ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 8, 2004

Applications of resampling approach to statistical problems of logical systems

Alexander Andronov and Maxim Fioshin

Abstract. The resampling approach to the problem of confidence interval construction is considered. An efficiency of this approach is investigated for logical system characteristics. Proposed method allows to calculate the probability that obtained interval covers true value of characteristic. Empirical results show that true covering probability is close to appointed value.

1. Introduction

Confidence intervals are main subject of mathematical statistics and reliability theory. Last years the bootstrap approach has been applied for confidence interval construction (Davison and Hinkley, 1997; DiCiccio and Efron, 1996). In this paper the resampling approach is used for this aim (see Andronov and Merkuryev, 2000; Andronov, 2001). As an example, a confidence interval calculation for characteristics of logical system is considered. Such system can be described in the following way.

Let $M = \{1, 2, \ldots, m\}$ be a set of integers, let I_l and I^r, $l, r \in \{1, 2, \ldots\}$, be subsets of M. Let $f_l(x_i : i \in I_l)$, $l = 1, 2, \ldots$, be real-valued functions of real variables $\{x_i : i \in I_l\}$, which define the $s(l)$-th order-statistics ($s(l) = 1, 2, \ldots$) of $\{x_i : i \in I_l\}$:

$$f_l(x_i : i \in I_l) = \min\{x_i \in \{x_i : i \in I_l\} : \#\{x_j \in \{x_j : j \in I_l\} : x_i \geq x_j\} \geq s(l)\},$$

where $\#A$ is the cardinal number of a set A.

In particular, if $s(l) = 1$ then $f_l(x_i : i \in I_l) = \min\{x_i : i \in I_l\}$; if $s(l) = \#I_l$ then $f_l(x_i : i \in I_l) = \max\{x_i : i \in I_l\}$.

Received October 26, 2003.
2000 *Mathematics Subject Classification.* 62G09, 62G15, 62G30, 62N05.
Key words and phrases. Resampling, nonparametric estimation, simulation, confidence intervals.

Let us also define predicates $P_1^{r_1}, P_2^{r_2}, \ldots$, where upper index r_i means that the range of definition of P_i is $R^{r_i} : P_j^r(x_i : i \in I_j^r)$. We consider predicates of the following types:

- "less (greater) than": $P^2(x_1, x_2) =$ "$x_1 \leq x_2$"; $P^{r+1}(x, x_1, \ldots, x_r) =$ "$x \leq \min\{x_1, x_2, \ldots, x_r\}$";
- "equal": $P^2(x_1, x_2) =$ "$x_1 = x_2$"; $P^r(x_1, \ldots, x_r) =$ "$x_1 = x_2 = \ldots = x_r$".

Note that we are allowed to use constants from $\mathbb{R}$, real variables $x, x_1, \ldots$ and functions $f_1, f_2, \ldots$ as arguments of the predicates.

Also, we consider logical functions over predicates, for example, disjunction $\vee$, conjunction &, negation $\neg$ and "at least l from k":

$$P^k(y_1, y_2, \ldots, y_k) = \begin{cases} 1 & \text{if } y_1 + y_2 + \ldots + y_k \geq l, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $y_i \in \{0, 1\}$.

Now we construct formula $Q(x_1, x_2, \ldots, x_m)$. It is a complex predicate that can contain primary predicates and logical functions.

In fact, arguments of the formula are independent random variables $X_1, X_2, \ldots, X_m$ with unknown distribution functions $F_1, F_2, \ldots, F_m$. Our aim is to estimate expectation of the formula Q:

$$E\,Q(X_1, \ldots, X_m) = \int \ldots \int Q(x_1, \ldots, x_m) dF_1(x_1) \ldots dF_m(x_m) \tag{2}$$

on the basis of sample populations $\{\mathbf{X}_i = \{X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}\}, i = 1, \ldots, m\}$.

Note that various problems of logical control, reliability etc. are described by this model. We use resampling approach for our aim. The point estimation of parameters of interest has been considered in previous papers of authors. Now we consider the problem of interval estimation. We use the approach discussed in (Andronov, 2002).

