Statistica Sinica Preprint No: SS-2021-0322	
Title	On Construction of Optimal Exact Confidence Intervals
Manuscript ID	SS-2021-0322
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0322
Complete List of Authors	Weizhen Wang
Corresponding Authors	Weizhen Wang
E-mails	weizhen.wang@wright.edu
Notice: Accepted version subject to English editing.	

Statistica Sinica

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

ON CONSTRUCTION OF OPTIMAL EXACT CONFIDENCE INTERVALS

Weizhen Wang

Wright State University

Tianjin University of Finance and Economics

Abstract: For a given confidence interval the central value is more likely equal to the parameter than a boundary one. However, when considering two null hypotheses whose hypothesized values are equal to these two values, none of the hypotheses should be rejected because both values are inside the interval. More broadly, is there a way to differentiate any two values in the interval? A general method, named the h-function method, is introduced to do so. It also improves any confidence interval as follows: (i) an approximate interval, including a point estimator, is modified to be exact; (ii) an exact interval is refined to be an exact one that is a subset of the previous interval. Three real datasets, including Johnson & Johnson's Janssen vaccine (2021), are used to illustrate the method. Simulation results are given in the Supplementary Materials.

Key words and phrases: Admissible confidence interval; Difference of two proportions; Infimum coverage probability; p-value; Vaccine efficacy.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

1. Introduction

Many $1-\alpha$ confidence intervals are approximate and their confidence coefficients may be much smaller than the nominal level $1-\alpha$. This results in unreliable inferences as shown in the well-known Wald interval for a proportion (Brown, Cai and DasGupta, 2001). How to improve such intervals for reliable inferences? When the underlying distributions are discrete, an exact two-sided interval is often conservative especially when it is equal to the intersection of two one-sided $1-\frac{\alpha}{2}$ intervals (Agresti, 2013). How to make an exact interval uniformly shorter without lowering $1-\alpha$? These two general questions, which are important in many fields, including clinical trials, motivate the current paper.

There is a one-to-one mapping between a family of tests and a confidence set. Let Θ be the range of a parameter of interest θ and let S be a sample space. For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level-α test of $H_0 : \theta = \theta_0$. Then

$$C(\underline{x}) = \{\theta_0 \in \Theta : \underline{x} \in A(\theta_0)\}. \tag{1.1}$$

is a $1-\alpha$ confidence set for θ. Conversely, let $C(\underline{x})$ be a $1-\alpha$ confidence set. Then

$$A(\theta_0) = \{\underline{x} \in S : \theta_0 \in C(\underline{x})\} \tag{1.2}$$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

is the acceptance region of level-α for H_0. A confidence set (or interval) is typically derived from the tests, but solving $C(\underline{x})$ from $A(\theta_0)$ as in (1.1) is complicated. A minor goal of this paper is to simplify this process.

A key feature of a confidence interval is that its confidence coefficient, defined as the infimum coverage probability over the entire parameter space (Casella and Berger, 2002), should be no smaller than the nominal level $1-\alpha$. To avoid ambiguity in discussion, a $1-\alpha$ exact confidence interval means that it has a confidence coefficient at least $1-\alpha$, i.e., the $1-\alpha$ interval by Casella and Berger (2002); while a $1-\alpha$ (approximate) confidence interval only means that the nominal level is set to be $1-\alpha$ but its confidence coefficient can be any number in $[0,1]$.

Ideally, a $1-\alpha$ approximate interval would have a confidence coefficient close to $1-\alpha$, which may not happen in practice even for a large sample size. For example, let $X \sim Bino(n,p)$ be a binomial with n trials and a success probability p, the well-known Wald internal for p,

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} \ \text{ for } \ \hat{p} = X/n \tag{1.3}$$

is used as a 95% interval. However, it is a zero exact interval since it has a zero confidence coefficient. In fact, a $1-\alpha$ Wald interval always has a zero confidence coefficient for any sample size n and any α in $[0,1]$ (Brown, Cai and DasGupta, 2001; Agresti, 2013). Also, a point estimator if used as

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

a confidence interval has a confidence coefficient zero but can be modified to be an exact interval as shown later. Therefore, it is safe to assume the confidence coefficient of a $1-\alpha$ approximate interval has a range of [0,1].

The requirement on confidence coefficient no smaller than $1-\alpha$ is often violated by an approximate interval. Thus, the major concern is whether its inferential conclusion is reliable as its confidence coefficient is seldom reported but can be much smaller than $1-\alpha$. For instance, Huwang (1995) proved this fact for Wilson interval (1927) for large samples. On the other hand, an approximate interval is easy to access. It is of great interest to build an exact interval through a given approximate interval – the first major goal of the paper. To the best of our knowledge, limited research has been done on this issue.

One common way to obtain a $1-\alpha$ exact two-sided interval is to take the intersection of two exact one-sided $1-\frac{\alpha}{2}$ intervals, for example, the conservative Clopper-Pearson interval (1934). It is also of great interest to shrink a given $1-\alpha$ exact interval to an optimal one – the second major goal of the paper. This can be easily applied to contingency tables, including but not limited to the establishment of a new treatment. Casella (1986), Wang (2014) and Casella and Roberts (1989) refined exact intervals for a proportion or a Poisson mean. However, their methods are valid only for a

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

single-parameter distribution family.

We address the above three problems by introducing an h-function that not only relates to the p-value for test construction but also is a function over the parameter of interest for interval construction. This idea was used by Blaker (2000) and Agresti and Min (2001) to derive exact intervals in some special cases. It now has a new use to modify and/or refine any interval (including the intervals in Blaker (2000) and Agresti and Min (2001)) – a solution to the challenging problem of improving any interval when nuisance parameters exist and the sample space is discrete. The main idea is to distinguish any two values outside (or inside) the interval using the function T_2 in (2.7). More precisely, for an approximate interval those parameter values that are outside but close to the interval are likely added to the interval; for an exact but conservative interval its boundary values are to be removed. In consequence, the resultant interval becomes either exact or shorter. The following example helps to understand the two major goals.

Example 1. Consider a two-arm randomized trial in Essenberg (1952) for testing the effect of tobacco smoking on tumor development in mice. In the smoking group tumors were observed on $21(= x)$ mice out of $23(= n_1)$ mice; in the control group $(y, n_2) = (19, 32)$. Here X and Y are two independent binomials with two tumor rates p_1 and p_2, respectively. The difference, $d =$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$p_1 - p_2$, is used to evaluate the smoking effect. As in Table 4, d is estimated by the 95% Wald type interval, the maximum likelihood estimator, the exact score-test interval (Agresti and Min, 2001), or the exact two-one-sided interval (Wang, 2010). Both approximate intervals have a zero confidence coefficient. How to improve them to have a confidence coefficient at least 0.95? For the two exact intervals, how to make them shorter while still keeping the confidence coefficient at least 0.95? $\square$

The paper is organized as follows: In Section 2, the h-function method is formally introduced to construct $1-\alpha$ exact optimal confidence intervals. Section 3 discusses three applications of the proposed method. Section 4 modifies any one-sided interval to the smallest interval. Discussions and the proofs are given in Section 5 and the Appendix, respectively.

2. A theory of deriving optimal two-sided intervals

Suppose $\underline{X}$ is observed from a distribution with joint cumulative distribution function $F_{(\theta,\underline{\eta})}(\underline{x})$ specified by a parameter vector $(\theta, \underline{\eta})$ in a parameter space H. Here θ is the parameter of interest and $\underline{\eta}$ is the nuisance parameter vector. The null hypothesis H_0 is one of the three forms: $\theta = \theta_0$, $\theta \leq \theta_0$ and $\theta \geq \theta_0$, for a fixed value θ_0, each corresponding to two-sided, lower and upper one-sided intervals, respectively. We introduce the h-function method

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

and show an important usage for constructing optimal exact intervals in the next three subsections.

2.1 The h-function method

A p-value $p(\underline{X})$ is valid for H_0 if, for every $0 \leq \alpha \leq 1$, $\sup_{\theta \in H_0} P_{(\theta,\underline{\eta})}(p(\underline{X}) \leq \alpha) \leq \alpha$ (Casella and Berger, 2002). For simplicity, we drop the subscript $(\theta, \underline{\eta})$ in the future discussion. The p-value at $\underline{x}$ can be defined by a given test statistic $T(\underline{X})$ using $p(\underline{x}) = \sup_{\theta \in H_0} P(T(\underline{X}) \leq T(\underline{x}))$ if a small value of $T(\underline{X})$ supports H_A. An example of $T(\underline{X})$ is the likelihood ratio test statistic. The p-value $p(\underline{X})$ indeed depends on both $\underline{X}$ and θ_0, so, is rewritten as

$$h(\underline{X}, \theta_0) = p(\underline{X}). \tag{2.4}$$

The left hand side is called the h-function, which is a function of both $\underline{X}$ and θ_0; while $p(\underline{X})$ is only a function of $\underline{X}$. Using $h(\underline{X}, \theta_0)$, the exact level-$\alpha$ acceptance region for H_0 and $1-\alpha$ exact confidence set for θ are given by

$$A(\theta_0) = \{\underline{x} : h(\underline{x}, \theta_0) > \alpha\} \text{ and } C(\underline{x}) = \{\theta_0 : h(\underline{x}, \theta_0) > \alpha\}, \tag{2.5}$$

respectively. Both are obtained by solving the same inequality $h(\underline{x}, \theta_0) > \alpha$ but in terms of two different arguments $\underline{x}$ and θ_0. Hence, the constructions of test and confidence set are unified. They are simpler than the approaches in (1.1) or (1.2) because of the intermediary h-function in (2.4). We name

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

it the h-function method. Blaker (2000) used a special h-function to derive confidence intervals in some discrete distributions of one parameter. This method is now applied to improve any given interval in a general case with nuisance parameters.

