QUANTITATIVE METHODS IN ECONOMICS
Vol. XIII, No 2, 2012, pp. 134 – 138

AN APPLICATION OF THE SHORTEST CONFIDENCE INTERVALS FOR FRACTION IN CONTROLS PROVIDED BY SUPREME CHAMBER OF CONTROL

Wojciech Zieliński
Department of Econometrics and Statistics
Warsaw University of Life Sciences – SGGW
e-mail: wojciech_zielinski@sggw.pl

Abstract: In statistical quality control objects are alternatively rated. It is of interest to estimate a fraction of negatively rated objects. One of such applications is a quality control provided by Supreme Chamber of Control (NIK) to find out a percentage of abnormalities in the work among others of tax offices. Mathematical details of experimental designs for alternatively rated phenomena are given in Karliński (2003). Zieliński (2010b) investigated statistical properties of those experimental designs. In the paper, the application of the shortest confidence intervals for fraction in experimental designs is shown. Those intervals were proposed by Zieliński (2010a).

Keywords: statistical quality control, alternative rating, experimental design, shortest confidence intervals for fraction

One of the problem of the statistical quality control is the problem of the estimation of the fraction of defective products. Generally speaking, the products are alternatively rating and one is interested in estimation of a fraction of negatively rated objects. In this approach, the binomial statistical model is applied, i.e. if ξ is a random variable counting negative rated in a sample of size n, then ξ is binomially distributed

$$P_\theta\{\xi = x\} = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x = 0,1,\ldots,n,$$

where $\theta \in (0,1)$ is a probability of drawing a defective product. The aim of the statistical quality control is to estimate θ.

In many norms and books devoted to different applications there are given exacts designs of experiments, i.e. requirements for sample sizes, number of negative rates in the sample, accuracy of estimation and error risks. One of such applications are quality controls provided by Supreme Chamber of Control, the goal of which is finding abnormalities in tax offices. Karliński (2003) gives mathematical details of such controls. There are given methods of providing experiments and rules of statistical inference. Statistical properties of given experimental designs were investigated by Zieliński (2010b). It was shown that proposed by Karliński solutions have at least two disadvantages: obtained confidence intervals for fraction may take on negative values and real confidence level may be significantly smaller than nominal one. Zieliński (2010b) proposed an application of Clopper-Pearson (1934) confidence intervals for estimation the proportion of negatively rating objects.

Clopper and Pearson (1934) give the confidence interval for θ, based on the exact distribution of ξ. Because

$$P_\theta\{\xi \le x\} = \beta(n-x, x+1; 1-\theta) \quad \text{oraz} \quad P_\theta\{\xi \ge x\} = \beta(x, n-x+1; \theta),$$

where $\beta(a,b;\cdot)$ denotes a CDF of a beta distribution with parameters (a,b), hence the confidence interval at the confidence level γ has the form $\left(\theta_L(x), \theta_U(x)\right)$, where

$$\theta_L(x) = \beta^{-1}\left(x, n-x+1; \gamma_1\right), \qquad \theta_U(x) = \beta^{-1}\left(x+1, n-x; \gamma_2\right).$$

Here $\gamma_1, \gamma_2 \in (0,1)$ are such that $\gamma_2 - \gamma_1 = \gamma$.

For $x = 0$ we take $\theta_L(0) = 0$, and for $x = n$ is taken $\theta_U(n) = 1$. Here $\beta^{-1}(a,b;\cdot)$ denotes the quantile of the Beta distribution with parameters (a,b).

Clopper and Pearson (1934) in their construction used $\gamma_1 = (1-\gamma)/2$, i.e. they applied the rule of symmetric division of $1-\gamma$ to both sides of the interval. The length of the confidence interval was not considered as a criterion. It is of interest to find the shortest confidence interval. So we want to find γ_1 and γ_2 such that the confidence interval is the shortest possible.

Consider the length of the confidence interval when $\xi = x$ is observed,

$$d(\gamma_1, x) = F^{-1}(x+1, n-x; \gamma+\gamma_1) - F^{-1}(x, n-x+1; \gamma_1).$$

Let x be given. We want to find $0 < \gamma_1 < 1-\gamma$ such that the length $d(\gamma_1, x)$ is minimal.

In Zieliński (2010a) the existence of the shortest confidence interval is proved and a numerical method of obtaining such intervals is shown. It is interesting, that for $x=0$ and $x=1$ as well as for $x=n-1$ and $x=n$ the shortest confidence interval is one-sided.

Karliński (2003) considered the following problem. For given confidence level γ and for given $\varepsilon>0$ find sample size n such that the length of obtained confidence interval is smaller than ε. Zieliński (2010b) compared Karliński's solution with those which is obtained by application of classical Clopper-Pearson confidence interval.

In what follows it is shown the solution for the shortest confidence interval.