Our aim is to construct a confidence interval for the expectation $\theta = E\,Q(X_1, X_2, \ldots, X_m)$, that corresponds to the confidence probability γ. A usual resampling approach realizes the following procedure (Gentle, 2002). We produce a series of experiments. Each experiment includes k trials. During the ν-th trial, $\nu = 1, 2, \ldots, k$, an element $X_{j_i(\nu)}^{(i)}$ is selected at random, with replacement or without replacement, from each sample $\mathbf{X}_i$, $i = 1, 2, \ldots, m$. After k trials we calculate empirical mean for the current, for example, l-th, experiment:

$$\theta_l = \frac{1}{k} \sum_{\nu=1}^{k} Q(X_{j_1(\nu)}^{(1)}, X_{j_2(\nu)}^{(2)}, \ldots, X_{j_m(\nu)}^{(m)}) \,. \tag{3}$$

Then we return all extracted elements into corresponding sample populations and repeat the described experiment r times, obtaining the sequence

$\theta_1, \theta_2, \ldots, \theta_r$. It gives order statistics $\theta_{(1)}, \theta_{(2)}, \ldots, \theta_{(r)}$ and corresponding α-quantile $\theta_{\lfloor \alpha r \rfloor}$ of their distribution, where $\lfloor \alpha r \rfloor = \max\{\xi = 1, 2, \ldots : \xi \leq \alpha r\}$. We set $\alpha = 1 - \gamma$ and accept $(\theta_{(\lfloor \alpha r \rfloor)}, \infty)$ as upper γ-confidence interval for the true value of θ.

Note that a confidence interval bound can depend on several order statistics, not only on one. For example, Gürtler and Henze (Gürtler and Henze, 2000) propose to use $\alpha\theta_{(\lfloor \alpha r \rfloor)} + (1-\alpha)\theta_{(\lfloor \alpha r \rfloor - 1)}$ as upper $(1-\alpha)$-confidence interval bound.

In this paper the efficiency of described approach is investigated. More exactly, our aim is to calculate the true value of the covering probability $R = P\{\theta_{(\lfloor \alpha r \rfloor)} \leq \theta\}$. For this we need to know the joint distribution of $\theta_1, \theta_2, \ldots, \theta_r$. We suppose that θk and αr are integer numbers. If mentioned numbers are not integer, the randomization can be applied to calculate intermediate values.

2. General approach

The function of our interest is described by the formula $Q(X_1, X_2, \ldots, X_m)$. It depends on m real arguments $X_1, X_2, \ldots, X_m$ and has two possible results only: one and zero. In fact, the value of the function depends only on the ordering of the arguments. For each m-dimensional real vector $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ let us define an ordered vector $(x_{(1)}, x_{(2)}, \ldots, x_{(m)})$ and corresponding permutation $\mathbf{p}(\mathbf{x}) = (j_1, j_2, \ldots j_m)$. Here $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(m)}$, $\{x_1, x_2, \ldots, x_m\} = \{x_{(1)}, x_{(2)}, \ldots, x_{(m)}\}$, and if $j_i = \nu$, then $x_i = x_{(\nu)}$. Now we are able to consider our function $Q(x_1, x_2, \ldots, x_m)$ as a function of permutations $\mathbf{p} \in \mathbf{\Pi}$, where $\mathbf{\Pi}$ is the set of all permutations of elements $1, 2, \ldots, m$. In order to illustrate the dependence of the formula on $\mathbf{p}$, we will write $\widetilde{Q}(\mathbf{p})$. We denote $\mathbf{\Pi}_1 = \{\mathbf{p} \in \mathbf{\Pi} : \widetilde{Q}(\mathbf{p}) = 1\}$, $\mathbf{\Pi}_0 = \{\mathbf{p} \in \mathbf{\Pi} : \widetilde{Q}(\mathbf{p}) = 0\}$. Therefore our parameter of interest can be written as

$$\theta = P\{\mathbf{p} \in \mathbf{\Pi}_1\}. \tag{4}$$

Our aim is to construct an upper confidence interval $(\hat{\theta}, 1)$ for θ on the basis of sample populations $\mathbf{X}_i = \{X_1^{(i)}, X_2^{(i)}, \ldots, X_{n_i}^{(i)}\}$, $i = 1, 2, \ldots, m$. Note that the total sample space is $\mathbb{R}^n$ where $n = n_1 + n_2 + \ldots + n_m$.

