한국통계학회논문집 제3권 1호
1996년 4월 pp. 21-27

Testing Procedure for Scale Shift at an Unknown Time Point[1)]

Il Seong Song[2)]

Abstract

A testing procedure is considered to the problem of testing whether there exists a shift in scale at an unknown time point when a fixed number of observations are drawn successively in time. A test statistic based on squared ranks test for equal variances is suggested and its aymptotic distribution is dereived. Small sample power comparisons are performed.

1. Introduction

Consider the situation where observations are generated successively and the distribution of this random sequence is subject to change in scale at possibly unknown time point. We investigate a nonparametric rank test for detecting possible change of a scale parameter.

Suppose that $X_1, X_2, \dots, X_N$, all having a constant known mean μ, are independent random variables that are successively drawn from continuous populations $F_i(x)$, $i=1,\cdots,N$, respectively. We assume that there exists a cdf $F(\cdot)$ such that

$$F_i(x) = F(\frac{x-\mu}{\theta_i}), \quad i=1,2,\cdots,N,$$

where θ_i are unknown. For the possible change of scale parameter θ_i, we consider the problem of testing the null hypothesis

$$H_0:\ \theta_1=\theta_2=\cdots=\theta_N=\theta_0 \quad (\theta_0 \textit{ is unknown})$$

against the alternative

$$\begin{aligned} H_1:\ &\theta_1=\cdots=\theta_\tau=\theta_0 \\ &\theta_{\tau+1}=\cdots=\theta_N=\theta_0+\Delta, \quad \Delta>0 \end{aligned} \qquad (1.1)$$

---

1) This paper was supported by the research grant of Sungshin Women's University in 1993.
2) Professor, Department of Statistics, Sungshin Women's University, Seoul, 136-742, Korea.

where Δ and τ ($1 \leq \tau < N$) are unknown.

This sequence of random variables is said to have change-point τ. We make no restrictions on the functional form of the underlying distribution $F(\cdot)$ except that it is continuous. In this paper we propose a rank test based on squared ranks statistic for this problem and derive its aymptotic distribution.

From the parametric viewpoint, Hsu(1977, 1979) suggested some change-point tests for detecting scale shift under the normal and the gamma distribution. Talwar and Gentle(1981) suggested a simple test for scale shift which is robust for heavy-tailed distributions. On the other hand, Ali and Giaccotto(1982) studied nonparametric tests for randomness against location and scale-shift alternatives through the analysis of stock prices. Hsieh(1984) suggested a class of rank tests for this change-point problem, and obtained their optimal properties. Gastwirth and Mahmoud(1986) considered the maximum efficiency robust test for the scale change for the gamma family.

In Section 2, a test procedure is developed and asymptotic null distribution of the proposed test statistic is derived. Section 3 is devoted to brief reviews of Hsu's(1977) normal theory test and modified Pettitt's(1979) test for power comparison with our rank test. The results of a small sample Monte Carlo study are presented in Section 4.

2. The Proposed Test Statistic and its Asymptotic Normality

Consider the change-point problem of (1.1). If the change-point $\tau = k$ is fixed, it can be considered as a testing problem for equal variances in two random samples. Let $X_1, X_2, \cdots, X_k$ denote the random sample of size k from population $F(\frac{x-\mu}{\theta_0})$ and let $X_{k+1}, \cdots, X_N$ denote the random sample of size $N-k$ from population $F(\frac{x-\mu}{\theta_0+\Delta})$. The squared ranks statistic $\sum_{i=k+1}^{N} R_i^2$ for testing equal variances can be used as the test statistic for testing the null hypothesis $\Delta = 0$, where R_i is the combined rank of $U_i = |X_i - \mu|$, $i = 1, \cdots, N$. We propose a sum-type test statistic based on the squared ranks statistic as defined by

$$S = \sum_{k=1}^{N-1} c_{N,k} \sum_{i=k+1}^{N} a_N(R_i) , \tag{2.1}$$

with large values being significant in favor of the alternative (1.1), where $c_{N,k}$ are the nonnegative weight coefficients and

$$a_N(R_i)=[\ R_i^2-(N+1)(2N+1)/6]\ /N^2 \tag{2.2}$$

is the nondecreasing rank score. From a Bayesian viewpoint, $c_{N,k}/\sum_{k=1}^{N-1} c_{N,k}$ can be interpreted as the prior probability that X_{k+1} is the initial shifted variable. If we take uniform weights according to an uninformative prior, the statistic (2.1) can be rewritten as

$$S=\sum_{i=1}^{N}(i-1)\ a_N(R_i)\ . \tag{2.3}$$

The asymptotic null distribution of the proposed test statistic S can be derived from the well-known theory of linear rank statistics. Our rank test S of (2.3) can be expressed as

$$S=\sum_{i=1}^{N} d_{N,i}\, a_N(R_i),$$

with $d_{N,i} = i-1$. We obtain the following result from Hajek and Sidak(1967).

