SEQUENTIAL ANALYSIS
2019, VOL. 38, NO. 2, 163–183
https://doi.org/10.1080/07474946.2019.1611315

Taylor & Francis
Taylor & Francis Group

Check for updates

Tandem-width sequential confidence intervals for a Bernoulli proportion

Tony Yaacoub[a], David Goldsman[a], Yajun Mei[a], and George V. Moustakides[b]

[a]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA; [b]Department of Computer Science, Rutgers University, Piscataway, New Jersey, USA

ABSTRACT
We propose a two-stage sequential method for obtaining tandem-width confidence intervals for a Bernoulli proportion p. The term "tandem-width" refers to the fact that the half-width of the $100(1-\alpha)\%$ confidence interval is not fixed beforehand; it is instead required to satisfy two different half-width upper bounds, h_0 and h_1, depending on the (unknown) values of p. To tackle this problem, we first propose a simple but useful sequential method for obtaining fixed-width confidence intervals for p, whose stopping rule is based on the minimax estimator of p. We observe Bernoulli(p) trials sequentially, and for some fixed half-width $h = h_0$ or h_1, we develop a stopping time T such that the resulting confidence interval for p, $[\hat{p}_T - h, \hat{p}_T + h]$, covers the parameter with confidence at least $100(1-\alpha)\%$, where $\hat{p}_T$ is the maximum likelihood estimator of p at time T. Furthermore, we derive theoretical properties of our proposed fixed-width and tandem-width methods and compare their performances with existing alternative sequential schemes. The proposed minimax-based fixed-width method performs similarly to alternative fixed-width methods, while being easier to implement in practice. In addition, the proposed tandem-width method produces effective savings in sample size compared to the fixed-width counterpart and provides excellent results for scientists to use when no prior knowledge of p is available.

ARTICLE HISTORY
Received 26 October 2018
Revised 15 January 2019
Accepted 15 April 2019

KEYWORDS
Bernoulli proportion; binomial distribution; confidence interval; sequential analysis; stopping time

SUBJECT CLASSIFICATIONS
62L12; 60G40

1. Introduction

Confidence interval (CI) estimation for a Bernoulli proportion p has a wide variety of important real-world applications such as those involving the prevalence of a rare disease (Sullivan et al., 2013), the overall response rate in clinical trials (Abramson et al., 2013), and accuracy assessment in remote sensing (Morisette and Khorram, 1998). Perhaps the most widely known fixed sample size $100(1-\alpha)\%$ CI is what is commonly referred to as Wald's CI of the form $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$, where $\hat{p}$ is the sample mean of n independent and identically distributed (i.i.d.) Bernoulli(p) observations, and $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. This CI is known to have poor

CONTACT Tony Yaacoub tryaacoub@gmail.com H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive NW, Atlanta, GA 30332-0205, USA.
Recommended by Nitis Mukhopadhyay
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsqa.
© 2019 Taylor & Francis Group, LLC

properties when $np(1-p)$ is small; see Vollset (1993), Agresti and Coull (1998), Newcombe (1998), and Brown et al. (2001). Research continues to be undertaken to improve this CI in the "offline" or "online" contexts. In particular, extensive research has been devoted to CIs for p in the offline context, in which the sample size is fixed a priori. Wilson (1927) proposed the "score" CI, which incorporates a correction term to Wald's CI to yield improved performance. Clopper and Pearson (1934) obtain an exact CI by inverting equal-tailed binomial tests; see also Sterne (1954), Crow (1956), Blyth and Still (1983), and Reiczigel (2003). Agresti and Coull (1998) and Brown et al. (2001, 2002) provide excellent surveys and comparisons of the methods.

In the online context, in which the sample size is not fixed beforehand and in fact depends on the observed data, a great deal of research has obviously been devoted to sequential CIs. In such cases, one continues to take observations until the $100(1-\alpha)\%$ CI satisfies a certain stopping criterion, often related to the length of the CI. For an estimator δ_T of p, with T being the time we stop sampling (i.e., the stopping time), one criterion for obtaining a CI is the *fixed-width* criterion, where the CI for p is given by $[\delta_T-h, \delta_T+h]$ for fixed half-width $h>0$. Example articles among the rich literature in this area include Armitage (1958), Tanaka (1961), Robbins and Siegmund (1974), Khan (1998), Zacks and Mukhopadhyay (2007; which gives two-stage and sequential modifications of Robbins and Siegmund (1974)), Frey (2010), and Yaacoub et al. (2019; a procedure satisfying certain optimality criteria). References involving broader methodology (e.g., a greater variety of distributions and/or functions of parameters) include Chow and Robbins (1965), Khan (1969), Siegmund (1985), and Mukhopadhyay and De Silva (2009).

A related basis for obtaining a CI is the *proportional accuracy* criterion, where the CI for p is given by $\{p : |\delta_T-p|<\eta p\}$ for some fixed $\eta \in (0,1)$. Huber (2017) and Malinovsky and Zacks (2018) cover the Bernoulli proportion; but see also Zacks (1966) and Nadas (1969). A third measure under which we can consider a CI is the *fixed-accuracy* criterion, where the CI for p is given by $\{p : p \in [d^{-1}\delta_T, d\delta_T]\}$ for some fixed $d>1$; see Mukhopadhyay and Banerjee (2015) for the Bernoulli case and Mukhopadhyay and Banerjee (2014) and Banerjee and Mukhopadhyay (2015), which discuss other cases.

In this article, we investigate a sequential CI for a Bernoulli proportion p but with the new twist that the CI is *tandem-width*. By this we mean that the half-width h of the $100(1-\alpha)\%$ CI is not fixed beforehand; it is instead required to satisfy two different upper bounds, h_0 and h_1, depending on the (unknown) values of p. Some motivating examples include the customer click-through rate to measure the efficacy of a new online ad marketing campaign and the statistical model checking approach adopted in complicated stochastic systems; see, for example, Jegourel et al. (2017). In both of these modern applications, it is very expensive and time-consuming to set up the experiments or simulations. Once they are set up, one wants to use the smallest number of samples to gain knowledge of the Bernoulli proportion p as *accurately and precisely* as possible due to the time or cost constraints. For instance, if the true (unknown) value of p were to be in $[0.2, 0.8]$, then one may feel that the half-width $h_0 = 0.1$ is precise enough and is acceptable. On the other hand, if the true p were to be in $[0, 0.1)$ or $(0.9, 1]$, then one may feel that $h_0 = 0.1$ is too crude, and the half-width $h_1 = 0.01$ might be more suitable. This inspires us to investigate the problem of tandem-width sequential interval estimation.

We propose to develop effective sequential methods for tandem-width interval estimation of the Bernoulli proportion p at the pre-specified confidence level $100(1-\alpha)\%$. It is intuitive to combine two sequential fixed-width CIs together, one for each fixed half-width h_0 or h_1, but unfortunately it might be very difficult to implement the combination in the sequential context if the two corresponding stopping times are not monotone decreasing in the half-width. Indeed, though many existing methods yield statistically efficient $100(1-\alpha)\%$ CIs in the sense of small expected sample sizes for a fixed half-width h, the stopping time $T(h)$ often depends heavily on h, and it is unclear whether the stopping boundary (i.e., termination criterion) of $T(h)$ at each time step is a monotone decreasing function of h or not, which would be a desirable property. As a concrete illustration, the stopping time of the sequential CI proposed by Frey (2010) is based on the Bayesian point estimator whose prior distribution depends on the half-width h when optimized for the smallest expected sample size, and thus the monotonicity property is unclear here. To circumvent the monotonicity issue, we propose to use the minimax point estimator of p instead of the Bayesian estimator to develop effective sequential methods for fixed-width sequential CIs. By doing so, it will be straightforward to show that the monotonicity properties hold. The methodology to be described in this article will allow us to conveniently combine two fixed-width sequential interval estimators together, yielding an efficient tandem-width sequential interval estimation method.

The remainder of this article is organized as follows. In Section 2, we formulate our problem on tandem-width sequential CIs for a Bernoulli proportion p and provide some background regarding different point estimators for p and, in particular, on the method proposed by Frey (2010). In Section 3, we describe our sequential stopping rules for the fixed-width CI and the tandem-width CI. We also discuss some asymptotic properties for our proposed methods. Section 4 presents simulation results for our tandem-width stopping rule. We also provide numerical results that compare our proposed fixed-width stopping rule to Frey's stopping rule. These numerical results are obtained through recursive formulas that were inspired by the methodologies given in Zacks (2017). Concluding remarks are included in Section 5.

2. Problem formulation and background

Assume that we observe a sequence of i.i.d. Bernoulli random variables, $X_1, X_2, \ldots$, sequentially; that is, one at a time. Suppose $\mathbf{P}(X_i = 1) = p$ and $\mathbf{P}(X_i = 0) = 1-p$, and we want to use as few samples as possible to make an accurate and precise interval estimate about the unknown parameter $p \in [0, 1]$ at the confidence level $100(1-\alpha)\%$ for some prespecified α. We assume that the $100(1-\alpha)\%$ CI for p is written in the form $[\delta_T - h, \delta_T + h]$, where h is the desired half-width of the CI, and δ_T can be thought as the point estimator of p when we stop taking observations at time T.

In the problem of formulating tandem-width sequential confidence intervals, we want to find a stopping time T and then a corresponding $100(1-\alpha)\%$ CI for p whose half-width is required to satisfy two different upper bounds, h_0 and h_1, depending on the unknown value of p and, in turn, the point estimate of p. On the one hand, when the estimate δ_T is not too small or large, say, when $\delta_T \in [p_0, 1-p_0]$ for some prespecified $p0$ —for example, $p_0 = 0.1$— we would like to set the half-width h of the CI to a relatively large value h_0 (e.g., $h_0 = 0.1$) to save time and sampling costs. On the other hand, when δ_T is quite small or large, say, when $\delta_T < p_0$ or $> 1-p_0$, we would like to set the half-width of the CI to a smaller value h_1 (e.g., $h_1 = 0.01$) in order for the CI to be more meaningful. In the latter case, it is useful to take more time (additional samples) to produce a meaningful CI instead of stopping earlier with a CI that is too wide to be practical.