How can we describe the total sample $\mathbf{X}_1 \cup \mathbf{X}_2 \cup \ldots \cup \mathbf{X}_m$ after its ordering? Let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be ordered sequence of elements of $\mathbf{X}_1 \cup \mathbf{X}_2 \cup \ldots \cup \mathbf{X}_m$. It is possible to use n-dimensional vector $\mathbf{W} = (W_1, W_2, \ldots, W_n)$ where $W_j \in \{1, 2, \ldots, m\}$, and $W_j = i$ means that the element $X_{(j)}$ belongs to $\mathbf{X}_i$. It is possible that neighboring members of $\mathbf{W}$, say W_j and W_{j+1}, have equal values. Therefore we can optimize this representation if we use a vector $\mathbf{V} = (V_1, V_2, \ldots)$ instead of $\mathbf{W}$. Each element V_j of the vector $\mathbf{V}$ is a pair (V_j', V_j''), where V_j' is the same as W_j, and V_j'' is the number of consequent repeatings of value V_j'.

For example, let $m = 3$, $\mathbf{X}_1 = \{2.5, 6.3, 1\}$, $\mathbf{X}_2 = \{0.5, 2.1, 5.3, 5.2, 0.9\}$, $\mathbf{X}_3 = \{6.1, 2.3\}$. Then $n = 10$, and ordered sequence is $\{0.5, 0.9, 1, 2.1, 2.3, 2.5, 5.2, 5.3, 6.1, 6.3\}$, $W = \{2, 2, 1, 2, 3, 1, 2, 2, 3, 1\}$, $V = \{(2,2), (1,1), (2,1), (3,1), (1,1), (2,2), (3,1), (1,1)\}$.

Another way to describe this ordering is the protocol notion introduced in (Andronov, 2002). The definition below generalizes the protocol notion for our case.

Definition 1. Let $(x_1^{(1)}, x_2^{(1)}, \ldots, x_{n_1}^{(1)})$, $(x_1^{(2)}, x_2^{(2)}, \ldots, x_{n_2}^{(2)})$, $\ldots$, $(x_1^{(m)}, x_2^{(m)}, \ldots, x_{n_m}^{(m)})$ be real-valued vectors, and $\mathbf{x}_1 = (x_{(1)}^{(1)}, x_{(2)}^{(1)}, \ldots, x_{(n_1)}^{(1)})$, $\mathbf{x}_2 = (x_{(1)}^{(2)}, x_{(2)}^{(2)}, \ldots, x_{(n_2)}^{(2)}), \ldots$, $\mathbf{x}_m = (x_{(1)}^{(m)}, x_{(2)}^{(m)}, \ldots, x_{(n_m)}^{(m)})$ corresponding ordered sequences: $x_{(1)}^{(i)} \leq x_{(2)}^{(i)} \leq \ldots \leq x_{(n_i)}^{(i)}$, $i = 1, 2, \ldots, m$. We call (n_2+1)-dimensional vector $\mathbf{C}(1) = (c_0(1), c_1(1), \ldots, c_{n_2}(1))$, $c_0(1) + c_1(1) + \ldots + c_{n_2}(1) = n_1$, *a subprotocol of the first level*, where $c_j(1) = \#\{x_\nu^{(1)} \in \mathbf{x}_1 : x_{(j)}^{(2)} < x_\nu^{(1)} \leq x_{(j+1)}^{(2)}\}$, $j = 0, 1, \ldots, n_2$, $x_{(0)}^{(2)} = -\infty$, $x_{(n_2+1)}^{(2)} = \infty$. *A subprotocol of the l-th level* $\mathbf{C}(l)$, $l = 2, 3, \ldots, m-1$, is determined analogously using the union $\mathbf{x}_1 \cup \mathbf{x}_2 \cup \ldots \cup \mathbf{x}_l$ instead of $\mathbf{x}_1$, and $\mathbf{x}_{l+1}$ instead of $\mathbf{x}_2$. We call a sequence of subprotocols *a protocol* $\mathbf{C} = (\mathbf{C}(1), \mathbf{C}(2), ..., \mathbf{C}(m-1))$.

