ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 196, 2006

*Dorota Pekasiewicz**

SEQUENTIAL TESTS FOR TRUNCATED DISTRIBUTION PARAMETERS

Abstract. Sequential probability ratio tests can be used to verify hypotheses about truncated distribution parameters when error probabilities of the first and the second type are set. For parameters of normal and exponential distributions truncated on both sides and on the left side, statistics of sequential probability ratio test were determined. Since the sample size in these tests is the random variable, formulas for expected values of sample size for considered tests were also defined.

Key words: sequential test, truncated distribution, expected value.

1. INTRODUCTION

Random variables, which are model probabilistically real phenomena, do not always assume values from the same set as observed variables. For example, in economic research it is assumed that considered variables have normal distribution although non-negative numbers make the set of values. We can do that if we omit sets of values for which the probability that the variable will take these values is very small, for example for random variable with a normal distribution from the rule of three sigma it results that $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.9973$, hence $P(-\infty < X < \mu - 3\sigma) + P(\mu + 3\sigma < X < \infty) \approx 0.0027$. That is why the limitation of the range of variability X from the infinite range to finite range $(\mu - 3\sigma, \mu + 3)$ is acceptable.

However, we do not always have the premises to make such limitation. At this stage we have to consider a distribution truncated on both sides, on the right side or on the left side.

The verification of the hypothesis about truncated parameters is possible when we apply sequential probability ratio tests. For normal distributions

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

parameters and exponential distributions truncated on the left side and on both sides, forms of testing statistics were determined. What is more, procedures of graphical verification of formulated hypotheses in all considered cases were introduced. Since the sample size is the random variable in sequential tests, formulas determining expected values of the sample size for all analyzed sequential tests were derived.

2. TRUNCATED DISTRIBUTIONS

Let X be a random variable with the distribution P_X taking values from the range (a, b) and let $\langle c, d\rangle \subset (a, b)$, while (a, b) is the finite or infinite distribution. We will be considering distributions truncated on both sides and on the left side but at the same time we will assume that $c, d \geqslant 0$. The assumption about positive values is connected with the usage of these distributions in investigations in which considered variable takes positive values.

Let us introduce the following definitions of truncated distributions (Domański, Pruska 2000).

Definition 1. The distribution of random variable $Y = X | c \leqslant X \leqslant d$ that is the variable whose values' set is made of values of the variable X belonging to the range $\langle c, d\rangle$, is called a truncated distribution P_X or a distribution P_X truncated on both sides.

Definition 2. The distribution of the random variable $Y = X | c \leqslant X < b$ that is the random variable whose values' set is made of values of variable X belonging to the range $\langle c, b)$, is called a distribution P_X truncated on the left side.

The form of the density of continuous distributions truncated on the left side and both sides is defined by the following theorem (Domański, Pruska 2000):

Theorem. Let the random variable X have the continuous distribution with the density f_X and take values from the range (a, b), where $a, b \in R$.

1. If $\langle c, d\rangle \subset (a, b)$, where $c, d \in R$, the density of the random variable $Y = X | c \leqslant X \leqslant d$ is expressed by a formula:

$$f_Y(y) = \begin{cases} \dfrac{f_X(y)}{F_X(d) - F_X(c)} & \text{for} \quad y \in \langle c, d\rangle, \\ 0 & \text{for} \quad y \notin \langle c, d\rangle. \end{cases} \tag{1}$$

2. If $\langle c, b) \subset (a, b)$, where $c \in R$, the density of the random variable $Y = X | c \leqslant X < b$ is expressed by a formula:

$$f_Y(y) = \begin{cases} \dfrac{f_X(y)}{1 - F_X(c)} & \text{for} \quad y \in \langle c, b), \\ 0 & \text{for} \quad y \notin \langle c, b). \end{cases} \tag{2}$$

Let us assume that the random variable X has the normal distribution with the expected value μ and the standard deviation σ. The random variable $Y = X | c \leqslant X \leqslant d$ has the normal distribution truncated on both sides with the density function:

$$f_Y(y) = \frac{\exp\left(-\dfrac{(y-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}\left(\Phi\left(\dfrac{d-\mu}{\sigma}\right) - \Phi\left(\dfrac{c-\mu}{\sigma}\right)\right)} \quad \text{for} \quad y \in \langle c, d \rangle. \tag{3}$$

