Exact confidence intervals and hypothesis tests for parameters of discrete distributions

MÅNS THULIN[1] and SILVELYN ZWANZIG[2]

[1]*Department of Statistics, Uppsala University, 751 05 Uppsala, Sweden.*
E-mail: mans.thulin@statistik.uu.se
[2]*Department of Mathematics, Uppsala University, 751 05 Uppsala, Sweden.*
E-mail: zwanzig@math.uu.se

We study exact confidence intervals and two-sided hypothesis tests for univariate parameters of stochastically increasing discrete distributions, such as the binomial and Poisson distributions. It is shown that several popular methods for constructing short intervals lack strict nestedness, meaning that accepting a lower confidence level not always will lead to a shorter confidence interval. These intervals correspond to a class of tests that are shown to assign differing p-values to indistinguishable models. Finally, we show that among strictly nested intervals, fiducial intervals, including the Clopper–Pearson interval for a binomial proportion and the Garwood interval for a Poisson mean, are optimal.

Keywords: binomial distribution; confidence interval; expected length; fiducial interval; hypothesis test; Poisson distribution

1. Introduction

Hypothesis testing and interval estimation of parameters in discrete distributions are two of the classic statistical problems, particularly for the binomial and Poisson distributions, which remain two of the most important statistical models. The fact that these distributions are discrete makes it impossible to construct non-randomized confidence intervals that have coverage equal to $1 - \alpha$ for all values of the unknown parameter θ, and, equivalently, impossible to construct two-sided tests with size equal to α for all pairs (α, θ_0), where θ_0 denotes the value of θ under the null hypothesis. It is however possible to construct confidence intervals that have coverage at least equal to $1 - \alpha$ for all values of the unknown parameter, and tests that have size at most equal to α. Such intervals and tests are called exact, and are the topic of this paper.

Given an observation x, the classic method of constructing exact confidence intervals for parameters of some common discrete distributions is to use the fiducial interval of Fisher [17,36]: (θ_L, θ_U) where θ_L and θ_U are such that

$$\sum_{k \leq x} P_{\theta_L}(X = k) = \alpha/2 \quad \text{and} \quad \sum_{k \geq x} P_{\theta_U}(X = k) = \alpha/2. \tag{1}$$

For the binomial parameter, the fiducial interval is known as the Clopper–Pearson interval [11] and for the mean of a Poisson distribution it is known as the Garwood interval [18].

The hypothesis $H_0 : \theta = \theta_0$ can be tested against the alternative $H_1 : \theta \neq \theta_0$ by checking whether θ_0 is contained in the fiducial interval. The p-value $\lambda_f(\theta_0, x)$ of this test is two times

the smaller p-value of two one-sided tests:

$$\lambda_f(\theta_0, x) = \min\left(2 \cdot \sum_{k \leq x} P_{\theta_0}(X = k), 2 \cdot \sum_{k \geq x} P_{\theta_0}(X = k), 1\right). \qquad (2)$$

In their seminal paper on binomial confidence intervals, Brown *et al.* [7] write: "The Clopper–Pearson interval is wastefully conservative and is not a good choice for practical use, unless strict adherence to the prescription $C(p, n) > 1 - \alpha$ is demanded," where $C(p, n)$ denotes the coverage probability. Instead they recommend using approximate intervals, which obtain the nominal confidence level $1 - \alpha$ in some average sense, but have lower coverage for some values of θ. Such intervals are typically shorter than exact intervals, and their corresponding tests typically have higher power. These advantages comes at the cost that the actual confidence levels may be much lower than stated and that the size of tests may be inflated. For popular approximate intervals, the deviations in coverage from $1 - \alpha$ may be non-negligible even for large sample sizes [31]. For this reason, some statistician prefer to use exact methods like those discussed in this paper, in order to guarantee that confidence levels are not exaggerated and type I error rates are not understated.

When other criteria than merely coverage levels and expected lengths are considered, exact confidence intervals can moreover compare favourably to approximate intervals [24,32]. Finally, even if one prefers to use average coverage as a criterion for comparing confidence intervals, it is of interest to study exact intervals due to the facts that these intervals can be adjusted to have coverage $1 - \alpha$ on average, and that such adjusted intervals tend to have shorter expected length than other approximate intervals [25,30]. For comparisons of exact and approximate intervals in the binomial setting, and further arguments for using exact methods for discrete distributions, see [31].

Regarding the fiducial Clopper–Pearson interval, Brown *et al.* [7] also write "better exact methods are available; see, for instance, [5] and [9]." Fiducial intervals are equal-tailed, meaning that the lower bound is a $1 - \alpha/2$ lower confidence bound and that the upper bound is a $1 - \alpha/2$ upper confidence bound. Several authors, including those mentioned by Brown *et al.* [7] in the above quote, have proposed shorter exact intervals that improve upon fiducial intervals by letting the tail-coverages vary for different x, so that their bounds no longer are $1 - \alpha/2$ confidence bounds [4,5,9,10,13,14,19,21,26,28,35]. Such intervals, known as strictly two-sided intervals, tend to have less conservative coverage and are typically shorter than fiducial intervals. Their use has been advocated by [1,2,15,16,20,25,27] and [22], among others.

Unlike the equal-tailed fiducial intervals, the p-values of tests corresponding to strictly two-sided confidence intervals can not be written as two times the smaller p-value of two one-sided tests. Instead, for some test statistic $T(\theta_0, X)$ satisfying mild regularity conditions detailed in Section 2, the p-value of a strictly two-sided test is defined as

$$\lambda(\theta_0, x) = P_{\theta_0}\big(T(\theta_0, X) \geq T(\theta_0, x)\big).$$

If the null distribution of $T(\theta_0, X)$ is asymmetric, the level α rejection region of such a test is not the intersection of the rejection regions of two one-sided level $\alpha/2$ tests.

The main goal of this paper is to show that strictly two-sided confidence intervals and hypothesis tests suffer from several problems. These are illustrated in Figure 1, in which the p-values
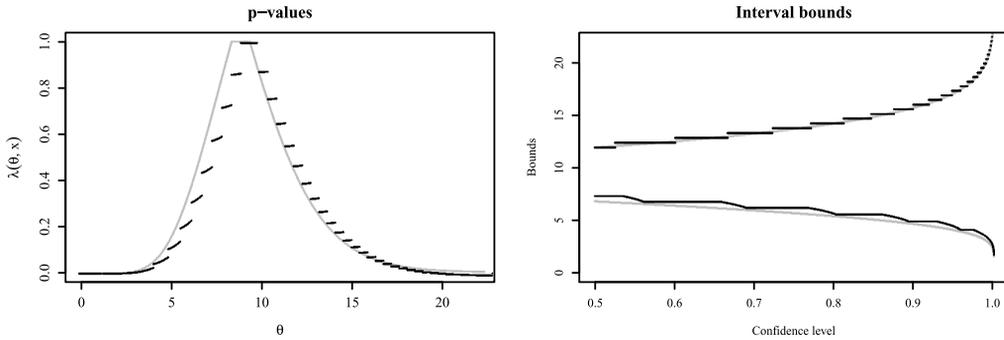
Figure 1. *p*-values and interval bounds for the mean of a Poisson distribution, when $x = 9$ has been observed. The strictly two-sided Sterne [28] method is shown in black, and the fiducial Garwood [18] method is shown in grey.

and interval bounds for the mean of a Poisson distribution are shown for two tests and their corresponding confidence intervals. The first of these is the strictly two-sided Sterne [28] interval, the other being the fiducial Garwood [18] interval.

In the spirit of Birnbaum [3], the *p*-values are plotted as a function of the value θ_0 of the parameter under the null hypothesis. In the Poisson model, it is reasonable to expect that a small change in the null value of θ should lead to a small change in the *p*-value, since $P_\theta(X = x)$ is continuous in θ, so that there is no concernable difference between the Poisson(θ) and Poisson($\theta + \varepsilon$) models when ε is infinitesimal. This is not the case for the strictly two-sided test: its *p*-value is discontinuous when viewed as a function of θ_0. The evidence against two models, which for all practical purposes are indistinguishable, can therefore differ greatly. Several examples of this are seen in Figure 1; the *p*-value for $\theta_0 = 4.954163$, for instance, is 0.0722, so that the null hypothesis is rejected at the 10% level, while the highly similar hypothesis $\theta_0 = 4.954164$ cannot be rejected as its *p*-value is 0.1071.

Moreover, we would expect that the *p*-value increases as θ_0 goes from 0 to the observed x, and that it thereafter decreases, since this would mean that the *p*-value becomes smaller when the null hypothesis agrees less with the data. This is not the case for the strictly two-sided test. Instead, the *p*-value sometimes increases when the null θ is changed to agree less with the observed x. As an example, consider the *p*-values shown in Figure 1. When $x = 9$ has been observed from a Poisson distribution, the *p*-value when $\theta_0 = 15.6$ is 0.0993, so that the null hypothesis is rejected at the 10% level. However, even though $x = 9$ disagrees even more with the null hypothesis $\theta_0 = 15.95$, the *p*-value for this θ_0 is 0.1011, and the hypothesis can not be rejected. The test corresponding to the fiducial interval does not suffer from either of these problems.

The strictly two-sided confidence interval is no better than its corresponding test. When the interval bounds are plotted as functions of the confidence level $1 - \alpha$, we see two phenomenons. The first is that the interval bounds are discontinuous in $1 - \alpha$, meaning that a small change in α can cause one of the interval bounds to leap. The second is that the bounds sometimes are constant, meaning that a change in α not necessarily will lead to a change in the bounds. For some α, both bounds remain unchanged in an interval $(\alpha - \varepsilon, \alpha + \varepsilon)$. There is therefore no

guarantee that accepting a larger α will lead to a shorter interval; we say that the interval is not *strictly nested*. The fiducial interval does not suffer from either of these problems.

These properties can also cause strictly two-sided test and intervals to behave strangely as more data is collected. As an example, consider the Blaker [4] test for the negative binomial proportion θ. When $k = 19$ successes are observed after $x = 38$ trials, the maximum likelihood estimator is $\hat{\theta} = 0.5$ and Blaker p-value for the test of the hypothesis $\theta = 0.625$ is 0.0.0929, causing us to reject the null hypothesis at the 10% level. If we then decide to collect more data by requiring that $k = 20$ successes should be observed, and observe one failure and one success so that $x = 40$, $\hat{\theta}$ is still 0.5. We would now expect the p-value to decrease as this outcome appears to be even less in line with $\theta = 0.625$. Instead, the Blaker p-value for $k = 20$ and $x = 40$ is 0.106, and we can no longer reject the null hypothesis at the 10% level. Analogous problems arise for confidence intervals. The 90% Blaker confidence interval for θ given $k = 19$ and $x = 38$ is (0.35992, 0.62279), while for $k = 20$ and $x = 40$ it is (0.36202, 0.62689). The latter interval is not, as we normally would expect, a subset of the former. Moreover, the interval based on more data is *wider* than the interval based on less data: the interval widths are 0.263 and 0.265, respectively.

