Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2013, Article ID 603297, 24 pages
http://dx.doi.org/10.1155/2013/603297

Research Article

Exact Group Sequential Methods for Estimating a Binomial Proportion

Zhengjia Chen[1] and Xinjia Chen[2]

[1] *Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322, USA*
[2] *Department of Electrical Engineering, Southern University and A&M College, Baton Rouge, LA 70813, USA*

Correspondence should be addressed to Zhengjia Chen; zchen38@emory.edu

Received 23 May 2012; Revised 1 October 2012; Accepted 9 October 2012

Academic Editor: Xuelin Huang

We first review existing sequential methods for estimating a binomial proportion. Afterward, we propose a new family of group sequential sampling schemes for estimating a binomial proportion with prescribed margin of error and confidence level. In particular, we establish the uniform controllability of coverage probability and the asymptotic optimality for such a family of sampling schemes. Our theoretical results establish the possibility that the parameters of this family of sampling schemes can be determined so that the prescribed level of confidence is guaranteed with little waste of samples. Analytic bounds for the cumulative distribution functions and expectations of sample numbers are derived. Moreover, we discuss the inherent connection of various sampling schemes. Numerical issues are addressed for improving the accuracy and efficiency of computation. Computational experiments are conducted for comparing sampling schemes. Illustrative examples are given for applications in clinical trials.

1. Introduction

Estimating a binomial proportion is a problem of ubiquitous significance in many areas of engineering and sciences. For economical reasons and other concerns, it is important to use as fewer as possible samples to guarantee the required reliability of estimation. To achieve this goal, sequential sampling schemes can be very useful. In a sequential sampling scheme, the total number of observations is not fixed in advance. The sampling process is continued stage by stage until a prespecified stopping rule is satisfied. The stopping rule is evaluated with accumulated observations. In many applications, for administrative feasibility, the sampling experiment is performed in a group fashion. Similar to group sequential tests [1, Section 8], [2], an estimation method based on taking samples by groups and evaluating them sequentially is referred to as a group sequential estimation method. It should be noted that group sequential estimation methods are general enough to include fixed-sample-size and fully sequential procedures as special cases. Particularly, a fixed-sample-size method can be viewed as a group sequential procedure of only one stage. If the increment between the sample sizes of consecutive stages is equal to 1, then the group sequential method is actually a fully sequential method.

It is a common contention that statistical inference, as a unique science to quantify the uncertainties of inferential statements, should avoid errors in the quantification of uncertainties, while minimizing the sampling cost. That is, a statistical inferential method is expected to be exact and efficient. The conventional notion of exactness is that no approximation is involved, except the round-off error due to finite word length of computers. Existing sequential methods for estimating a binomial proportion are dominantly of asymptotic nature (see, e.g., [3–7] and the references therein). Undoubtedly, asymptotic techniques provide approximate solutions and important insights for the relevant problems. However, any asymptotic method inevitably introduces unknown error in the resultant approximate solution due to the necessary use of a finite number of samples. In the direction of nonasymptotic sequential estimation, the primary goal is to ensure that the true coverage probability is above the prespecified confidence level for any value of

the associated parameter, while the required sample size is as low as possible. In this direction, Mendo and Hernando [8] developed an inverse binomial sampling scheme for estimating a binomial proportion with relative precision. Tanaka [9] developed a rigorous method for constructing fixed-width sequential confidence intervals for a binomial proportion. Although no approximation is involved, Tanaka's method is very conservative due to the bounding techniques employed in the derivation of sequential confidence intervals. Franzén [10] studied the construction of fixed-width sequential confidence intervals for a binomial proportion. However, no effective method for defining stopping rules is proposed in [10]. In his later paper [11], Franzén proposed to construct fixed-width confidence intervals based on sequential probability ratio tests (SPRTs) invented by Wald [12]. His method can generate fixed-sample-size confidence intervals based on SPRTs. Unfortunately, he made a fundamental flaw by mistaking that if the width of the fixed-sample-size confidence interval decreases to be smaller than the prespecified length as the number of samples is increasing, then the fixed-sample-size confidence interval at the termination of sampling process is the desired fixed-width sequential confidence interval guaranteeing the prescribed confidence level. More recently, Frey published a paper [13] in The American Statistician (TAS) on the classical problem of sequentially estimating a binomial proportion with prescribed margin of error and confidence level. Before Frey submitted his original manuscript to TAS in July 2009, a general framework of multistage parameter estimation had been established by Chen [14–18], which provides exact methods for estimating parameters of common distributions with various error criterion. This framework is also proposed in [19]. The approach of Frey [13] is similar to that of Chen [14–18] for the specific problem of estimating a binomial proportion with prescribed margin of error and confidence level.

In this paper, our primary interests are in the exact sequential methods for the estimation of a binomial proportion with prescribed margin of error and confidence level. We first introduce the exact approach established in [14–18]. In particular, we introduce the inclusion principle proposed in [18] and its applications to the construction of concrete stopping rules. We investigate the connection among various stopping rules. Afterward, we propose a new family of stopping rules which are extremely simple and accommodate some existing stopping rules as special cases. We provide rigorous justification for the feasibility and asymptotic optimality of such stopping rules. We prove that the prescribed confidence level can be guaranteed uniformly for all values of a binomial proportion by choosing appropriate parametric values for the stopping rule. We show that as the margin of error tends to be zero, the sample size tends to the attainable minimum as if the binomial proportion were exactly known. We derive analytic bounds for distributions and expectations of sample numbers. In addition, we address some critical computational issues and propose methods to improve the accuracy and efficiency of numerical calculation. We conduct extensive numerical experiment to study the performance of various stopping rules. We determine parametric values for the proposed stopping rules to achieve unprecedentedly efficiency while guaranteeing prescribed confidence levels. We attempt to make our proposed method as user-friendly as possible so that it can be immediately applicable even for layer persons.

The remainder of the paper is organized as follows. In Section 2, we introduce the exact approach proposed in [14–18]. In Section 3, we discuss the general principle of constructing stopping rules. In Section 4, we propose a new family of sampling schemes and investigate their feasibility, optimality, and analytic bounds of the distribution and expectation of sample numbers. In Section 5, we compare various computational methods. In particular, we illustrate why the natural method of evaluating coverage probability based on gridding parameter space is neither rigorous nor efficient. In Section 6, we present numerical results for various sampling schemes. In Section 7, we illustrate the applications of our group sequential method in clinical trials. Section 8 is the conclusion. The proofs of theorems are given in appendices. Throughout this paper, we shall use the following notations. The empty set is denoted by $\varnothing$. The set of positive integers is denoted by $\mathbb{N}$. The ceiling function is denoted by $\lceil\cdot\rceil$. The notation $\Pr\{E \mid \theta\}$ denotes the probability of the event E associated with parameter θ. The expectation of a random variable is denoted by $\mathbb{E}[\cdot]$. The standard normal distribution is denoted by $\Phi(\cdot)$. For $\alpha \in (0,1)$, the notation $\mathcal{Z}_\alpha$ denotes the critical value such that $\Phi(\mathcal{Z}_\alpha) = 1-\alpha$. For $n \in \mathbb{N}$, in the case that $X_1, \ldots, X_n$ are i.i.d. samples of X, we denote the sample mean $(\sum_{i=1}^{n} X_i)/n$ by $\overline{X}_n$, which is also called the relative frequency when X is a Bernoulli random variable. The other notations will be made clear as we proceed.

2. How Can It Be Exact?

In many areas of scientific investigation, the outcome of an experiment is of dichotomy nature and can be modeled as a Bernoulli random variable X, defined in probability space $(\Omega, \Pr, \mathcal{F})$, such that

$$\Pr\{X=1\} = 1 - \Pr\{X=0\} = p \in (0,1), \tag{1}$$

where p is referred to as a binomial proportion. In general, there is no analytic method for evaluating the binomial proportion p. A frequently used approach is to estimate p based on i.i.d. samples $X_1, X_2, \ldots$ of X. To reduce the sampling cost, it is appropriate to estimate p by a multistage sampling procedure. More formally, let $\epsilon \in (0,1)$ and $1-\delta$, with $\delta \in (0,1)$, be the prespecified margin of error and confidence level, respectively. The objective is to construct a sequential estimator $\widehat{\mathbf{p}}$ for p based on a multistage sampling scheme such that

$$\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} \geq 1-\delta, \tag{2}$$

for any $p \in (0,1)$. Throughout this paper, the probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ is referred to as the *coverage probability*. Accordingly, the probability $\Pr\{|\widehat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ is referred to as the *complementary coverage probability*. Clearly, a complete construction of a multistage estimation scheme needs to determine the number of stages, the sample sizes for all stages, the stopping rule, and the estimator for p. Throughout this paper, we let s denote the number of stages and let n_ℓ denote the number of samples at the ℓth stages. That is, the sampling process consists of s stages with sample sizes $n_1 < n_2 < \cdots < n_s$. For $\ell = 1, 2, \ldots, s$, define $K_\ell = \sum_{i=1}^{n_\ell} X_i$ and $\widehat{\mathbf{p}}_\ell = K_\ell / n_\ell$. The stopping rule is to be defined in terms of $\widehat{\mathbf{p}}_\ell$, $\ell = 1, \ldots, s$. Of course, the index of stage at the termination of the sampling process, denoted by $\mathbf{l}$, is a random number. Accordingly, the number of samples at the termination of the experiment, denoted by $\mathbf{n}$, is a random number which equals $n_{\mathbf{l}}$. Since for each ℓ, $\widehat{\mathbf{p}}_\ell$ is a maximum-likelihood and minimum-variance unbiased estimator of p, the sequential estimator for p is taken as

$$\widehat{\mathbf{p}} = \widehat{\mathbf{p}}_{\mathbf{l}} = \frac{\sum_{i=1}^{n_{\mathbf{l}}} X_i}{n_{\mathbf{l}}} = \frac{\sum_{i=1}^{\mathbf{n}} X_i}{\mathbf{n}}. \tag{3}$$

In the above discussion, we have outlined the general characteristics of a multistage sampling scheme for estimating a binomial proportion. It remains to determine the number of stages, the sample sizes for all stages, and the stopping rule so that the resultant estimator $\widehat{\mathbf{p}}$ satisfies (2) for any $p \in (0,1)$.

Actually, the problem of sequential estimation of a binomial proportion has been treated by Chen [14–18] in a general framework of multistage parameter estimation. The techniques of [14–18] are sufficient to offer exact solutions for a wide range of sequential estimation problems, including the estimation of a binomial proportion as a special case. The central idea of the approach in [14–18] is the control of coverage probability by a single parameter ζ, referred to as the *coverage tuning parameter*, and the adaptive rigorous checking of coverage guarantee by virtue of bounds of coverage probabilities. It is recognized in [14–18] that, due to the discontinuity of the coverage probability on parameter space, the conventional method of evaluating the coverage probability for a finite number of parameter values is neither rigorous not computationally efficient for checking the coverage probability guarantee.