The random variable $Y = X | X \geqslant c$ has a normal distribution truncated on the left side with the density function:

$$f_Y(y) = \frac{\exp\left(-\dfrac{(y-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}\left(1 - \Phi\left(\dfrac{c-\mu}{\sigma}\right)\right)} \quad \text{for} \quad y \geqslant c. \tag{4}$$

If X is the random variable with the exponential distribution and the expected value λ and this variable takes values only from the range $\langle c, d \rangle$, $c, d \geqslant 0$, the density function of the random variable $Y = X | c \leqslant X \leqslant d$ will be expressed by a formula:

$$f_Y(y) = \frac{\dfrac{1}{\lambda}\exp\left(-\dfrac{y}{\lambda}\right)}{\exp\left(-\dfrac{c}{\lambda}\right) - \exp\left(-\dfrac{d}{\lambda}\right)} \tag{5}$$

If we consider the random variable $Y = X | X \geqslant c$, where $c \geqslant 0$ its density function will be in the form of:

$$f_Y(y) = \frac{1}{\lambda}\exp\left(\frac{c-y}{\lambda}\right). \tag{6}$$

3. PARAMETRIC SEQUENTIAL TESTS FOR TRUNCATED NORMAL AND EXPONENTIAL DISTRIBUTIONS

Let Y be the random variable with the density function $f_Y(x)$, and μ the reliable parameter of its distribution. Let us formulate the null hypothesis:

$$H_0: \mu = \mu_0 \tag{7}$$

and alternative hypothesis:

$$H_1: \mu = \mu_1, \quad \text{where} \quad \mu_1 > \mu_0. \tag{8}$$

We can verify the above hypotheses applying the sequential probability ratio test.

The statistics of the sequential test has the following form (Marek, Noworol 1982):

$$I_n = \ln \frac{f_Y(Y_1, Y_2, ..., Y_n; \mu_1)}{f_Y(Y_1, Y_2, ..., Y_n; \mu_0)}, \tag{9}$$

where $f_Y(Y_1, ..., Y_n; \mu)$ is the total density function of the random sample $Y_1, ..., Y_n$.

We compare the statistics I_n calculated for n-element sample to the following values:

$$A = \ln \frac{1-\beta}{\alpha} \quad \text{and} \quad B = \ln \frac{\beta}{1-\alpha},$$

determined on the ground of α and β – fixed error probabilities of the first and the second type.

If $I_n < B$, we accept the hypothesis H_0. If $I_n > A$ we reject the hypothesis H_0 and accept the hypothesis H_1. If $B \leqslant I_n \leqslant A$ we sample the next element.

For random variable $Y = X | c \leqslant X \leqslant d$, if X is the variable with the distribution $N(\mu, \sigma)$ the statistics of the sequential test, which verifies hypotheses about the parameter μ, is expressed by the formula:

$$I_n = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^{n} y_i + \frac{(\mu_0^2 - \mu_1^2)n}{2\sigma^2} + n \ln \frac{\Phi((d-\mu_0)/\sigma) - \Phi((c-\mu_0)/\sigma)}{\Phi((d-\mu_1)/\sigma) - \Phi((c-\mu_1)/\sigma)}. \tag{10}$$

For the random variable $Y = X | c \leqslant X < b$:

$$I_n = \frac{\mu - \mu_0}{\sigma^2} \sum_{i=1}^{n} y_i + \frac{(\mu_0^2 - \mu_1^2)n}{2\sigma^2} + n \ln \frac{1 - \Phi((c-\mu_0)/\sigma)}{1 - \Phi((c-\mu_1)/\sigma)}. \tag{11}$$

Another procedure of sequential verification of hypotheses is the graphic procedure consisting in the introduction of acceptance regions of hypotheses H_0, H_1 and the region of continuation of sampling on the plane OXY.

