Mathematics and Statistics 1(2): 53-58, 2013
DOI: 10.13189/ms.2013.010207 http://www.hrpub.org

0-event: Point and Interval Estimates

Sergey Gurov

Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991, Russia

*Corresponding Author: sgur@cs.msu.ru

Abstract Point and interval probability estimates for an event that has never been observed in a Bernoulli trial series (0-event) are proposed and validated. In this case, the classical statistical methods yield a zero point estimate, which is often unacceptable in practice. Nonzero point and interval probability estimates for a 0-event are proposed and validated. A classification of samples by size for the case of a 0-event is proposed.

Keywords Mathematical statistics, Point Estimator, 0-event, Interval Estimator, Consistency Principle, Small Sample

1 Introduction. Problem formulation

Estimation of a nonrandom but unknown probability p of a certain random event X occurring in a single test is considered. The random variable of the number of successes $m \in \{0, 1, \ldots, n\}$ in $n > 0$ Bernoulli trials has the binomial distribution $Bi_m(n, p) = \binom{n}{m} p^m (1-p)^{n-m}$, $p \in \Theta$, where $\Theta = (0, 1)$ is the range of the parameter p (open one-dimensional interval).

The maximum likelihood point estimator $\widehat{p}_{ml}$ of p is given by the classical formula (the last equality below) for computing probabilities, which was proposed as early as in the 17th century:

$$\widehat{p}_{ml} = \arg\max_{p \in \overline{\Theta}} L(p, x) = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{m}{n}. \quad (1)$$

Here,

$L(p, x) = L(p \mid m, n) = p^m(1-p)^{n-m}$ is the likelihood function for the binomial statistical model, where $x = (x_1, \ldots, x_n)$ is the sample obtained as a result of performing n elementary independent experiments of observing the event X $(x_i \in \{0, 1\}, i = \overline{1, n})$, where 1 occurs in x m times and 0 occurs $n - m$ times (as usual, 1 means observation, and 0 means that the event did not occur in this experiment);

$\overline{\Theta} = [0, 1]$ is the closure of the set Θ.

This estimator is unbiased, efficient, and consistent. An unbiased estimator of its variance (vide [17, Example 17.9]) is

$$\frac{m(n-m)}{n^3}. \quad (2)$$

If $m = 0$, we say that a *0-event* has occurred (vide, e.g. [15, Section 4.5.4]). More precisely, by the 0-event we mean the random event X that has never been observed in a series of Bernoulli trials (rather than the fact of obtaining a zero sample as in [10]). In that case, formula (1) yields a zero point estimator of the probability of observing X, and formula (2) yields a zero estimated value of its variance.

However, this estimate $\widehat{p} = 0$ is often unacceptable in practice. This situation occurs, for example, in the construction of correct classifying algorithm: it is clear that the approval of the zero-error algorithm for the classification of all newly imposed objects is too categorically to be true. Author does not know the special works on the subject. This fact was the reason for the research.

In this paper, which is a further development of [9], a nonzero point estimator of a 0-event is proposed and validated.

2 Available estimators

2.1 Frequency approach. Confidence estimation

In the case of a 0-event, the classical statistical methods based on the frequency approach [2, p. 107, Table 5.2], [15, Section 4.5.4, (4.26)] determine the lower bound $p^-(n)$ of the confidence interval with the confidence coefficient η as zero, and the upper bound $p^+(n)$ is determined as a solution (for x) of the equation

$$I_x(1, n) = \eta.$$

Here, $I_x(\cdot, \cdot)$ is the ratio of the incomplete B (beta) Euler to the complete B-function with appropriate parameters:

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1}(1-t)^{b-1} dt,$$

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

($\Gamma(\cdot)$ is the gamma-function). For practical purposes, it is usually assumed that $\eta = 0.95$; other values (0.99 or 0.90) are used much more rarely depending on the specific practical situation. Thus, we have

$$I_x(1,n) = n\int_0^x (1-t)^{n-1}dt = 1-(1-x)^n = \eta,$$

whence

$$x = p^+(n) = 1-\sqrt[n]{1-\eta}.$$

Thus, for $\eta = 0.95$, we have $p^+(10) = 0.2589$ and $p^+(100) = 0.0295$. For $n > 50$, it may be assumed that $p^+(n) \approx 3/n$.