As mentioned in the introduction, Frey published an article [13] in TAS on the sequential estimation of a binomial proportion with prescribed margin of error and confidence level. For clarity of presentation, the comparison of the works of Chen and Frey is given in Section 5.4. In the remainder of this section, we shall only introduce the idea and techniques of [14–18], which had been precedentially developed by Chen before Frey submitted his original manuscript to TAS in July 2009. We will introduce the approach of [14–18] with a focus on the special problem of estimating a binomial proportion with prescribed margin of error and confidence level.

2.1. Four Components Suffice. The exact methods of [14–18] for multistage parameter estimation have four main components as follows.

(i) Stopping rules parameterized by the coverage tuning parameter $\zeta > 0$ such that the associated coverage probabilities can be made arbitrarily close to 1 by choosing $\zeta > 0$ to be a sufficiently small number.

(ii) Recursively computable lower and upper bounds for the complementary coverage probability for a given ζ and an interval of parameter values.

(iii) Adapted branch and bound algorithm.

(iv) Bisection coverage tuning.

Without looking at the technical details, one can see that these four components are sufficient for constructing a sequential estimator so that the prescribed confidence level is guaranteed. The reason is as follows. As lower and upper bounds for the complementary coverage probability are available, the global optimization technique, branch and bound (B&B) algorithm [20], can be used to compute exactly the maximum of complementary coverage probability on the whole parameter space. Thus, it is possible to check rigorously whether the coverage probability associated with a given ζ is no less than the prespecified confidence level. Since the coverage probability can be controlled by ζ, it is possible to determine ζ as large as possible to guarantee the desired confidence level by a bisection search. This process is referred to as bisection coverage tuning in [14–18]. Since a critical subroutine needed for bisection coverage tuning is to check whether the coverage probability is no less than the prespecified confidence level, it is not necessary to compute exactly the maximum of the complementary coverage probability. Therefore, Chen revised the standard B&B algorithm to reduce the computational complexity and called the improved algorithm as the adapted B&B Algorithm. The idea is to adaptively partition the parameter space as many subintervals. If for all subintervals, the upper bounds of the complementary coverage probability are no greater than δ, then declare that the coverage probability is guaranteed. If there exists a subinterval for which the lower bound of the complementary coverage probability is greater than δ, then declare that the coverage probability is not guaranteed. Continue partitioning the parameter space if no decision can be made. The four components are illustrated in the sequel under the headings of stopping rules, interval bounding, adapted branch and bound, and bisection coverage tuning.

2.2. Stopping Rules. The first component for the exact sequential estimation of a binomial proportion is the stopping rule for constructing a sequential estimator such that the coverage probability can be controlled by the coverage tuning

parameter ζ. For convenience of describing some concrete stopping rules, define

$$\mathscr{M}(z,\theta)=\begin{cases} z\ln\dfrac{\theta}{z} & \\ \quad +(1-z)\ln\dfrac{1-\theta}{1-z} & \text{for } z\in(0,1),\\ & \theta\in(0,1),\\ \ln(1-\theta) & \text{for } z=0,\ \theta\in(0,1),\\ \ln\theta & \text{for } z=1,\ \theta\in(0,1),\\ -\infty & \text{for } z\in[0,1],\\ & \theta\notin(0,1),\end{cases}$$

$$S(k,l,n,p)=\begin{cases}\displaystyle\sum_{i=k}^{l}\binom{n}{i}p^{i}(1-p)^{n-i} & \text{for } p\in(0,1),\\ 0 & \text{for } p\notin(0,1),\end{cases} \tag{4}$$

where k and l are integers such that $0\le k\le l\le n$. Assume that $0<\zeta\delta<1$. For the purpose of controlling the coverage probability $\Pr\{|\widehat{\mathbf{p}}-\mathrm{p}|<\epsilon\mid p\}$ by the coverage tuning parameter, Chen has proposed four stopping rules as follows.

Stopping Rule A. Continue sampling until $\mathscr{M}((1/2)-|(1/2)-\widehat{\mathbf{p}}_\ell|,(1/2)-|(1/2)-\widehat{\mathbf{p}}_\ell|+\epsilon)\le(\ln(\zeta\delta))/n_\ell$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule B. Continue sampling until $(|\widehat{\mathbf{p}}_\ell-(1/2)|-(2/3)\epsilon)^2\ge(1/4)+(\epsilon^2 n_\ell/2\ln(\zeta\delta))$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule C. Continue sampling until $S(K_\ell,n_\ell,n_\ell,\widehat{\mathbf{p}}_\ell-\epsilon)\le\zeta\delta$ and $S(0,K_\ell,n_\ell,\widehat{\mathbf{p}}_\ell+\epsilon)\le\zeta\delta$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule D. Continue sampling until $n_\ell\ge\widehat{\mathbf{p}}_\ell(1-\widehat{\mathbf{p}}_\ell)(2/\epsilon^2)\ln(1/\zeta\delta)$ for some $\ell\in\{1,\dots,s\}$.

Stopping Rule A was first proposed in [14, Theorem 7] and restated in [15, Theorem 16]. Stopping Rule B was first proposed in [16, Theorem 1] and represented as the third stopping rule in [21, Section 4.1.1]. Stopping Rule C originated from [17, Theorem 1] and was restated as the first stopping rule in [21, Section 4.1.1]. Stopping Rule D was described in the remarks following Theorem 7 of [22]. All these stopping rules can be derived from the general principles proposed in [18, Section 3] and [19, Section 2.4].

Given that a stopping rule can be expressed in terms of $\widehat{\mathbf{p}}_\ell$ and n_ℓ for $\ell=1,\dots,s$, it is possible to find a bivariate function $\mathscr{D}(\cdot,\cdot)$ on $\{(z,n):z\in[0,1],\ n\in\mathbb{N}\}$, taking values from $\{0,1\}$, such that the stopping rule can be stated as the following: continue sampling until $\mathscr{D}(\widehat{\mathbf{p}}_\ell,n_\ell)=1$ for some $\ell\in\{1,\dots,s\}$. It can be checked that such representation applies to Stopping Rules A, B, C, and D. For example, Stopping Rule B can be expressed in this way by virtue of function $\mathscr{D}(\cdot,\cdot)$ such that

$$\mathscr{D}(z,n)=\begin{cases}1 & \text{if } \left(\left|z-\dfrac{1}{2}\right|-\dfrac{2}{3}\epsilon\right)^2\ge\dfrac{1}{4}+\dfrac{\epsilon^2 n}{2\ln(\zeta\delta)},\\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

The motivation of introducing function $\mathscr{D}(\cdot,\cdot)$ is to parameterize the stopping rule in terms of design parameters. Function $\mathscr{D}(\cdot,\cdot)$ determines the form of the stopping rule and, consequently, the sample sizes for all stages can be chosen as functions of design parameters. Specifically, let

$$N_{\min}=\min\left\{n\in\mathbb{N}:\mathscr{D}\left(\frac{k}{n},n\right)=1 \text{ for some nonnegative integer } k \text{ not exceeding } n\right\}, \tag{6}$$

$$N_{\max}=\min\left\{n\in\mathbb{N}:\mathscr{D}\left(\frac{k}{n},n\right)=1 \text{ for all nonnegative integer } k \text{ not exceeding } n\right\}. \tag{7}$$

To avoid unnecessary checking of the stopping criterion and thus reduce administrative cost, there should be a possibility that the sampling process is terminated at the first stage. Hence, the minimum sample size n_1 should be chosen to ensure that $\{\mathbf{n}=n_1\}\neq\varnothing$. This implies that the sample size n_1 for the first stage can be taken as $N_{\min}$. On the other hand, since the sampling process must be terminated at or before the sth stage, the maximum sample size n_s should be chosen to guarantee that $\{\mathbf{n}>n_s\}=\varnothing$. This implies that the sample size n_s for the last stage can be taken as $N_{\max}$. If the number of stages s is given, then the sample sizes for stages in between 1 and s can be chosen as $s-2$ integers between $N_{\min}$ and $N_{\max}$. Particularly, if the group sizes are expected to be approximately equal, then the sample sizes can be taken as

$$n_\ell=\left\lceil N_{\min}+\frac{\ell-1}{s-1}(N_{\max}-N_{\min})\right\rceil,\quad \ell=1,\dots,s. \tag{8}$$

Since the stopping rule is associated with the coverage tuning parameter ζ, it follows that the number of stages s and the sample sizes $n_1,n_2,\dots,n_s$ can be expressed as functions of ζ. In this sense, it can be said that the stopping rule is parameterized by the coverage tuning parameter ζ. The above method of parameterizing stopping rules has been used in [14–17] and proposed in [21, Section 2.1, page 9].

2.3. Interval Bounding. The second component for the exact sequential estimation of a binomial proportion is the method of bounding the complementary coverage probability $\Pr\{|\widehat{\mathbf{p}}-p|\ge\epsilon\mid p\}$ for p in an interval $[a,b]$ contained by interval $(0,1)$. Applying Theorem 8 of [15] to the special case of a Bernoulli distribution immediately yields

$$\begin{aligned}&\Pr\{\widehat{\mathbf{p}}\le a-\epsilon\mid b\}+\Pr\{\widehat{\mathbf{p}}\ge b+\epsilon\mid a\}\\ &\quad\le\Pr\{|\widehat{\mathbf{p}}-p|\ge\epsilon\mid p\}\\ &\quad\le\Pr\{\widehat{\mathbf{p}}\le b-\epsilon\mid a\}+\Pr\{\widehat{\mathbf{p}}\ge a+\epsilon\mid b\},\end{aligned} \tag{9}$$

∇ Let $k \leftarrow 0$, $l_0 \leftarrow \Psi_{\text{lb}}(\mathcal{I}_{\text{init}})$ and $u_0 \leftarrow \Psi_{\text{ub}}(\mathcal{I}_{\text{init}})$.
∇ Let $\mathcal{S}_0 \leftarrow \{\mathcal{I}_{\text{init}}\}$ if $u_0 > \delta$. Otherwise, let $\mathcal{S}_0$ be empty.
∇ While $\mathcal{S}_k$ is nonempty, $l_k < \delta$ and u_k is greater than $\max\{l_k + \eta, \delta\}$, do the following:
⋄ Split each interval in $\mathcal{S}_k$ as two new intervals of equal length.
Let S_k denote the set of all new intervals obtained from this splitting procedure.
⋄ Eliminate any interval $\mathcal{I}$ from S_k such that $\Psi_{\text{ub}}(\mathcal{I}) \leq \delta$.
⋄ Let $\mathcal{S}_{k+1}$ be the set S_k processed by the above elimination procedure.
⋄ Let $l_{k+1} \leftarrow \max_{\mathcal{I} \in \mathcal{S}_{k+1}} \Psi_{\text{lb}}(\mathcal{I})$ and $u_{k+1} \leftarrow \max_{\mathcal{I} \in \mathcal{S}_{k+1}} \Psi_{\text{ub}}(\mathcal{I})$. Let $k \leftarrow k + 1$.
∇ If $\mathcal{S}_k$ is empty and $l_k < \delta$, then declare $\max \Psi(\mathcal{I}_{\text{init}}) \leq \delta$.
Otherwise, declare $\max \Psi(\mathcal{I}_{\text{init}}) > \delta$.