Theorem 2.1. Suppose that F is absolutely continuous with density f,

$$\lim_{N\to\infty}\sum_{i=1}^{N}(d_{N,i}-\bar{d}_N)^2/\max(d_{N,i}-\bar{d}_N)^2=\infty \ \textit{with}\ \bar{d}_N=\sum_{i=1}^{N} d_{N,i}/N,$$

and there exists a square integrable function $\phi:[\ 0,1]\ \rightarrow \mathrm{R}$ such that

$$\sigma_\phi^2=\int\{\phi(u)-\bar{\phi}(u)\}^2 du<\infty \quad \text{with} \quad \bar{\phi}(u)=\int\phi(u)\,du$$

and

$$\lim_{N\to\infty}\int[\ a_N(1+[\ Nu]\ )-\phi(u)]\ ^2 du=0.$$

Then under H_0, $S=\sum_{i=1}^{N} d_{N,i}\,a_N(R_i)$ is asymptotically distributed $N(0,\sigma_N^{*2})$ with

$$\sigma_N^{*2}=\sum_{i=1}^{N}(d_{N,i}-\bar{d}_N)^2\int\{\phi(u)-\bar{\phi}(u)\}^2 du.$$

According to Theorem 2.1, our rank statistic S is asymptotically normal with mean zero and variance $N(N^2-1)/135$.

3. Parametric Test and Modified Pettitt Test

First, we present a brief review of the parametric test proposed by Hsu(1977). He described two tests for the change-point problem of (1.1) under the assumption of a normal distribution. One test is based on the statistic

$$T=\left\{\sum_{i=1}^{N}(i-1)Y_i\right\}/\left\{(N-1)\sum_{i=1}^{N}Y_i\right\}, \tag{3.1}$$

where $Y_i=(X_i-\mu)^2$. Hsu showed the test statistic T is locally most powerful, and provided approximate critical values for this test statistic.

Next, Talwar and Gentle(1981) proposed a distribution-free test, which is modified Pettitt's(1979) test for the change-point problem of (1.1). Let $Z_i=|X_i-X^*|$, $i=1,2,\cdots,N$, where X^* is the median of $X_1,X_2,\cdots,X_N$, and let $D_{ij}=\mathrm{sgn}(Z_j-Z_i)$, where sgn(y)=1 if y>0, sgn(y)=0 if y=0, and sgn(y)=-1 if y<0. For the change-point problem of (1.1), they suggested the following test statistic

$$K=\max_{1\le k<N}\left\{\sum_{i=1}^{k}\sum_{j=k+1}^{N}D_{ij}\right\}. \tag{3.2}$$

From the simulation work, they compared the test K with normal theory test T only for normally distributed data.

In the next section, we will compare our rank test S with the tests K and T for several underlying distributions through simulation work.

4. Small Sample Power Comparisons

To compare the performance of the tests T, K and S, we investigate the empirical power of the tests for the sample size N=30 when the alternative is specified by the change-point τ, amount of shift Δ and the underlying distribution F. We choose the change-point τ such that

$$\tau=3(3)27$$

and take θ_0=1 and $\Delta=0.5$. Since the performances of tests might depend on the tailweight and skewness in the underlying distribution, we consider the normal, logistic, double exponential, Cauchy and exponential distributions. All the testing procedures were run at a nominal significance level α=.05. The critical values were estimated based on 1,000 replications, and the power estimates were obtained based on 1,000 replications for each alternative through the Monte Carlo method.

The computation was done on the IBM PC and the uniform random numbers were generated by the subroutine of IMSL. The inverse transform method was used to generate logistic, double exponential, Cauchy and exponential random variables, and the Box-Muller method to generate normally distributed random variables.