To be more rigorous, we would like to find a stopping time T and the corresponding estimator δ_T that minimize the average run lengths (ARLs), $\mathbf{E}_p(T)$, simultaneously for all $0 \leq p \leq 1$, subject to the coverage probability (CP) constraints that

$$\mathrm{CP}_p(h_0) = \mathbf{P}_p\big(p \in [\delta_T - h_0, \delta_T + h_0]\big) \geq 1-\alpha, \quad \text{when } p_0 \leq p \leq 1-p_0, \tag{2.1}$$

and

$$\mathrm{CP}_p(h_1) = \mathbf{P}_p\big(p \in [\delta_T - h_1, \delta_T + h_1]\big) \geq 1-\alpha, \quad \text{when } p < p_0 \text{ or } p > 1-p_0, \tag{2.2}$$

where $0 < h_1 < h_0 < 1$ and $\alpha \in (0, 1)$ are prespecified. Note that $\mathbf{E}_p$ and $\mathbf{P}_p$ denote the expectation and the probability measure, respectively, when p is the true Bernoulli parameter.

Let us now provide some background information on point and interval estimation of the Bernoulli proportion p. For this purpose, we first review three different kinds of point estimators of p in the offline context when the complete set of observations is $\{X_1, X_2, ..., X_n\}$: the maximum likelihood estimator (MLE), Bayes estimator, and minimax estimator, denoted by $\hat{p}_n$, $\tilde{p}_n$, $p_n^\star$, respectively, to emphasize their dependence on the sample size n. First of all, the MLE of p is the sample mean,

$$\hat{p}_n = \hat{p}_{\mathrm{MLE}} = \frac{S_n}{n}, \quad \text{where } S_n = \sum_{i=1}^{n} X_i. \tag{2.3}$$

Below we follow the literature to assume that the point estimator $\hat{p}_T$ from (2.1) is the MLE estimator from (2.3) when implemented with the (random) stopping time T. This will allow us to make a fair, apples-to-apples comparison between our proposed stopping time T and other sequential methods in the literature.

As for the Bayes estimator of p, it is well known that if the prior distribution of p is the Beta(a, b) distribution for some prespecified $a, b > 0$, then the posterior of p given observed $(X_1, X_2, ..., X_n)$ is the Beta($a + S_n, b - S_n + n$) distribution. Thus, the mean of the posterior distribution, $(S_n + a)/(n + a + b)$, is the Bayes estimator of p under the standard squared error loss function. One important special case of the prior Beta distribution is when $b = a > 0$, so that the corresponding Bayes estimator of p becomes

$$\tilde{p}_{n,a} = \tilde{p}_{\mathrm{Bayes}} = \frac{S_n + a}{n + 2a}. \tag{2.4}$$

Meanwhile, under the squared error loss function, the minimax framework is to find an estimator $\delta = \delta(X_1, ..., X_n)$ that minimizes the largest mean square error over the entire space $[0, 1]$ of the true parameter p. In other words, the minimax estimator minimizes $\max_{0 \leq p \leq 1} \mathbf{E}_p[(\delta - p)^2]$. For Bernoulli random variables and for fixed sample size n, the minimax estimator is known to be given by

$$p_n^\star = p_{\text{minimax}}^\star = \frac{S_n + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}; \tag{2.5}$$

see, for example, Lehmann and Casella (1998, pp. 311–312). Note that $p_n^\star$ is minimax in the offline context because it is Bayes with respect to the (least favorable) prior distribution Beta$(\sqrt{n}/2, \sqrt{n}/2)$ and has a constant risk or mean square error of $1/(4(\sqrt{n}+1)^2)$.

It is useful to compare the Bayes estimator $\tilde{p}_{n,a}$ from (2.4) with the minimax estimator $p_n^\star$ from (2.5). On the one hand, for a fixed sample size n, the minimax estimator $p_n^\star$ can be thought of as a special case of the Bayes estimator with $a = \sqrt{n}/2$. On the other hand, when the sample size n is variable, the estimators are fundamentally different: The minimax estimator incorporates the sample size n adaptively in the estimator itself, whereas the Bayes estimator involves a constant parameter a that can be tuned for optimization depending on the problem context.

Next, we review the well-known offline sample size formula for estimating the Bernoulli proportion p. Recall that in the offline context with a fixed sample size n, the central limit theorem gives $(\hat{p}_n - p)/\sqrt{p(1-p)/n} \approx N(0,1)$ for large n, where $\hat{p}_n$ is the MLE from (2.3). Thus, an (approximate) $100(1-\alpha)\%$ CI for p is $\hat{p}_n \pm z_{\alpha/2}\sqrt{p(1-p)/n}$. If we would like the half-width of this CI to be at most h, then the sample size n needs to satisfy

$$z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leq h, \tag{2.6}$$

so that the fixed-sample lower bound on the required sample size for the two-sided $100(1-\alpha)\%$ CI is

$$n_0 = \left\lceil p(1-p)\left(\frac{z_{\alpha/2}}{h}\right)^2 \right\rceil, \tag{2.7}$$

where $\lceil \cdot \rceil$ is the ceiling function. When we do not have any prior knowledge of p, it is often conservative to set the fixed sample size $n_{\text{cons}} = 0.25(z_{\alpha/2}/h)^2$ by using the fact that $p(1-p) \leq 0.25$ for any $0 \leq p \leq 1$. For instance, for survey polls, one typically sets $\alpha = 5\%$ and $h = 3\%$ (often called the margin of error), and thus the conservative required sample size will be $n_{\text{cons}} \approx 1068$.

Note that the conservative required fixed sample size n_{cons} depends heavily on the half-width h. In modern applications, a smaller half-width often makes sense only when the true p value is very small or very large, and this occasionally allows us to significantly reduce the sample size from the conservative required fixed sample size obtained from (2.7) when we have prior knowledge on the bounds of p. For instance, for the half-width $h = 0.01$, if we have prior knowledge that p is very small or very large, say, $p \leq 0.03$ or $p \geq 0.97$, then we can significantly reduce the required sample size from the conservative value $n_{\text{cons}} = 9604$ to $n_0 = (0.03)(0.97)(1.96/0.01)^2 \approx 1118$, which is more manageable. This is exactly the main idea in the sequential context, where we are able to update our estimate of p over time as we collect data, which in turn may occasionally allow us to identify opportunities to reduce the required sample size.

Table 1. Optimal choices of a and γ for 100(1−α)% CIs with fixed half-width h from Frey (2010).

$1-\alpha$	90%		95%		99%	
h	a	γ	a	γ	a	γ
0.10	4	0.0754	4	0.0356	6	0.0068
0.05	4	0.0859	6	0.0433	8	0.0083
0.01	8	0.0972	10	0.0487	14	0.0097

Finally, let us review the existing methods for sequential fixed-width CIs for p. The fixed-width sequential CI problem can be thought of as a special case of our proposed tandem-width CI problem when $h_0 = h_1 = h$ in (2.1). In other words, for the fixed-width $100(1-\alpha)\%$ sequential CI, we would like to find a stopping time T that (asymptotically) minimizes the ARLs, $\mathbf{E}_p(T)$, simultaneously for all $0 \le p \le 1$, subject to the CP constraint that

$$\inf_{0\le p\le 1} \mathbf{P}_p\big(p \in [\hat{p}_T - h, \hat{p}_T + h]\big) \ge 1-\alpha, \tag{2.8}$$

where $\alpha>0$ and $h>0$ are prespecified (e.g., $\alpha = 5\%$ and $h=0.1$).

In the context of sequential CIs with fixed half-width h, one often writes the sequential CI in the form $[\hat{p}_T - h, \hat{p}_T + h]$, where $\hat{p}_T$ is the MLE from (2.3). Of course, when the lower bound $\hat{p}_T - h \le 0$ or the upper bound $\hat{p}_T + h \ge 1$, we can threshold these values to 0 and 1, respectively. Note that no statistical procedure can exactly and simultaneously optimize over all $0 \le p \le 1$, and thus it is reasonable to adopt an asymptotic approach as $h, \alpha \to 0$; for example, finding a family of stopping times $T = T_{h,\alpha}$ such that $\mathbf{E}_p(T)$ is asymptotically equivalent to the fixed-sample lower bound from (2.7) at each $0 < p < 1$.

Most existing methods for fixed-width sequential CIs work with the relationship (2.6) by estimating the unknown true p carefully, especially at the early stages when few samples are available. To highlight the challenge of sequential CIs, let us estimate the unknown p from (2.6) by the MLE $\hat{p}_n$ (2.3). This will yield a naive stopping time based on Wald's CI:

$$T_{\mathrm{W}} = \inf\left\{ n \ge 1 : \frac{\hat{p}_n(1-\hat{p}_n)}{n} \le \left(\frac{h}{z_{\alpha/2}}\right)^2 \right\}. \tag{2.9}$$

Unfortunately, T_{W} from (2.9) is not effective. In fact, when $n = 1$, the MLE $\hat{p}_n = 0$ or 1; and thus T_{W} will always stop at time 1. There are many ways to improve this stopping time, say, implementing it only after taking $m_0 \ge 2$ observations or setting lower bounds on $\hat{p}_n(1-\hat{p}_n)$; but the corresponding new stopping times often require various tuning parameters and become very complicated.