Algorithm 1

for all $p \in [a, b] \subseteq (0, 1)$. The bounds of (9) can be shown as follows. Note that $\Pr\{\hat{\mathbf{p}} \leq a - \epsilon \mid p\} + \Pr\{\hat{\mathbf{p}} \geq b + \epsilon \mid p\} \leq \Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\} = \Pr\{\hat{\mathbf{p}} \leq p - \epsilon \mid p\} + \Pr\{\hat{\mathbf{p}} \geq p + \epsilon \mid p\} \leq \Pr\{\hat{\mathbf{p}} \leq b - \epsilon \mid p\} + \Pr\{\hat{\mathbf{p}} \geq a + \epsilon \mid p\}$ for $p \in [a, b] \subseteq (0, 1)$. As a consequence of the monotonicity of $\Pr\{\hat{\mathbf{p}} \geq \vartheta \mid p\}$ and $\Pr\{\hat{\mathbf{p}} \leq \vartheta \mid p\}$ with respect to p, where ϑ is a real number independent of p, the lower and upper bounds of $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ for $p \in [a, b] \subseteq (0, 1)$ can be given as $\Pr\{\hat{\mathbf{p}} \leq a - \epsilon \mid b\} + \Pr\{\hat{\mathbf{p}} \geq b + \epsilon \mid a\}$ and $\Pr\{\hat{\mathbf{p}} \leq b - \epsilon \mid a\} + \Pr\{\hat{\mathbf{p}} \geq a + \epsilon \mid b\}$, respectively.

In page 15, equation (1) of [15], Chen proposed to apply the recursive method of Schultz et al. [23, Section 2] to compute the lower and upper bounds of $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ given by (9). It should be pointed out that such lower and upper bounds of $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ can also be computed by the recursive path-counting method of Franzén [10, page 49].

2.4. Adapted Branch and Bound. The third component for the exact sequential estimation of a binomial proportion is the adapted B&B algorithm, which was proposed in [15, Section 2.8], for quick determination of whether the coverage probability is no less than $1 - \delta$ for any value of the associated parameter. Such a task of checking the coverage probability is also referred to as checking the coverage probability guarantee. Given that lower and upper bounds of the complementary coverage probability on an interval of parameter values can be obtained by the interval bounding techniques, this task can be accomplished by applying the B&B algorithm [20] to compute exactly the maximum of the complementary coverage probability on the parameter space. However, in our applications, it suffices to determine whether the maximum of the complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ with respect to $p \in (0, 1)$ is greater than the confidence parameter δ. For fast checking whether the maximal complementary coverage probability exceeds δ, Chen proposed to reduce the computational complexity by revising the standard B&B algorithm as the Adapted B&B Algorithm in [15, Section 2.8]. To describe this algorithm, let $\mathcal{I}_{\text{init}}$ denote the parameter space $(0, 1)$. For an interval $\mathcal{I} \subseteq \mathcal{I}_{\text{init}}$, let $\max \Psi(\mathcal{I})$ denote the maximum of the complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ with respect to $p \in \mathcal{I}$. Let $\Psi_{\text{lb}}(\mathcal{I})$ and $\Psi_{\text{ub}}(\mathcal{I})$ be, respectively, the lower and upper bounds of $\Psi(\mathcal{I})$, which can be obtained by the interval bounding techniques introduced in Section 2.3. Let $\eta > 0$ be a prespecified tolerance, which is much smaller than δ. The adapted B&B algorithm of [15] is represented with a slight modification as in Algorithm 1.

It should be noted that for a sampling scheme of symmetrical stopping boundary, the initial interval $\mathcal{I}_{\text{init}}$ may be taken as $(0, 1/2)$ for the sake of efficiency. In Section 5.1, we will illustrate why the adapted B&B algorithm is superior than the direct evaluation based on gridding parameter space. As will be seen in Section 5.2, the objective of the adapted B&B algorithm can also be accomplished by the Adaptive Maximum Checking Algorithm due to Chen [21, Section 3.3] and rediscovered by Frey [13, Appendix]. An explanation is given in Section 5.3 for the advantage of working with the complementary coverage probability.

2.5. Bisection Coverage Tuning. The fourth component for the exact sequential estimation of a binomial proportion is Bisection Coverage Tuning. Based on the adaptive rigorous checking of coverage probability, Chen proposed in [14, Section 2.7] and [15, Section 2.6] to apply a bisection search method to determine maximal ζ such that the coverage probability is no less than $1 - \delta$ for any value of the associated parameter. Moreover, Chen has developed asymptotic results in [15, page 21, Theorem 18] for determining the initial interval of ζ needed for the bisection search. Specifically, if the complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ associated with $\zeta = \zeta_0$ tends to δ as $\epsilon \to 0$, then the initial interval of ζ can be taken as $[\zeta_0 2^i, \zeta_0 2^{i+1}]$, where i is the largest integer such that the complementary coverage probability associated with $\zeta = \zeta_0 2^i$ is no greater than δ for all $p \in (0, 1)$. By virtue of a bisection search, it is possible to obtain $\zeta^* \in [\zeta_0 2^i, \zeta_0 2^{i+1}]$ such that the complementary coverage probability associated with $\zeta = \zeta^*$ is guaranteed to be no greater than δ for all $p \in (0, 1)$.

3. Principle of Constructing Stopping Rules

In this section, we shall illustrate the inherent connection between various stopping rules. It will be demonstrated that a lot of stopping rules can be derived by virtue of the inclusion principle proposed by Chen [18, Section 3].

3.1. Inclusion Principle. The problem of estimating a binomial proportion can be considered as a special case of parameter estimation for a random variable X parameterized by $\theta \in \Theta$, where the objective is to construct a sequential estimator $\widehat{\theta}$ for θ such that $\Pr\{|\widehat{\theta} - \theta| < \epsilon \mid \theta\} \geq 1 - \delta$ for any $\theta \in \Theta$. Assume that the sampling process consists of s stages with sample sizes $n_1 < n_2 < \cdots < n_s$. For $\ell = 1, \ldots, s$, define an estimator $\widehat{\theta}_\ell$ for θ in terms of samples $X_1, \ldots, X_{n_\ell}$ of X. Let $[L_\ell, U_\ell]$, $\ell = 1, 2, \ldots, s$ be a sequence of confidence intervals such that for any ℓ, $[L_\ell, U_\ell]$ is defined in terms of $X_1, \ldots, X_{n_\ell}$ and that the coverage probability $\Pr\{L_\ell \leq \theta \leq U_\ell \mid \theta\}$ can be made arbitrarily close to 1 by choosing $\zeta > 0$ to be a sufficiently small number. In Theorem 2 of [18], Chen proposed the following general stopping rule:

$$\text{Continue sampling until } U_\ell - \epsilon \leq \widehat{\theta}_\ell \leq L_\ell + \epsilon \text{ for some } \ell \in \{1, \ldots, s\}. \tag{10}$$

At the termination of the sampling process, a sequential estimator for θ is taken as $\widehat{\theta} = \widehat{\theta}_{\mathbf{l}}$, where $\mathbf{l}$ is the index of stage at the termination of sampling process.

Clearly, the general stopping rule (10) can be restated as follows.

Continue sampling until the confidence interval $[L_\ell, U_\ell]$ is included by interval $[\widehat{\theta}_\ell - \epsilon, \widehat{\theta}_\ell + \epsilon]$ for some $\ell \in \{1, \ldots, s\}$.

The sequence of confidence intervals are parameterized by ζ for purpose of controlling the coverage probability $\Pr\{|\widehat{\theta} - \theta| < \epsilon \mid \theta\}$. Due to the inclusion relationship $[L_\ell, U_\ell] \subseteq [\widehat{\theta}_\ell - \epsilon, \widehat{\theta}_\ell + \epsilon]$, such a general methodology of using a sequence of confidence intervals to construct a stopping rule for controlling the coverage probability is referred to as the *inclusion principle*. It is asserted by Theorem 2 of [18] that

$$\Pr\left\{\left|\widehat{\theta} - \theta\right| < \epsilon \mid \theta\right\} \geq 1 - s\zeta\delta, \quad \forall \theta \in \Theta, \tag{11}$$

provided that $\Pr\{L_\ell < \theta < U_\ell \mid \theta\} \geq 1 - \zeta\delta$ for $\ell = 1, \ldots, s$ and $\theta \in \Theta$. This demonstrates that if the number of stages s is bounded respective to ζ, then the coverage probability $\Pr\{|\widehat{\theta} - \theta| < \epsilon \mid \theta\}$ associated with the stopping rule derived from the inclusion principle can be controlled by ζ. Actually, before explicitly proposing the inclusion principle in [18], Chen had extensively applied the inclusion principle in [14–17] to construct stopping rules for estimating parameters of various distributions such as binomial, Poisson, geometric, hypergeometric, and normal distributions. A more general version of the inclusion principle is proposed in [19, Section 2.4]. For simplicity of the stopping rule, Chen had made effort to eliminate the computation of confidence limits.

In the context of estimating a binomial proportion p, the inclusion principle immediately leads to the following general stopping rule:

$$\text{Continue sampling until } \widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon \text{ for some } \ell \in \{1, \ldots, s\}. \tag{12}$$

Consequently, the sequential estimator for p is taken as $\widehat{\mathbf{p}}$ according to (3). It should be pointed out that the stopping rule (12) had been rediscovered by Frey in Section 2, the 1st paragraph of [13]. The four stopping rules considered in his paper follow immediately from applying various confidence intervals to the general stopping rule (12).

In the sequel, we will illustrate how to apply (12) to the derivation of Stopping Rules A, B, C, and D introduced in Section 2.2 and other specific stopping rules.