As we will see, intervals lacking strict nestedness is equivalent to their corresponding p-values being discontinuous in θ. Consequently, intervals which are not strictly nested correspond to tests that attach widely differing evidence to indistinguishable hypotheses. We believe that this is an unacceptable property of a hypothesis test, and argue that such intervals and tests should be avoided.

In this paper, we show that these problems are universal for strictly two-sided intervals and tests, when the data is generated by a class of discrete distributions that includes the binomial, Poisson and negative binomial distributions. They also carry over to exact analysis of contingency tables and discrete models with nuisance parameters, when such analyses are based on conditioning that reduces the problem to a one-parameter framework.

In Section 2, we give a formal description of the setting for our results. We then show that the p-values of strictly two-sided tests are discontinuous, and that their corresponding intervals have bounds that are not strictly monotone. Finally, we show that strictly two-sided intervals never are strictly nested, meaning that both interval bounds simultaneously may remain unchanged when α is changed. Section 3 is devoted to showing that strictly two-sided intervals typically have bounds that moreover are discontinuous in α, and that the corresponding p-values lack desirable monotonicity properties. In Section 4, it is then demonstrated that fiducial intervals not only are strictly nested but also are the shortest equal-tailed intervals. The paper concludes with a discussion in Section 5. Most proofs and some technical details are contained in two appendices.

2. The lack of strict nestedness and its implications

2.1. Setting

This section is concerned with nestedness. We start by defining this concept.

Definition 1. *A confidence interval is* nested *if the* $1 - \alpha$ *interval is a subset of the* $1 - \alpha_0$ *interval when* $1 > \alpha > \alpha_0 > 0$, *and* strictly nested *if the* $1 - \alpha$ *interval always is a proper subset of the* $1 - \alpha_0$ *interval.*

If an interval is not strictly nested, accepting a lower confidence level does not always yield a shorter interval, so that sometimes nothing is gained by increasing α. Despite the importance of nestedness, this property has not been discussed much in the literature, likely because it is taken for granted. Notable exceptions are Blaker [4], who proved that the binomial Blyth–Still–Casella interval is not strictly nested and Vos and Hudson [33], who showed by example that the Blaker interval for a binomial proportion lacks strict nestedness.

Next, we give some definitions and state the assumptions under which strictly two-sided intervals are not strictly nested. We will limit our study to parameters of discrete distributions P_θ belonging to a class $\mathcal{P}(\Theta, \mathcal{X})$.

Definition 2. *Let* $\theta \in \Theta$ *denote an unknown parameter, with* Θ *being a connected open subset of* $\mathbb{R}$, *and let* $\mathcal{X} \subseteq \mathbb{Z}$ *be a sample space consisting of consecutive integers. A family of distributions* P_θ *on* $\mathcal{X}$ *parameterized by* $\theta \in \Theta$ *belongs to* $\mathcal{P}(\Theta, \mathcal{X})$ *if*

A1. $\forall (\theta, x) \in \Theta \times \mathcal{X}, P_\theta(X = x) > 0$,
A2. P_θ *is stochastically increasing, i.e.* $P_\theta(X \le x)$ *is strictly decreasing in* θ *for any fixed* $x \in \mathcal{X} \setminus \sup \mathcal{X}$,
A3. *For any fixed* $x \in \mathcal{X}$, $P_\theta(X = x)$ *is differentiable in* θ.

Conditions A1–A3 are satisfied by for instance the binomial, Poisson and negative binomial distributions as long as Θ is the natural parameter space, that is, as long as it has not been restricted. This follows directly from the proposition below, the proof of which is given in Appendix B. The conditions are typically also satisfied for other common parameterizations.

Proposition 1. *If* P_θ *constitutes a regular discrete one-parameter exponential family with an increasing likelihood ratio, where* θ *is the natural parameter, then* $P_\theta \in \mathcal{P}(\Theta, \mathcal{X})$.

To fully understand the implications of the lack of nestedness, we will study the hypothesis tests to which non-nested intervals correspond, so-called strictly two-sided tests:

Definition 3. *Consider a two-sided test of* $H_0 : \theta = \theta_0$ *versus* $H_1 : \theta \ne \theta_0$, *with a test statistic* $T(\theta_0, x)$. *The test is called* strictly two-sided *if the* p-value *of the test is* $\lambda(\theta_0, x) = P_{\theta_0}(T(\theta_0, X) \ge T(\theta_0, x))$ *and it satisfies conditions* B1–B2 *below. Moreover, in case* $\lambda(\theta, x)$, *viewed as a function of* θ, *has a jump at* θ_0 *we define* $\lambda(\theta_0, x) = \liminf_{\theta \to \theta_0} \lambda(\theta, x)$.

B1. *For any* $x \in \mathcal{X}$, *there exists a* $\theta_x \in \Theta$ *such that* $T(\theta_x, x) < T(\theta_x, y)$ *for all* $y \in \mathcal{X} \setminus \{x\}$.
B2. *There exists a* $\theta_0 \in \Theta$ *such that there does not exist a* $\mu \in \Theta$ *for which* $P_{\theta_0}(T(\theta_0, X) = \mu - k) = P_{\theta_0}(T(\theta_0, X) = \mu + k)$ *for all* $k : \mu \pm k \in \mathcal{X}$.

Condition B1 is included to ensure that the test does not yield the same result for all x and θ. The name strictly two-sided comes from condition B2, which ensures that the p-value must be computed by comparing the test statistic to *both* tails of the null distribution simultaneously.

The *p*-value of a strictly two-sided test can be written as

$$\lambda(\theta, x) = \sum_{k \in \mathcal{A}_{\theta,x}} P_\theta(X = k) \qquad \text{where } \mathcal{A}_{\theta,x} = \left\{ k \in \mathcal{X} : T(\theta, k) \geq T(\theta, x) \right\}. \tag{3}$$

For simplicity, we will assume that the test statistic is such that

B3. For any $\theta \in \Theta$, there exists $x_\theta \in \mathcal{X}$ such that $T(\theta, x)$ is decreasing in x when $x < x_\theta$ and increasing in x when $x > x_\theta$.

Under B3, the set $\mathcal{A}_{\theta,x}$ has a particularly simple form.

Proposition 2. *Under* B3, *the functions* $k_1(\theta, x) := \min\{k \geq x_\theta : T(\theta, k) \geq T(\theta, x)\}$ *and* $k_2(\theta, x) := \max\{k \leq x_\theta : T(\theta, k) \geq T(\theta, x)\}$ *are such that*

$$\mathcal{A}_{\theta,x} = \left\{ k \in \mathcal{X} : k \geq k_1(\theta, x) \right\} \cup \left\{ k \in \mathcal{X} : k \leq k_2(\theta, x) \right\}. \tag{4}$$

For any x, *at least one of* $k_1(\theta, x)$ *and* $k_2(\theta, x)$ *is non-constant in* θ.

The proof of the proposition is given in Appendix B.

When x is fixed and θ is varying we will refer to $\lambda(\theta, x)$ as the *p*-value function. We define the corresponding confidence interval using the convex hull of $\{\theta : \lambda(\theta, x) > \alpha\}$ to ensure that it in fact is an interval; as we will see in Section 3, $\{\theta : \lambda(\theta, x) > \alpha\}$ itself is not always connected. The interval in the following definition is guaranteed to be nested: if $\alpha > \alpha_0$ the convex hull of $\{\theta : \lambda(\theta, x) > \alpha\}$ is a subset of the convex hull of $\{\theta : \lambda(\theta, x) > \alpha_0\}$.

Definition 4. *The* $1 - \alpha$ *confidence interval* $I_\alpha(x) = (L_\alpha(x), U_\alpha(x))$ *corresponding to a test is*

$$I_\alpha(x) = \left(\inf\{\theta : \lambda(\theta, x) > \alpha\}, \sup\{\theta : \lambda(\theta, x) > \alpha\} \right). \tag{5}$$

A confidence interval is said to be strictly two-sided *if it is based on the inversion of a strictly two-sided test.*

2.2. Examples of strictly two-sided tests

We will focus on four commonly used strictly two-sided tests, which satisfy conditions B1, B2 and B3 for some common discrete distributions, including the binomial, Poisson and negative binomial distributions. These tests are briefly described below. Further details, as well as conditions for B1–B3 to hold, are given in Appendix A.

The likelihood ratio test, for which $T(\theta, x)$ is the likelihood ratio statistic [20,27].

The score test, for which $T(\theta, x)$ is the score statistic [20,27].

The Sterne test, for which $T(\theta, x) = 1/P_\theta(X = x)$ [28].

The Blaker test, which in fact is a class of tests. Given a statistic $S(x)$, the Blaker statistic is $T(\theta, x) = 1/\min\{P_\theta(S(X) \leq S(x)), P_\theta(S(X) \geq S(x))\}$, was introduced in Blaker [4]. See also [37] for a interpretation based on confidence curves. In the binomial, negative binomial and Poisson settings, we will use the sufficient statistic $S(x) = x$, as is common.

In Section 2.5, we will discuss confidence intervals that have varying tail-coverage but are based on minimization algorithms rather than test inversion. Because these intervals do not fall under Definition 4 we will refer to them as being of strictly two-sided-type rather than as being strictly two-sided.

2.3. Lack of strict nestedness and its interpretation

We will now show that strictly two-sided intervals lack strict nestedness, and that this is caused by jumps in the *p*-value function $\lambda(\theta, x)$, viewed as a function of θ.

Proposition 3. *Assume that $P_\theta \in \mathcal{P}(\Theta, \mathcal{X})$. Let $\lambda(\theta, x)$ be the p-value function of a strictly two-sided test and let $I_\alpha(x)$ denote its corresponding strictly two-sided confidence interval. Then for any $x \in \mathcal{X}$:*

- (a) $\lambda(\theta, x)$ *is not continuous in θ,*
- (b) *the bounds of $I_\alpha(x)$ are not strictly monotone in α,*
- (c) $I_\alpha(x)$ *is not strictly nested.*

First, we show that $\lambda(\theta, x)$ has jumps. For any fixed $x \in \mathcal{X}$, by Proposition 2 we have, under B3,

$$\lambda(\theta, x) = \sum_{k \in \mathcal{A}_{\theta,x}} P_\theta(X = k) = \sum_{k \geq k_1(\theta,x)} P_\theta(X = k) + \sum_{k \leq k_2(\theta,x)} P_\theta(X = k), \qquad (6)$$

where at least one of the $k_i(\theta, x)$ is non-constant in θ. $k_i(\theta, x)$ are integer-valued step-functions. Thus, for $\varepsilon > 0$ whenever $k_i(\theta, x) < k_i(\theta + \varepsilon, x)$, k_i must have a jump between θ and $\theta + \varepsilon$. This induces a jump in the *p*-value function as well. To see this, assume without loss of generality that $k_1(\theta + \varepsilon, x) = k_1(\theta, x)$ and $k_2(\theta + \varepsilon, x) = k_2(\theta, x) + 1$. Then

$$\lambda(\theta + \varepsilon, x) = \sum_{k \geq k_1(\theta,x)} P_{\theta+\varepsilon}(X = k) + \sum_{k \leq k_2(\theta,x)} P_{\theta+\varepsilon}(X = k) + P_{\theta+\varepsilon}(X = k_2(\theta, x) + 1),$$

but by A1 and A3,

$$\lim_{\varepsilon \searrow 0} \lambda(\theta + \varepsilon, x) = \lambda(\theta, x) + P_\theta(X = k_2(\theta, x) + 1) > \lambda(\theta, x).$$

