APPLICATIONES MATHEMATICAE
41,1 (2014), pp. 43–49

Wojciech Zieliński (Warszawa)

THE SHORTEST RANDOMIZED CONFIDENCE INTERVAL FOR PROBABILITY OF SUCCESS IN A NEGATIVE BINOMIAL MODEL

Abstract. Zieliński (2012) showed the existence of the shortest confidence interval for a probability of success in a negative binomial distribution. The method of obtaining such an interval was presented as well. Unfortunately, the confidence interval obtained has one disadvantage: it does not keep the prescribed confidence level. In the present article, a small modification is introduced, after which the resulting shortest confidence interval does not have that disadvantage.

Consider the negative binomial (or Pascal) statistical model

$$(\{0,1,2,\ldots\},\{\mathrm{NB}(r,\pi),0<\pi<1\}),$$

where $\mathrm{NB}(r,\pi)$ denotes the negative binomial distribution with pdf

$$\binom{r+x-1}{x}\pi^r(1-\pi)^x,\qquad x=0,1,2,\ldots.$$

It is known that

$$\sum_{x=0}^{t}\binom{r+x-1}{x}\pi^r(1-\pi)^x=F(r,t+1;\pi),$$

where $F(a,b;\cdot)$ denotes the cdf of the beta distribution with parameters (a,b).

Let X denote a negative binomial $\mathrm{NB}(r,\pi)$ random variable. A confidence interval for probability π at confidence level γ is of the form (see Clopper and Pearson's (1934) construction of the confidence interval for π in a binomial statistical model)

$$\big(F^{-1}(r,X+1;\gamma_1);F^{-1}(r,X;\gamma_2)\big),$$

2010 *Mathematics Subject Classification*: Primary 62F25.
Key words and phrases: probability of success, negative binomial distribution, Pascal distribution, confidence interval, shortest confidence interval.

DOI: 10.4064/am41-1-4

where $\gamma_1, \gamma_2 \in (0,1)$ are such that $\gamma_2 - \gamma_1 = \gamma$ and $F^{-1}(a,b;\alpha)$ is the α quantile of the beta distribution with parameters (a,b), i.e.

$$P_\pi\left\{\pi \in \left(F^{-1}(r, X+1; \gamma_1); F^{-1}(r, X; \gamma_2)\right)\right\} \geq \gamma, \quad \forall \pi \in (0,1).$$

For $X = 0$ the right end is taken to be 1.

Zieliński (2012) considered the length of the confidence interval when $X = x$ is observed:

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

Let x be given. The existence as well as the method of finding $0 < \gamma_1 < 1-\gamma$ such that $d(\gamma_1, x)$ is minimal was shown. Exemplary solutions are given in Table 1.

Table 1. $r = 5$

	Shortest c.i.			
x	γ_1	$\text{left}_{\text{short}}$	$\text{right}_{\text{short}}$	$\text{length}_{\text{short}}$
1	0.05000	0.41820	1.00000	0.58180
5	0.02303	0.18339	0.78408	0.60070
10	0.01515	0.10436	0.56211	0.45775
15	0.01263	0.07289	0.43500	0.36210
20	0.01141	0.05600	0.35417	0.29817
25	0.01069	0.04546	0.29849	0.25302
30	0.01021	0.03826	0.25786	0.21960
35	0.00988	0.03303	0.22694	0.19391
40	0.00963	0.02905	0.20262	0.17356
45	0.00944	0.02593	0.18300	0.15706
50	0.00928	0.02342	0.16684	0.14342

The confidence level of the shortest confidence interval for probability π equals

$$\sum_{x=0}^{\infty} \binom{r+x-1}{x} \pi^r (1-\pi)^x \mathbf{1}(x, \pi),$$

where

$$\mathbf{1}(x,\pi) = \begin{cases} 1 & \text{if } \pi \in (\text{left}_{\text{short}}(x), \text{right}_{\text{short}}(x)), \\ 0 & \text{otherwise.} \end{cases}$$

For $r = 5$ and $\gamma = 0.95$ the confidence level is shown in Figure 1.

