

Comparative analysis of methods for calculating the mode of a stationary random process

Vladimir Marchuk, Igor Shrayfel, Gleb Hripkov, Ivan Nikishin

Don State Technical University, Institute of Service and Business (Branch) DSTU in Shakhty, Russia

Abstract. The paper considers methods for calculating the mode of a stationary random process when solving the problem of processing measurement results under conditions of a priori uncertainty. The results of the conducted studies allowed us to conclude that the most effective method for calculating the mode for a given sample is the proposed method, which allows to increase the accuracy of its calculation by at least 8 times, compared to methods based on the construction of histograms. It should be noted that the proposed method allows to provide an estimate with an error of at least 5% for samples with a number of measurements of about 5 value.

1 Introduction

Currently, when processing random stationary processes, in some cases, it is necessary to calculate the mode of a random process. One example of such a need is the use of the method of multiplying estimates (RAZOTS) when processing measurement results under conditions of a priori uncertainty [1-3]. As follows from the analysis of the research results presented in the paper [4, 5], the distribution of the error in each of the cross sections of the multiplication set of the original sample is assumed according to the normal law. However, a more detailed analysis of the research results showed that the error distribution at the beginning and end of the sample differs significantly from the normal one and approaches the Rayleigh distribution. It is known that under the Rayleigh distribution law, the values of the mode and the mean are already significantly different compared to the normal law, where they are equal. This difference is the source of the increase in the error of processing the measurement results, which can be reduced by calculating not the average, but the mode in each of the cross-sections of processing the measurement results. Keeping in mind that a mode is understood as the value of a random variable, the probability of which is maximal.

Thus, replacing the mean with the mode will significantly increase the accuracy of processing the measurement results. This task is very urgent and practically significant.

2 Methods and materials

Suppose that we are given a sample of a random stationary process distributed according to Rayleigh's law

$$w(y) = \frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}}; y \geq 0. \tag{1}$$

The algorithm for modeling random numbers distributed according to Rayleigh's law, according to the work [6] it can be described by the following expression

$$y = \sigma\sqrt{-2\ln x}, \tag{2}$$

where x – a random variable evenly distributed in the interval $[0,1]$.

Various methods can be used to estimate the mode value, including the value calculated from the analytical expression of the Rayleigh distribution law, which is equal to the derivative of the expression (1), and it is equal to zero. The calculation shows that the mode value is equal to the value σ . The second well – known method [7-9] for calculating the mode by estimating the differential density of the distribution is histograms. And here we can distinguish three approaches for calculating the mode. 1 – build a histogram and estimate the range of values that have the maximum value, take the extreme values of the interval, average and consider this value as a mode. 2 – determine the interval that has the maximum value, select all the values of the random process that fall into this interval and build a new histogram based on these values. determine the maximum interval, average it by the extreme values, and consider it a mode. It is known from the literature [10-20] that in order to find the mode, you need to find the modal interval and use the following formula:

$$M_0 = x_0 + \frac{n_M - n_{M-1}}{(n_M - n_{M-1}) + (n_M - n_{M+1})} \cdot h, \tag{3}$$

where x_0 – lower limit of the modal interval; h – length of the modal interval; n_M – frequency of the modal interval; n_{M-1} – frequency of the previous interval; n_{M+1} – frequency of the next interval.

As follows from the above, the modal interval method is also based on the construction of a histogram, which significantly complicates the calculations, especially when the sample size is limited. In this regard, the following method of determining the mode from the sample itself without constructing a histogram, proposed by the authors of this work, is of considerable interest.

Let's assume that for a random variable X , distributed according to Rayleigh's law, the implementation of the sample was obtained x_1, x_2, \dots, x_n . We find an estimate of the maximum likelihood of the parameter σ of this distribution.

It is known that, its density is different from zero only on the interval $(0, +\infty)$, with

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \text{ при } x > 0. \tag{4}$$

Therefore, the likelihood function in this case has the form

$$L(\sigma) = \prod_{k=1}^n \frac{x_k}{\sigma^2} e^{-\frac{x_k^2}{2\sigma^2}}, \tag{5}$$

and its logarithm is the form

$$\ln L(\sigma) = \sum_{k=1}^n (\ln x_k - 2\ln \sigma - \frac{x_k^2}{2\sigma^2}) = A - 2n \ln \sigma - B\sigma^{-2};$$

the following notation is used here $A = \sum_{k=1}^n \ln x_k$, $B = 0,5 \sum_{k=1}^n x_k^2$.

Find the point of the highest value of the logarithmic likelihood function. Equating its derivative to zero

$$(\ln L(\sigma))' = (A)' - \frac{2n}{\sigma} + 2B\sigma^{-3} = \frac{2B}{\sigma^3} - \frac{2n}{\sigma} = \frac{2(B-n\sigma^2)}{\sigma^3}, \quad (6)$$

we get the equation $B - n\sigma^2 = 0$. From its two roots $\sigma = \pm\sqrt{\frac{B}{n}}$ we choose the positive one $\sigma^* = \sqrt{\frac{B}{n}}$. Since the function is positive $(\ln L(\sigma))'$ on the interval $(0; \sigma^*)$ and is negativ at $\sigma > \sigma^*$, it can be argued that σ^* represents the maximum likelihood estimate of the parameter σ .