Thus $\lambda(\theta + \varepsilon, x) \searrow \lambda(\theta, x)$ as $\varepsilon \searrow 0$ and the function is hence not continuous in θ. In particular, we have shown that $\lambda(\theta, x)$ has the following property:

Lemma 1. *Under the assumptions of Theorem 3, $\lambda(\theta, x)$ as a function of θ has a jump whenever a point is added to or removed from $\mathcal{A}_{\theta,x}$.*

Values of α for which $I_\alpha(x)$ is not strictly nested correspond to the jumps in $\lambda(\theta, x)$. To see this, note that if the interval $(\alpha_0, \alpha_1) \subseteq (0, 1)$ is such that

$$\{\theta : \lambda(\theta, x) \in (\alpha_0, \alpha_1)\} = \varnothing \qquad (7)$$

then for $\alpha \in (\alpha_0, \alpha_1)$, we have $\lambda(\theta, x) > \alpha$ if and only if $\lambda(\theta, x) > \alpha_1$, which means that the lower interval bound

$$L_\alpha(x) = \inf\{\theta : \lambda(\theta, x) > \alpha\} = \inf\{\theta : \lambda(\theta, x) > \alpha_1\} = L_{\alpha_1}(x)$$

so that $L_\alpha(x)$ is not strictly monotone in α. By definition, the interval is not strictly nested if there exists an α such that both $L_\alpha(x)$ and the upper interval bound $U_\alpha(x)$ simultaneously are constant in a neighbourhood of α. The proof that there always exists such an α is somewhat technical, and is deferred to Appendix B.

In particular, Proposition 3 holds when the test and its corresponding confidence interval are exact. The proposition is illustrated for exact tests and intervals in Figures 2–3. In Figure 2, p-values for the strictly two-sided [4,28], likelihood ratio and score tests [20,27] are compared to the p-values of the non-strictly two-sided test that corresponds to the fiducial interval in the Poisson and binomial settings. It is readily verified that the strictly two-sided tests satisfy B1–B3; see Appendix A. In Figure 3, the interval bounds of some strictly two-sided intervals are compared to the bounds of the fiducial interval. In the Poisson case, the Sterne, Blaker, likelihood ratio, score, Crow–Gardner [10,14] and Kabaila–Byrne [21] (the latter two being of strictly two-sided-type) intervals are compared to the Garwood interval. In the binomial case, the Sterne, Blaker, likelihood ratio, score, Crow [5,9,13] (which is of strictly two-sided-type) and Göb and Lurz [19] intervals are compared to the Clopper–Pearson interval.

2.4. The largest α for which an interval is strictly nested

Proposition 3 tells us that strictly two-sided confidence intervals lack strict nestedness and that their bounds are not strictly monotone in α. This may however not be a great problem if the lack of strict nestedness and monotonicity occurs only for α close to 1.

Under some stronger assumptions on $T(\theta, x)$, $\mathcal{X}$ and P_θ we can derive expressions for the largest α for which $I_\alpha(x)$ is strictly nested and the largest α for which each interval bound is strictly monotone. As we will see, these bounds for α are usually close to 0, meaning that the lack of strict nestedness and monotonicity occurs also for α that are used in practice.

We restrict our attention to samples spaces of the form $\mathcal{X} = \{0, 1, 2, \ldots\}$ or $\mathcal{X} = \{0, 1, 2, \ldots, n\}$, for some known $n < \infty$. Moreover, we will require some additional conditions, which essentially make up stronger versions of A2 and B3:

A2$^+$. $P_\theta(X \le x)$ is strictly decreasing in θ for any $x \in \mathcal{X} \setminus \sup \mathcal{X}$.
B3$^+$.   (i) For any $\theta \in \Theta$, there exists $x_\theta \in \mathcal{X}$ such that $T(\theta, x)$ is strictly decreasing in x when $x < x_\theta$ and strictly increasing in x when $x > x_\theta$.
      (ii) For any $x \in \mathcal{X}$, there exists a $\theta_x \in \Theta$ such that $\lambda(\theta_x, x) = 1$ and $T(\theta, x)$ is strictly decreasing in θ when $\theta < \theta_x$ and strictly increasing in θ when $\theta > \theta_x$.
      (iii) x_θ is an increasing function of θ.

Proposition 4. *Assume that $\mathcal{X} = \{0, 1, 2, \ldots\}$ or $\mathcal{X} = \{0, 1, 2, \ldots, n\}$. Under A2$^+$, B3$^+$ and the assumptions of Proposition 3 it holds that*

(a) *There exists an $\alpha_{\text{nest}} > 0$ such that $I_\alpha(x)$ is strictly nested for all $x \in \mathcal{X}$ and $\alpha \le \alpha_{\text{nest}}$.*
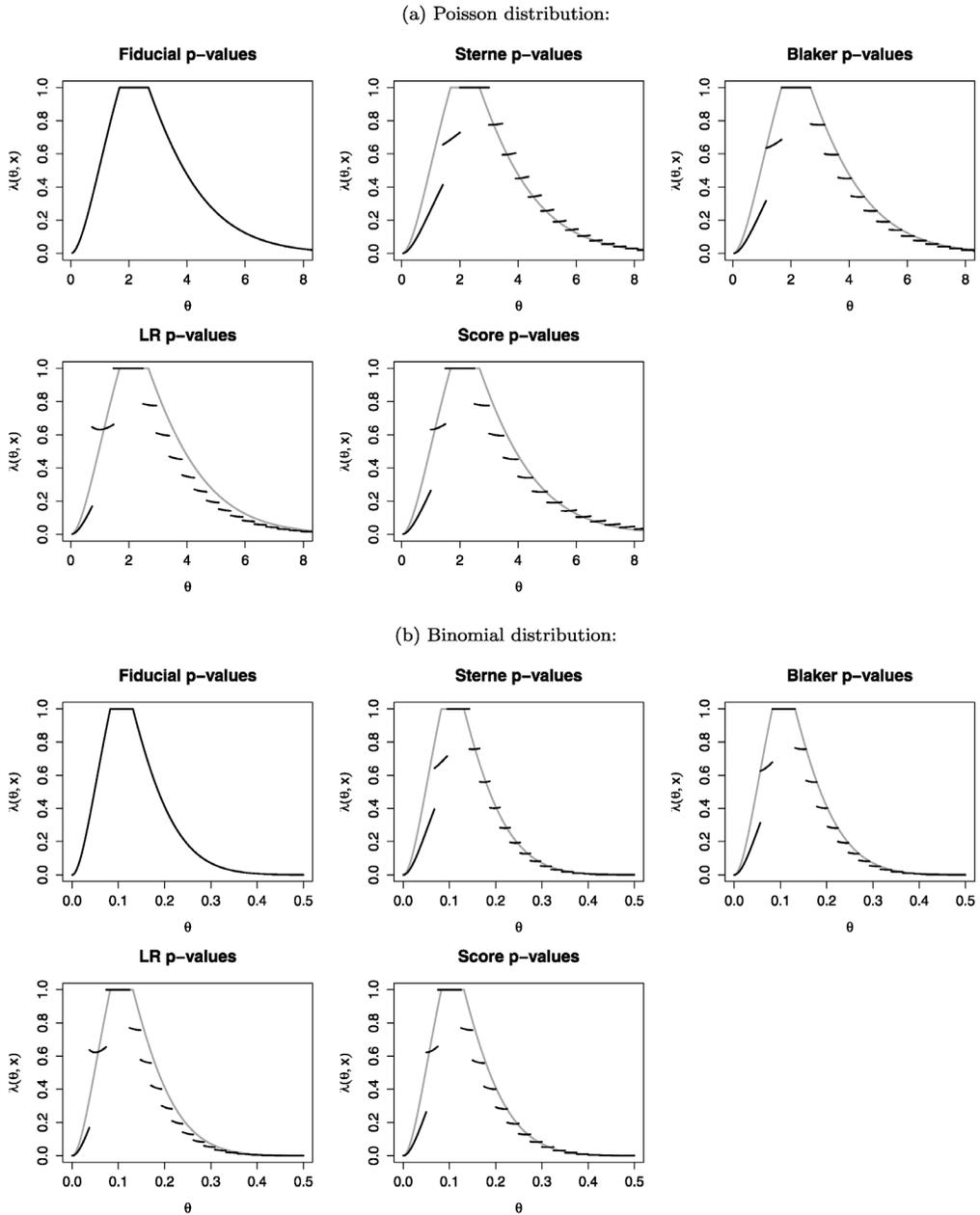
(a) Poisson distribution:

(b) Binomial distribution:

Figure 2. Unlike the p-values for the fiducial test (shown in grey in all plots), the strictly two-sided Sterne, Blaker, likelihood ratio (LR) and score p-values are discontinuous and not bimonotone. In (a), the p-values are shown when $x = 2$ is an observation from a Poisson distribution with null mean θ. In (b), the p-values are shown when $x = 2$ is an observation from a null Bin$(20, \theta)$-distribution.