3.2. Stopping Rule from Wald Intervals. By virtue of Wald's method of interval estimation for a binomial proportion p, a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be constructed such that

$$L_\ell = \widehat{\mathbf{p}}_\ell - \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widehat{\mathbf{p}}_\ell(1 - \widehat{\mathbf{p}}_\ell)}{n_\ell}}, \quad U_\ell = \widehat{\mathbf{p}}_\ell + \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widehat{\mathbf{p}}_\ell(1 - \widehat{\mathbf{p}}_\ell)}{n_\ell}}, \quad \ell = 1, \ldots, s, \tag{13}$$

and that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \approx 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. Note that, for $\ell = 1, \ldots, s$, the event $\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon\}$ is the same as the event $\{(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) - n_\ell(\epsilon/\mathcal{Z}_{\zeta\delta})^2\}$. So, applying this sequence of confidence intervals to (12) results in the stopping rule "continue sampling until $(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) - n_\ell(\epsilon/\mathcal{Z}_{\zeta\delta})^2$ for some $\ell \in \{1, \ldots, s\}$". Since for any $\zeta \in (0, 1/\delta)$, there exists a unique number $\zeta' \in (0, 1/\delta)$ such that $\mathcal{Z}_{\zeta\delta} = \sqrt{2\ln(1/\zeta'\delta)}$, this stopping rule is equivalent to "Continue sampling until $(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))$ for some $\ell \in \{1, \ldots, s\}$." This stopping rule is actually the same as Stopping Rule D, since $\{(\widehat{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))\} = \{n_\ell \geq \widehat{\mathbf{p}}_\ell(1 - \widehat{\mathbf{p}}_\ell)(2/\epsilon^2)\ln(1/\zeta\delta)\}$ for $\ell \in \{1, \ldots, s\}$.

3.3. Stopping Rule from Revised Wald Intervals. Define $\widetilde{\mathbf{p}}_\ell = (n_\ell\,\widehat{\mathbf{p}}_\ell + a)/(n_\ell + 2a)$ for $\ell = 1, \ldots, s$, where a is a positive number. Inspired by Wald's method of interval estimation for p, a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ can be constructed such that

$$L_\ell = \widehat{\mathbf{p}}_\ell - \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widetilde{\mathbf{p}}_\ell(1 - \widetilde{\mathbf{p}}_\ell)}{n_\ell}}, \qquad U_\ell = \widehat{\mathbf{p}}_\ell + \mathcal{Z}_{\zeta\delta}\sqrt{\frac{\widetilde{\mathbf{p}}_\ell(1 - \widetilde{\mathbf{p}}_\ell)}{n_\ell}} \tag{14}$$

and that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \approx 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. This sequence of confidence intervals was applied by Frey [13] to the general stopping rule (12). As a matter of fact, such idea of revising Wald interval $[\,\overline{X}_n - \mathcal{Z}_{\zeta\delta}\sqrt{(\overline{X}_n(1 - \overline{X}_n))/n},\ \overline{X}_n + \mathcal{Z}_{\zeta\delta}\sqrt{(\overline{X}_n(1 - \overline{X}_n))/n}\,]$ by replacing the relative frequency $\overline{X}_n = (\sum_{i=1}^n X_i)/n$ involved in the confidence limits with $\widetilde{p}_a = (n\overline{X}_n + a)/(n + 2a)$ had been proposed by Chen [24, Section 4].

As can be seen from Section 2, page 243, of Frey [13], applying (12) with the sequence of revised Wald intervals yields the stopping rule "Continue sampling until

$(\widetilde{\mathbf{p}}_\ell - 1/2)^2 \geq (1/4) + (\epsilon^2 n_\ell / 2\ln(\zeta\delta))$ for some $\ell \in \{1, \ldots, s\}$." Clearly, replacing $\widehat{\mathbf{p}}_\ell$ in Stopping Rule D with $\widetilde{\mathbf{p}}_\ell = (a + n_\ell \widehat{\mathbf{p}}_\ell)/(n_\ell + 2a)$ also leads to this stopping rule.

3.4. Stopping Rule from Wilson's Confidence Intervals. Making use of the interval estimation method of Wilson [25], one can obtain a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p such that

$$L_\ell = \max\left\{0, \frac{\widehat{\mathbf{p}}_\ell + \left(\mathcal{Z}^2_{\zeta\delta}/2n_\ell\right)}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)} - \frac{\mathcal{Z}_{\zeta\delta}\sqrt{\left(\widehat{\mathbf{p}}_\ell\left(1 - \widehat{\mathbf{p}}_\ell\right)/n_\ell\right) + \left(\mathcal{Z}_{\zeta\delta}/2n_\ell\right)^2}}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)}\right\},$$

$$U_\ell = \min\left\{1, \frac{\widehat{\mathbf{p}}_\ell + \left(\mathcal{Z}^2_{\zeta\delta}/2n_\ell\right)}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)} + \frac{\mathcal{Z}_{\zeta\delta}\sqrt{\left(\widehat{\mathbf{p}}_\ell\left(1 - \widehat{\mathbf{p}}_\ell\right)/n_\ell\right) + \left(\mathcal{Z}_{\zeta\delta}/2n_\ell\right)^2}}{1 + \left(\mathcal{Z}^2_{\zeta\delta}/n_\ell\right)}\right\} \tag{15}$$

and that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \approx 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. It should be pointed out that the sequence of Wilson's confidence intervals has been applied by Frey [13, Section 2, page 243] to the general stopping rule (12) for estimating a binomial proportion.

Since a stopping rule directly involves the sequence of Wilson's confidence intervals is cumbersome, it is desirable to eliminate the computation of Wilson's confidence intervals in the stopping rule. For this purpose, we need to use the following result.

Theorem 1. *Assume that* $0 < \zeta\delta < 1$ *and* $0 < \epsilon < 1/2$. *Then, Wilson's confidence intervals satisfy* $\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon\} = \{(|\widehat{\mathbf{p}}_\ell - 1/2| - \epsilon)^2 \geq (1/4) - n_\ell(\epsilon/\mathcal{Z}_{\zeta\delta})^2\}$ *for* $\ell = 1, \ldots, s$.

See Appendix A for a proof. As a consequence of Theorem 1 and the fact that for any $\zeta \in (0, 1/\delta)$, there exists a unique number $\zeta' \in (0, 1/\delta)$ such that $\mathcal{Z}_{\zeta\delta} = \sqrt{2\ln(1/\zeta'\delta)}$, applying the sequence of Wilson's confidence intervals to (12) leads to the following stopping rule.

Continue sampling until

$$\left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}, \tag{16}$$

for some $\ell \in \{1, \ldots, s\}$.

3.5. Stopping Rule from Clopper-Pearson Confidence Intervals. Applying the interval estimation method of Clopper-Pearson [26], a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be obtained such that $\Pr\{L_\ell \leq p \leq U_\ell \mid p\} \geq 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$, where the upper confidence limit U_ℓ satisfies the equation $S(0, K_\ell, n_\ell, U_\ell) = \zeta\delta$ if $K_\ell < n_\ell$; and the lower confidence limit L_ℓ satisfies the equation $S(K_\ell, n_\ell, n_\ell, L_\ell) = \zeta\delta$ if $K_\ell > 0$. The well-known equation (10.8) in [27, page 173] implies that $S(0, k, n, p)$, with $0 \leq k < n$, is decreasing with respect to $p \in (0, 1)$ and that $S(k, n, n, p)$, with $0 < k \leq n$, is increasing with respect to $p \in (0, 1)$. It follows that

$$\begin{aligned}
\left\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell\right\} &= \left\{0 < \widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell \leq \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell > \epsilon,\ S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \leq \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell > \epsilon,\ S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \leq \epsilon,\ S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\} \\
&= \left\{S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta\right\}, \\
\left\{\widehat{\mathbf{p}}_\ell + \epsilon \geq U_\ell\right\} &= \left\{1 > \widehat{\mathbf{p}}_\ell + \epsilon \geq U_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell \geq 1 - \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell < 1 - \epsilon,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \geq 1 - \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell < 1 - \epsilon,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \geq 1 - \epsilon,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\} \\
&= \left\{S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\},
\end{aligned} \tag{17}$$

for $\ell = 1, \ldots, s$. Consequently,

$$\begin{aligned}
&\left\{\widehat{\mathbf{p}}_\ell - \epsilon \leq L_\ell \leq U_\ell \leq \widehat{\mathbf{p}}_\ell + \epsilon\right\} \\
&\quad = \left\{S\left(K_\ell, n_\ell, n_\ell, \widehat{\mathbf{p}}_\ell - \epsilon\right) \leq \zeta\delta,\ S\left(0, K_\ell, n_\ell, \widehat{\mathbf{p}}_\ell + \epsilon\right) \leq \zeta\delta\right\},
\end{aligned} \tag{18}$$

for $\ell = 1, \ldots, s$. This demonstrates that applying the sequence of Clopper-Pearson confidence intervals to the general stopping rule (12) gives Stopping Rule C.

It should be pointed out that Stopping Rule C was rediscovered by Frey as the third stopping rule in Section 2, page 243 of his paper [13].

3.6. Stopping Rule from Fishman's Confidence Intervals. By the interval estimation method of Fishman [28], a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be obtained such that

$$L_\ell = \begin{cases} 0 & \text{if } \widehat{\mathbf{p}}_\ell = 0, \\ \left\{\theta_\ell \in \left(0, \widehat{\mathbf{p}}_\ell\right) : \mathcal{M}\left(\widehat{\mathbf{p}}_\ell, \theta_\ell\right) = \dfrac{\ln(\zeta\delta)}{n_\ell}\right\} & \text{if } \widehat{\mathbf{p}}_\ell > 0, \end{cases}$$

$$U_\ell = \begin{cases} 1 & \text{if } \widehat{\mathbf{p}}_\ell = 1, \\ \left\{\theta_\ell \in \left(\widehat{\mathbf{p}}_\ell, 1\right) : \mathcal{M}\left(\widehat{\mathbf{p}}_\ell, \theta_\ell\right) = \dfrac{\ln(\zeta\delta)}{n_\ell}\right\} & \text{if } \widehat{\mathbf{p}}_\ell < 1. \end{cases} \tag{19}$$

Under the assumption that $0 < \zeta\delta < 1$ and $0 < \epsilon < 1/2$, by similar techniques as the proof of Theorem 7 of [22], it can be shown that $\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le U_\ell \le \widehat{\mathbf{p}}_\ell + \epsilon\} = \{\mathscr{M}((1/2) - |(1/2) - \widehat{\mathbf{p}}_\ell|, (1/2) - |(1/2) - \widehat{\mathbf{p}}_\ell| + \epsilon) \le (\ln(\zeta\delta))/n_\ell\}$ for $\ell = 1, \ldots, s$. Therefore, applying the sequence of confidence intervals of Fishman to the general stopping rule (12) gives Stopping Rule A.

