2014 Hawaii University International Conferences
Science, Technology, Engineering, Math & Education
June 16, 17, & 18 2014
Ala Moana Hotel, Honolulu, Hawaii

The Tail Functions Approach to Confidence Estimation of a Uniform Parameter

Puza, Borek
Australian National University
Research Sch. of Finance, Actuarial Studies &
Applied Statistics

Borek Puza
Research Sch. of Finance, Actuarial Studies & Applied Statistics,
Australian National University

The Tail Functions Approach to Confidence Estimation of a Uniform Parameter

Synopsis:

This paper reviews the method of tail functions (TFs) for constructing confidence intervals (CIs). The idea is to generalize the constant cutoffs (e.g. 0.025 and 0.975) in the definition of a CI to functions of the target parameter. The ordinary 2- and 1-sided CIs are defined by the constant TFs 1/2, 0 and 1. Under prior information, the TFs approach is an alternative to the Bayesian and can lead to improved CIs, whilst retaining frequentist coverage probabilities.

The tail functions approach to confidence estimation of a uniform parameter

Borek Puza

Australian National University, Canberra

13 May 2014

Abstract

This paper illustrates how the method of tail functions can be used to construct confidence intervals for the upper bound parameter of a uniform distribution. It shows how this method generalizes classical confidence intervals and can be used to solve problems which arise when the uniform parameter is constrained in an interval. Other approaches are also discussed.

Key words and phrases

Bayesian inference; confidence interval; constraint; optimization; prior; tail function.

MSC 2000

Primary 62F25; secondary 62F15.

1. Introduction

The method of tail functions (Puza and O'Neill, 2006b) is an alternative to the classical (or standard) approach to confidence estimation (Stuart, Ord and Arnold, 1999). The classical theory involves inverting a probability statement regarding a pivotal quantity so as to yield two statistics which form the lower and upper bounds of a confidence interval for the parameter of interest. This theory requires the specification of a confidence level (e.g. 95%) which 'cuts off' equal areas (e.g. 2.5%) from either tail of the pivotal quantity's distribution. The tail functions approach involves specifying these areas not as constants but in terms of a function of the target parameter itself - called the tail function - and then proceeding to invert the probability statement so as to yield an alternative confidence interval for the parameter.

A key feature of the tail functions approach is that it is highly flexible - by virtue of the many ways in which a tail function may be specified - and can often be applied to produce confidence intervals with attractive properties, in some sense. For example, in the presence of prior information, tail functions can be used to engineer an interval with smaller prior expected size than the standard interval, whilst retaining its frequentist coverage probabilities. Tail functions thereby provide an alternative to the Bayesian approach (Lee, 1997), which typically yields an even shorter interval but one without the frequentist coverage properties of a classical confidence interval. Another and related use of tail functions is to address the problem of a confidence interval being empty or being a single point with nonzero probability. This problem sometimes arises when a model parameter is constrained within some interval.

Previous studies have explored how tail functions can be used to improve confidence intervals for the normal mean (Puza and O'Neill, 2005, 2006b, 2008), the binomial proportion (Puza and O'Neill, 2006a, 2009), the Poisson mean (Puza and Yang, 2011) and the exponential mean (Puza and Yang, 2013). Applications of the theory include survey sampling (Puza, 2011) and studies of mortality and air pollution (Puza, Roberts and Yang, 2011).

This paper further illustrates the theory of tail functions by applying it to confidence estimation of the upper bound parameter of a uniform distribution with lower bound zero. The theory is found to be especially useful in situations where the parameter is constrained from below or in a finite interval. In these cases, the classical approach leads to a confidence interval which could be empty. This problem can be solved via the tail functions approach.

Section 2 describes the classical approach to confidence estimation of the upper bound parameter of a uniform distribution whose lower bound is zero, and discusses the issues which arise when that parameter is constrained from below, from above, and in a finite positive interval, respectively. Section 3 shows how these issues can be resolved via the tail functions approach. Section 4 shows how tail functions can be used to engineer a suitable confidence interval when prior information is available and one wishes to minimize the prior expected size of the interval over a class of tail functions. Section 5 is a summary, with discussion of other approaches, including the Bayesian, and suggestions for further research.

2. Classical confidence intervals for a uniform parameter

Suppose we observe *x*, a value of the random variable X which is uniformly distributed between 0 and θ, where θ is an unknown nonnegative constant. Thus $X \sim U(0,\theta)$, where $\theta \geq 0$. We wish to construct a $1-\alpha$ confidence interval (CI) for the uniform parameter θ. To this end, we form the pivotal quantity $Y = X/\theta \sim U(0,1)$ and observe that for any $\theta > 0$:

$$1-\alpha = P\left(\frac{\alpha}{2} \leq Y \leq 1-\frac{\alpha}{2}\right) = P\left(\frac{\alpha}{2} \leq \frac{X}{\theta} \leq 1-\frac{\alpha}{2}\right) \qquad (1)$$

To obtain a $1-\alpha$ CI for θ, we now invert this probability statement, meaning manipulate it so that θ is in the middle and 'straddled' by two statistics, $L(X)$ (the lower bound, LB) and $U(X)$ (the upper bound, UB). This can be done as follows:

$$1-\alpha = P\left(\frac{\alpha}{2}\theta \leq X \leq \theta\left(1-\frac{\alpha}{2}\right)\right) \qquad \text{(after multiplying both sides by } \theta\text{)}$$

$$= P\left(\frac{\alpha}{2}\theta \leq X,\ X \leq \theta\left(\frac{2-\alpha}{2}\right)\right) \qquad \text{(where the comma means "and")}$$

$$= P\left(\theta \leq X\frac{2}{\alpha},\ X\left(\frac{2}{2-\alpha}\right) \leq \theta\right)$$

$$= P\left(X\left(\frac{2}{2-\alpha}\right) \leq \theta \leq X\frac{2}{\alpha}\right)$$

$$= P\left(L(X) \leq \theta \leq U(X)\right)$$

where $L(X) = \dfrac{2X}{2-\alpha}$ and $U(X) = \dfrac{2X}{\alpha}$ are the two statistics required.

We see that a two-sided $1-\alpha$ CI for θ is

$$[l,u] = [L(x),U(x)] = \left[\frac{x}{1-\alpha/2}, \frac{2x}{\alpha}\right]$$

The size (or equivalently, length or width or volume) of this CI is

$$s = S(x) = U(x) - L(x) = x\left(\frac{2}{\alpha} - \frac{1}{1-\alpha/2}\right) = \frac{4(1-\alpha)}{\alpha(2-\alpha)}x$$

Note

Generally, suppose that we observe x, the realized value of a random variable X whose distribution depends on a parameter θ, where each of x and θ may be a scalar or a vector. Then a $1-\alpha$ confidence set (CS) for θ is defined as any set $CS(x)$ such that

$$P\left(\theta \in CS(X)\right) = 1-\alpha$$

for all possible values of θ. The size (or equivalently, volume) of $CS(x)$ is defined as

$$S(x) = \int I(\theta \in CS(x))d\theta$$

In the special (and very common) case where $CS(x)$ is a single interval, we may refer to $CS(x)$ as a confidence interval (CI) and write $CS(x)$ as $CI(x) = [L(x), U(x)]$, where $L(x)$ is the lower bound (LB) and $U(x)$ is the upper bound (UB). In that case, the size of $CI(x)$ is

$$S(x) = \int_{L(x)}^{U(x)} 1 d\theta = U(x) - L(x)$$

and may be referred to as the length (or equivalently, width) of the CI.

A special case of a CS is the null set $\varnothing$ (phi), also known as the empty set $\{\}$. Whenever a CS is $\varnothing$, we may refer to that CS as being empty. In that case, the size of the CS is 0 (zero).

Another special case of a CS is the singleton set $\{t\}$, where t is a single value (scalar or vector). In that case the size of the CS is also 0 (zero) (i.e. the same as if the CS were empty).

In this paper we are primarily interested in CIs rather than CSs. However, as we shall see, when attempting to calculate a CI we sometimes obtain a set which is not actually an interval. But rather than labor the distinction, we will use the terms "confidence interval" and "confidence set" interchangeably. Likewise, the words "length", "width", "size" and "volume" will be used synonymously, as also will descriptors such as "short", "narrow" and "small", etc.

Other classical confidence intervals

Other CIs for θ can be constructed by choosing a constant $\tau \in [0,1]$ and rewriting (1) as

$$1-\alpha = P\left(\alpha\tau \le Y \le 1-\alpha+\alpha\tau\right) = P\left(\alpha\tau \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau\right) \quad (2)$$

(since $Y = X/\theta \sim U(0,1)$ and $\left(1-\alpha+\alpha\tau\right)-\alpha\tau = 1-\alpha$)

$$= P\left(\alpha\tau\theta \le X \le \theta\left(1-\alpha+\alpha\tau\right)\right)$$

$$= P\left(\alpha\tau\theta \le X, X \le \theta\left(1-\alpha+\alpha\tau\right)\right)$$

$$= P\left(\theta \le \frac{X}{\alpha\tau}, \frac{X}{1-\alpha+\alpha\tau} \le \theta\right)$$

$$= P\left(\frac{X}{1-\alpha+\alpha\tau} \le \theta \le \frac{X}{\alpha\tau}\right)$$

$$= P\left(L(X) \le \theta \le U(X)\right)$$

where $L(X) = \dfrac{X}{1-\alpha(1-\tau)}$ and $U(X) = \dfrac{X}{\alpha\tau}$

So another $1-\alpha$ CI (or rather, class of CIs) for θ is

$$[l,u] = [L(x),U(x)] = \left[\frac{x}{1-\alpha(1-\tau)}, \frac{x}{\alpha\tau}\right]$$

This formula defines an infinite number of $1-\alpha$ CIs for θ, where each one is indexed by a value of $\tau \in [0,1]$. Thus, we could also write *L*, *U*, *l* and *u* as L_τ, U_τ, l_τ and u_τ, respectively.

We see that the two-sided CI above is a special case of this class, defined by τ = 1/2. Two other special cases are defined by τ = 0 and τ = 1. These specifications lead to the following one-sided CIs, respectively:

$\left[\dfrac{x}{1-\alpha}, \infty\right)$ (the upper range $1-\alpha$ CI for θ, consisting of the largest possible values)

$\left[x, \dfrac{x}{\alpha}\right]$ (the lower range $1-\alpha$ CI for θ, consisting of the smallest possible values)

An important observation here is that the CI indexed by τ has size

$$s = \frac{x}{\alpha\tau} - \frac{x}{1-\alpha(1-\tau)} = x\left(\frac{1-\alpha}{\alpha}\right)\tau^{-1}[1-\alpha(1-\tau)]^{-1}$$

By the chain rule for differentiation, the slope of this size (as a function of τ) is

$$\frac{ds}{d\tau} = x\left(\frac{1-\alpha}{\alpha}\right)\left\{\tau^{-1}(-1)[1-\alpha(1-\tau)]^{-2}\alpha + (-\tau^{-2})[1-\alpha(1-\tau)]^{-1}\right\}$$

which is negative for all values of τ (and x and α). So $\tau = 1$ uniformly minimizes the length.

So, in this sense, the optimal $1-\alpha$ CI for θ is $[l,u] = \left[x, \frac{x}{\alpha}\right]$ with a size of $s = \left(\frac{1-\alpha}{\alpha}\right)x$

Note: This CI can also be arrived at in other ways within the framework of classical inference.

