Available online at: https://ijact.in

Date of Submission	05/08/2019
Date of Acceptance	07/10/2019
Date of Publication	31/10/2019
Page numbers	3440-3443(4 Pages)

Cite This Paper: Oleg J. Kravets et.al. Creating a model of resource saving for diagnosing various deviations, 8(10), COMPUSOFT, An International Journal of Advanced Computer Technology. PP. 3440-3443.

ISSN:2320-0790

An International Journal of Advanced Computer Technology

CREATING A MODEL OF RESOURCE SAVING FOR DIAGNOSING VARIOUS DEVIATIONS

Oleg Jakovlevich Kravets[1], Igor Viktorovich Atlasov[2], Vladimir Dmitriyevich Sekerin[3], Anna Evgenievna Gorokhova[3], Sergey Kurbanovich Gasanbekov[3]

[1]Voronezh State Technical University, Moscow Ave., 14, Voronezh, 394026, Russia
[2]Moscow University of Ministry of Internal Affairs of Russian Federation named by V.Ja. Kikot, Volgin Academician St., 12, Moscow, 107061, Russia
[3]Moscow Polytechnic University, Bolshaya Semenovskaya St., 38, Moscow, 107023, Russia

Abstract: When examining a large group of people and provided that very few people get sick (the probability of a disease p ≈ 0), there is a technique to diagnose the disease that allows reducing the use of drugs necessary for diagnosing the disease. A set of all people (total - n) is divided arbitrarily into groups of k people (k>1). Then the blood of these k people is mixed and a part is drawn for a single analysis. If it is possible to find k so that practically in all of these groups there will be no patients with a disease, the quantity of the chemicals used for diagnosing can be reduced approximately k times. Section I (introduction) describes this problem. In section II the proposed methodology is discussed. The first improved model include the problem of finding the maximum natural value of k arises, so that almost all of these groups of k people include no sick people. The second improved model shows that it is possible to further reduce the use of resources due to the fact that if a virus is found in the fluid collected from a group of k people and the first k-1 people are healthy, then the last person should not be checked, since it is definitely infected. The idea of the third improved model is based on the fact that the group can also be divided into smaller groups and, under certain circumstances, the blood will be drawn not from each individual, but in small groups. The fourth improved model consider the generalized technique. The section III (conclusion) confirms that the scope of the proposed methods is wider than that of the existing ones, using a numerical example.

Keywords: extremum, mathematical expectation, random function

I. INTRODUCTION

When examining a large group of people and provided that very few people get sick (the probability of a disease $p \approx 0$), there is a technique to diagnose the disease that allows reducing the use of drugs necessary for diagnosing the disease. A set of all people (total - n) is divided arbitrarily into groups of k people ($k> 1$). Then the blood of these k people is mixed and a part is drawn for a single analysis. If it is possible to find k so that practically in all of these groups there will be no patients with a disease, the quantity of the chemicals used for diagnosing can be reduced approximately k times.

II. PROPOSED METHODOLOGY

A. The first improved model

The problem of finding the maximum natural value of k arises, so that almost all of these groups of k people include no sick people. In [1] the following probabilistic model is proposed. Let Υ_i be an integer-valued random variable equal to the number of procedures performed for the i -th party. Let n has no remainder divided by$k$. The total number of procedures performed is$\Upsilon = \sum_{i=1}^{\frac{n}{k}} \Upsilon_i$. If the mathematical expectation $M(\Upsilon) = \sum_{i=1}^{\frac{n}{k}} M(\Upsilon_i) = \frac{n}{k} M(\Upsilon_1)$ is less than n, then we can assume that some savings were achieved. That is, it makes sense to calculate the function $\varphi(k) = M(\Upsilon_i)$ and find its minimum by$k$. A random value $\Upsilon_i = 1$ provided that all k people are healthy and$\Upsilon_i = k + 1$, provided that at least one person is sick. It is obvious that if the sick patients are independent of each other, we have

$$\Upsilon_i = \begin{cases} 1 & where \quad (1-p)^k \\ k+1 & where \quad 1-(1-p)^k \end{cases}$$

Then, calculate the mathematical expectation of a random variable Υ_i

$M(\Upsilon_i) = (1-p)^k + (k+1)(1-(1-p)^k) = (k+1) - k(1-p)^k$and, accordingly, the mathematical expectation of a random variable Υ is equal to

