Journal of Physics: Conference Series **666** (2016) 012034 doi:10.1088/1742-6596/666/1/012034

Testing composite parametrical hypotheses without applying the reduction

V Vlasov[1], S Vlasova[2] and A Tolokonsky[1,3]
[1] National Research Nuclear University MEPhI (MEPhI), Kashirskoe shosse 31, Moscow, 115409, Russian Federation
[2] All-Russian Research Institute for Nuclear Power Plants, Ferganskaya st., 25, Moscow, 109507, Russian Federation

E-mail: AOTolokonskij@mephi.ru

Abstract. Usually when the parametrical hypotheses are being tested the Wald reduction from composite hypothesis to the simple one is used. However, in order to apply the reduction method it is needed to know the distribution law of unknown parameter. Practically such a law cannot be determined precisely using the experimental data. First of all, it requires long supervision over the controlled stationary process to provide an inalterability of probability characteristics of the process. In practice the modifications in technological process could be made, and therefore the probability characteristics of the process can also change. Using the example of exponential distribution the algorithm of testing composite parametrical hypothesis about the distribution parameter which does not exceed the declared threshold value without use of reduction is considered in article. Such approach is based on the fact that the partition boundary of the sample space depends monotonously on unknown value of the interest parameter.

1. Introduction

Usually when the parametrical hypotheses are being tested the Wald reduction from composite hypothesis to the simple one is used. However, in order to apply the reduction method it is needed to know the distribution law of unknown parameter. Practically such a law cannot be determined precisely using the experimental data. First of all, it requires long supervision over the controlled stationary process to provide an inalterability of probability characteristics of the process. In practice the modifications in technological process could be made, and therefore the probability characteristics of the process can also change.

2. Experimental setup

Let us assume that the threshold value of parameter λ is λ_0. On the basis of the sample $\xi_1, \xi_2, ..., \xi_n$ of the population which is described by the following probability density function:

$p(x) = \begin{cases} \lambda \exp(-\lambda x), x \geq 0 \\ 0, x < 0 \end{cases}$, it is needed to test the statistical hypothesis H_1, which claims the inequality $\lambda < \lambda_0$ is true (this inequality defines the composite parametrical hypothesis H_1). In order

[3] To whom any correspondence should be addressed.

to solve the task we will consider the alternative composite parametrical hypothesis H_2, that states the other inequality $\lambda > \lambda_0$ is true.

Such a task can appear in a wide range of application areas. For example, in the analysis of scientific experiments or in the analysis of unsuccessful changes in technological processes resulting in deterioration of production etc.

At first we will consider the order of constructing the critical region basing on Pearson's theorem in order to test two simple competing hypotheses H_1, H_2. The critical region for a hypothesis H_1 (when this hypothesis is rejected) is defined in the sample space by the inequality:

$$p(x_1, x_2, ..., x_n, H_2) > Cp(x_1, x_2, ..., x_n, H_1), \tag{1}$$

where: $p(x_1, x_2, ..., x_n, H_2)$ - the distribution density of a sample set in the conditions of the hypothesis H_2, $p(x_1, x_2, ..., x_n, H_1)$ - the distribution density of a sample set in the conditions of the hypothesis H_1. A value of a constant C could be chosen both according to the values of probabilities of type I and II errors [1,2], or according to the values of weights (costs) of type I and II errors on the basis of variation principles [3].

It is important to note that when the decision between two parametrical hypotheses H_1 (when $\lambda = \lambda_1$) and H_2 (when $\lambda = \lambda_2$) is being made the values λ_1 and λ_2 should be known. In the considering case the hypothesis H_1 is completely defined: $\lambda = \lambda_0$. In the conditions of the opposite hypothesis H_2, that states the truthfulness of inequality $\lambda > \lambda_0$, the value of parameter λ remains unknown. Therefore the simple parametrical hypothesis competes with the composite one.

We will further use the property of monotonous dependence of the critical region's boundary on the value of unknown parameter λ in the conditions of the composite parametrical hypothesis H_2 which include the set of simple hypotheses. The property of monotonous dependence is considered in the work [4].

Let us assume that in the conditions of the exponential distribution when the hypothesis H_1 is true the value of parameter is known: $\lambda_1 = \lambda_0$. Therefore for any value of the parameter λ in the conditions of the hypothesis H_2 the expression (1) can be formulated in the following way:

$$\lambda^n \exp(-\lambda S) > C\lambda_0^n \exp(-\lambda_0 S), \tag{2}$$

where $S = \sum_{i=1}^{n} x_i$.

