5243 |

respectively. For these $\phi$ functions we have the following corollaries.

**Corollary 2.1** Under $H_0$, the statistic $\frac{1}{(n+1)^2\sqrt{n}}\left(S_M^* - E_0(S_M^*)\right)$ is asymptotically normal with mean 0 and variance

$$\sigma_M^2 = \frac{1}{45} + \frac{1}{16}\left(\frac{b^+_M}{f(0)} - \frac{2}{3}\right)^2 \tag{2.5}$$

where

$$b^+_M = 8\int_0^\infty (2F(x)-1)f^2(x)\,dx.$$

**Corollary 2.2** Under $H_0$, the statistic $\frac{1}{\sqrt{n}}\left(S_N^* - E_0(S_N^*)\right)$ is asymptotically normal with mean 0 and variance

$$\sigma_N^2 = \frac{1}{2} + \frac{1}{16}\left(\frac{b^+_N}{f(0)} - 2\right)^2 \tag{2.6}$$

where

$$b^+_N = 4\int_0^\infty \frac{\Phi^{-1}(F(x))f^2(x)}{\phi(\Phi^{-1}(F(x)))}\,dx.$$

Since $F$ and $f$ appear in the variance, (2.4) is not distribution-free. Kim(1992) Proposed to use $F_0$ = logistic as a reference point to studentize the statistics $S_M^*$ and $S_N^*$. When the underlying distribution varies over the whole family of unimodal symmetric densities, the bounds of variances are given in the following Lemma.

**Lemma.** We have :

( i ) $\frac{1}{45} \leq \sigma_M{}^2 \leq \frac{2}{15}$,

( ii ) $\frac{1}{48} \leq \sigma_N{}^2 < \infty$.

The proof is easy and left to the reader. Table 2.1 gives the asymptotic variances of $S_M{}^*$ and $S_N{}^*$ and from the values we find that logistic is a reasonable choice. Through a Monte Carlo study we show in Section 3 that the dependence on $F_0$ is small.

## 3. Monte Carlo Results

We performed Monte Carlo studies to evaluate the ability of the tests based on the statistics $S_M{}^*$ and $S_M{}^*$. For comparative purposes, a classical test is included. Let $m_k$ be the $k$-th sample moment about the sample mean. Consider the statistic

$$b_1 = m_3 / m_2{}^{3/2}.$$

Provided $\int_{-\infty}^{\infty} x^6 dF(x) < \infty$, $b_1$ can be used to test the null hypothesis that the coefficient of skewness is zero. Being often used to test normality, the statistic $b_1$ is well-known measure of asymmetry, and a test for asymmetry. Gupta (1967) studied this statistic as a testing procedure of symmetry and proved $b_1 / Var_0(b_1)$ has asymptotic normal distribution. Under the hypothesis of symmetry

$$E_0(b_1) = 0 \quad : \quad Var_0(b_1) = (\mu_6 - 6\mu_2\mu_4 + 9\mu_2{}^3) / n\mu_2{}^3,$$

where $\mu_k$ is the $k$-th central moment of the null underlying distribution. $b_1$-test is not distribution-free, but standardized by corresponding function of the sample moments it can be made asymptotically distribution-free. Large (absolute) value of $b_1$ leads to the rejection of the hypothesis that the underlying distribution is symmetric.

We evaluate the stability of the level in normal, logistic, double exponential, and Cauchy distribution. As the alternatives we selected 6 members of the generalized lambda family (GLF) discussed in Ramberg and Schmeiser (1974). This family is easily generated since it is defined in terms of the inverse cumulative distribution function :

$$F^{-1}(u) = \lambda_1 + [u\lambda_3 - (1-u)\lambda_4] / \lambda_2, \quad 0 < u < 1,$$

where $\lambda_1$ is a location parameter, $\lambda_2$ is a scale parameter and $\lambda_3$ and $\lambda_4$ are shape parameters. The values of the parameter $\lambda$'s of the selected 6 GLF distributions are listed in Table 3.1 along with the associated skewness ($\alpha_3$) and kurtosis ($\alpha_4$) values.

We also included familiar asymmrtric distributions such as the Chi-square distribution $\chi^2(v)$, $v$ = 1, 2, 4, and 6. The performance of tests was investigated for the sample sizes of $n = 55$ and 101 using significance levels $\alpha = 0.05$ and 0.1. For each combination of underlying distributions and sample sizes, 500 random samples were generated. To generate

〈 Table 3.1 GLF Distributions Used in the Monte Carlo Study 〉

| Distribution | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\alpha_3$ | $\alpha_4$ |
|---|---|---|---|---|---|---|
| GLF 1 | 0.0 | 1.0 | 1.4 | 0.25 | 0.5 | 2.2 |
| GLF 2 | 0.0 | 1.0 | 0.00007 | 0.1 | 1.5 | 5.8 |
| GLF 3 | 3.586508 | 0.04306 | 0.025213 | 0.094029 | 0.9 | 4.2 |
| GLF 4 | 0.0 | −1.0 | −0.0075 | −0.03 | 1.5 | 7.5 |
| GLF 5 | 0.0 | −1.0 | −0.001 | −0.13 | 3.16 | 23.8 |
| GLF 6 | 0.0 | −1.0 | −0.0001 | −0.17 | 3.88 | 40.7 |

