DMPS Page

Discussiones Mathematicae
Probability and Statistics 38 (2018) 15–19
doi:10.7151/dmps.1200

EXACT CONFIDENCE INTERVALS AND RECTANGLES FOR THE ENDPOINTS OF THE UNIFORM DISTRIBUTION

Iosif Pinelis

Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931, USA

e-mail: ipinelis@mtu.edu

Abstract

Exact confidence intervals for each of the endpoints a and b of the uniform distribution on the interval $[a, b]$ with unknown a and b, as well as an exact confidence rectangle for the pair (a, b), are given.

Keywords: confidence intervals, confidence rectangles, uniform distribution.

2010 Mathematics Subject Classification: 62F25.

Let $X_1, \ldots, X_n$ be independent identically distributed (iid) random variables, each uniformly distributed on the interval $[a, b]$, for some unknown real a and b such that $a < b$. We are assuming that $n \geqslant 2$.

Let $Y_i := \frac{X_i - a}{b - a}$, so that the Y_i's are iid, each uniformly distributed on $[0, 1]$, and for the corresponding order statistics one has $X_{(i)} = a + (b - a)Y_{(i)}$. Let

$$R_n := X_{(n)} - X_{(1)},$$

the sample range. Then, for any real $c > 0$,

$$\alpha := \mathsf{P}\left(X_{(1)} > a + cR_n\right) = \mathsf{P}\left(Y_{(1)} > (Y_{(n)} - Y_{(1)})c\right) = \mathsf{P}\,\mathfrak{B}\left(Y_{(n)} < Y_{(1)}\frac{1 + c}{c}\right).$$

The joint probability density of $(Y_{(1)}, Y_{(n)})$ is given by the formula $g(z_1, z_n) = n(n-1)(z_n - z_1)^{n-2}\,\mathrm{I}\{0 < z_1 < z_n < 1\}$, where $\mathrm{I}\{A\}$ denotes the indicator of an assertion A. So,

$$
\begin{aligned}
\alpha &= \int_0^1 n\,dz_1 \int_{z_1}^{1\wedge[z_1(1+c)/c]} dz_n\,(n-1)(z_n-z_1)^{n-2} \\
&= \int_0^1 n\,dz_1\,[(1-z_1)\wedge(z_1/c)]^{n-1} \qquad (1)\\
&= \int_{c/(c+1)}^1 n\,dz_1\,(1-z_1)^{n-1} + \int_0^{c/(c+1)} n\,dz_1\,(z_1/c)^{n-1} = \frac{1}{(1+c)^{n-1}},
\end{aligned}
$$

whence

$$c = c_\alpha := \alpha^{-1/(n-1)} - 1$$

and

$$\mathsf{P}\big(X_{(1)} - c_\alpha R_n < a < X_{(1)}\big) = \mathsf{P}\big(X_{(1)} - c_\alpha R_n \leqslant a\big) = 1-\alpha,$$

for each $\alpha \in (0,1)$. Therefore and in view of symmetry, one has

Proposition 1.

$[X_{(1)} - c_\alpha R_n, X_{(1)}]$ *is an exact* $(1-\alpha)$*-confidence interval for* a*.*

$[X_{(n)}, X_{(n)} + c_\alpha R_n]$ *is an exact* $(1-\alpha)$*-confidence interval for* b*.*

Using the Bonferroni rule, we obtain the following.

Corollary 2. *For any* $\alpha \in (0,1)$*,*

$$[X_{(1)} - c_{\alpha/2} R_n, X_{(1)}] \times [X_{(n)}, X_{(n)} + c_{\alpha/2} R_n] \tag{2}$$

is a $(1-\alpha)$*-confidence rectangle for the pair* (a,b)*, in the sense that the point* (a,b) *is contained in this (random) rectangle with probability* at least $1-\alpha$.