$M(\Upsilon) = \sum_{i=1}^{\frac{n}{k}} M(\Upsilon_i) = \frac{n}{k} M(\Upsilon_1) = n\left(\left(1+\frac{1}{k}\right) - (1-p)^k\right)$. To study the function $M(\Upsilon)$ on an extremum in [1], the following function is studied on extremum with an actual argument *x*.

$$K(x) = 1 + \frac{1}{x} - (1-p)^x$$

The function $\bar{K}(x) = \frac{1}{x} + xp$ approximate to $K(x)$ is investigated. Since $\bar{K}^*(x) = -\frac{1}{x^2} + p$ then the minimum point is $\bar{x}_k = \frac{1}{\sqrt{p}}$ and $\bar{K}\left(\frac{1}{\sqrt{p}}\right) = 2\sqrt{p}$. It makes sense to use the model at $M(\Upsilon) < n$, or

$$\min_x K(x) \approx \bar{K}\left(\frac{1}{\sqrt{p}}\right) = 2\sqrt{p} < 1,$$

or $p < \frac{1}{4} = 0.25$.

Similar problems were considered in works [1-8].

B. The second improved model

In [8] an analog of a random variable Υ_i is built - the random variable $\hbar_i$ has the form

$$\hbar_i = \begin{cases} 1 & where \quad (1-p)^k \\ k & where \quad p(1-p)^{k-1} \\ k+1 & where \quad 1-(1-p)^k - p(1-p)^{k-1} \end{cases}$$

The probability of the event is $(1-p)^{k-1}p$. If everyone in the party is healthy, then one study is needed, with the probability $(1-p)^k$. And if the previous two conditions are not fulfilled, then a k+1 study is necessary, with the probability $1-(1-p)^k-(1-p)^{k-1}p$. It is proved that almost everywhere $P(\hbar_i \le \Upsilon_i) = 1$. The mathematical expectation of a random variable $\hbar = \sum_{i=1}^{\frac{n}{k}} \hbar_i$ is

$$M(\hbar) = n\left(\left(1+\frac{1}{k}\right) - \left(\frac{p}{k(1-p)} + 1\right)(1-p)^k\right) =$$

$$= M(\Upsilon) - \frac{np(1-p)^k}{k(1-p)} < M(\Upsilon)$$

The following function was also considered

$$H(x) = 1 + \frac{1}{x} - \left(\frac{p}{(1-p)x} + 1\right)(1-p)^x$$

which was investigated for an extremum. It is proved that the equation $H^*(x) = 0$ has a root x_h, in which a maximum is attained and $1 < x_h < x_k$, where x_k. - is the smallest root of the equation $K^*(x) = 0$. Using an approximate equality $(1-p)^x \approx 1 - xp$, it is seen that

$$H(x) = 1 + \frac{1}{x} - \left(\frac{p}{(1-p)x} + 1\right)(1-p)^x \approx$$

$$\approx 1 + \frac{1}{x} - \left(\frac{p}{(1-p)x} + 1\right)(1-xp) = \frac{1-2p}{(1-p)x} + xp + \frac{p^2}{(1-p)}$$

and the following function is studied.

$$\bar{H}(x) = \frac{1-2p}{(1-p)x} + xp + \frac{p^2}{(1-p)}.$$

The minimum of this function is achieved at the point $\bar{x}_h = \frac{1}{\sqrt{p}}\sqrt{\frac{1-2p}{1-p}}$. The last inequality is valid for small quantities of p and provided the fulfilling of two conditions $p < 0.5$, and

$$\min_p H(p) \approx \bar{H}\left(\frac{1}{\sqrt{p}}\sqrt{\frac{1-2p}{1-p}}\right) = 2\sqrt{p}\sqrt{\frac{1-2p}{1-p}} + \frac{p^2}{1-p} < 1.$$

Then, the following function was investigated for an extremum:

$y = p^4 - 6p^3 + 11p^2 - 6p + 1$.