Set $C(\underline{x})$ may not be an interval. Let $\overline{A}$ denote the smallest simply connected set containing set A. Thus, $\overline{C(\underline{x})}$ is always an interval and its infimum coverage probability over H is not smaller than $1-\alpha$. Casella and Berger (2002, p. 431) gave a concrete example to show a difference between $C(\underline{x})$ and $\overline{C(\underline{x})}$. Throughout the paper, we use $C(\underline{X})$ to denote $\overline{C(\underline{X})}$ and $ICP(C)$ to denote the confidence coefficient of $C(\underline{X})$.

In general, a test statistic may also depend on θ_0 and thus has a form of $T(\underline{X}, \theta_0)$. Let $K(\underline{x}, \theta_0) = \{\underline{y} : T(\underline{y}, \theta_0) \le T(\underline{x}, \theta_0)\}$. Then

$$h(\underline{x}, \theta_0) = \sup_{(\theta,\underline{\eta})\in H_0} P(K(\underline{x}, \theta_0)) = \begin{cases} \sup_{(\theta,\underline{\eta})\in H_0} \sum_{\underline{y}\in K(\underline{x},\theta_0)} f_{(\theta,\underline{\eta})}(\underline{y}) \\ \sup_{(\theta,\underline{\eta})\in H_0} \int_{K(\underline{x},\theta_0)} f_{(\theta,\underline{\eta})}(\underline{y}) d\underline{y}, \end{cases} \tag{2.6}$$

where $f_{(\theta,\underline{\eta})}$ is either the joint probability mass function or probability density function of $\underline{X}$ and the probability $P(K(\underline{x}, \theta_0))$ is a function of the nuisance parameter vector $\underline{\eta}$.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

2.2 Modifying a given two-sided confidence interval

For convenience, consider the closed interval $C_0(\underline{X}) = [L_0(\underline{X}), U_0(\underline{X})]$ for θ, which is to be improved by the h-function method. Consider hypotheses $H_0 : \theta = \theta_0$ vs $H_A : \theta \neq \theta_0$ for a given θ_0. Introduce a test statistic

$$T_2(\underline{X}, \theta_0) = T(L_0(\underline{X}), U_0(\underline{X}), \theta_0) \tag{2.7}$$

for some function $T(l, u, \theta_0)$, where the subscript '2' means 'two-sided'. They may satisfy some or all of the following three conditions:

(a) a small value of $T_2(\underline{X}, \theta_0)$ is in favor of H_A;

(b) $T(l, u, \theta_0) \geq 0$ if and only if $\theta_0 \in [l, u]$;

(c) For fixed $l_1 \leq l_2 \leq u_2 \leq u_1$, $T(l_2, u_2, \theta_0) \leq T(l_1, u_1, \theta_0)$ for any θ_0.

Here are three choices of T_2 that satisfy the three conditions. The first $T_2^D(\underline{X}, \theta_0)$ uses $T(l, u, \theta_0) = \min\{\theta_0 - l, u - \theta_0\}$. When the range of θ is nonnegative, define $0/0 = 1$. The second $T_2^R(\underline{X}, \theta_0)$ has $T(l, u, \theta_0) = \min\{\frac{\theta_0}{l}, \frac{u}{\theta_0}\} - 1$. The third $T_2^I(\underline{X}, \theta_0)$ has $T(l, u, \theta_0) = I_{[l,u]}(\theta_0) - 1$ using the indicator function for interval $[l, u]$.

The h-function based on $T_2(\underline{x}, \theta_0)$ is

$$h_2(\underline{x}, \theta_0) = \sup_{H_0} P(T_2(\underline{X}, \theta_0) \leq T_2(\underline{x}, \theta_0)). \tag{2.8}$$

Following (2.5), the level-α acceptance region and $1 - \alpha$ exact confidence

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

interval are

$$A_2(\theta_0) = \{\underline{x} : h_2(\underline{x}, \theta_0) > \alpha\}, C_0^M(\underline{x}) = \overline{\{\theta_0 : h_2(\underline{x}, \theta_0) > \alpha\}}, \tag{2.9}$$

respectively. In the rest of the paper, let $\overline{A}$ denote the smallest closed simply connected set that contains set A. The superscript "M" refers a modification. For a nonnegative integer k, $C_0^{Mk}(\underline{x})$ is the resultant interval when the modification process of (2.7), (2.8) and (2.9) is applied to $C_0(\underline{X})$ for k consecutive times. For example, $C_0^{M(k+1)}(\underline{x}) = (C_0^{Mk})^M(\underline{x})$ for any $k \geq 0$. The following theorems discuss the properties of $C_0^M(\underline{X})$ and its variant. First, what is the confidence coefficient of $C_0^M(\underline{X})$?

Theorem 1. *Suppose T_2 in (2.7) satisfies Condition (a). For a given interval $C_0(\underline{X})$,*

(i) the h-function $h_2(\underline{X}, \theta_0)$ in (2.8) is a valid p-value for test statistic $T_2(\underline{X}, \theta_0)$.

(ii) Interval $C_0^M(\underline{X})$ given in (2.9) is a $1-\alpha$ exact interval. i.e., $ICP(C_0^M) \geq 1-\alpha$.

The theorem modifies interval $C_0(\underline{X})$ of any level, including a point estimator, to an exact interval $C_0^M(\underline{X})$. Also, unlike either reject or accept H_0 by checking whether $C_0^M(\underline{X})$ includes θ_0, a p-value can be calculated by $h_2(\underline{x}, \theta_0)$ – a new usage of confidence interval.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

2.3 Refining a $1-\alpha$ exact two-sided confidence interval

When $C_0(\underline{X})$ is exact, the modified interval $C_0^M(\underline{X})$ is also exact duo to Theorem 1. However, what is the relationship between $C_0(\underline{X})$ and $C_0^M(\underline{X})$?

Theorem 2. *Suppose T_2 in (2.7) satisfies Conditions (a) and (b).*

(i) If $C_0(\underline{X})$ is a $1-\alpha$ exact interval, then, $C_0^M(\underline{x})$ is a subset of $C_0(\underline{X})$. In particular, if $T_2 = T_2^I$, then $C^M(\underline{X}) = C_0(\underline{X})$. i.e., one should not use T_2^I to improve an exact interval.

(ii) For any interval $C_0(\underline{X})$, $C_0^{M2}(\underline{X})$ is a subset of $C_0^M(\underline{X})$.

For a fixed θ_0, the p-value, $h(\underline{X}, \theta_0)$, differentiates two sample points $\underline{x}_1$ and $\underline{x}_2$ regarding $H_0 : \theta = \theta_0$. If $h(\underline{x}_1, \theta_0) > h(\underline{x}_2, \theta_0) > \alpha$, then both sample points fail to reject H_0 but $\underline{x}_1$ is more supportive of H_0 by its larger p-value. However, if using an acceptance region of level α, both points belong to the region and we cannot tell which point supports H_0 more. The h-function in (2.8) plays a similar role to differentiate two parameter values θ_1 and θ_2 through the test statistic T_2 in (2.7). For an observed $\underline{x}$, traditionally, one tests on $H_0 : \theta = \theta_1$ (or θ_2) by checking whether θ_1 (or θ_2) belongs to $C_0(\underline{x})$. If both of them belong to $C_0(\underline{x})$, one fails to reject $\theta = \theta_1$ and $\theta = \theta_2$ but cannot tell which statement is more likely to be true. We now quantify this by introducing $T_2(\underline{x}, \theta_0)$ as a function of θ_0. i.e., $\theta = \theta_1$ is more likely if

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$T_2(\underline{x}, \theta_1) > T_2(\underline{x}, \theta_2)$. Thus, T_2^D or T_2^R but not T_2^I (which is a constant over $C_0(\underline{x})$) should be used to shrink an exact interval as in Theorem 2.

The modification process can be applied to an exact interval multiple times, and each time a subset-interval is generated. What is the smallest interval out of the process?

Theorem 3. *Suppose T_2 in (2.7) satisfies Conditions (a) and (b). For an exact interval $C_0(\underline{X})$ and a sample point $\underline{x}$, let $C_0^{M\infty}(\underline{x}) = \cap_{k=0}^{+\infty} C_0^{Mk}(\underline{x})$. Then,*

(i) interval $C_0^{Mk}(\underline{x})$, as a set of θ, is nonincreasing in k for $k \geq 0$.

(ii) $C_0^{M\infty}(\underline{X})$, contained in $C_0^{Mk}(\underline{X})$ for any k, is a $1-\alpha$ exact interval.

(iii) If $C_0^{Mk}(\underline{X}) = C_0^{M(k+1)}(\underline{X})$ for some $k \geq 0$, then $C_0^{M\infty}(\underline{X}) = C_0^{Mk}(\underline{X})$.

One concern for deriving $C_0^{M\infty}(\underline{X}) = C_0^{Mk}(\underline{X})$ is a possibly large k. In Theorem 3, the constant $k = k_{\underline{X}}$ is independent of the sample points. Next, we state a fact that this k indeed depends on the sample point $\underline{x}$, i.e., $k = k(\underline{x})$ and $k_{\underline{X}} = \sup_{\{\text{all } \underline{x}\}} k(\underline{x})$. This makes the computation of $C_0^{M\infty}(\underline{X})$ at $\underline{X} = \underline{x}$ simpler because of $k(\underline{x}) \leq k_{\underline{X}}$.

Theorem 4. *Suppose T_2 in (2.7) satisfies Conditions (a), (b) and (c). For an exact interval $C_0(\underline{X})$ and a fixed sample point $\underline{x}$, if $C_0^{Mk}(\underline{x}) =$*

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$C_0^{M(k+1)}(\underline{x})$ *for some* $k = k(\underline{x}) \geq 0$, *then* $C_0^{M(k+2)}(\underline{x}) = C_0^{M(k+1)}(\underline{x})$. *Therefore,* $C_0^{M\infty}(\underline{x}) = C_0^{Mk}(\underline{x})$.

Can $C_0^{M\infty}(\underline{X})$ be shortened anymore? A sufficient condition for an admissible $C_0^{M\infty}(\underline{X})$ is stated below. A $1-\alpha$ exact interval $C(\underline{X})$ is admissible if any interval $C'(\underline{X})$, which is a subset of but not equal to $C(\underline{X})$, has a confidence coefficient strictly less than $1-\alpha$.