Let $\gamma=0.95$ and $\varepsilon=0.1$. Assume that the true fraction of negatively rated objects is $\theta=0.05$. For given x the sample size n is seek such that the expected length of the shortest confidence interval is smaller than ε. Numerical solutions for $n=81$ and $n=82$ are given in Tables 1 and 2, respectively. In the column before last one it is denoted whether the obtained interval covers the estimated value 0.05.

Table 1. Shortest confidence interval for sample size 81

X	γ_1	left	right	length		$P_{0.05}\{\xi=m\}$
0	0	0	0.03631	0.03631	0	0.01569
1	0	0	0.05723	0.05723	1	0.06689
2	0.00079	0.00050	0.07594	0.07544	1	0.14082
3	0.00371	0.00378	0.09426	0.09048	1	0.19517
4	0.00635	0.00902	0.11191	0.10288	1	0.20030
5	0.00844	0.01540	0.12892	0.11352	1	0.16235
6	0.01010	0.02254	0.14542	0.12288	1	0.10823
7	0.01146	0.03024	0.16151	0.13127	1	0.06103
8	0.01260	0.03838	0.17726	0.13887	1	0.02971
9	0.01357	0.04688	0.19271	0.14583	1	0.01268
10	0.01442	0.05567	0.20791	0.15225	0	0.00481
11	0.01517	0.06471	0.22289	0.15818	0	0.00163
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Source: own computations

Table 2. Shortest confidence interval for sample size 82

X	γ_1	left	right	length		$P_{0.05}\{\xi = m\}$
0	0	0	0.03587	0.03587	0	0.01491
1	0	0	0.05655	0.05655	1	0.06433
2	0.00079	0.00049	0.07504	0.07455	1	0.13712
3	0.00370	0.00373	0.09314	0.08941	1	0.19245
4	0.00634	0.00891	0.11058	0.10168	1	0.20004
5	0.00842	0.01520	0.12740	0.11219	1	0.16425
6	0.01008	0.02225	0.14371	0.12146	1	0.11094
7	0.01144	0.02986	0.15961	0.12976	1	0.06339
8	0.01258	0.03789	0.17518	0.13729	1	0.03128
9	0.01355	0.04627	0.19045	0.14418	1	0.01354
10	0.01439	0.05495	0.20548	0.15053	0	0.00520
11	0.01514	0.06388	0.22029	0.15642	0	0.00179
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Source: own computations

Multiplying columns *length* and $P_{0.05}\{\xi = m\}$ we obtain the expected length. For $n = 81$ it is 0.100108 and for $n = 82$: 0.0995025. To obtain expected length exactly equal to prescribed precision 0.1 a randomization is needed. The sample size should be applied in the following way

$$n = \begin{cases} 81, & \textit{with probability } 0.821635, \\ 82, & \textit{with probability } 0.178365. \end{cases}$$

Expected length equals now

$$0.100108 \cdot 0.821635 + 0.0995025 \cdot 0.178365 = 0.1.$$

Of course, drawing sample size should be done before realization of the proper experiment. Any random number generator may be applied, for example the one in Excel.

Zieliński (2010b) showed that the application of the classical Clopper-Pearson confidence interval need a sample of size 90 to fulfill above requirements. Hence, application of the shortest confidence interval needs smaller sample sizes.

As it was mentioned, all calculations may be done in Excel. There are following useful functions.

BETADISTRIBUTION(x;alfa;beta;A;B): where alfa and beta are the parameters of the distribution. The function gives a values of CDF at point x. Numbers A and B defines a support of the distribution: default values are 0 and 1.

BETAINV(probability;alpha;beta;A;B): where alfa and beta are parameters of the distribution. The function gives the probability quantile of the Beta distribution. Numbers A and B defines a support of the distribution: default values are 0 and 1.

The shortest confidence interval in the binomial model may be calculated in the following way.

	A	B
1	100	sample size
2	10	number of successes
3	0.95	confidence level
4	0.01	probability gamma1
5	=IF(OR(A2=0;A2=1);0;BETAINV(A4;A2;A1-A2+1))	left end
6	=IF(OR(A2=A1-1;A2=A1);1;BETAINV(A3+A4;A2+1;A1-A2))	right end
7	=A6-A5	length

To obtain the shortest confidence interval the Addin Solver should be used. The goal is cell A7 by changing A4.

REFERENCES

Clopper C. J., Pearson E. S. (1934) The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, Biometrika 26, 404-413.

Karliński W.(2003) http://nik.gov.pl/docs/kp/kontrola$_$panstwowa-2003$_$05.pdf

Zieliński W. (2010a) The shortest Clopper-Pearson confidence interval for binomial probability, Communications in Statistics - Simulation and Computation 39, 188-193.

Zieliński W. (2010b) Statistical properties of a control design of controls provided by Supreme Chamber of Control, Quantitative Methods in Economics XI, 158-167.