It is proved that the derivative equated to zero is equal to

$$y^* = \left(p - \frac{3}{2}\right)\left(p - \frac{3+\sqrt{5}}{2}\right)\left(p - \frac{3-\sqrt{5}}{2}\right).$$

It is proved that the function y on the interval $\left[0, \frac{1}{2}\right]$ is positive and equals zero only at the point $\frac{3-\sqrt{5}}{2}$. It is proved that the condition is fulfilled for all $0 < p < 0.5$. Thus, the scope of the new technique is wider than the scope of the previous one.

$$\min_p H(p) \approx \bar{H}\left(\frac{1}{\sqrt{p}}\sqrt{\frac{1-2p}{1-p}}\right) = 2\sqrt{p}\sqrt{\frac{1-2p}{1-p}} + \frac{p^2}{1-p} < 1$$

C. The third improved model

The idea of the method is based on the fact that the group can also be divided into smaller groups and, under certain circumstances, the blood will be drawn not from each individual, but in small groups. Take some numbers *w* and *v* and denote *k*=*wv*. Break *k* individuals into *w* groups of *v* people. In each mini-group we drew blood, mix it, take one drop and store it. Then we take bioactive liquid from the storehouse, and in each group of *k* individuals we mix it and provide only one analysis. If none of the people in this group is sick, then instead of *k* analyses, it is enough to do one with the probability $(1-p)^{wv}$, otherwise in each group should be made more than *w*+1 analysis, that is, in each group, blood drops will be taken, mixed and checked for infections.

If only one group of v people has a sick patient, then it will be necessary to make w+1+v analyses with probability $C_w^1(1-p)^{v(w-1)}(1-(1-p)^v)$ - it is necessary to take an additional test in the group with a sick patient.

If there is at least one sick patient in two groups of v, then w+1+2v tests will have to be done with probability $C_w^2(1-p)^{v(w-2)}(1-(1-p)^v)^2$ - it is necessary to take tests in two groups with the sick patients.

If in s<w groups of v people there is a minimum one sick patient, then w+1+v analyses will have to be done with probability $C_w^s(1-p)^{v(w-s)}(1-(1-p)^v)^s$ even in s groups where sick patients are.

Note that the method for v=1 completely coincides with the previous technique.

In this work an analog of a random variable Υ_i - Θ_i for all $1 \le s \le w$ is constructed taking the following form

$$\Theta_i = \begin{cases} 1 & where \quad (1-p)^{wv} \\ w+1+sv & where \quad C_w^s(1-p)^{v(w-s)}(1-(1-p)^v)^s \\ w+1+wv & where \quad (1-(1-p)^v)^w \end{cases}$$

.

Let us find the mathematical expectation of this random variable:

$$M(\Theta_i) = \sum_{s=0}^{w} C_w^s(1-p)^{v(w-s)}(1-(1-p)^v)^s + \\ +w\sum_{s=1}^{w} C_w^s(1-p)^{v(w-s)}(1-(1-p)^v)^s + \\ +v\sum_{s=0}^{w} sC_w^s(1-p)^{v(w-s)}(1-(1-p)^v)^s.$$

Using the binomial distribution and expectation formulas for $0 < a < 1$ we have:

$$M(\Theta_i) = 1 + w(1-(1-p)^{vw}) + vw(1-(1-p)^v)$$

at n multiple of wv we have

$M(\Theta_i) = \sum_{i=1}^{\frac{n}{wv}} M(\Theta_i) = n\left(\frac{1}{wv} + \frac{1}{v}(1-(1-p)^{vw}) + 1 - 1-pv\right.$. Let us investigate the extremum function

$$G(w,v) = \frac{1}{wv} + \frac{1}{v}(1-(1-p)^{vw}) + 1 - (1-p)^v.$$

Since for sufficiently small positive p it is true that $(1-p)^{vw} = 1 - vwp$ then the approximate value of the function $G(w,v)$ has the form:

$$G(w,v) \simeq \frac{1}{wv} + p(w+v) = U(w,v).$$

It must be shown that the values $v_0 = w_0 = \frac{1}{\sqrt[3]{p}} > 0$ are at least approximately the solution of the system

$$\begin{cases} \frac{1}{w^2v} + (1-p)^{vw}\ln(1-p) = 0 \\ \frac{1}{wv^2} + \frac{1}{v^2} + \frac{w}{v}(1-p)^{vw}\ln(1-p) - \frac{1}{v^2}(1-p)^{vw} + \\ +(1-p)^v\ln(1-p) = 0 \end{cases}$$

It is possible to consider approximately that $v = w = \frac{1}{\sqrt[3]{p}}$ is the solution of the last system.