Theorem 5. *Let* $\underline{X}$ *be a random observation on a finite sample space* S *and let* T_2 *in (2.7) be* T_2^D *or* T_2^R. *For an interval* $C_0(\underline{X})$, *if the confidence limits* $L_0^{M\infty}(\underline{X})$ *and* $U_0^{M\infty}(\underline{X})$ *of interval* $C_0^{M\infty}(\underline{X})$ *are both one-to-one functions, then* $C_0^{M\infty}(\underline{X})$ *is admissible.*

When $L_0^{M\infty}$ and $U_0^{M\infty}$ are not one-to-one functions, Theorem 5 indicates that an improvement upon $C_0^{M\infty}$ may only occur at those $\underline{x}$'s at which $L_0^{M\infty}$ or $U_0^{M\infty}$ are tied. The modification process is still helpful to derive admissible intervals as shown in Section 3.2.

3. Applications of improving a given two-sided interval

In this section, we focus on the choice of $T_2 = T_2^D$. Three parameters are to be estimated: (i) a proportion p based on a binomial $X \sim Bino(n, p)$; (ii) the difference of two proportions $d = p_1 - p_2$ based on two independent

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

binomials; (iii) the difference of two proportions d_m based on a match-paired multinomial observation. These parameters are widely used in practice but the best intervals have not been recognized yet.

3.1 Estimating a proportion

Consider an interval $C_p(x) = [L_p(x), U_p(x)]$ that satisfies

$$U_p(x) = 1 - L_p(n - x), \ \ \forall x \in [0, n]. \tag{3.10}$$

The modification process in (2.7), (2.8) and (2.9) is applied to each of the six intervals C_{pi} repeatedly to generate the modified intervals C_{pi}^M and $C_{pi}^{M\infty}$ for $i = 1, ..., 6$: (1) Wald interval C_{p1} (approximate, given in (1.3)); (2) Wilson interval (1927) C_{p2} (approximate, Agresti, 2013, p. 14); (3) the maximum likelihood estimator $C_{p3}(X) = \frac{X}{n}$ (approximate); (4) Clopper-Pearson interval (1934) C_{p4} (exact, Agresti, 2013, p. 603); (5) Blaker interval (2000) C_{p5} (exact, Agresti, 2013, p. 605); (6) Wang interval (2014) C_{p6} (exact).

Intervals C_{p1} and C_{p2} are derived using the asymptotic normality and C_{p3} is just a point estimator. Their confidence coefficients are much lower than $1 - \alpha$. Intervals C_{p4} and C_{p5} themselves are generated by the h-function method (2.5) using an h-function $h_{p4}(x, p_0) = \min\{2 \min\{P_{p_0}(X \leq x), P_{p_0}(X \geq x)\}, 1\}$ and a test statistic $T_{p5}(x, p_0) = \min\{P_{p_0}(X \leq x), P_{p_0}(X \geq x)\}$, respectively. Interval C_{p6} is derived from a refining algorithm over

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

C_{p4} and is admissible. In fact, C_{p5} and C_{p6} are subsets of C_{p4}.

Table 1 contains these intervals over all sample points, their confidence coefficients and total interval lengths for $n = 16$. The confidence coefficients of the three approximate intervals are equal to 0, 0.8362 and 0, respectively, where the second value is given by Huwang (1995).

Following Theorem 1, the modified intervals C_{pi}^{M} and $C_{pi}^{M\infty}$ are all exact. For $i = 4, 5, 6$, C_{pi}^{M} is a subset of C_{pi} due to Theorem 2. The reduction in interval length of C_{p4}^{M} over C_{p4} is noticeable; $C_{p5}^{M}(x)$ shrinks $C_{p5}(x)$ at $x = 6, 8, 10$; no improvement over C_{p6} can be found since C_{p6} is proved to be admissible for any n and α in Wang (2014). In this sense, C_{p6} is the best among these exact intervals.

We also report C_{pi}^{MS}, the simulated version of C_{pi}^{M}, for $i = 1, ..., 6$ in the Supplementary Materials to confirm theoretical results, including Theorems 1 and 2. As expected, $C_{pi}^{MS} \approx C_{pi}^{M}$ for all i's.

The final refined intervals $C_{pi}^{M\infty}$, except $C_{p1}^{M\infty}$, are all admissible following Theorem 5 since their confidence limits have no ties, and $C_{pi}^{M\infty} = C_{pi}^{Mk}$ for a small k (=1 or 0). For a given $k \geq 1$, the ratio of the two total interval lengths of C_{pi}^{Mk+1} and C_{pi}^{Mk} is not larger than one. If it is equal to one, then $C_{pi}^{M\infty} = C_{pi}^{Mk}$ by Theorem 3. The ratio is accurate up to the seventh decimal place. The five admissible intervals are different but have

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 1: The lower confidence limits of: (1) 95% Wald interval C_{p1}, the modification C_{p1}^{M}, the 22nd modification $C_{p1}^{M22}(=C_{p1}^{M\infty})$; (2) 95% Wilson interval C_{p2}, $C_{p2}^{M}(=C_{p2}^{M\infty})$; (3) the sample-proportion estimator C_{p3}, $C_{p3}^{M}(=C_{p3}^{M\infty})$; (4) Clopper-Pearson interval C_{p4}, $C_{p4}^{M}(=C_{p4}^{M\infty})$; (5) Blaker interval C_{p5}, $C_{p5}^{M}(=C_{p5}^{M\infty})$; (6) Wang interval $C_{p6}(=C_{p6}^{M\infty})$; and the infimum coverage probability (ICP) and total interval length (TIL) when $n=16$. The upper limits are given by (3.10).

X	0	1	2	3	4	5	6	7	8	ICP
C_{p1}	0.0000	-0.0562	-0.0371	-0.0038	0.0378	0.0853	0.1377	0.1944	0.2550	0
C_{p1}^{M}	0.0000	0.0000	0.0000	0.0000	0.0189	0.0426	0.0688	0.0972	0.1275	0.9500
C_{p1}^{M22}	0.0000	0.0000	0.0000	0.0000	0.0902	0.1321	0.1708	0.1708	0.1708	0.9500
C_{p2}	0.0000	0.0111	0.0349	0.0659	0.1018	0.1416	0.1848	0.2309	0.2799	0.8362
C_{p2}^{M}	0.0000	0.0032	0.0226	0.0531	0.0902	0.1321	0.1777	0.2122	0.2719	0.9500
C_{p3}	0.0000	0.0625	0.1250	0.1875	0.2500	0.3125	0.3750	0.4375	0.5000	0
C_{p3}^{M}	0.0000	0.0032	0.0226	0.0531	0.0902	0.1321	0.1777	0.2187	0.2719	0.9500
C_{p4}	0.0000	0.0015	0.0155	0.0404	0.0726	0.1101	0.1519	0.1975	0.2465	0.9578
C_{p4}^{M}	0.0000	0.0032	0.0226	0.0531	0.0902	0.1321	0.1777	0.2017	0.2719	0.9500
C_{p5}	0.0000	0.0032	0.0226	0.0531	0.0902	0.1321	0.1746	0.2011	0.2717	0.9500
C_{p5}^{M}	0.0000	0.0032	0.0226	0.0531	0.0902	0.1321	0.1777	0.2011	0.2719	0.9500
C_{p6}	0.0000	0.0032	0.0226	0.0531	0.0902	0.1321	0.1777	0.2059	0.2719	0.9500
X	9	10	11	12	13	14	15	16		TIL
C_{p1}	0.3194	0.3877	0.4603	0.5378	0.6212	0.7129	0.8188	1.0000		6.0559
C_{p1}^{M}	0.1597	0.3374	0.4195	0.5000	0.5705	0.6478	0.7326	0.8291		7.8957
C_{p1}^{M22}	0.1708	0.3521	0.4294	0.5000	0.5705	0.6478	0.7326	0.8291		7.0650
C_{p2}	0.3317	0.3864	0.4440	0.5050	0.5699	0.6397	0.7167	0.8063		6.0974
C_{p2}^{M}	0.3075	0.3733	0.4370	0.5000	0.5629	0.6266	0.6924	0.7877		6.4978
C_{p3}	0.5625	0.6250	0.6875	0.7500	0.8125	0.8750	0.9375	1.0000		0
C_{p3}^{M}	0.3125	0.3750	0.4375	0.5000	0.5625	0.6250	0.6875	0.7812		6.4978
C_{p4}	0.2987	0.3543	0.4133	0.4762	0.5435	0.6165	0.6976	0.7940		6.9380
C_{p4}^{M}	0.3005	0.3689	0.4349	0.5000	0.5650	0.6310	0.6994	0.7982		6.4978
C_{p5}	0.3004	0.3682	0.4344	0.5000	0.5655	0.6317	0.6995	0.7988		6.5043
C_{p5}^{M}	0.3004	0.3682	0.4344	0.5000	0.5655	0.6317	0.6995	0.7988		6.4978
C_{p6}	0.3023	0.3834	0.4415	0.5000	0.5584	0.6165	0.6976	0.7940		6.4978

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

the same total interval length 6.4978. Four of them are generated by the newly proposed modification process. These suggest that the best interval for p may not exist. It is worth mentioning that a point estimator for p can be modified to an admissible interval without using its standard error.