Example 1 Some classical confidence intervals

Figure 1 below shows three sets of classical 80% CIs for θ. (Thus, we are setting $\alpha = 0.2$.) These sets are defined by τ = 0, 0.5 and 1. For the data x = 1.5, the CIs work out as:

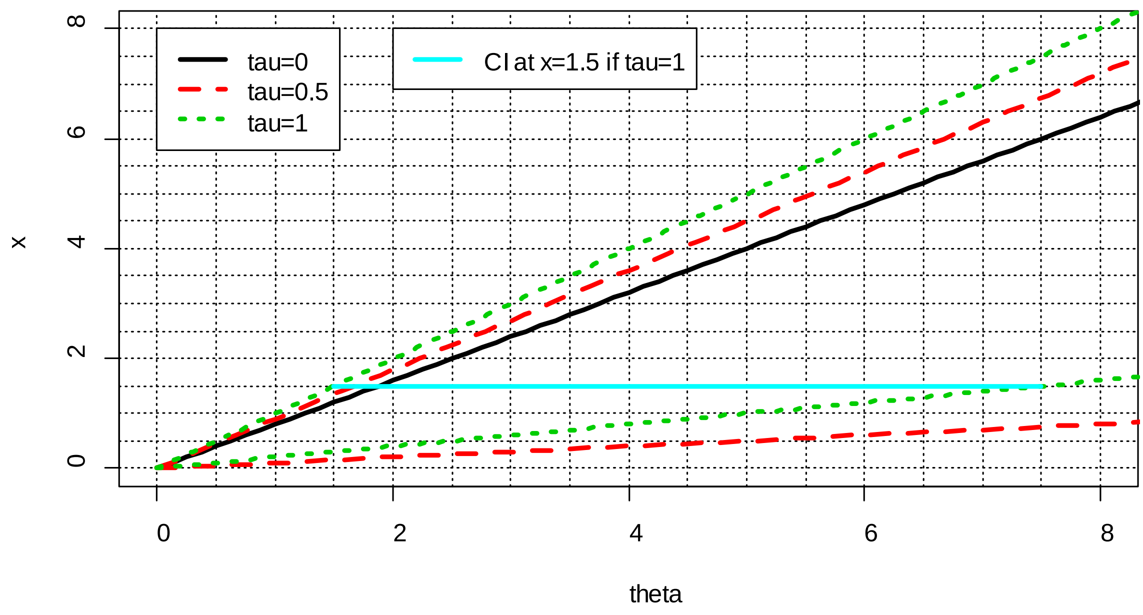
$$\left[\frac{x}{1-\alpha}, \infty\right) = \left[\frac{1.5}{1-0.2}, \infty\right) \qquad = [1.875, \infty) \quad \text{if } \tau = 0$$

$$\left[\frac{x}{1-\alpha(1-\tau)}, \frac{x}{\alpha\tau}\right] = \left[\frac{1.5}{1-0.2(1-0.5)}, \frac{1.5}{0.2\times 0.5}\right] \qquad = [1.6667, 15] \text{ if } \tau = 0.5$$

$$\left[x, \frac{x}{\alpha}\right] = \left[1.5, \frac{1.5}{0.2}\right] \qquad = [1.5, 7.5] \quad \text{if } \tau = 1$$

The CI implied by x = 1.5 and τ = 1 is shown in the figure as a light blue horizontal line. This CI is the smallest (or narrowest) of the three, and its size (or length or width) is $s = S(1.5) = 7.5 - 1.5$ = 6.

Figure 1 Three classical 80% confidence intervals



Classical interval estimation under constraints

So far we have described the classical approach to estimation of θ when nothing is known about that parameter. We will now consider how this approach adapts to when θ is constrained, first from below, secondly from above, and thirdly from both above and below.

Case 1: When the uniform parameter is bounded from below

Suppose that the uniform parameter θ is bounded from below by some constant a > 0. Then the unconstrained interval $[l,u]=\left[\frac{x}{1-\alpha(1-\tau)},\frac{x}{\alpha\tau}\right]$ is still a $1-\alpha$ CI for θ, but with an understanding that all impossible values (namely those less than a) should be excluded.

To determine an explicit formula for this interval, we need to consider two possibilities for τ .

First, we find that if $\tau = 0$, the constrained $1-\alpha$ CI for θ is given by

$$\begin{cases} [a, \infty), & 0 \le x \le (1-\alpha)a \\ \left[\dfrac{x}{1-\alpha}, \infty\right), & (1-\alpha)a < x < \infty \end{cases}$$

and secondly, if $\tau > 0$, the CI is given by

$$\begin{cases} \varnothing, & 0 \le x < \alpha\tau a \\ \left[a, \dfrac{x}{\alpha\tau}\right], & \alpha\tau a \le x \le [1-\alpha(1-\tau)]a \\ \left[\dfrac{x}{1-\alpha(1-\tau)}, \dfrac{x}{\alpha\tau}\right], & [1-\alpha(1-\tau)]a < x < \infty \end{cases}$$

We see there is now a possibility of the CI being empty, and the probability of this is

$$P(0 \le x < a\alpha\tau) = \frac{a\alpha\tau}{\theta} \quad \text{(since } x \sim U(0,\theta)\text{)}$$

This may be small if $a\alpha\tau$ is small (relative to θ) but is guaranteed to be zero only if $\tau = 0$.

It follows that the only constrained CI which cannot be empty is the one-sided interval

$$\begin{cases} [a, \infty), & 0 \le x \le (1-\alpha)a \\ \left[\dfrac{x}{1-\alpha}, \infty\right), & (1-\alpha)a < x < \infty \end{cases}$$

But this CI has an obvious problem, in that it is infinitely wide for all possible values of *x*.

One way to obtain a finite CI which cannot be empty is to set τ > 1 and then redefine the CI as the single point *a* (equivalently the interval $[a,a]$, or the singleton set $\{a\}$) if $0 \le x < \alpha\tau a$. (Note that previously the classical CI was already defined as the single point *a* at $x = \alpha\tau a$.)

However, a problem with this ad hoc solution is that the probability of the CI being a single point (i.e. a singleton set) is then the same as it previously had of being empty, namely $a\alpha\tau / \theta$. Also, the resulting CI then has a frequentist coverage probability of 100% (rather than the nominal or intended level of $1-\alpha$) at $\theta = a$.

Case 2: When the uniform parameter is bounded from above

Next suppose that the uniform parameter θ is bounded from above by some constant $b > 0$. Then $[l,u]=\left[\frac{x}{1-\alpha(1-\tau)},\frac{x}{\alpha\tau}\right]$ is again a $1-\alpha$ CI for θ, but with an understanding that all impossible values (now those greater than b) should be excluded.

As for Case 1, we need to consider two possibilities for τ. We find that if $\tau=0$, the constrained $1-\alpha$ CI for θ is

$$\begin{cases}\left[\frac{x}{1-\alpha},b\right] & 0\le x\le(1-\alpha)b \\ \varnothing, & (1-\alpha)b<x\le b\end{cases}$$

and if $\tau>0$, the CI is given by

$$\begin{cases}\left[\frac{x}{1-\alpha(1-\tau)},\frac{x}{\alpha\tau}\right], & 0\le x\le\alpha\tau b \\ \left[\frac{x}{1-\alpha(1-\tau)},b\right], & \alpha\tau b<x\le[1-\alpha(1-\tau)]b \\ \varnothing, & [1-\alpha(1-\tau)]b<x\le b\end{cases}$$

As for Case 1, this CI could be empty, and the probability of this happening is now

$$P(x>[1-\alpha(1-\tau)]b)=\begin{cases}0, & a\le\theta\le[1-\alpha(1-\tau)]b \\ \frac{\theta-[1-\alpha(1-\tau)]b}{\theta}, & [1-\alpha(1-\tau)]b<\theta\le b\end{cases}$$

This is guaranteed to be zero if and only if $\tau=1$.

So the only CI which cannot be empty is $[l,u]=\begin{cases}\left[x,\frac{x}{\alpha}\right], & 0\le x\le\alpha b \\ [x,b], & \alpha b<x\le b\end{cases}$

This CI is also attractive in that it is relatively the shortest amongst all CIs indexed by $0\le\tau\le1$ and has zero probability of being a single point. Note that the CI is a single point only if $X = 0$ or $X = b$; and for any $\theta>0$, each of these events occurs with probability zero.

Case 3: When the uniform parameter is bounded from above and below

Now suppose that the uniform parameter θ is bounded from below by some constant $a > 0$ and also from above by some finite constant $b > a$. Then $[l,u]=\left[\frac{x}{1-\alpha(1-\tau)},\frac{x}{\alpha\tau}\right]$ is again a $1-\alpha$ CI for θ, but with an understanding that all impossible values should be excluded.

It is now possible to combine the results from Cases 1 and 2, as follows.

If $\tau=0$ then the constrained $1-\alpha$ CI for θ is

$$\begin{cases} [a,b], & 0 \le x \le (1-\alpha)a \\ \left[\frac{x}{1-\alpha},b\right], & (1-\alpha)a < x \le (1-\alpha)b \\ \varnothing, & (1-\alpha)b < x \le b \end{cases}$$

and if $\tau > 0$ then the CI is given by $[l,u]$, where:

$$l=\begin{cases} NA, & 0 \le x < \alpha\tau a \\ a, & \alpha\tau a \le x \le [1-\alpha(1-\tau)]a \\ \frac{x}{1-\alpha(1-\tau)}, & [1-\alpha(1-\tau)]a < x \le [1-\alpha(1-\tau)]b \\ NA, & [1-\alpha(1-\tau)]b < x \le b \end{cases}$$

$$u=\begin{cases} NA, & 0 \le x < \alpha\tau a \\ \frac{x}{\alpha\tau}, & \alpha\tau a \le x < \alpha\tau b \\ b, & \alpha\tau b \le x \le [1-\alpha(1-\tau)]b \\ NA, & [1-\alpha(1-\tau)]b < x \le b \end{cases}$$

Here, *NA* stands for not available (or not defined), with an understanding that $[NA, NA]=\varnothing$.

Observe that for $\tau=0$, the CI is empty if $(1-\alpha)b < x \le b$, and the probability of this is

$$P(x > (1-\alpha)b)=\begin{cases} \frac{\theta-(1-\alpha)b}{\theta}, & (1-\alpha)b < \theta \le b \\ 0, & a \le \theta \le (1-\alpha)b \end{cases}$$

Also, for $\tau > 0$, the CI is empty if $x < \alpha\tau a$ or $x > [1-\alpha(1-\tau)]b$, and the probability of this is

$$P(x < \alpha\tau a) + P(x > [1-\alpha(1-\tau)]b) = \begin{cases} \dfrac{\alpha\tau a}{\theta}, & a \le \theta \le [1-\alpha(1-\tau)]b \\ \dfrac{\alpha\tau a}{\theta} + \dfrac{\theta - [1-\alpha(1-\tau)]b}{\theta}, & [1-\alpha(1-\tau)]b \le \theta \le b \end{cases}$$

We see there exists no single value of τ for which the CI can be guaranteed to not be empty.

As in Case 1, we could redefine the CI as the single point a if $x < \alpha\tau a$ and as the single point b if $x > [1-\alpha(1-\tau)]b$. But then (as in Case 1) there would be a nonzero probability of the CI being a single point, and a frequentist coverage probability of 100% at $\theta = a$ and $\theta = b$.

An important special case is where $\tau = 1$. This leads to the optimal unconstrained $1-\alpha$ CI $\left[x, \dfrac{x}{\alpha}\right]$ but truncated inside the interval $[a,b]$. This interval is empty if $x < \alpha a$, and the probability of this happening is $\alpha a / \theta$ (a number between $\alpha a / b$ and $\alpha a / a = \alpha$, inclusive).

The problems here will be addressed in the next section via the tail functions approach. However, before that we will present an example which illustrates the above concepts.

Example 2 Some constrained confidence intervals

Figure 2 below shows three sets of 80% CIs for θ when that parameter is constrained between $a = 1$ and $b = 2$. These three sets are defined by $\tau = 0$, 1/2 and 1, respectively.

For the data $x = 1.5$, the three CIs work out as:

$$\left[\frac{1.5}{1-0.2(1-0)}, 2\right] = [1.875, 2] \quad \text{if } \tau = 0$$

$$\left[\frac{1.5}{1-0.2(1-0.5)}, 2\right] = [1.6667, 2] \quad \text{if } \tau = 0.5$$

$$\left[\frac{1.5}{1-0.2(1-1)}, 2\right] = [1.5, 2] \quad \text{if } \tau = 1$$

The CI implied by x = 1.5 and τ = 1 is shown in the figure as a light blue horizontal line. This line is the same as the light blue line in Figure 1 but with the part greater than 2 removed.