Then $G\left(\frac{1}{\sqrt[3]{p}}, \frac{1}{\sqrt[3]{p}}\right) \approx 3\sqrt[3]{p^2}$ and

$$\min_{w,v} G(w,v) \approx U\left(\frac{1}{\sqrt[3]{p}}, \frac{1}{\sqrt[3]{p}}\right) = 3p^{\frac{2}{3}} < 1,$$

or $p < \frac{1}{\sqrt{27}} \simeq 0.19245$.

Thus, for arbitrary, $\varepsilon > 0$ the condition $\min_{w,v} G(w,v) < \varepsilon \min_x K(x)$, is satisfied when $p < \left(\frac{2}{3}\varepsilon\right)^6$. The effectiveness of the new method in comparison with the old one is achieved already at $p < \left(\frac{2}{3}\right)^6 = 0{,}08779$.

Thus, the application range of the new methodology is narrower than of the previous two. Application area is $0 < p < 0.1952$. The technique is more effective than the two previous ones on the interval $0 < p < 0.0877$.

D. The fourth improved model

Consider the generalized technique. We note that the probability that in a group of v people for which it is known that there is a sick person in it, there is only one infected and he/she is the last in check, is equal $(1-p)^{v-1}p$ and drawing blood from all patients is not necessary. It is possible to take blood from the first v-1, since the latter is apparently sick. One of three events is possible: A_i - in the i-th group everyone is healthy; B_i - in the i-th group, the first v-1 tested people are healthy and v is sick; C_i- in an i-th group at least one is sick and the event B_i is not fulfilled. The probabilities of these events are respectively equal –

$$P(A_i) = (1-p)^v,\ P(B_i) = p(1-p)^{v-1}, P(C_i) = 1 - (1-p)^v - p(1-p)^{v-1}.$$

Then, take a biologically active liquid from the storage, and in each group, we mix it and do just one analysis. If none of the people in this group are sick, then instead of k analyses it is enough to conduct one with probability $(1-p)^{wv}$, otherwise in each group more than w+1 analyses should be made.

If exactly in s<w mini groups there is one sick patient who is checked last, h other cases involve sick patients, and in w-h-s cases everyone is healthy, we will have to make $w+1+s(v-1)+hv$ analyses with probability

$$\frac{w!}{s!\,h!\,(w-s-h)!}(p(1-p)^{v-1})^s((1-p)^v)^{w-s-h} \\ (1-p(1-p)^{v-1}-(1-p)^v)^h$$

Let us find the mathematical expectation of this random variable.

$$M(\Psi_i) = (1-p)^{wv} + \sum_{s=0}^{w}\sum_{h}^{w-s}(w+1+s(v-1)+hv) * \\ * \frac{w!}{s!\,h!\,(w-s-h)!}(p(1-p)^{v-1})^s((1-p)^v)^{w-s-h} * \\ * (1-p(1-p)^{v-1}-(1-p)^v)^h$$

Simplify the above-obtained expression by applying formulas for the polynomial distribution. Using these formulas, we have

$$M(\Psi_i) = n\left(-\frac{1}{v}\frac{p}{1-p}(1-p)^v + (1-(1-p)^v) - \right.$$

$$-\frac{1}{v}\frac{p}{1-p}(1-p)^v + (1-(1-p)^v)$$

Let us test the function for an extremum

$$Q(w,v) = \frac{1}{wv} + \frac{1}{v}(1-(1-p)^{wv}) -$$
$$-\frac{1}{v}\frac{p}{1-p}(1-p)^v + (1-(1-p)^v)$$

where $w > 0$ and $v > 0$ are positive real numbers. It's obvious that $Q(w,v) < U(w,v)$. That is, the mathematical expectation has been somewhat reduced. Let's try to simplify the task a little. Since for sufficiently small positive p the following equality is true $(1-p)^{vw} = 1 - vwp$, which follows from the limit $\lim_{p\to 0}\frac{(1-p)^{vw}-1}{-p} = vw$, then the approximate value of the function $Q(w,v)$ has the form

$$Y(w,v) = \frac{1}{wv} + \frac{1}{v}\big(1-(1-wvp)\big) -$$
$$-\frac{1}{v}\frac{p}{1-p}(1-vp) + \big(1-(1-vp)\big) =$$
$$= \frac{1}{wv} + (w+v) - \frac{1}{v}\frac{p}{1-p}(1-vp)$$

Further, find the partial derivatives

$$Y_w^*(w,v) = -\frac{1}{w^2v} + p$$

$$Y_v^*(w,v) = -\frac{1}{wv^2} + p + \frac{1}{v^2}\frac{p}{1-p}$$

It is quite difficult to solve even this simplified system the usual way, by equating the derivatives to zero.