For the previous example we have $\mathbf{C}(1) = (0, 0, 1, 1, 0, 1)$, $\mathbf{C}(2) = (4, 3, 1)$, $\mathbf{C} = (\mathbf{C}(1), \mathbf{C}(2))$. Obviously, there is an one-to-one correspondence between the protocol $\mathbf{C}$ and vector $\mathbf{W}$ (or $\mathbf{V}$). Often we prefer the protocols, because they can be recursively calculated. In general, we use both notions.

Also each point $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ of the sample space $\mathbb{R}^n$ can be described by vectors $\mathbf{W}$ and $\mathbf{V}$ or the protocol $\mathbf{C}$.

Now we must find the following:

- the algorithm for enumeration of all protocols,
- the probability $P_\mathbf{C}$ to get a fixed protocol $\mathbf{C}$,
- the conditional probability $q_\mathbf{C}$ of the event $\{Q(X_1, X_2, \ldots, X_m) = 1\}$ given the protocol $\mathbf{C}$,
- the conditional probability $\rho_\mathbf{C}$ of the event $\{\theta_l < \theta\}$ for the l-th experiment with fixed protocol $\mathbf{C}$,
- conditional cover probability $R_\mathbf{C}$ given the protocol $\mathbf{C}$.

Then the unconditional coverage probability can be calculated as follows:

$$R = \sum_{\mathbf{C}} P_\mathbf{C} R_\mathbf{C}. \tag{5}$$

The first problem is a standard combinatorial problem (Brualdi, 1999), so we do not consider it. The other problems are considered in the following sections.

3. Probability distributions of protocols

Our task is to calculate the probability to get a fixed protocol $\mathbf{C}$. We assume that $X_1, X_2, \ldots, X_m$ are independent continuous random variables with distribution functions $F_1(x), F_2(x), \ldots, F_m(x)$.

To solve this problem, it is convenient to consider the vector $\mathbf{W} = (W_1, W_2, \ldots, W_n)$ instead of the protocol $\mathbf{C}$. Let $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$ be ordered sequence of the original data. We denote via $S_i(j) = \#\{W_\nu \in \{W_1, W_2, \ldots, W_{j-1}\} : W_\nu = i\}$ the number of elements from $\mathbf{X}_i$ that are less than $X_{(j)}$. We consider the extraction procedure in different time moments. The time moment $X_{(j)}$ is called j-th point of the protocol. Let $U_{\mathbf{C}}(j,t)$ be the probability that at the time moment t the protocol $\mathbf{C}$ is realized till the j-th point, where $t \in (0,\infty)$, $j = 1, 2, \ldots, n$. Note that the probability to get $\mathbf{C}$-protocol is equal to $P_{\mathbf{C}} = U_{\mathbf{C}}(n,\infty)$.

The probabilities $U_{\mathbf{C}}(j,t)$, $j = 1, 2, \ldots, n$, are calculated recurrently. Obviously,

$$U_{\mathbf{C}}(1,t) = n_{W_1} \int_0^t \prod_{i \neq W_1} (1 - F_i(z))^{n_i} (1 - F_{W_1}(z))^{n_{W_1}-1} dF_{W_1}(z), \quad t > 0\,. \tag{6}$$

Other values of $U_{\mathbf{C}}(j,t)$, $j = 2, 3, ..., n$, $t > 0$, are calculated by the formula

$$\begin{aligned} U_{\mathbf{C}}(j,t) &= \left(n_{W_j} - S_{W_j}(j)\right) \int_0^t \int_z^t \prod_{i \neq W_j} \left(\frac{1 - F_i(y)}{1 - F_i(z)}\right)^{n_i - S_i(j)} \\ &\times \frac{1}{1 - F_{W_j}(z)} \left(\frac{1 - F_{W_j}(y)}{1 - F_{W_j}(z)}\right)^{n_{W_j} - S_{W_j}(j) - 1} dF_{W_j}(y) dU_{\mathbf{C}}(j-1, z)\,. \end{aligned} \tag{7}$$