Note that for some probabilities π the confidence level is smaller than the nominal one. This contradicts the definition of the confidence interval.

Let Y be a random variable conditionally distributed on the interval $[0,1]$ with cdf $G_{Y|X=x}(\cdot)$. The confidence interval will be constructed on the basis of $T_g = X + Y$ and $T_d = X - (1 - Y)$. The distribution of the r.v. T_g

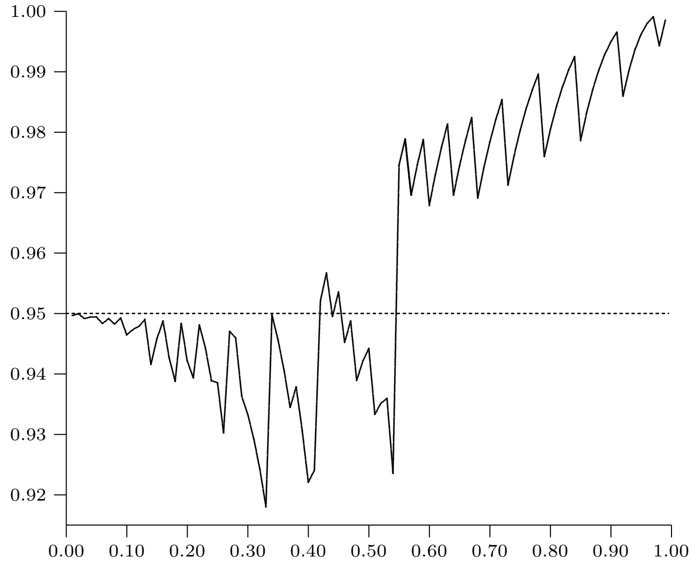

Fig. 1. Confidence level of the shortest confidence interval: $r = 5$, $\gamma = 0.95$

is easy to obtain:

$$P_\pi\{T_g \le t\} = P_\pi\{X + Y \le t\}$$
$$= \begin{cases} \alpha(\lfloor t \rfloor, \lceil t \rceil) F(r, 1; \pi) & \text{for } \lfloor t \rfloor = 0, \\ (1 - \alpha(\lfloor t \rfloor, \lceil t \rceil)) F(r, \lfloor t \rfloor; \pi) + \alpha(\lfloor t \rfloor, \lceil t \rceil) F(r, \lfloor t \rfloor + 1; \pi) & \text{for } \lfloor t \rfloor \ge 1, \end{cases}$$

where $\lfloor t \rfloor$ is the greatest integer not greater than t, $\lceil t \rceil = t - \lfloor t \rfloor$ and $\alpha(x, y) = \int_0^y dG_{Y|X=x}(t)$.

The distribution of T_d may be obtained in a similar way.

Let $X = x$ and $Y = y$ be observed. The shortest confidence interval (π_L, π_U) at confidence level γ will be obtained as a solution with respect to π of the following problem:

$$\begin{cases} \pi_U - \pi_L = \min!, \\ P_{\pi_U}\{T_g \le x + y\} = \gamma_2, \\ P_{\pi_L}\{T_d \ge x - (1 - y)\} = 1 - \gamma_1, \\ \gamma_2 - \gamma_1 = \gamma. \end{cases}$$

Hence, we have to find π_L and π_U such that

$$\begin{cases} \pi_U - \pi_L = \min!, \\ (1 - \alpha(x, y)) F(r, x; \pi_U) + \alpha(x, y) F(r, x + 1; \pi_U) = \gamma_2, \\ (1 - \alpha(x, y)) F(r, x + 1; \pi_L) + \alpha(x, y) F(r, x + 2; \pi_L) = \gamma_1, \\ \gamma_2 - \gamma_1 = \gamma. \end{cases}$$

It is easy to note that the distribution of Y may be taken to be uniform $U(0,1)$ independently of X. Let

$$G(\pi; r, x, y) = (1-y)F(r, x; \pi) + yF(r, x+1; \pi).$$

Then

$$\pi_L = G^{-1}(\gamma_1; r, x+1, y) \quad \text{and} \quad \pi_U = G^{-1}(\gamma+\gamma_1; r, x, y).$$