In practice, the use of $p^+(n)$ as a point estimator of p is reasonable if the occurrence of the event X entails significant consequences that require appropriate safety measures (e.g. in the assessment of various risks). Otherwise, such an estimator, which gives a too high probability estimate, almost certainly entails the inequality $p \le p^+$. However, it is not required of the point estimator that this inequality be almost always true.

2.2 Bayesian approach

When the Bayesian approach is used to solve statistics problems, there is a question of providing more details on the *a priori* distribution.

We consider the most interesting situation when no results of earlier experiments are availabile, i.e. when no data for reconstructing the *a priori* distribution is available (the empirical Bayesian approach). In such cases, Laplace's principle of insufficient reason, which establishes that, if nothing is known about a parameter and this parameter varies within a finite interval, then its *a priori* distribution is assumed to be uniform.

As is customary, we will select the *a priori* distribution from the family of conjugate *a priori* distributions [12] formed by the densities of B-distributions (or Bernoulli distributions) (the so-called binomial statistical model)

$$Be_p(a,\, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, p^{a-1}(1-p)^{b-1}$$

with the parameters $a,\, b > 0$. The uniform distribution $U(0,1)$ on the interval $(0,1)$ is the B-distribution with the parameters $a = b = 1$. Since the 0-event likelihood function is $L(p \mid 0,\, n) = (1-p)^n$, the probability density of the a posteriori distribution is $f(p)_{a\_post} = Be_p(1,\, n+1) = (n+1)(1-p)^n \propto L(p \mid 0,\, n)\cdot U(0,1)$. The expectation of the a posteriori distribution thus obtained is

$$\mu = (n+1)\int_0^1 p\,(1-p)^n\,dp = \\ = \frac{I_1(2,n+1)}{n+2} = \frac{1}{n+2}, \quad (3)$$

and the median is $med = 1 - 1/\sqrt[n]{2}$.

A Bayesian point estimator is typically set to be equal to the expectation or the median of the a posteriori distribution because they minimize the root-mean-square losses and the mean deviation, respectively. Thus, we have two estimates

$$\widehat{p}_{B^U_\mu}(n) = \frac{1}{n+2} \quad \text{and} \quad \widehat{p}_{B^U_{med}}(n) = 1-\sqrt[n]{0.5}\,.$$

The estimate $\widehat{p}_{B^U_\mu}(n)$ corresponds to the the so-called Laplace's law of succession [23]. Since $1-\sqrt[n]{0.5} \to \ln 2/n$ as $n \to \infty$, we have $\widehat{p}_{B_{med}}(n) \simeq 1/(1.443\,n)$. Note that the median-based estimator is robuster [22]. A valuable quality of the Bayesian estimators is their independence of the confidence probability concept.

In any case, it is clear that both estimates are too high for not very small n because they are based on the assumption of the uniform *a priori* distribution of p on the interval $(0,1)$, which is hardly consistent with the occurrence of the 0-event.

3 Estimator $\widehat{p}_0$

A 0-event occurs when a *0-sample* $x^0 = (0,\, \ldots,\, 0)$ of length $n \ge 1$ is obtained as a result of n elementary experiments of observing the event X. We assume that no other information about the event X is available and can be obtained.

Below, we use the concept of the confidence coefficient $\eta \in (0,1)$ to estimate the probability p of the occurrence of the event X in a single experiment. Let $\widehat{p}$ be the chosen estimate of the probability p of the event X, and $P(n,\widehat{p})$ be the probability of a certain event related to the observed 0-event; on the basis of the latter event, conclusions are drawn regarding X. We assume that $P = P(n,\widehat{p})$ is greater than the chosen confidence coefficient:

$$P \ge \eta\,. \quad (4)$$

As we do so, we face an unusual relatiion $P(n,\widehat{p}) \to 1$ as $\widehat{p} \to 0$, which is due to the zero estimate of p according to (1). Therefore, the confidence coefficient here (we do not change the terminology) represents not the degree of the event confidence but rather the extent of "compromise" we can afford to obtain an estimate different from the theoretically true but unacceptable (in the specific situation) value. Due to this fact, we are interested in the maximum estimate possible under the given assumptions (the most distant from 0).