If the exponential case takes place then it is possible to get an explicit formula for the probability of the $\mathbf{C}$-protocol. Let

$$F_i(x) = 1 - e^{-\lambda_i x}, \quad x > 0, \quad i = 1, 2, \ldots, m.$$

Then, by using memoryless property of the exponential distribution, we have

$$P_{\mathbf{C}} = \prod_{j=1}^{n} \frac{\lambda_{W_j}(n_{W_j} - S_{W_j}(j))}{\sum_{i=1}^{m} \lambda_i (n_i - S_i(j))}\,. \tag{8}$$

4. Conditional probability of unit value for fixed protocol

The protocol $\mathbf{C}$ corresponds to sample populations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$ that contains n_1 elements from $\mathbf{X}_1$, n_2 elements from $\mathbf{X}_2$, ..., n_m elements from $\mathbf{X}_m$. During resampling procedure we extract elements from $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m$, one by one, and form a resample (subsample) $(X_1, X_2, \dots, X_m)$. Some resamples give unit value of the formula, $Q(X_1, X_2, \dots, X_m) = 1$, other resamples give zero value. We will calculate the probability of event $\{Q(X_1, X_2, \dots, X_m) = 1\}$ for a fixed protocol $\mathbf{C}$.

Resamples are formed at random, therefore the conditional probability is equal to the ratio of two numbers: the number $h_1(\mathbf{C})$ of resamples, which give unit value of formula, and the number $h_\Sigma(\mathbf{C})$ of all resamples:

$$q_{\mathbf{C}} = \frac{h_1(\mathbf{C})}{h_\Sigma(\mathbf{C})}\,. \tag{9}$$

If each variable X_i, $i = 1, 2, \dots, m$, enters into formula Q one time only, then the number of all resamples $h_\Sigma(\mathbf{C}) = n_1 n_2 \dots n_m$, and therefore

$$q_{\mathbf{C}} = \frac{h_1(\mathbf{C})}{n_1 n_2 \dots n_m}\,. \tag{10}$$

The calculation of $h_1(\mathbf{C})$ is more difficult. In general, it is necessary to consider each permutation from $\mathbf{\Pi}_1 = \{\mathbf{p} \in \mathbf{\Pi} : \widetilde{Q}(\mathbf{p}) = 1\}$. If $\mathbf{p} \in \mathbf{\Pi}_1$, then we must calculate the number $h_{\mathbf{p}}(\mathbf{C})$ of resamples that correspond to this permutation. Then

$$h_1(\mathbf{C}) = \sum_{\mathbf{p} \in \mathbf{\Pi}_1} h_{\mathbf{p}}(\mathbf{C})\,. \tag{11}$$

Direct use of this formula requires a huge amount of calculation, therefore we consider a special case where the amount of calculation can be decreased.

5. Special case

The general case considered above can only be realized for small dimensions of the problem because the number of protocols is too large. We will consider a partial case, where the dimension can be decreased. In this case separate predicates do not depend on all variables but only on a part of them. It allows us to use another type of protocols which are shorter than those given in Definition 1.

Let us consider a predicate $P^m =$ "$X_m < \min\{X_1, X_2, \dots, X_{m-1}\}$". We can interpret it as follows (Gertsbakh, 2000): X_m is the operation period of a system, $X_1, X_2, \dots, X_{m-1}$ are lifetimes of system elements. Considered inequality means reliability of the system during its operation period. In this case the ordering between $X_1, X_2, \dots, X_{m-1}$ is not important. It allows us to use shorter protocol that contains only one subprotocol of the $(m-1)$-th level. Let $\mathbf{x}_{(m)} = (x^{(m)}_{(1)}, x^{(m)}_{(2)}, \dots, x^{(m)}_{(n_m)})$ be ordered sequence of the values of the variable X_m. We now define a protocol as $(n_m + 1)$-dimensional vector $\mathbf{C} = \mathbf{C}(m-1) = (\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{n_m})$, where $\mathbf{c}_j = (c_{j,1}, c_{j,2}, \dots, c_{j,m-1})$ is $(m-1)$-dimensional vector with elements $c_{j,i} = \#\{x^{(i)}_\nu \in x_i : x^{(m)}_{(j)} < x^{(i)}_\nu < x^{(m)}_{(j+1)}\}$,

