ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 206, 2007

*Dorota Pekasiewicz**

# SEQUENTIAL TESTS WITH POWER EQUAL TO 1 FOR CHOSEN LOCATION PARAMETERS

**Abstract**. We often verify hypotheses about random variable distribution parameters, when the variable distribution is unknown. In these cases we apply nonparametric tests, in particular nonparametric sequential tests.

This paper presents sequential tests for the mean and median. These tests have important property – their power is equal to 1.

**Key words**: sequential test, power of test, mean, median.

## 1. INTRODUCTION

There exist such situations like continuous manufacturing process in which there is no need to take any actions as long as manufactured elements meet particular requirements (null hypothesis). However, it is necessary to take particular action if case is opposite (when the alternative hypothesis is true). Such a situation may take place for example in clinical research concerning newly introduced medicines. As long as the new medicine does not work considerably better than the existing one, there is no reason to launch it. On the other hand, if it works considerably better, then it should be proposed as soon as possible to be launched for patients' good. At the same time, if the new medicine is not better than the existing one, it should not be launched on the market. Therefore, under such circumstances we need to apply a sequential procedure which:

a) would stop very rarely if the null hypothesis is true

b) would stop with probability equal to 1 as soon as possible if the alternative hypothesis is true.

* Ph.D., Chair of Statistical Methods, University of Łódź.

The paper presents examples of sequential tests which have such properties. In order to construct tests of this type we apply the law of iterated logarithm.

## 2. SEQUENTIAL TEST WITH POWER EQUAL TO 1

The verification of statistical hypotheses with the use sequential tests is usually connected with establishing the probabilities of the errors of first and second kind, which are used to define the regions of acceptance of the null hypothesis, the acceptance of alternative and sampling continuation. The test's statistic calculated at every stage of the sequential procedure leads to one of the three decisions.

In most of the problems considered we have different attitude towards the null hypothesis and alternative which leads to modifications and constructing classes of tests for which only one error probability is established (e.g. for significance tests – the probability of the first kind error). If we want the probability of the second kind error to be close to 0, the power of such test will be equal to 1.

Let $X$ be a random variable and $\theta$ its parameter. Let us consider the null hypothesis

$$H_0 : \theta \in \Theta_0 \tag{1}$$

against the alternative

$$H_1 : \theta \in \Theta - \Theta_0 \tag{2}$$

where $\Theta$ is a parameter space and $\Theta_0$ is its subset.

The above hypothesis will be verified with the test whose power function will hold the condition: $M(\theta) = 1$ for $\theta \in \Theta - \Theta_0$.

For the tests of this kind we define the $H_0$ rejection region and what follows the $H_1$ acceptance region (the probability of the acceptance of $H_1$ when its true is equal to1) as well as the sampling continuation region.

Let us we denote by $T_n$ the test statistic, whose value will be determined from an $n$ element simple sample. If $T_n \in J_n$, where $J_n$ is a certain defined interval, we enlarge the sample. If $T_n \notin J_n$ we stop the verification process, accepting $H_1$. In case when $T_n \in J_n$ for $n \to \infty$ we do not rejection $H_0$ (see: R a o 1982).

In these tests the probability of stopping the sample enlargement process after the finite number of steps then the alternative is true is equal to one.

### 3. NONPARAMETRIC SEQUENTIAL TESTS FOR MEAN

Let us assume that $X$ is a random variable with unknown continuous distribution. Let $\mu$ be its expectation. Let us consider the following hypotheses about the value of $\mu$:

$$H_0 : \mu = \mu_0 \tag{3}$$

$$H_1 : \mu \neq \mu_0 \tag{4}$$

where $\mu_0$ is a fixed constant.

Let us assume first that variance $\sigma^2$ of random variable $X$ is known. We start sequential sampling from a $k_0$-element sample ($k_0 \geqslant 3$). In the $k$-th stage of the sampling sequential procedure verifying the formulated hypothesis we will arrive at $n$-element simple sample $X_1, ..., X_n$, where $n = k_0 + k - 1$, from which we calculate the values of the statistics

$$T_n = \frac{\overline{X}_n - \mu_0}{\sigma} \quad \left(T_n \underset{n \to \infty}{\longrightarrow} \frac{\mu - \mu_0}{\sigma}\right) \tag{5}$$

The set $J_n$ is defined in the following way:

$$J_n = \left\{x : |x| \leqslant \sqrt{\frac{2}{n} \ln \ln n}\right\} \quad \text{for} \quad n \geqslant 3 \tag{6}$$

If $T_n \in J_n$, we go on with the sampling. If $T_n \notin J_n$, we reject $H_0$ and accept the alternative $H_1$. If $T_n \in J_n$ for big $n$, we accept $H_0$.

If the alternative hypothesis has the form

$$H_1 : \mu > \mu_0 \tag{7}$$

then

$$J_n = \left\{x : x \leqslant \sqrt{\frac{2}{n} \ln \ln n}\right\} \quad \text{for} \quad n \geqslant 3 \tag{8}$$

In a similar way we define set $J_n$, when alternative has the form $H_1 : \mu < \mu_0$.

