

UDK 519.226.3

doi: 10.26907/2541-7746.2019.1.145-151

SEQUENTIAL d -GUARANTEED ESTIMATE OF THE NORMAL MEAN WITH BOUNDED RELATIVE ERROR

R.F. Salimov, I.N. Volodin, N.F. Nasibullina

Kazan Federal University, Kazan, 420008 Russia

Abstract

In this paper, we continue our research on evaluation of the mean value of the normal distribution with prior information that this parameter is positive and very small. These data are obtained by using a prior exponential distribution with a large intensity parameter. The estimation problem with guaranteed relative error is considered. This issue is more important when small fractions are estimated. In addition to restrictions on the relative error, the procedure must have a given level of d -risk. We suggest a sequential procedure based on the first achievement by posterior probability of estimate reliability of a given level $1 - \beta$. The procedure is adapted to the problem of estimating harmful impurities in food products.

Keywords: first crossing procedure, normal mean estimation, d -posterior approach, sequential estimation

Introduction

As in [1], the problem of estimating the mean value θ of the normal distribution with a known variance σ^2 is considered under the assumption that the unknown value θ is an implementation of ϑ with a prior exponential distribution $F(x) = 1 - \exp\{-\lambda\theta\}$. An estimation $\hat{\theta}_\nu = \theta_\nu(X_1, \dots, X_\nu)$ by a random sample $X^\nu = X_1, \dots, X_\nu$ with an observation stopping moment ν should satisfy the following restriction:

$$P \left\{ \frac{|\vartheta - \hat{\theta}_\nu|}{d} \leq \Delta \mid \hat{\theta}_\nu = d \right\} \geq 1 - \beta.$$

The posterior reliability of the estimate for a given relative accuracy is calculated. A Bayesian estimate for θ in the case of a fixed number of observations n is found and its “ d -reliability” is calculated. Assuming that the values σ^2 and λ are known, a sequential estimation procedure is constructed. This procedure is based on the first crossing of the given constraint $1 - \beta$ by the posterior reliability (see [2, 3] for examples of using this procedure).

The constructed procedure is adapted to the problem of estimating harmful impurities in food products. It should be noted that we know nothing about the guaranteed procedure based on a fixed number of observations for the case when constraints are given by the relative estimation error. The d -posterior approach for solving this problem enables construction of a sequential guaranteed procedure (see [4–9] for the classical solutions of sequential guaranteed procedures). The simulation results show that the observations volume with a sequential procedure will be with high probability unacceptably large for practical use. It turned out that these large volumes correspond to too small values of θ , which gives the prior distribution. We considered a more

practical situation with truncation of the prior distribution at a certain threshold close to the standardized value of the parameter. In this case, the number of observations becomes plausible for practical application.

1. Bayesian estimate θ for a fixed volume of observations

Formally, within the framework of the general theory of statistical inference, the problem consists in estimating of the mean θ of the normal (θ, σ^2) distribution with a loss function: $L(\theta, d) = 1$, if $|\theta - d|/d > \Delta$, and is 0 otherwise. Here, Δ is a given restriction on the accuracy of estimation. Since the statistical experiment has sufficient statistics $S = \sum_{k=1}^n X_k$, the family of distributions of the random sample is reduced to a family of normal $(\theta, \sigma^2/n)$ density functions of sufficient statistics:

$$p(s | \theta) = \frac{1}{\sqrt{2\pi n\sigma}} \exp \left\{ -\frac{1}{2n\sigma^2} (s - n\theta)^2 \right\}, \quad \theta \in R, \quad s \in R.$$

It is assumed that the prior distribution of the random parameter ϑ has a density function

$$g(\theta) = \lambda e^{-\lambda\theta}, \quad \theta > 0, \quad \lambda > 0.$$

Now, we turn to the problem of θ estimation with the prior information taken into account. This problem consists in constructing a procedure for estimating (φ, ν) with a decision function $\hat{\theta}_\nu$ satisfying the inequality

$$\mathfrak{R}(d; \varphi) = \inf_{d \in \mathcal{D}} \mathbf{P} \left\{ \frac{1}{1 + \Delta} \leq \frac{\vartheta}{\hat{\theta}_\nu} \leq 1 + \Delta \mid \hat{\theta}_\nu = d \right\} \geq 1 - \beta.$$

