

INTERVAL ESTIMATION OF DISTRIBUTION PARAMETER BY STATISTICAL TRIALS OF EXPECTED VALUE

V.O. Barannik

O.M. Beketov National University of Urban Economy, Kharkiv, Ukraine

E-mail: v_barannik@ukr.net

The distribution parameter interval estimators are obtained by direct numerical approximation of the expected value for infinite and finite populations under the known upper and lower bounds of the random variable domain. Like in Bayesian approach, the distribution parameters are treated as random variables, and their uncertainty is described as a distribution. The Monte Carlo procedure is involved to get the correspondent confidence interval endpoints. The model does not impose any restrictions on the type of distributions. In contrast to other nonparametric interval assessments of distribution parameters, the model is operable for samples of any size.

PACS: 02.50.Ng

INTRODUCTION

Data analysis lies in the basement of physical sciences. The significantly increased cost of modern physical experiment makes us pay attention to the lack of reliable methods for analyzing small-volume samples. Actually, the sample methods [1] are widely used to study states of large population. In this case, the collective properties of the population, being quantified through the parameters of the distribution of individual properties among the elements, are usually of great interest. In order to evaluate a specific distribution parameter, a sampling plan for elements is determined and implemented, which ensures representativeness of the sample for estimation, taking into account the known properties of the population.

In the analysis of sample data, both parametric and nonparametric methods can be used, depending on whether hypotheses are involved in the analysis regarding the type of distribution or not. An example of a parametric approach is the maximum likelihood method [2], in which an interval estimation of the distribution parameter is performed using the likelihood function constructed from the probability distributions of sample elements. At the same time the consequences of specifying the wrong distribution may prove very costly. If such distribution does not hold, then the confidence levels of the confidence intervals (or of hypotheses tests) may be completely off.

An example of a nonparametric approach is the bootstrap method, which is currently widespread [3, 4], in which it is assumed that a representative sample adequately reflects the distribution structure of a property in a population and can replace a population. This allows one to obtain interval estimate of the distribution parameter using statistical trials (Monte Carlo method [5]), consisting in multiple selections of samples with replacement from the original sample. At the same time, for all the asymptotic advantages of method [6], one can hardly expect reliable estimates if the size of the original sample is small or extremely small, because bootstrap operates under assumption that all possible different values of population have been observed [7].

Another approach to statistical inference may be related to the fact that any real property of a real population is a limited variable distributed in the local domain of the real line. In this regard, the sample can be considered as a random segmentation of the domain. Domain

partition becomes complete if the sample is supplemented with numerical values of the upper and lower domain boundaries. If the distribution parameter can be represented as an integral or a sum for the expected value, then such a domain structure can serve as the basis for the interval parameter estimation by approximating this representation. A variant of the model of expected value approximation is described below.

1. PROBLEM FORMULATION

Infinite population. We consider an infinite population of elements having a measurable property X . The result x of measuring this property in a randomly selected element is a random variable with an unknown integrable probability density function (**pdf**) $\rho(x)$ having limited domain $x_{\min} \leq x \leq x_{\max}$, where x_{\min} and x_{\max} are known lower and upper endpoints of the domain, respectively. We conditionally call population to be infinite if it is possible to draw the simple random sample of any size n without introducing any distortions to the **pdf**. In this sense the finite population that allows extracting samples with replacement can be referred to infinite.