Also, for the data x = 1.95, the three CIs work out as:

$\varnothing$ (the empty set) if $\tau = 0$

$\varnothing$ if $\tau = 0.5$

$\left[\frac{1.95}{1-0.2(1-0.5)}, 2\right] = [1.95, 2]$ if $\tau = 1$

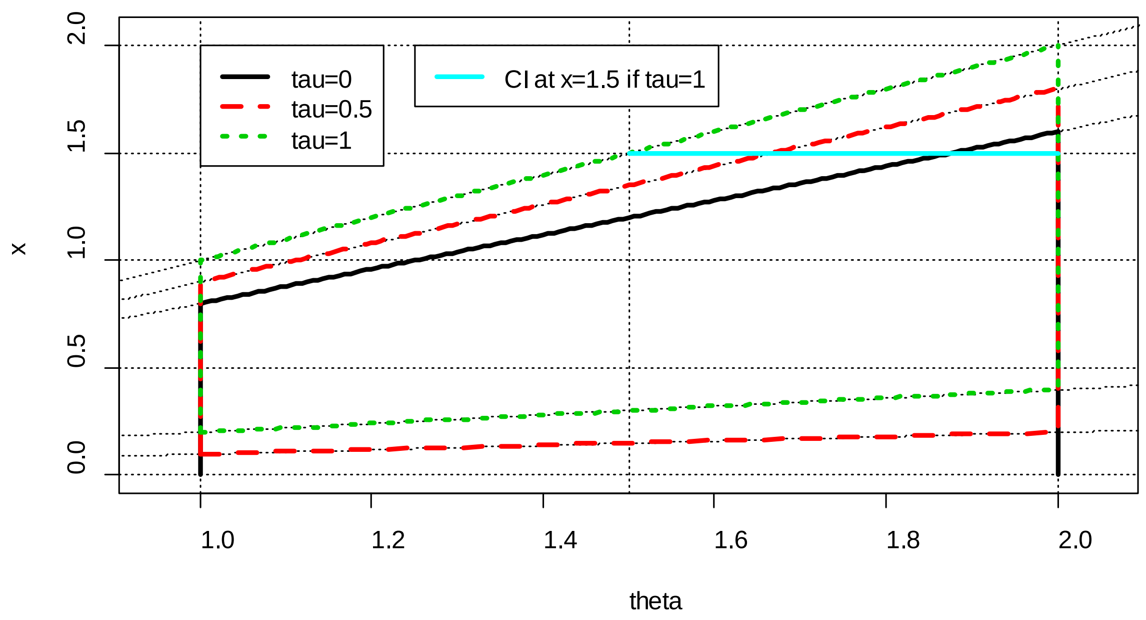
For the data x = 0.1, the CIs work out as:

[1, 2] (the entire parameter space) if $\tau = 0$

[1,1] (i.e., the singleton set, {1}) if $\tau = 0.5$

$\varnothing$ if $\tau = 1$

Figure 2 Three constrained 80% confidence intervals



3. Confidence intervals for the uniform parameter via tail functions

The classical theory in Section 2 can be generalized as follows. Recall Equation (2), namely

$$1-\alpha = P\left(\alpha\tau \le Y \le 1-\alpha+\alpha\tau\right) = P\left(\alpha\tau \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau\right)$$

Now observe that this statement is true not only for any constant $\tau \in [0,1]$ but any function $\tau = \tau(\theta)$ (i.e. function of the target parameter itself), subject to the range of that function being a subset of the unit interval $[0,1]$. In our case, $\tau : [0,\infty) \to D \subseteq [0,1]$.

The key point here is that for any such function $\tau(\theta)$ and any $\theta > 0$, it is true that

$$1-\alpha = P\left(\alpha\tau(\theta) \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau(\theta)\right) \quad (3)$$

Note: This is true because $\frac{X}{\theta} \sim U(0,1)$ and $\left(1-\alpha+\alpha\tau(\theta)\right) - \alpha\tau(\theta) = 1-\alpha$, for any $\tau(\theta)$.

In this context, we call $\tau(\theta)$ the tail function (TF), because it is a function which determines the tail probabilities (originally $\alpha/2$ and $1-\alpha/2$, and subsequently generalized to the constants $\alpha\tau$ and $1-\alpha+\alpha\tau$). For simplicity, we will assume that $\tau(\theta)$ is nondecreasing.

Due to this regularity condition, it is always possible to invert Equation (3) and equate it to

$$P\left(L(X) \le \theta \le U(X)\right)$$

for some well-defined functions *L* and *U*.

Once these functions have been determined, a $1-\alpha$ CI for θ is $[l,u] = [L(x), U(x)]$.

Calculation of the new confidence bounds

In general, and except in some special cases, this CI is not straightforward to calculate. However, the bounds of the interval, *l* and *u*, are always the solutions in θ of the equations

$$\frac{x}{\theta} = 1-\alpha+\alpha\tau(\theta) \quad \text{and} \quad \alpha\tau(\theta) = \frac{x}{\theta} \quad \text{respectively}$$

These can typically be solved via an iterative search technique, such as trial and error or the Newton-Raphson algorithm (NRA).

For example, the upper bound *u* is the solution in θ of the equation $m_U(\theta) = 0$, where

$$m_U(\theta) = \alpha\tau(\theta) - x\theta^{-1}$$

This can be obtained via the NRA by choosing a suitable initial estimate θ_0 of *u* and iterating

$$\theta_j = \theta_{j-1} - \frac{m_U(\theta_{j-1})}{m_U'(\theta_{j-1})} \qquad \text{for } j = 1,2,3,\ldots, \text{ until convergence}$$

An easy way to graph the confidence bounds

Although the above NRA for deriving the confidence bounds $L(x)$ and $U(x)$ may be somewhat involved, there is a technique for easily graphing these bounds in the $(\theta, x)-$plane. This technique provides another way (in addition to substitution, as described above) to check that $L(x)$ and $U(x)$ have been calculated correctly.

The method involves extending Equation (3) and writing

$$1-\alpha = P\left(\alpha\tau(\theta) \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau(\theta)\right)$$

$$= P\left(L(X) \le \theta \le U(X)\right)$$

$$= P\left(A(\theta) \le X \le B(\theta)\right)$$

where *A* and *B* are some functions.

But clearly:

$$A(\theta) = U^{-1}(\theta) = \theta\alpha\tau(\theta))$$

$$B(\theta) = L^{-1}(\theta) = \theta[1-\alpha+\alpha\tau(\theta)]$$

So $\theta = L(x)$ and $\theta = U(x)$ can also be graphed using the equations $x = B(\theta)$ and $x = A(\theta)$.

The linear tail function

There exists a TF which has two attractive properties, in that it can be used to solve the problem of a possibly empty CI in Section 2 and it leads to a CI which can be expressed easily in closed form. This is the linear TF with parameters *c* and *d*, defined by

$$\tau(\theta)=\begin{cases} 0, & 0\le\theta<c \\ \dfrac{\theta-c}{d-c}, & c\le\theta<d \\ 1, & d\le\theta<\infty \end{cases} \qquad (4)$$

where $0\le c\le d<\infty$.

For this TF, assuming $c < d$, we see that:

$$A(\theta)=\theta\alpha\left(\frac{\theta-c}{d-c}\right)$$

$$B(\theta)=\theta\left(1-\alpha+\alpha\left(\frac{\theta-c}{d-c}\right)\right)$$

These two functions define quadratic equations that are easy to solve.

Explicitly, $x=A(\theta)$ has inverse $\theta=G(x)$, where

$$G(x)=\frac{\alpha c+\sqrt{\alpha^2c^2+4\alpha x(d-c)}}{2\alpha}$$

and $x=B(\theta)$ has inverse $\theta=H(x)$, where

$$H(x)=\frac{\alpha c-(1-\alpha)(d-c)+\sqrt{[\alpha c-(1-\alpha)(d-c)]^2+4\alpha x(d-c)}}{2\alpha}$$

Note: The only relevant quadratic roots here are the positive ones.

By considering the various ranges of the data *x*, we find that the unconstrained $1-\alpha$ CI for θ implied by the linear tail function (Equation 4), is given by $[l,u]=[L(x),U(x)]$, where:

$$l=\begin{cases}\dfrac{x}{1-\alpha}, & 0\le x\le(1-\alpha)c\\ H(x), & (1-\alpha)c<x<d\\ x, & d\le x<\infty\end{cases}$$

$$u=\begin{cases}G(x), & 0\le x<\alpha d\\ x/\alpha, & \alpha d\le x<\infty\end{cases}$$

This is the CI for θ when that parameter is unconstrained over the positive real line. If θ is constrained over an interval $[a,b]$, then the formulae above still apply but truncated, i.e. with all values of *l* less than *a* replaced by *a*, and with all values of *u* greater than *b* replaced by *b*.

A special case is where *c* = *a* and *d* = *b*. Then, the associated CI is $[l,u]$, where:

$$l=\begin{cases}a, & 0\le x\le(1-\alpha)a\\ H(x), & (1-\alpha)a<x\le b\end{cases}$$

$$u=\begin{cases}G(x), & 0\le x<\alpha b\\ b, & \alpha b\le x<\infty\end{cases}$$

and where *G* and *H* are as before but with *c* and *d* changed to *a* and *b* throughout, respectively.

Some attractive features of the CI in this special case are that:

- the CI cannot be empty
- the CI has zero probability of consisting of a single point
- the CI converges to the single point *a* as the data *x* converges down towards 0
- the CI converges to the single point *b* as the data *x* converges up towards *b*
- the CI has frequentist coverage probability of exactly $1-\alpha$ for all possible values of $\theta\in[a,b]$ (including *a* and *b*)

This CI is illustrated in the next example.

Example 3 A constrained confidence interval implied by the linear tail function

Figure 3 shows all of the 80% CIs for θ when that parameter is constrained between a = 1 and b = 2 and the tail function is taken to be linear with parameters c = a = 1 and d = b = 2.

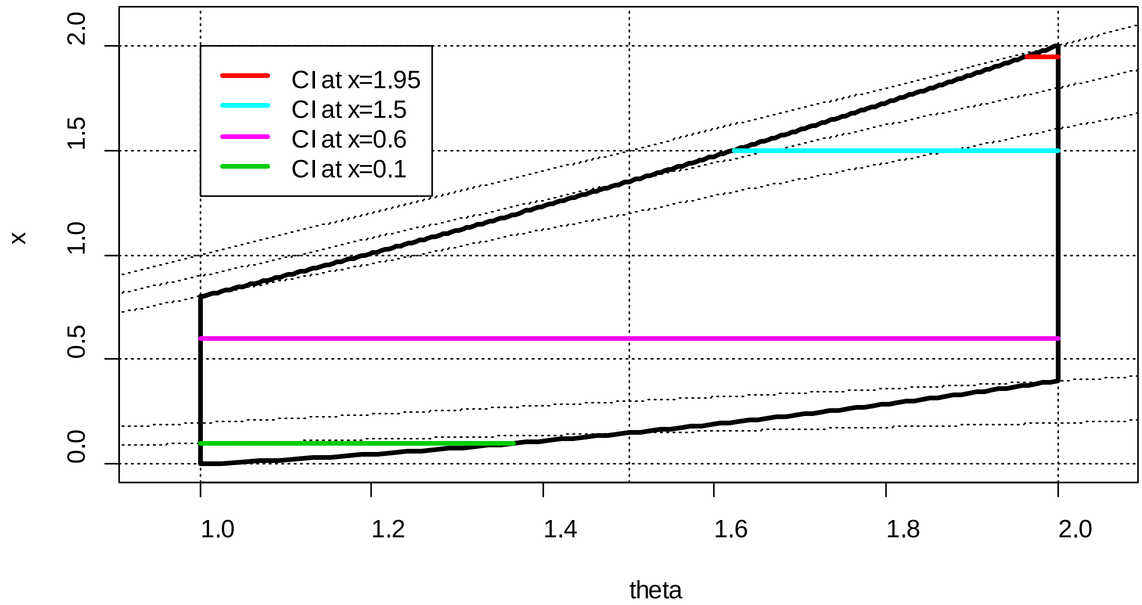
For x = 0.1, 0.6, 1.5 and 1.95, this CI equals [1, 1.3660], [1,2], [1.6225, 2] and [1.9641, 2], respectively, as also depicted in Figure 3.

It will be noted that none of these CIs are empty, whereas this was not the case in Example 2.

For instance, the CI at x = 0.1 was in that example computed as [1, 2] for τ = 0, as [1, 1] for τ = 0.5 and as ∅ (the empty set) for τ = 1.

The new CIs for θ in Figure 3 may be described as a 'continuously weighted average' or 'twisted blend' of the two one-sided CIs depicted in Figure 2, that is, those defined by the constant tail functions $\tau(\theta) = 0$ and $\tau(\theta) = 1$, respectively.

Figure 3 Constrained 80% CIs implied by the linear TF with c = a = 1 and d = b = 2



Other choices for the linear tail function

We have shown that under the interval constraint $a \le \theta \le b$, a suitable CI for θ may be obtained by applying the linear tail function

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ \dfrac{\theta - c}{d - c}, & c \le \theta < d \\ 1, & d \le \theta < \infty \end{cases}$$

with $c = a$ and $d = b$. As we have seen, this choice leads to a CI with attractive properties.

However, these properties can also be obtained with other choices for c and d, so long as $a \le c \le d \le b$. Some special cases (besides $c = a$ and $d = b$ above) are as follows.