Note that the Rayleigh distribution mode is equal to σ ; this is easily verified by equating $f'(x)$ to zero. Thus, the value of

$$\sigma^* = \sqrt{\frac{1}{2n} \sum_{k=1}^n x_k^2} \quad (7)$$

is also an estimate of the maximum likelihood of its mode.

So, the work considers four practical methods for calculating the mode of a random process that has a Rayleigh distribution law. We will conduct research on a comparative analysis of their effectiveness in order to obtain recommendations for their practical use. During the research, the mode estimation was carried out on a sample size of N samples by 100, 500, 1000, 2000 and 10000 repetitions. In this case, the estimation of the mode was the mathematical expectation over all realizations. This approach allows not only to obtain an estimate of the mean value, but also the spread of the values of the estimate of the mode relative to the simulated samples. The results of the research are presented in Figure 1.

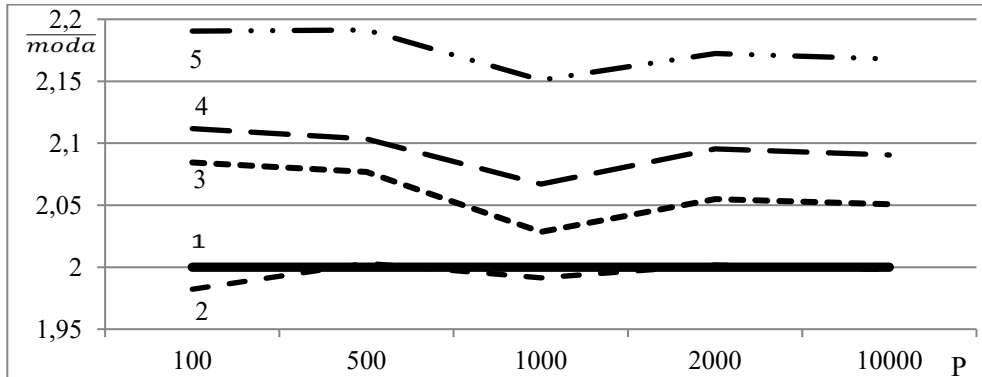


Fig. 1. Comparative analysis of the results of the study of the effectiveness of methods for evaluating modes for the implementation of a random process, distributed according to Rayleigh's law

3 Results and discussion

The analysis of the presented results shows that the accuracy of the mode estimation for each method is different. In Figure 1, a solid line (1) marks the value of the theoretical mode value for Rayleigh's law. Curve 2 characterizes the dependence of the mode estimate on the number of repetitions calculated by expression 7, i.e. according to the method proposed by the authors. Curve 3 characterizes the estimation of the mode from the histogram by averaging the extreme values of the maximum modal interval. The use of the modal interval and the calculation by expression 3 is shown in Figure 1 by line 4, and finally, line 5, the mode estimate is calculated by re-constructing the histogram from the values that fall within the maximum estimation interval of the original histogram. Comparative analysis of the obtained research results allows us to conclude that the

proposed method possesses the smallest error, with 1000 repetitions, the error of this method for curve 2 is 0.43%, for curve 3 is 1.4%, for curve 4 is 8% and finally for 5 is 7.6%. It should be noted that the variation of values for each of the methods differs, for example, for 2 curves – about 5%, for 3-about 27%, for 4-about 20% and for 5-about 26%. From the above studies, we can conclude that the proposed method has not only the best accuracy, but also the smallest variation of values depending on the specific values in the sample.

In conclusion, let us consider the dependence of the mode estimation error on the sample length for the proposed method, which are presented in Figure 2. The results are presented for processing 1000 realizations of length N , which varies from 5,10,20,30,50 and 70. The results obtained fully confirm the previously obtained results, even with small samples $N = 5$, the error does not exceed 5%.