Table 2 reports the confidence coefficient and total interval length of the intervals in Table 1 for two other values of n. The two quantities measure the reliability and precision of the interval, respectively. Among a group of $1-\alpha$ exact intervals, the one with the smallest total interval length is preferred. This criterion is also applied in Tables 3 and 5. Here, the confidence coefficients of exact intervals should not be smaller than 0.95. To confirm these numerically, the confidence coefficient of an interval $C(X)$ for p with a nondecreasing lower confidence limit $L(X)$ is achieved at one of the values $(L(x))^-$ for $x = 1, ..., n$, where a^- denotes the left limit of y when y approaches a, see Wang (2007). The confirmation is necessary to prevent potential errors in numerical calculation. The modification process is able to generate admissible intervals, $C_{pi}^{M\infty}$ for $i = 2, ..., 6$, following Theorem 5. Intervals $C_{p4}^{M\infty}$, $C_{p5}^{M\infty}$ and C_{p6} have the smallest total interval length for each n, and the last one, already admissible, does not need to be modified.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 2: The infimum coverage probability (ICP) and total interval length (TIL) for confidence intervals for p: C_{pi}, C_{pi}^M, and $C_{pi}^{M\infty}(=C_{pi}^{Mk})$ for $i=1$ (Wald), 2 (Wilson), 3 (the sample proportion), 4 (Clopper-Pearson), 5 (Blaker), 6 (Wang), when $1-\alpha=0.95$ and n varies. The smallest TIL for each n is marked by * and an admissible interval is marked by †.

n	C_p	ICP	TIL	C_p^M	ICP	TIL	$C_p^{M\infty}$	ICP	TIL
30	C_{p1}	0	8.3772	C_{p1}^M	0.9500	10.0999	C_{p1}^{M21}	0.9500	9.4420
	C_{p2}	0.8371	8.3933	C_{p2}^M	0.9500	8.7975	$^\dagger C_{p2}^{M13}$	0.9500	8.7960
	C_{p3}	0	0	C_{p3}^M	0.9500	8.8279	$^\dagger C_{p3}^{M8}$	0.9500	8.8278
	C_{p4}	0.9505	9.2705	C_{p1}^M	0.9500	8.7784	$^\dagger C_{p4}^{M16}$	0.9500	8.7726*
	C_{p5}	0.9500	8.7814	C_{p5}^M	0.9500	8.7770	$^\dagger C_{p5}^{M16}$	0.9500	8.7726*
	$^\dagger C_{p6}$	0.9500	8.7726*		$C_{p6}^M=C_{p6}$			$C_{p6}^{M\infty}=C_{p6}$	
100	C_{p1}	0	15.3772	C_{p1}^M	0.9500	16.7763	C_{p1}^{M20}	0.9500	16.4196
	C_{p2}	0.8379	15.3803	C_{p2}^M	0.9500	15.8488	$^\dagger C_{p2}^{M13}$	0.9500	15.8465
	C_{p3}	0	0	C_{p3}^M	0.9500	15.8648	$^\dagger C_{p3}^{M12}$	0.9500	15.8637
	C_{p4}	0.9503	16.3057	C_{p1}^M	0.9500	15.8214	$^\dagger C_{p4}^{M15}$	0.9500	15.8146*
	C_{p5}	0.9500	15.8243	C_{p2}^M	0.9500	15.8176	$^\dagger C_{p5}^{M14}$	0.9500	15.8146*
	$^\dagger C_{p6}$	0.9500	15.8146*		$C_{p6}^M=C_{p6}$			$C_{p6}^{M\infty}=C_{p6}$	

3.2 Intervals for the difference of two independent proportions

The difference $d=p_1-p_2$ is often used for comparison of two proportions based on two independent binomials $X\sim Bino(n_1,p_1)$ and $Y\sim Bino(n_2,p_2)$. Consider $H_0: d=d_0$ vs $H_A: d\neq d_0$ for a fixed $d_0\in[-1,1]$. Under H_0, $p_1=d_0+p_2$ for $p_2\in D(d_0)$, where $D(d_0)=[0,1-d_0]$ if $d_0\in[0,1]$ and $D(d_0)=[-d_0,1]$ if $d_0\in[-1,0)$. Suppose $T_d(x,y,d_0)$ is a test statistic

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

satisfying

$$T_d(x, y, d_0) = T_d(n_1 - x, n_2 - y, -d_0), \ \forall \ (x, y) \in S_d = [0, n_1] \times [0, n_2], \tag{3.11}$$

and a small value of $T_d(x, y, d_0)$ supports H_A. Its h-function is

$$h_d(x, y, d_0) = \sup_{p_2 \in D(d_0)} \sum_{\{(u,v) \in S_d : T_d(u,v,d_0) \leq T_d(x,y,d_0)\}} p_B(u, n_1, p_2 + d_0) p_B(v, n_2, p_2), \tag{3.12}$$

where $p_B(x, n, p)$ is the probability mass function of $Bino(n, p)$. The acceptance region of level-α for H_0 and $1 - \alpha$ exact confidence interval for d are

$$A_d(d_0) = \{(x, y) : h_d(x, y, d_0) > \alpha\} \ \text{and} \ C_d(x, y) = \overline{\{d_0 : h_d(x, y, d_0) > \alpha\}}. \tag{3.13}$$

The fact below simplifies interval calculation by half.

Proposition 1. *For a test statistic T_d satisfying (3.11), we have*

$$U_d(x, y) = -L_d(n_1 - x, n_2 - y), \ \forall \ (x, y) \in S_d. \tag{3.14}$$

Four exact and approximate intervals C_{di} satisfying (3.14) are described below for $i = 1, ..., 4$. For each $C_{di}(x, y) = [L_{di}(x, y), U_{di}(x, y)]$, introduce $T_{di}(x, y, d_0) = T_2^D(L_{di}(x, y), U_{di}(x, y), d_0)$. The modification process of (3.11), (3.12) and (3.13) is applied to T_{di} to produce C_{di}^M and $C_{di}^{M\infty}$.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Theorems 1 and 3 assure that the modified intervals are exact and $C_{di}^{M\infty}$ is a subset of C_{di}^{M}. All intervals satisfy (3.14).

First, consider the score test statistic

$$T_{d1}^{*}(x,y,d_0) = \frac{-|\hat{p}_1 - \hat{p}_2 - d_0|}{\sqrt{\frac{\hat{p}_{1d}(x,y,d_0)(1-\hat{p}_{1d}(x,y,d_0))}{n_1} + \frac{\hat{p}_{2d}(d_0)(1-\hat{p}_{2d}(x,y,d_0))}{n_2}}},$$

where $\hat{p}_1 = x/n_1$, $\hat{p}_2 = x/n_2$, $\hat{p}_{2d}(x,y,d_0) = arg\max_{p_2 \in D(d_0)} p_B(x,n_1,p_2 + d_0)p_B(y,n_2,p_2)$, and $\hat{p}_{1d}(x,y,d_0) = \hat{p}_{2d}(x,y,d_0) + d_0$. When $(x,y,d_0) = (n_1,0,1)$ or $(0,n_2,-1)$, the above ratio is $0/0$, so it is defined to be 0. Its h-function h_{d1}^{*} follows (3.12) and generates an exact interval C_{d1}, which is recommended by Agresti and Min (2001) and Fay (2010).

Secondly, following a series of works, originated by Buehler (1957), the smallest exact one-sided interval is derived under a given order on the sample space. A $1-\alpha$ exact two-sided interval C_{d2} for d is easily obtained by taking the intersection of two smallest lower and upper one-sided $1-\frac{\alpha}{2}$ intervals in Wang (2010). However, such intervals may be conservative (Agresti, 2013). When there exists a nuisance parameter, including the current case, it is challenging to improve an exact but conservative interval. The main difficulty is that the method of finding a confidence coefficient in Wang (2007) fails. The proposed modification process, however, provides a promising effort to shrink C_{d2}.

The following approximate intervals are also considered: the Wald type

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

interval and the maximum likelihood estimator $\hat{p}_1 - \hat{p}_2$ (used as an interval)

$$C_{d3}(X,Y) = [\hat{p}_1 - \hat{p}_2 \mp z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}],\ C_{d4}(X,Y) = [\hat{p}_1 - \hat{p}_2 \mp 0],$$

where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$th percentile of the standard normal distribution. They both have a zero confidence coefficient for any n_1, n_2 and α, and C_{d4} has a zero total interval length.

Table 3: The infimum coverage probability (ICP) and total interval length (TIL) for twelve 95% intervals C_{di}, C_{di}^M, and $C_{di}^{M\infty}(=C_{di}^{Mk})$ for $i=1$ (Score, exact), 2 (Wang, exact), 3 (Wald, approximate), 4 (the maximum likelihood estimator, approximate) when (n_1, n_2) varies. The smallest TIL for exact intervals is marked by * for each (n_1, n_2).

(n_1,n_2)	C_{di}	ICP	TIL	C_{di}^M	ICP	TIL	$C_{di}^{M\infty}$	ICP	TIL
(5,6)	C_{d1}	0.9500	38.7295	C_{d1}^M	0.9500	38.5384	C_{d1}^{M17}	0.9500	38.4253
	C_{d2}	0.9511	41.7394	C_{d2}^M	0.9500	38.6381	C_{d2}^{M19}	0.9500	38.3833*
	C_{d3}	0	34.2758	C_{d3}^M	0.9500	52.2953	C_{d3}^{M22}	0.9500	45.4999
	C_{d4}	0	0	C_{d4}^M	0.9500	39.5540	C_{d4}^{M18}	0.9500	39.1202
(10,15)	C_{d1}	0.9500	113.3737	C_{d1}^M	0.9500	112.1987	C_{d1}^{M18}	0.9500	111.5613*
	C_{d2}	0.9515	116.8048	C_{d2}^M	0.9500	112.6569	C_{d2}^{M18}	0.9500	111.7894
	C_{d3}	0	106.2471	C_{d3}^M	0.9500	148.7108	C_{d3}^{M30}	0.9500	131.8738
	C_{d4}	0	0	C_{d4}^M	0.9500	120.7789	C_{d4}^{M2}	0.9500	120.7007
(23,32)	C_{d1}	0.9500	346.4825	C_{d1}^M	0.9500	344.3728	C_{d1}^{M20}	0.9500	342.6230
	C_{d2}	0.9503	347.4601	C_{d2}^M	0.9500	342.6516	C_{d2}^{M17}	0.9500	341.4697*
	C_{d3}	0	332.3962	C_{d3}^M	0.9500	399.0738	C_{d3}^{M47}	0.9500	375.2666
	C_{d4}	0	0	C_{d4}^M	0.9500	372.2596	C_{d4}^{M17}	0.9500	370.0785

Table 3 reports the confidence coefficient and total interval length of

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

these 95% intervals for different (n_1, n_2). Each confidence coefficient is obtained through a large number of calculations: select 201^2 pairs of (p_1, p_2), where both p_1 and p_2 are the multiples of 0.005, and 50000 pairs of (p_1, p_2) following a uniform distribution; compute these coverage probabilities; use the minimum as the confidence coefficient.