We construct two estimates of the probability of the 0-event that are free of those drawbacks and are based on different approaches.

3.1 Estimator $\widehat{p}_\eta$

If the true value of the estimated probability is p, then the the probability P of the 0-event is $P = (1-p)^n$. According to (4), we assume that

$$P = (1-p)^n \ge \eta\,,$$

whence

$$p \le \widehat{p}_\eta = 1-\sqrt[n]{\eta} \simeq \frac{\ln(1/\eta)}{n}\,.$$

3.2 Estimator $\widehat{p}_r$

A random event X observed in a single Bernoulli trial with a probability $p \in (0,1)$ is said to determine the discrete-time random process $\mathfrak{X}_p$ that produces the sample x^0 as its realization.

The idea underlying the estimate $\widehat{p}_r(n)$ is to replace the consideration of the realization x^0 of the process $\mathfrak{X}_p$ by some other realization of this process that contains at least one unit value.

Let us construct the required realization x^1. Consider the process $\mathfrak{X}_q$ determined by the probability q of observing the event X in a single Bernoulli trial, and let x^1 be a realization of this process. Let the size of the sample x^1 be $N \geq 1$ of which $M \geq 1$ values are zero. Next, we use estimate (1). Let us determine the feasible values M and N assuming that the equality $p = q$ holds with the confidence degree no less than η.

To solve the problem thus formulated, we use the exact Fisher test for the comparison of the probabilities underlying two binomial distributions [15, Section 4.6.7]. The method is based on analyzing the so-called 2×2 tables. In our case, we have the table

0	n	n
M	N-M	N
M	N-M+n	N+n

(5)

The use of this criterion is based on the fact that the general criterion for analyzing 2×2 tables can be used only when the table elements are sufficiently large, which is not the case here because one such value is zero.

The probability $P = P(N, M; n)$ that the table is generated by the same value of probability is

$$P = \frac{n!\,N!\,M!\,(N-M+n)!}{(N+n)!} \cdot \frac{1}{n!\,M!\,(N-M)!} = \\ = \frac{N!\,(N-M+n)!}{(N-M)!\,(N+n)!} = \frac{\binom{N}{M}}{\binom{N+n}{M}}. \quad (6)$$

The asymptotics (vide, e.g.[4])

$$\frac{\binom{n-s}{k}}{\binom{n}{k}} \sim \exp\Big\{-\frac{s\,k}{n} - \frac{s^2 k + s k^2}{2\,n^2}\Big\},$$

is known to ve valid for $s + k = o(n^{3/4})$ and $n \to \infty$. In our case, this gives

$$P = \frac{\binom{(N+n)-n}{M}}{\binom{N+n}{M}} \sim \exp\Big\{-\frac{n\,M}{N+n}\Big(1 + \frac{M+n}{N+n}\Big)\Big\}$$

with preserving the representation condition (it is easy to show that $M^2 = o(N)$ as $P \to \max$; hence, $n + M = o((N+n)^{3/2})$ as $N \to \infty$, $n = const$). Then, according to (4), we have

$$-\frac{n\,M}{N+n}\Big(1 + \frac{M+n}{N+n}\Big) \lesssim \ln\eta\,;$$

furthermore, setting $\widehat{p}_r = \dfrac{M}{N}$ in accordance with (1) and assuming that $N \gg 1$, we obtain

$$n\,\widehat{p}_r\,(1 + \widehat{p}_r) \gtrsim \ln\frac{1}{\eta}\,.$$

Therefore, neglecting the term $\widehat{p}_r^2$, we finally obtain $\widehat{p}_r \simeq \dfrac{\ln(1/\eta)}{n} = \widehat{p}_\eta$.

Thus, both estimates are almost identical. We denote this estimate by $\widehat{p}_0$:

$$\widehat{p}_0(n) = 1 - \sqrt[n]{\eta} \simeq \frac{\ln(1/\eta)}{n} \simeq \frac{1-\eta^2}{2\,\eta\,n} \simeq \frac{1-\eta}{\eta\,n}\,; \quad (7)$$

and we propose to regard this estimate as a point estimate of the probability of the 0-event. The above asymptotics (listed in the order of reduced accuracy with overestimation) are valid for the practical values of η and of not too low n.