Frey (2010) proposes an interesting idea to salvage (2.9) by using the Bayes estimate $\tilde{p}_{n,a}$, and this yields the stopping time

$$T_{\mathrm{F}} = \inf\left\{ n \ge 1 : \frac{\tilde{p}_{n,a}(1-\tilde{p}_{n,a})}{n} \le \left(\frac{h}{z_{\gamma/2}}\right)^2 \right\}, \tag{2.10}$$

where the parameter $\gamma = \gamma(a, h, \alpha)$ is chosen to satisfy the CP constraint from (2.8). The main advantage of Frey's method T_{F} is that it is intuitively appealing and avoids the trivial stopping scenario of (2.9). Unfortunately, in Frey's method T_F, both the decision statistic and the threshold $(h/z_{\gamma/2})^2$ depend on the tuning parameter a, which needs to be optimized according to the specific half-width h and confidence level $1-\alpha$; see Table 1 for the optimal values of a and γ recommended by Frey (2010). As a result, it is challenging to combine two fixed-width sequential CIs, each individually arising from Frey's method, together in the tandem-width sequential CI context.

3. Proposed sequential methods

For the problem of tandem-width sequential CIs, we propose to develop sequential methods by combining two efficient sequential methods that are designed for fixed-width CIs. For efficiency and easy implementation, we require that these two sequential methods for fixed-width CIs have the same decision statistics, with the only difference being the thresholds of the decision statistics. For this purpose, we will use the minimax estimator $p_n^\star$ from (2.5) to estimate the unknown p used in (2.6). This allows us to develop effective stopping times that do not involve tuning parameters.

To better present our methods, this section is divided into three parts: Section 3.1 develops our proposed stopping times for sequential CIs; Section 3.2 derives the asymptotic properties of those sequential methods; and Section 3.3 discusses finite-sample numerical issues, particularly how to accurately compute the ARL and CP properties of our proposed sequential methods by non–Monte Carlo numerical methods. This will allow us to validate our theoretical results.

3.1. Proposed stopping times

To simplify our notation, below we fix the α value in the CP constraint from (2.8) and write the proposed fixed-width and tandem-width stopping times as a function of the half-width h of the CI.

Let us begin with the proposed stopping time for a sequential $100(1-\alpha)\%$ CI with the fixed half-width h. The key idea is to apply the minimax estimator $p_n^\star$ from (2.5) to estimate p from (2.6). This motivates us to propose the following stopping time:

$$T_M(c) = \inf\left\{ n \geq 1 : \frac{p_n^\star(1-p_n^\star)}{n} \leq c \right\}, \tag{3.1}$$

where the threshold $c = c_h$ is chosen to satisfy the CP constraint from (2.8). We report the fixed-width sequential CI for p as $[\hat{p}_{T_M(c)} - h, \hat{p}_{T_M(c)} + h]$ or, more accurately, as

$$\left[\max\left(0, \hat{p}_{T_M(c)} - h\right), \min\left(1, \hat{p}_{T_M(c)} + h\right)\right].$$

An alternative way to consider confidence intervals is introduced in Mukhopadhyay and Banerjee (2014) where we adopt the *fixed-accuracy* criterion. This selection proposes a continuously variable CI with respect to the true parameter and forms a "cone of confidence" over all of the values that it can take. Even though such an approach is interesting, for our problem, when p is close to 1, it will result in the largest possible CI.

Table 2. Choices of $\gamma = \gamma(h, \alpha)$ and $c = c(h, \alpha)$ for 90%, 95%, and 99% CIs of fixed half-width h for our method.

	$1-\alpha = 90\%$		$1-\alpha = 95\%$		$1-\alpha = 99\%$	
h	γ	c	γ	c	γ	c
0.10	0.0736	3.12417×10^{-3}	0.0351	2.25210×10^{-3}	0.0051	1.27492×10^{-3}
0.09	0.0762	2.57622×10^{-3}	0.0373	1.86780×10^{-3}	0.0057	1.05982×10^{-3}
0.08	0.0801	2.08954×10^{-3}	0.0394	1.50818×10^{-3}	0.0064	8.60900×10^{-4}
0.07	0.0826	1.62629×10^{-3}	0.0412	1.17569×10^{-3}	0.0071	6.76096×10^{-4}
0.06	0.0851	1.21429×10^{-3}	0.0426	8.75656×10^{-4}	0.0078	5.08559×10^{-4}
0.05	0.0877	8.57312×10^{-4}	0.0436	6.13951×10^{-4}	0.0086	3.62106×10^{-4}
0.04	0.0901	5.56989×10^{-4}	0.0450	3.98145×10^{-4}	0.0089	2.33823×10^{-4}
0.03	0.0925	3.17985×10^{-4}	0.0462	2.26456×10^{-4}	0.0092	1.32673×10^{-4}
0.02	0.0950	1.43496×10^{-4}	0.0475	1.01844×10^{-4}	0.0095	5.94678×10^{-5}
0.01	0.0975	3.64170×10^{-5}	0.0488	2.57585×10^{-5}	0.0097	1.49495×10^{-5}

Instead, our approach treats values close to 0 and 1 similarly and switches to a more stringent CI in these two cases.

It is also important to point out that the threshold $c = c_h$ from (3.1) is an increasing function of h, in order to satisfy the CP constraint in (2.8). To see this, note that

$$\mathbf{P}_p\left(p \in \left[\hat{p}_{T_{\mathrm{M}}(c)} - h, \hat{p}_{T_{\mathrm{M}}(c)} + h\right]\right) = \mathbf{P}_p\left(\hat{p}_{T_{\mathrm{M}}(c)} \in [p - h, p + h]\right),$$

and thus the CP constraint in (2.8) implies that $\hat{p}_{T_{\mathrm{M}}(c)}$ needs to be closer to the true p with high probability for a smaller half-width h. This can only happen if the sample size $T_{\mathrm{M}}(c)$ becomes larger. Meanwhile, the stopping time $T_{\mathrm{M}}(c)$ from (3.1) or the (expected) sample size is clearly increasing as the threshold $c = c_h$ decreases. This implies that c_h is increasing in h.

For the purpose of comparison with relation (2.6) and Frey's method from (2.10), we can rewrite the threshold c from (3.1) as

$$c = c_h = \left(\frac{h}{z_{\gamma/2}}\right)^2, \tag{3.2}$$

where $\gamma = \gamma(h, \alpha)$ depends on both h and α. In the finite-sample setting, we usually have $0 < \gamma < \alpha$ due to the repeated estimation over time from (3.1), although asymptotically $\gamma/\alpha \to 1$ as $h \to 0$; see Theorem 3.1 in the next subsection. Our extensive numerical experiments also suggest that γ is a decreasing function of h (see Table 2) but, unfortunately, we have been unable to prove the result rigorously.

Now we are ready to present our proposed tandem-width sequential CI. Denote by $T_{\mathrm{M}}(c_0)$ and $T_{\mathrm{M}}(c_1)$ the stopping times $T_{\mathrm{M}}(c)$ from (3.1) with $h = h_0$ (e.g., $= 0.1$) and $h = h_1$ (e.g., $= 0.01$), respectively. Furthermore, based on the values of h_0 and h_1, we can write

$$c_0 = c_{h_0} = \left(\frac{h_0}{z_{\gamma_0/2}}\right)^2 \quad \text{and} \quad c_1 = c_{h_1} = \left(\frac{h_1}{z_{\gamma_1/2}}\right)^2, \tag{3.3}$$

where $\gamma_0 = \gamma(h_0, \alpha)$ and $\gamma_1 = \gamma(h_1, \alpha)$. At a high level, our proposed stopping time is a two-stage procedure: The first stage uses our stopping time $T_{\mathrm{M}}(c_0)$ to derive a sequential CI with a larger half-width h_0; and if the estimate $\hat{p}_{T_{\mathrm{M}}(c_0)}$ at the end of the first stage is too small or too large, then we continue to conduct the second stage by using $T_{\mathrm{M}}(c_1)$

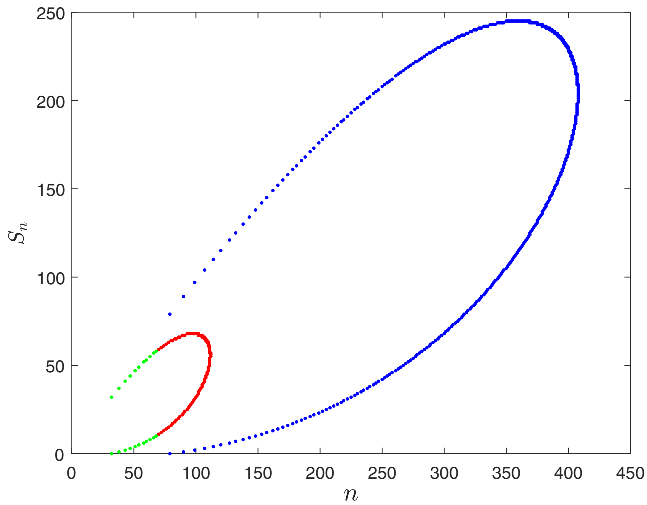


Figure 1. The stopping points for T_{TW} from (3.5) with $h_0 = 0.1, h_1 = 0.05, \gamma_0 = 0.0351, \gamma_1 = 0.0436$, and $p_0 = 0.15$.

to derive another sequential CI with a smaller half-width h_1. Note that $T_{\mathrm{M}}(c_0) \le T_{\mathrm{M}}(c_1)$, and the observations in the first stage are kept and used in $T_{\mathrm{M}}(c_1)$ in the second stage.