The final refined intervals, $C_{d1}^{M\infty}$ and $C_{d2}^{M\infty}$, for exact intervals are shorter than those, $C_{d3}^{M\infty}$ and $C_{d4}^{M\infty}$, for approximate intervals. Originally from a point estimator, $C_{d4}^{M\infty}$ is surprisingly shorter than $C_{d3}^{M\infty}$. Although C_{d2} is wider than C_{d1}, $C_{d2}^{M\infty}$ performs better than or as well as $C_{d1}^{M\infty}$. These indicate that the modification process is effective in generating both accurate and precise intervals.

To determine whether C_{di}^{Mk} for an integer k is equal to $C_{di}^{M\infty}$, the ratio of the total interval lengths of two consecutive intervals $C_{di}^{M(k+1)}$ and C_{di}^{Mk} is used as in Section 3.1. If it is equal to one, then, by Theorem 3, $C_{di}^{M\infty} = C_{di}^{Mk}$. But on a sample point (x, y), $k(x, y)$ in Theorem 4 may be much smaller than k. For example, when $(n_1, n_2) = (5, 6)$, $C_{d2}^{M\infty} = C_{d2}^{M19}$ for $k = 19$; but $C_{d2}^{M}(0, 5) = C_{d2}^{M2}(0, 5) = [-0.9915, -0.2587]$. Thus, $C_{d2}^{M\infty}(0, 5) = C_{d2}^{M}(0, 5)$ for $k(0, 5) = 1$, much smaller than 19.

Being a conservative interval, C_{d2} has a larger confidence coefficient and total interval length than C_{d1}. But the modification process makes

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$C_{d2}^{M\infty}$ having the smallest total interval length in general. If an interval has a small number of ties, then its modified interval tends to be short. This happens on C_{d2}. In contrast, $C_{d4} = \hat{p}_1 - \hat{p}_2$ has many ties. For instance, when $(n_1, n_2) = (10, 15)$, $C_{d4}(2i, 3i) = 0$ for $i = 0, ..., 5$. Then, $C_{d4}^{M\infty}(2i, 3i) = [-0.3834, 0.3834]$. In consequence, $C_{d4}^{M\infty}$ is much longer than $C_{d2}^{M\infty}$.

Intervals $C_{di}^{M\infty}$'s in Table 3 are not admissible due to ties in their confidence limits. However, they can be modified to be admissible by breaking the ties. When $(n_1, n_2) = (5, 6)$, the lower limits of $C_{d1}^{M\infty}$ are equal to -0.1942 at points (3,1) and (5,4). i) Break the ties by lifting the lower limit at one of the two points, say (3,1), to -0.19419, just a little larger than -0.1942. i.e., introduce an interval C_{new} which has the same lower limits as $C_{d1}^{M\infty}$ except the lower limit at $(3, 1)$. ii) Compute the confidence coefficient of C_{new}. iii) If this confidence coefficient is less than 0.95, then $C_{d1}^{M\infty}$ cannot be shortened at $(3, 1)$; otherwise, apply the modification process to C_{new} and obtain $C_{new}^{M\infty}$, a subset of $C_{d1}^{M\infty}$. Repeat i), ii) and iii) to all other tied points and obtain an admissible interval. The total interval lengths for the admissible intervals that are obtained by improving $C_{d1}^{M\infty}$ and $C_{d2}^{M\infty}$ are equal to 38.4077 and 38.3728, respectively.

Example 1 (continued). The twelve intervals in Table 3 at $(x, y) =$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 4: Four 95% confidence intervals and their modifications at $(x, y) = (21, 19)$: (C_{di}, C_{di}^M, $C_{di}^{M\infty}$), for $i = 1, ..., 4$, and their lengths when $(n_1, n_2) = (23, 32)$. The smallest length of the exact intervals is marked by *.

C_{di}	lower	upper	length	C_{di}^M	lower	upper	length	$C_{di}^{M\infty}$	lower	upper	length
C_{d1}	0.0794	0.5228	0.4434	C_{d1}^M	0.0794	0.5223	0.4429	$C_{d1}^{M\infty}$	0.0794	0.5218	0.4424
C_{d2}	0.0946	0.5126	0.4180	C_{d2}^M	0.0968	0.5081	0.4113	$C_{d2}^{M\infty}$	0.0968	0.5038	0.4070*
C_{d3}	0.1138	0.5248	0.4110	C_{d3}^M	0.0569	0.5486	0.4917	$C_{d3}^{M\infty}$	0.1185	0.5468	0.4283
C_{d4}	0.3193	0.3193	0	C_{d4}^M	0.0523	0.5442	0.4919	$C_{d4}^{M\infty}$	0.0529	0.5438	0.4909

$(21, 19)$ are reported in Table 4. Interval $C_{d2}^{M\infty}(x, y)$ is equal to $[0.0968, 0.5038]$ and has the shortest length 0.4070, a confidence coefficient 0.95, and the shortest total interval length 341.4697. $\square$

3.3 Intervals for the difference of two dependent proportions

Consider a 2×2 contingency table with two binary variables A (row) and B (column), where 1 is a success and 0 is a failure. The random observation $(N_{11}, N_{10}, N_{01}, N_{00})$ follows a multinomial distribution with n trials and probabilities $(p_{11}, p_{10}, p_{01}, p_{00})$. The parameter of interest here is $d_m = P(A = 1) - P(B = 1) = p_{10} - p_{01}$. Let $T = N_{11} + N_{00}$ and $p_t = p_{11} + p_{00}$. The conditional distribution of N_{ij}'s for given (N_{10}, T) does not involve p_{10} and p_{01}, so the inferences about d_m should be based on (N_{10}, T) if following the similar reasoning of the sufficiency principle. The reduced sample and

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

parameter spaces are $S_M = \{(n_{10}, t) : n_{10} + t \in [0, n]\}$ and $H_M = \{(d_m, p_t) : d_m \in [-1, 1], p_t \in [0, 1 - |d_m|]\}$, respectively. The probability mass function for (N_{10}, T), in terms of (d_m, p_t), is

$$p_M(n_{10}, t, d_m, p_t) = \frac{n!}{n_{10}!t!n_{01}!}(\frac{1 + d_m - p_t}{2})^{n_{10}} p_t^t (\frac{1 - d_m - p_t}{2})^{t - n_{10}}.$$

Wang (2012) proposed the smallest $1 - \frac{\alpha}{2}$ lower and upper one-sided intervals for d_m. Then, their intersection, denoted by $C_{d_m1}(N_{10}, T) = [L_{d_m1}(N_{10}, T), U_{d_m1}(N_{10}, T)]$, is of level $1 - \alpha$ and can be computed by an R-package "ExactCIdiff" (Shan and Wang, 2013). To derive $C^M_{d_m1}(n_{10}, t)$, let $T_{m1}(n_{10}, t, d_m) = T^D_2(L_{d_m1}(n_{10}, t), U_{d_m1}(n_{10}, t), d_m)$ and

$$h_{m1}(n_{10}, t, d_m) = \sup_{p_t \in [0, 1-|d_m|]} \sum_{\{(n'_{10}, t') \in S_M : T_{m1}(n'_{10}, t', d_m) \le T_{m1}(n_{10}, t, d_m)\}} p_M(n_{10}, t, d_m, p_t). \tag{3.15}$$

Following Theorem 2, interval $C^M_{d_m1}(n_{10}, t) = \overline{\{d_m : h_{m1}(n_{10}, t, d_m) > \alpha\}}$ is exact and is a subset of $C_{d_m1}(n_{10}, t)$. The upper limit can be computed by the lower limit using $U^M_{d_m1}(n_{10}, t) = -L^M_{d_m1}(n_{01}, t)$. Repeat the modification process for k times so that $C^{M\infty}_{d_m1} = C^{Mk}_{d_m1}$.

Fagerland, Lydersen and Laake (2013) provided a good summary on the approximate and exact intervals for d_m and recommended Tango approximate score interval C_{d_m2} (Tango, 1998) and the Wald interval with Bonett–Price adjustment C_{d_m3} (Bonett and Price, 2012). The modifica-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

tion process generates the improved intervals for C_{d_m2} and C_{d_m3}. Next, we present a numerical comparative study for C_{d_m1}, C_{d_m2}, C_{d_m3} and their modifications.

Example 2. Bentur et al. (2009) measured airway hyper-responsiveness status: (Yes = 1, No = 0) in $n(= 21)$ children before (A) and after (B) stem cell transplantation and observed $(n_{11}, n_{10}, n_{01}, n_{00}) = (1, 1, 7, 12)$. So, $(n_{10}, t) = (1, 13)$. Then, the maximum likelihood estimate for d_m is -0.2857. Table 5 reports nine 95% confidence intervals at $(1, 13)$ for individual performance and their confidence coefficient and total interval length for overall performance.

As expected, C_{d_m1} has the largest length and total interval length. However, the small total interval lengths for C_{d_m2} and C_{d_m3} are due to their incorrect confidence coefficients: 0.8367 and 0.9145. The modified intervals all have confidence coefficients no less than 0.95; $C^{M\infty}_{d_m1}$ is the shortest at (1,13) and has a little larger total interval length than $C^{M\infty}_{d_m2}$. One reason for a large total interval length of $C^{M\infty}_{d_m3}$ is that C_{d_m3} has many ties in its confidence limits, especially when n_{10} is close to n or 0.