The Bayesian estimation for θ , which we will use for constructing sequential procedure, is defined as a maximum point of the posterior reliability over a . The density function of the posterior distribution is calculated in [1]. It is the density function of the truncated normal $(T, \sigma^2/n)$ distribution:

$$h(\theta | T) = \frac{\sqrt{n}}{\Phi(T\sqrt{n}/\sigma)\sqrt{2\pi\sigma}} \exp \left\{ -\frac{n}{2\sigma^2} (\theta - T)^2 \right\}, \quad \theta > 0,$$

where

$$T = \frac{S - \lambda\sigma^2}{n} = \bar{X} - \frac{\lambda\sigma^2}{n},$$

and $\Phi(\cdot)$ is the distribution function of the standard normal law.

Notably, that the inequality $|\vartheta - a|/a \leq \Delta$ is equivalent to $1 - \Delta \leq \vartheta/a \leq 1 + \Delta$, and, since posterior distribution ϑ is concentrated on the positive semi-axis, the posterior reliability of the solution a (estimation of θ) is

$$\begin{aligned} H_n(T) &= P \left(\frac{1}{1 + \Delta} \leq \frac{\vartheta}{a} \leq 1 + \Delta \mid T \right) = \\ &= \frac{\sqrt{n}}{\sqrt{2\pi\sigma} \Phi(T\sqrt{n}/\sigma)} \int_{a/(1+\Delta)}^{a(1+\Delta)} \exp \left\{ -\frac{n}{2\sigma^2} (\theta - T)^2 \right\} d\theta = \\ &= \frac{\Phi((a(1 + \Delta) - T)\sqrt{n}/\sigma) - \Phi((a/(1 + \Delta) - T)\sqrt{n}/\sigma)}{\Phi(T\sqrt{n}/\sigma)}. \quad (1) \end{aligned}$$

In [10], it was shown that the estimates that provide the minimum of d -risk should be d -minimax. Unfortunately, the methods for constructing such estimates are still not known and the minimax estimate for θ is found only for the case of normal prior distribution (see [11]). The form of expression (1) for posterior reliability does not show any hope to obtain the estimate with the uniformly minimal d -risk in the closed form. So, the only thing that we can do is to find the Bayesian estimate, which is necessary to construct the first crossing sequential procedure.

The Bayesian estimate $\hat{\theta}_G = a(T)$ is the point of attaining the maximum of the function

$$W(a) = \int_{a/(1+\Delta)}^{a(1+\Delta)} \exp \left\{ -\frac{n}{2\sigma^2} (\theta - T)^2 \right\} d\theta,$$

over a . Using the traditional methods of differential calculation to find the function extremum, we obtain the Bayesian estimate

$$\hat{\theta}_G = \hat{\theta}_G(T) = \frac{T}{c_2} + \sqrt{\frac{T^2}{c_2^2} + \frac{4\sigma^2}{nc_1c_2} \log(1 + \Delta)},$$

where

$$c_1 = 1 + \Delta - \frac{1}{1 + \Delta}, \quad c_2 = 1 + \Delta + \frac{1}{1 + \Delta}.$$

It is easy to see that when $\Delta \rightarrow 0$,

$$\hat{\theta}_G \sim \frac{1}{2} \left(T + \sqrt{T^2 + \frac{\sigma^2}{n}} \right).$$

Thus, the posterior reliability of the Bayesian estimate is

$$\bar{H}_n(T) = \frac{\Phi(\hat{\theta}_G(1 + \Delta) - T)\sqrt{n}/\sigma) - \Phi(\hat{\theta}_G/(1 + \Delta) - T)\sqrt{n}/\sigma)}{\Phi(T\sqrt{n}/\sigma)}. \tag{2}$$