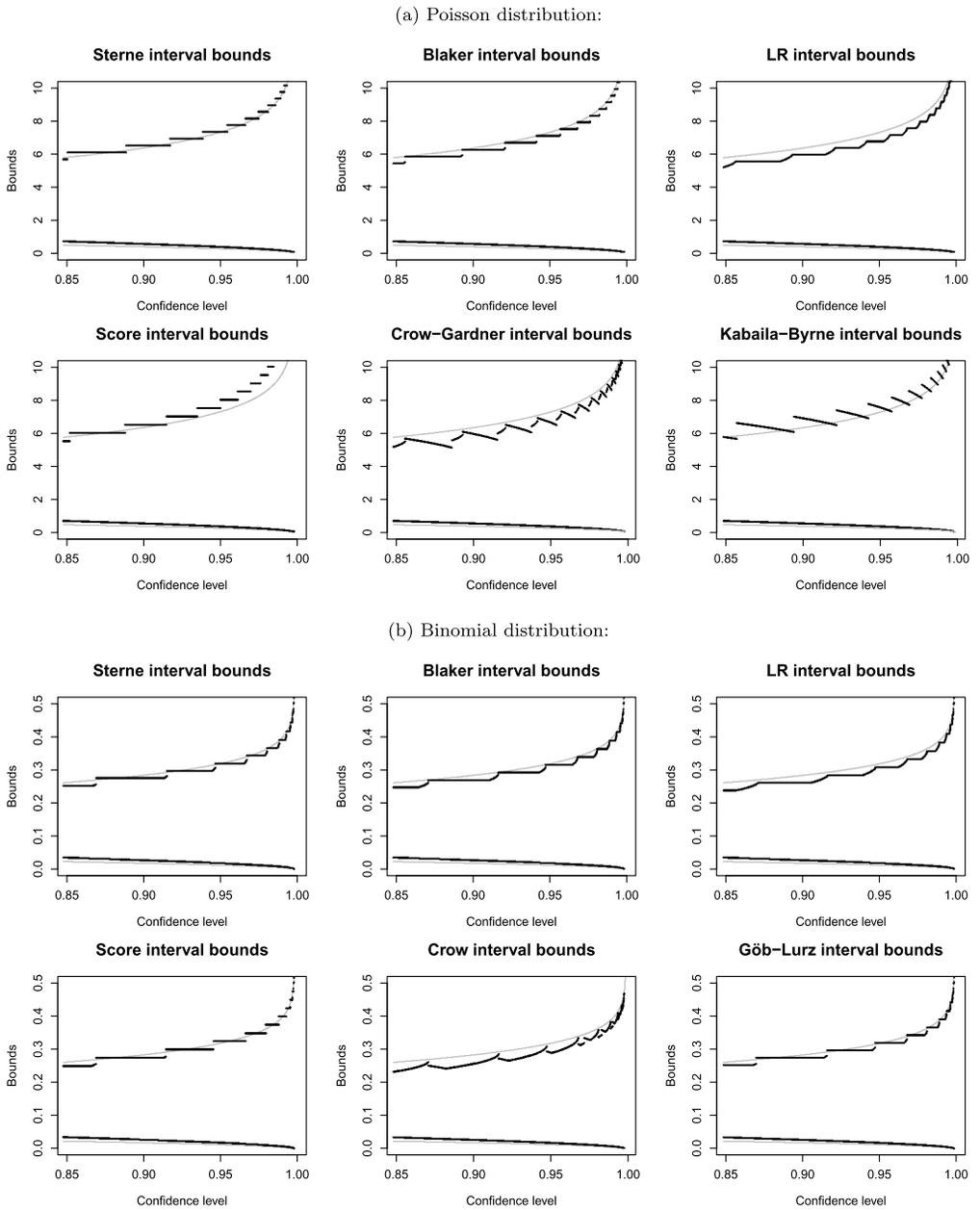(a) Poisson distribution:

(b) Binomial distribution:

Figure 3. Interval bounds of several strictly two-sided and strictly two-sided-type confidence intervals. The intervals are compared to the fiducial interval, the bounds of which are plotted in grey. In (a), the intervals are shown when $x = 2$ is an observation from a Poisson distribution with mean θ. In (b), the intervals are shown when $x = 2$ is an observation from a $\text{Bin}(20, \theta)$-distribution.
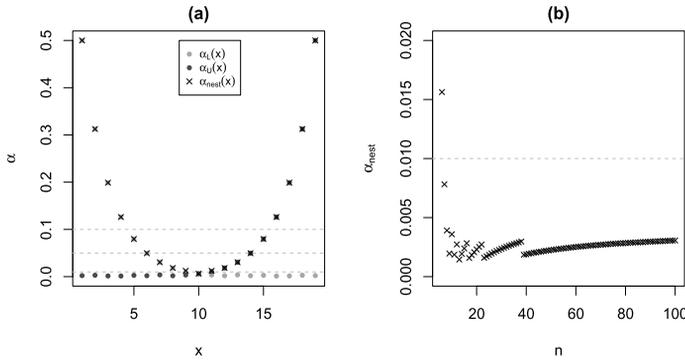
Figure 4. (a) The largest α for which the lower and upper bounds of the Blaker interval for a binomial proportion are strictly monotone ($\alpha_L(x)$ and $\alpha_U(x)$), and the largest α for which the interval is nested conditioned on x ($\alpha_{\mathrm{nest}}(x)$), when $n = 20$. The common choices $\alpha \in \{0.01, 0.05, 0.1\}$ are shown as dashed lines. (b) α_{nest}, the largest α for which the Blaker interval for a binomial proportion is strictly nested, as a function of n.

(b) *Let $\alpha_L = \inf_{x \in \mathcal{X}} \inf_{\theta \in \{\theta : T(\theta,0) > T(\theta,x)\}} \lambda(\theta, x)$. Then* (i) $\alpha_L > 0$, (ii) *for all $x > 0$, $L_\alpha(x)$ is continuous and strictly increasing in α when $\alpha \leq \alpha_L$, and* (iii) *there exists an $x > 0$ and an $\varepsilon > 0$ such that $L_\alpha(x)$ is constant in $(\alpha_L, \alpha_L + \varepsilon)$.*

(c) *For $\mathcal{X} = \{0, 1, 2, \ldots, n\}$, let $\alpha_U = \inf_{x \in \mathcal{X}} \sup_{\theta \in \{\theta : T(\theta,n) > T(\theta,x)\}} \lambda(\theta, x)$. Then* (i) $\alpha_U > 0$, (ii) *for all $x < n$, $U_\alpha(x)$ is continuous and strictly decreasing in α when $\alpha \leq \alpha_U$, and* (iii) *there exists an $x > 0$ and an $\varepsilon > 0$ such that $U_\alpha(x)$ is constant in $(\alpha_U, \alpha_U + \varepsilon)$.*

Proposition 4 deals with α guaranteeing strict monotonicity and nestedness for all x. We can also study monotonicity and nestedness for fixed x. For any $x \in \mathcal{X}$, let $\alpha_L(x)$ denote the largest α for which $L_\alpha(x)$ is strictly monotone, and $\alpha_U(x)$ denote the largest α for which $U_\alpha(x)$ is strictly monotone. Finally, let $\alpha_{\mathrm{nest}}(x)$ be the largest α for which $I_\alpha(x)$ is strictly nested. In Figure 4(a), these quantities are shown for the Blaker interval for a binomial proportion, with $n = 20$ and $x \in \{1, 2, \ldots, 19\}$. In this example, $\alpha_{\mathrm{nest}}(x) < 0.1$ for most x. As is seen, $\alpha_{\mathrm{nest}}(x)$ is often equal to or very close to $\max(\alpha_L(x), \alpha_U(x))$. Figures for other intervals, other n and other distributions are similar.

Figure 4(b) shows α_{nest} for the binomial Blaker interval as a function of the sample size n. It is seen that when $7 \leq n \leq 100$ we have $\alpha_{\mathrm{nest}} < 0.01$ for the Blaker interval, meaning that the interval lacks strict nestedness for virtually all values of α that actually are used in practice for these sample sizes.

2.5. Confidence intervals not based on test-inversion

An interesting class of confidence intervals are based on minimization algorithms. This class includes [5,9,10,13,14,21] and [26] intervals. For such intervals, the shortest interval is determined for each α. What typically occurs for these intervals is that they correspond to inversion of different tests for different α. Often this will result in intervals that lack nestedness (and not only strict
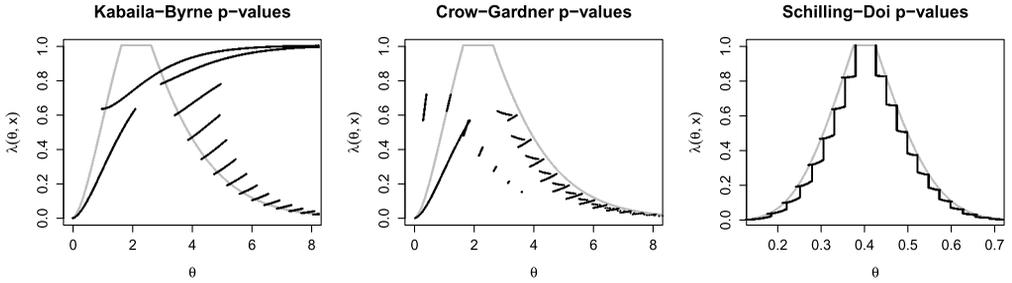
Figure 5. Comparison between *p*-values corresponding to fiducial intervals (grey) and *p*-values corresponding to some minimization-based intervals (black). The *p*-values of the tests corresponding to the Kabaila–Byrne and Crow–Gardner intervals are shown for $x = 2$ being an observation from a Poisson distribution with null mean θ, and the *p*-values of the test corresponding to the Schilling–Doi interval are shown for $x = 8$ being an observation from a null $\mathrm{Bin}(20, \theta)$-distribution.

nestedness), as it leads to some values of θ having multiple *p*-values attached to them. This can be seen in Figure 3: neither the Crow interval for the binomial parameter nor the Crow–Gardner and Kabaila–Byrne intervals for the Poisson parameter are nested.

If a two-sided $1 - \alpha$ interval is (θ_ℓ, θ_u), then the *p*-values for the corresponding two-sided tests of the hypotheses $\theta_0 = \theta_\ell$ and $\theta_0 = \theta_u$ are α. Using this relationship, we can plot the *p*-value functions of tests corresponding to intervals that are not defined in terms of test inversion, such as minimization-based intervals. The lack of nestedness means that the *p*-value function $\lambda(\theta, x)$ of the corresponding test is not a proper function for $x \in \mathcal{X}$ fixed, since some values of θ are mapped to more than one *p*-value. For some intervals, this problem becomes extreme. Two examples of this are the Kabaila–Byrne and Crow–Gardner intervals for a Poisson mean, shown in Figure 5. For other intervals, the lack of nestedness results in less extreme *p*-value functions. An example of this is the Schilling–Doi interval for a binomial proportion; in Figure 5 the jumps in its *p*-value function are shown as vertical lines, in order to make the consequences of the non-nestedness easier to spot.

3. Continuity and bimonotonicity

For $\Theta \subseteq \mathbb{R}$, we say that a function $f : \Theta \to \mathbb{R}$ is strictly bimonotone on Θ if there exist $\theta_0, \theta_1 \in \Theta$ such that f is strictly increasing on $(\inf \Theta, \theta_0)$, constant on (θ_0, θ_1) and strictly decreasing on $(\theta_1, \sup \Theta)$.

As have been argued for example, by Hirji [20] and Vos and Hudson [33], this type of bimonotonicity is a highly desirable property of *p*-values when viewed as a function of θ. Ideally $\lambda(\theta, x)$ should increase monotonically from 0 to 1 and then decreases monotonically to 0, just like the *p*-values of the tests corresponding to fiducial intervals do in Figure 2. One reason that this property is desirable is the following result.

Proposition 5. *The bounds of a confidence interval are discontinuous in α if their corresponding p-value function is not strictly bimonotone in θ.*

Proof. Assume without loss of generality that there exist $\theta_0 < \theta_1 < \inf\{\theta : \lambda(\theta, x) = 1\}$ such that $\lambda(\theta, x)$ is increasing in θ in the interval $(\inf \Theta, \theta_0)$ and decreasing or constant in the interval (θ_0, θ_1). Let $\alpha_0 = \lambda(\theta_0, x)$. Then $\theta_1 = \inf\{\theta > \theta_0 : \lambda(\theta, x) > \alpha_0\}$. Thus $L_{\alpha_0}(x) = \theta_0$ but for all $\varepsilon > 0$, $L_{\alpha+\varepsilon}(x) \geq \theta_1$, meaning that $L_\alpha(x)$ has a jump of length $\theta_1 - \theta_0 > 0$ at $\alpha = \alpha_0$. An analogous argument holds for the upper bound. □

Hirji [20] mentions that p-value functions of strictly two-sided tests need not be bimonotone, whereas Vos and Hudson [33] showed by example that the Blaker test for a binomial proportion lacks bimonotonicity. Upon closer inspection of Figures 2 and 3, it can be seen that all the strictly two-sided tests considered here suffer from this problem.