Table 5 also reports the p-values for testing $H_0 : d_m = 0$ that are associated with the modified intervals $C^{M}_{d_mi}$ and $C^{M\infty}_{d_mi}$ at $(1, 13)$. The p-value corresponding to $C^{M}_{d_m1}$ is equal to $h_{m1}(1, 13, 0) = 0.04125$, where h_{m1}

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Table 5: Nine 95% confidence intervals: $(C_{d_m i}, C^M_{d_m i}, C^{M\infty}_{d_m i})$ for $i = 1$ (Wang, exact), 2 (Tango, approximate), 3 (Wald with Bonett and Price adjustment, approximate), the interval length at $(n_{10}, t) = (1, 13)$, the infimum coverage probability (ICP) and total interval length (TIL) when $n = 21$. The smallest length and TIL of exact intervals are marked by *.

Part I: The nine intervals at $(n_{10}, t) = (1, 13)$

	lower	upper	length		lower	upper	length		lower	upper	length
						the p-value				the p-value	
C_{d_m1}	-0.5214	-0.0126	0.5088	$C^M_{d_m1}$	-0.5065	-0.0155	0.4910	$C^{M\infty}_{d_m1}$	-0.4923	-0.0155	0.4768*
						0.04125				0.04393	
C_{d_m2}	-0.5173	-0.0260	0.4913	$C^M_{d_m2}$	-0.5320	-0.0182	0.5138	$C^{M\infty}_{d_m2}$	-0.5287	-0.0182	0.5105
						0.04125				0.04215	
C_{d_m3}	-0.5084	-0.0133	0.4951	$C^M_{d_m3}$	-0.5000	0.0122	0.5122	$C^{M\infty}_{d_m3}$	-0.4997	0.0122	0.5119
						0.07835				0.07835	

Part II: The nine interval's ICPs and TILs

	ICP	TIL		ICP	TIL		ICP	TIL
C_{d_m1}	0.9500	152.4780	$C^M_{d_m1}$	0.9500	147.8739	$C^{M\infty}_{d_m1}$	0.9500	146.8296
C_{d_m2}	0.8376	144.1614	$C^M_{d_m2}$	0.9500	147.7267	$C^{M\infty}_{d_m2}$	0.9500	146.2317*
C_{d_m3}	0.9146	147.7201	$C^M_{d_m3}$	0.9501	152.3374	$C^{M\infty}_{d_m3}$	0.9500	149.2308

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

is given in (3.15). This is consistent to the fact that $C^M_{d_m1}(1,13)$ excludes zero; while $C^M_{d_m3}(1,13)$ and $C_{d_m3}(1,13)$ provide two different conclusions on including zero. $\square$

4. Modifying one-sided confidence intervals

Assume the range of parameter θ is $[A,B]$ for two known constants A and B. Consider $H_0 : \theta \leq \theta_0$ vs $H_A : \theta > \theta_0$. For a lower one-sided interval for θ, $C_l(\underline{X}) = [L_l(\underline{X}), B]$ with $L_l(\underline{X}) \geq A$, let $T_{1l}(\underline{x}, \theta_0) = \theta_0 - L_l(\underline{x})$. A small value of T_{1l} supports H_A and $T_{1l}(\underline{x}, \theta_0) \geq 0$ if and only if $\theta_0 \geq L_l(\underline{x})$. The h-function is

$$h_{1l}(\underline{x}, \theta_0) = \sup_{H_0} P(T_{1l}(\underline{X}, \theta_0) \leq T_{1l}(\underline{x}, \theta_0)) = \sup_{\theta \leq \theta_0} P(L_l(\underline{x}) \leq L_l(\underline{X})). \tag{4.16}$$

Following (2.5), the level-α acceptance region for H_0 and $1-\alpha$ exact lower one-sided interval for θ are

$$A_{1l}(\theta_0) = \{\underline{x} : h_{1l}(\underline{x}, \theta_0) > \alpha\} \text{ and } C^M_l(\underline{X}) = \overline{\{\theta_0 : h_{1l}(\underline{x}, \theta_0) > \alpha\}}, \tag{4.17}$$

respectively. As mentioned in Section 3.2, the smallest $1-\alpha$ exact one-sided confidence interval under a given order can be automatically constructed. In the current case, the order is given by the function $L_l(\underline{X})$. More precisely,

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

consider a class of $1-\alpha$ exact intervals

$$\mathcal{C}_l = \{C(\underline{X}) = [L(\underline{X}), B] : L(\underline{x}') \le (=) L(\underline{x}) \ \text{ if } L_l(\underline{x}') \le (=) L_l(\underline{x}), \ \forall \ \underline{x}' \text{ and } \underline{x}\}.$$

The smallest interval that is contained in any interval in $\mathcal{C}_l$ is of interest. Different from Section 2, only one time of modification yields the smallest interval – a much stronger result than Theorems 1 through 4. This also establishes a connection between the h-function method and the construction of the smallest one-sided interval under an order.

Theorem 6. *For a lower one-sided confidence interval $C_l(\underline{X}) = [L_l(\underline{X}), B]$ of any level,*

(i) interval $C_l^M(\underline{X})$ given in (4.17) is a $1-\alpha$ exact interval;

(ii) $C_l^{M\infty}(\underline{X}) = C_l^{Mk}(\underline{X})$ for $k=1$;

(iii) $C_l^M(\underline{X}) = [L_l^M(\underline{X}), B]$ is the smallest in $\mathcal{C}_l$. i.e., $L_l^M(\underline{X}) \ge L(\underline{X})$ for $[L(\underline{X}), B] \in \mathcal{C}_l$.

Example 3 (estimating the vaccine efficacy) Under the setting of Section 3.2, the vaccine efficacy (VE) is defined as $VE = 1 - r = 1 - \frac{p_1}{p_2}$, where r is the relative risk, and p_1 and p_2 are the rates of developing the disease for vaccinated people and unvaccinated people, respectively. A lower one-sided interval $C_{ve}(X,Y) = [L_{ve}(X,Y), 1]$ for VE is wanted since a large VE is needed. Let $U_r(X,Y)$ be the upper limit of the $1-2\alpha$ two-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

sided Koopman interval (1984) for r, which was recommended by Fagerland, Lydersen and Laake (2015). Then $C_{ve}(X,Y)$ with the lower limit $L_{ve}(X,Y) = 1 - U_r(X,Y)$ is a $1-\alpha$ approximate interval for VE. Following Theorem 6, $C_{ve}^M(x,y)$ is derived by solving

$$h(x,y,VE_0) \;=\; \sup_{\{1-\frac{p_1}{p_2}\le VE_0\}} \sum_{\{(u,v):L_{ve}(x,y)\le L_{ve}(u,v)\}} p_B(u,n_1,p_1)p_B(v,n_2,p_2) > \alpha.$$

Janssen Biotech, Inc. (2021) reported that Johnson & Johnson's Janssen Vaccine for Coronavirus Disease 2019 has a point estimate of 81.7142% for VE for the Severe/Critical group in South Africa based on the data $(x,n_1,y,n_2) = (4,2449,22,2463)$. Solve $h(4,22,VE_0) > 0.05$ and obtain the smallest 95% exact lower one-sided interval $C_{ve}^M(4,22) = [0.56564,1]$. In contrast, $C_{ve}(4,22) = [0.56566,1]$, however, C_{ve} only has a confidence coefficient of 0.8000. Therefore, $C_{ve}^M(4,22)$ is as precise as but much more reliable than $C_{ve}(4,22)$. For the Moderate to Severe/Critical group, $(x,n_1,y,n_2) = (23,2449,64,2463)$, $\widehat{VE} = 63.8571\%$, $C_{ve}^M(23,64) = [0.46386,1]$, and $C_{ve}(23,64) = [0.46388,1]$. Again, $C_{ve}^M(23,64)$ dominates $C_{ve}(23,64)$. $\square$

Example 4 (the stochastically nondecreasing distribution family). Suppose X has a cumulative distribution function $F(x,\theta)$ that satisfies $F(x,\theta_1) \ge F(x,\theta_2)$ for any x and $\theta_1 \le \theta_2$. This family includes all important single-parameter distributions. The modified interval $C_l^M(X)$ for the one-sided interval $C_l(X) = [X,B]$ is of interest. Interval $C_l(X)$ itself may be mean-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

ingless for estimating θ. For example, one would not use $[X, 1]$ to estimate p when $X \sim Bino(n, p)$. Following (4.16), $h_{1l}(x, \theta_0) = \max_{\theta \leq \theta_0} P(x \leq X) = 1 - F(x^-, \theta_0)$, where x^- denotes the largest value of X less than x. The lower limit of $C_l^M(x)$ is $L_l^M(x) = \inf\{\theta_0 : 1 - F(x^-, \theta_0) > \alpha\}$. Theorem 6 assures that $C_l^M(X)$ is the smallest interval among all $1-\alpha$ exact intervals of form $[L(X), B]$ with a nondecreasing $L(X)$. In particular, if $X \sim Bino(n, p)$, then $C_l^M(X)$ is the lower one-sided Clopper-Pearson interval of level $1-\alpha$.

This example shows the importance of selecting a good order on interval construction. In particular, C_l is not a meaningful interval, but the good order by X still generates the smallest interval. $\square$

For an upper one-sided interval $C_u(\underline{X}) = [A, U_u(\underline{X})]$, let $T_{1u}(\underline{x}, \theta_0) = U_u(\underline{x}) - \theta_0$ and

$$h_{1u}(\underline{x}, \theta_0) = \sup_{H_0} P(T_{1u}(\underline{X}, \theta_0) \leq T_{1u}(\underline{x}, \theta_0)) = \sup_{\theta \geq \theta_0} P(U_u(\underline{X}) \leq U_u(\underline{x})). \tag{4.18}$$

The level-α acceptance region for $H_0 : \theta \geq \theta_0$ and $1-\alpha$ exact upper one-sided interval for θ are

$$A_{1u}(\theta_0) = \{\underline{x} : h_{1u}(\underline{x}, \theta_0) > \alpha\} \text{ and } C_u^M(\underline{x}) = \overline{\{\theta_0 : h_{1u}(\underline{x}, \theta_0) > \alpha\}}, \tag{4.19}$$

respectively. The following is a parallel result to Theorem 6 and the proof is skipped.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Theorem 7. *For an upper one-sided interval $C_u(\underline{X}) = [A, U_u(\underline{X})]$ of any level,*

(i) interval $C_u^M(\underline{X})$ given in (4.19) is a $1-\alpha$ exact interval;

(ii) $C_u^{M\infty}(\underline{X}) = C_u^M(\underline{X})$;

(iii) Define a class of $1-\alpha$ exact upper one-sided intervals $\mathcal{C}_u = \{C(\underline{X}) = [A, U(\underline{X})] : U(\underline{x}') \le U(\underline{x})$ if $U_u(\underline{x}') \le U_u(\underline{x}),\ \forall\ \underline{x}'$ and $\underline{x}\}$. Then, $C_u^M(\underline{X}) = [A, U_u^M(\underline{X})]$ is the smallest interval in $\mathcal{C}_u$.