Note that $v_0 = w_0 = \frac{1}{\sqrt[3]{p}} > 0$ are approximately the roots of the last two equations described above, since

$$Y_w^*\left(\frac{1}{\sqrt[3]{p}}, \frac{1}{\sqrt[3]{p}}\right) = 0$$
$$Y_v^*\left(\frac{1}{\sqrt[3]{p}}, \frac{1}{\sqrt[3]{p}}\right) = \sqrt[3]{p^2}\frac{p}{1-p}$$

Then,

$$Y\left(\frac{1}{\sqrt[3]{p}}, \frac{1}{\sqrt[3]{p}}\right) = 3\sqrt[3]{p^2} - \frac{\sqrt[3]{p^4}}{1-p} + \frac{p^2}{1-p} =$$
$$= \sqrt[3]{p^2}\left(3 - \frac{\sqrt[3]{p^2}}{1-p}\left(1-\sqrt[3]{p^2}\right)\right) =$$
$$= \sqrt[3]{p^2}\left(3 - \sqrt[3]{p^2}\frac{1+\sqrt[3]{p}}{1+\sqrt[3]{p}+\sqrt[3]{p^2}}\right) < 3\sqrt[3]{p^2}$$

Note that the latter method can be used for large $p \leq 0.2$. It's easy to calculate that $Y\left(\frac{1}{\sqrt[3]{0.2}}, \frac{1}{\sqrt[3]{0.2}}\right) = 0.929 < 1$.

III. Conclusion

The article proposed four models that provide resource saving for diagnosing various deviations. Area of its application $0 < p < 0.2$. Further, this technique is more effective than the previous ones on the interval $0 < p < 0.0877$. Therefore, the questions of the application of the techniques on the interval $0 < p < 0.5$ can be clarified as follows - on the interval $0 < p < 0.0877$ the third method should be applied, on the interval $0.0877 < p < 0.25$ - the first one, and on the interval $0.25 < p < 0.5$ - the second one.Thus, the application scope of the new method is somewhat wider than of the previous one.

References

[1]. Lagutin, M.B. 2009. Naglyadnaya matematicheskaya statistika [Visual mathematical statistics Tutoria l, Binom, Laboratory of knowledge, Moscow]. Moscow: BINOM.

[2]. Hakrama, I., Frashëri, N.Agent-Based Modelling and Simulation of an Artificial Economy with Repast. International Journal on Information Technologies & Security, 10(2) (2018) 47-56.

[3]. Hassani, M.M., Berang, R. An Analytical model to calculate blocking probability of secondary user in cognitive radio sensor networks. International Journal on Information Technologies & Security, 10(2) (2018) 3-12.

[4]. Gnedenko, B.V. (2001), Kurs teorii veroyatnostey [Course of Probability Theory], Editorial URSS, Moscow, Russia.

[5]. Atlasov I.V. About efficiency of work of several interchangeable devices. Vestnik TvGU. Seriya: Prikladnaya matematika [Herald of Tver State University. Series: Applied Mathematics], 2015, no. 4, pp. 85-101. (in Russian). http://pmk-vestnik.tversu.ru/issues/2015-4/vestnik-pmk-2015-4-atlasov.pdf

[6]. Atlasov I.V. Operation of two parallel devices with regard to the time of their replacement. Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics], 2016, no. 2, pp. 49-79. (in Russian). http://pmk-vestnik.tversu.ru/issues/2016-2/vestnik-pmk-2016-2-atlasov.pdf

[7]. Borovkov A.A. Probability Theory. Springer-Verlag London, 2013. https://www.springer.com/gp/book/9781447152002#otherversion=9781447152019.

[8]. Atlasov, I.V., Dubinina, N.M., Construction of a model that saves resources needed to diagnose self-replicating organisms in a biologically active environment. Sistemy upravleniya i informatsionnye tekhnologii [Control systems and information technologyes], 2017, no. 3(69), pp. 49-53. (in Russian). https://elibrary.ru/item.asp?id=29851953.