- Specifying $c > a$ leads to a CI at $x = 0$ which is an interval rather than a point (as when $c = a$).
- Specifying d as slightly larger than a yields a CI which is similar to the truncated classical CI defined by $\tau = 1$, except for a small perturbation near $\theta = a$ which ensures that the interval cannot be empty and has zero chance of being a single point.
- Specifying $d = a$ implies the truncated classical CI defined by $\tau = 1$, as already discussed in Section 2.
- Specifying $a < c = d < b$ implies the piecewise-constant TF

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ 1, & c \le \theta < \infty \end{cases}$$

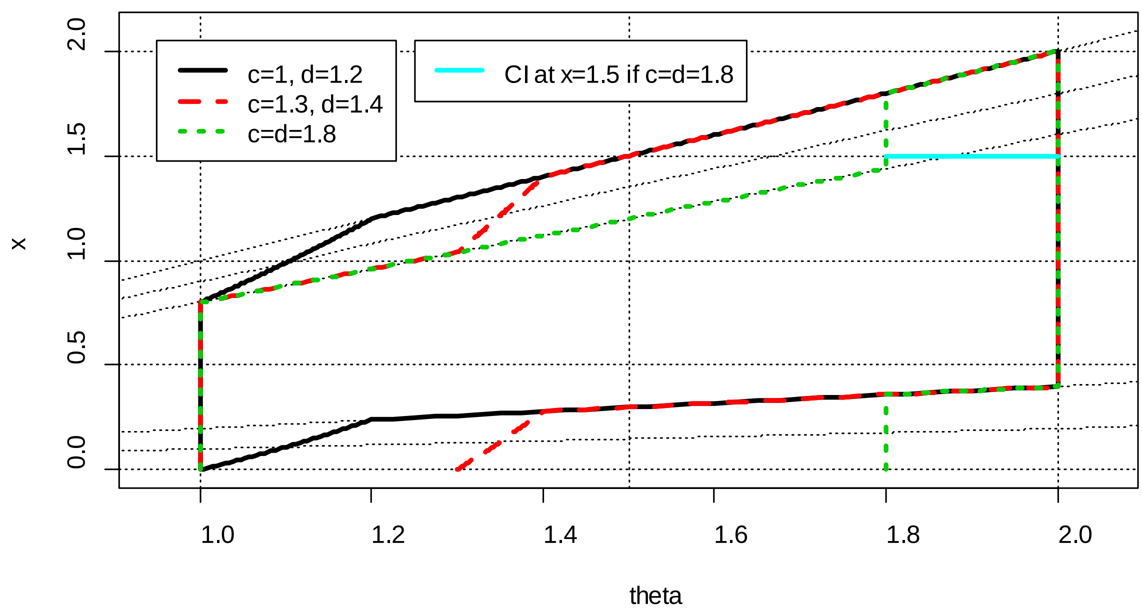
which leads to a CI that has a particularly simple form and can be calculated without the need for solving quadratic equations. Explictly in that case, the CI is [l,u], where:

$$l = \begin{cases} a, & 0 \le x \le (1-\alpha)a \\ \dfrac{x}{1-\alpha}, & (1-\alpha)a < x < (1-\alpha)c \\ c, & (1-\alpha)c \le x \le c \\ x, & c < x \le d \end{cases} \quad \text{and} \quad u = \begin{cases} c, & 0 \le x \le \alpha c \\ \dfrac{x}{\alpha}, & \alpha c < x < \alpha b \\ b, & \alpha b \le x \le d \end{cases}$$

Example 4 Some other CIs implied by the linear TF

Figure 4 shows three sets of 80% CIs for θ when that parameter is constrained between a = 1 and b = 2 and when the tail function is linear with parameters (c, d) = (1, 1.2), (1.3, 1.4) and (1.8, 1.8), respectively. For the last of these three specifications, the CI is simply [1.8, 2], as shown by a light blue horizontal line. We see that none of the CIs in the figure are empty and none have a nonzero probability of containing only a single point.

Figure 4 Three other sets of constrained 80% CIs implied by the linear TF



4. Engineering a confidence interval via tail functions

We have seen how TFs generalize the class of classical CIs for a uniform parameter, and how that class may be thought of as a special case corresponding to the set of constant TFs. We have also shown how linear TFs can be used to solve the problem of a CI being empty after truncation due to the parameter being constrained, from below or from above and below.

Examples 3 and 4 illustrated this solution and revealed that many TFs lead to a constrained CI that cannot be empty and moreover has zero probability of consisting of a single value. This now raises the question of which amongst these many suitable CIs is the most preferable.

The answer to this question partly depends on what criterion, or criteria, the quality of a CI is to be assessed. For definiteness, suppose that any $1-\alpha$ CI for θ under consideration shall:

(i) have zero probability of being empty or consisting of a single point

(ii) converge to the single point *a* as the data *x* converges down towards 0

(iii) converge to the single point *b* as the data *x* converges up towards *b*

(iv) be assessed as better than another $1-\alpha$ CI for θ if it has a smaller size

With these criteria in mind, consider the class of $1-\alpha$ CIs for θ defined by the linear TF

$$\tau(\theta)=\begin{cases} 0, & 0\le\theta<c \\ \dfrac{\theta-c}{d-c}, & c\le\theta<d \\ 1, & d\le\theta<\infty \end{cases}$$

with parameters $c = a$ and $d \in [a,b]$.

Note that $c = a$ ensures Criterion (ii) above. Also note that the possibility $d = a$ violates Criterion (i) but is included because it defines an important special case, namely the truncated classical CI defined by $\tau=1$, as discussed earlier in Section 2.

Then define the conditional expected size (CES) of the CI as

$$CES(\theta)=E\{S(X)\,|\,\theta\}=\int S(x)f(x\,|\,\theta)dx$$

where

$$f(x\,|\,\theta)=\frac{1}{\theta},\ 0\le x\le\theta$$

is the model density of the data *X* for any given value of θ.

Let us further suppose that we have information which can be expressed in terms of a prior density $f(\theta)$. Then define the prior expected size (PES) of the CI as

$$PES = ES(X) = EE\{S(X) \mid \theta\} = \int CES(\theta) f(\theta) d\theta = \int S(x) f(x) dx$$

where $f(x) = \int f(x \mid \theta) f(\theta) d\theta$ is the unconditional (or prior) density of the data *X*.

Note that $S(x)$, $CES(\theta)$ and *PES* are all implictly functions of *d*, as well as of α, *a* and *b*.

It is now of interest to analyze these measures of size, with a view to obtaining guidance regarding a good choice of *d*. The analysis will of course depend on the specification of $f(\theta)$.

Example 5 Analysis of size

Suppose that $\alpha = 0.2$, *a* = 1, *b* = 2 and we apply the linear tail function (Equation 4) with *c* = *a* and $a \le d \le b$; that is, we specify

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < 1 \\ \dfrac{\theta - 1}{d - 1}, & 1 \le \theta < d \\ 1, & d \le \theta < \infty \end{cases}$$

with $1 \le d \le 2$.

Figure 5 shows the size of the associated CI as a function of *x* for selected values of *d*.

Also, Figure 6 shows the conditional expected size as a function of θ for selected values of *d*, and Figure 7 shows the prior expected size as a function of *d*, assuming the flat prior $\theta \sim U(1,2)$. (Note: This prior may be considered uninformative, or noninformative, for θ.)

We see in Figure 5 that some trade-offs are involved when deciding on the value of *d* based on the size of the corresponding CI as a function of *x*. Specifically: a small value of *d* is associated with narrow CIs when *x* is small and wide CIs when *x* is large (relatively); whereas a large value of *d* is associated with wide CIs when *x* is small and narrow CIs when *x* is large.

However, Figure 6 reveals that for any fixed value of θ the CES of the CI decreases uniformly as *d* decreases. This then implies that when the CES is averaged over any prior distribution, the result will be a PES which also decreases uniformly with *d*. This effect is apparent in Figure 7.

It follows that if we wish to choose the CI with minimum PES, and nothing else matters, the best choice of *d* is *a* = 1. However, as previously noted, the implied CI has a nonzero probability of being empty. Explictly, that probability is $a\alpha / \theta$, and in our case with $\alpha = 0.2$ this could be anything from 10% (if θ happens to equal 2) to 20% (if θ happens to equal 1).

Numerically, we find that the PESs of the 80% CIs defined by *d* = 1, 1.2 and 2 are 0.6955, 0.6962 and 0.7285, respectively. The last two of these CIs are similarly attractive on account of both having zero probability of being empty. However, the second CI may be considered as the better of the two because it is only 0.1% wider than the first CI (in prior expectation), compared to the third CI, which is 4.7% wider.