5. Discussions

A confidence interval can be obtained by converting a family of tests and vice versa. However, we formally introduce a middle function, the h-function, that yields both the confidence interval and test – a simpler but more general approach. This idea was used by Blaker (2000) and Agresti and Min (2001), but the process was not defined as in the general setting of the current paper. More importantly, the proposed h-function method is now used for the first time to improve any confidence interval.

The effectiveness of the method is demonstrated in Sections 3 and 4. In particular, the method differentiates the parameter values in a given confidence interval. It can be used in many applications, especially when the underlying distribution is discrete and contains nuisance parameters. The-

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

orem 1 is easy to follow and powerful in its ability to modify any intervals, including asymptotic intervals, point estimators and credible intervals, to exact intervals. This is a solution for the important problem that approximate intervals are easy to obtain but not reliable. Through the modification process, the reliability of these intervals are greatly enhanced since an invalid inferential procedure is converted to be a valid one. When nuisance parameters exist, it is also important to improve an existing conservative interval. Theorem 2 is a successful effort to resolve the problem because it delivers uniformly shorter exact intervals, while Theorem 3 provides the smallest interval that can be generated by the modification process. When the final interval $C_0^{M\infty}$ is not admissible due to ties, one can break the ties and then apply the modification process to obtain an admissible interval. Furthermore, Theorems 6 and 7 establish a connection between the h-function method and the construction of the smallest one-sided interval based on an order. All these results build a solid foundation for deriving optimal exact confidence intervals.

From a theoretical point of view, the interval construction in (2.7), (2.8) and (2.9) is an automatic process. However, it is computationally complex, particularly for precisely finding the globe maximum of $P(K(\underline{x}, \theta_0))$ in (2.6), as a function of $\underline{\eta}$, and solving the smallest and largest roots of the equation,

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$h(\underline{x}, \theta_0) > \alpha$, as a function of θ_0. To our best knowledge, no software can accomplish the two tasks both quickly and accurately. It is worth mentioning that $h(\underline{x}, \theta_0)$ is not continuous in θ_0 in general. Our best effort for global optimization is based on a combination of grid search and local optimization. This reduces down to a question of whether the resultant interval (e.g., $C_0^{M\infty}$) is truly of level $1-\alpha$ due to a grid search which may not be fine enough. We intend to select a large number of points for the search, which inevitably takes more time in computation. For example, this number is between 200 and 4000 for a range of $[-1, 1]$ when deriving intervals for d. Additionally, each of the tables reports the confidence coefficient to assure that an exact interval has a correct confidence coefficient.

Future research questions include: i) which is the best choice of T_2 in (2.7) so that C_0^{Mk} converges to $C_0^{M\infty}$ fast? ii) how to construct an optimal confidence interval for a function of several parameters through some existing intervals of those parameters? iii) how to combine several confidence intervals for the same parameter θ, where each of these intervals only uses a part of the data set, to form an optimal interval that utilizes the whole data set?

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Acknowledgments

The author is indebted to Professor Xiaotong Shen for the interpretation of Theorem 2.

Appendix: the proofs

Proof of Theorem 1. Following (2.8), $h_2(\underline{x}, \theta_0)$ is a valid p-value (Casella and Berger 2002, p. 397). Thus, $A_2(\theta_0)$ in (2.9) is the acceptance region of a level-α test, implying that interval $C_0^M(\underline{x})$ is of level $1-\alpha$. $\square$

Proof of Theorem 2. We only need to prove the first claim because the second claim follows Theorem 1 and the first claim. To prove the first claim, it suffices to show $h_2(\underline{x}, \theta_0) \leq \alpha$ for any $\theta_0 \notin C_0(\underline{x})$. Let $Cover_{C_0}(\theta, \underline{\eta})$ be the coverage probability function of $C_0(\underline{X})$. So, $\inf_H Cover_{C_0}(\theta, \underline{\eta}) \geq 1-\alpha$.

First consider the case of $\theta_0 < L_0(\underline{x})$. For $H_0: \theta = \theta_0$,

$$\begin{aligned} h_2(\underline{x}, \theta_0) &\overset{\theta_0 < L_0(\underline{x})}{\leq} \sup_{H_0} P(T_2(\underline{X}, \theta_0) < 0) \overset{(b)}{=} \sup_{H_0}(1 - P(\theta_0 \in C_0(\underline{X}))) \\ &\leq 1 - \inf_H Cover_{C_0}(\theta, \underline{\eta}) \leq \alpha. \end{aligned}$$

Similarly, $h_2(\underline{x}, \theta_0) \leq \alpha$ when $\theta_0 > U_0(\underline{x})$. Hence, $C_0^M(\underline{x}) \subset C_0(\underline{x})$.

When $T_2 = T_2^I$, $h_2(\underline{x}, \theta_0) = 1$ for any $\theta_0 \in C_0(\underline{x})$. Thus, $C_0^M(\underline{x}) = \{\theta_0 : h_2(\underline{x}, \theta_0) > \alpha\} \supset C_0(\underline{x})$. $\square$

Proof of Theorem 3. We only prove part ii) as the other claims are

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

straightforward. Let $Cover_C(\theta, \underline{\eta})$ be the coverage probability function for an interval $C(\underline{X})$. Note that the indicator functions satisfy

$$I_{C_0^{M\infty}(\underline{x})}(\theta) = \lim_{k \to +\infty} I_{C_0^{Mk}(\underline{x})}(\theta), \ \ \forall \underline{x}$$

because $C_0^{Mk}(\underline{x})$ is nonincreasing and note that $Cover_{C_0^{Mk}}(\theta, \underline{\eta}) \geq 1 - \alpha$ for any $(\theta, \underline{\eta})$ because each interval $C_0^{Mk}(\underline{X})$ is of level $1 - \alpha$. Then, following the Dominated Convergence Theorem

$$Cover_{C_0^{M\infty}}(\theta, \underline{\eta}) \ = \ \lim_{k \to +\infty} E_{(\theta, \underline{\eta})}[I_{C_0^{Mk}(\underline{x})}(\theta)] = \lim_{k \to +\infty} Cover_{C_0^{Mk}}(\theta, \underline{\eta}) \geq 1 - \alpha.$$

□

Proof of Theorem 4. It suffices to prove the case of $k = 0$. i.e., if $C_0(\underline{x}) = C_0^M(\underline{x})$, then $C_0^M(\underline{x}) = C_0^{M2}(\underline{x})$. Denote $C_0(\underline{x}) = [L_0(\underline{x}), U_0(\underline{x})]$ and $C_0^M(\underline{x}) = [L_0^M(\underline{x}), U_0^M(\underline{x})]$. By definition, $T_2^M(\underline{x}, \theta_0) = T_2(\underline{x}, \theta_0)$. Also, $T_2^M(\underline{y}, \theta_0) \leq T_2(\underline{y}, \theta_0)$ for any $\underline{y}$ due to $C_0^M(\underline{X}) \subset C_0(\underline{X})$ and Condition (c). Then,

$$h_2^M(\underline{x}, \theta_0) = \sup_{H_0} P(\underline{y} : T_2^M(\underline{y}, \theta_0) \leq T_2^M(\underline{x}, \theta_0))$$

$$\geq \sup_{H_0} P(\underline{y} : T_2(\underline{y}, \theta_0) \leq T_2^M(\underline{x}, \theta_0)) = h_2(\underline{x}, \theta_0).$$

So, $C_0^M(\underline{x}) = \overline{\{\theta_0 : h_2(\underline{x}, \theta_0) > \alpha\}} \subset \overline{\{\theta_0 : h_2^M(\underline{x}, \theta_0) > \alpha\}} = C_0^{M2}(\underline{x})$ and $C_0^M(\underline{x}) = C_0^{M2}(\underline{x})$. □

Proof of Theorem 5. We only prove the case of $T_2 = T_2^D$. The proof for $T_2 = T_2^R$ is similar. Suppose the claim of theorem is not true. There exists

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

a $1-\alpha$ exact interval $C_1(\underline{X}) = [L_1(\underline{X}), U_1(\underline{X})]$ and a sample point $\underline{x}_0$ so that $C_1(\underline{x}_0) \subsetneqq C_0^{M\infty}(\underline{x}_0)$ and $C_1(\underline{x}) = C_0^{M\infty}(\underline{x})$ if $\underline{x} \neq \underline{x}_0$. Without loss of generality, assume $L_0^{M\infty}(\underline{x}_0) < L_1(\underline{x}_0)$ and $U_0^{M\infty}(\underline{x}_0) = U_1(\underline{x}_0)$.