It should be noted that Fishman's confidence intervals are actually derived from the Chernoff bounds of the tailed probabilities of the sample mean of Bernoulli random variable. Hence, Stopping Rule A is also referred to as the stopping rule from Chernoff bounds in this paper.

3.7. Stopping Rule from Confidence Intervals of Chen et al. Using the interval estimation method of Chen et al. [29], a sequence of confidence intervals $[L_\ell, U_\ell]$, $\ell = 1, \ldots, s$ for p can be obtained such that

$$L_\ell = \max\left\{0, \widehat{\mathbf{p}}_\ell + \frac{3}{4} \times \frac{1 - 2\widehat{\mathbf{p}}_\ell - \sqrt{1 + (9n_\ell/2\ln(1/\zeta\delta))\widehat{\mathbf{p}}_\ell(1-\widehat{\mathbf{p}}_\ell)}}{1 + (9n_\ell/8\ln(1/\zeta\delta))}\right\},$$

$$U_\ell = \min\left\{1, \widehat{\mathbf{p}}_\ell + \frac{3}{4} \times \frac{1 - 2\widehat{\mathbf{p}}_\ell + \sqrt{1 + (9n_\ell/2\ln(1/\zeta\delta))\widehat{\mathbf{p}}_\ell(1-\widehat{\mathbf{p}}_\ell)}}{1 + (9n_\ell/8\ln(1/\zeta\delta))}\right\} \tag{20}$$

and that $\Pr\{L_\ell \le p \le U_\ell \mid p\} \ge 1 - 2\zeta\delta$ for $\ell = 1, \ldots, s$ and $p \in (0, 1)$. Under the assumption that $0 < \zeta\delta < 1$ and $0 < \epsilon < 1/2$, by similar techniques as the proof of Theorem 1 of [30], it can be shown that $\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le U_\ell \le \widehat{\mathbf{p}}_\ell + \epsilon\} = \{(|\widehat{\mathbf{p}}_\ell - 1/2| - (2/3)\epsilon)^2 \ge (1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))\}$ for $\ell = 1, \ldots, s$. This implies that applying the sequence of confidence intervals of Chen et al. to the general stopping rule (12) leads to Stopping Rule B.

Actually, the confidence intervals of Chen et al. [29] are derived from Massart's inequality [31] on the tailed probabilities of the sample mean of Bernoulli random variable. For this reason, Stopping Rule B is also referred to as the stopping rule from Massart's inequality in [21, Section 4.1.1].

4. Double-Parabolic Sequential Estimation

From Sections 2.2, 3.2, and 3.7, it can be seen that, by introducing a new parameter $\rho \in [0, 1]$ and letting ρ take values 2/3 and 0, respectively, Stopping Rules B and D can be accommodated as special cases of the following general stopping rule.

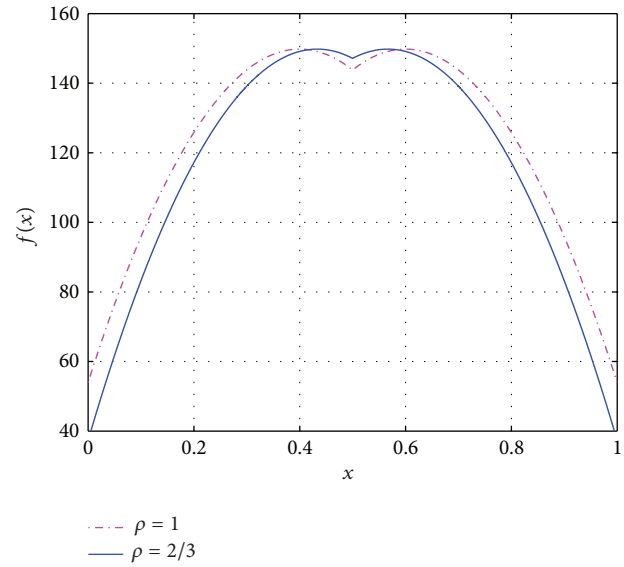

Figure 1: Double-parabolic sampling.

Continue the sampling process until

$$\left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \rho\epsilon\right)^2 \ge \frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)} \tag{21}$$

for some $\ell \in \{1, 2, \ldots, s\}$, where $\zeta \in (0, 1/\delta)$.

Moreover, as can be seen from (16), the stopping rule derived from applying Wilson's confidence intervals to (12) can also be viewed as a special case of such general stopping rule with $\rho = 1$.

From the stopping condition (21), it can be seen that the stopping boundary is associated with the double-parabolic function $f(x) = (2/\epsilon^2)\ln(\zeta\delta)[(1/4) - (|x - 1/2| - \rho\epsilon)^2]$ such that x and $f(x)$ correspond to the sample mean and sample size, respectively. For $\epsilon = 0.1$, $\delta = 0.05$, and $\zeta = 1$, stopping boundaries with various ρ are shown by Figure 1.

For fixed ϵ and δ, the parameters ρ and ζ affect the shape of the stoping boundary in a way as follows. As ρ increases, the span of stopping boundary is increasing in the axis of sample mean. By decreasing ζ, the stopping boundary can be dragged toward the direction of increasing sample size. Hence, the parameter ρ is referred to as the *dilation coefficient*. The parameter ζ is referred to as the *coverage tuning parameter*. Since the stopping boundary consists of two parabolas, this approach of estimating a binomial proportion is referred to as the *double-parabolic sequential estimation* method.

4.1. Parametrization of the Sampling Scheme. In this section, we shall parameterize the double-parabolic sequential sampling scheme by the method described in Section 2.2. From the stopping condition (21), the stopping rule can be restated as follows. Continue sampling until $\mathscr{D}(\widehat{\mathbf{p}}_\ell, n_\ell) = 1$

for some $\ell \in \{1, \ldots, s\}$, where the function $\mathscr{D}(\cdot,\cdot)$ is defined by

$$\mathscr{D}(z,n) = \begin{cases} 1 & \text{if } \left(\left|z-\frac{1}{2}\right| - \rho\epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n}{2\ln(\zeta\delta)}, \\ 0 & \text{otherwise.} \end{cases} \quad (22)$$

Clearly, the function $\mathscr{D}(\cdot,\cdot)$ associated with the double-parabolic sequential sampling scheme depends on the design parameters ρ, ζ, ϵ and δ. Applying the function $\mathscr{D}(\cdot,\cdot)$ defined by (22) to (6) yields

$$N_{\min} = \min\left\{ n \in \mathbb{N} : \left(\left|\frac{k}{n}-\frac{1}{2}\right| - \rho\epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n}{2\ln(\zeta\delta)} \text{ for some nonnegative integer } k \text{ not exceeding } n \right\}. \quad (23)$$

Since ϵ is usually small in practical applications, we restrict ϵ to satisfy $0 < \rho\epsilon \leq 1/4$. As a consequence of $0 \leq \rho\epsilon \leq 1/4$ and the fact that $|z - 1/2| \leq 1/2$ for any $z \in [0,1]$, it must be true that $(|z-1/2| - \rho\epsilon)^2 \leq ((1/2) - \rho\epsilon)^2$ for any $z \in [0,1]$. It follows from (23) that $((1/2) - \rho\epsilon)^2 \geq (1/4) + (\epsilon^2 N_{\min}/2\ln(\zeta\delta))$, which implies that the minimum sample size can be taken as

$$N_{\min} = \left\lceil 2\rho\left(\frac{1}{\epsilon} - \rho\right)\ln\frac{1}{\zeta\delta} \right\rceil. \quad (24)$$

On the other hand, applying the function $\mathscr{D}(\cdot,\cdot)$ defined by (22) to (7) gives

$$N_{\max} = \min\left\{ n \in \mathbb{N} : \left(\left|\frac{k}{n}-\frac{1}{2}\right| - \rho\epsilon\right)^2 \geq \frac{1}{4} + \frac{\epsilon^2 n}{2\ln(\zeta\delta)} \text{ for all nonnegative integer } k \text{ not exceeding } n \right\}. \quad (25)$$

Since $(|z-1/2| - \rho\epsilon)^2 \geq 0$ for any $z \in [0,1]$, it follows from (25) that $(1/4) + (\epsilon^2 N_{\max}/2\ln(\zeta\delta)) \leq 0$, which implies that maximum sample size can be taken as

$$N_{\max} = \left\lceil \frac{1}{2\epsilon^2}\ln\frac{1}{\zeta\delta} \right\rceil. \quad (26)$$

Therefore, the sample sizes $n_1, \ldots, n_s$ can be chosen as functions of ρ, ζ, ϵ, and δ which satisfy the following constraint:

$$N_{\min} \leq n_1 < \cdots < n_{s-1} < N_{\max} \leq n_s. \quad (27)$$

In particular, if the number of stages s is given and the group sizes are expected to be approximately equal, then the sample sizes, $n_1, \ldots, n_s$, for all stages can be obtained by substituting $N_{\min}$ defined by (24) and $N_{\max}$ defined by (26) into (8). For example, if the values of design parameters are $\epsilon = 0.05$, $\delta = 0.05$, $\rho = 3/4$, $\zeta = 2.6759$ and $s = 7$, then the sample sizes of this sampling scheme are calculated as

$$n_1 = 59, \quad n_2 = 116, \quad n_3 = 173, \quad n_4 = 231, \quad n_5 = 288, \quad n_6 = 345, \quad n_7 = 403. \quad (28)$$

The stopping rule is completely determined by substituting the values of design parameters into (21).

4.2. Uniform Controllability of Coverage Probability. Clearly, for prespecified ϵ, δ, and ρ, the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ depends on the parameter ζ, the number of stages s, and the sample sizes $n_1, \ldots, n_s$. As illustrated in Section 4.1, the number of stages s and the sample sizes $n_1, \ldots, n_s$ can be defined as functions of $\zeta \in (0, 1/\delta)$. That is, the stopping rule can be parameterized by ζ. Accordingly, for any $p \in (0,1)$, the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ becomes a function of ζ. The following theorem shows that it suffices to choose $\zeta \in (0, 1/\delta)$ small enough to guarantee the prespecified confidence level.