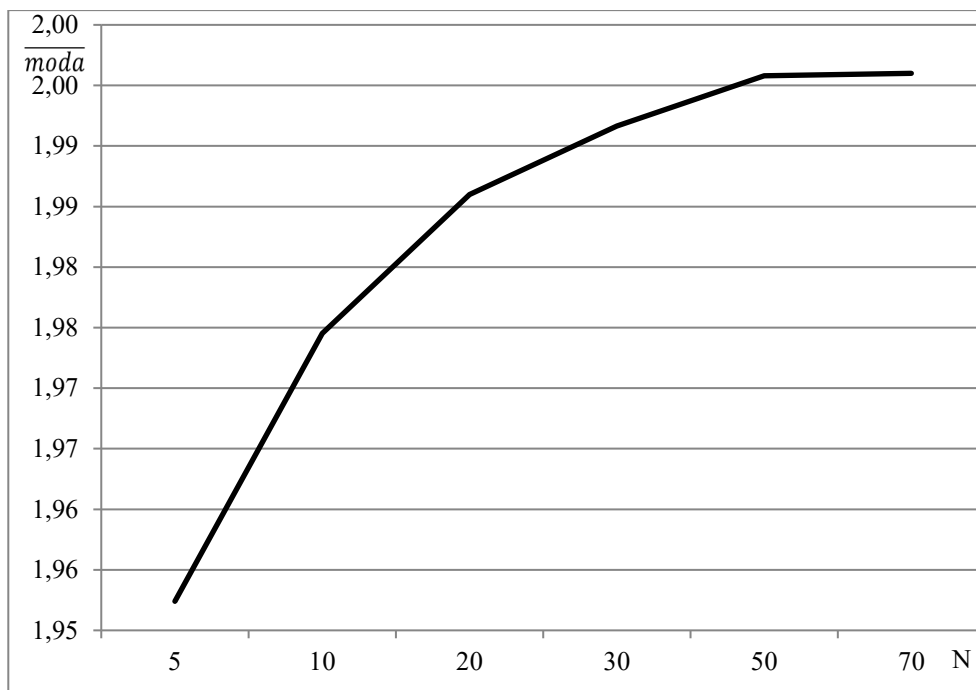


Fig. 2. Mode estimation for small samples over 1000 realizations according to the proposed method

Thus, on the basis of the studies carried out, we can recommend expression 7 for calculating the mode in practice, which makes it possible to provide an estimate of the mode with a minimum error for almost any sample of a random process.

4 Conclusion

1. A new method for calculating the mode of a random process, which is distributed according to Rayleigh's law, is proposed.
2. The conducted research shows a high efficiency using the proposed method, with a calculation error within 0.43%, which is 8 times less than the most commonly used modal interval method.

3. When the sample is reduced to 5 values, the estimation error does not exceed 5%, this case cannot be compared with methods that require the construction of a histogram, which, with such sample values, makes no sense.

References

1. V.V. Bykov. *NASA STI*, **75**, 10294 (1973)
2. V.I. Marchuk, G.R. Saakyan, A.P. Ulanov. Method for identification of a trend by replicating the estimates of its single initial implementation (RAZOTS) and a device for its implementation. Patent No. 2207622, IPC G06F 17 / 18, applicant and patentee «South-Russian State University of Economics and Service» - No. 2000127308/09; declared 10/30/2000; publ. 06/27/2003 Bul. No. 11.
3. V.I. Marchuk, I.S. Shrayfel. Methods for isolating the useful component with a priori uncertainty and a limited amount of measurement results. Publishing House SSSU Shakhty (2008)
4. V. Marchuk. *Serbian journal of electrical engineering*, **15**, 3, 365-370 (2018)
5. V. Marchuk, D Chernyshov, I. Sadrtidinov, A. Minaev. *E3S Web of Conferences*, **104**, 02003 (2019)
6. V. I. Marchuk, G.L. Hripkov, I.O. Nikishin, I.S. Shrayfel. *IOP Conference Series: Materials Science and Engineering*, **1029**, 1, 012094 (2021)
7. M. A. Polidanov, O.D. Eroshina, I. S. Blokhin, A. A. Skorokhod, S. G. Alieva, I. V. Shcherbakova. *Modern Science*, 2-1, 230-233 (2020)
8. E. V. Kabaeva. *Scientific almanac*, 1-1, 439-441 (2016)
9. D. Efimov, L. Fridman, T. Raïssi, A. Zolghadri, R. Seydoud. *Automatica*, 48, 9, 2365-2371 (2012)
10. M. Asadulagi, M. Amal, I. M. Pershin. *Modern Science and Innovations*, 1, 27-32 (2017)
11. M. A. Polidanov, I. S. Blokhin, S. I. Trofimova. *Modern achievements of youth science*, 370-374 (2020)
12. I. M. Kosachev, K. N. Chugai, K. A. Rybakov. *Proceedings of the MAI*, 105, 18-18 (2019)
13. V. D. VERESKUN, M. A. BUTAKOVA, A. N. GUDA, A. V. CHERNOV. *Bulletin of the Rostov State University of Communications*, 1, 125-135 (2019)
14. K. V. Semenikhin. *Autom*, **3**, 103-122 (2019)
15. Sh. M. Chabdarov, A. A. Korobkov. *Systems of synchronization, formation and processing of signals*, **9**, **2**, 162-169 (2018)
16. R. Kh. Vakhitov, M. V. Kaltyrina. *Some questions of analysis, algebra, geometry and mathematical education*, **7-2**, 31-32 (2017)
17. I. M. Kosachev, K. N. Chugay, K. A. Rybakov. *Modeling and data analysis*, 2019, **3**, 73-79 (2019)
18. M. A. Polidanov. *Path to science: collection of student works/ed. I.V. Shcherbakova*. Saratov: Nauka, 145-147 (2019)
19. E. L. Kuleshov. *Avtometriya*, **52**, **1**, 30 (2016)
20. Yu. I. Pshigotskiy, P. N. Fedorov, D. V. Chistyakov, N. A. Shtertser, V. I. Yashin. *Actual problems of infotelecommunications in science and education (APINO 2020)*, 319-324 (2020)