Since $L_0^{M\infty}(\underline{X})$ is a one-to-one function and assumes finite many values, we choose $L_1(\underline{x}_0)$ close to $L_0^{M\infty}(\underline{x}_0)$ so that none of the $L_0^{M\infty}(\underline{x})$'s belongs to interval $(L_0^{M\infty}(\underline{x}_0), L_1(\underline{x}_0))$. Denote $\epsilon = L_1(\underline{x}_0) - L_0^{M\infty}(\underline{x}_0)$, which can be any small positive number. Pick $\theta_0^* = L_1(\underline{x}_0) - \epsilon/m \in (L_0^{M\infty}(\underline{x}_0), L_1(\underline{x}_0))$ for a large positive integer m. Define $T_0(\underline{x}, \theta_0) = T_2^D(L_0^{M\infty}(\underline{x}), U_0^{M\infty}(\underline{x}), \theta_0)$ and $T_1(\underline{x}, \theta_0) = T_2^D(L_1(\underline{x}), U_1(\underline{x}), \theta_0)$. Then,

$$T_0(\underline{x}, \theta_0) = T_1(\underline{x}, \theta_0) \ \forall \underline{x} \neq \underline{x}_0; \quad T_1(\underline{x}_0, \theta_0^*) < 0 < T_0(\underline{x}_0, \theta_0^*). \tag{5.20}$$

Let $K_j(\underline{x}_0, \theta_0^*) = \{\underline{y} : T_j(\underline{y}, \theta_0^*) \leq T_j(\underline{x}_0, \theta_0^*)\}$ for $j = 0, 1$. Claim

$$K_0(\underline{x}_0, \theta_0^*) = K_1(\underline{x}_0, \theta_0^*). \tag{5.21}$$

Suppose the claim (5.21) is true. Let h_0 and h_1 be the h-functions for T_0 and T_1, respectively. Then,

$$h_0(\underline{x}_0, \theta_0^*) = \sup_{(\theta_0^*, \underline{\eta})} P(K_0(\underline{x}_0, \theta_0^*)) \overset{(5.21)}{=} \sup_{(\theta_0^*, \underline{\eta})} P(K_1(\underline{x}_0, \theta_0^*)) = h_1(\underline{x}_0, \theta_0^*).$$

Since $\theta_0^* \in C_0^{M\infty}(\underline{x}_0) = (C_0^{M\infty})^M(\underline{x}_0)$, $h_0(\underline{x}_0, \theta_0^*) > \alpha$. On the other hand, since $\theta_0^* \notin C_1(\underline{x}_0)$ and $C_1^M(\underline{x}_0) \subset C_1(\underline{x}_0)$, $\theta_0^* \notin C_1^M(\underline{x}_0)$, which implies $h_1(\underline{x}_0, \theta_0^*) \leq \alpha$. Therefore, $h_0(\underline{x}_0, \theta_0^*) \neq h_1(\underline{x}_0, \theta_0^*)$, a contradiction. Therefore, the claim of the theorem is true.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Now we prove the claim (5.21).

Case i). Suppose $U_0^{M\infty}(\underline{x}) \neq L_0^{M\infty}(\underline{x}_0)$ for any $\underline{x}$. Thus,

$$\begin{aligned}
& K_0(\underline{x}_0, \theta_0^*) \overset{(5.20)}{=} \{\underline{y} : \underline{y} \neq \underline{x}_0, T_1(\underline{y}, \theta_0^*) \leq T_0(\underline{x}_0, \theta_0^*)\} \cup \{\underline{y} : \underline{y} = \underline{x}_0\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}_0, T_1(\underline{y}, \theta_0^*) \leq \theta_0^* - L_0^{M\infty}(\underline{x}_0)\} \cup \{\underline{y} : \underline{y} = \underline{x}_0\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}_0, T_1(\underline{y}, \theta_0^*) \leq \theta_0^* - L_1(\underline{x}_0)\} \\
& \cup \{\underline{y} : \underline{y} \neq \underline{x}_0, \theta_0^* - L_1(\underline{x}_0) < T_1(\underline{y}, \theta_0^*) \leq \theta_0^* - L_0^{M\infty}(\underline{x}_0)\} \cup \{\underline{y} : \underline{y} = \underline{x}_0\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}_0, T_1(\underline{y}, \theta_0^*) \leq T_1(\underline{x}_0, \theta_0^*)\} \cup \emptyset \cup \{\underline{y} : \underline{y} = \underline{x}_0\} = K_1(\underline{x}_0, \theta_0^*).
\end{aligned}$$

Case ii). Suppose $U_0^{M\infty}(\underline{x}^*) = L_0^{M\infty}(\underline{x}_0)$ for some $\underline{x}^*(\neq \underline{x}_0)$. Such x^* must be unique.

$$\begin{aligned}
& K_0(\underline{x}_0, \theta_0^*) = \{\underline{y} : T_0(\underline{y}, \theta_0^*) \leq T_0(\underline{x}_0, \theta_0^*)\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq T_1(\underline{x}_0, \theta_0^*)\} \cup \{\underline{y} : \underline{y} = \underline{x}^*, T_0(\underline{y}, \theta_0^*) \leq T_0(\underline{x}_0, \theta_0^*)\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq T_1(\underline{x}_0, \theta_0^*)\} \cup \{\underline{y} : \underline{y} = \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq \theta_0^* - L_0^{M\infty}(\underline{x}_0)\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq T_1(\underline{x}_0, \theta_0^*)\} \cup \{\underline{y} : \underline{y} = \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq \theta_0^* - L_1^{M\infty}(\underline{x}_0)\} \\
= \ & \{\underline{y} : \underline{y} \neq \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq T_1(\underline{x}_0, \theta_0^*)\} \cup \{\underline{y} : \underline{y} = \underline{x}^*, T_1(\underline{y}, \theta_0^*) \leq T_1(\underline{x}_0, \theta_0^*)\} \\
= \ & K_1(\underline{x}_0, \theta_0^*).
\end{aligned}$$

The proof is complete. □

Proof of Proposition 1. Note (3.11), $D(-d_0) = 1 - D(d_0)$ and $p_B(x, n, p) =$

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

$p_B(n-x,n,1-p)$. Then,

$$\begin{aligned}
& h_d(n_1-x,n_2-y,-d_0) \\
= & \sup_{p_2\in D(-d_0)} \sum_{\{(u,v)\in S_d:T_d(u,v,-d_0)\le T_d(x,y,d_0)\}} p_B(n_1-u,n_1,1-p_2+d_0) \\
& \qquad\qquad * p_B(n_2-v,n_2,1-p_2) \\
= & \sup_{p_2'\in D(d_0)} \sum_{\{(u',v')\in S_d:T_d(u',v',d_0)\le T_d(x,y,d_0)\}} p_B(u',n_1,p_2'+d_0)p_B(v',n_2,p_2') \\
= & h(x,y,d_0).
\end{aligned}$$

Therefore, $h_d(x,y,U(x,y)) = h_d(n_1-x,n_2-y,-U(x,y))$, establishing (3.14). □

Proof of Theorem 6. Part i) is similar to the proof of Theorem 1. Part iii) is similar to the proof of Theorem 4 in Wang (2010) and is skipped. Part ii) follows part iii). □

References

Agresti, A. (2013). *Categorical Data Analysis*, 3rd edition. John Wiley & Sons.

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

Bentur, L., Lapidot, M., Livnat, G., Hakim, F., Lidroneta-Katz, C., Porat, I., Vilozni, D., and Elhasid, R. (2009). Airway reactivity in children before and after stem cell transplantation. *Pediatric Pulmonology*, 44, 845-850.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *Canadian Journal of Statistics*, 28, 783-798.

Bonett, D. G. and Price, R. M. (2012). Adjusted Wald confidence interval for a difference of binomial proportions based on paired data. *Journal of Educational and Behavioral Statistics*, 37(4), 479-488.

Brown, L. D., Cai, T. T. and DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16, 101-133.

Buehler, R. J. (1957). Confidence intervals for the product of two binomial parameters. *Journal of the American Statistical Association*, 52, 482-493.

Casella, G. (1986). Refining binomial confidence intervals. *Canadian Journal of Statistics*, 14, 113-129.

Casella, G. and Berger, R. L. (2002), *Statistical Inference,* 2nd ed. Duxbury Press, Pacific Grove.

Casella, G. and Robert, C. (1989). Refining Poisson confidence intervals. *Canadian Journal of Statistics*, 17, 45-57.

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

Essenberg, J. M. (1952). Cigarette smoke and the incidence of primary neoplasm of the lung in albino mice. *Science*, 116, 561-562.

Fagerland, M. W., Lydersenb, S. and Laakec, P. (2013). Recommended tests and confidence

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

intervals for paired binomial proportions. *Statistics in Medicine*, 33(16), 2850-2875.

Fagerland, M. W., Lydersenb, S. and Laakec, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Statistical Methods in Medical Research*, 24(2), 224-254.

Fay, M. P. (2010). Two-sided exact tests and matching confidence intervals for discrete data. *The R Journal*, 2(1), 53-58.

Janssen Biotech, Inc. (2021). FDA Briefing Document: Janssen Ad26.COV2.S Vaccine for the Prevention of COVID-19. *Vaccines and Related Biological Products Advisory Committee Meeting February 26, 2021*.

Huwang, L. (1995). A note on the accuracy of an approximate interval for the binomial parameter. *Statistics and Probability Letters* 24, 177-180.

Koopman, P. A. R. (1984). Confidence intervals for the ratio of two binomial proportions. *Biometrics* 40, 513–517.

Shan, G. and Wang, W. (2013). ExactCIdiff: An R package for computing exact confidence intervals for the difference of two proportions. *The R Journal*, 5, 62-70.

Tango, T. (1998). Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics in Medicine*, 17, 891-908.

Wang, H. (2007). Exact confidence coefficients of confidence intervals for a binomial proportion. *Statistica Sinica*, 17, 361-368.

Statistica Sinica: Newly accepted Paper
(accepted author-version subject to English editing)

Wang, W. (2010). On construction of the smallest one-sided confidence interval for the difference of two proportions. *The Annals of Statistics*, 38, 1227-1243.

Wang, W. (2012). An inductive order construction for the difference of two dependent proportions. *Statistics and Probability Letters*, 82, 1623-1628.

Wang, W. (2014). An iterative construction of confidence interval for a proportion. *Statistica Sinica*, 24, 1389-1410.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 209-212.

Weizhen Wang

Department of Mathematics and Statistics

Wright State University

Dayton, OH 45435, USA

E-mail: weizhen.wang@wright.edu

School of Statistics

Tianjin University of Finance and Economics

Tianjin 300222, China