$j = 0, 1, \ldots, n_m$, $x_{(0)}^{(m)} = -\infty$, $x_{(n_m+1)}^{(m)} = \infty$, $c_{0,i} + c_{1,i} + \ldots + c_{n_m,i} = n_i$, $i = 1, 2, \ldots, m-1$. The total number of such protocols is equal to

$$\prod_{i=1}^{m-1} \binom{n_m + n_i}{n_i}. \tag{12}$$

Let us consider a fixed protocol $\mathbf{C} = \mathbf{C}(m-1) = (\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{n_m})$. For a fixed value $x_{(j)}^{(m)}$, $j = 1, 2, ..., n_m$, of the random variable X_m we have $\sum_{\nu=0}^{j-1} c_{\nu,i}$ values of X_i that are less than $x_{(j)}^{(m)}$. Therefore, if the value of X_m equals $x_{(j)}^{(m)}$, then the number of resamples $(X_1, X_2, \ldots, X_m)$ which give unit value of predicate P^m is calculated by the formula:

$$\prod_{i=1}^{m-1} \left(\sum_{\nu=j}^{n_m} c_{\nu,i} \right). \tag{13}$$

Finally, we have

$$h_1(\mathbf{C}) = \sum_{j=1}^{n_m} \prod_{i=1}^{m-1} \left(\sum_{\nu=j}^{n_m} c_{\nu,i} \right). \tag{14}$$

Now it is possible to calculate the probability $q_{\mathbf{C}}$ by the formula (10).

In order to calculate the probability $P_{\mathbf{C}}$ of the protocol $\mathbf{C}$ we can modify a formula from (Andronov, 2002). Using notations from Section 3, we have the following formula for the first point of the protocol:

$$U_{\mathbf{C}}(1,t) = \int_0^t n_m (1 - F_m(x))^{n_m - 1}$$
$$\times \prod_{i=1}^{m-1} \binom{n_i}{c_{0,i}} (F_i(x))^{c_{0,i}} (1 - F_i(x))^{n_i - c_{0,i}} dF_m(x).$$

Let $\tilde{c}_{0,i} = n_i$, $\tilde{c}_{j,i} = n_i - S_i(j) = n_i - \sum_{\nu=0}^{j-1} c_{\nu,i} = \tilde{c}_{j-1,i} - c_{j-1,i}$. Then we have for $j = 2, 3, \ldots, n_m$:

$$U_{\mathbf{C}}(j,t) = \int_0^t (n_m - j + 1) \int_x^t \left(1 - \frac{F_m(y) - F_m(x)}{1 - F_m(x)}\right)^{n_m - j} \frac{1}{1 - F_m(x)}$$
$$\times \prod_{i=1}^{m-1} \binom{\tilde{c}_{j-1,i}}{c_{j-1,i}} \left(\frac{F_i(y) - F_i(x)}{1 - F_i(x)}\right)^{c_{j-1,i}}$$
$$\times \left(1 - \frac{F_i(y) - F_i(x)}{1 - F_i(x)}\right)^{\tilde{c}_{j-1,i} - c_{j-1,i}} dF_m(y) dU_{\mathbf{C}}(j-1, x).$$

Remind that the probability $P_{\mathbf{C}} = U_{\mathbf{C}}(n_m, \infty)$.

If elements have exponential distribution, then $P_{\mathbf{C}}$ can be calculated using the following formula:

$$P_{\mathbf{C}} = \prod_{j=0}^{n_m-1} \int_0^\infty (n_m-j)\lambda_m e^{-\lambda_m(n_m-j)t} \prod_{i=0}^{m-1} \left(\begin{array}{c} \tilde{c}_{j,i} \\ c_{j,i} \end{array} \right) (1-e^{-\lambda_i t})^{c_{j,i}} e^{-\lambda_i t(\tilde{c}_{j,i}-c_{j,i})} dt\,.$$

6. Conditional coverage probability

The conditional probability of the event $\theta_l < \theta$ can be calculated by the formula

$$\rho_{\mathbf{C}} = P_{\mathbf{C}}\{\theta_l < \theta\} = \sum_{\xi=0}^{\theta k-1} \left(\begin{array}{c} k \\ \xi \end{array} \right) q_{\mathbf{C}}^{\xi}(1-q_{\mathbf{C}})^{k-\xi}\,. \qquad (15)$$