A somewhat coarser reasoning based on fixing a certain value of N yields (when $P \to \max$)

$$M = 1\,.$$

In such a case, $P = N/(N+n)$, and, due to (4), we have

$$N = \Big]\frac{\eta\,n}{1-\eta}\Big[\,; \quad (8)$$

due to (1), we immediately obtain $p \leq \widehat{p} = M/N = (1-\eta)/(\eta\,n)$, which coincides with (7).

It is clear that, for practical values of η and $n > 3$, we have

$$\widehat{p}_0(n) < \widehat{p}_{B^U_{med}}(n) < \widehat{p}_{B^U_\mu}(n) < p^+(n)\,.$$

4 Interval consistent estimation

The point estimator $\widehat{p}_0$ obtained above makes it possible to give an interval estimate of p based on the consistency principle [8, 9]. This principle based on E. Leman's concept [19, Ch. 4, Section 2, Example 2.7] enables us to specify the *a priori* distribution of the parameter being estimated in the framework of the Bayesian approach . The proposed method is designed specifically for low probabilities of events.

Based on the principle of consistency, the *a priori* distribution is to be chosen, in particular, from the condition of coincidence of the Bayesian and frequency point estimators of the parameter being determined. The resultant *a priori* distribution $f_{a\_priori}(p) = Be_p(1, b)$ (where b is a certain parameter) agrees with the observed 0-event better than the uniform distribution. Furthermore, the a posteriori distribution is $f_{a\_post}(p) = Be_p(1, b+n)$, and the upper bound p_c^+ of the confidence interval $(0,\, p_c^+)$ for the probability p is the solution to the equation

$$I_x(1,\, n+b-1) = \eta\,.$$

By the consistency principle, the parameter b is determined from the condition $\widehat{p} = 1/N = 1/(b+n+1)$; therefore, $b = N - n - 1$. Then, the equation for determining $x = \widehat{p}_c^+$ takes the form

$$I_x(1,\, \frac{1}{\widehat{p}} - 2) = \eta \quad \text{or} \quad I_x(1,\, N-2) = \eta\,; \quad (9)$$

in the latter case, the value of N is taken from (8).

For example, for $\eta = 0.95$ and $n = 10$, we have $N = 190$ and $\widehat{p}_0 = 0.0052$. Equation (9) is specified as $I_x(1, 188) = 0.95$, whence, using Table 5.2 from [2], we obtain $\widehat{p}_c^+ \approx 0.016$. For comparison, the classical methods give the confidence interval (0, 0.024) for the parameters M and N.

5 The case of small sample

Intuitively, the proposed estimator $\widehat{p}_0$ seems to be too low for small values of n. Let us construct an estimate $\widehat{p}(n)$ for this case.

In the case of very small n, the occurrence of a 0-event does not contradict the assumption that the probability p is high. Therefore, it seems reasonable to use the following approach. For a certain N-element sample with M unit values, use (6) to find the probability $P(N, M; n)$ that table (5) was generated by one and the same probability value; then, average the estimate M/N in accordance with the introduced probability distribution on the samples. The mean value

$$\widehat{p}_N(n) \;=\; \frac{\sum\limits_{M=0}^{N} \frac{M}{N} \cdot P(N, M; n)}{\sum\limits_{M=0}^{N} P(N, M; n)} \qquad (10)$$

is taken as the desired estimate for the given $N \geq 1$.

By induction, it is easy to show that

$$\sum_{M=0}^{N} P(N, M; n) \;=\; \sum_{M=0}^{N} \frac{\binom{N}{M}}{\binom{N+n}{M}} \;= \;=\; \sum_{M=0}^{N} \frac{N!\,(N-M+n)!}{(N-M)!\,(N+n!)} \;=\; \frac{N+n+1}{n+1}\,;$$

thus, the denominator in (10) is determined. Similarly, for the numerator, we have

$$\sum_{M=0}^{N} \frac{M}{N} \cdot P(N, M; n) \;= \;=\; \sum_{M=1}^{N} \frac{M}{N} \cdot \frac{N!\,(N-M+n)!}{(N-M)!\,(N+n!)} \;=\; \frac{N+n+1}{(n+1)(n+2)}\,.$$

Hence,

$$\widehat{p}_N(n) \;=\; \frac{1}{n+2} \;=\; \widehat{p}_{B_\mu}(n)$$

(and it is a pleasant surprise that $\widehat{p}_N(n)$ is independent of N; this saves us the calculation of N or one more calculation of the average over its values).