Next, we give a condition under which the p-value function of a strictly two-sided test is strictly bimonotone for fixed x, the derivation of which is given in Appendix B. The bimonotonicity condition requires the following additional assumptions, which are satisfied by the binomial, negative binomial and Poisson distributions.

A4. For $x \in \mathcal{X} \setminus \sup \mathcal{X}$, $\lim_{\theta \to \inf \Theta} P_\theta(X \leq x) = 1$ and $\lim_{\theta \to \sup \Theta} P_\theta(X \leq x) = 0$.

A5. For $k_1, k_2 \in \mathcal{X}$ such that $k_1 \geq k_2 + 2$, $\sum_{k \geq k_1} P_\theta(X = k) + \sum_{k \leq k_2} P_\theta(X = k)$ has a unique minimum in the interior of Θ.

Proposition 6. *Under the assumptions and notation of Proposition 3, assume that P_θ satisfies conditions A4 and A5. Let $\theta_r(\theta_0, x)$ be the solution to*

$$\sum_{k=k_2(\theta_0,x)+1}^{k_1(\theta_0,x)-1} \frac{d}{d\theta} P_\theta(X = k) = 0 \tag{8}$$

in the interior of Θ. Then

(a) *$\lambda(\theta, x)$ is strictly bimonotone in θ for any fixed $x \in \mathcal{X} \setminus \sup \mathcal{X}$.*
(b) *The bounds of $I_\alpha(x)$ are continuous in α,*

if and only if there does not exist (θ_0, x) such that either

$$\theta_0 < \inf\{\theta : \lambda(\theta, x) = 1\} \quad and \quad \theta_0 < \theta_r(\theta_0, x), \quad or$$
$$\theta_0 > \sup\{\theta : \lambda(\theta, x) = 1\} \quad and \quad \theta_0 > \theta_r(\theta_0, x). \tag{9}$$

For any given P_θ, we can evaluate numerically whether the bimonotonicity condition (9) is violated for a pair (θ_0, x). We have not been able to find a strictly two-sided test that passes (9) for any x. Proposition 6 is illustrated in the Poisson and binomial settings in Figures 2–3. When $x = 2$ from a Poisson random variable has been observed, the p-value functions of the Sterne and Blaker tests are non-bimonotone for the first time when $\theta = 3$. For the likelihood ratio test, the first occurrence is at $\theta = 1$ and for the score test the first occurrence is at $\theta = \sqrt{12}$.

A consequence of $\lambda(\theta, x)$ lacking bimonotonicity is that the confidence "interval" $\{\theta : \lambda(\theta, x) > \alpha\}$ may contain holes, and therefore not be an interval at all. The common remedy for this is to redefine the intervals as the convex hull of $\{\theta : \lambda(\theta, x) > \alpha\}$, as we did in Definition 4. This does not change the infimum or supremum of the set, and does therefore not

affect nestedness or continuity of the bounds. Similarly, Fay [15] proposed handling the problem of non-bimonotone p-value functions by redefining the p-values using the convex hull of $\{\theta : \lambda(\theta, x) > \alpha\}$. The redefined p-values are constant where they previously were non-monotone. By Proposition 5, the bounds of the corresponding intervals are however still discontinuous in α.

For the binomial and negative binomial distributions, the left-hand side of (8) is a polynomial of order $k_1(\theta_0, x) - k_2(\theta_0, x) - 1$. For the Poisson distribution, it is straightforward to find a general solution to (8), which yields the following proposition, the proof of which is omitted.

Proposition 7. *For $X \sim \text{Poisson}(\theta)$, the p-value function $\lambda(\theta, x)$ belonging to a strictly two-sided test is bimonotone in θ if and only if there does not exist (θ, x) such that either*

- $\theta < \inf\{\theta : \lambda(\theta, x) = 1\}$ *and* $\theta < (\frac{(k_1(\theta,x)-1)!}{k_2(\theta,x)!})^{1/(k_1(\theta,x)-k_2(\theta,x)-1)}$, *or*
- $\theta > \sup\{\theta : \lambda(\theta, x) = 1\}$ *and* $\theta > (\frac{(k_1(\theta,x)-1)!}{k_2(\theta,x)!})^{1/(k_1(\theta,x)-k_2(\theta,x)-1)}$.

Note that if we let $n = k_1(\theta, x) - k_2(\theta, x) - 1$ then

$$\left(\frac{(k_1(\theta, x) - 1)!}{k_2(\theta, x)!} \right)^{1/(k_1(\theta,x)-k_2(\theta,x)-1)} = \left(\prod_{k=k_2(\theta,x)+1}^{k_1(\theta,x)-1} k \right)^{1/n},$$

the geometric mean of $\mathcal{A}^c_{\theta,x}$.

4. Some results for fiducial intervals

4.1. Fiducial intervals are strictly nested and have continuous bounds

The test corresponding to the fiducial intervals is not strictly two-sided. Its p-values are defined by (2). The following proposition, the proof of which can be found in Appendix B, states that fiducial intervals do not suffer from the problems associated with strictly two-sided intervals.

Proposition 8. *Under* A1–A3 *and* A4, *fiducial intervals are strictly nested. Moreover, for any $x \in \mathcal{X}$ the bounds of the interval are continuous in α and $\lambda_f(\theta, x)$ is continuous in θ.*

4.2. Optimality results

For a binomial proportion, Wang [34] presented results claiming that under certain conditions on α and n the fiducial Clopper–Pearson interval is the shortest interval in the class of exact confidence intervals with monotone bounds. A counterexample to the optimality result of [34] is the strictly two-sided Blaker interval [4], which always is contained in the Clopper–Pearson interval. Among equal-tailed intervals however, fiducial intervals posses length optimality properties. We expect that this is known, but have not been able to find such results in the literature, for which reason we briefly cover length optimality below.

Our main tool for showing length optimality is a theorem due to [6]. Under assumptions A1, A2 and A3, consider the class $\mathcal{M}_{L,\alpha}$ of one-sided $1 - \alpha$ confidence bounds $(L_\alpha(x), \infty) \cap \Theta$ for $\theta \in \Theta$ based on an observation x of $X \sim P_\theta$ satisfying the following three criteria:

C1. $L_\alpha(x) \le L_\alpha(x + 1)$,
C2. $\inf_{\theta \in \Theta} P_\theta(L_\alpha(x) \le \theta) \ge 1 - \alpha$,
C3. $L_\alpha(x)$ only depends on x, α and P_θ.

Criterion C3 rules out randomized bounds, which can be shorter while maintaining exact coverage, but rely on conditioning on information not contained in the sufficient statistic; see, for example, [29]. C3 is implicit in Bolshev's paper; we have added it here for clarity. $\mathcal{M}_{L,\alpha}$ is the class of monotone exact lower confidence bounds. We call an interval (or a bound) $I_\alpha(x)$ in a class of intervals $\mathcal{K}$ the smallest interval in $\mathcal{K}$ if, for any other interval $I_\alpha^*(x) \in \mathcal{K}$, $I_\alpha(x) \setminus I_\alpha^*(x) = \varnothing$. For the $\mathcal{M}_{L,\alpha}$ class, Bolshev [6] proved that the one-sided lower fiducial bound is the smallest bound in $\mathcal{M}_{L,\alpha}$. Under analogous conditions, the upper fiducial bound is similarly the smallest bound in the set $\mathcal{M}_{U,\alpha}$ of exact monotone upper confidence bounds.

The extension of Bolshev's theorem to two-sided confidence intervals is straightforward and does not require the additional conditions that Wang [34] used in the binomial setting. Consider the class $\mathcal{M}_\alpha$ of exact equal-tailed confidence intervals $(L_{\alpha/2}(x), U_{\alpha/2}(x))$ for θ based on an observation x of $X \sim P_\theta$ satisfying

D1. $L_{\alpha/2}(x) \le L_{\alpha/2}(x + 1)$ and $U_{\alpha/2}(x) \le U_{\alpha/2}(x + 1)$,
D2. $\inf_{\theta \in \Theta} P_\theta(L_{\alpha/2}(x) \le \theta) \ge 1 - \alpha/2$ and $\inf_{\theta \in \Theta} P_\theta(U_{\alpha/2}(x) \ge \theta) \ge 1 - \alpha/2$,
D3. $(L_{\alpha/2}(x), U_{\alpha/2}(x))$ only depends on x, α and P_θ.

Note that if an interval belongs to $\mathcal{M}_\alpha$ then it is the intersection of a bound in $\mathcal{M}_{L,\alpha/2}$ and a bound in $\mathcal{M}_{U,\alpha/2}$.

Proposition 9. *The fiducial interval is the smallest interval in $\mathcal{M}_\alpha$.*

Proof. Let $I_\alpha(x) = (L_{\alpha/2}(x), U_{\alpha/2}(x))$ denote the fiducial interval and assume that there is an interval $I_\alpha^*(x) = (L_{\alpha/2}^*(x), U_{\alpha/2}^*(x))$ in $\mathcal{M}_\alpha$ such that $I_\alpha(x) \setminus I_\alpha^*(x) \ne \varnothing$. Then $L_{\alpha/2}^*(x) > L_{\alpha/2}(x)$ or $U_{\alpha/2}^*(x) < U_{\alpha/2}(x)$. Consequently, at least one of the one-sided bounds $(L_{\alpha/2}^*(x), \infty) \cap \Theta$ or $(-\infty, U_{\alpha/2}^*(x)) \cap \Theta$ is smaller than the corresponding fiducial bound. By Bolshev's theorem, this means that $I_\alpha^*(x)$ is not in $\mathcal{M}_\alpha$, which is a contradiction. $\qquad \square$

Similar results can be obtained for intervals with fixed but unequal tails, in a completely analogue manner.

Finally, the fact that the fiducial interval is the smallest interval in $\mathcal{M}_\alpha$ leads to the following proposition, in which the smallness is expressed in the more familiar terms of the interval length $U_{\alpha/2}(x) - L_{\alpha/2}(x)$.

Proposition 10. *Among the intervals in $\mathcal{M}_\alpha$, the fiducial interval minimizes the expected length for all $\theta \in \Theta$ as well as the length for all $x \in \mathcal{X}$.*

Proof. For an interval $(L^*_{\alpha/2}(x), U^*_{\alpha/2}(x)) \in \mathcal{M}_\alpha$ to have shorter length than the fiducial interval $(L_{\alpha/2}(x), U_{\alpha/2}(x))$ it must hold that $L^*_{\alpha/2}(x) > L_{\alpha/2}(x)$ or $U^*_{\alpha/2}(x) < U_{\alpha/2}(x)$. By Proposition 9 neither condition can be fulfilled. Since the fiducial interval therefore minimizes the length for each x, it also minimizes the expected length $\sum_k P_\theta(X = k)(U_{\alpha/2}(k) - L_{\alpha/2}(k))$. $\qquad\square$

A consequence of Proposition 10 is that, in the class of equal-tailed two-sided tests of $\theta = \theta_0$, the test that corresponds to the fiducial interval is admissible in the sense of Cohen and Strawderman [12].