Proposition 1. *The Bayesian estimate $\hat{\theta}_G$, based on a fixed number of observations n , has d -risk*

$$\mathfrak{R}(d; \hat{\theta}_G) = 1 - \frac{\Phi((d(1 + \Delta) - T_d)\sqrt{n}/\sigma) - \Phi((d/(1 + \Delta) - T_d)\sqrt{n}/\sigma)}{\Phi(T_d\sqrt{n}/\sigma)},$$

where

$$T_d = \frac{c_2}{2d} \left[d^2 - \frac{4\sigma^2}{nc_1c_2} \log(1 + \Delta) \right].$$

Proof. The d -risk of estimation is the conditional mathematical expectation of the posterior risk of estimation with respect to the decision function. So, it is equal to the substitution in a posteriori risk (4) d for a and the root of the equation $\hat{\theta}_G = d$ for T . The simple calculations show that this root is

$$T_d = \frac{c_2}{2d} \left[d^2 - \frac{4\sigma^2}{nc_1c_2} \log(1 + \Delta) \right].$$

□

It is easy to check that the d -risk of the Bayesian estimate has a range of values that fills the entire interval $[0; 1]$. Thus, the Bayesian estimate, like the \bar{X} in the classical approach, does not solve the problem of estimating the mean of the normal distribution with guaranteed limitations on the accuracy and reliability of the assessment.

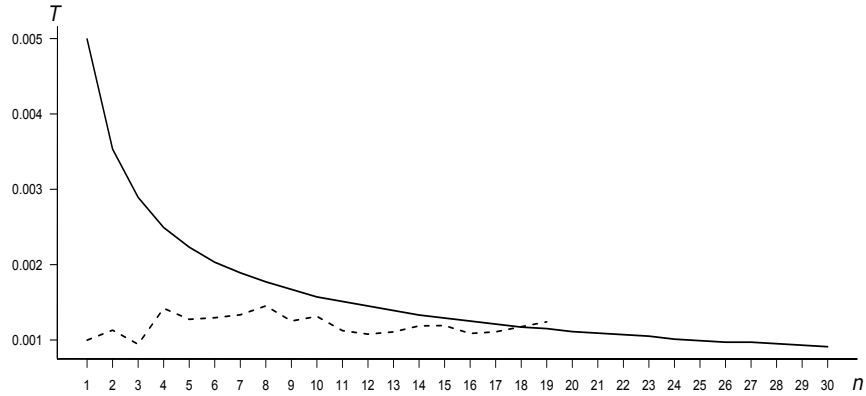


Fig. 1. Graphical illustration for first crossing procedure

2. Sequential first crossing procedure

The first crossing procedure is defined by the stopping moment

$$\nu = \min\{n : \bar{H}_n(T) \geq 1 - \beta\}, \quad (3)$$

when the posterior probability (2) crosses the given reliability level $1 - \beta$ for the first time. After the experiment is stopped at some step n , the unknown value θ is defined by the calculation of the Bayesian estimate.

In contrast to the analogous problem for estimation with the absolute error, the first crossing procedure (see [1]) does not stop with probability 1.

The boundary of the regions of continuation and stopping of observations is shown in Fig. 1 with an illustration of the trajectory of reaching this boundary.

The construction of an empirical analogue of this evaluation procedure is set in paragraph 4 of [1]. There are also mentioned applications to evaluation of the content of arsenic in food products, which are applicable without change to the procedure with the relative error.

3. Investigating the properties of the first intersection procedure by statistical modeling

The distribution of the stopping moment of the presented sequential procedure is investigated using the statistical modeling method within the same formulation of the problem as in [1]. It should be reminded that the sanitary-epidemiological rules and regulations of the Russian Federation assign the upper threshold θ_0 of arsenic content as 0.1–0.2 mg/kg. Assuming that the input quality level is $Q = 0.99$, we get the value $\lambda = 25$ at the rate of 0.2. Using the same arguments as in [1] for selecting the values for variance and guaranteed accuracy, we can set $\Delta = \sigma = 0.01$.