Let us introduce the following denotations for sequential test verifying hypotheses about the value of the parameter μ if the random variable Y has normal distribution truncated on both sides:

$$s_n = \sum_{i=1}^{n} y_i, \tag{12}$$

$$a = \frac{\mu_0 + \mu_1}{2} - \frac{\sigma^2}{\mu_1 - \mu_0} \ln \frac{\Phi\left(\frac{d-\mu_0}{\sigma}\right) - \Phi\left(\frac{c-\mu_0}{\sigma}\right)}{\Phi\left(\frac{d-\mu_1}{\sigma}\right) - \Phi\left(\frac{c-\mu_1}{\sigma}\right)}, \tag{13}$$

$$b_1 = \frac{\sigma^2 A}{\mu_1 - \mu_0}, \tag{14}$$

$$b_2 = \frac{\sigma^2 B}{\mu_1 - \mu_0}. \tag{15}$$

If the inequality $s_n < an + b_2$, on the n-th stage is satisfied we accept the hypothesis H_0. If the inequality $s_n > an + b_1$, is satisfied we accept the hypothesis H_1. Otherwise, we enlarge the sample by additional sampling of at least one element. Graphically determining parallels $s = an + b_1$ and $s = an + b_2$, we can mark the acceptance regions of considered hypotheses on the plane (Figure 1).

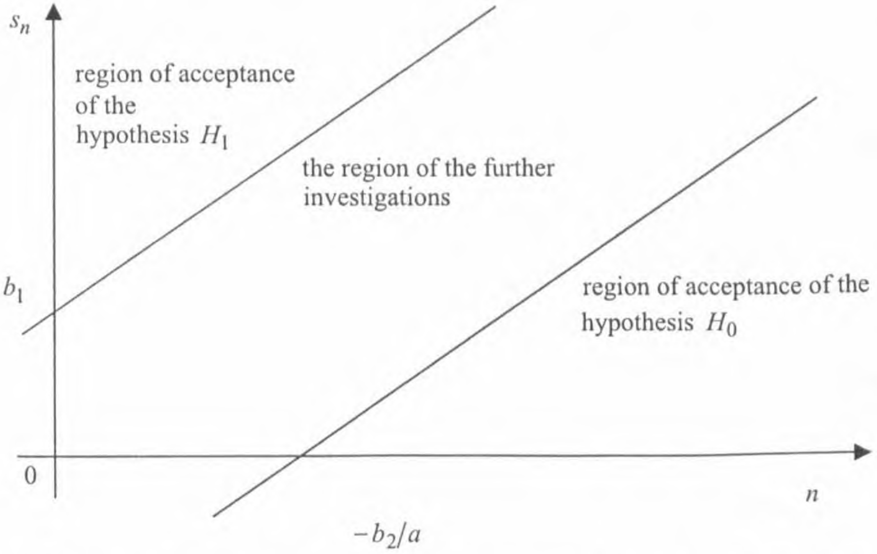

Fig. 1. Graphical representation of the regions of acceptance of the null hypothesis, alternative hypothesis and the continuation of the sampling

For sequential test verifying hypotheses about the value of the parameter μ of the random variable with normal distribution truncated on the left side only the directional factor a will have a different form.

$$a = \frac{\mu_0 + \mu_1}{2} - \frac{\sigma^2}{\mu_1 - \mu_0} \ln \frac{1 - \Phi\left(\frac{c - \mu_0}{\sigma}\right)}{1 - \Phi\left(\frac{c - \mu_1}{\sigma}\right)}. \tag{16}$$

In sequential tests the sample size is the random variable. That is why we can determine its expected value from the formula (Marek, Noworol, 1982):

$$E_\mu(N) = \begin{cases} \dfrac{(1-\alpha)B + \alpha A}{E_{\mu_0}(Z)} & \text{for} \quad \mu = \mu_0, \\ \dfrac{\beta B + (1-\beta)A}{E_{\mu_1}(Z)} & \text{for} \quad \mu = \mu_1, \end{cases} \tag{17}$$

where $E_\mu(Z)$ is conditional expected value of the random variable $Z = \ln \frac{f(Y, \mu_1)}{f(Y, \mu_0)}$.