This result makes us conclude that, for low values of n, a validated point estimate of the 0-event probability is the expectation-based Bayesian estimate under the condition of the uniform *a priori* distribution.

It is of interest to note that formula (3) may be interpreted as the computation of the mean probability value p when the distribution is $Be_p(1,\, n+1)$; in the framework of the fiducial Fisher approach, this distribution expresses the degree of confidence that the current value of p is equal to the true value of the 0-event probability. Besides, it is easily shown that the asymptotics

$$P(N, M; n) \;=\; \frac{\binom{N}{M}}{\binom{N+n}{M}} \;\sim\; \left(1 - \frac{M}{N}\right)^{n}$$

holds for $n \ll N$. Thus, (10) turns out to be a discrete analog of (3), which accounts for the coincidence of the estimates $\widehat{p}_N(n)$ and $\widehat{p}_{B_\mu^U}(n)$, as well as $\widehat{p}_\eta$ and $\widehat{p}_r$.

For "medium" (not too low and not too high) n, it seems appropriate to use a step function as an *a priori* distribution taking into account not all possible values of $p$but only those that do not contradict the assumption that the current value of p is equal to the true value of the 0-event probability with the confidence degree η:

$$f_{a\_priori}(p) \;=\; \begin{cases} 1/\widehat{p}_\eta & \text{if } 0 \leq p \leq \widehat{p}_\eta\,, \\ 0, & \text{otherwise}\,. \end{cases}$$

As a result, we obtain (cf. (3)) the estimate $\widehat{p}_{B_\eta}(n)$:

$$\widehat{p}_{B_\eta}(n) \;=\; \frac{1}{(n+2)\cdot \widehat{p}_\eta}\, I_{\widehat{p}_\eta}(2, n+1)\,,$$

where $\widehat{p}_\eta = 1 - \sqrt[n]{\eta}$. For example, for $\eta = 0.95$, we have $\widehat{p}_{B_\eta}(10) = 0.0272$ and $\widehat{p}_{B_\eta}(20) = 0.0259$.

Note that in our case, the values of the incomplete B-function cannot be determined from the available tables (e.g. [2]) for the parameter values of interest. For that purpose, the formula

$$I_x(2, n+1) \;=\; \sum_{k=2}^{n+2} C_{n+2}^{k}\, x^k (1-x)^{n+2-k}$$

can be used (where the terms rapidly decrease with increasing k).

6 When different estimators should be used?

Let us state from the outset that we did not consider cases when the principles (or, rather, advisability of more likely deviation) underlying the choice of a point estimate in the subject-matter area of investigation in question are clear: here everything depends on how desirable or undesirable the occurrence of the given rare event is.

Let us immediately note that we do not consider the the cases when the principles for the choice of the desired point estimate in a specific subject area are clear; in this case, everything depends on how desirable or undesirable is the occurrence of the rare event under examination.

If no such principles are available, then, in order to answer the question appearing in the heading, we should clarify the concept of "small sample".

Different authors define this notion differently: a sample is considered small if its size does not exceed 200 [16], or 50 [26], or 30 [6], or 10–20 [18], or 10–15 [21], or "is smaller than a calculated number determined by using a special nomogram of sufficiently large numbers" [20], or if "there are no stable informational properties and statistical characteristics" [25]. Often, this notion is not defined at all.

The standpoint of the author of this paper was presented in [7]: a sample is considered to be small if its processing by methods based on grouping of observations and by approximation methods cannot ensure the desired accuracy and confidence. For the case of a 0-event, this statement needs to be made more precise; namely it should be indicated for which values of n the estimate $\widehat{p}_{B^U_\mu}(n)$ should be used, and for which n $\widehat{p}_0(n)$ is preferable. It is clear that there can be no absolutely objective criteria of such a choice. However, we propose a partition based on the statistical significance of results.