Consider now the "joint probability"

$$p(c) := \mathsf{P}\big(a \in [X_{(1)} - cR_n, X_{(1)}], b \in [X_{(n)}, X_{(n)} + cR_n]\big), \tag{3}$$

again for real $c > 0$. By a calculation similar to, but a bit more involved than, (1), one can obtain the following rather simple expression for $p(c)$:

$$p(c) = 1 - 2(1+c)^{1-n} + (1+2c)^{1-n}. \tag{4}$$

Indeed, letting $\ell(z_1) := 1 \wedge \left(\frac{(1+c)z_1}{c} \vee \frac{1+cz_1}{1+c}\right)$, by (3) we have

$$\begin{aligned}
p(c) &= \mathsf{P}\left(Y_{(1)} < c(Y_{(n)} - Y_{(1)}), Y_{(n)} > 1 - c(Y_{(n)} - Y_{(1)})\right) \\
&= \int_0^1 n\,dz_1 \int_{\ell(z_1)}^1 dz_n\,(n-1)(z_n - z_1)^{n-2} \\
&= \int_0^{c/(1+2c)} n\,dz_1\,(1-z_1)^{n-1}\left[1-(1+c)^{1-n}\right] \\
&+ \int_{c/(1+2c)}^{c/(1+c)} n\,dz_1\,\left[(1-z_1)^{n-1} - z_1^{n-1}c^{1-n}\right] \\
&= 1 - 2(1+c)^{1-n} + (1+2c)^{1-n}.
\end{aligned}$$

Using (3) or (4), it is easy to see that $p(c)$ continuously increases from 0 to 1 as c increases from 0 to ∞. So, given any natural $n \geqslant 2$ and any real $\alpha \in (0,1)$, it is easy to find (numerically) the unique positive real root, $\tilde{c}_\alpha = \tilde{c}_{n,\alpha}$, of the equation

$$p(\tilde{c}_\alpha) = 1 - \alpha, \tag{5}$$

and $\tilde{c}_\alpha$ continuously decreases in $\alpha \in (0,1)$. Thus, we have

Theorem 3.

$$[X_{(1)} - \tilde{c}_\alpha R_n, X_{(1)}] \times [X_{(n)}, X_{(n)} + \tilde{c}_\alpha R_n]$$

is an exact $(1-\alpha)$*-confidence rectangle for the pair* (a,b)*.*

Cf. the "excessive", "conservative" $(1-\alpha)$-confidence rectangle (2).

Let us now give explicit upper bounds on the solution $\tilde{c}_\alpha$ of equation (5); we shall see that these upper bounds are actually quite close to $\tilde{c}_\alpha$. First here, note that we always have $\tilde{c}_\alpha < c_{\alpha/2}$: the coefficient $\tilde{c}_\alpha$ for the exact $(1-\alpha)$-confidence rectangle is less than the "excessive" Bonferroni coefficient $c_{\alpha/2}$; this can also be supported by the following simple analytical argument: in view of (4),

$$p(c_{\alpha/2}) = 1 - \alpha + (1 + 2c_{\alpha/2})^{1-n} > 1 - \alpha = p(\tilde{c}_\alpha).$$

Moreover,

$$\tilde{c}_\alpha < c_{1-\sqrt{1-\alpha}} < c_{\alpha/2+\alpha^2/8} < c_{\alpha/2}. \tag{6}$$

Indeed, the latter two inequalities in (6) follow because $1 - \sqrt{1-\alpha} > \alpha/2 + \alpha^2/8 > \alpha/2$ and c_α decreases in α. Concerning the first inequality in

(6), note that, by (5) and (4), for $c = \tilde{c}_\alpha$ we have $1-\alpha = p(c) > 1-2(1+c)^{1-n} + (1+c)^{2(1-n)} = (1-(1+c)^{1-n})^2$, whence $(1+c)^{1-n} > 1-\sqrt{1-\alpha}$, so that indeed $\tilde{c}_\alpha = c < c_{1-\sqrt{1-\alpha}}$.