If the variable Y has the normal distribution truncated on both sides the variable Z has the following form:

$$Z = \frac{\mu_1 - \mu_0}{\sigma^2} Y + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \ln \frac{\Phi\left(\frac{d - \mu_0}{\sigma}\right) - \Phi\left(\frac{c - \mu_0}{\sigma}\right)}{\Phi\left(\frac{d - \mu_1}{\sigma}\right) - \Phi\left(\frac{c - \mu_1}{\sigma}\right)}. \tag{18}$$

Since

$$E_\mu(Y) = \mu + \frac{\sigma\left(\exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right) - \exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right)\right)}{\sqrt{2\pi}\left(\Phi\left(\frac{d-\mu}{\sigma}\right) - \Phi\left(\frac{c-\mu}{\sigma}\right)\right)} \tag{19}$$

hence

$$E_\mu(Z) = \frac{\mu_1 - \mu_0}{\sigma^2} \cdot E_\mu(Y) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \ln \frac{\Phi\left(\frac{d - \mu_0}{\sigma}\right) - \Phi\left(\frac{c - \mu_0}{\sigma}\right)}{\Phi\left(\frac{d - \mu_1}{\sigma}\right) - \Phi\left(\frac{c - \mu_1}{\sigma}\right)}. \tag{20}$$

If the random variable Y has the normal distribution truncated on the left side, then:

$$E_\mu(Y) = \mu + \frac{\sigma \exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}\left(1 - \Phi\left(\frac{c-\mu}{\sigma}\right)\right)} \tag{21}$$

hence

$$E_\mu(Z) = \frac{\mu_1 - \mu_0}{\sigma^2} \cdot E_\mu(Y) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \ln \frac{1 - \Phi\left(\frac{c-\mu_0}{\sigma}\right)}{1 - \Phi\left(\frac{c-\mu_1}{\sigma}\right)}. \tag{22}$$

Let us consider another case of the verification, using sequential test, of hypotheses about the value of the parameter λ of truncated exponential distributions.

Let

$$H_0 : \lambda = \lambda_0 \tag{23}$$

$$H_1 : \lambda = \lambda_1, \quad \text{where} \quad \lambda_1 > \lambda_0. \tag{24}$$

Let us assume that the variable Y has the exponential distribution truncated on both sides which means that the density function has the form of (5). The statistics of the sequential probability ratio test calculated on n-th stage of the sequential procedure is expressed by the formula:

$$I_n = n \ln \frac{\lambda_0\left(\exp\left(-\frac{c}{\lambda_0}\right) - \exp\left(-\frac{d}{\lambda_0}\right)\right)}{\lambda_1\left(\exp\left(-\frac{c}{\lambda_1}\right) - \exp\left(-\frac{d}{\lambda_1}\right)\right)} + \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)\sum_{i=1}^{n} y_i. \tag{25}$$

If the random variable Y has the distribution truncated on the left side the test's statistics has the following form:

$$I_n = n\left(\ln\frac{\lambda_0}{\lambda_1} + \frac{c}{\lambda_1} - \frac{c}{\lambda_0}\right) + \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)\sum_{i=1}^{n} y_i. \tag{26}$$

We introduce the following denotations for the graphical procedure of the verification of the hypothesis (23) against (24):

$$s_n = \sum_{i=1}^{n} y_i, \tag{27}$$

$$b_1 = \frac{\lambda_1 \lambda_0}{\lambda - \lambda_0} A, \tag{28}$$

$$b_2 = \frac{\lambda_1 \lambda_0}{\lambda - \lambda_0} B, \tag{29}$$

and

$$a = \begin{cases} \dfrac{\lambda_1 \lambda_0}{\lambda_1 - \lambda_0} \ln \dfrac{\lambda_0 \left(\exp\left(-\frac{c}{\lambda_0} \right) - \exp\left(-\frac{d}{\lambda_0} \right) \right)}{\lambda_1 \left(\exp\left(-\frac{c}{\lambda_1} \right) - \exp\left(-\frac{d}{\lambda_1} \right) \right)}, & \text{if } Y \text{ has the exponential distribution truncated on both sides,} \\ \dfrac{\lambda_1 \lambda_0}{\lambda_1 - \lambda_0} \ln \left(\dfrac{\lambda_0}{\lambda_1} + \dfrac{c}{\lambda_1} - \dfrac{c}{\lambda_0} \right), & \text{if } Y \text{ has the exponential distribution truncated on the left side.} \end{cases} \tag{30}$$

The regions of acceptance of the null and alternative hypotheses and the region of the necessity of enlarging the sample look analogously to Fig. 1.