Figure 5 Size of a confidence interval

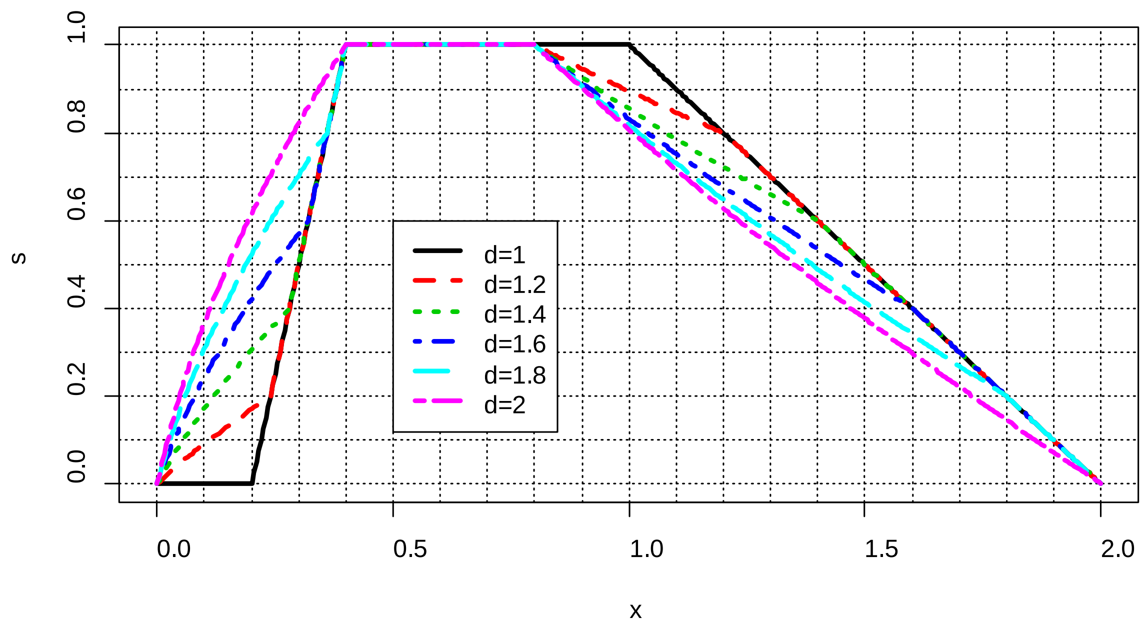


Figure 6 Conditional expected size of a confidence interval

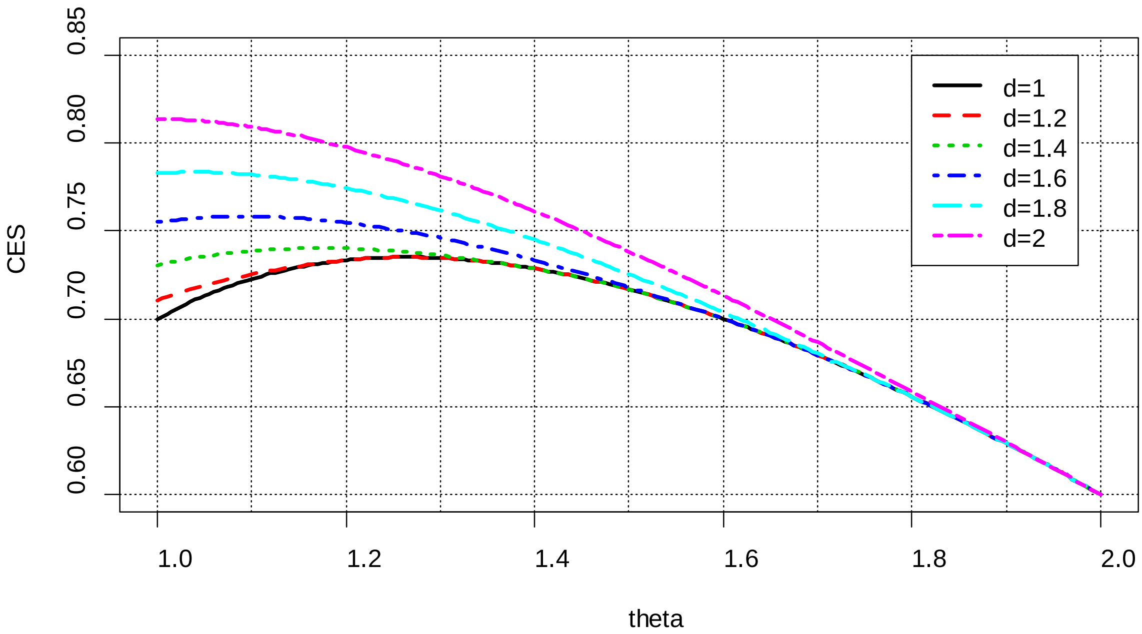
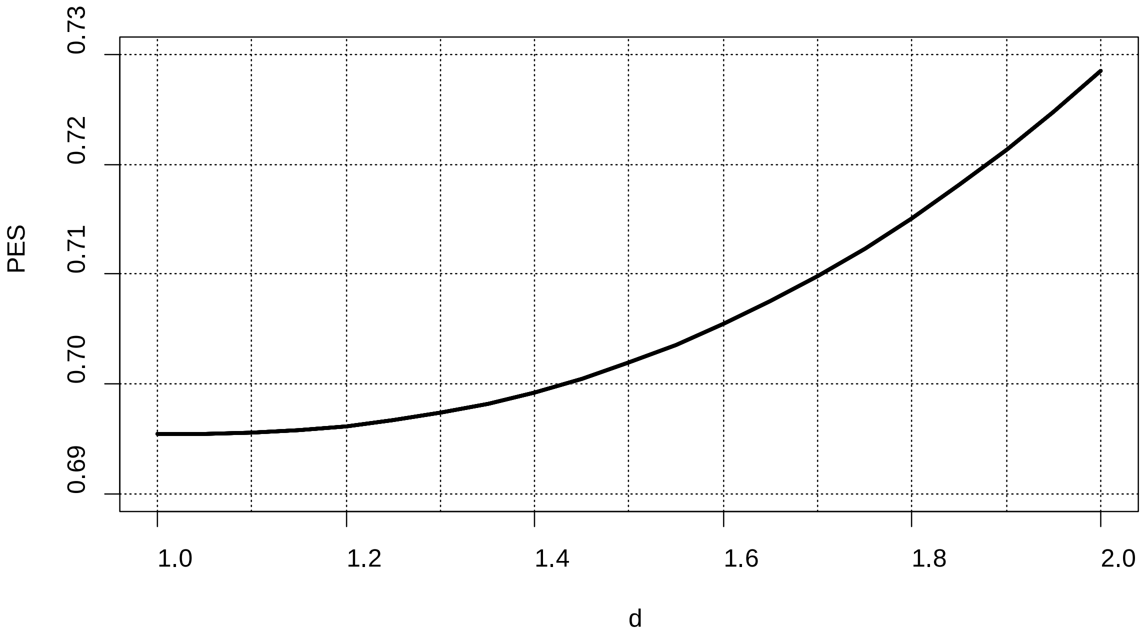


Figure 7 Prior expected size of a confidence interval



5. Summary and discussion

In this paper we have shown how the method of tail functions can be used to generalize the classical or ordinary $1-\alpha$ confidence interval for the upper bound parameter θ of a uniform distribution with lower bound zero, based on a single observed value *x* from that distribution. When this parameter is constrained in an interval $[a,b]$, the truncated classical confidence interval has the problem of possibly being empty.

We illustrated how this problem can be addressed by choosing a confidence interval in the class defined by the linear tail function

$$\tau(\theta)=\begin{cases} 0, & 0\le\theta<c \\ \dfrac{\theta-c}{d-c}, & c\le\theta\le d \\ 1, & d<\theta<\infty \end{cases}$$

where $0\le c\le d<\infty$. Under the constraint $a\le\theta\le b$, a suitable subclass is given by *c* = *a* and $a\le d\le b$, and one attractive confidence interval in particular is defined by *c* = *a* and *d* = *b*.

We also showed how prior information can be used within the tail functions framework so as to engineer an attractive confidence interval in some sense. Example 5, with $a=c=1$, $b=2$ and $\theta\sim U(1,2)$, suggested that if minimizing prior expected size is the sole criterion then the truncated classical confidence interval, defined by *d* = *a*, is optimal. A compromise solution with *d* slightly larger than *a* (e.g. *d* = 1.2) was suggested. This choice yields an interval with zero probability of being empty and with a prior expected size that is very nearly minimal.

That the truncated classical confidence interval is in fact optimal amongst all confidence intervals for θ can be shown using Equation 17 in Brown, Casella and Hwang (1995). This interval is also the one implied by the unified approach of Feldman and Cousins (1998) (see also Mandelkern, 2002) under the constraint $a\le\theta\le b$ but with no prior distribution for θ.

The question of why the Bayesian approach should not be used when prior information is available can be answered by saying that it leads to a confidence interval which is not proper, meaning one with a frequentist coverage probability that is not at least the desired level $1-\alpha$ for all possible values of θ. This general fact can be illustrated by finding the 80% highest posterior density region for θ in Example 5 and the frequentist coverage probability of that region. It turns out that this probability tapers down to zero as θ converges up to *b* = 2.

It also turns out that the prior expected size of the highest posterior density region is less than than that of the truncated classical confidence interval. Thus, the Bayesian approach leads to an even shorter interval in prior expectation but at the expense of that interval being improper.

The work in this paper could easily be extended to confidence estimation of θ based on a random sample of size *n*, in which case a sufficient statistic would be the sample maximum. We focussed on the special case $n = 1$ here because it led to simple formulae which aided in the discussion of central issues. Other forms of prior distribution could also be considered.

Also, other criteria could be used to define an optimal confidence interval. For example, it might be desirable to minimize the prior probability of an interval being wider than some specified amount. This could be illustrated in the context of Example 5 by searching for the value of *d* which minimizes the prior probability of the interval being wider than 0.8 (say).

Furthermore, the class of linear tail functions could be generalized so as to improve some features of the associated confidence intervals. The number of parameters of the tail function could be increased from two to three or four, thereby providing greater flexibility which could, for example, lead to lower and upper bounds that are smoother than those in Figure 4. This would of course come at the cost of a greater computational burden and might require an iterative search procedure, such as the Newton-Raphson algorithm described in Section 3.

References

Brown, L.D., Casella, G., and Hwang, J.T.G. (1995). Optimal confidence sets, bioequivalence, and the Limaçon of Pascal. *Journal of the American Statistical Association*, **90**, 880-889.

Feldman, G. J., and Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals. *Physical Reviews D*, **57**, 3873-3889.

Lee, P. (1997). *Bayesian Statistics: An Introduction*. New York: Oxford University Press

Mandelkern, M. (2002). Setting confidence intervals for bounded parameters. *Statistical Science*, **17**, 149-172 (with Discussion).

Puza, B., and O'Neill, T. (2005). *Interval Estimation via Tail Functions*. Working Paper STAT0005WP, College of Business and Economics, Australian National University, Canberra. Website: cbe.anu.edu.au/research/papers/pdf/STAT0005WP.pdf

Puza, B.D., and O'Neill, T.J. (2006a). Generalised Clopper-Pearson confidence intervals for the binomial proportion. *Journal of Statistical Computation and Simulation*, **76**, 489-508.

Puza, B., and O'Neill, T. (2006b). Interval estimation via tail functions. *The Canadian Journal of Statistics*, **34**, 299-310.

Puza, B., and O'Neill, T. (2008). Optimal constrained confidence estimation via tail functions. *The Mathematical Scientist*, **33**, 134-140.

Puza, B., and O'Neill, T. (2009). Constrained confidence estimation of the binomial *p* via tail functions. *The Mathematical Scientist*, **34**, 43-48.

Puza, B., Roberts, S., and Yang, M. (2011). Constrained confidence intervals in time series studies of mortality and air pollution. *Environment International*, **37**, 204-209.

Puza, B. (2011). The tail function approach to confidence estimation in survey sampling. *Pakistan Journal of Statistics*, **27**, 543-554.

Puza, B., and Yang, Mo. (2011). Optimal constrained confidence estimation of the Poisson mean via tail functions. *The Mathematical Scientist*, **36**, 95-104.

Puza, B., and Yang, M. (2013). Improved confidence estimation for the exponential mean via tail functions. *Communications in Statistics - Theory and Methods*. DOI:10.1080/03610926.2013.833235. Website: www.tandfonline.com/doi/abs/10.1080/03610926.2013.833235#.UxkQMfna701

Stuart, A., Ord, K., and Arnold, S. (1999). *Kendall's Advanced Theory of Statistics, Sixth Edition, Volume 2A, Classical Inference and the Linear Model*. London: Arnold.

Borek Puza
Research Sch. of Finance, Actuarial Studies & Applied Statistics,
Australian National University

The Tail Functions Approach to Confidence Estimation

Synopsis:

This paper reviews the method of tail functions (TFs) for constructing confidence intervals (CIs). The idea is to generalize the constant cutoffs (e.g. 0.025 and 0.975) in the definition of a CI to functions of the target parameter. The ordinary 2- and 1-sided CIs are defined by the constant TFs 1/2, 0 and 1. Under prior information, the TFs approach is an alternative to the Bayesian and can lead to improved CIs, whilst retaining frequentist coverage probabilities.

The tail functions approach to confidence estimation of a uniform parameter

Borek Puza

Australian National University, Canberra

Hawaii Education & STEM Conference, 16-18 June 2014

Classical confidence intervals for a uniform parameter

We observe *x*, a value of the random variable $X \sim U(0,\theta)$

We wish to construct a $1-\alpha$ confidence interval (CI) for θ

Form the pivotal quantity $Y = \dfrac{X}{\theta} \sim U(0,1)$

For any θ: $$1-\alpha = P\left(\frac{\alpha}{2} \le \frac{X}{\theta} \le 1-\frac{\alpha}{2}\right) = P\big(L(X) \le \theta \le U(X)\big)$$

for some functions *L* (the lower bound) and *U* (the upper bound)

Working to get θ in the middle:

$$1-\alpha = P\left(\frac{\alpha}{2} \le \frac{X}{\theta} \le 1-\frac{\alpha}{2}\right)$$

$$= P\left(\frac{\alpha}{2}\theta \le X \le \theta\left(1-\frac{\alpha}{2}\right)\right) \qquad = P\left(\frac{\alpha}{2}\theta \le X,\ X \le \theta\left(\frac{2-\alpha}{2}\right)\right)$$

$$= P\left(\theta \le X\frac{2}{\alpha},\ X\left(\frac{2}{2-\alpha}\right) \le \theta\right) \qquad = P\left(X\left(\frac{2}{2-\alpha}\right) \le \theta \le X\frac{2}{\alpha}\right)$$

$$= P\left(L(X) \le \theta \le U(X)\right)$$

So $L(X) = \dfrac{2X}{2-\alpha}$ and $U(X) = \dfrac{2X}{\alpha}$

So if the observed value of X is x, a $1-\alpha$ CI for θ is

$$[l,u] = [L(x), U(x)] = \left[\frac{x}{1-\alpha/2}, \frac{2x}{\alpha}\right]$$

The size (or length or width or volume) of this CI is

$$s = S(x) = U(x) - L(x) = x\left(\frac{2}{\alpha} - \frac{1}{1-\alpha/2}\right) = \frac{4(1-\alpha)}{\alpha(2-\alpha)}x$$

Other classical confidence intervals

Choose a constant $\tau \in [0,1]$ and note that for any θ :

$$1-\alpha = P\left(\alpha\tau \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau\right)$$ (we again try to get θ in the middle)

$$= P\left(\alpha\tau\theta \le X \le \theta\left(1-\alpha+\alpha\tau\right)\right) \quad = P\left(\alpha\tau\theta \le X, X \le \theta\left(1-\alpha+\alpha\tau\right)\right)$$

$$= P\left(\theta \le \frac{X}{\alpha\tau}, \frac{X}{1-\alpha+\alpha\tau} \le \theta\right) \quad = P\left(\frac{X}{1-\alpha+\alpha\tau} \le \theta \le \frac{X}{\alpha\tau}\right)$$

$$= P\left(L(X) \le \theta \le U(X)\right)$$

Thus $L(X) = \dfrac{X}{1-\alpha(1-\tau)}$ and $U(X) = \dfrac{X}{\alpha\tau}$

So another $1-\alpha$ CI (or rather class of CIs) for θ is

$$[l,u] = [L(x), U(x)] = \left[\frac{x}{1-\alpha(1-\tau)}, \frac{x}{\alpha\tau}\right]$$

The previous CI $\left[\dfrac{x}{1-\alpha/2}, \dfrac{2x}{\alpha}\right]$ is defined by τ = 1/2

This special case is a symmetric and two-sided CI for θ

Two other special cases are the one-sided CIs:

$$\left[\frac{x}{1-\alpha},\infty\right) \qquad \text{defined by } \tau = 0 \quad \text{(the upper-range CI)}$$

$$\left[x,\frac{x}{\alpha}\right] \qquad \text{defined by } \tau = 1 \quad \text{(the lower-range CI)}$$