5. Conclusion

There exist a large number of methods for obtaining exact confidence intervals that are shorter than the equal-tailed fiducial intervals. The use of such an interval comes at the cost of losing control over the balance between the coverage levels of the corresponding lower and upper confidence bounds. In many situations it is preferable to use an equal-tailed interval, in order to guard equally against overestimation and underestimation and not to bias the inference in some direction. The case for equal-tailed intervals is further strengthened by the fact that strictly two-sided confidence intervals lack strict nestedness. This causes difficulties with the interpretation of the intervals: what does it mean that, for a particular x, the 92% interval equals the 95% interval? Which confidence level should be reported for such an interval? More seriously, we have also seen that such intervals may yield highly disparate conclusions for two indistinguishable models P_θ and $P_{\theta+\varepsilon}$. From a hypothesis testing perspective, this occurs when the null hypothesis θ_0 is changed slightly. From a confidence interval perspective, it can occur for small changes in α, since the bounds of strictly two-sided intervals typically are discontinuous in α. These problems have been pointed out for specific intervals in the past [4,33]. We have shown that they in fact are inherent to strictly two-sided confidence intervals.

The problems discussed in this paper arise also for strictly two-sided methods for discrete distributions not covered by Definition 2. Examples include the hypergeometric distribution and the joint distribution of two binomial proportions. We have restricted our attention to the class of distributions given by Definition 2 in order to keep the proofs reasonably short.

Strictly two-sided and equal-tailed confidence intervals are the most commonly used types of two-sided confidence intervals. We have seen that strictly-two sided intervals lack strict nestedness and that an extension of Bolshev's theorem shows that the standard fiducial intervals are the shortest equal-tailed exact intervals. While fiducial intervals have been criticized for being overly conservative and too wide [1,7,8], the conclusion of this paper is that they for practical purposes in fact are the optimal strictly nested intervals.

Appendix A: Strictly two-sided tests

A.1. The likelihood ratio and Sterne tests

Let $L(\theta, x) = P_\theta(X = x)$ be the likelihood function of P_θ. The likelihood ratio statistic is

$$T_{\mathrm{LR}}(\theta_0, x) = \frac{\sup_{\theta \in \Theta} L(\theta, x)}{L(\theta_0, x)}$$

and the Sterne statistic is

$$T_{St}(\theta_0, x) = 1/L(\theta_0, x).$$

Both these statistics are minimized when θ_0 is the maximum likelihood estimator of θ given x. Thus for B1 to be satisfied it suffices that the maximum likelihood estimator of θ is well-defined and strictly monotone in x. B2 is satisfied when there exists a θ such that $L(\theta, x)$ is an asymmetric function of x. By definition, B3 is satisfied if there exists an x_0 such that $L(\theta, x)$ is increasing when $x < x_0$ and decreasing when $x > x_0$. This is guaranteed if P_θ has a monotone likelihood ratio.

The binomial, negative binomial and Poisson distributions all have well-defined and strictly monotone maximum likelihood estimators and monotone likelihood ratios. Moreover, their probability functions are in general asymmetric in x. The likelihood ratio and Sterne tests therefore satisfy conditions B1–B3 for these models.

A.2. The score test

Let $U(\theta, x) = \frac{\partial}{\partial \theta} \ln L(\theta, x)$ and let $I(\theta)$ be the Fisher information of P_θ. The score test statistic is

$$T_{Sc}(\theta_0, x) = \frac{(U(\theta_0, x))^2}{I(\theta_0)}.$$

If the maximum likelihood estimator of θ exists and is unique, then B1 is satisfied, with θ_x being the maximum likelihood estimator of θ given x. B2 is satisfied if the distribution of $U(\theta, x)^2$ is asymmetric for some θ. B3 is satisfied if there exists an x_0 such that $U(\theta, x)^2$ is decreasing when $x < x_0$ and increasing when $x > x_0$.

If P_θ is a regular exponential family with natural parameter θ, then $U(\theta, x) = x - E_\theta(X)$ and $I(\theta) = \text{Var}_\theta(X)$. B2 is satisfied if the distribution of X^2 is asymmetric for some θ and B3 is satisfied since $(x - E_\theta(X))^2$ is convex in x. B1–B3 are therefore satisfied for the binomial, negative binomial and Poisson distributions, using the natural parametrizations. These conditions are also satisfied for the most commonly used alternative parametrizations.

A.3. The Blaker test

The Blaker statistic is

$$T_B(\theta_0, x) \propto \left(\lambda_T(\theta_0, x)\right)^{-1},$$

where $\lambda_T(\theta_0, x)$ is the p-value of a test with a rejection region that is the union of the rejection regions of two one-sided level $\alpha/2$ tests. The properties of $T(\theta_0, x)$ therefore depend on the choice of $\lambda_T(\theta_0, x)$. A typical choice is the fiducial p-value (2).

Under A3 and A4, for any $x \in \mathcal{X}$ there exist θ_x such that $\sum_{k \leq x-1} P_{\theta_x}(X = k) < 1/2$ and $\sum_{k \geq x+1} P_{\theta_x}(X = k) < 1/2$. Then we have $\sum_{k \leq x} P_{\theta_x}(X = k) \geq 1/2$ and $\sum_{k \geq x} P_{\theta_x}(X = k) \geq 1/2$, so that $\lambda_f(\theta, x) = 1$. Let $\Theta_1(x)$ denote the set of such θ_x.

Now, let $y = x - 1$. Then $\sum_{k \le y-1} \mathrm{P}_{\theta_x}(X = k) < 1/2$ but $\sum_{k \ge y+1} \mathrm{P}_{\theta_x}(X = k) = \sum_{k \ge x} \mathrm{P}_{\theta_x}(X = k) \ge 1/2$, so if $\theta_x \in \Theta_1(x)$ then $\theta_x \notin \Theta_1(y)$. Similarly, if we let $y = x + 1$, $\sum_{k \le y-1} \mathrm{P}_{\theta_x}(X = k) = \sum_{k \le x} \mathrm{P}_{\theta_x}(X = k) \ge 1/2$, so if $\theta_x \in \Theta_1(x)$ then $\theta_x \notin \Theta_1(y)$. Thus, A3 and A4 are sufficient for B1 to hold for the Blaker statistic based on the fiducial p-value.

B2 holds if the distribution of $\lambda_T(\theta, x)$ is asymmetric in x for some θ. For $\lambda_f(\theta, x)$ this holds if $\mathrm{P}_\theta(X = x)$ is asymmetric as a function of x for some θ.

Finally, B3 is satisfied since the monotonicity of $\mathrm{P}_\theta(X \le x)$ in x implies that $\lambda_T(\theta, x)$ is a bimonotone function of x. B1–B3 are therefore satisfied for the binomial, negative binomial and Poisson distributions.

Appendix B: Proofs

B.1. Proof of Proposition 1

If P_θ is a discrete one-parameter exponential family in natural form, for $(\theta, x) \in \Theta \times \mathcal{X}$ its probability function can be written as

$$p_\theta(x) = \mathrm{P}_\theta(X = x) = \exp\{\theta T(x) - K(\theta)\} h(x), \tag{B.1}$$

where $T : \mathcal{X} \mapsto \mathbb{R}$ is a function that does not depend on θ and $K : \Theta \mapsto \mathbb{R}$ is infinitely often differentiable in Θ since P_θ is regular [23], Theorem 1.17. θ, $T(x)$ and $K(\theta)$ are all finite for $(\theta, x) \in \Theta \times \mathcal{X}$, and thus (B.1) is strictly positive when $(\theta, x) \in \Theta \times \mathcal{X}$, yielding A1. Moreover, A3 follows from the fact that when x is fixed (B.1) is differentiable in θ since $\theta T(x)$ and $K(\theta)$ are infinitely differentiable.

To see that an increasing likelihood ratio implies A2, let $\ell(x) = p_{\theta_2}(x)/p_{\theta_1}(x)$ for $\theta_1 < \theta_2$ in Θ. The likelihood ratio $\ell(x)$ is increasing in x. Let

$$F_\theta(x) = \mathrm{P}_\theta(X \le x) = \sum_{k \le x} \mathrm{P}_\theta(X = k)$$

and

$$G_\theta(x) = \mathrm{P}_\theta(X > x) = \sum_{k > x} \mathrm{P}_\theta(X = k).$$

We consider the cases when $\ell(x) \le 1$ and $\ell(x) \ge 1$ separately.

If $\ell(t) \le 1$ then for $s < t$, $F_{\theta_2}(s) \le F_{\theta_1}(s)$ since $p_{\theta_2}(x) \le p_{\theta_1}(x)$ for all $x < t$.

If $\ell(t) \ge 1$ then $p_{\theta_2}(x) \ge p_{\theta_1}(x)$ when $x > t$ and for $s > t$, $G_{\theta_2}(s) \ge G_{\theta_1}(s)$. Since $F_\theta(x) = 1 - G_\theta(x)$, it follows that $F_{\theta_2}(s) \le F_{\theta_1}(s)$.

B.2. Proof of Proposition 2

First, assume that $k \ge x_\theta$. Then by B3 $T(\theta, \cdot)$ is increasing at k. There are two possible scenarios:

(i) $k \geq k_1(\theta, x)$: By definition, $T(\theta, k_1(\theta, x)) \geq T(\theta, x)$. Since $T(\theta, \cdot)$ is increasing for $x \geq x_\theta$ it follows that $T(\theta, k) \geq T(\theta, k_1(\theta, x)) \geq T(\theta, x)$, meaning that $k \in \mathcal{A}_{\theta, x}$.

(ii) $k < k_1(\theta, x)$: it follows from the definition of $k_1(\theta, x)$ that $T(\theta, k) < T(\theta, x)$, so $k \notin \mathcal{A}_{\theta, x}$.

In summary, if $k \geq x_\theta$ then $k \in \mathcal{A}_{\theta, x}$ if and only if $k \geq k_1(\theta, x)$. An analogous argument shows that if $k \leq x_\theta$ then $k \in \mathcal{A}_{\theta, x}$ if and only if $k \leq k_2(\theta, x)$, and the first part of the proposition follows.

To see that at least one of $k_1(\theta, x)$ and $k_2(\theta, x)$ is non-constant in θ, note that by B1, for any pair $(x, y) \in \mathcal{X}^2$ there exist $(\theta_x, \theta_y) \in \Theta^2$ such that $x \notin \mathcal{A}_{\theta_x, y}$ but $x \in \mathcal{A}_{\theta_y, y}$. The set $\mathcal{A}_{\theta, x}$ is therefore not constant in θ, and thus at least one of $k_1(\theta, x)$ and $k_2(\theta, x)$ must be non-constant in θ.