We present the simulation results in Tables 4.1-4.5, which deal with our rank test S, Hsu's test T and modified Pettitt's test K. Table 4.1 shows the power estimates of the tests when the underlying distribution F is normal. As expected, the normal theory test T has good power in most instances when the underlying distribution is normal.

Table 4.1. Empirical powers of the tests T, K and S when F is N(0,1) and the sample size N=30.

τ	Test T	Test K	Test S
3	.096	.095	.094
6	.180	.145	.169
9	.212	.227	.240
12	332	.287	.317
15	.352	.283	.327
18	.363	.278	.289
21	.381	.265	.310
24	.279	.152	.191
27	.219	.119	.155

Table 4.2. Empirical powers of the tests T, K and S when F is logistic with mean zero, $\sigma^2 = \pi^2/3$ and the sample size N=30.

τ	Test T	Test K	Test S
3	.113	.083	.104
6	.148	.126	.167
9	.185	.182	.218
12	.275	.225	.278
15	.326	.238	.295
18	.349	.241	.285
21	.360	.216	.285
24	.282	.137	.210
27	.187	.089	.131

Table 4.2 presents the empirical powers of the same tests when the underlying distribution is logistic. The powers of Hsu's test T are higher than those of rank tests K and S in

most instances. We note that Hsu's test T is effective in the medium-tailed distributions.

Tables 4.3-4.4 give the empirical powers of the same tests when the underlying distributions are double exponential and Cauchy, respectively. Our rank test S has very good powers in most of the cases. It is noted that the test S is quite effective in detecting scale shift particularly if the underlying distribution is rather heavy-tailed.

Finally, Table 4.5 shows the power performances of tests when the underlying distribution is asymmetric like exponential. From this table, we find the parametric Hsu's test T is better than the rank tests K and S when the underlying distribution is asymmetric.

Table 4.3. Empirical powers of the tests T, K and S when F is double exponential with density $f(x)=\frac{1}{2}e^{-|x|}$ and the sample size N=30.

τ	Test T	Test K	Test S
3	.081	.098	.099
6	.109	.131	.139
9	.122	.168	.172
12	.192	.220	.239
15	.201	.222	.224
18	.220	.231	.245
21	.221	.240	.243
24	.131	.174	.215
27	.119	.139	.164

Table 4.4. Empirical powers of the tests T, K and S when F is Cauchy with density $f(x)=\frac{1}{\pi(1+x^2)}$ and the sample size N=30.

τ	Test T	Test K	Test S
3	.052	.062	.087
6	.070	.087	.114
9	.061	.136	.142
12	.080	.132	.156
15	.078	.165	.199
18	.080	.139	.150
21	.088	.140	.165
24	.090	.104	.136
27	.080	.076	.097

Table 4.5. Empirical powers of the tests T, K and S when F is exponential with density $f(x)=e^{-x}$ and the sample size N=30.

τ	Test T	Test K	Test S
3	.092	.081	.082
6	.130	.085	.087
9	.173	.128	.140
12	.187	.115	.131
15	.225	.151	.174
18	.226	.118	.133
21	.273	.130	.161
24	.225	.104	.134
27	.184	.085	.082

References

[1] Ali, M.M. and Giaccotto, C.(1982). The identical distribution hypothesis for stock market prices: location- and scale-shift alternatives, *Journal of the American Statistical Association,* Vol. 77, 19-28.

[2] Gastwirth, J.L. and Mahmoud, H.(1986). An efficiency robust nonparametric test for scale change for data from a gamma distribution, *Technometrics,* Vol. 28, 81-84.

[3] Hájek, J. and Šidák, Z.(1967). *Theory of Rank Tests*, Academic Press, New York.

[4] Hsieh, H.K.(1984). Nonparametric tests for scale shift at an unknown time point, *Communications in Statistics-Theory and Methods.*, Vol. 13, 1335-1355.

[5] Hsu, D.A.(1977). Tests for variance shift at an unknown time point, *Applied Statistics,* Vol. 26, 279-284.

[6] Hsu, D.A.(1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis, *Journal of the American Statistical Association,* Vol. 74, 31-40.

[7] Pettitt, A.N.(1979). A non-parametric approach to the change-point problem, *Applied Statistics*, 28, 126-135.

[8] Talwar, P.P. and Gentle, J.E.(1981). Detecting a scale shift in a random sequence at an unknown time point, *Applied Statistics,* Vol. 30, 301-304.