The general CI indexed by τ has size

$$s = S(x) = \frac{x}{\alpha\tau} - \frac{x}{1-\alpha(1-\tau)} = x\left(\frac{1-\alpha}{\alpha}\right)\tau^{-1}[1-\alpha(1-\tau)]^{-1}$$

By the chain rule for differentiation, the slope of the size is

$$\frac{ds}{d\tau} = x\left(\frac{1-\alpha}{\alpha}\right)\left\{\tau^{-1}(-1)[1-\alpha(1-\tau)]^{-2}\alpha + (-\tau^{-2})[1-\alpha(1-\tau)]^{-1}\right\}$$

This slope is negative for all τ (and x and α)

So $\tau = 1$ uniformly minimizes the size of the CI

So, in this sense, the optimal $1-\alpha$ CI for θ is $[l,u] = \left[x, x/\alpha\right]$

(which is one-sided), and the size of this CI is $s = u - l = \left(\frac{1-\alpha}{\alpha}\right)x$

Example 1 Some classical confidence intervals

Figure 1 shows three sets of classical 80% CIs for θ (thus $\alpha = 0.2$)

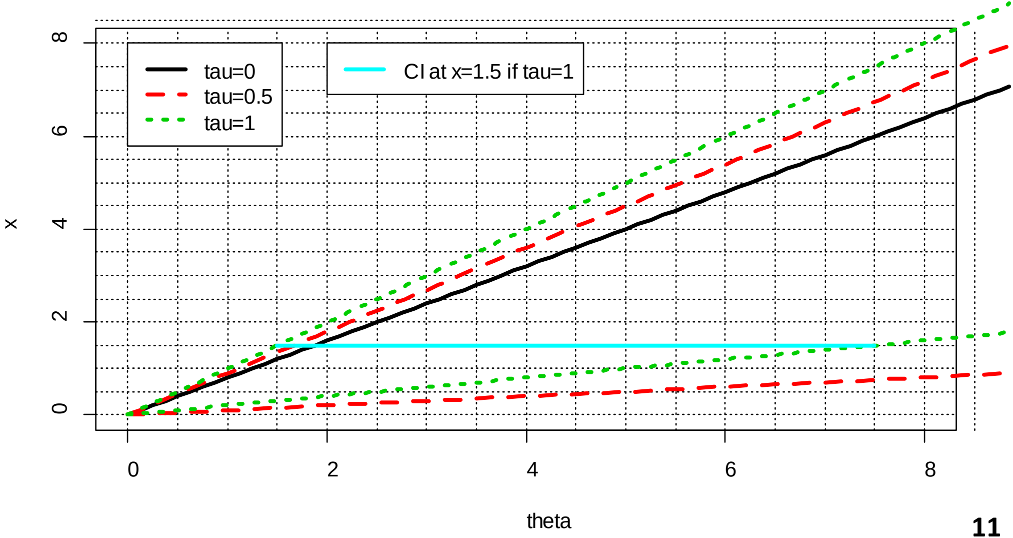
These are defined by τ = 0, 0.5 and 1. For x = 1.5, the CIs are:

$$\left[\frac{x}{1-\alpha}, \infty\right) = \left[\frac{1.5}{1-0.2}, \infty\right) = [1.875,\ \infty) \quad \text{if } \tau = 0$$

$$\left[\frac{x}{1-\alpha(1-\tau)}, \frac{x}{\alpha\tau}\right] = \left[\frac{1.5}{1-0.2(1-0.5)}, \frac{1.5}{0.2\times 0.5}\right] = [1.6667,\ 15] \quad \text{if } \tau = 0.5$$

$$\left[x, \frac{x}{\alpha}\right] = \left[1.5, \frac{1.5}{0.2}\right] = [1.5,\ 7.5] \quad \text{if } \tau = 1 \quad \text{(shown light blue)}$$

Figure 1 Three classical 80% confidence intervals



Classical interval estimation under constraints

Case 1: When the uniform parameter is bounded from below

Suppose $\theta \geq a$, where $a > 0$

Then $[l,u] = \left[\dfrac{x}{1-\alpha(1-\tau)}, \dfrac{x}{\alpha\tau} \right]$ is still a $1-\alpha$ CI for θ

but with all impossible values (those less than *a*) excluded

To revise the formula, we need to consider two possibilities for τ

If $\tau = 0$ the CI is $\begin{cases} [a, \infty), & 0 \le x \le (1-\alpha)a \\ \left[\dfrac{x}{1-\alpha}, \infty\right), & (1-\alpha)a < x < \infty \end{cases}$

If $\tau > 0$ the CI is $\begin{cases} \varnothing, & 0 \le x < \alpha\tau a \\ \left[a, \dfrac{x}{\alpha\tau}\right], & \alpha\tau a \le x \le [1-\alpha(1-\tau)]a \\ \left[\dfrac{x}{1-\alpha(1-\tau)}, \dfrac{x}{\alpha\tau}\right], & [1-\alpha(1-\tau)]a < x < \infty \end{cases}$

Here, $\varnothing$ (phi) denotes the null set, also known as the empty set { }

So the CI could be empty, and the probability of this happening is

$$P(0 \le x < a\alpha\tau) = \frac{a\alpha\tau}{\theta} \qquad \text{(since } X \sim U(0,\theta)\text{)}$$

This probability is zero only if $\tau = 0$, that is if the CI is

$$\begin{cases} [a,\infty), & 0 \le x \le (1-\alpha)a \\ \left[\dfrac{x}{1-\alpha},\infty\right), & (1-\alpha)a < x < \infty \end{cases}$$

But this CI has is infinitely wide for all possible values of *x*

One way to obtain a finite CI which cannot be empty is to set $\tau > 1$ and then redefine the CI as the single point *a* if $0 \le x < \alpha\tau a$

(Note that the CI already was the single point *a* at $x = \alpha\tau a$)

But this ad hoc solution has two problems:

1. The probability of the CI being a single point is then the same as it previously had of being empty, i.e. $a\alpha\tau / \theta$

2. The CI then has a frequentist coverage probability of 100% (not the nominal or intended level of $1-\alpha$) at $\theta = a$

Case 2: When the uniform parameter is bounded from above

Suppose $\theta \le b$ where $b > 0$

Then $[l,u] = \left[\frac{x}{1-\alpha(1-\tau)}, \frac{x}{\alpha\tau} \right]$ is still a $1-\alpha$ CI for θ

but with all impossible values (those more than b) excluded

As for Case 1, we need to consider two possibilities for τ

If $\tau = 0$ the CI is $$\begin{cases} \left[\dfrac{x}{1-\alpha}, b\right] & 0 \le x \le (1-\alpha)b \\ \varnothing, & (1-\alpha)b < x \le b \end{cases}$$

If $\tau > 0$ the CI is $$\begin{cases} \left[\dfrac{x}{1-\alpha(1-\tau)}, \dfrac{x}{\alpha\tau}\right], & 0 \le x \le \alpha\tau b \\ \left[\dfrac{x}{1-\alpha(1-\tau)}, b\right], & \alpha\tau b < x \le [1-\alpha(1-\tau)]b \\ \varnothing, & [1-\alpha(1-\tau)]b < x \le b \end{cases}$$

So, as for Case 1, the CI could be empty, but now with probability

$$P(x > [1-\alpha(1-\tau)]b) = \begin{cases} 0, & a \le \theta \le [1-\alpha(1-\tau)]b \\ \dfrac{\theta - [1-\alpha(1-\tau)]b}{\theta}, & [1-\alpha(1-\tau)]b < \theta \le b \end{cases}$$

This is guaranteed to be zero only if $\tau = 1$, that is if the CI is

$$\begin{cases} \left[x, \dfrac{x}{\alpha}\right], & 0 \le x \le \alpha b \\ [x, b], & \alpha b < x \le b \end{cases}$$

(We see no problems with this CI)

Case 3: When the uniform parameter is bounded from above and below

Suppose $a \le \theta \le b$ where $0 < a < b$

Then we combine results for Cases 1 and 2 to get the following formulae

If $\tau = 0$ the CI is $\begin{cases} \left[a, b\right], & 0 \le x \le (1-\alpha)a \\ \left[\dfrac{x}{1-\alpha}, b\right], & (1-\alpha)a < x \le (1-\alpha)b \\ \varnothing, & (1-\alpha)b < x \le b \end{cases}$

If $\tau > 0$ the CI is $[l, u]$, where:

$$
l = \begin{cases} NA, & 0 \le x < \alpha\tau a \\ a, & \alpha\tau a \le x \le [1-\alpha(1-\tau)]a \\ \dfrac{x}{1-\alpha(1-\tau)}, & [1-\alpha(1-\tau)]a < x \le [1-\alpha(1-\tau)]b \\ NA, & [1-\alpha(1-\tau)]b < x \le b \end{cases}
$$

$$
u = \begin{cases} NA, & 0 \le x < \alpha\tau a \\ \dfrac{x}{\alpha\tau}, & \alpha\tau a \le x < \alpha\tau b \\ b, & \alpha\tau b \le x \le [1-\alpha(1-\tau)]b \\ NA, & [1-\alpha(1-\tau)]b < x \le b \end{cases}
$$

Here, NA stands for not available, meaning that $[NA, NA] = \varnothing$

So if $\tau = 0$, the CI is empty if $(1-\alpha)b < x \le b$, with probability

$$P(x > (1-\alpha)b) = \begin{cases} \dfrac{\theta - (1-\alpha)b}{\theta}, & (1-\alpha)b < \theta \le b \\ 0, & a \le \theta \le (1-\alpha)b \end{cases}$$

And if $\tau > 0$, the CI is empty if $x < \alpha\tau a$ or $x > [1-\alpha(1-\tau)]b$,

with probability

$$P(x < \alpha\tau a) + P(x > [1-\alpha(1-\tau)]b]$$

$$= \begin{cases} \dfrac{\alpha\tau a}{\theta}, & a \le \theta \le [1-\alpha(1-\tau)]b \\ \dfrac{\alpha\tau a}{\theta} + \dfrac{\theta - [1-\alpha(1-\tau)]b}{\theta}, & [1-\alpha(1-\tau)]b \le \theta \le b \end{cases}$$

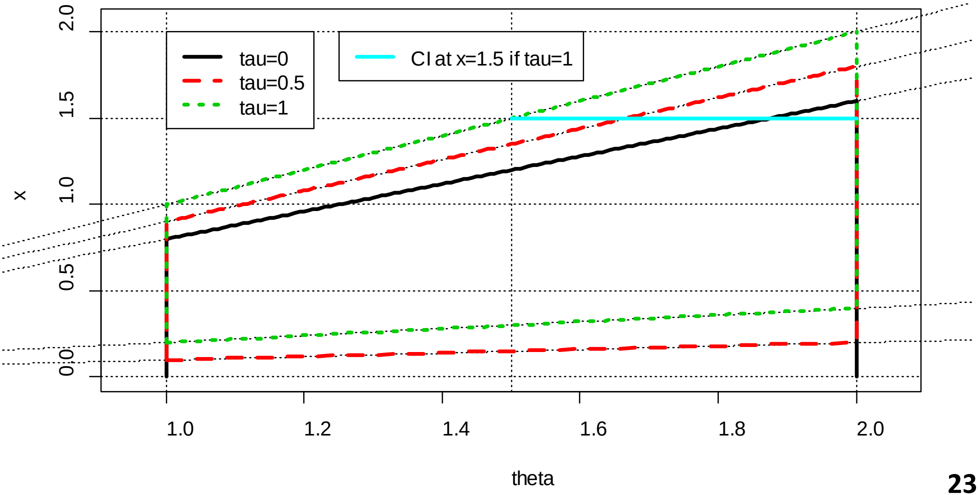
So there is no value of τ for which the CI is guaranteed to not be empty

Example 2 Some constrained confidence intervals

Figure 2 shows 3 sets of 80% CIs for θ under the constraint $1 \le \theta \le 2$

These 3 sets are defined by τ = 0, 1/2 and 1, respectively

Figure 2 Three constrained 80% confidence intervals



For x = 1.5, the 3 CIs work out as:

$$\left[\frac{1.5}{1-0.2(1-0)}, 2\right] \quad = [1.875, 2] \quad \text{if } \tau = 0$$

$$\left[\frac{1.5}{1-0.2(1-0.5)}, 2\right] \quad = [1.6667, 2] \quad \text{if } \tau = 0.5$$

$$\left[\frac{1.5}{1-0.2(1-1)}, 2\right] \quad = [1.5, 2] \quad \text{if } \tau = 1$$

This last CI is shown light blue and is a subset of the same line in Figure 1

For x = 1.95, the 3 CIs are: $\varnothing$ (the empty set) if $\tau = 0$

$\varnothing$ (the empty set) if $\tau = 0.5$

$$\left[\frac{1.95}{1-0.2(1-0.5)}, 2\right] = [1.95, 2] \quad \text{if } \tau = 1$$

For x = 0.1, the 3 CIs are: [1, 2] (the whole parameter space) if $\tau = 0$

[1,1] (the singleton set, {1}) if $\tau = 0.5$

$\varnothing$ (the empty set) if $\tau = 1$

3. CIs for the uniform parameter via tail functions

Recall that $1-\alpha = P\left(\alpha\tau \le \dfrac{X}{\theta} \le 1-\alpha+\alpha\tau\right)$

Now observe that this statement is true not only for any constant $\tau \in [0,1]$

but for any ***function*** $\tau = \tau(\theta)$ (i.e., function of the target parameter itself)

subject to the range of that function being a subset of $[0,1]$

In our case, $\tau : [0,\infty) \to D \subseteq [0,1]$

For any such function $\tau(\theta)$ and any $\theta > 0$ it is true that

$$1-\alpha = P\left(\alpha\tau(\theta) \leq \frac{X}{\theta} \leq 1-\alpha+\alpha\tau(\theta)\right)$$

We call $\tau(\theta)$ the ***tail function*** (TF)

Why?