Let $u(x)$ is integrable monotonic (non-decreasing or non-increasing) function (generator of distribution parameter) determining the contribution of the individual property x of element to the collective property U of population, which we call distribution parameter. Then we consider distribution parameter that can be presented as integral

$$U = \int_{x_{\min}}^{x_{\max}} u(x) \rho(x) dx . \quad (1)$$

It may be m -th moment of distribution if $u(x)=x^m$, or proportion of population with property $x \leq x^*$ if $u(x) = Y(x^* - x)$, or something else ($Y(*)$ is Heaviside function). If there is no additional information then one can estimate the value of distribution parameter as

$$u_{\min} \leq U \leq u_{\max} , \quad (2)$$

where

$$u_{\min} = \min(u(x_{\min}), u(x_{\max})) , \\ u_{\max} = \max(u(x_{\min}), u(x_{\max})) .$$

To improve the precision of this estimate, if it is deemed insufficient, the measurement data are added to data available on the property of the elements from a

simple random sample of size n . Thus, the new data set in ascending order looks like this:

$$x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n. \quad (3)$$

This data are combined with endpoints $x_0 = x_{\min}$ and $x_{n+1} = x_{\max}$; the corresponding series of values of the generator function being either non-decreasing or non-increasing:

$$u_0 = u(x_{\min}), \{u_i = u(x_i)\}_{i=1}^n, u_{n+1} = u(x_{\max}). \quad (4)$$

It is required to reassess the value (2) of the distribution parameter (1) in the light of new data.

Finite population. We also consider a finite population x_1, x_2, \dots, x_N with a known size N , for which the distribution parameter is determined by the sum

$$S = \frac{1}{N} \sum_{j=1}^N s(x_j), \quad (5)$$

where $s(x)$ is another monotonic generator function.

Without loss of generality, we assume that the elements of the population are numbered in order of increasing property x . If sampling were made with replacement from this population, then the population could be considered as infinite population having pdf as

$$\rho_N(x) = \frac{1}{N} \sum_{j=1}^N \delta(x - x_j), \quad (6)$$

where $\delta(*)$ is Dirac delta function.

Let (3) be the simple random sample extracted from this population without replacement. The correspondent values of the generator function are

$$\{s_i = s(x_i)\}_{i=1}^n. \quad (7)$$

In this case initial interval estimation of S is also just a range of function $s(x)$:

$$s_{\min} \leq S \leq s_{\max}, \quad (8)$$

where

$$s_{\min} = \min(s(x_{\min}), s(x_{\max})), \\ s_{\max} = \max(s(x_{\min}), s(x_{\max})).$$

The task is the same as before: it is required to reassess the value (8) of the distribution parameter (5) in the light of new data.

2. PROBLEM ANALYSIS

Infinite population. We introduce the cumulative distribution function (cdf) into consideration in a usual way

$$f(x) = \int_{x_{\min}}^x \rho(x) dx.$$

In particular for (6) it will be

$$f_N(x) = \frac{1}{N} \sum_{j=1}^N Y(x - x_j).$$

Integral (1) can be presented as following

$$U = \int_0^1 u[x(f)] df = \sum_{i=1}^{n+1} \int_{f_{i-1}}^{f_i} u[x(f)] df, \quad (9)$$

where $x(f)$ is inverse cdf and $\{f_i = f(x_i)\}_{i=1}^n$, $f_0 = 0$, $f_{n+1} = 1$.

In accordance to the integral mean value theorem equation (9) can be presented as following:

$$U = \sum_{i=1}^{n+1} \hat{u}_i (f_i - f_{i-1}), \quad (10)$$

where, if $u(x)$ is entirely non-decreasing or non-increasing function,

$\min(u_{i-1}, u_i) \leq \hat{u}_i \leq \max(u_{i-1}, u_i)$; $u_i = u(x_i)$. Then following conditions are fulfilled:

$$U \geq U_{\min} = \sum_{i=1}^{n+1} \min(u_{i-1}, u_i) (f_i - f_{i-1}), \quad (11)$$

$$U \leq U_{\max} = \sum_{i=1}^{n+1} \max(u_{i-1}, u_i) (f_i - f_{i-1}). \quad (12)$$

Although values $\{f_i\}_{i=1}^n$ are unknown they have well-known posterior pdf as following

$$\rho(f_1, f_2, \dots, f_n) = \begin{cases} n! & \text{if } 0 \leq f_1 \leq f_2 \leq \dots \leq f_n \leq 1 \\ 0 & \text{otherwise} \end{cases}. \quad (13)$$

It means that any set $\{f_i\}_{i=1}^n$ of random independent uniformly distributed on $[0, 1]$ and ordered in ascending order numbers is equally probable and can be considered to be likely true set. Now, taking into account (13) and representations (11), (12) for endpoints, sample (3) can be considered as deterministic and parameter U as random.