B.3. Proof of Proposition 3

(a) and (b) were proved in Section 2.3. We will now prove (c). Let $L_\alpha(x)$ and $U_\alpha(x)$ denote the lower and upper bounds of the interval. We will show that for any $x \in \mathcal{X}$ there exists an $\alpha_0 \in (0, 1)$ such that $L_\alpha(x)$ and $U_\alpha(x)$ simultaneously are constant in a neighbourhood of α_0, so that the confidence interval is not strictly nested.

We introduce the mutually disjoint sets

$$
\begin{aligned}
\Theta_1(x) &:= \{\theta : \lambda(\theta, x) = 1\}, \\
\Theta_\ell(x) &:= \{\theta : \theta \leq \inf \Theta_1(x)\} \quad \text{and} \\
\Theta_u(x) &:= \{\theta : \theta \geq \sup \Theta_1(x)\},
\end{aligned}
\tag{B.2}
$$

which are such that $\Theta_\ell(x) \cup \Theta_1(x) \cup \Theta_u(x) = \Theta$. We also define

$$
\Theta_\alpha(x) := \{\theta : \lambda(\theta, x) > \alpha\}.
$$

By condition B1, given $x \in \mathcal{X}$ there exists $\theta_x \in \Theta$ such that

$$
\lambda(\theta_x, x) = P_{\theta_x}\left(T(\theta_x, X) \geq T(\theta_x, x)\right) = 1 - P_{\theta_x}\left(T(\theta_x, X) < T(\theta_x, x)\right) = 1,
$$

so $\Theta_1(x)$ is non-empty.

First, we investigate the behaviour of the bounds when either $\Theta_\ell(x)$ or $\Theta_u(x)$ is empty. Let $\bar{\Theta}_\alpha(x)$ be the closure of $\Theta_\alpha(x)$. Since $\Theta_1(x) \subseteq \bar{\Theta}_\alpha(x)$ for all $\alpha \in (0, 1)$,

$$
\sup \Theta_1(x) \in \bar{\Theta}_\alpha(x) \quad \text{and} \quad \inf \Theta_1(x) \in \bar{\Theta}_\alpha(x).
$$

If $\inf \Theta_1(x) = \inf \Theta$, then $\Theta_\ell(x) = \{\theta : \theta < \inf \Theta\} = \varnothing$. Then

$$
L_\alpha(x) = \inf \Theta_\alpha(x) \leq \inf \Theta_1(x) = \inf \Theta \leq \inf \Theta_\alpha(x),
$$

so $L_\alpha(x) = \inf \Theta$ for all $\alpha \in (0, 1)$. Similarly, if $\sup \Theta_1(x) = \sup \Theta$ then $\Theta_u(x) = \{\theta : \theta > \sup \Theta\} = \varnothing$, and

$$
U_\alpha(x) = \sup \Theta_\alpha(x) \geq \sup \Theta_1(x) = \sup \Theta \geq \sup \Theta_\alpha(x),
$$

so $U_\alpha(x) = \sup \Theta$ for all $\alpha \in (0, 1)$. Thus, when $\Theta_\ell(x)$ is empty $L_\alpha(x)$ is constant and when $\Theta_u(x)$ is empty $U_\alpha(x)$ is constant. In this case, whether or not the interval is strictly nested therefore depends on whether there exists an $\alpha \in (0, 1)$ such that the other bound is constant in a neighbourhood of α. We will therefore without loss of generality assume that neither $\Theta_\ell(x)$ nor $\Theta_u(x)$ are empty.

Let

$$\alpha_\ell = \liminf_{\theta \to \inf \Theta_1(x)} \lambda(\theta, x) \quad \text{and} \quad \alpha_u = \liminf_{\theta \to \sup \Theta_1(x)} \lambda(\theta, x).$$

Since $\mathcal{A}_{\theta,x} \neq \mathcal{X}$ for $\theta < \inf \Theta_1(x)$, by A1 $\alpha_\ell < 1$, and similarly $\alpha_u < 1$. Thus, a point is added to or removed from $\mathcal{A}_{\theta,x}$ at $\theta = \inf \Theta_1(x)$ and $\theta = \sup \Theta_1(x)$, and by Lemma 1 the p-value function $\lambda(\theta, x)$ must have jumps at $\theta = \inf \Theta_1(x)$ and at $\theta = \sup \Theta_1(x)$. Then for $\alpha_0 = \max(\alpha_\ell, \alpha_u)$, there is an $\alpha_1 \in (\alpha_0, 1)$ for which there exists $\delta > 0$ such that

$$\{\theta \in \Theta : \lambda(\theta, x) \in (\alpha_1 - \delta, \alpha_1 + \delta)\} = \varnothing.$$

Thus both the upper and the lower bound of $I_\alpha(x)$ are constant in a neighbourhood of $\alpha = \alpha_1$, and the interval is not strictly nested.

B.4. An auxiliary lemma

The following auxiliary lemma will be used in the proof of Proposition 4.

Lemma 2. *With $\mathcal{X}$ as in Proposition 4, for any $y, x \in \mathcal{X}$ such that $0 \leq y < x$, let*

$$\theta_{y,x} = \inf\{\theta : T(\theta, x) \leq T(\theta, y)\} \tag{B.3}$$

and define $\theta_{x,x} := \inf \Theta_1(x)$. Under A2$^+$ and B3$^+$,

$$\theta_{0,x} \leq \theta_{1,x} \leq \cdots \leq \theta_{x-1,x} \leq \theta_{x,x}. \tag{B.4}$$

Moreover,

$$\mathcal{A}_{\theta,x} = \{k \in \mathcal{X} : k \geq x\} \quad \text{if and only if} \quad \theta \in (\inf \Theta, \theta_{0,x}), \tag{B.5}$$

and

$$\mathcal{A}_{\theta,x} = \{k \in \mathcal{X} : k \leq y\} \cup \{k \in \mathcal{X} : k \geq x\} \quad \text{if and only if} \quad \theta \in [\theta_{y,x}, \theta_{y+1,x}). \tag{B.6}$$

Proof. First, we establish some facts about $\theta_{y,x}$ and the behaviour of $T(\theta, \cdot)$ for such θ. If $\theta \in \Theta_1(x)$, then it follows from (3) that $T(\theta, x) \leq T(\theta, y)$. Thus, by (B.3) we have $\theta_{y,x} \leq \inf \Theta_1(x)$, so by (B.2), $\theta_{y,x} \in \Theta_\ell(x)$.

With x_θ as defined in B3$^+$(i), $T(\theta, \cdot)$ is increasing at x if $x > x_\theta$. If $\theta \in \Theta_1(x)$ then $x_\theta = x$. By B3$^+$(iii), x_θ is an increasing function of θ. Thus, if $\theta \leq \inf \Theta_1(x)$, i.e. $\theta \in \Theta_\ell(x)$, we have $x > x_\theta$. Since $\theta_{y,x} \in \Theta_\ell(x)$, $T(\theta_{y,x}, \cdot)$ is increasing at x.

It now follows that for any $y < x$, $T(\theta, x) < T(\theta, y)$ can happen only if $T(\theta, \cdot)$ is decreasing at y. Whenever $T(\theta, x) < T(\theta, y)$ and $T(\theta, \cdot)$ is decreasing at y, we have $T(\theta, x) < T(\theta, y) < T(\theta, y - 1)$, and (B.4) follows since $\{\theta : T(\theta, x) \leq T(\theta, y)\} \subset \{\theta : T(\theta, x) \leq T(\theta, y - 1)\}$.

Let $x \geq 1$ be fixed. If $\theta \leq \theta_{0,x}$ then $T(\theta, x) < T(\theta, 0)$ and (B.4) ensures that $T(\theta, x) < T(\theta, y)$ for all other $y < x$ as well. (B.5) now follows from (3).

Next, for some $y < x$, let $\theta \in \Theta_\ell(x)$ be such that $\theta > \theta_{y,x}$. Under B3$^+$(ii) we have $\theta < \theta_x \in \Theta_1(x)$, so $T(\cdot, x)$ is decreasing in $\Theta_\ell(x)$. However, if $T(\theta, x) < T(\theta, y)$ then $y < x_\theta$, so $\theta \geq \sup \Theta_1(y)$. Thus $\theta_{x,y} > \theta_y$, implying that $T(\cdot, y)$ is increasing at $\theta_{y,x}$. Thus $T(\theta, x) < T(\theta, y)$ for all $\theta > \theta_{y,x}$. Equations (B.5) and (B.6) now follow from (3). $\qquad\square$

B.5. Proof of Proposition 4

We start by showing (b) and finish by proving (a). The proof of (c) is analogous to the proof of (b), and is therefore omitted.

(b) We wish to find the largest α_L such that, for all $x \in \mathcal{X}$, $L_\alpha(x)$ is strictly monotone in α when $\alpha < \alpha_L$. For a given x, let $\alpha_L(x)$ be the largest α such that $L_\alpha(x)$ is strictly monotone in α when $\alpha < \alpha_L(x)$. Then $\alpha_L \leq \alpha_L(x)$ for all x, with equality for some x. We therefore show the statement by showing that $\alpha_L(x) = \inf_{\theta \in \{\theta : T(\theta,0) > T(\theta,x)\}} \lambda(\theta, x) > 0$.

As in the proof of Lemma 2, it suffices to study $\theta \in \Theta_\ell(x)$, where $\Theta_\ell(x)$ is defined as in (B.2).

(i) Let $x \geq 1$ be fixed. If $\theta \leq \theta_{0,x}$, defined as in (B.3), then by (B.5), $\mathcal{A}_{\theta,x} = \{k : k \geq x\}$. Thus, by (3), $\lambda(\theta, x) = P_\theta(X \geq x)$. The p-value function $\lambda(\cdot, x)$ is therefore non-negative (by A1), yielding (i).

(ii) $\lambda(\cdot, x)$ is strictly increasing (by A2$^+$) and continuous (by A3). We extend the p-value function by defining $\lambda(\inf \Theta, x) := \lim_{\theta \searrow \inf \Theta} \lambda(\theta, x)$. Then $\lambda(\theta, x)$ is a continuous strictly monotone bijection from the compact set $[\inf \Theta, \theta_{0,x}]$ to the compact set $[\lambda(\inf \Theta, x), \alpha_L(x)]$. It is therefore a homeomorphism, and it follows that its inverse $L_\alpha(x)$ is continuous and strictly monotone in α, which yields (ii).

(iii) From Lemmas 2 and 1, it follows that the first discontinuity in $\lambda(\theta, x)$ occurs at $\theta_{0,x}$. From Definition 4, it follows that there exists an $\varepsilon > 0$ such that $L_\alpha(x)$ is constant in $(\alpha_L(x), \alpha_L(x) + \varepsilon)$ if and only if $\lambda(\theta, x) > \alpha_L(x)$ for all $\theta \in [\theta_{0,x}, \inf \Theta_1(x)]$.