Consider the length of the confidence interval when $X = x$ and $Y = y$ are observed:

$$d(\gamma_1; r, x, y) = G^{-1}(\gamma+\gamma_1; r, x, y) - G^{-1}(\gamma_1; r, x+1, y).$$

THEOREM 1. *For $x \geq 2$ and for all $y \in [0,1]$ there exists a two-sided shortest confidence interval.*

Proof. We have to show that for $x \geq 2$ and for all $y \in [0,1]$ there exists $0 < \gamma_1 < 1-\gamma$ such that $d(\gamma_1; r, x, y)$ is minimal. The derivative of $d(\gamma_1; r, x, y)$ with respect to γ_1 equals (in what follows, $B(\cdot,\cdot)$ denotes the beta function)

$$\frac{\partial d(\gamma_1; r, x, y)}{\partial \gamma_1} = \frac{1}{\mathrm{LHS}(\gamma_1; r, x, y)} - \frac{1}{\mathrm{RHS}(\gamma_1; r, x, y)}$$

where

$$\mathrm{LHS}(\gamma_1; r, x, y) = \frac{(1-G^{-1}(\gamma+\gamma_1; r, x, y))^{x-1}(G^{-1}(\gamma+\gamma_1; r, x, y))^{r-1}}{B(r,x)} \cdot \left(1 - y + y\frac{x+r}{x}(1-G^{-1}(\gamma+\gamma_1; r, x, y))\right),$$

$$\mathrm{RHS}(\gamma_1; r, x, y) = \frac{(1-G^{-1}(\gamma_1; r, x+1, y))^{x}(G^{-1}(\gamma_1; r, x+1, y))^{r-1}}{B(r,x+1)} \cdot \left(1 - y + y\frac{x+r+1}{x+1}(1-G^{-1}(\gamma_1; r, x+1, y))\right).$$

Because

$$G^{-1}(0; r, x, y) = 0 \quad \text{and} \quad G^{-1}(1; r, x, y) = 1,$$

for $x \geq 2$ we have:

- if $\gamma_1 \to 0$ then $\mathrm{LHS}(\gamma_1; r, x, y) > 0$ and $\mathrm{RHS}(\gamma_1; r, x, y) \to 0^+$,
- if $\gamma_1 \to 1-\gamma$ then $\mathrm{LHS}(\gamma_1; r, x, y) \to 0^+$ and $\mathrm{RHS}(\gamma_1; r, x, y) > 0$.

Therefore, the equation

$$(*) \qquad \frac{\partial d(\gamma_1; r, x, y)}{\partial \gamma_1} = 0$$

has a solution.

It is easy to see that $\mathrm{LHS}(\cdot; r, x, y)$ and $\mathrm{RHS}(\cdot; r, x, y)$ are unimodal and concave on the interval $(0, 1-\gamma)$. Hence, the solution of $(*)$ is unique. Let

γ_1^* denote the solution. Because $\partial d(\gamma_1; r, x, y)/\partial\gamma_1 < 0$ for $\gamma_1 < \gamma_1^*$ and $\partial d(\gamma_1; r, x, y)/\partial\gamma_1 > 0$ for $\gamma_1 > \gamma_1^*$, we have

$$d(\gamma_1^*; r, x, y) = \inf\{d(\gamma_1; r, x, y) : 0 < \gamma_1 < 1 - \gamma\}.$$

THEOREM 2. *For $x = 1$ there exists $y^* \in (0, 1)$ such that if $Y < y^*$ then the shortest confidence interval is one-sided, and is two-sided otherwise.*