The 10^4 replications of the sequential first intersection procedure were performed, where the stopping moment is determined by (3). An estimate of the distribution of ν is presented in Table 1. The obtained data indicate an unacceptably large amount of observations in comparison with the fact that we got under restrictions on the absolute error in [1]. The analysis of the values θ , which produce the prior distribution in our modeling, shows that the large values ν arise only for very small values θ ($< 10^{-3}$). It is clear that such values of the arsenic content in a series of products are hardly possible in real production practice. Usually the content of arsenic, as well as other harmful

Table 1. Stopping moment distribution estimate

n	1	2	3	4	5	6	7	8	9	10
\hat{P}	0.007	0.021	0.029	0.029	0.027	0.023	0.022	0.021	0.019	0.016
n	11-20	21-30	31-40	41-50	>51					
\hat{P}	0.113	0.074	0.056	0.041	0.50					

Table 2. Stopping moment distribution estimate in case of truncation

n	1	2	3	4
\hat{P}	0.091	0.287	0.333	0.289

impurities, is close to half of the standardized value. In this regard, if we assume that the value θ in the experiment comes from a truncated exponential distribution, such as $1 - \exp\{-\lambda(\theta - 0.1)\}$, then the random number of observations in the experiment with a high probability becomes indeed acceptable (see Table 2).

Conclusions

The paper presents the solution of the statistical problem of estimating the mean θ of the normal distribution for prior information on the positivity and small value of the estimated parameter, when the volume of observations is determined by the given restrictions on the relative accuracy and reliability of the estimate. As noted in the Introduction, we do not know the guarantee procedure for estimating θ on a fixed number of observations, when the restrictions relate to the relative estimation error. The d -posterior approach to solving this problem led us to the construction of a consistent guaranteed procedure. Notably, this approach does not require a small probability of the given deviations Δ of the estimate (random variable-statistics) from the fixed (given) value θ , but provide the guaranteed probability $(1 - \beta)$ of a hitting random parameter ϑ to the given neighbourhood of the estimate obtained in the statistical experiment.

The results of simulation of the stopping moment distribution show that the amount of observations using a sequential procedure with a high probability may be unacceptably large for practical use. It was revealed that these large volumes arise only in cases when the prior distribution “throws out” an excessively small value of θ . If we assume that in practice the values θ are extremely small, which is quite true for a number of food products containing arsenic, then the proposed sequential procedure has quite acceptable volumes of observations.

Acknowledgements. This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities (project no. 1.7629.2017/8.9).

References

1. Salimov R.F., Su-Fen Yang, Turilova E.A., Volodin I.N. Estimation of the mean value for the normal distribution with constraints on d -risk. *Lobachevskii J. Math.*, 2018, vol. 39, no. 3, pp. 377–387. doi: 10.1134/S1995080218030174.
2. Volodin I.N. Guaranteed statistical inference procedure (determination of the sample size). *J. Sov. Math.*, 1989, vol. 44, no. 5, pp. 568–600. doi: 10.1007/BF01095166.