6.1 Lower bound

First, it seems clear that at low values of n no statistical conclusions can be drawn at all. Note that, for $n < 4$, we have $\widehat{p}_{B_{med}}(n) < \widehat{p}_{B^U_\mu}(n)$, and the reverse inequality holds for higher n. This value appears to be a natural boundary for separating the "small sample" notion from the case of insufficient data for any statistical conclusions. Thus, I am of the opinion that, for $1 \le n \le 3$, one can only ascertain the occurrence of a 0-event for the given number of trials.

A similar conclusion is drawn in [11]: *One of the main issues of mathematical statistics is to determine what should be the minimum necessary information to obtain the required validity of the result. ... If the conditions are interpreted as the absence of any limitations on the accuracy of the final result of the statistical analysis, then the answer to this question was given by R. Fisher [27, 24]. The minimum size of the sample cannot be less than four. Otherwise, a systematic error (bias) is inevitable. A bias is the first indication that the statistic is insufficient [19]. A number of authors supported Fisher's conclusion.* I would add for myself: cf. the above-quoted extract from [25].

Also, when testing the hypothesis on the value of the ratio ξ of observed absolute frequencies a and b using Pearson's χ^2 test with the statistical reliability of 95%, it is required (vide, e.g., [15, (4.33)]) that

$$\widehat{\chi}^2 = \frac{(\xi a - b)^2}{\xi(a+b)} < \chi^2 = 3.841\,.$$

When determining the equality of probabilities producing samples as realizations of random processes, we assume that $\xi = 1$, which leads to the inequality $|a - b| < 3.841$. Since the use of this test suggests that $0 < a \le b$, we consider the complementary *complete event*, for which $b = n$, instead of the 0-event. Thus, in order to be able to regard the sample with a unit values as a different realization of the same random process that produced the zero sample $x^0 = (0, \ldots, 0)$ of the same length, $n - a$ must be not greater than 3[1]. Thus, the difference can be determined statistically only if the sample length is $4 \le n$.

6.2 Upper bound

It is natural to set the upper bound for a small sample in the case of a 0-event equal to N according to (8) with $n = 1$. This value virtually coincides with the condition $\widehat{p}_{B^U_\mu}(n) < 1 - \eta$ for the same confidence value η. Lower values of n do not make it possible to determine with statistical certainty the coincidence of the unit events' probabilities related to the samples in question.

For $\eta = 0.95$, we have $n = 19$ for the desired bound. Note that it entails $\widehat{p}_{B^U_\mu}(n) \approx 5\%$. Furthermore, 5% is often used in practice as the boundary of a "rare event" (vide, e.g.[1]).

[1] Interestingly, the boundary value 3 (the so-called the "Bongard three") often emerges in combinatorial investigations on non-randomness of events [3, 13, 14].

7 Conclusions

As a result, the following classification of 0"=samples by size is proposed; the point estimate $\widehat{p}$ of the 0"=event probability is also indicated.

n	0-sample type, $\widehat{p}$
$1, 2, 3$	no estimates can be given
from 4 to 19	"small" 0-sample, $\widehat{p} = \widehat{p}_{B^U_\mu}(n)$
more than 20	"large" 0-sample, $\widehat{p} = \widehat{p}_0(n)$

An abrupt jump of the proposed estimate when a "small" sample is replaced by a "large" one is due to the extreme nature of the very concept of 0-event: it becomes possible to statistically register with sufficient certainty the coincidence or non-coincidence of unit events' probabilities that determine samples under the Bernoulli scheme. In the case of small samples, the occurrence of a 0-event is quite possible even if p is not close to zero; on the contrary, in the case of large samples, it inevitably implies either an extremely low value of p or impossibility of the occurrence of event X.

Acknowledgements

I am grateful to Yu. I. Zhuravlev for support and to V. E. Bening for helpful advice.

REFERENCES

[1] N. Beili. Mathematics in biology and medicine, Online available from http://www.biometrica.tomsk.ru/beili_2_2.htm [in Russian].