By (B.6) for any $\theta \in [\theta_{0,x}, \inf \Theta_1(x)]$, there exists a $y < x$ such that $\theta \in [\theta_{y,x}, \theta_{y+1,x})$, so that

$$
\begin{aligned}
\lambda(\theta, x) = P_\theta(X \leq y) + P_\theta(X \geq x) &> \varepsilon + P_\theta(X \geq x) \\
&> \varepsilon + P_{\theta_{0,x}}(X \geq x) = \varepsilon + \alpha_L(x) > \alpha_L(x),
\end{aligned}
\tag{B.7}
$$

where the first inequality follows from A1 and the second inequality follows from A2$^+$. (iii) now follows.

(a) For any $x \in \mathcal{X}$, let $A_L(x)$ denote the set of α for which $L_\alpha(x)$ is locally constant in α and $A_U(x)$ denote the set of α for which $U_\alpha(x)$ is locally constant in α. By definition, the largest α for which $I_\alpha(X)$ is strictly nested is

$$
\alpha_{\text{nest}} = \inf \left\{ \alpha : \exists \varepsilon > 0 \text{ for which } (\alpha, \alpha + \varepsilon) \subseteq \bigcup_{x \in \mathcal{X}} \left(A_L(x) \cap A_U(x) \right) \right\}.
$$

Using Proposition 3(c), $A_L(x) \cap A_U(x)$ has a connected uncountable subset for all x, so α_{nest} always exists. By part (b) of Proposition 4, $\alpha_{\text{nest}} \geq \alpha_L > 0$.

B.6. Proof of Proposition 6

By Proposition 3(a) and A3, $\lambda(\theta, x)$ is a piecewise continuous function. By Lemma 1, it is not continuous at the boundaries of the set $\Theta_1(x)$. Hence $\lambda(\theta, x)$ can only be bimonotone if it is monotone whenever it is continuous. Each of its continuous parts can be represented by equation (6) with fixed k_1 and k_2. Such a part can be written as

$$1 - \sum_{k \leq k_1 - 1} P_\theta(X = k) + \sum_{k \leq k_2} P_\theta(X = k). \tag{B.8}$$

By A2, $1 - \sum_{k \leq k_1 - 1} P_\theta(X = k)$ is strictly increasing and $\sum_{k \leq k_2} P_\theta(X = k)$ is strictly decreasing. By condition A4 (B.8) equals 1 at the boundaries of Θ. If it is not constant, it must therefore by condition A5 have a unique minimum in the interior of Θ. Rewriting the expression again, we have

$$(\text{B.8}) = 1 - \sum_{k_2 + 1 \leq k \leq k_1 - 1} P_\theta(X = k),$$

so that the minimum is given by the root θ_r of the equation

$$\frac{d}{d\theta} \sum_{k = k_2 + 1}^{k_1 - 1} P_\theta(X = k) = \sum_{k = k_2 + 1}^{k_1 - 1} \frac{d}{d\theta} P_\theta(X = k) = 0 \tag{B.9}$$

that is in the interior of Θ. Next, we let k_1 and k_2 vary as functions of (θ, x) and use $\theta_r(\theta, x)$ to denote the solution of (B.9) with $k_1 = k_1(\theta, x)$ and $k_2 = k_2(\theta, x)$.

By Proposition 3(a), $\lambda(\theta, x)$ has jumps corresponding to changes in $k_1(\theta, x)$ or $k_2(\theta, x)$. $\lambda(\theta, x)$ fails to be bimonotone if

$$\big(k_1(\theta, x), \ k_2(\theta, x)\big) = \big(k_1\big(\theta_r(\theta, x) + \varepsilon, x\big), \ k_2\big(\theta_r(\theta, x) + \varepsilon, x\big)\big) \qquad \text{for some } \varepsilon > 0,$$

i.e. if it does not jump before the root $\theta_r(\theta, x)$ that corresponds to $(k_1(\theta, x), k_2(\theta, x))$, since for $\lambda'(\theta, x) = \frac{d}{d\theta} \lambda(\theta, x)$,

$$\text{sign}\big(\lambda'\big(\theta_r(\theta, x) - \varepsilon\big)\big) \neq \text{sign}\big(\lambda'\big(\theta_r(\theta, x) + \varepsilon\big)\big).$$

Assume that $\theta \in \Theta_u(x)$. Then $\lambda(\theta, x)$ should be decreasing in θ. If $\theta > \theta_r(\theta, x)$ then (B.8) with $k_1 = k_1(\theta, x)$ and $k_2 = k_2(\theta, x)$ is increasing, so that $\lambda(\theta, x)$ is not bimonotone. If instead we assume that $\theta \in \Theta_\ell(x)$ so that $\lambda(\theta, x)$ is in its increasing part, we similarly get that $\lambda(\theta, x)$ is not bimonotone if $\theta < \theta_r(\theta, x)$. This establishes (a). Part (b) then follows from Proposition 5.

B.7. Proof of Proposition 8

Let $\overline{\mathbb{R}}$ be the extended real line and $\overline{\Theta} \subseteq \overline{\mathbb{R}}$ be the closure of Θ. Let $F(\theta, x) = \mathrm{P}_\theta(X \leq x)$. By A1–A4, $F(\cdot, x)$ is a continuous monotone bijection from $\overline{\Theta}$ to $[0, 1]$ for all $x \in \mathcal{X} \setminus \sup \mathcal{X}$. Since $\overline{\Theta}$ and $[0, 1]$ both are compact, it follows that $F(\cdot, x)$ is a homeomorphism, which ensures that the bounds given by (1) are continuous in α. The monotonicity of $F_\theta(\cdot, x)$ ensures that both $F_\theta^{-1}(\cdot, x)$ and the bounds are monotone, so that the interval is strictly nested.

Finally, by condition A3, the p-value function (2) is continuous in θ when $x \in \mathcal{X}$ is fixed.

Acknowledgments

References

[1] Agresti, A. (2003). Dealing with discreteness: Making "exact" confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Stat. Methods Med. Res*. **12** 3–21. MR1977232

[2] Agresti, A. and Min, Y. (2001). On small-sample confidence intervals for parameters in discrete distributions. *Biometrics* **57** 963–971. MR1863460

[3] Birnbaum, A. (1961). Confidence curves: An omnibus technique for estimation and testing statistical hypotheses. *J. Amer. Statist. Assoc*. **56** 246–249. MR0121904

[4] Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canad. J. Statist*. **28** 783–798. MR1821434

[5] Blyth, C.R. and Still, H.A. (1983). Binomial confidence intervals. *J. Amer. Statist. Assoc*. **78** 108–116. MR0696854

[6] Bol'šev, L.N. (1965). On the construction of confidence limits. *Teor. Verojatnost. i Primenen*. **10** 187–192. MR0210218

[7] Brown, L.D., Cai, T.T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statist. Sci*. **16** 101–133. MR1861069

[8] Byrne, J. and Kabaila, P. (2005). Comparison of Poisson confidence intervals. *Comm. Statist. Theory Methods* **34** 545–556. MR2181200

[9] Casella, G. (1986). Refining binomial confidence intervals. *Canad. J. Statist*. **14** 113–129. MR0849867

[10] Casella, G. and Robert, C. (1989). Refining Poisson confidence intervals. *Canad. J. Statist*. **17** 45–57. MR1014090

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

[12] Cohen, A. and Strawderman, W.E. (1973). Admissibility implications for different criteria in confidence estimation. *Ann. Statist*. **1** 363–366. MR0348875

[13] Crow, E.L. (1956). Confidence intervals for a proportion. *Biometrika* **43** 423–435. MR0093077

[14] Crow, E.L. and Gardner, R.S. (1959). Confidence intervals for the expectation of a Poisson variable. *Biometrika* **46** 441–453. MR0125681

[15] Fay, M.P. (2010). Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics* **11** 373–374.

[16] Fay, M.P. (2010). Two-sided exact tests and matching confidence intervals for discrete data. *R Journal* **2** 53–58.

[17] Fisher, R.A. (1930). Inverse probability. *Proc. Camb. Philos. Soc.* **26** 528–535.

[18] Garwood, F. (1936). Fiducial limits for the Poisson distribution. *Biometrika* **28** 437–442.

[19] Göb, R. and Lurz, K. (2014). Design and analysis of shortest two-sided confidence intervals for a probability under prior information. *Metrika* **77** 389–413. MR3175131

[20] Hirji, K.F. (2006). *Exact Analysis of Discrete Data*. Boca Raton, FL: Chapman & Hall/CRC. MR2193238

[21] Kabaila, P. and Byrne, J. (2001). Exact short Poisson confidence intervals. *Canad. J. Statist.* **29** 99–106. MR1834489

[22] Lecoutre, B. and Poitevineau, J. (2014). New results for computing Blaker's exact confidence interval for one parameter discrete distributions. *Comm. Statist. Simulation Comput.* DOI:10.1080/03610918.2014.911900.

[23] Liese, F. and Miescke, K.-J. (2008). *Statistical Decision Theory*: *Estimation*, *Testing*, *and Selection*. *Springer Series in Statistics*. New York: Springer. MR2421720

[24] Newcombe, R.G. (2011). Measures of location for confidence intervals for proportions. *Comm. Statist. Theory Methods* **40** 1743–1767. MR2781501

[25] Reiczigel, J. (2003). Confidence intervals for the binomial parameter: Some new considerations. *Stat. Med.* **22** 611–621.

[26] Schilling, M.F. and Doi, J.A. (2014). A coverage probability approach to finding an optimal binomial confidence procedure. *Amer. Statist.* **68** 133–145. MR3246552

[27] Sommerville, M.C. and Brown, R.S. (2013). Exact likelihood ratio and score confidence intervals for the binomial proportion. *Pharmaceutical Statistics* **12** 120–128.

[28] Sterne, T.E. (1954). Some remarks on confidence or fiducial limits. *Biometrika* **41** 275–278. MR0062387

[29] Thulin, M. (2014). On split sample and randomized confidence intervals for binomial proportions. *Statist. Probab. Lett.* **92** 65–71. MR3230474

[30] Thulin, M. (2014). Coverage-adjusted confidence intervals for a binomial proportion. *Scand. J. Stat.* **41** 291–300. MR3207171

[31] Thulin, M. (2014). The cost of using exact confidence intervals for a binomial proportion. *Electron. J. Stat.* **8** 817–840. MR3217790

[32] Vos, P.W. and Hudson, S. (2005). Evaluation criteria for discrete confidence intervals: Beyond coverage and length. *Amer. Statist.* **59** 137–142. MR2133560

[33] Vos, P.W. and Hudson, S. (2008). Problems with binomial two-sided tests and the associated confidence intervals. *Aust. N. Z. J. Stat.* **50** 81–89. MR2414657

[34] Wang, W. (2006). Smallest confidence intervals for one binomial proportion. *J. Statist. Plann. Inference* **136** 4293–4306. MR2323417

[35] Wang, W. (2014). Exact optimal confidence intervals for hypergeometric parameters. *J. Amer. Statist. Assoc.* To appear. DOI:10.1080/01621459.2014.966191.

[36] Wang, Y.H. (2000). Fiducial intervals: What are they? *Amer. Statist.* **54** 105–111. MR1803120

[37] Xie, M.-g. and Singh, K. (2013). Confidence distribution, the frequentist distribution estimator of a parameter: A review. *Int. Stat. Rev.* **81** 3–39. MR3047496