3. Salimov R.F. A sequential d -guaranteed test for distinguishing two interval hypotheses. *Lobachevskii J. Math.*, 2016, vol. 37, no. 4, pp. 500–503. doi: 10.1134/S1995080216040156.
4. Roughani G., Mahmoudi E. Exact risk evaluation of the two-stage estimation of the gamma scale parameter under bounded risk constraint. *Sequential Anal.*, 2015, vol. 34, no. 3, pp. 387–405. doi: 10.1080/07474946.2015.1030303.
5. Mukhopadhyay N. Improved sequential estimation of means of exponential distributions. *Ann. Inst. Stat. Math.*, 1994, vol. 46, no. 3, pp. 509–519. doi: 10.1007/BF00773514.
6. Mukhopadhyay N., Datta S. On fine-tuned bounded risk sequential point estimation of the mean of an exponential distribution. *S. Afr. Stat. J.*, 1995, vol. 29, no. 1, pp. 9–27.
7. Mukhopadhyay N., Pepe W. Exact bounded risk estimation when the terminal sample size and estimator are dependent: The exponential case. *Sequential Anal.*, 2006, vol. 25, no. 1, pp. 85–101. doi: 10.1080/07474940500452254.
8. Zacks S., Mukhopadhyay N. Bounded risk estimation of the exponential parameter in a two-stage sampling. *Sequential Anal.*, 2006, vol. 25, no. 4, pp. 437–452. doi: 10.1080/07474940600934896.
9. Zacks S., Mukhopadhyay N. Exact risks of sequential point estimators of the exponential parameter. *Sequential Anal.*, 2006, vol. 25, no. 2, pp. 203–220. doi: 10.1080/07474940600596703.
10. Volodin I.N. Optimum sample size in statistical inference procedures. *Izv. Vyssh. Uchebn. Zaved. Mat.*, 1978, no. 12, pp. 33–45. (In Russian)
11. Simushkin S.V., Volodin I.N. Statistical inference with a minimal d -risk. *Lect. Notes Math.*, 1983, vol. 1021, pp. 629–636.

Received
December 6, 2017

Salimov Rustem Faridovich, Assistant Lecturer of Department of Mathematical Statistics

Kazan Federal University
ul. Kremlevskaya, 18, Kazan, 420008 Russia
E-mail: rustem.salimov@gmail.com

Volodin Igor Nikolaevich, Doctor of Physics and Mathematics, Professor of Department of Mathematical Statistics

Kazan Federal University
ul. Kremlevskaya, 18, Kazan, 420008 Russia
E-mail: igorvolodin@gmail.com

Nasibullina Nailya Fergatevna, Student of Institute of Computational Mathematics and Information Technologies

Kazan Federal University
ul. Kremlevskaya, 18, Kazan, 420008 Russia
E-mail: nasibusha@bk.ru

УДК 519.226.3

doi: 10.26907/2541-7746.2019.1.145-151

**Последовательная d -гарантийная оценка
среднего значения нормального распределения**

Р.Ф. Салимов, И.Н. Володин, Н.Ф. Насибуллина

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

Аннотация

Мы продолжаем исследования по проблеме оценки среднего значения нормального распределения с априорной информацией о положительности и малости этого параметра. Эта информация доставляется априорным показательным распределением с большим значением параметра интенсивности. Рассматривается задача оценки с гарантированной относительной ошибкой, что более важно при оценке малых долей. Кроме ограничений на относительную ошибку процедура должна иметь заданный уровень d -риска. Предлагается последовательная процедура оценки, основанная на первом достижении апостериорной вероятностью надежности оценки заданного уровня $1 - \beta$. Построенная процедура адаптируется к проблеме оценивания содержания вредных примесей в пищевых продуктах.

Ключевые слова: процедура первого пересечения, оценка среднего значения нормального распределения, d -апостериорный подход, последовательное оценивание

Поступила в редакцию
06.12.17

Салимов Рустем Фаридович, ассистент кафедры математической статистики

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *rustem.salimov@gmail.com*

Володин Игорь Николаевич, доктор физико-математических наук, профессор кафедры математической статистики

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *igorvolodin@gmail.com*

Насибуллина Наиля Фергатовна, студент Института вычислительной математики и информационных технологий,

Казанский (Приволжский) федеральный университет
ул. Кремлевская, д. 18, г. Казань, 420008, Россия
E-mail: *nasibusha@bk.ru*

For citation: Salimov R.F., Volodin I.N., Nasibullina N.F. Sequential d -guaranteed estimate of the normal mean with bounded relative error. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki*, 2019, vol. 161, no. 1, pp. 145–151. doi: 10.26907/2541-7746.2019.1.145-151.

Для цитирования: Salimov R.F., Volodin I.N., Nasibullina N.F. Sequential d -guaranteed estimate of the normal mean with bounded relative error // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2019. – Т. 161, кн. 1. – С. 145–151. – doi: 10.26907/2541-7746.2019.1.145-151.