[2] L. N. Bolshev, N. V. Smirnov. Tables of Mathematical Statistics, Nauka, Moscow, 1983 [in Russian].

[3] M. M. Bongard. Recognition problem, Nauka, Moscow, 1967 [in Russian].

[4] G. P. Gavrilov, A. A. Sapozhenko. Discrete mathematics problems and exercises, FISMATLIT, Moscow, 2004 [in Russian].

[5] D. V. Gaskarov, V. I. Shapovalov. Small sample, Statistika, Moscow, 1978 [in Russian].

[6] V. E. Gmurman. Theory of probabilities and mathematical statistics, Vysshaya Shkola, Moscow, 1977 [in Russian].

[7] S. I. Gurov. Estimation of classifying algorithms' reliability, Publications office of MSU VMK Dept., Moscow, 2002 [in Russian].

[8] S. I. Gurov. Consistency principle and Bayesian interval estimator, Tavrian Informatics and Mathematics Newsletter, 2003, Issue 2, 14-27 [in Russian].

[9] S. I. Gurov. Point estimator based on the consistency Principle, Vestnik Tverskogo Gosudarstvenogo Universiteta, "Applied Mathematics" series, No. 14 (74), Issue 9, 77-93 (2008) [in Russian].

[10] S. I. Gurov. Consistency and Bayesian Interval Estimation, Tavricheskii Vestnik Informatiki i Matematiki, No. 2, 14-27 (2003) [in Russian].

[11] A. V. Gusev, E. A. Lidskii, O. V. Mironenko. Small samples in estimating operability and dependability of electronic components. Part I, Chip news, Engineering microelectronics, 2002, No. 1, 52-56. Online available from http://www.chipinfo.ru/literature/chipnews/about.html. [in Russian].

[12] M. H. DeGroot. Optimal Statistical Decisions, McGraw-Hill, New York, 1970; Mir, Moscow, 1974.

[13] V. I. Donskoi, A. I. Bashta. Discrete models of decision-making with incomplete data, Tavria, Simferopol, 1992 [in Russian].

[14] A. D. Zakrevskii. Recognition logic, URSS Editorial, Moscow, 2003 [in Russian].

[15] L. Zacks. The Theory of Statistical Inference, Wiley, New York, 1971; Statistika, Moscow, 1976.

[16] M. Kendal, A. Stuart. Distribution theory, Mac Publishing Co, New York, 1997; Nauka, Moscow, 1966.

[17] M. G. Kendal, A. Stuart. Statistical Inference and Relationships, Charles Griffin, London, 1969; Nauka, Moscow, 1973.

[18] N. Sh. Kremer. Theory of probabilities and mathematical statistics, UNITY-DANA, Moscow, 2000 [in Russian].

[19] E. L. Lehmann. Theory of Point Estimation, Springer, New York, 1988; Nauka, Moscow, 1991.

[20] Methods of statistical analysis and processing low number of observations, when auditing the quality and dependability of instruments and machines, LDNTP Publ., Leningrad, 1974 [in Russian].

[21] N. V. Smirnov, I. V. Dunin-Barkovskii. Course of probabilities theory and mathematical statistics for technical applications, Nauka, Moscow, 1965 [in Russian].

[22] S. A. Smolyak, B. P. Titarenko. Sound estimator methods: (Statistical ptrocessing of heterogeneous populations), Statistika, Moscow, 1980 [in Russian].

[23] V. Feller. Introducton into the theory of probabilities and its applications, Vol. 1-2, Wiley, 1971; Mir, Moscow, 1984.

[24] R. Fisher, Statistical methods for research workers, Oliver and Boyd, 1925; Gostekhizdat, Moscow, 1958.

[25] V. A. Fursov. Identification of imaging system models by a low numbder of observations, Publications office of Samar. state aerospace Univ., Samara, 1998 [in Russian].

[26] Ya. B. Shor. Statistical conclusions of analysis and of dependability/quality audit, Sov. Radio, Moscow, 1962.

[27] R. A. Fisher. On the mathematical foundations of theoretical statistics, Phil. Trans. Roy. Soc., Ser. A, Vol. 222, 1921.