Theorem 2. *Let $\epsilon, \delta \in (0,1)$ and $\rho \in (0,1]$ be fixed. Assume that the number of stages s and the sample sizes $n_1, \ldots, n_s$ are functions of $\zeta \in (0, 1/\delta)$ such that the constraint* (27) *is satisfied. Then,* $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ *is no less than* $1 - \delta$ *for any* $p \in (0,1)$ *provided that*

$$0 < \zeta \leq \frac{1}{\delta}\exp\left(\frac{\ln(\delta/2) + \ln\left[1 - \exp\left(-2\epsilon^2\right)\right]}{4\epsilon\rho(1-\rho\epsilon)}\right). \quad (29)$$

See Appendix B for a proof. For Theorem 2 to be valid, the choice of sample sizes is very flexible. Particularly, the sample sizes can be arithmetic or geometric progressions or any others, as long as the constraint (27) is satisfied. It can be seen that for the coverage probability to be uniformly controllable, the dilation coefficient ρ must be greater than 0. Theorem 2 asserts that there exists $\zeta > 0$ such that the coverage probability is no less than $1 - \delta$, regardless of the associated binomial proportion p. For the purpose of reducing sampling cost, we want to have a value of ζ as large as possible such that the prespecified confidence level is guaranteed for any $p \in (0,1)$. This can be accomplished by the technical components introduced in Sections 2.1, 2.3, 2.4, and 2.5. Clearly, for every value of ρ, we can obtain a corresponding value of ζ (as large as possible) to ensure the desired confidence level. However, the performance of resultant stopping rules are different. Therefore, we can try a number of values of ρ and pick the best resultant stopping rule for practical use.

4.3. Asymptotic Optimality of Sampling Schemes. Now we shall provide an important reason why we propose the sampling scheme of that structure by showing its asymptotic optimality. Since the performance of a group sampling scheme will be close to its fully sequential counterpart, we investigate the optimality of the fully sequential sampling

scheme. In this scenario, the sample sizes $n_1, n_2, \ldots, n_s$ are consecutive integers such that

$$\left\lceil 2\rho\left(\frac{1}{\epsilon}-\rho\right)\ln\frac{1}{\zeta\delta}\right\rceil = n_1 < n_2 < \cdots < n_{s-1} < n_s = \left\lceil \frac{1}{2\epsilon^2}\ln\frac{1}{\zeta\delta}\right\rceil. \tag{30}$$

The fully sequential sampling scheme can be viewed as a special case of a group sampling scheme of $s = n_s - n_1 + 1$ stages and group size 1. Clearly, if δ, ζ and ρ are fixed, the sampling scheme is dependent only on ϵ. Hence, for any $p \in (0,1)$, if we allow ϵ to vary in $(0,1)$, then the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ and the average sample number $\mathbb{E}[\mathbf{n}]$ are functions of ϵ. We are interested in knowing the asymptotic behavior of these functions as $\epsilon \to 0$, since ϵ is usually small in practical situations. The following theorem provides us the desired insights.

Theorem 3. *Assume that $\delta \in (0,1)$, $\zeta \in (0, 1/\delta)$ and $\rho \in (0,1]$ are fixed. Define $N(p,\epsilon,\delta,\zeta) = (2p(1-p)\ln(1/\zeta\delta))/\epsilon^2$ for $p \in (0,1)$ and $\epsilon \in (0,1)$. Then,*

$$\Pr\left\{\lim_{\epsilon\to 0}\frac{\mathbf{n}}{N(p,\epsilon,\delta,\zeta)} = 1 \mid p\right\} = 1, \tag{31}$$

$$\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = 2\Phi\left(\sqrt{2\ln\frac{1}{\zeta\delta}}\right) - 1, \tag{32}$$

$$\lim_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)} = 1, \tag{33}$$

for any $p \in (0,1)$.

See Appendix C for a proof. From (32), it can be seen that $\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = 1 - \delta$ for any $p \in (0,1)$ if $\zeta = (1/\delta)\exp(-(1/2)\mathscr{Z}^2_{\delta/2})$. Such value can be taken as an initial value for the coverage tuning parameter ζ. In addition to providing guidance on the coverage tuning techniques, Theorem 3 also establishes the optimality of the sampling scheme. To see this, let $\mathscr{N}(p,\epsilon,\delta)$ denote the minimum sample size n required for a fixed-sample-size procedure to guarantee that $\Pr\{|\overline{X}_n - p| < \epsilon \mid p\} \geq 1 - \delta$ for any $p \in (0,1)$, where $\overline{X}_n = (\sum_{i=1}^n X_i)/n$. It is well known that from the central limit theorem,

$$\lim_{\epsilon\to 0}\frac{\mathscr{N}(p,\epsilon,\delta)}{p(1-p)\left(\mathscr{Z}_{\delta/2}/\epsilon\right)^2} = 1. \tag{34}$$

Applying (33), (34), and letting $\zeta = (1/\delta)\exp(-(1/2)\mathscr{Z}^2_{\delta/2})$, we have $\lim_{\epsilon\to 0}(\mathscr{N}(p,\epsilon,\delta)/N(p,\epsilon,\delta,\zeta)) = 1$ for $p \in (0,1)$ and $\delta \in (0,1)$, which implies the asymptotic optimality of the double-parabolic sampling scheme. By virtue of (33), an approximate formula for computing the average sample number is given as follows:

$$\mathbb{E}[\mathbf{n}] \approx N(p,\epsilon,\delta,\zeta) = \frac{2p(1-p)\ln(1/\zeta\delta)}{\epsilon^2}, \tag{35}$$

for $p \in (0,1)$ and $\epsilon \in (0,1)$. From (34), one obtains $\mathscr{N}(p,\epsilon,\delta) \approx p(1-p)(\mathscr{Z}_{\delta/2}/\epsilon)^2$, which is a well-known result in statistics. In situations that no information of p is available, one usually uses

$$N_{\text{normal}} \stackrel{\text{def}}{=} \left\lceil \frac{1}{4}\left(\frac{\mathscr{Z}_{\delta/2}}{\epsilon}\right)^2\right\rceil \tag{36}$$

as the sample size for estimating the binomial proportion p with prescribed margin of error ϵ and confidence level $1-\delta$. Since the sample size formula (36) can lead to under-coverage, researchers in many areas are willing to use a more conservative but rigorous sample size formula

$$N_{\text{ch}} \stackrel{\text{def}}{=} \left\lceil \frac{\ln(2/\delta)}{2\epsilon^2}\right\rceil, \tag{37}$$

which is derived from the Chernoff-Hoeffding bound [32, 33]. Comparing (35) and (37), one can see that under the premise of guaranteeing the prescribed confidence level $1-\delta$, the double-parabolic sampling scheme can lead to a substantial reduction of sample number when the unknown binomial proportion p is close to 0 or 1.

4.4. Bounds on Distribution and Expectation of Sample Number. We shall derive analytic bounds for the cumulative distribution function and expectation of the sample number $\mathbf{n}$ associated with the double-parabolic sampling scheme. In this direction, we have obtained the following results.

Theorem 4. *Let $p \in (0, 1/2]$. Define $a_\ell = (1/2) - \rho\epsilon - \sqrt{(1/4) + (\epsilon^2 n_\ell/2\ln(\zeta\delta))}$ for $\ell = 1, \ldots, s$. Let τ denote the index of stage such that $a_{\tau-1} \leq p < a_\tau$. Then, $\Pr\{\mathbf{n} > n_\ell \mid p\} \leq \exp(n_\ell \mathscr{M}(a_\ell, p))$ for $\tau \leq \ell < s$. Moreover, $\mathbb{E}[\mathbf{n}] \leq n_\tau + \sum_{\ell=\tau}^{s-1}(n_{\ell+1} - n_\ell)\exp(n_\ell \mathscr{M}(a_\ell, p))$.*

See Appendix D for a proof. By the symmetry of the double-parabolic sampling scheme, similar analytic bounds for the distribution and expectation of the sample number can be derived for the case that $p \in [1/2, 1)$.

5. Comparison of Computational Methods

In this section, we shall compare various computational methods. First, we will illustrate why a frequently used method of evaluating the coverage probability based on gridding the parameter space is not rigorous and is less efficient as compared to the adapted B&B algorithm. Second, we will introduce the Adaptive Maximum Checking Algorithm of [21] which has better computational efficiency as compared to the adapted B&B algorithm. Third, we will explain that it is more advantageous in terms of numerical accuracy to work with the complementary coverage probability as compared to direct evaluation of the coverage probability. Finally, we will compare the computational methods of Chen [14–18] and Frey [13] for the design of sequential procedures for estimating a binomial proportion.

∇ Choose initial step size $d > \eta$.
∇ Let $F \leftarrow 0$, $T \leftarrow 0$ and $b \leftarrow \overline{\theta}$.
∇ While $F = T = 0$, do the following:
  ⋄ Let st $\leftarrow 0$ and $\ell \leftarrow 2$;
  ⋄ While st = 0, do the following:
    ⋆ Let $\ell \leftarrow \ell - 1$ and $d \leftarrow d2^{\ell}$.
    ⋆ If $b - d > \underline{\theta}$, then let $a \leftarrow b - d$ and $T \leftarrow 0$.
      Otherwise, let $a \leftarrow \underline{\theta}$ and $T \leftarrow 1$.
    ⋆ If $\overline{C}(a, b) < \delta$, then let st $\leftarrow 1$ and $b \leftarrow a$.
    ⋆ If $d < \eta$, then let st $\leftarrow 1$ and $F \leftarrow 1$.
∇ Return F.

Algorithm 2

5.1. Verifying Coverage Guarantee without Gridding Parameter Space. For purpose of constructing a sampling scheme so that the prescribed confidence level $1 - \delta$ is guaranteed, an essential task is to determine whether the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ associated with a given stopping rule is no less than $1 - \delta$. In other words, it is necessary to compare the infimum of coverage probability with $1 - \delta$. To accomplish such a task of checking coverage guarantee, a natural method is to evaluate the infimum of coverage probability as follows:

(i) choose m grid points $p_1, \ldots, p_m$ from parameter space $(0, 1)$;

(ii) compute $c_j = \Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p_j\}$ for $j = 1, \ldots, m$;

(iii) Take $\min\{c_1, \ldots, c_m\}$ as $\inf_{p\in(0,1)}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$.

This method can be easily mistaken as an exact approach and has been frequently used for evaluating coverage probabilities in many problem areas.

It is not hard to show that if the sample size $\mathbf{n}$ of a sequential procedure has a support $\mathcal{S}$, then the coverage probability $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\}$ is discontinuous at $p \in \mathcal{P} \cap (0, 1)$, where $\mathcal{P} = \{(k/n) \pm \epsilon : k$ is a nonnegative integer no greater than $n \in \mathcal{S}\}$. The set $\mathcal{P}$ typically has a large number of parameter values. Due to the discontinuity of the coverage probability as a function of p, the coverage probabilities can differ significantly for two parameter values which are extremely close. This implies that an intolerable error can be introduced by taking the minimum of coverage probabilities of a finite number of parameter values as the infimum of coverage probability on the whole parameter space. So, if one simply uses the minimum of the coverage probabilities of a finite number of parameter values as the infimum of coverage probability to check the coverage guarantee, the sequential estimator $\widehat{\mathbf{p}}$ of the resultant stopping rule will fail to guarantee the prescribed confidence level.