Proof. For $x = 1$ we have

$$\begin{aligned}
&\mathrm{LHS}(\gamma_1; r, 1, y) \\
&\quad = r(G^{-1}(\gamma + \gamma_1; r, 1, y))^{r-1}\big(1 - y + y(r+1)(1 - G^{-1}(\gamma + \gamma_1; r, 1, y))\big) \\
&\mathrm{RHS}(\gamma_1; r, 1, y) = r(r+1)(1 - G^{-1}(\gamma_1; r, 2, y))(G^{-1}(\gamma_1; r, 2, y))^{r-1} \\
&\qquad \cdot \left(1 - y + y\frac{r+2}{2}(1 - G^{-1}(\gamma_1; r, 2, y))\right).
\end{aligned}$$

It can be seen that if $\gamma_1 \to 0$, then

$$\begin{aligned}
\mathrm{LHS}(\gamma_1; r, 1, y) &\to r(G^{-1}(\gamma; r, 1, y))^{r-1} \\
&\qquad \cdot \big(1 - y + y(r+1)(1 - G^{-1}(\gamma; r, 1, y))\big), \\
\mathrm{RHS}(\gamma_1; r, 1, y) &\to 0,
\end{aligned}$$

and if $\gamma_1 \to 1 - \gamma$, then

$$\begin{aligned}
\mathrm{LHS}(\gamma_1; r, 1, y) &\to r(1 - y), \\
\mathrm{RHS}(\gamma_1; r, 1, y) &\to r(r+1)(1 - G^{-1}(1-\gamma; r, 2, y))(G^{-1}(1-\gamma; r, 2, y))^{r-1} \\
&\qquad \cdot \left(1 - y + y\frac{r+2}{2}(1 - G^{-1}(1-\gamma; r, 2, y))\right).
\end{aligned}$$

Because $\mathrm{LHS}(1-\gamma; r, 1, 0) > \mathrm{RHS}(1-\gamma; r, 1, 0)$ and $\mathrm{LHS}(1-\gamma; r, 1, 1) < \mathrm{RHS}(1-\gamma; r, 1, 1)$, there exists y^* such that

$$\mathrm{LHS}(1-\gamma; r, 1, y^*) = \mathrm{RHS}(1-\gamma; r, 1, y^*).$$

So, for $y < y^*$ the shortest confidence interval is one-sided, and it is two sided otherwise.

The probability y^* may be found numerically as a solution of

$$\mathrm{LHS}(1-\gamma; r, 1, y^*) = \mathrm{RHS}(1-\gamma; r, 1, y^*).$$

In Table 2 the values of y^* for different r and confidence levels γ are given.

The confidence level of the randomized shortest confidence interval for $r = 5$ and $\gamma = 0.95$ is shown in Figure 2.

Below we give a short program in the `R` language for calculating γ_1^* and the ends of the shortest randomized confidence interval. Of course, one can also use any other mathematical or statistical package (in a similar way) to find the values of γ_1^* as well as the ends of the shortest randomized confidence interval (cf. Zieliński 2010).

Table 2. Values of y^*

r	0.9	0.95	0.99	r	0.9	0.95	0.99
3	0.67728	0.77537	0.91232	17	0.90240	0.94544	0.98672
4	0.76309	0.84507	0.94730	18	0.90414	0.94657	0.98708
5	0.80700	0.87858	0.96201	19	0.90568	0.94757	0.98739
6	0.83322	0.89782	0.96979	20	0.90706	0.94846	0.98767
7	0.85051	0.91017	0.97452	21	0.90829	0.94926	0.98792
8	0.86271	0.91871	0.97765	22	0.90940	0.94998	0.98814
9	0.87174	0.92495	0.97987	23	0.91041	0.95062	0.98834
10	0.87870	0.92969	0.98152	24	0.91133	0.95121	0.98852
11	0.88421	0.93341	0.98279	25	0.91217	0.95175	0.98868
12	0.88868	0.93640	0.98379	26	0.91294	0.95224	0.98883
13	0.89238	0.93886	0.98460	27	0.91365	0.95269	0.98897
14	0.89548	0.94091	0.98527	28	0.91431	0.95311	0.98909
15	0.89813	0.94265	0.98583	29	0.91491	0.95350	0.98921
16	0.90041	0.94415	0.98631	30	0.91548	0.95386	0.98932