To get interval assessment of population parameter (1) by use of statistical trials [8] one can generate K sets of above-mentioned numbers $\{f_{i,k}\}_{i=1}^n$ ($k=1, 2, \dots, K$) and calculate correspondent posterior statistics:

$$\tilde{U}_{\min, k} = \sum_{i=1}^{n+1} \min(u_{i-1}, u_i) (f_{i,k} - f_{i-1,k}), \quad (14)$$

$$\tilde{U}_{\max, k} = \sum_{i=1}^{n+1} \max(u_{i-1}, u_i) (f_{i,k} - f_{i-1,k}). \quad (15)$$

Arranging the results of statistical trials (14) and (15) in ascending order $\{U_{\min, m}\}_{m=1}^K$ and $\{U_{\max, m}\}_{m=1}^K$ one can get the resultant interval estimation in the form of confidence interval

$$U_{\min, \frac{\alpha}{2}K} \leq U \leq U_{\max, (1-\frac{\alpha}{2})K}, \quad (16)$$

where α is appropriate level of significance.

Finite population. It is suitable to rewrite the sum (5) in the following way:

$$S = \frac{1}{N} \sum_{j=1}^N s(x_j) = \frac{N+1}{N} V - \frac{s(x_0) + s(x_{N+1})}{2N}, \quad (17)$$

$$V = \frac{1}{N+1} \sum_{j=1}^{N+1} \frac{s(x_j) + s(x_{j-1})}{2}, \quad (18)$$

where $x_0 = x_{\min}$ and $x_{N+1} = x_{\max}$.

Let $r(i)$ be the serial number in the population of the element, which in the sample takes position i . Let also the serial number of elements $x_0 = x_{\min}$ and $x_{N+1} = x_{\max}$ attached to the population be $r(0) = 0$ and $r(n+1) = N+1$. Then expression (18) can be represented as the sum of elements within the segments of real line, the ends of which are the elements of the sample:

$$V = \frac{1}{N+1} \sum_{i=1}^{n+1} \frac{s(x_j) + s(x_{j+1})}{2}. \quad (19)$$

It is now possible to express the value V in terms of the average values of the generator function \hat{s}_i within each segment i :

$$V = \frac{1}{N+1} \sum_{i=1}^{n+1} \hat{s}_i \cdot [r(i) - r(i-1)], \quad (20)$$

$$\hat{s}_i = \frac{1}{r(i) - r(i-1)} \sum_{j=r(i-1)}^{r(i)-1} \frac{s(x_j) + s(x_{j+1})}{2}. \quad (21)$$

Taking into account the monotonic property of generator $s(x)$ the value (20) can be assessed as

$$V \geq V_{\min} = \frac{1}{N+1} \sum_{i=1}^{n+1} \min[s(x_{i-1}), s(x_i)] \cdot [r(i) - r(i-1)], \quad (22)$$

$$V \leq V_{\max} = \frac{1}{N+1} \sum_{i=1}^{n+1} \max[s(x_{i-1}), s(x_i)] \cdot [r(i) - r(i-1)]. \quad (23)$$

In these expressions, now random variables are actually not the elements of the sample but the places that the elements of the sample occupy in an ordered population.

Obviously Monte Carlo procedure can be also applied to find the lower and upper bounds of the confidence interval of appropriate significance level α for the population parameter S . For this purpose K random samples of n positive integers should be extracted from the set $\{r\}_{r=1}^N$ without replacement, ordered from the bottom to the top to represent likely true sets $\{r(i, k)\}_{i=1}^n$, ($k=1, 2, \dots, K$), and substituted to the equations (22), (23) to calculate likely true values of correspondent endpoints.