In the case of verifying hypothesis about the mean of random variable $X$, when the variance is unknown, at every stage it is estimated with the value of statistic

$$S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}.$$

The test statistic is calculated from the formula

$$T_n = \frac{\overline{X}_n - \mu_0}{S_n} \quad \left(T_n \underset{n\to\infty}{\longrightarrow} \frac{\mu - \mu_0}{\sigma}\right) \tag{9}$$

## 4. NONPARAMETRIC SEQUENTIAL TESTS FOR MEDIAN

Let $\xi_{0.5}$ be a median of the distribution of continuous random variable $X$ with unknown distribution function $F(x)$ i.e. $F(\xi_{0.5}) = 0.5$.

Let us consider the null hypothesis of the form

$$H_0: \xi_{0.5} = \xi_0 \tag{10}$$

against the alternative

$$H_1: \xi_{0.5} \neq \xi_0 \quad (\text{or } \xi_{0.5} > \xi_0) \tag{11}$$

where $\xi_0$ is a fixed number.

Let $X_1, \ldots, X_n$ be an $n$-element random sample at the $k$-th stage of sequential verification of the above hypotheses ($n = k_0 + k - 1$, where $k_0$ denotes the number of elements drawn in the first stage). We define random variables $y_i$, for $i = 1, \ldots, n$, in the following way:

$$Y_i \begin{cases} 0, & \text{if } X_i \leqslant \xi_0 \\ 1, & \text{if } X_i > \xi_0 \end{cases} \tag{12}$$

Random variables $Y_i$, follow the two point distribution ($E(Y_i) = 0.5$, $D^2(Y_i) = 0.5$), when the null hypothesis is true.

Let

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{13}$$

The sequential test statistic verifying hypothesis (10) against (11) has the form

$$T_n = \frac{\overline{Y}_n - 0.5}{\sqrt{0.5(1-0.5)}} = 2\overline{Y}_n - 1 \tag{14}$$

This test may be generalized to verify hypotheses about the quantilies of order $p$, substituting $p$ for 0.5 in formula (14). In this test, the set $J_n$ is defined by the formula (6) (or (8)).

Another test verifying hypotheses about the value of the median is based on ranks (see: Sen 1981). Let us assume that $\xi_{0.5}$ is the median of the distribution of random variable $X$ with unknown distribution function $F(x)$ symmetric about the median. Let $Y = X - \xi_0$. Let us denote the median of the variable $Y$ by $\xi_Y$. We may formulate hypotheses equivalent to the hypotheses (11) and (12) in the following way:

$$\mathrm{H}_0 : \xi_Y = 0 \tag{15}$$

$$\mathrm{H}_1 : \xi_Y \neq 0 \quad (\text{or } \xi_Y > 0) \tag{16}$$

In this case, the test statistic determined from $n$-element simple sample $Y_1, ..., Y_n$ has the form

$$Z_n = \sum_{i=1}^{n} \operatorname{sgn} Y_i a_n(R_{ni}^+) \tag{17}$$

where $R_{ni}^+ = \sum_{j=1}^{n} c(|Y_i| - |Y_j|)$ ($c(u) = 1$ or 0, if $u \geqslant 0$ or $u < 0$ respectively) and $a_n(i) = \phi\left(\frac{i}{n+1}\right)$.

When the null hypothesis is true, statistic $Z_n$ is distributed symetrically about 0, with variance $nA_n^2$, where $A_n^2 = \frac{1}{n}\sum_{i=1}^{n} a_n^2(i)$.

In this case the set $J_n$ is defined in the following way:

$$J_n = \{x : |x| \leqslant A_n^2 \sqrt{2n \ln \ln n}\} \quad \text{for} \quad n \leqslant 3 \tag{18}$$

If the alternative hypothesis has the form

$$\mathrm{H}_1 : \xi_{0.5} = \xi_0 \tag{19}$$

then

$$J_n = \{x : x \leqslant A_n^2\sqrt{2n\, ln\, ln\, n}\} \quad \text{for} \quad n \geqslant 3 \tag{20}$$

## 5. FINAL REMARKS

Sequential tests presented above can be used to verify hypotheses about the mean value or the median of a random variable, if the random variable distribution is unknown.

Application of these tests can cause a problem connected with the completion of sequential enlarging of the sample, if $T_n \in J_n$ for big $n$.

We come across do with such a situation, if the null hypothesis is true. Therefore, it is necessary to determine the sample size $n_0$ up to which we complete extra sampling of elements for the sample and alternatively we assume the null hypothesis.

We should give more considerable amount of thought to this problem and carry out many Monte Carlo analyses.

## REFERENCES

R a o C. R. (1982), *Modele liniowe statystyki matematycznej*, PWN, Warszawa.

S e n P. K. (1981), *Sequential Nonparametric: Invariante Principles and Statistical Inference*, Wiley, New York.

*Dorota Pekasiewicz*

## SEKWENCYJNE TESTY O MOCY RÓWNEJ 1 DLA WYBRANYCH PARAMETRÓW POŁOŻENIA

Nieparametryczne testy sekwencyjne mogą służyć do weryfikacji hipotez o wartościach parametrów zmiennej losowej, takich jak wartość oczekiwana i mediana, w przypadku gdy nie znamy klasy rozkładu badanej zmiennej.

W pracy przedstawione zostały przykłady testów sekwencyjnych, których moc, przy dużej liczebności próby, jest równa 1. Testy tego typu mogą znaleźć zastosowanie zarówno w kontroli jakości produkcji, jak i badaniach medycznych.