〈 Table 3.2 Monte Carlo Estimates of the True Level 〉

| | Normal | Logistic | Double Exponential | Cauchy |
|---|---|---|---|---|
| | | $\alpha = 0.10$ ($n = 55$) | | |
| $b$ | 0.132 | 0.096 | 0.140 | 0.094 |
| $S_M^*$ | 0.116 | 0.092 | 0.084 | 0.098 |
| $S_N^*$ | 0.090 | 0.064 | 0.054 | 0.054 |
| | | $\alpha = 0.10$ ($n = 101$) | | |
| $b$ | 0.096 | 0.082 | 0.108 | 0.090 |
| $S_M^*$ | 0.106 | 0.080 | 0.072 | 0.076 |
| $S_N^*$ | 0.092 | 0.062 | 0.036 | 0.064 |
| | | $\alpha = 0.05$ ($n = 55$) | | |
| $b$ | 0.050 | 0.030 | 0.054 | 0.018 |
| $S_M^*$ | 0.064 | 0.058 | 0.038 | 0.038 |
| $S_N^*$ | 0.040 | 0.026 | 0.020 | 0.018 |
| | | $\alpha = 0.05$ ($n = 101$) | | |
| $b$ | 0.052 | 0.022 | 0.048 | 0.012 |
| $S_M^*$ | 0.062 | 0.036 | 0.026 | 0.032 |
| $S_N^*$ | 0.042 | 0.020 | 0.020 | 0.022 |

the random numbers we used IMSL FORTRAN Subroutine GGUBT, GGNPM and transformed these variates into observations from each of the selected distributions using the appropriate inverse cumulative distribution functions.

To compute the empirical levels and powers of tests, we counted the number of times that the null hypothesis $H_0$ was rejected. Here we present the results of two-sided test and which are given in Table 3.2～Table 3.4. Table 3.2 shows the empirical levels of the tests. It is apparent that the test $S_M{}^*$ holds the $\alpha$ level reasonably well for all four of the symmetric distributions while the test $S_N{}^*$ somewhat conservative. We think that $S_M{}^*$ is reasonably valid test for asymmetry over the set of symmetric distributions likely to occur in practice. Table 3.3～Table 3.4 give the empirical powers of the tests. Since the results are similar, the empirical pewers for the significance level $\alpha = 0.1$ are not presented here. The values in Table 3.3～Table 3.4 show that the tests $S_M{}^*$ and $S_N{}^*$ are efficient when the underlying distribution is nearly J-shaped. Considering both robustness of validity and good efficiency, $S_M{}^*$ is recommended.

〈 Table 3.3 Empirical Power of Tests at $\alpha = 0.05(\times 500)$ 〉

| | GLF 1 | GLF 2 | GLF 3 | GLF 4 | GLF 5 | GLF 6 |
|---|---|---|---|---|---|---|
| | | | $n = 55$ | | | |
| $b$ | 377 | 434 | 316 | 333 | 337 | 326 |
| $S_M{}^*$ | 272 | 404 | 184 | 237 | 434 | 446 |
| $S_N{}^*$ | 326 | 443 | 207 | 276 | 473 | 477 |
| | | | $n = 101$ | | | |
| $b$ | 473 | 483 | 437 | 424 | 381 | 368 |
| $S_M{}^*$ | 389 | 481 | 293 | 365 | 489 | 497 |
| $S_N{}^*$ | 460 | 497 | 381 | 452 | 499 | 500 |

〈 Table 3.4 Empirical Power of Tests at $\alpha = 0.05(\times 500)$ 〉

| | $\chi^2(1)$ | $\chi^2(2)$ | $\chi^2(4)$ | $\chi^2(6)$ |
|---|---|---|---|---|
| | | $n = 55$ | | |
| $b$ | 358 | 406 | 402 | 374 |
| $S_M{}^*$ | 489 | 434 | 317 | 241 |
| $S_N{}^*$ | 496 | 465 | 361 | 272 |
| | | $n = 101$ | | |
| $b$ | 423 | 443 | 468 | 452 |
| $S_M{}^*$ | 500 | 482 | 413 | 340 |
| $S_N{}^*$ | 500 | 498 | 475 | 438 |

# REFERENCES


1. Antille, A., Kersting, G. and Zucchini, W. (1982). "*Testing Symmetry*," Journal of the American Statistical Association, 77, 639−646.
2. Bhattacharya, P.K., Gastwirth, J.L. and Wright, A.L. (1982). "*Two modified Wilcoxon tests for symmetry about an unknown location parameter*," Biometrika, 69, 377−382.
3. Gupta, M.K. (1967). "*An Asymptotically Nonparametric Test of Symmetry*," The Annals of Mathematical Statistics, 38, 849−866.
4. Kim, S.O. (1992). "*Tests for Symmetry Based on Linear Signed Rank Statistics*," Ph.D. Dissertation, Department of Computer Science and Statistics, Seoul National University, Seoul Korea.
5. Ramberg, J.S. and Schmeiser, B.W (1974). "*An Approximate Method for Generating Asymmetric Random Variables*," Communications of the ACM, 17, 78−82.