The hypothesis H_1 (when $\lambda = \lambda_0$) will be rejected if the inequality (2) is true. If for any value of parameter λ that exceeds the value λ_0 the opposite inequality is true:

$$\lambda^n \exp(-\lambda S) < C\lambda_0^n \exp(-\lambda_0 S), \tag{3}$$

then any simple hypothesis that is included in composite hypothesis H_2 will be rejected. Therefore the composite hypothesis H_2 should be rejected and we come to the following conclusion: when the last inequality is true the value of unknown parameter λ does not exceed the value λ_0. Further we will consider the simplest case when $C = 1$. In this case the inequality (3) can be expressed by the formula:

$S > \dfrac{n \ln \gamma}{\lambda_0 (\gamma - 1)}$, where $\gamma = \dfrac{\lambda}{\lambda_0}$.

Using the concept of sufficient statistics for exponential distribution: $S = \sum_{i=1}^{n} \xi_i$, where ξ_i − sample values, the received formula can be written in a more compact way.

The boundary value S_g for the variable S , which splits the sample space into two parts, could be found from the expression (2) when the equation is reached:

$$S_g = \frac{n}{\lambda_0}\frac{\ln\gamma}{\gamma-1}, \tag{4}$$

where $\gamma = \dfrac{\lambda}{\lambda_0}$, $\lambda > \lambda_0$.

Therefore the criterion according to which it is necessary to reject any simple parametrical hypothesis when $\lambda > \lambda_0$, could be simply expressed by the inequality:

$$S > S_g. \tag{5}$$

This directly means that the parameter λ does not exceed the value λ_0. The simple analysis shows that the variable S_g decreases when the parameter γ in its turn increases. Therefore if for some value γ^* the criterion (5) is true, it is also be true for every values of γ, for which $\gamma > \gamma^*$. If $C = 1$ and γ tends to the value 1 the expression (4) is uncertain. This problem can be easily overcome by means of application of the wide known Lopital rule [6], and the variable S_g could be found from the expression: $S_g^* = \dfrac{n}{\lambda_0}$. Actually, when $\gamma \to 1$ the limit will be equal to the expression:

$$\lim \frac{n}{\lambda_0}\frac{\ln\gamma}{\gamma-1} = \frac{n}{\lambda_0}\lim\frac{\frac{1}{\gamma}}{1} = \frac{n}{\lambda_0}.$$

Therefore the criterion that the parameter λ does not exceed the value λ_0 in this case has a simple expression:

$$S \le \frac{n}{\lambda_0}. \tag{6}$$

The condition $C = 1$ corresponds to the case when the weights (costs) of type I and II errors are equal [7].

The dependence of variable S_g on parameter λ for the case when $C = 1$ is presented in figure 1.

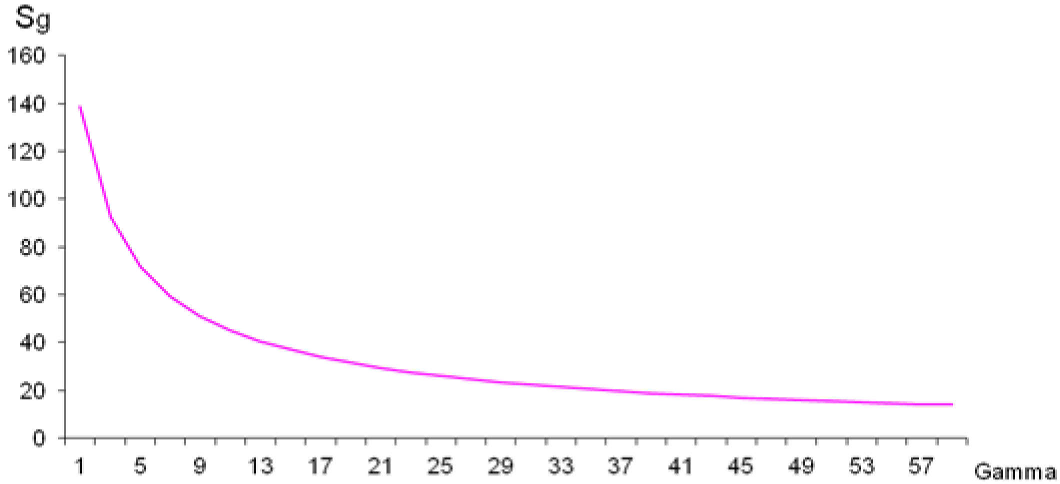

Figure 1. The dependence of variable S_g on parameter λ for the case when $C = 1$.

The situation when $S > \frac{n}{\lambda_0}$ also needs detailed consideration. In this case it is necessary to consider the parametrical hypotheses which claims that $\lambda < \lambda_0$. In the work [4] it was shown that when λ decreases (that is when $\lambda < \lambda_0$) the variable S_g increases, and if the parameter λ tends to zero the variable S_g tends monotonously to infinity (in this work the following mistake was made: the incorrect sign < was used instead of correct sign > when the hypotheses were tested, and therefore this led to the opposite conclusion). When the values of the parameter λ are small this hypothesis should be rejected (when the condition $S < S_g$ is true, so it cannot be accepted). It could be simply explained by the following illustration: let us assume that the value of the parameter λ is very small, then the sample values have to be very big and the hypothesis claiming that the values of parameter λ are small will be true if the sum S of sampled values is rather great. We will show it mathematically. Let us denote H as the hypothesis that states $\lambda = \lambda_1$, where $\lambda_1 < \lambda_0$, and by symbol $\overline{H}$ the hypothesis that states $\lambda = \lambda_0$.