In other words, our proposed stopping time for the tandem-width sequential CI is defined by

$$T_{\mathrm{TW}} \equiv \begin{cases} T_{\mathrm{M}}(c_0), & \text{if } \hat{p}_{T_{\mathrm{M}}(c_0)} \in [p_0, 1-p_0]; \\ T_{\mathrm{M}}(c_1), & \text{otherwise.} \end{cases} \tag{3.4}$$

When $T_{\mathrm{TW}} = T_{\mathrm{M}}(c_0)$, we have $\hat{p}_{T_{\mathrm{M}}(c_0)} \in [p_0, 1-p_0]$, and thus we report the $100(1-\alpha)\%$ CI as the one with a larger half-width h_0; that is, $[\max(0, \hat{p}_{T_{\mathrm{M}}(c_0)} - h_0), \min(1, \hat{p}_{T_{\mathrm{M}}(c_0)} + h_0)]$. When $T_{\mathrm{TW}} = T_{\mathrm{M}}(c_1)$, we have $\hat{p}_{T_{\mathrm{M}}(c_0)} \notin [p_0, 1-p_0]$ and typically report the $100(1-\alpha)\%$ CI as the one with a smaller half-width h_1; that is, $[\max(0, \hat{p}_{T_{\mathrm{M}}(c_1)} - h_1), \min(1, \hat{p}_{T_{\mathrm{M}}(c_1)} + h_1)]$. In the finite-sample setting, it is possible, though very rare, that $\hat{p}_{T_{\mathrm{M}}(c_0)} \notin [p_0, 1-p_0]$ but $\hat{p}_{T_{\mathrm{M}}(c_1)} \in [p_0, 1-p_0]$. In such rare cases, when $T_{\mathrm{TW}} = T_{\mathrm{M}}(c_1)$, one may choose to report the $100(1-\alpha)\%$ CI by using $\hat{p}_{T_{\mathrm{M}}(c_1)}$ with a larger half-width h_0; for example, report CI as $[\max(0, \hat{p}_{T_{\mathrm{M}}(c_1)} - h_0), \min(1, \hat{p}_{T_{\mathrm{M}}(c_1)} + h_0)]$.

For the purpose of numerical computations, it is useful to rewrite T_{TW} from (3.4) as

$$T_{\mathrm{TW}} = T_{\mathrm{M}}(c_1) - \big(T_{\mathrm{M}}(c_1) - T_{\mathrm{M}}(c_0)\big) \cdot \mathbb{1}\Big\{\hat{p}_{T_{\mathrm{M}}(c_0)} \in [p_0, 1-p_0]\Big\}, \tag{3.5}$$

which allows us to investigate the properties of T_{TW} by conditioning on the sufficient statistics S_n from (2.3) when (n, S_n) is on the boundary of the stopping region of $T_{\mathrm{M}}(c_0)$.

3.2. Asymptotic properties

In this subsection, we present asymptotic properties of the proposed tandem-width sequential CI defined by the stopping time T_{TW} from (3.4), including both the asymptotic expressions of ARL and the asymptotic CP. The main theoretical challenge is to investigate the asymptotic properties of the stopping time $T_{\mathrm{M}}(c)$ from (3.1) for the fixed-width sequential CI as $h \to 0$ or, equivalently, as $c = c_h \to 0$. It is useful to point out that our technique is applicable to derive the asymptotic properties of Frey's stopping time $T_{\mathrm{F}}(a, h)$ from (2.10); this complements Frey (2010), which only reports finite-sample numerical performance results.

Let us begin with the investigation of the asymptotic properties of our proposed stopping time $T_{\mathrm{M}}(c)$ from (3.1) for the fixed-width CI — including the CP in the unconstrained scenario as the threshold $c \to 0$. This will later allow us to investigate the constrained scenario by finding c that satisfies the CP constraint in (2.8). The following theorem summarizes the main results for $T_{\mathrm{M}}(c)$.

Theorem 3.1. *As $c \to 0$, we have $cT_{\mathrm{M}}(c) \to p(1-p)$ almost surely for each $p \in (0, 1)$, and*

$$\mathbf{E}_p[T_{\mathrm{M}}(c)] = (1 + o(1))\frac{p(1-p)}{c}. \tag{3.6}$$

Moreover, denote by $\hat{p}_{T_{\mathrm{M}}}$ the MLE of p from (2.3) at time $T_{\mathrm{M}}(c)$. Then, as $c \to 0$,

$$\frac{1}{\sqrt{c}}\left(\hat{p}_{T_{\mathrm{M}}} - p\right) \to N(0, 1) \text{ in distribution,} \tag{3.7}$$

and thus an asymptotic $100(1-\alpha)\%$ CI for p is $\hat{p}_{T_{\mathrm{M}}} \pm z_{\alpha/2}\sqrt{c}$.

Before detailing the proof of this theorem, we comment on its usefulness. First, for $T_{\mathrm{M}}(c)$, if we set the half-width of the asymptotic $100(1-\alpha)\%$ CI for p to be h, then $z_{\alpha/2}\sqrt{c} = h$, and thus $c = (h/z_{\alpha/2})^2$. This justifies the form of $c = c_h$ from (3.2) and shows that $\gamma \sim \alpha$ as $h \to 0$. Moreover, for $T_{\mathrm{M}}(c_0)$, with the threshold $c_0 = \rho_0 c$ for some constant $\rho_0 > 0$, as $c_0 \to 0$, we have that $\mathbf{P}_p(\hat{p}_{T_{\mathrm{M}}(c_0)} \in [p_0, 1-p_0])$ is equal to $1 - o(1)$ if $p \in [p_0, 1-p_0]$ and $o(1)$ otherwise. When the sample sizes of these two cases are of the same order, then the $o(1)$ term will become negligible. Thus, for the proposed stopping time, $T_{\mathrm{TW}}(c)$ for the tandem-width CI, the asymptotic properties follow directly from the theorem if the thresholds c_0 and c_1 in the two stages are of the same order. Such results are summarized by the following corollary.

Corollary 3.1. *Let T_0 and T_1 denote the stopping times $T_{\mathrm{M}}(c)$ from (3.1) with the thresholds $c_0 = \rho_0 c$ and $c_1 = \rho_1 c$, respectively, for some $\rho_0 > \rho_1 > 0$. Then for the proposed stopping time $T_{\mathrm{TW}}(c)$ from (3.4), we have with probability 1 under $\mathbf{P}_p$ that as $c \to 0$,*

$$T_{\mathrm{TW}}(c) = \begin{cases} T_0, & \text{if } p \in [p_0, 1-p_0]; \\ T_1, & \text{if } p < p_0 \textit{ or if } p > 1-p_0 \end{cases} \tag{3.8}$$

and

$$\mathbf{E}_p[T_{\mathrm{TW}}(c)] = \begin{cases} (1 + o(1))\dfrac{p(1-p)}{c_0}, & \textit{if } p_0 < p < 1-p_0; \\[2ex] (1 + o(1))\dfrac{p(1-p)}{c_1}, & \textit{if } p < p_0 \textit{ or } p > 1-p_0; \\[2ex] (1 + o(1))p(1-p)\left(\dfrac{1}{2c_0} + \dfrac{1}{2c_1}\right), & \textit{if } p = p_0 \textit{ or } p = 1-p_0. \end{cases} \tag{3.9}$$

Two simple lemmas will enable us to prove the theorem. One shows that $T_{\mathrm{M}}(c)$ in Theorem 3.1 is bounded above, and the other shows that $T_{\mathrm{M}}(c)$ is bounded below. Both bounds are non-asymptotic and hold for any threshold $c > 0$.

Lemma 3.1. *For $T_{\mathrm{M}}(c)$ in Theorem 3.1, we have $T_{\mathrm{M}}(c) \le \max(1, 1/(4c))$ for any $c > 0$.*

Proof. The key idea is to note that $p_n^\star(1-p_n^\star) \le 1/4$ regardless of the value of $p_n^\star$. When $n > 1/(4c) \ge 1$, we have

$$\frac{p_n^\star(1-p_n^\star)}{n} \le \frac{1}{4n} < \frac{1}{4 \cdot \left(\frac{1}{4c}\right)} = c.$$

The lemma then follows directly from the definition of the stopping time from (3.1). □

Lemma 3.2. *For $T_{\mathrm{M}}(c)$ in Theorem 3.1, we have $T_{\mathrm{M}}(c) \ge \left(\frac{1}{8c}\right)^{2/3}$ for any $c > 0$.*

Proof. By the definition of the minimax estimator $p_n^\star$ from (2.5), an elementary argument shows that for all $n \ge 1$,

$$\frac{p_n^\star(1-p_n^\star)}{n} = \frac{S_n(n-S_n) + n\frac{\sqrt{n}}{2} + n/4}{n\left(n+\sqrt{n}\right)^2} \ge \frac{0 + n\frac{\sqrt{n}}{2} + 0}{n\left(n+\sqrt{n}\right)^2} > \frac{n\frac{\sqrt{n}}{2}}{n(n+n)^2} = \frac{1}{8}n^{-1.5}.$$

Here we use the fact that $S_n(n-S_n) \ge 0$ since $0 \le S_n = \sum_{i=1}^n X_i \le n$. Hence, whenever $n \le \left(\frac{1}{8c}\right)^{2/3}$, we have $\frac{p_n^\star(1-p_n^\star)}{n} > c$, and thus $T_{\mathrm{M}}(c)$ will not stop at time n. This proves the lemma. □

Remark 3.1. Lemmas 3.1 and 3.2 provide nonasymptotic bounds that allow us to prove the asymptotic results in Theorem 3.1 as $c \to 0$ for our stopping time $T_{\mathrm{M}}(c)$. However, these results also apply to Frey's procedure $T_{\mathrm{F}}(c, a)$ from (2.10). In particular, by the elementary arguments in Lemmas 3.1 and 3.2, we can show that for $a > 0$ and $c > 0$,

$$\sqrt{\frac{a}{c}} - 2a \le T_{\mathrm{F}}(c, a) \le \max\left(1, \frac{1}{4c}\right), \tag{3.10}$$

which results in $T_{\mathrm{F}}(c, a) \to \infty$ almost surely as $c \to 0$.