After sorting resultant sets in ascending order $\{V_{\min, k}\}_{k=1}^K$, $\{V_{\max, k}\}_{k=1}^K$ the confidence interval of desirable significance level α for the distribution parameter S can be presented as following:

$$\frac{N+1}{N} V_{\min, K \frac{\alpha}{2}} - \frac{s(x_{\min}) + s(x_{\max})}{2N} \leq S \leq \frac{N+1}{N} V_{\max, K \left(1 - \frac{\alpha}{2}\right)} - \frac{s(x_{\min}) + s(x_{\max})}{2N}. \quad (24)$$

3. SIMULATIONS

Here we consider some special cases of the above mentioned model application to demonstrate its useful properties.

Infinite population – continuous distribution. Let $\rho(x)$ be the pdf of the uniformly distributed within closed interval $[0, 1]$ random variable, so that

$$\rho(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

then m -th moment of this distribution is $M = (1 + m)^{-1}$. To test the described approach for statistical inference we generated 100 samples of different size $n=3, 7, 15, 30, 50$ from this distribution and did $K=1000$ statistical trials on the every sample to get interval estimation for the following moment: $m = 3, 7, 15, 30$; the generator function being $u(x)=x^m$. The endpoints of confidence interval (16) with confidence level $1-\alpha=0.95$ for the every sample and minimal D_{\min} and maximal D_{\max} widths of the confidence interval for the every set of samples of definite size were calculated together with the number of faults when true value of

moment went outside the correspondent interval. These results are presented in table below.

Simulation Output for the Uniform Distribution

n	Parameters	Order of Moment, m				
		1	3	7	15	30
3	D_{\min}	0.677	0.677	0.656	0.673	0.669
	D_{\max}	0.861	0.861	0.861	0.881	0.88
	Faults	0	0	2	0	0
7	D_{\min}	0.423	0.419	0.378	0.388	0.37
	D_{\max}	0.65	0.631	0.582	0.652	0.595
	Faults	0	0	1	1	2
15	D_{\min}	0.27	0.233	0.224	0.207	0.2
	D_{\max}	0.415	0.436	0.399	0.336	0.354
	Faults	2	0	0	1	1
30	D_{\min}	0.204	0.154	0.127	0.12	0.104
	D_{\max}	0.276	0.293	0.264	0.266	0.206
	Faults	3	2	2	6	1
50	D_{\min}	0.153	0.128	0.099	0.082	0.071
	D_{\max}	0.207	0.21	0.191	0.172	0.151
	Faults	3	2	2	2	1

One can see from table that described numerical algorithm adequately worked independently on the sample size and moment order providing the statistical inference on the chosen significance level. These results demonstrate expected decrease of the confidence interval width (increase of the assessment precision) with increase of the sample size and order of the moment.

At the same time, the graphs (Figs. 1-5) of the empirical distribution functions for the confidence interval boundaries show that the reduction of the width of the confidence interval at a high value of the moment order ($m = 30$) is achieved mainly by moving the right boundary. The distribution function of the left boundary of the confidence interval remains sandwiched between the left endpoint of the domain and the exact value of the moment when the sample size grows.

The cdf of the generator $u_m(x)=x^m$ for the m -th moment of considered random variable is $\eta_m(u)=\sqrt[m]{u}$.

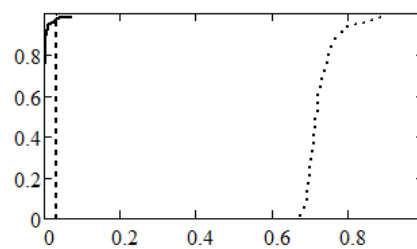


Fig. 1. Distribution functions of the left (solid line), and right (dotted line) confident interval boundaries for the sample size $n = 3$. Dashed line shows the point of true moment value $M = 1/31$