According to condition (1) the hypothesis H is rejected in the case when the following inequality is true:

$$\lambda_0^n \exp -(\lambda_0 S) > C\lambda_1^n \exp -(\lambda_1 S).$$

Applying simple algebraic transformations to the last expression lead to the result:

$$S < \frac{n \ln \frac{\lambda_0}{\lambda_1} - \ln C}{\lambda_0 - \lambda_1} = S_g$$

The main result about the monotonous dependence of the critical region's boundary on the value of the parameter λ and possibility of its applying for testing composite parametrical hypotheses about the value of the parameter λ remain true. Therefore any simple parametrical hypothesis should be rejected when $\lambda = \lambda_1$ is quite small in case of truthfulness of the inequality $S < S_g$. It is clear, that if the hypothesis λ^* is rejected, then any of the simple hypotheses claiming that $\lambda < \lambda^*$ should be also rejected. Obviously the acceptance of the hypothesis about small values of the parameter λ requires large sample size and considerable time for supervision.

The dependence of the boundary S_g on the parameter λ in case when $C=1$ and the value of the parameter λ is quite small is shown in figure 2.

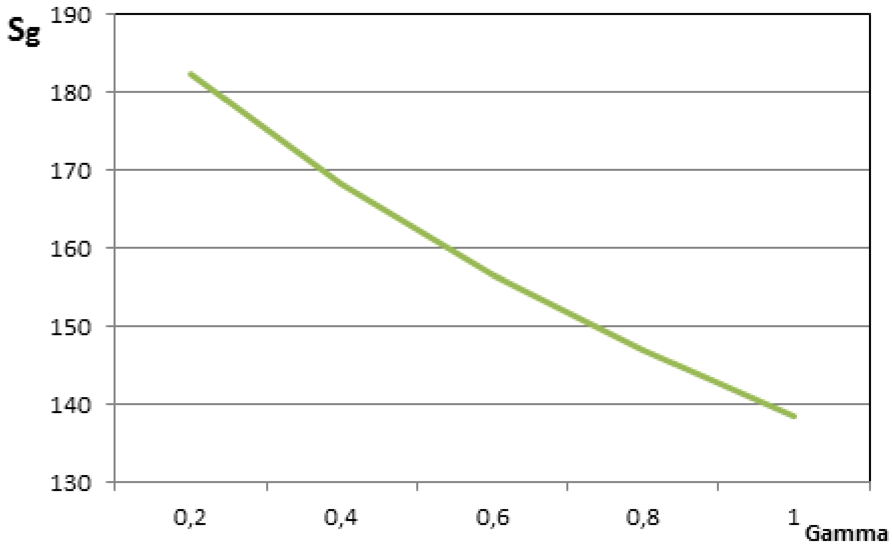

Figure 2. The dependence of the boundary S_g on the parameter λ in case when $C=1$ and the value of the parameter λ is quite small.

3. Discussion

Therefore the whole sample space is split into two regions and in every region the certain hypotheses are accepted. It should be noted that notwithstanding the acceptance of any of composite hypotheses the true value of the parameter λ remains unknown. For research purposes in order to avoid difficulties of the analysis when C is not equal to 1, it is possible to consider the auxiliary parametrical hypothesis which claims that the value of parameter λ is equal to the value λ_0', which is close to λ_0. This question deserves further more detailed discussion.

References

[1] Wilks S S 1961 A combinatorical tests for the problem of two samples from continuous distribution *Proc. Fourth Berkeley Symp. On Math. Stat. and Prob.* vol 1 (University of California Press.)
[2] Nikitin A I, Leipunsky I O and Nikitina T F 2010 *Prikladnaya Fizika* **2** 15-22 (in Russian)
[3] Rao C R 1968 *Linear Statistical Methods and its Applications* (M: Science) (in Russian)
[4] Vlasov V A 2012 *Estimates, Decisions, Risks* (M., Binom) (in Russian)
[5] Vlasov V A, Vlasova S V and Tolokonsky A O 2014 Estimates when parametrical hypotheses are accepted and comparison them with maximum likelihood estimates *Life Science Journal* **11**.
[6] Seber G A F 1977. *Linear Regression Analysis, Wiley Series in Probability and Mathematical Statistics* (New York-London-Sydney-Toronto) pp:456
[7] Fikhtengol'ts G M 2003 *Course of Differential and Integral Calculus* (M., Fizmathlit) (in Russian).
[8] Vlasov V A and Vlasova S V 2007 Software Support for Testing Composite Hypotheses about the Parameters of Exponential Distribution taking into Account the Weighs of Errors *Instruments and Systems: Monitoring, Control, and Diagnostics* **8** 25-27 (in Russian)