Given the nonasymptotic bounds in Lemmas 3.1 and 3.2, we are now ready to prove the asymptotic results in Theorem 3.1 as $c \to 0$.

Proof of Theorem 3.1. By Lemma 3.2, as $c \to 0$, we have $T_{\mathrm{M}}(c) \to \infty$ with probability 1. To find an accurate asymptotic expression of $T_{\mathrm{M}}(c)$, it is useful to rewrite its stopping rule in terms of the MLE $\hat{p}_n$ from (2.3), whose asymptotic properties are well known. A simple mathematical argument shows that

$$\frac{p_n^\star(1-p_n^\star)}{n} = \frac{p(1-p) + \frac{1}{2\sqrt{n}} + \frac{1}{4n} + \left(\hat{p}_n(1-\hat{p}_n) - p(1-p)\right)}{n\left(1 + \frac{1}{\sqrt{n}}\right)^2}. \tag{3.11}$$

At a high level, the proof is based on two disjoint events related to $\hat{p}_n$, depending on how close the term from (3.11) is to $p(1-p)/n$. By the law of large numbers, for any given $0 < p < 1$, the term from (3.11) is asymptotically equivalent to $p(1-p)/n$ with probability that tends to 1 for large n. This turns out to capture the first-order asymptotic analysis, as the corresponding complement event is negligible, since $T_{\mathrm{M}}(c)$ is bounded from above by Lemma 3.1.

Below is the detailed, rigorous proof. Fix $0<p<1$ and $\epsilon>0$. Then there exists an integer $n_\epsilon>0$ such that for all $n\ge n_\epsilon$,

$$\left(1+\frac{1}{\sqrt{n}}\right)^2 \le 1+\epsilon. \tag{3.12}$$

Furthermore, all $n\ge n_\epsilon$, denote the event

$$\mathcal{A}_{n,\epsilon}=\left\{\left|\frac{1}{2\sqrt{n}}+\frac{1}{4n}+\left(\hat{p}_n\left(1-\hat{p}_n\right)-p(1-p)\right)\right|\le \epsilon\cdot p(1-p)\right\}. \tag{3.13}$$

First consider the case when the event $\mathcal{A}_{n,\epsilon}$ does not hold. Then for $\epsilon>0$ and $\delta>0$, the weak law of large numbers implies that there exists $n_{\epsilon,\delta}>0$ such that for $n\ge n_{\epsilon,\delta}$, we have $\mathbf{P}(\mathcal{A}^c_{n,\epsilon})<\delta$. Moreover,

$$\mathbf{E}[T_{\mathrm{M}}(c)]=\mathbf{E}\left[T_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right]+\mathbf{E}\left[T_{\mathrm{M}}(c);\mathcal{A}^c_{n,\epsilon}\right], \tag{3.14}$$

Wherein the previous equation and throughout the proof, we use $\mathbf{E}[T_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}]$ to denote $\mathbf{E}[T_{\mathrm{M}}(c)]$ when the event $\mathcal{A}_{n,\epsilon}$ holds. By Lemma 3.1, $T_{\mathrm{M}}(c)\le 1/(4c)$, so

$$\mathbf{E}\left[T_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right]\le \mathbf{E}[T_{\mathrm{M}}(c)]\le \mathbf{E}\left[T_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right]+\frac{1}{4c}\mathbf{P}\left(\mathcal{A}^c_{n,\epsilon}\right)<\mathbf{E}\left[T_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right]+\frac{1}{4c}\delta. \tag{3.15}$$

Now, we prove the case when the event $\mathcal{A}_{n,\epsilon}$ is true. In this case, a combination of (3.11) and the fact that the event $\mathcal{A}_{n,\epsilon}$ holds yields that for all $n\ge n_\epsilon$,

$$\frac{(1-\epsilon)p(1-p)}{(1+\epsilon)n}\le\frac{p^\star_n(1-p^\star_n)}{n}\le\frac{(1+\epsilon)p(1-p)}{n}. \tag{3.16}$$

Note that such n_ϵ might depend on p and ϵ, but relation (3.16) holds for all $n\ge n_\epsilon$. Now by Lemma 3.2, there exists a $c^*>0$ such that for all $c\le c^*$, we have $T_{\mathrm{M}}(c)\ge n_\epsilon+1$ and thus relation (3.16) holds for both $n=T_{\mathrm{M}}(c)$ and $n=T_{\mathrm{M}}(c)-1$.

By the definition from (3.1), when $n=T_{\mathrm{M}}(c)$, we have $\frac{p^\star_n(1-p^\star_n)}{n}\le c$. Combining this with the first inequality in (3.16) for $n=T_{\mathrm{M}}(c)$ yields that for all $c\le c_\epsilon$,

$$\frac{(1-\epsilon)p(1-p)}{(1+\epsilon)T_{\mathrm{M}}(c)}\le c \qquad \text{or, equivalently,} \qquad cT_{\mathrm{M}}(c)\ge\frac{1-\epsilon}{1+\epsilon}p(1-p).$$

Letting $c\to 0$ yields

$$\liminf_{c\to 0}\left\{cT_{\mathrm{M}}(c)\right\}\ge\frac{1-\epsilon}{1+\epsilon}p(1-p)$$

with probability $1-\delta$ for any given $\epsilon>0$. Now the inf-limit on the left-hand side does not depend on ϵ. Thus, letting $\epsilon\to 0$, we have, with probability $1-\delta$,

$$\liminf_{c\to 0}\left\{cT_{\mathrm{M}}(c)\right\}\ge p(1-p). \tag{3.17}$$

On the other hand, by the definition from (3.1), when $n=T_{\mathrm{M}}(c)-1$, we have $\frac{p^\star_n(1-p^\star_n)}{n}>c$. Combining this with the second inequality in (3.16) yields that for all $c\le c_\epsilon$,

$$c < \frac{(1+\epsilon)p(1-p)}{T_{\mathrm{M}}(c)-1} \qquad \text{or, equivalently,} \qquad cT_{\mathrm{M}}(c) < (1+\epsilon)p(1-p) + c$$

with probability $1-\delta$. Letting $c \to 0$, we have

$$\lim_{c\to 0}\sup\{cT_{\mathrm{M}}(c)\} \le (1+\epsilon)p(1-p)$$

for any $\epsilon > 0$, which results in

$$\lim_{c\to 0}\sup\{cT_{\mathrm{M}}(c)\} \le p(1-p) \tag{3.18}$$

with probability $1-\delta$. Combining (3.17) and (3.18), we obtain

$$\lim_{c\to 0}\{cT_{\mathrm{M}}(c)\} = p(1-p)$$

with probability $1-\delta$. By Lemma 3.1 and Lebesgue's dominated convergence theorem, we have

$$\lim_{c\to 0}\mathbf{E}_p\left[cT_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right] = \mathbf{E}_p\left[\lim_{c\to 0} cT_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right] = p(1-p),$$

and so

$$\mathbf{E}_p\left[T_{\mathrm{M}}(c);\mathcal{A}_{n,\epsilon}\right] = (1+o(1))\frac{p(1-p)}{c}. \tag{3.19}$$

Now, letting $\delta \to 0$ and using (3.19),

$$(1+o(1))\frac{p(1-p)}{c} \le \mathbf{E}[T_{\mathrm{M}}(c)] \le (1+o(1))\frac{p(1-p)}{c} + o(1/c) = (1+o(1))\frac{p(1-p)}{c}. \tag{3.20}$$

This proves (3.6).

To prove (3.7), a crucial step is to define an integer-valued constant $m = m_c = \lfloor p(1-p)/c \rfloor$ as $c \to 0$. On the one hand, by the central limit theorem, $\sqrt{m}(\hat{p}_m - p)/\sqrt{p(1-p)} = (\hat{p}_m - p)/\sqrt{c}$ is asymptotically normally $N(0, 1)$ distributed, as $c \to 0$. On the other hand, for $T = T_{\mathrm{M}}(c)$, we just showed that $T/m \to 1$ almost surely. By Equation (2.43) in Theorem 2.40 of Siegmund (1985, p. 23), we have

$$\sqrt{m}\left(\hat{p}_T - \hat{p}_m\right) \to 0 \text{ in probability.} \tag{3.21}$$

Combining these two results together yields (3.7), thus completing the proof of the theorem. □

3.3. Finite-sample numerical computation

In this subsection, we discuss the numerical computation of the finite-sample performance properties of our proposed stopping times $T_{\mathrm{M}}(c)$ from (3.1) and $T_{\mathrm{TW}}(c)$ from (3.8), including the ARL, $\mathbf{E}_p[T]$, and the CP, $\mathbf{P}_p(|\hat{p}_T - p| \le h)$, at each p. This allows us to validate the asymptotic properties of our stopping times from the previous subsection as well as compare properties of different methods.

For a given stopping time T and its corresponding sequential CI, there are two approaches to compute its finite-sample properties, $\mathbf{E}_p[T]$ and $\mathbf{P}_p(|\hat{p}_T - p| \le h)$ for all

$0 < p < 1$. The first one is an approximate Monte Carlo method based on repeated random sampling of Bernoulli(p) random variables for each $0 < p < 1$. It is straightforward to implement such a Monte Carlo method, although it is very time consuming to obtain accurate estimates of the ARL or CP properties over the whole interval $p \in (0, 1)$, especially when the true p is close to 0 or 1. The second approach is an accurate non–Monte Carlo numerical method based on the path-counting ideas in Girshick et al. (1946) and Schultz et al. (1973); see also Frey (2010). This non–Monte Carlo numerical method is validated against the Monte Carlo method, and both yield the same results.