In addition to the lack of rigorousness, another drawback of checking coverage guarantee based on the method of gridding parameter space is its low efficiency. A critical issue is on the choice of the number, m, of grid points. If the number m is too small, the induced error can be substantial. On the other hand, choosing a large number for m results in high computational complexity.

In contrast to the method based on gridding parameter space, the adapted B&B algorithm is a rigorous approach for checking coverage guarantee as a consequence of the mechanism for comparing the bounds of coverage probability with the prescribed confidence level. The algorithm is also efficient due to the mechanism of pruning branches.

5.2. Adaptive Maximum Checking Algorithm. As illustrated in Section 2, the techniques developed in [14–18] are sufficient to provide exact solutions for a wide range of sequential estimation problems. However, one of the four components, the adapted B&B algorithm, requires computing both the lower and upper bounds of the complementary coverage probability. To further reduce the computational complexity, it is desirable to have a checking algorithm which needs only one of the lower and upper bounds. For this purpose, Chen had developed the Adaptive Maximum Checking Algorithm (AMCA) in [21, Section 3.3] and [19, Section 2.7]. In the following introduction of the AMCA, we shall follow the description of [21]. The AMCA can be applied to a wide class of computational problems dependent on the following critical subroutine.

Determine whether a function $C(\theta)$ is smaller than a prescribed number δ for every value of θ contained in interval $[\underline{\theta}, \overline{\theta}]$.

Particularly, for checking the coverage guarantee in the context of estimating a binomial proportion, the parameter θ is the binomial proportion p and the function $C(\theta)$ is actually the complementary coverage probability. In many situations, it is impossible or very difficult to evaluate $C(\theta)$ for every value of θ in interval $[\underline{\theta}, \overline{\theta}]$, since the interval may contain infinitely many or an extremely large number of values. Similar to the adapted B&B algorithm, the purpose of AMCA is to reduce the computational complexity associated with the problem of determining whether the maximum of $C(\theta)$ over $[\underline{\theta}, \overline{\theta}]$ is less than δ. The only assumption required for AMCA is that, for any interval $[a, b] \subseteq [\underline{\theta}, \overline{\theta}]$, it is possible to compute an upper bound $\overline{C}(a, b)$ such that $C(\theta) \leq \overline{C}(a, b)$ for any $\theta \in [a, b]$ and that the upper bound converges to $C(\theta)$ as the interval width $b - a$ tends to 0. The backward AMCA proceeds as in Algorithm 2.

The output of the backward AMCA is a binary variable F such that "$F = 0$" means "$C(\theta) < \delta$" and "$F = 1$" means "$C(\theta) \geq \delta$." An intermediate variable T is introduced in the description of AMCA such that "$T = 1$" means that the left endpoint of the interval is reached. The backward AMCA starts from the right endpoint of the interval (i.e., $b = \overline{\theta}$) and attempts to find an interval $[a, b]$ such that $\overline{C}(a, b) < \delta$. If such an interval is available, then, attempt to go backward to find the next consecutive interval with twice width. If doubling the interval width fails to guarantee $\overline{C}(a, b) < \delta$, then try to repeatedly cut the interval width in half to ensure that $\overline{C}(a, b) < \delta$. If the interval width becomes smaller than a prescribed tolerance η, then AMCA declares that "$F = 1$." For our relevant statistical problems, if $C(\theta) \geq \delta$ for some $\theta \in [\underline{\theta}, \overline{\theta}]$, it is sure that "$F = 1$" will be declared. On the other hand, it is possible that "$F = 1$" is declared even though $C(\theta) < \delta$ for any $\theta \in [\underline{\theta}, \overline{\theta}]$. However, such situation can be made extremely rare and immaterial if we choose η to be a very small number. Moreover, this will only introduce negligible conservativeness in the evaluation of $C(\theta)$ if η is chosen to be sufficiently small (e.g., $\eta = 10^{-15}$). Clearly, the backward AMCA can be easily modified as forward AMCA. Moreover, the AMCA can also be easily modified as Adaptive Minimum Checking Algorithm (forward and backward). For checking the maximum of complementary coverage probability $\Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$, one can use the AMCA with $C(p) = \Pr\{|\hat{\mathbf{p}} - p| \geq \epsilon \mid p\}$ over interval $[0, 1/2]$. We would like to point out that, in contrast to the adapted B&B algorithm, it seems difficult to generalize the AMCA to problems involving multidimensional parameter spaces.

5.3. Working with Complementary Coverage Probability. We would like to point out that, instead of evaluating the coverage probability as in [13], it is better to evaluate the complementary coverage probability for purpose of reducing numerical error. The advantage of working on the complementary coverage probability can be explained as follows. Note that, in many cases, the coverage probability is very close to 1 and the complementary coverage probability is very close to 0. Since the absolute precision for computing a number close to 1 is much lower than the absolute precision for computing a number close to 0, the method of directly evaluating the coverage probability will lead to intolerable numerical error for problems involving small δ. As an example, consider a situation that the complementary coverage probability is in the order of 10^{-5}. Direct computation of the coverage probability can easily lead to an absolute error of the order of 10^{-5}. However, the absolute error of computing the complementary coverage probability can be readily controlled at the order of 10^{-9}.

5.4. Comparison of Approaches of Chen and Frey. As mentioned in the introduction, Frey published a paper [13] in The American Statistician (TAS) on the sequential estimation of a binomial proportion with prescribed margin of error and confidence level. The approaches of Chen and Frey are based on the same strategy as follows. First, construct a family of stopping rules parameterized by γ (and possibly other design parameters) so that the associated coverage probability $\Pr\{|\hat{\mathbf{p}} - p| < \epsilon \mid p\}$ can be controlled by parameter γ in the sense that the coverage probability can be made arbitrarily close to 1 by increasing γ. Second, apply a bisection search method to determine the parameter γ so that the coverage probability is no less than the prescribed confidence level $1 - \delta$ for any $p \in (0, 1)$.

For the purpose of controlling the coverage probability, Frey [13] applied the inclusion principle previously proposed in [18, Section 3] and used in [14–17]. As illustrated in Section 3, the central idea of inclusion principle is to use a sequence of confidence intervals to construct stopping rules so that the sampling process is continued until a confidence interval is included by an interval defined in terms of the estimator and margin of error. Due to the inclusion relationship, the associated coverage probability can be controlled by the confidence coefficients of the sequence of confidence intervals. The critical value γ used by Frey plays the same role for controlling coverage probabilities as that of the coverage tuning parameter ζ used by Chen. Frey [13] stated stopping rules in terms of confidence limits. This way of expressing stopping rules is straightforward and insightful, since one can readily see the principle behind the construction. For convenience of practical use, Chen proposed to eliminate the necessity of computing confidence limits.

Similar to the AMCA proposed in [21, Section 3.3], the algorithm of Frey [13, Appendix] for checking coverage guarantee adaptively scans the parameter space based on interval bounding. The adaptive method used by Frey for updating step size is essentially the same as that of the AMCA. Ignoring the number 0.01 in Frey's expression "$\varepsilon_i = \min\{0.01, 2(p_{i-1} - p_{i-2})\}$," which has very little impact on the computational efficiency, Frey's step size ε_i can be identified as the adaptive step size d in the AMCA. The operation associated with "$\varepsilon_i = \min\{0.01, 2(p_{i-1} - p_{i-2})\}$" has a similar function as that of the command "Let st $\leftarrow 0$ and $\ell \leftarrow 2$" in the outer loop of the AMCA. The operation associated with Frey's expression "$p_{i-1} + \varepsilon_i/2^j,\ j \geq 0$" is equivalent to that of the command "Let $\ell \leftarrow \ell - 1$ and $d \leftarrow d2^{\ell}$" in the inner loop of the AMCA. Frey proposed to declare a failure of coverage guarantee if "the distance from p_{i-1} to the candidate value for p_i falls below 10^{-14}." The number "10^{-14}" actually plays the same role as "η" in the AMCA, where "$\eta = 10^{-15}$" is recommended by [21].

6. Numerical Results

In this section, we shall illustrate the proposed double-parabolic sampling scheme through examples. As demonstrated in Sections 2.2 and 4, the double-parabolic sampling scheme can be parameterized by the dilation coefficient ρ and the coverage tuning parameter ζ. Hence, the performance of the resultant stopping rule can be optimized with respect to $\rho \in (0, 1]$ and ζ by choosing various values of ρ from interval $(0, 1]$ and determining the corresponding values of

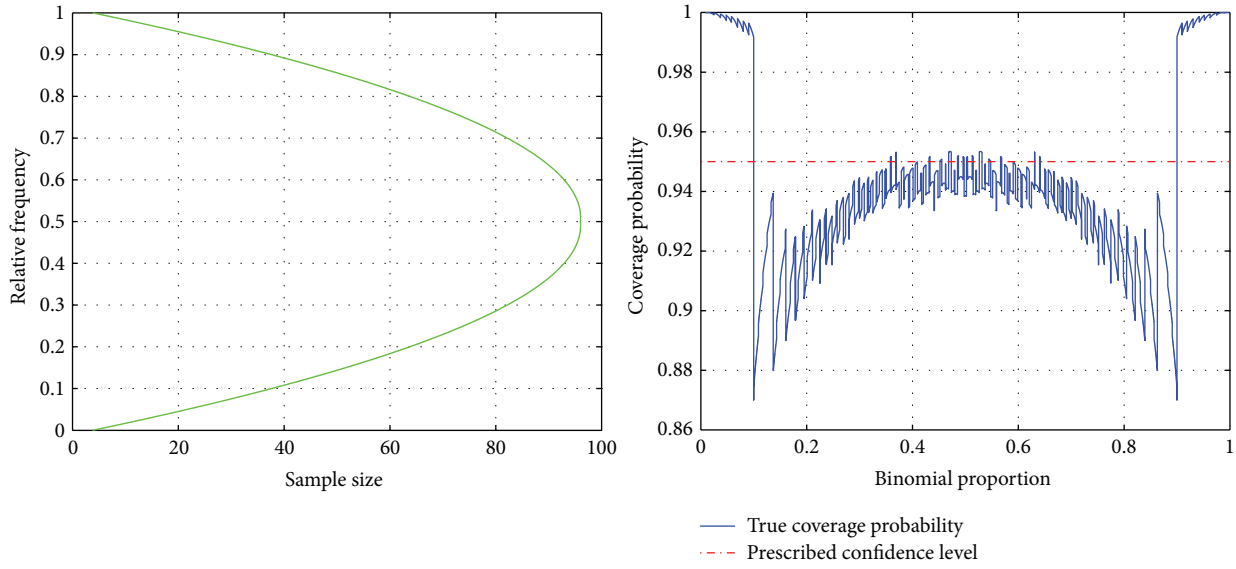

Figure 2: Double-parabolic sampling with $\epsilon = 0.1$, $\delta = 0.05$, $\rho = 1/10$, and $\zeta = 2.93$.