The expected value of the sample N size in this sequential test is expressed by a formula (17).

For the random variable Y with the distribution truncated on both sides the random variable Z has the following form:

$$Z = Y\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1} \right) + \ln \frac{\lambda_0}{\lambda_1} + \ln\left(\exp\left(-\frac{c}{\lambda_0} \right) - \exp\left(\frac{d}{\lambda_0} \right) \right)$$
$$- \ln\left(\exp\left(-\frac{c}{\lambda_1} \right) - \exp\left(-\frac{d}{\lambda_1} \right) \right), \tag{31}$$

hence

$$E_\lambda Z = E_\lambda(Y)\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1} \right) + \ln \lambda_0 \left(\exp\left(-\frac{c}{\lambda_0} \right) - \exp\left(\frac{d}{\lambda_0} \right) \right)$$
$$- \ln \lambda_1 \left(\exp\left(-\frac{c}{\lambda_0} \right) - \exp\left(-\frac{d}{\lambda_0} \right) \right), \tag{32}$$

where:

$$E_\lambda(Y) = \lambda + \frac{c \exp\left(-\frac{c}{\lambda} \right) - d \exp\left(-\frac{d}{\lambda} \right)}{\exp\left(-\frac{c}{\lambda} \right) - \exp\left(-\frac{d}{\lambda} \right)}.$$

For the random variable Y with the exponential distribution truncated on the left side the random variable Z has the following form:

$$Z = Y\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right) + \ln\frac{\lambda_0}{\lambda_1} + \frac{c(\lambda_0 - \lambda_1)}{\lambda_1\lambda_0}. \tag{33}$$

Since $E_\lambda(Y) = \lambda + c$, then

$$E_\lambda(Z) = E_\lambda(Y)\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right) + \ln\frac{\lambda_0}{\lambda_1} + \frac{c(\lambda_0 - \lambda_1)}{\lambda_1\lambda_0} = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right)(\lambda + c) + \ln\frac{\lambda_0}{\lambda_1} + \frac{c(\lambda_0 - \lambda_1)}{\lambda_1\lambda_0}. \tag{34}$$

4. FINAL REMARKS

Sequential probability ratio tests can be used to verify hypotheses about truncated distribution parameters. The test's statistics is modified depending on the form of truncated distributions. The paper exemplifies sequential tests for parameters of normal and exponential distribution truncated on both sides and on the left side. Testing statistics (analytical method of hypotheses verification) and acceptance regions of particular hypotheses and the continuation of sampling (graphical method) were determined for considered tests. Analogously, we can introduce formulas and graphically determine acceptance regions of hypotheses and additional elements' sampling for sequential probability ratio tests for parameters (among others of the mean) of other random variables distributions which can be put into practice.

In sequential tests the sample size is the random variable, that is why the formula for expected values of the sample size which is necessary to take a decision about the acceptance of one of the verified hypotheses with the errors of the first and the second type were determined.

REFERENCES

Domański Cz., Pruska K. (2000), *Nieklasyczne metody statystyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Marek T., Noworol Cz. (1982), *Analiza sekwencyjna w badaniach empirycznych*, Państwowe Wydawnictwo Naukowe, Warszawa.

Dorota Pekasiewicz

TESTY SEKWENCYJNE DLA PARAMETRÓW ROZKŁADÓW UCIĘTYCH

(Streszczenie)

Ilorazowe testy sekwencyjne można stosować do weryfikacji hipotez o parametrach rozkładów uciętych, przy ustalonych prawdopodobieństwach błędów I-go i II-go rodzaju. Dla parametrów obustronnie i lewostronnie uciętych rozkładów normalnych i wykładniczych zostały wyznaczone statystyki ilorazowych testów sekwencyjnych. Ponieważ liczebność próby w tych testach jest zmienną losową, określone zostały również wzory na wartości oczekiwane liczebności prób dla rozważanych testów.