Let us provide a more-detailed discussion on the accurate non–Monte Carlo numerical method. Note that $S_n = \sum_{i=1}^n X_i = S_{n-1} + X_n$ is a sufficient statistic for the Bernoulli proportion p, and conditional on S_{n-1}, the value of S_n has only two choices: S_{n-1} or $S_{n-1} + 1$, depending on whether $X_n = 0$ or 1. Then, the key idea of the non–Monte Carlo numerical method is to count the number of paths, denoted by $H(a, n)$, from $S_0 = 0$ at time 0 to $S_n = a$ at time n without hitting any earlier stopping boundaries of T before time n. For many reasonable stopping times T, including our proposed stopping time $T = T_{\mathrm{M}}(h)$, the stopping points/boundaries of T can be written as the set of discrete points, $(S_{n_1} = a_1, n_1), \ldots, (S_{n_k} = a_k, n_k)$, for some (possibly large) $k \geq 1$. Furthermore, in our proposed stopping time and many other stopping times, $(S_{n_i} = a, n_i)$ is a stopping point if and only if $(n_i - a, n_i)$ is also a stopping point, due to the fact that the problem is symmetric at $p = 1/2$. Then, when the stopping time T stops at time n_i with the observed value $S_{n_i} = a_i$—that is, when (a_i, n_i) is a stopping point—we estimate p by $\hat{p}_T = \hat{p}_i = a_i/n_i$ and report the confidence interval as $[\max(0, \hat{p}_T - h), \min(1, \hat{p}_T + h)]$.

Now once we have counted the number $H(a_i, n_i)$ of sample paths from (0, 0) to $(S_n = a_i, n = n_i)$ without hitting any earlier stopping regions for all stopping points of T, we can compute the finite-sample properties of T simultaneously for all p by

$$\mathbf{P}_p\left(|\hat{p}_T - p| \leq h\right) = \sum_{i=1}^{k} H(a_i, n_i) p^{a_i} (1-p)^{n_i - a_i} \mathbb{1}\left\{|p - \hat{p}_i| \leq h\right\}, \tag{3.22}$$

and

$$\mathbf{E}_p[T] = \sum_{i=1}^{k} H(a_i, n_i) p^{a_i} (1-p)^{n_i - a_i} n_i. \tag{3.23}$$

Numerically, we can use (3.22) and (3.23) to compute $\mathbf{P}_p(|\hat{p}_T - p| \leq h)$ and $\mathbf{E}_p[T]$ as a function of p as p varies from 0 to 1 (or to 1/2 due to symmetric properties) with a small step size.

For each threshold c or tuning parameter γ from (3.2), we will be able to derive the corresponding finite-sample properties, $\mathbf{E}_p[T]$ and $\mathbf{P}_p(|\hat{p}_T - p| \leq h)$, of our proposed stopping times $T = T_{\mathrm{M}}(c)$ from (3.1) or $T = T_{\mathrm{TW}}(c)$ from (3.8) for all $0<p<1$. To satisfy the $1-\alpha$ CP constraints from (2.1) or (2.8), we propose to use the bisection search method to obtain the desired threshold c or γ.

We split the remainder of this subsection into two parts: (a) the numerical computation of the finite-sample properties of $T = T_{\mathrm{M}}(c)$ from (3.1) and (b) the numerical computation of the finite-sample properties of $T = T_{\mathrm{TW}}(c)$ from (3.8). The latter part uses the numerical computations of part (a) but is more involved in computation because the stopping region for the tandem method involves two stopping regions, one from the first stage using h_0 and another from the second stage using h_1.

3.3.1. Finite-sample properties of $T_{\mathrm{M}}(c)$

Let us first focus on how to count the number of paths for a stopping time T such as $T = T_{\mathrm{M}}(c)$ from (3.1) whose stopping region boundary is *convex*. Without loss of generality, assume that the stopping time is defined as $T = \inf\{n \geq 1 : S_n \in \mathcal{R}_n\}$, where $\mathcal{R}_n = \mathcal{R}_n(\gamma)$ is the stopping region at time n. Note that $0 \leq S_n \leq n$ for all $n \geq 1$. Now for each n and each possible value $S_n = a$, we define two functions: (i) the indicator function $I(a, n) = 1$ if $S_n = a$ is an interior (non-stopping) point at time n and $I(a, n) = 0$ if $S_n = a$ belongs to the stopping region $\mathcal{R}_n$, and (ii) is the counting function $H(a,\ n)$ that denotes the number of ways to get to $S_n = a$ successes at time step n without hitting any earlier stopping regions $\mathcal{R}_k$s at times $1 \leq k \leq n{-}1$. Note that $H(0, 1) = H(1, 1) = 1$, since we only have one way to obtain $S_1 = 0$ or 1 at time $n = 1$.

To compute the counting function $H(S_n = a, n)$ in general, note that $S_{n-1} = a$ or $a - 1$ if $S_n = a$, depending on whether $X_n = 1$ or 0, and thus the number of path counts for points $(S_n = a, n)$ can be computed by the number of paths to either $(S_{n-1} = a, n{-}1)$ or $(S_{n-1}{-}1 = a{-}1, n{-}1)$, when at least one of them is an interior (non-stopping) point. In other words, the counting function $H(S_n = a, n)$ can be recursively computed by:

$$H(a, n) = H(a, n{-}1)I(a, n{-}1) + H(a{-}1, n{-}1)I(a{-}1, n{-}1). \tag{3.24}$$

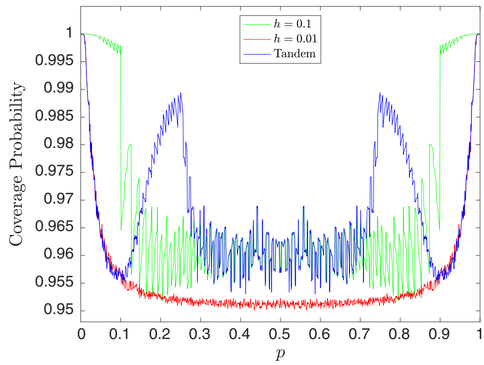
For the purpose of numerical computation, the value $H(a,\ n)$ can be large for large n and, in such case, this recursion can be implemented on the log scale to avoid overflow problems by using the equality $\log(c + d) = \log c + \log(1 + \exp(\log d - \log c))$.

Table 2 presents the numerical values of γ for different choices of α and h that guarantee that the coverage probability of the confidence interval is at least $1{-}\alpha$.

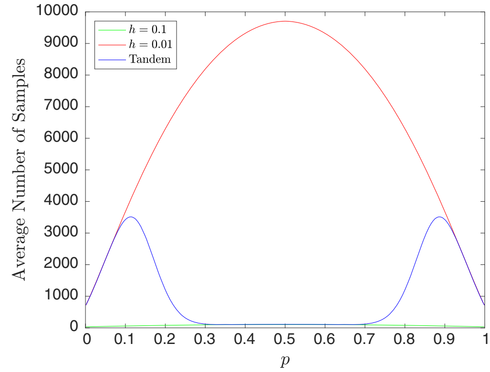
3.3.2. Finite-sample properties of $T_{\mathrm{TW}}(c)$

It is much more challenging to count the number of paths for the stopping $T_{\mathrm{TW}}(c)$ from (3.8) for tandem-width sequential CIs, since its stopping region boundary is *non-convex*.

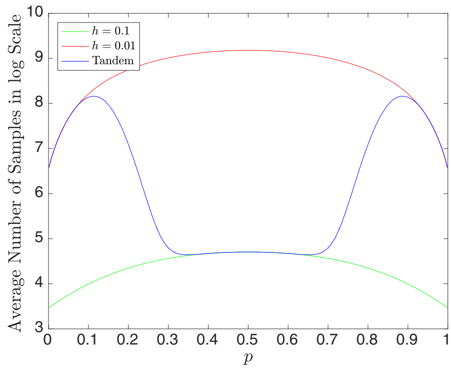
To better illustrate the challenges, consider Figure 1, which plots the stopping points for our proposed tandem method with $h_0 = 0.1, h_1 = 0.05, \gamma_0 = 0.0351, \gamma_1 = 0.0436$, and $p_0 = 0.15$. Equivalently, $c_0 = 2.2521 \times 10^{-3}$ and $c_1 = 2.5758 \times 10^{-5}$. The stopping points in red represent the stopping points for $T_{\mathrm{M}}(c_0)$ when $h_0 = 0.1$ and $\hat{p}_{T_{\mathrm{M}}(c_0)} \in [p_0, 1{-}p_0]$. This means that if we hit these red stopping points, then we stop sampling and report the $100(1{-}\alpha)$% CI as $[\max(0, \hat{p}_{T_{\mathrm{M}}(c_0)}{-}h_0), \min(1, \hat{p}_{T_{\mathrm{M}}(c_0)} + h_0)]$. However, if we do not hit these points in the first stage and instead hit the green stopping points for $T_{\mathrm{M}}(c_0)$ where $\hat{p}_{T_{\mathrm{M}}(c_0)} \notin [p_0, 1{-}p_0]$, then we need to keep on sampling until we reach the blue stopping points for $T_{\mathrm{M}}(c_1)$ and report the $100(1{-}\alpha)$% CI as $[\max(0, \hat{p}_{T_{\mathrm{M}}(c_1)}{-}h_1), \min(1, \hat{p}_{T_{\mathrm{M}}(c_1)} + h_1)]$. As a result, the stopping region boundary of $T_{\mathrm{TW}}(c)$ from (3.8) consists of both red and blue stopping times, which form a non-convex set. The good news is that this non-convex set is the difference of two convex boundaries, which allows us to simplify the computations.