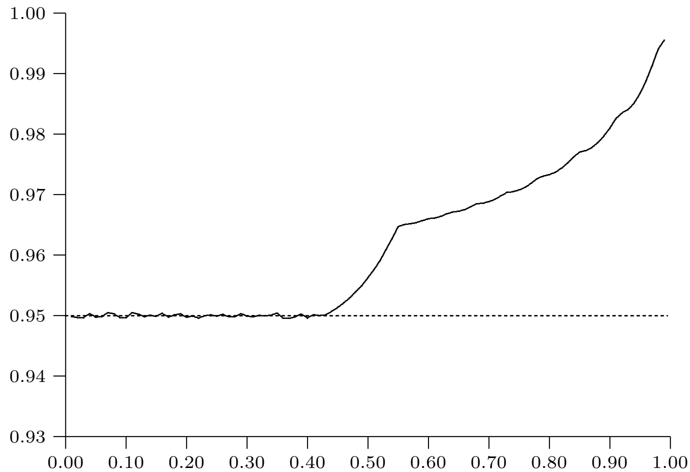

Fig. 2. Confidence level of the randomized shortest confidence interval: $r = 5$, $\gamma = 0.95$

```
Bet = function(a,b,q){pbeta(q,a,b)} #Beta CDF
Dys = function(r,x,q,y){if (x>1) (1-y)*Bet(r,x,q)+y*Bet(r,x+1,q)
  else y*Bet(r,1,q)}
Left = function(r,x,y,p){uniroot(function(q) Dys(r,x+1,q,y)-p,
  lower = 0, upper = 1, tol = 1e-20)$root}
Right = function(r,x,y,p){uniroot(function(q) Dys(r,x,q,y)-p,
  lower = 0, upper = 1, tol = 1e-20)$root}
Leng = function(r,x,y,q,s){Right(r,x,y,q+s)-Left(r,x,y,s)}
```

```
Ystar=function(r,level){uniroot(function(a)
  (1 - a)*r - r*(r + 1)*(1 - Left(r,1,a,1-level))*exp((r - 1)
    *log(Left(r,1,a,1-level)))*
  (1 - a + a*(r + 2)*(1 - Left(r,1,a,1-level))/2),
  lower = 0,upper = 1,tol = 1e-20)$root}
FindMinimumLeng = function(r,x,y,q){optimize(Leng,interval=c(0,1-q), r=r,
      x=x, q=q, y=y, tol=1e-20)$minimum}
r=5; #input number of successes
x=2; #input number of fails
level=0.95; #input confidence level
y=runif(1, 0, 1); #random U(0,1) number
y #output y
ss=if (x+y<1+Ystar(r,level)) 1-level else FindMinimumLeng(r,x,y,level)
ss #output γ1*
Left(r,x,y,ss) #output left end
if (x+y<1+Ystar(r,level)) 1 else Right(r,x,y,level+ss) #output right end
```

Because calculating confidence intervals is very easy with the aid of computer software, using the shortest confidence interval is recommended, especially for small values of r. To avoid problems with wrong inference due to the confidence level, one should use randomized shortest confidence intervals. Of course, the generated value y of the $U(0,1)$ r.v. must be attached to the final report. So results now are given by three numbers: number of successes, number of fails and the value y.

References

C. J. Clopper and E. S. Pearson (1934), *The use of confidence or fiducial limits illustrated in the case of the binomial*, Biometrika 26, 404–413.

W. Zieliński (2010), *The shortest Clopper–Pearson confidence interval for binomial probability*, Comm. Statist. Simulation Comput. 39, 188–193.

W. Zieliński (2012), *The shortest confidence interval for the probability of success in a negative binomial model*, Appl. Math. (Warsaw) 30, 143–149.

Wojciech Zieliński
Department of Econometrics and Statistics
Warsaw University of Life Sciences
Nowoursynowska 159
02-776 Warszawa, Poland
E-mail: wojciech_zielinski@sggw.pl

Received on 29.3.2013;
revised version on 16.1.2014 (2176)