Because it is a function which determines the tail probabilities

(originally $\alpha/2$ and $1-\alpha/2$, and later $\alpha\tau$ and $1-\alpha+\alpha\tau$)

For simplicity, we will assume that $\tau(\theta)$ is nondecreasing

Due to this regularity condition, it is always possible to write

$$1-\alpha = P\left(\alpha\tau(\theta) \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau(\theta)\right)$$

$$= P\big(L(X) \le \theta \le U(X)\big) \quad \text{for some functions } L \text{ and } U$$

Once these functions have been determined, a $1-\alpha$ CI for θ is

$$[l,u] = [L(x), U(x)]$$

Calculation of the new confidence bounds

The bounds l and u of the CI are the solutions in θ of the equations

$$\frac{x}{\theta} = 1 - \alpha + \alpha\tau(\theta) \quad \text{and} \quad \alpha\tau(\theta) = \frac{x}{\theta} \quad \text{respectively}$$

These can be solved via the Newton-Raphson algorithm (NRA)

E.g., u is the solution of $m_U(\theta) = 0$ where $m_U(\theta) = \alpha\tau(\theta) - x\theta^{-1}$

The NRA says choose an initial estimate θ_0 of u and iterate

$$\theta_j = \theta_{j-1} - \frac{m_U(\theta_{j-1})}{m'_U(\theta_{j-1})} \qquad \text{for } j = 1,2,3,..., \quad \text{until convergence}$$

An easy way to graph the confidence bounds

Observe that $$1-\alpha = P\left(\alpha\tau(\theta) \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau(\theta)\right)$$

$$= P\big(L(X) \le \theta \le U(X)\big) = P\big(A(\theta) \le X \le B(\theta)\big)$$

where $A(\theta) = U^{-1}(\theta) = \theta\alpha\tau(\theta))$ and $B(\theta) = L^{-1}(\theta) = \theta[1-\alpha+\alpha\tau(\theta)]$

So $\theta = L(x)$ and $\theta = U(x)$ can be graphed using $x = B(\theta)$ and $x = A(\theta)$

The linear tail function

The linear TF is $$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ \dfrac{\theta - c}{d - c}, & c \le \theta < d \\ 1, & d \le \theta < \infty \end{cases} \qquad (0 \le c \le d < \infty)$$

This TF has two attractive properties:

1. It can be used to solve the problem of CIs being empty

2. It always leads to a CI with a simple closed-form expression

For the linear TF, assuming $c < d$, we see that:

$$1-\alpha = P\left(\alpha\tau(\theta) \le \frac{X}{\theta} \le 1-\alpha+\alpha\tau(\theta)\right)$$

$$= P\big(L(X) \le \theta \le U(X)\big)$$

$$= P\big(A(\theta) \le X \le B(\theta)\big)$$

where: $$A(\theta) = \theta\alpha\left(\frac{\theta-c}{d-c}\right), \qquad B(\theta) = \theta\left(1-\alpha+\alpha\left(\frac{\theta-c}{d-c}\right)\right)$$

Each of these functions defines a quadratic equation that is easy to solve

Explicitly, $x = A(\theta)$ has inverse $\theta = G(x)$, where

$$G(x) = \frac{\alpha c + \sqrt{\alpha^2 c^2 + 4\alpha x(d - c)}}{2\alpha}$$

and $x = B(\theta)$ has inverse $\theta = H(x)$, where

$$H(x) = \frac{\alpha c - (1 - \alpha)(d - c) + \sqrt{[\alpha c - (1 - \alpha)(d - c)]^2 + 4\alpha x(d - c)}}{2\alpha}$$

We find that the resulting CI is $[l, u] = [L(x), U(x)]$, where:

$$l = \begin{cases} \dfrac{x}{1-\alpha}, & 0 \le x \le (1-\alpha)c \\ H(x), & (1-\alpha)c < x < d \\ x, & d \le x < \infty \end{cases}, \qquad u = \begin{cases} G(x), & 0 \le x < \alpha d \\ x/\alpha, & \alpha d \le x < \infty \end{cases}$$

If θ is constrained over an interval $[a,b]$, we simply truncate the CI (i.e., if $l < a$ then we change l to a, and if $u > b$ then we change u to b)

A special case is $c = a$ and $d = b$, leading to the CI $[l,u]$, where:

$$l = \begin{cases} a, & 0 \le x \le (1-\alpha)a \\ H(x), & (1-\alpha)a < x \le b \end{cases}, \qquad u = \begin{cases} G(x), & 0 \le x < \alpha b \\ b, & \alpha b \le x < \infty \end{cases}$$

This CI:

- cannot be empty
- has zero probability of consisting of a single point
- converges to a as x converges down towards 0
- converges to b as x converges up towards b
- has frequentist coverage probability $1-\alpha$ for all $\theta \in [a,b]$

Example 3 A constrained CI implied by the linear tail function

Figure 3 shows all 80% CIs for θ under the constraint $1 \le \theta \le 2$

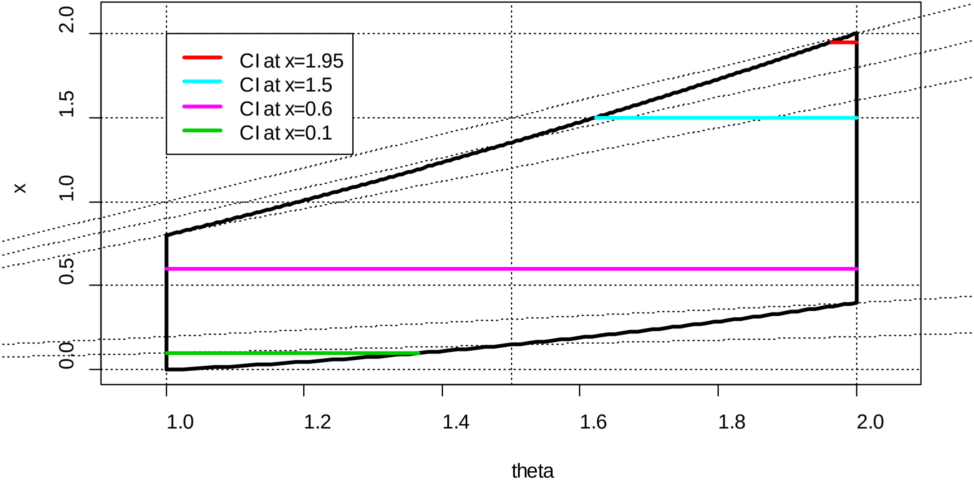
and corresponding to the linear tail function with c = 1 and d = 2

For x = 0.1, 0.6, 1.5 and 1.95, this CI works out (respectively) as

[1, 1.3660], [1,2], [1.6225, 2] and [1.9641, 2]

Note that none of these CIs is empty, unlike in Example 2

Figure 3 80% CIs implied by the linear TF with $c = a = 1$ and $d = b = 2$



Other choices for the linear tail function

We have considered the class of linear TFs defined by

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ \dfrac{\theta - c}{d - c}, & c \le \theta < d \\ 1, & d \le \theta < \infty \end{cases}$$

with $0 \le c \le d < \infty$, and shown that under the constraint $a \le \theta \le b$, a CI with attractive features is the one defined by $c = a$ and $d = b$

But other choices for c and d also lead to attractive features:

- $c > a$ leads to a CI at $x = 0$ which is an interval (not a point)
- $d = a$ implies the classical CI defined by $\tau = 1$ (i.e. $[x, x/\alpha]$) but truncated in the interval $[a, b]$ (as discussed previously)
- d slightly larger than a yields a CI similar to the truncated classical CI except for a slight perturbation near $\theta = a$ which ensures the CI cannot be empty and has zero chance of being a single point

- $a < c = d < b$ implies the piecewise-constant TF

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ 1, & c \le \theta < \infty \end{cases}$$

This leads to a very simple expression for the CI, namely [l,u], where:

$$l = \begin{cases} a, & 0 \le x \le (1-\alpha)a \\ \dfrac{x}{1-\alpha}, & (1-\alpha)a < x < (1-\alpha)c \\ c, & (1-\alpha)c \le x \le c \\ x, & c < x \le d \end{cases}, \quad u = \begin{cases} c, & 0 \le x \le \alpha c \\ \dfrac{x}{\alpha}, & \alpha c < x < \alpha b \\ b, & \alpha b \le x \le d \end{cases}$$

Example 4 Some other CIs implied by the linear TF

Figure 4 shows 3 sets of 80% CIs for θ if $1 \leq \theta \leq 2$

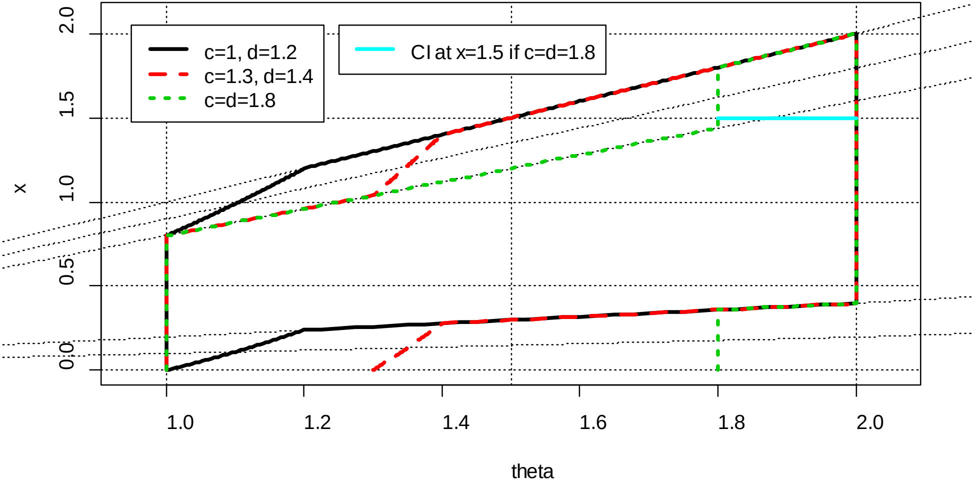
implied by the linear TF with parameters (respectively):

(c, d) = (1, 1.2), (1.3, 1.4) and (1.8, 1.8)

For c = d = 1.8, the CI at x = 1.5 is [1.8, 2] (shown light blue)

Note that none of the CIs in the figure are empty and none have a nonzero probability of containing only a single point

Figure 4 Three other sets of 80% CIs implied by the linear TF



4. Engineering a confidence interval via tail functions

Consider a CI implied by the linear TF

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ \dfrac{\theta - c}{d - c}, & c \le \theta < d \\ 1, & d \le \theta < \infty \end{cases}$$

with $c = a$ and $a \le d \le b$

We will now explore what happens as d changes within the range a to b

Define the conditional expected size (CES) of the CI as

$$CES(\theta) = E\{S(X)\,|\,\theta\} \qquad \left(= E\{U(X)\,|\,\theta\} - E\{L(X)\,|\,\theta\} \right)$$

$$= \int S(x) f(x\,|\,\theta)dx \quad \left(= \int U(x) f(x\,|\,\theta)dx - \int L(x) f(x\,|\,\theta)dx \right)$$

where $f(x\,|\,\theta) = \dfrac{1}{\theta},\ 0 \le x \le \theta$ (the model pdf of X given θ)