(a) Coverage Probability

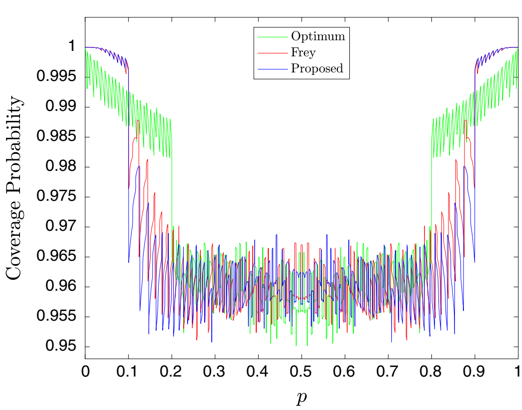


(b) Expected Sample Size

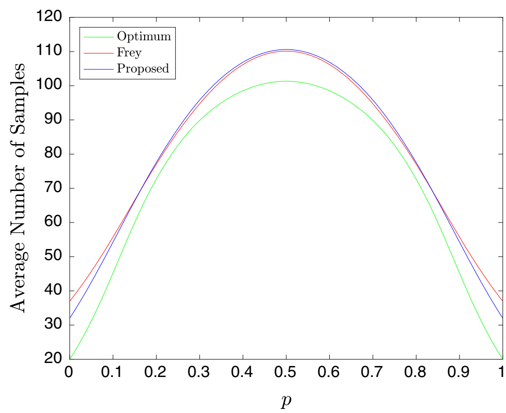


(c) Expected Sample Size in log Scale

Figure 2. A comparison of coverage probability and average run length for three sequential methods: (i) our proposed tandem-width CI with $h_0 = 0.01$ and $h_1 = 0.1$ (blue line); (ii) our proposed fixed-width CI with $h = h_0 = 0.01$ (red line); and (iii) our proposed fixed-width CI with $h = h_1 = 0.1$ (green line).

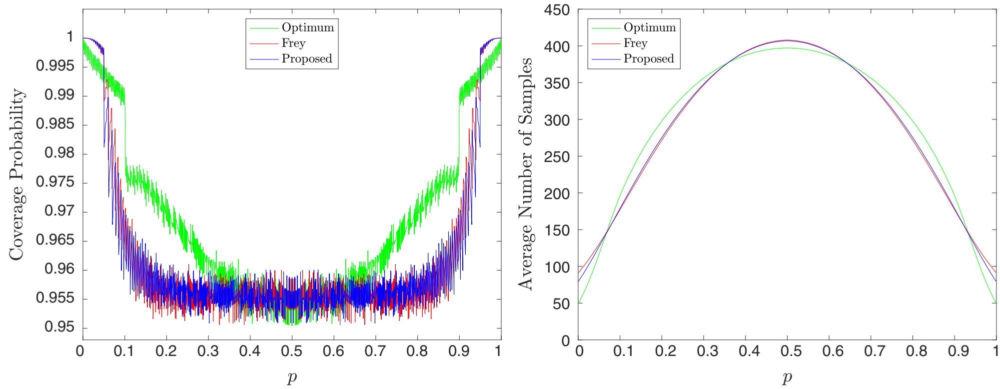


(a) Coverage Probability



(b) Expected Sample Size

Figure 3. A comparison of coverage probabilities and expected sample sizes of three methods (Optimum, Frey, and Proposed), for $h = 0.1$.



(a) Coverage Probability

(b) Expected Sample Size

Figure 4. A comparison of coverage probabilities and expected sample sizes of three methods (Optimum, Frey, and Proposed), for $h = 0.05$.

To be more concrete, we use the definition of our tandem stopping time from (3.5) to split the CP and ARL for the tandem procedure into three parts as follows. First, we compute CP and ARL achieved by using $T_M(c_1)$, hitting the blue stopping points when the blue region is the only stopping region. Second, compute CP and ARL achieved by hitting the red stopping points; that is, the stopping points for the first stage where we stop sampling using the equations in Section 3.3.1. The third part is the more demanding part, as we need to compute the number of ways to hit the blue stopping points starting from the red stopping points without hitting any stopping points in the interim process. We start the recursion (3.24) from each red stopping point as the origin and continue recursively until we hit the blue region. Then, we finish the third step by computing the CP and ARL as from (3.22) and (3.23) but with the modified number of ways reaching these blue points. The CP and ARL for the tandem procedure can be combined by adding the CP and ARL from the first and third parts and subtracting the second part.

4. Numerical studies

In this section, we report on numerical study results to further demonstrate the usefulness of our proposed stopping times. In Section 4.1, we illustrate the performance of the tandem-width stopping time T_{TW} from (3.4). In Section 4.2, we compare our proposed fixed-width stopping time T_M from (3.1) with Frey's method T_F from (2.10) that involves an additional tuning parameter of the Bayes prior.

4.1. Tandem-width CI

Suppose that we are interested in deriving a 95% tandem-width sequential CI with half-width $h_0 = 0.1$ if $\hat{p} \in [p_0, 1-p_0]$ for $p_0 = 0.15$ and with half-width $h_1 = 0.01$ if $\hat{p} < p_0 = 0.15$ or $> 1-p_0 = 0.85$. For our proposed tandem-width CI method, two threshold values are $c_0 = 2.2521 \times 10^{-3}$ and $c_1 = 2.5758 \times 10^{-5}$ or, equivalently, $\gamma_0 = 0.0351$ and $\gamma_1 = 0.0488$ based on Table 2. Next, we obtain the the coverage probability and ARL through simulation, with 500,000 replications at each value of $p = 0.001, 0.002, \ldots, 0.999$. Note that we could also use the path-counting ideas in the previous section to obtain CP and ARL analytically, but in this case it is easier to verify our results through simulation. We report the estimate of $\mathbf{P}_p(|\hat{p}_{T_{\mathrm{TW}}} - p| \le h)$ as the number of instances that p is within the reported confidence interval divided by the total number of replications. Furthermore, we report the estimate of $\mathbf{E}_p[T_{\mathrm{TW}}]$ as the average run length over each replication for each value of p. In Figure 2, we compare the tandem-width CI simulation results (blue line) versus the analytical results (obtained from the finite-sample numerical computational methods in Section 3.3) of the fixed-width CI based on $p_n^\star$ obtained with $h = 0.1$ (green line) and $h = 0.01$ (red line).

We notice that by not choosing to use a fixed-width CI for $h = 0.01$, such as that based on T_{M} from (3.1), we can save in the worst case about 60% of the sampling cost and time if we are willing to report a $100(1-\alpha)\%$ CI for p with larger half-width $h = 0.1$ when p is not close to 0 or close to 1. This savings in sampling cost becomes more obvious as we get closer to $p = 0.5$. This illustrates the importance of our tandem-width methodology, because when resources are scarce or when no historical data are available to gain prior knowledge about p, then we do not need to spend so much time to report a very accurate CI with a very small half-width when p is close to 1/2.

Now that we illustrated the usefulness of our tandem-width methodology, we compare the performance of the minimax-based method from (3.1) versus Frey's method from (2.10).

4.2. Fixed-width CI comparisons

In this subsection, we compare our proposed fixed-width method with Frey's method T_{F} from (2.10), and with the optimum scheme in our earlier work in Yaacoub et al. (2019). Using the numerical iterations from Section 3.3, we calculate numerically $\mathbf{P}_p(|\hat{p}_T - p| \le h)$ and $\mathbf{E}_p[T]$ for $p = 1/2001, 2/2001, \ldots, 2000/2001$. Note that the requirement is to be able to guarantee a minimal coverage probability *for all* p. Therefore, parameters were selected so that all competing schemes guaranteed *the same worst-case coverage probability;* that is, coverage of at least $1-\alpha$ for all p. Here the tuning parameter γ is chosen from Table 2 for our method and from Table 1 for Frey's method T_{F}. The optimum scheme in Yaacoub et al. (2019) requires two tuning parameters: one is the parameter u that sets the Beta(u, u) as the prior distribution of p, and the other is the parameter κ for the cost per observation. Here, the choice of $u = 1$ (uniform prior) and cost $\kappa = 0.00097$ will satisfy the coverage probability constraint in Yaacoub et al. (2019) for $\alpha = 0.05$ and $h = 0.1$.

In Figure 3(a), we plot the coverage probability for each method versus p and in Figure 3(b) we plot the corresponding average sample size required to obtain this performance for $\alpha = 0.05$ and $h = 0.1$. We can draw the following conclusions from the figures: Our proposed scheme and Frey's require about the same sample sizes for most values of p, although our fixed-width scheme is slightly more parsimonious when p is close to 0 or 1. Moreover, the two procedures exhibit similar coverage probability profiles. The optimum scheme in Yaacoub et al. (2019) is the best in terms of the smallest number of samples to guarantee the worst-case CP of at least 0.95.

We also ran numerical experiments for many other combinations of (α, h), and we make similar conclusions. For instance, in Figure 4(a), we plot the coverage probability for each method versus p, and in Figure 4(b) we plot the corresponding average sample size required to obtain this performance for $\alpha = 0.05$ and $h = 0.05$. Our proposed method and Frey's method perform almost identically, whereas the optimum method has a smaller sample size and larger coverage probability if the true p is not too close to 0 or 1. Notice that the behavior of the optimum scheme differs between different values of h. For instance, for $h = 0.1$, the optimum scheme has a lower expected sample size than both methods, whereas for the case of $h = 0.05$ the expected sample size of the optimum scheme is sometimes larger than those of both methods, even though in such cases the coverage probability is larger. One possible explanation for this phenomenon is that the optimum scheme puts more weights on the expected sample size when h is larger but more weight on CP when h is smaller. However, we are unable to prove such a claim.