Now we can find conditional probability to cover the true value of θ:

$$R_{\mathbf{C}} = P_{\mathbf{C}}\{\theta_{(\lfloor \alpha r \rfloor)} \leq \theta\} = \sum_{\xi=\lfloor \alpha r \rfloor}^{r} \left(\begin{array}{c} r \\ \xi \end{array} \right) \rho_{\mathbf{C}}^{\xi}(1-\rho_{\mathbf{C}})^{r-\xi}\,. \qquad (16)$$

7. Numerical example

As formula Q from (2) we consider a predicate P^m = "$X_m < \min\{X_1, X_2, \ldots, X_{m-1}\}$". Let $m = 3$, which means that we have a predicate P^3 = "$X_3 < \min\{X_1, X_2\}$". Let variables X_1, X_2 and X_3 have exponential distribution with parameters λ_1, λ_2 and λ_3. Then

$$\theta = P\{X_3 < \min(X_1, X_2)\} = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}\,.$$

Let $\lambda_1 = 3$, $\lambda_2 = 3$, $\lambda_3 = 2$. In this case true value of $\theta = 0.25$. Let also the number of experiments $r = 10$ and the number of trials in each experiment $k = 16$. We will use a protocol described in Section 5 for the calculation of the coverage probability R according to the formula (5).

The results are presented in Table 1.

	Coverage probability R				
(n_1, n_2, n_3)	γ=0.5	γ=0.6	γ=0.7	γ=0.8	γ=0.9
(3,3,3)	0.533	0.576	0.625	0.686	0.770
(9,9,3)	0.519	0.571	0.630	0.701	0.793
(4,4,4)	0.521	0.578	0.640	0.709	0.797
(6,6,4)	0.516	0.576	0.642	0.715	0.807
(5,5,5)	0.515	0.579	0.646	0.722	0.817
(3,3,8)	0.516	0.581	0.651	0.728	0.823
(4,4,7)	0.512	0.580	0.652	0.732	0.830

TABLE 1. Actual coverage probabilities

8. Conclusion

In this paper we applied the resampling approach to construct a confidence interval for logical system characteristics. The precision of constructed interval has been investigated. The numerical results show that actual coverage probability is close to appointed value even if sample sizes are small.

Acknowledgement. We are very thankful to the referee for a number of corrections which helped to improve the article.

References

Andronov, A. and Merkuryev, Y. (2000). Optimisation of statistical sample sizes in simulation. *J. Statist. Plann. Inference* **85**, 93–102.

Andronov, A. (2001). On moments of empirical mean of random variable function. In: *The 10th International Symposium on Applied Stochastic Models and Data Analysis*, Compegne, France, 106–111.

Andronov, A. (2002). On resampling approach to a construction of approximate confidence intervals for system reliability. In: *Third International Conference on Mathematical Methods in Reliability. Methodology and Practice.* June 17–20, 2002, Trondheim, Norway, Norwegian University of Science and Technology, 39–42.

Brualdi, R. A. (1999). *Introductory Combinatorics*. Prentice Hall, New Jersey.

Davison, A. C. and Hinkley, D V. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press, Cambridge.

DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science* **11(3)**, 189–228.

Feigin, P. D. and Lumelskii, Y. P. (2001). On Monte-Carlo simulation of confidence bounds for reliability problems. In: *Proceedings of 15-th European Simulation Multiconference*, SCS, Prague, 719–721.

Gentle, J. E. (2002). *Elements of Computational Statistics*. Springer, New York.

Gertsbakh, I. (2000). *Reliability Theory: with Applications to Preventive Maintenance.* Springer, Berlin.

Gürtler, N. and Henze, N. (2000). Recent and classical goodness-of-fit test for the Poisson distribution. *J. Statist. Plann. Inference* **90**, 207–225.

Riga Technical University, Kalku 1, LV-1658 Riga, Latvia
E-mail address: lora@mailbox.riga.lv
E-mail address: mf@tsi.lv