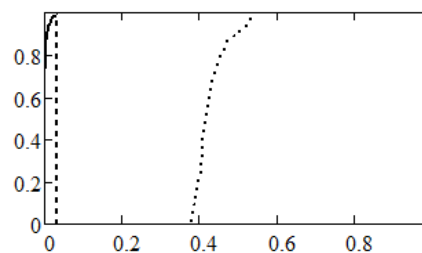


Fig. 2. The same for the sample size $n = 7$

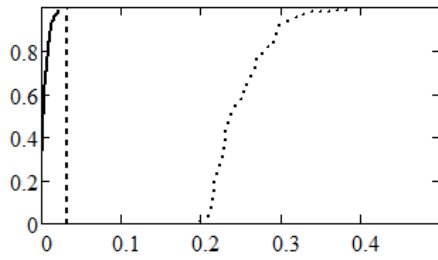


Fig. 3. The same for the sample size $n = 15$

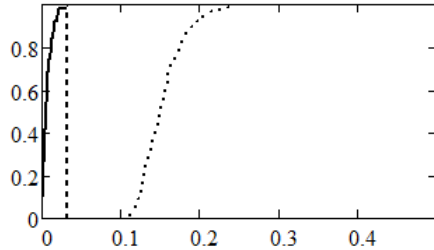


Fig. 4. The same for the sample size $n = 30$

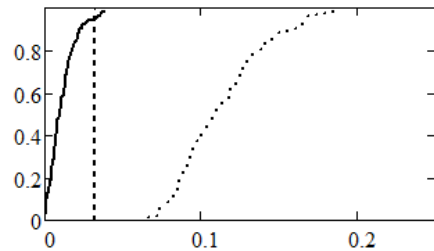


Fig. 5. The same for the sample size $n = 50$

It can be understood from Fig. 6, where dependence of moment distribution skewness and kurtosis versus moment order m is graphically presented, that moment distribution has right tail; being right-skewed the distribution is concentrated near the left boundary of the interval $[0, 1]$.

Infinite population – discrete distribution. Here we consider dichotomous population that consists of elements having one of two possible signs $x=1$ or $x=0$, in other words – “success” or “failure”, and p is proportion of population elements having sign 1.

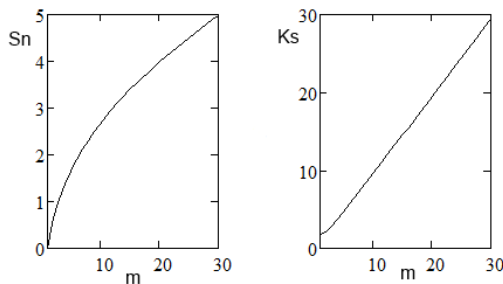


Fig. 6. Skewness (Sn) and kurtosis (Ks) vs order (m) of the moment of variable uniformly distributed on $[0, 1]$

If the random value is the sign x of the element randomly extracted from the population (Bernoulli trial) then correspondent **pdf** is

$$\rho(x) = (1-p)\delta(x) + p\delta(x-1),$$

and first moment $u(x)=x$ of this distribution is

$$\int_{-\infty}^{\infty} x\rho(x)dx = p.$$

Let x_1, x_3, \dots, x_n be the signs of elements in the ordered simple random sample from this population, so

that $n-s$ elements have sign 0 and s elements have sign 1: $x_{1 \leq i \leq n-s} = 0$, $x_{n-s+1 \leq i \leq n} = 1$. Then

$$U_{\min} = \sum_{i=1}^{n+1} \min(u_{i-1}, u_i)(f_i - f_{i-1}) =$$

$$\sum_{i=n-m+2}^{n+1} (f_i - f_{i-1}) = 1 - f_{n-m+1},$$

$$U_{\max} = \sum_{i=1}^{n+1} \max(u_{i-1}, u_i)(f_i - f_{i-1}) =$$

$$\sum_{i=n-m+1}^{n+1} (f_i - f_{i-1}) = 1 - f_{n-m}.$$

Taking into account that the i -th order statistic f_i is a beta-distributed random variable having **pdf**