Next, suppose we have prior information in terms of a prior density $f(\theta)$

Then define the prior expected size (PES) of the CI as

$$PES = ES(X) = EE\{S(X) \mid \theta\} = \int CES(\theta) f(\theta) d\theta = \int S(x) f(x) dx$$

where $f(x) = \int f(x \mid \theta) f(\theta) d\theta$ (the unconditional pdf of *X*)

NB: $S(x)$, $CES(\theta)$ and *PES* are all implictly functions of *d*, α, *a* and *b*

One reasonable criterion for choosing *d* is to minimize the PES of the CI

Example 5 Analysis of size

Suppose $\alpha = 0.2$, a = 1, b = 2, $\theta \sim U(1,2)$ and we apply the linear TF

$$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < 1 \\ \dfrac{\theta - 1}{d - 1}, & 1 \le \theta < d \\ 1, & d \le \theta < \infty \end{cases} \qquad \text{with } 1 \le d \le 2$$

Figure 5 shows the size of the CI as a function of x for selected values of d

Figure 6 shows the CES as a function of θ for selected values of d

Figure 7 shows the PES as a function of d

Figure 5 Size of a confidence interval

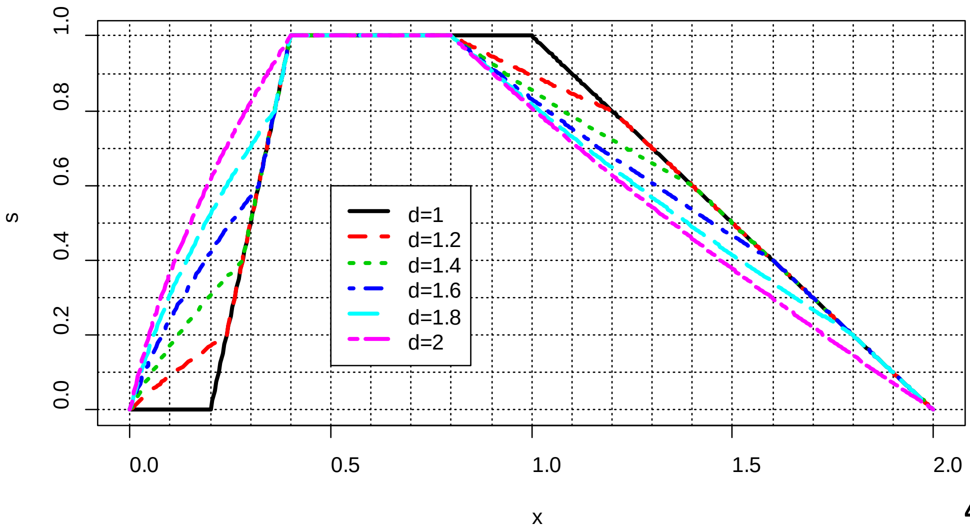


Figure 6 Conditional expected size of a confidence interval

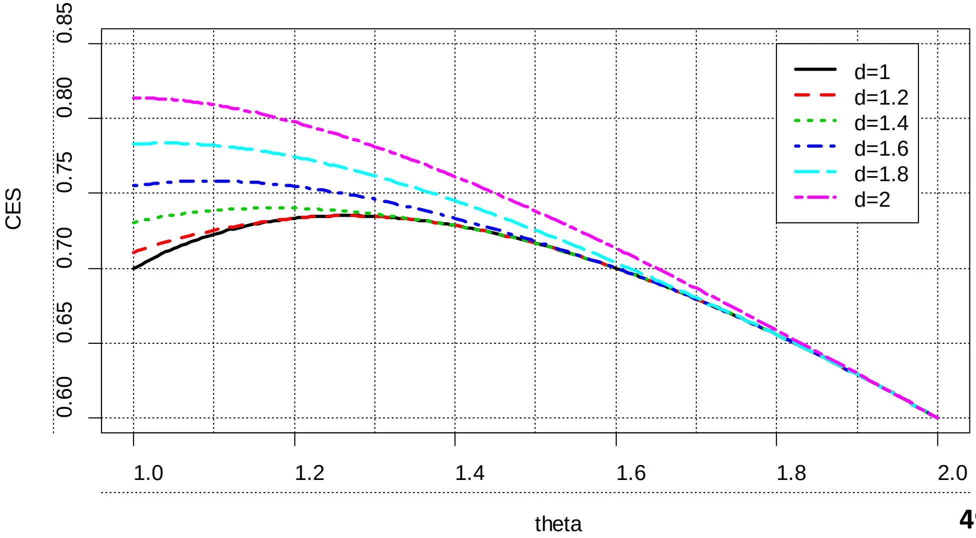
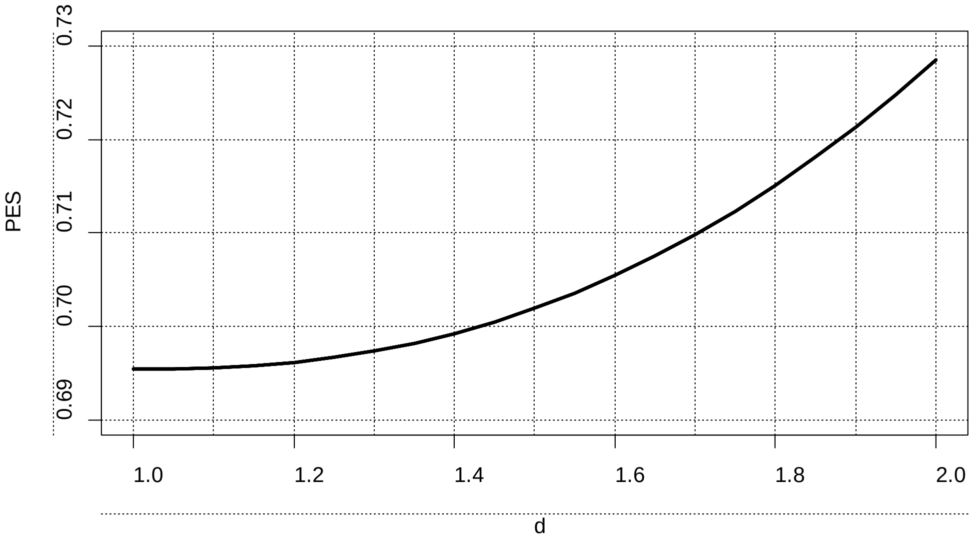


Figure 7 Prior expected size of a confidence interval



In Figure 5 we see some trade-offs:

- Small *d* implies narrow CIs if *x* is small and wide CIs if *x* is large
- Large *d* implies wide CIs if *x* is small and narrow CIs if *x* is large

But in Figure 6, the CES decreases uniformly as *d* decreases, for any θ

So if the CES is averaged over any prior (e.g. the one specified), the resulting PES must also decrease uniformly as *d* decreases

This pattern is evident in Figure 7, and so if we wish to minimize PES (and this is our one and only criterion), then the best choice of *d* is $a = 1$

But this 'optimal' d = 1 CI has a nonzero probability of being empty equal to $a\alpha / \theta$ (between 10% if θ = 2 and 20% if θ = 1, inclusive)

The PESs of the 3 CIs defined by:	d = 1,	d = 1.2,	d = 2
are (respectively):	0.6955,	0.6962,	0.7285

The d = 1.2 CI and the d = 2 CI are equally attractive on account of both having zero probability of being empty (or consisting of a single point)

But the d = 1.2 CI may be considered 'better' than the d = 2 CI. Why?
Because it is only 0.1% wider than the possibly empty d = 1 CI, whereas the d = 2 CI is 4.7% wider (i.e. relatively much wider)

5. Summary and discussion

In this paper we have shown how tail functions can be used to:

- generalize the classical $1-\alpha$ CIs for θ in the model $X \sim U(0,\theta)$
- solve the problem of a classical CI possibly being empty, which arises under the constraint $a \le \theta \le b$ (where $0 < a < b \le \infty$)
- engineer a CI with attractive properties, e.g. define a class of TFs and associated CIs, and then find a CI in that class which cannot be empty and has a small prior expected size under a given prior $f(\theta)$

We studied the linear tail function $$\tau(\theta) = \begin{cases} 0, & 0 \le \theta < c \\ \dfrac{\theta - c}{d - c}, & c \le \theta \le d \\ 1, & d < \theta < \infty \end{cases}$$

and focussed on the special class defined by $c = a$ and $a \le d \le b$, with illustrations including the particular CI implied by $d = b$

We showed that the value of d which minimizes PES is a, and this leads to the 'optimal' classical CI $[x, x/\alpha]$ but truncated inside $[a, b]$ (with the problem then arising that this CI is empty when $x < \alpha a$)

This truncated CI is in fact optimal (as regards minimizing PES) amongst all CIs for θ under the constraint $a \le \theta \le b$ and under all priors $f(\theta)$

This can be shown using Equation 17 in Brown, Casella and Hwang (1995)

This truncated CI is also implied by the unified approach of Feldman and Cousins (1998), which is based only on the constraint $a \le \theta \le b$ and does not involve a prior distribution (see also Mandelkern, 2002)

A question which may be asked:

Why not use the Bayesian approach when a prior is available?

Because the Bayesian approach leads to an 'improper' confidence interval

This means that under an informative prior, any Bayesian interval estimate for θ, e.g. the highest posterior density region (HPDR), has a frequentist coverage probability (FCP) that is not the desired level of $1-\alpha$ for all θ

Typically the FCP is close to zero for values of θ far from the prior mean (or mode or median)

This can be illustrated by finding the 80% HPDR for θ in Example 5 and the FCP of that HPDR, and then showing that FCP $\rightarrow$ 0 as $\theta \rightarrow b = 2$

Also in this context, the PES of the HPDR works out as (slightly) less than that of the truncated classical CI.

Thus, the Bayesian approach leads to an even shorter CI a priori
− but this at the expense of that CI being 'improper'

In general, the tail functions approach provides an alternative to the Bayesian when prior information is available − one which can use that information to produce shorter intervals, without compromising coverage

Some avenues for further research

- CIs for θ based on a random sample $X_1,...,X_n \sim iid\, U(0,\theta)$
  (leading to more complicated equations involving
  $X_{(n)} = \max(X_1,...,X_n)$ and a need for the NRA)

- Other criteria for optimality (e.g. choosing the CI with smallest
  prior probability of being wider than some specified amount)

- Other TFs (e.g. a smoother version of the linear TF which leads to
  smoother lines for the lower and upper bounds in Figure 4)

References

Brown, L.D., Casella, G., and Hwang, J.T.G. (1995). Optimal confidence sets, bioequivalence, and the Limaçon of Pascal. *Journal of the American Statistical Association*, **90**, 880-889.

Feldman, G. J., and Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals. *Physical Reviews D*, **57**, 3873-3889.

Lee, P. (1997). *Bayesian Statistics: An Introduction*. New York: Oxford University Press.

Mandelkern, M. (2002). Setting confidence intervals for bounded parameters. *Statistical Science*, **17**, 149-172 (with Discussion).

Puza, B., and O'Neill, T. (2005). *Interval Estimation via Tail Functions*. Working Paper STAT0005WP, College of Business and Economics, Australian National University, Canberra. Website:

cbe.anu.edu.au/research/papers/pdf/STAT0005WP.pdf

Puza, B.D., and O'Neill, T.J. (2006a). Generalised Clopper-Pearson confidence intervals for the binomial proportion. *Journal of Statistical Computation and Simulation*, **76**, 489-508.

Puza, B., and O'Neill, T. (2006b). Interval estimation via tail functions. *The Canadian Journal of Statistics*, **34**, 299-310.

Puza, B., and O'Neill, T. (2008). Optimal constrained confidence estimation via tail functions. *The Mathematical Scientist*, **33**, 134-140.

Puza, B., and O'Neill, T. (2009). Constrained confidence estimation of the binomial *p* via tail functions. *The Mathematical Scientist*, **34**, 43-48.

Puza, B., Roberts, S., and Yang, M. (2011). Constrained confidence intervals in time series studies of mortality and air pollution. *Environment International*, **37**, 204-209.

Puza, B. (2011). The tail function approach to confidence estimation in survey sampling. *Pakistan Journal of Statistics*, **27**, 543-554.

Puza, B., and Yang, Mo. (2011). Optimal constrained confidence estimation of the Poisson mean via tail functions.

The

Mathematical Scientist, **36**, 95-104.

Puza, B., and Yang, M. (2013). Improved confidence estimation for the exponential mean via tail functions. *Communications in Statistics - Theory and Methods*. DOI:10.1080/03610926.2013.833235. Website:
www.tandfonline.com/doi/abs/

10.1080/03610926.2013.833235#.UxkQMfna701

Stuart, A., Ord, K., and Arnold, S. (1999). *Kendall's Advanced Theory of*

Statistics, Sixth Edition, Volume 2A, Classical Inference and the

Linear Model. London: Arnold.

Thank You