Actually, $1-\sqrt{1-\alpha}$ is the best (that is, largest) possible value (say β) such that inequality $\tilde{c}_\alpha < c_\beta$ holds for all n. More specifically, for each $\beta \in (1-\sqrt{1-\alpha}, 1)$ one has $\tilde{c}_\alpha > c_\beta$ eventually – that is, for all large enough n. Indeed, take any $\beta \in (1-\sqrt{1-\alpha}, 1)$. Letting $n \to \infty$, by l'Hospital's rule we have

$$c_\beta = \beta^{-1/(n-1)} - 1 = \frac{\ln \beta}{1-n}\,(1+o(1)), \tag{7}$$

so that $(1+2c_\beta)^{1-n} = (1+\frac{(2+o(1))\ln\beta}{1-n})^{1-n} \to \beta^2$, and hence $p(c_\beta) \to 1-2\beta+\beta^2 = (1-\beta)^2 < 1-\alpha = p(\tilde{c}_\alpha)$. Therefore and because $p(c)$ is increasing in $c > 0$, it is now confirmed that the inequality $\tilde{c}_\alpha > c_\beta$ holds eventually.

It appears that the coefficient $\tilde{c}_\alpha$ differs rather little from the somewhat excessive "Bonferroni" coefficient $c_{\alpha/2}$, and then of course $\tilde{c}_\alpha$ differs even less from $c_{\alpha/2+\alpha^2/8}$ and, especially, $c_{1-\sqrt{1-\alpha}}$. This is illustrated in Table 1.

α	n	$\tilde{c}_\alpha$	$c_{1-\sqrt{1-\alpha}}$	$c_{\alpha/2+\alpha^2/8}$	$c_{\alpha/2}$
0.05	10	0.500243	0.504499	0.504552	0.50663
	100	0.0378116	0.0378307	0.0378341	0.0379643
	1000	0.00368642	0.0036866	0.00368692	0.0036994
0.01	10	0.797947	0.801146	0.801148	0.801648
	100	0.0549413	0.0549496	0.0549498	0.0549764
	1000	0.00531511	0.00531518	0.0053152	0.00531771

Table 1. Approximate values of $\tilde{c}_\alpha, c_{1-\sqrt{1-\alpha}}, c_{\alpha/2+\alpha^2/8}, c_{\alpha/2}$ for $\alpha \in \{0.05, 0.01\}$ and $n \in \{10, 100, 1000\}$.

So, it appears reasonable to use $c_{1-\sqrt{1-\alpha}} = (1-\sqrt{1-\alpha}\,)^{-1/(n-1)} - 1$ as the initial, and already good, approximation to the root $\tilde{c}_\alpha$ of equation (5).

One may also note that, in view of (7), the coefficient c_α and hence the length $c_\alpha R_n$ of the confidence interval decrease, roughly, inversely proportionally to n for large n.

In conclusion, let us mention some related results found in the literature. In [2], following [4], an exact confidence interval for the real parameter θ was given, for a family of densities of the form $f_\theta(x) = \frac{g(x)}{h(\theta)}\,\mathrm{I}\{-a(\theta) \leqslant x \leqslant b(\theta)\}$, where $a(\theta)$ and $b(\theta)$ are either both increasing or both decreasing in θ. A particular case of this setting is that of the family of the uniform distributions on the interval

$[-\theta, \theta]$ with a unknown $\theta > 0$. In [1], an exact confidence interval for the standard deviation of a uniform distribution was obtained, which was an improvement on earlier results in [3]. However, this author is not aware of any results concerning confidence rectangles for the two completely unknown endpoints of a uniform distribution.

REFERENCES

[1] H. Leon Harter, *The use of sample ranges in setting exact confidence bounds for the standard deviation of a rectangular population*, J. Amer. Statist. Assoc. **56** (1961) 601–609.

[2] V.S. Huzurbazar, *Confidence intervals for the parameter of a distribution admitting a sufficient statistic when the range depends on the parameter*, J. Roy. Statist. Soc. Ser. B **17** (1955) 86–90.

[3] F.C. Leone, Y.H. Rutenberg and C.W. Topp, *The use of sample quasi-ranges in setting confidence intervals for the population standard deviation*, J. Amer. Statist. Assoc. **56** (1961) 260–272.

[4] E.J.G. Pitman, *Sufficient statistics and intrinsic accuracy*, Proc. Camb. Phil. Soc. **32** (1936) 567–579.

Received 2 January 2018
Revised 23 February 2018