$$\rho(f_i) = \frac{n!}{(i-1)!(n-i)!} f_i^{i-1} (1-f_i)^{n-i} =$$

$$\text{Beta}(i, n-i+1),$$

the **pdf** of the confidence interval endpoints are

$$\rho(U_{\min}) = \text{Beta}(m, n-m+1),$$

$$\rho(U_{\max}) = \text{Beta}(m+1, n-m).$$

Then the $1-\alpha$ confidence interval for the probability of success is just Clopper-Pearson interval [9] (“exact” confidence interval) that is an early and very common method for calculating binomial confidence intervals:

$$B\left(\frac{\alpha}{2}, m, n-m+1\right) \leq p \leq B\left(1-\frac{\alpha}{2}, m+1, n-m\right),$$

where $B(r, s, z)$ is r -th quantile from the beta distribution with shape parameters s and z .

Finite population. Here we consider dichotomous population consisted of N elements, where every element x has also one of two possible signs 1 or 0. Then the sum (5), where generator $s(x_j) = x_j$, turns to the proportion

$$S = \frac{1}{N} \sum_{j=1}^N s(x_j) = \frac{F}{N} = p,$$

where F is number of elements having sign 1 and p is probability to extract element having sign 1 if Bernoulli trial is applied.

Let $x_{1 \leq i \leq n-m} = 0$ and $x_{n-m+1 \leq i \leq n} = 1$ is ordered simple random sample of size n without replacement where m elements have sign “1”. Then equations (22), (23) take the following view

$$V \geq V_{\min} = \frac{1}{N+1} [N+1-r(n-m+1)],$$

$$V \leq V_{\max} = \frac{1}{N+1} [N+1-r(n-m)].$$

The total number of ordered simple random samples of size n that can be drawn without replacement from the population including N elements is $\binom{N}{n}$. The number of such samples with fixed value $i \leq r(i) \leq N-n+i$ is $\binom{r(i)-1}{i-1} \binom{N-r(i)}{n-i}$. That means that probability distribution for the discrete random value $r(i)$ can be presented as

$$P[r(i)] = \binom{N}{n}^{-1} \binom{r(i)-1}{i-1} \binom{N-r(i)}{n-i}. \quad (25)$$

Equations (26), (27) enable to calculate two endpoints, $r_{\max, \alpha}$ and $r_{\min, \alpha}$, of the confidence interval (28) for the desirable level of significance α , though the discontinuous nature of the distribution (25) may preclude any interval with exact probability coverage for all population proportions.

$$P[r(n-m+1) \leq r_{\max, \alpha}] = \sum_{r(i)=n-m+1}^{r_{\max, \alpha}} \binom{N}{n}^{-1} \binom{r(i)-1}{n-m} \binom{N-r(i)}{m-1} = 1 - \frac{\alpha}{2}, \quad (26)$$

$$P[r(n-m) \leq r_{\min, \alpha}] = \sum_{r(i)=n-m}^{r_{\min, \alpha}} \binom{N}{n}^{-1} \binom{r(i)-1}{n-m-1} \binom{N-r(i)}{m} = \frac{\alpha}{2}, \quad (27)$$

$$1 + \frac{1}{2N} \frac{r_{\max, \alpha}}{N} \leq p \leq 1 + \frac{1}{2N} \frac{r_{\min, \alpha}}{N}. \quad (28)$$