We should emphasize that the optimum scheme in our earlier work (Yaacoub et al., 2019) becomes computationally expensive as h gets smaller—for example, $h = 0.01$—as it involves dynamic programming and involves matrices of dimension of order $1/h^2$; see Yaacoub et al. (2019). For the fixed half-width $h = 0.01$, the performance of our fixed-width method and Frey's is also similar, although Frey's method gives a slightly smaller (i.e., better) ARL, whereas our proposed method gives a slightly larger coverage probability.

In summary, as compared to Frey's method that needs to optimize the tuning parameter for the Bayes prior, our proposed method has similar finite-sample properties but is much simpler to implement since the minimax estimator does not involve any tuning parameters. In other words, our new tandem-width sequential CIs is a simple but useful method for fixed-width sequential CIs that is fast and efficient with performance characteristics that are comparable to or only slightly worse than those of the optimum scheme.

5. Conclusions

We proposed two sequential schemes for obtaining confidence intervals for a binomial proportion p using the minimax estimator of p: a fixed-width scheme, and a tandem-width scheme. We also established upper and lower bounds for our stopping times, presented their asymptotic properties, and compared our proposed schemes with other existing methods. We found that our proposed sequential schemes are computationally simple and also enjoy nice theoretical properties.

Our proposed tandem-width method can be extended in a couple of different directions. First, it can be extended to three or more half-widths or stages, allowing better flexibility with the choice of half-widths based on the true value of p. For instance, one may prefer a half-width size of (i) $h_0 = 0.10$ if the true $p \in [0.4, 0.6]$, (ii) $h_1 = 0.05$ if $p \in [0.1, 0.4)$ or $p \in (0.6, 0.9]$, or (iii) $h_2 = 0.01$ if $p < 0.1$ or $p > 0.9$. In such cases, we can extend our proposed stopping time in (3.4) to develop a three-stage stopping time, depending on the value of $\hat{p}$ at the end of each stage. Moreover, it is of interest to combine our method with a proportional accuracy (aka relative-width) CI, where the half-width h is a function of p; for example, $h = h(p) = \eta p$ for some $\eta \in (0, 1)$. This allows us to overcome a disadvantage of relative-width schemes that often become very costly for small p.

Acknowledgment

The authors thank the Editor for his thoughtful and constructive comments that greatly improved the quality and presentation of this article.

Funding

This research was supported in part by National Science Foundation (NSF) grants CMMI-1233141, CMMI-1362876, DMS-1613258, and DMS-1830344, through the Georgia Institute of Technology, and in part by NSF grant CIF-1513373, through Rutgers University.

References

Abramson, J. S., Takvorian, R. S., Fisher, D. C., Feng, Y., Jacobsen, E. D., Brown, J. R., Barnes, J. A., Neuberg, D. S., and Hochberg, E. P. (2013). Oral Clofarabine for Relapsed/Refractory Non-Hodgkin Lymphomas: Results of a Phase 1 Study, *Leukemia & Lymphoma* 54: 1915–1920.

Agresti, A. and Coull, B. A. (1998). Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions, *American Statistician* 52: 119–126.

Armitage, P. (1958). Numerical Studies in the Sequential Estimation of a Binomial Parameter, *Biometrika* 45: 1–15.

Banerjee, S. and Mukhopadhyay, N. (2015). A General Sequential Fixed-Accuracy Confidence Interval Estimation Methodology for a Positive Parameter: Illustrations Using Health and Safety Data, *Annals of Institute of Statistical Mathematics* 68: 541–570.

Blyth, C. R. and Still, H. A. (1983). Binomial Confidence Intervals, *Journal of American Statistical Association* 78: 108–116.

Brown, L. D., Cai, T., and Dasgupta, A. (2001). Interval Estimation for a Binomial Proportion, *Statistical Science* 16: 101–133.

Brown, L. D., Cai, T., and Dasgupta, A. (2002). Interval Estimation for a Binomial Proportion and Asymptotic Expansions, *Annals of Statistics* 30: 160–201.

Chow, Y. S. and Robbins, H. (1965). On the Asymptotic Theory of Fixed Width Sequential Confidence Intervals for the Mean, *Annals of Mathematical Statistics* 36: 457–462.

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.

Crow, E. L. (1956). Confidence Intervals for a Proportion, *Biometrika* 43: 423–435.

Frey, J. (2010). Fixed-Width Sequential Confidence Intervals for a Proportion, *American Statistician* 64: 242–249.

Girshick, M. A., Mosteller, F., and Savage, L. J. (1946). Unbiased Estimates for Certain Binomial Sampling Problems with Applications, *Annals of Mathematical Statistics* 17: 13–23.

Huber, M. (2017). A Bernoulli Mean Estimate with Known Relative Error Distribution, *Random Structures & Algorithms* 50: 173–182.

Jegourel, C., Sun, J., and Dong, J. S. (2017). Sequential Schemes for Frequentist Estimation of Properties in Statistical Model Checking, in *Quantitative Evaluation of Systems (QEST 2017). Lecture Notes in Computer Science*, N. Bertrand and L. Bortolussi, eds., pp. 333–350, Cham: Springer.

Khan, R. A. (1969). A General Method for Determining Fixed-Width Confidence Intervals, *Annals of Mathematical Statistics* 40: 704–709.

Khan, R. A. (1998). Fixed Width Confidence Sequences for the Normal Mean and the Binomial Probability, *Sequential Analysis* 17: 205–217.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. New York: Springer.

Malinovsky, Y. and Zacks, S. (2018). Proportional Closeness Estimation of Probability of Contamination under Group Testing, *Sequential Analysis* 37: 145–157.

Morisette, J. T. and Khorram, S. (1998). Exact Binomial Confidence Interval for Proportions, *Photogrammetric Engineering & Remote Sensing* 64: 281–283.

Mukhopadhyay, N. and Banerjee, S. (2014). Purely Sequential and Two-Stage Fixed-Accuracy Confidence Interval Estimation Methods for Count Data from Negative Binomial Distributions in Statistical Ecology: One-Sample and Two-Sample Problems, *Sequential Analysis* 33: 251–285.

Mukhopadhyay, N. and Banerjee, S. (2015). Purely Sequential and Two-Stage Bounded-Length Confidence Intervals for the Bernoulli Parameter with Illustrations from Health Studies and Ecology, in *Ordered Data Analysis, Modeling and Health Research Methods, in Honor of H. N. Nagaraja's 60th Birthday*, P. Choudhary, C. Nagaraja, and H. K. T. Ng, eds., pp. 211–234, New York: Springer.

Mukhopadhyay, N. and De Silva, B. M. (2009). *Sequential Methods and Their Applications*, Boca Raton: CRC.

Nadas, A. (1969). An Extension of a Theorem of Chow and Robbins on Sequential Confidence Intervals for the Mean, *Annals of Mathematical Statistics* 40: 667–671.

Newcombe, R. G. (1998). Two-sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods, *Statistics in Medicine* 17: 857–872.

Reiczigel, J. (2003). Confidence Intervals for the Binomial Parameter: Some New Considerations, *Statistics in Medicine* 22: 611–621.

Robbins, H. and Siegmund, D. (1974). Sequential Estimation of p in Bernoulli trials, in *Studies in Probability and Statistics*, E. J. G. Pitman and E. J. Williams, eds., pp. 103–107, Jerusalem: Jerusalem Academic Press.

Schultz, J. R., Nichol, F. R., Elfring, G. L., and Weed, S. D. (1973). Multiple-Stage Procedures for Drug Screening, *Biometrics* 29: 293–300.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer.

Sterne, T. E. (1954). Some Remarks on Confidence of Fudicial Limits, *Biometrika* 41: 275–278.

Sullivan, A., Raben, D., Reekie, J., Rayment, M., Mocroft, A., Esser, S., Leon, A., Begovac, J., Brinkman, K., Zangerle, R., Grzeszczuk, A., Vassilenko, A., Hadziosmanovic, V., Krasnov, M., Sonnerborg, A., Clumeck, N., Gatell, J., Gazzard, B., Monforte, A. D., Rockstroh, J., and Lundgren, J. D. (2013). Feasibility and Effectiveness of Indicator Condition-Guided Testing for HIV: Results from HIDES I (HIV Indicator Diseases across Europe Study), *PLoS ONE* 8: e52845.

Tanaka, M. (1961). On a Confidence Interval for Given Length for the Parameter of the Binomial and the Poisson Distributions, *Annals of Institute of Statistical Mathematics* 13: 201–215.

Vollset, S. E. (1993). Confidence Intervals for a Binomial Proportion, *Statistics in Medicine* 12: 809–824.

Wilson, E. B. (1927). Probable Inference, the Law of Succession, and Statistical Inference, *Journal of American Statistical Association* 22: 209–212.

Yaacoub, T., Moustakides, G. V., and Mei, Y. (2019). Optimal Stopping for Interval Estimation in Bernoulli Trials, *IEEE Transactions on Information Theory* 65: 3022–3033.

Zacks, S. (1966). Sequential Estimation of the Mean of a Log-Normal Distribution Having a Prescribed Proportional Closeness, *Annals of Mathematical Statistics* 37: 1439–1888.

Zacks, S. (2017). *Sample Path Analysis and Distributions of Boundary Crossing Times*, Cham: Springer.

Zacks, S. and Mukhopadhyay, N. (2007). Distributions of Sequential and Two-Stage Stopping Times for Fixed-Width Confidence Intervals in Bernoulli Trials: Application in Reliability, *Sequential Analysis* 26: 425–441.