As example we consider situation when $N=100$, $n=30$, $m=3$ and $\alpha=0.05$. Then we get $r_{\max, \alpha}=97$, $r_{\min, \alpha}=76$, $0.035 \leq p \leq 0.245$. For comparison correspondent Clopper-Pearson confidence interval has another endpoints $0.021 \leq p \leq 0.266$, so the width of the "exact" confidence interval is greater than the width of the interval (26), (27) for a population of known size. Moreover the Monte Carlo algorithm, being applied directly to (22), (23), allows to get more valuable information about properties of distribution of random values $\{p_{\min, k}\}_{k=1}^K$ and $\{p_{\max, k}\}_{k=i}^K$:

$$p_{\min, k} = \frac{N+1}{N} \left\{ V_{\min, k} - \frac{1}{2(N+1)} \right\},$$

$$p_{\max, k} = \frac{N+1}{N} \left\{ V_{\max, k} - \frac{1}{2(N+1)} \right\},$$

which define the endpoints of confidence interval

$$p_{\min, K} \frac{\alpha}{2} \leq p \leq p_{\max, K} \left(1 - \frac{\alpha}{2}\right).$$

Actually, under $K=10^6$ statistical trials the correspondent endpoints were found $p_{\min, 25000}=0.035$ and $p_{\max, 975000}=0.245$. At the same time it was happened that set of quantiles from the empirical distribution function have the same value $p_{\min, 24446 \leq k \leq 77448}=0.035$. Then one can conclude that level of significance for the lower endpoint is less then $\alpha=0.05$ and it can be estimated as $\alpha=2 \cdot 24446 : K=0.0489$. The same is true for the right boundary of the confidence interval because $p_{\max, 973365 \leq k \leq 980677}=0.245$, then we have to reassess the level of significance for this quantile as $\alpha=2 \cdot (1-980677 : K)=0.0386$. It is also possible to test the nearest quantile $p_{\max, 973364}=0.235$ for the right endpoint. The correspondent level of significance for the interval $0.035 \leq p \leq 0.235$ can be estimated as $\alpha=1+24446 : K-973364 : K=0.051$. It seems that this

interval is in the best degree satisfying a significance level of 0.05. It's worse also pointing that all 10^6 statistical trials gave proportions in the absolute limits $0.025 \leq p \leq 0.525$.

These results are in agreement with notation [10] that for interval estimation of a proportion, coverage probability tends to be too large for "exact" confidence intervals based on inverting the binomial test.

CONCLUSIONS

It is shown that interval estimation of the distribution parameter can be fulfilled by the direct approximation of expected value integral or sum if the upper x_{\max} and lower x_{\min} bounds of random value domain are known. In this model it is assumed that the sample to be drawn divides the domain of a random variable into fixed segments: $[x_{i-1}, x_i]$, if population is infinite, or $[x_{r(i-1)}, x_{r(i)}]$, if population is finite. At the same time, the statistical weights of the segments, $p_i = f_i - f_{i-1}$ or $P_i = [r(i) - r(i-1)] \cdot (N+1)^{-1}$, and, therefore, the distribution parameter are considered as random variables, which resembles the Bayesian approach [11]. However, this all the similarity ends. The model of the expected value approximation does not need hypotheses about a priori distribution of the parameter, since the probability distributions of these statistical weights are known if the sample satisfies the i.i.d. conditions. It is easily implemented numerically using the Monte Carlo method, and it does not impose any restrictions on the sample size. In contrast to the bootstrap method, the model is formally operable for samples of any size $n \geq 1$.

It is clear that the possibilities of practical application of the model are directly related to the availability of information regarding the boundaries of the domain of the variable, the distribution parameter of which is estimated. The practical attractiveness of the described approach is stipulated for the circumstance that some measurable properties of the physical, biological and social populations have known bounds. There may be various options for information support. First of all, it should be pointed out that there are variables with naturally well-defined boundaries. A classic example is the dichotomous population with a variable taking two possible values. Another example is the correlation coefficient K . If there are no expectations, then the natural boundaries for the correlation coefficient are $-1 \leq K \leq 1$, if there is confidence in a positive relationship, then $0 \leq K \leq 1$, and $-1 \leq K \leq 0$ in opposite case.

Generally speaking, if the problem concerns the extreme values of the observed variable in a large system, then one can expect that the addresses of such "extreme" elements in the system are known and this makes it possible to target these extreme values. Note that the use of the model does not require knowledge of the exact upper and lower bounds of the variable domain. These boundaries, if necessary, can be assigned with a margin. For example, if it is known that the upper bound exceeds the lower bound by hundreds or more times, then the lower bound can be set equal to zero. Obviously, this will lead to a slight broadening of the confidence

interval. Furthermore, the special options could be envisaged in the frames of the sampling plan in order to find appropriate population elements and to estimate the measured variable bounds. These are just the cases, when the described method of expected value approximation could be applied.

REFERENCES

1. A. Bevan. *Statistical Data Analysis for the Physical Sciences*. Queen Mary University of London, 2013, 229 p.
2. R.A. Fisher. On the mathematical foundations of theoretical statistics // *Philosophical Transactions of the Royal Society*. 1922, A. 222 (594-604), p. 309-368.
3. B. Efron. Bootstrap methods: another look at the jackknife // *The Annals of Statistics*. 1979, v. 7, № 1, p. 1-26.
4. B. Efron. The bootstrap and modern statistics // *Journal of the American Association*. 2000, v. 95, № 452, p. 1293-1296.
5. N. Metropolis, S. Ulam. The Monte Carlo Method // *Journal of the American Statistical Association*. 1949, v. 44, № 247, p. 335-341.
6. T.G. DiCiccio, B. Efron. Bootstrap confidence intervals (with Discussion) // *Statistical Science*. 1996, v. 11, № 3, p. 189-228.
7. D.B. Rubin. The Bayesian Bootstrap // *The Annals of Statistics*. 1981, v. 9, № 1, p. 130-134.
8. D.P. Kroese, T. Taimre, Z.I. Botev. *Handbook of Monte Carlo Methods*. New York, John Wiley and Sons. 2011, 772 p.
9. C. Clopper, E.S. Pearson. The use of confidence or fiducial illustrated in the case of binomial // *Biometrika*. 1934, v. 26 (4), p. 404-413.
10. A. Agresti, Alan, B. Coull, Brent. Approximate Is Better than "Exact" for Interval Estimation of Binomial Proportions // *The American Statistician*. 1998, v. 52 (2), p. 119-126.
11. J.A. Bernardo, Adrian F.M. Smith. *Bayesian Theory*. New York, John Wiley and Sons. 1994, 586 p.

Article received 04.10.2019

ИНТЕРВАЛЬНАЯ ОЦЕНКА ПАРАМЕТРА РАСПРЕДЕЛЕНИЯ СТАТИСТИЧЕСКИМИ ИСПЫТАНИЯМИ ОЖИДАЕМОЙ ВЕЛИЧИНЫ

В.А. Баранник

Получены интервальные оценки параметров распределения аппроксимацией ожидаемых значений бесконечной или конечной генеральной совокупности с известными границами. Аналогично методу Байеса параметры распределения интерпретируются как случайные переменные, и их неопределенность выражается в терминах распределений. Для нахождения границ доверительного интервала используется метод Монте-Карло. Модель не накладывает каких-либо ограничений на вид распределений. В отличие от других непараметрических интервальных оценок параметров распределений модель работает с выборками любого размера.

ИНТЕРВАЛЬНА ОЦІНКА ПАРАМЕТРА РОЗПОДІЛУ СТАТИСТИЧНИМИ ВИПРОБУВАННЯМИ ОЧІКУВАНОЇ ВЕЛИЧИНИ

В.О. Бараннік

Отримано інтервальні оцінки параметрів розподілу апроксимацією очікуваних значень нескінченної або скінченної генеральної сукупності з відомими границями. Аналогічно методу Байеса параметри розподілу розглядаються як випадкові величини, а їх невизначеність виражається в термінах розподілу. Для знаходження границь довірчого інтервалу застосовується метод Монте-Карло. Модель не накладає будь-яких обмежень на вид розподілів. На відміну від інших непараметричних інтервальних оцінок параметрів розподілу модель працює з вибірками будь-якого розміру.