Table 1: Coverage tuning parameter.

ϵ	δ	ζ	ϵ	δ	ζ	ϵ	δ	ζ
0.1	0.1	2.0427	0.1	0.05	2.4174	0.1	0.01	3.0608
0.05	0.1	2.0503	0.05	0.05	2.5862	0.05	0.01	3.3125
0.02	0.1	2.1725	0.02	0.05	2.5592	0.02	0.01	3.4461
0.01	0.1	2.1725	0.01	0.05	2.5592	0.01	0.01	3.4461

ζ by the computational techniques introduced in Section 2 to guarantee the desired confidence interval.

6.1. Asymptotic Analysis May Be Inadequate. For fully sequential cases, we have evaluated the double-parabolic sampling scheme with $\epsilon = 0.1$, $\delta = 0.05$, $\rho = 0.1$, and $\zeta = (1/\delta)\exp(-(1/2)\mathcal{Z}^2_{\delta/2}) \approx 2.93$. The stopping boundary is displayed in the left side of Figure 2. The function of coverage probability with respect to the binomial proportion is shown in the right side of Figure 2, which indicates that the coverage probabilities are generally substantially lower than the prescribed confidence level $1 - \delta = 0.05$. By considering $\epsilon = 0.1$ as a small number and applying the asymptotic theory, the coverage probability associated with the sampling scheme is expected to be close to 0.95. This numerical example demonstrates that although the asymptotic method is insightful and involves virtually no computation, it may not be adequate.

In general, the main drawback of an asymptotic method is that there is no guarantee of coverage probability. Although an asymptotical method asserts that if the margin of error ϵ tends to 0, the coverage probability will tend to the prespecified confidence level $1 - \delta$, it is difficult to determine how small the margin of error ϵ is sufficient for the asymptotic method to be applicable. Note that $\epsilon \to 0$ implies the average sample size tends to ∞. However, in reality, the sample sizes must be finite. Consequently, an asymptotic method inevitably introduces unknown statistical error. Since an asymptotic method does not necessarily guarantee the prescribed confidence level, it is not fair to compare its associated sample size with that of an exact method, which guarantees the prespecified confidence level.

This example also indicates that, due to the discrete nature of the problem, the coverage probability is a discontinuous and erratic function of p, which implies that Monte Carlo simulation is not suitable for evaluating the coverage performance.

6.2. Parametric Values of Fully Sequential Schemes. For fully sequential cases, to allow direct application of our double-parabolic sequential method, we have obtained values of coverage tuning parameter ζ, which guarantee the prescribed confidence levels, for double-parabolic sampling schemes with $\rho = 3/4$ and various combinations of (ϵ, δ) as shown in Table 1. We used the computational techniques introduced in Section 2 to obtain this table.

To illustrate the use of Table 1, suppose that one wants a fully sequential sampling procedure to ensure that $\Pr\{|\widehat{\mathbf{p}} - p| < 0.1 \mid p\} > 0.95$ for any $p \in (0, 1)$. This means that one can choose $\epsilon = 0.1$, $\delta = 0.05$ and the range of sample size is given by (30). From Table 1, it can be seen that the value of ζ corresponding to $\epsilon = 0.1$ and $\delta = 0.05$ is 2.4174. Consequently, the stopping rule is completely determined by substituting the values of design parameters $\epsilon = 0.1$,

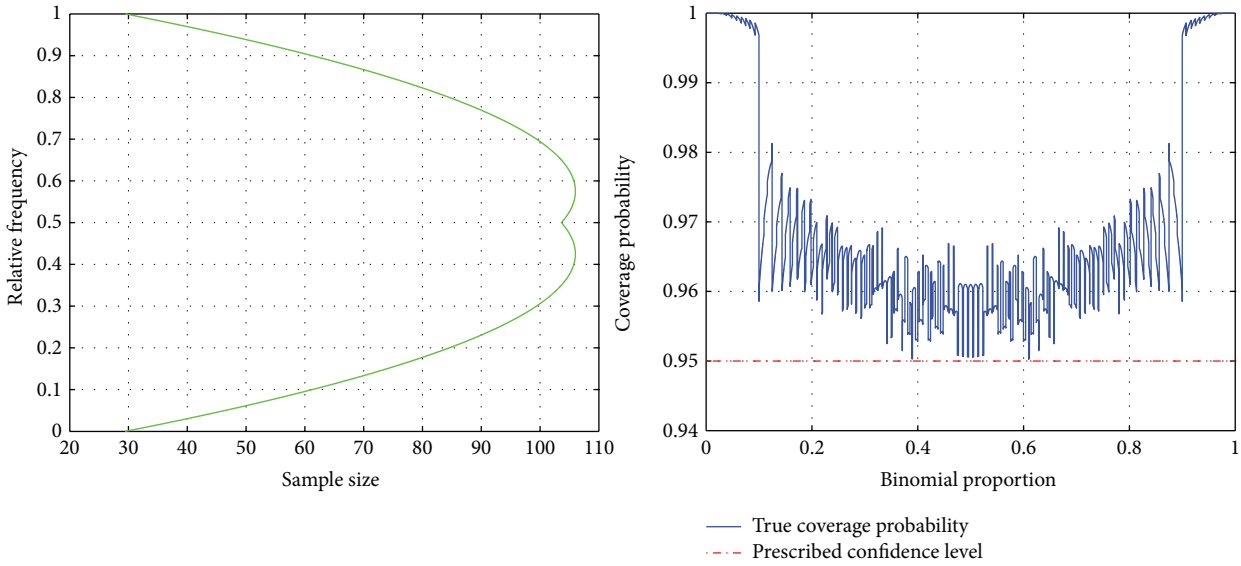

Figure 3: Double-parabolic sampling with $\epsilon = 0.1$, $\delta = 0.05$, $\rho = 3/4$, and $\zeta = 2.4174$.

Table 2: Coverage tuning parameter.

	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$	$s = 8$	$s = 9$	$s = 10$
$\epsilon = 0.1$	2.6583	2.6583	2.5096	2.5946	2.4459	2.6512	2.5096	2.4459
$\epsilon = 0.05$	2.6759	2.6759	2.6759	2.6759	2.6759	2.6759	2.6759	2.6759
$\epsilon = 0.02$	2.6725	2.6725	2.6725	2.6725	2.6725	2.6725	2.6725	2.6725
$\epsilon = 0.01$	2.6796	2.6796	2.6796	2.6796	2.6796	2.5875	2.6796	2.6796

$\delta = 0.05$, $\rho = 3/4$, and $\zeta = 2.4174$ into its definition. The stopping boundary of this sampling scheme is displayed in the left side of Figure 3. The function of coverage probability with respect to the binomial proportion is shown in the right side of Figure 3.

6.3. Parametric Values of Group Sequential Schemes. In many situations, especially in clinical trials, it is desirable to use group sequential sampling schemes. In Tables 2 and 3, assuming that sample sizes satisfy (8) for the purpose of having approximately equal group sizes, we have obtained parameters for concrete schemes by the computational techniques introduced in Section 2.

For dilation coefficient $\rho = 3/4$ and confidence parameter $\delta = 0.05$, we have obtained values of coverage tuning parameter ζ, which guarantee the prescribed confidence level 0.95, for double-parabolic sampling schemes, with the number of stages s ranging from 3 to 10, as shown in Table 2.

For dilation coefficient $\rho = 3/4$ and confidence parameter $\delta = 0.01$, we have obtained values of coverage tuning parameter ζ, which guarantee the prescribed confidence level 0.99, for double-parabolic sampling schemes, with the number of stages s ranging from 3 to 10, as shown in Table 3.

To illustrate the use of these tables, suppose that one wants a ten-stage sampling procedure of approximately equal group sizes to ensure that $\Pr\{|\widehat{\mathbf{p}} - p| < 0.01 \mid p\} > 0.99$ for any $p \in (0, 1)$. This means that one can choose $\epsilon = \delta = 0.01$, $s = 10$ and sample sizes satisfying (8). To obtain appropriate parameter values for the sampling procedure, one can look at Table 3 to find the coverage tuning parameter ζ corresponding to $\epsilon = 0.01$ and $s = 10$. From Table 3, it can be seen that ζ can be taken as 3.5753. Consequently, the stopping rule is completely determined by substituting the values of design parameters $\epsilon = 0.01$, $\delta = 0.01$, $\rho = 3/4$, $\zeta = 3.5753$, and $s = 10$ into its definition and (8). The stopping boundary of this sampling scheme and the function of coverage probability with respect to the binomial proportion are displayed, respectively, in the left and right sides of Figure 4.

6.4. Comparison of Sampling Schemes. We have conducted numerical experiments to investigate the impact of dilation coefficient ρ on the performance of our double-parabolic sampling schemes. Our computational experiences indicate that the dilation coefficient $\rho = 3/4$ is frequently a good choice in terms of average sample number and coverage probability. For example, consider the case that the margin of error is given as $\epsilon = 0.1$ and the prescribed confidence level is $1 - \delta$ with $\delta = 0.05$. For the double-parabolic sampling scheme with the dilation coefficient ρ chosen as $2/3, 3/4$, and 1, we have determined that, to ensure the prescribed confidence level $1 - \delta = 0.95$, it suffices to set the coverage tuning parameter ζ as 2.1, 2.4 and 2.4, respectively. The average sample numbers of these sampling schemes and the coverage probabilities as functions of the binomial proportion are shown, respectively, in the left and right sides

Table 3: Coverage tuning parameter.

	$s = 3$	$s = 4$	$s = 5$	$s = 6$	$s = 7$	$s = 8$	$s = 9$	$s = 10$
$\epsilon = 0.1$	3.3322	3.3322	3.3322	3.3322	3.3322	3.2709	3.0782	3.3322
$\epsilon = 0.05$	3.5074	3.5074	3.5074	3.5074	3.5074	3.5074	3.5074	3.5074
$\epsilon = 0.02$	3.5430	3.5430	3.5430	3.5430	3.5430	3.5430	3.5430	3.5430
$\epsilon = 0.01$	3.5753	3.5753	3.5753	3.5753	3.5753	3.5753	3.5753	3.5753

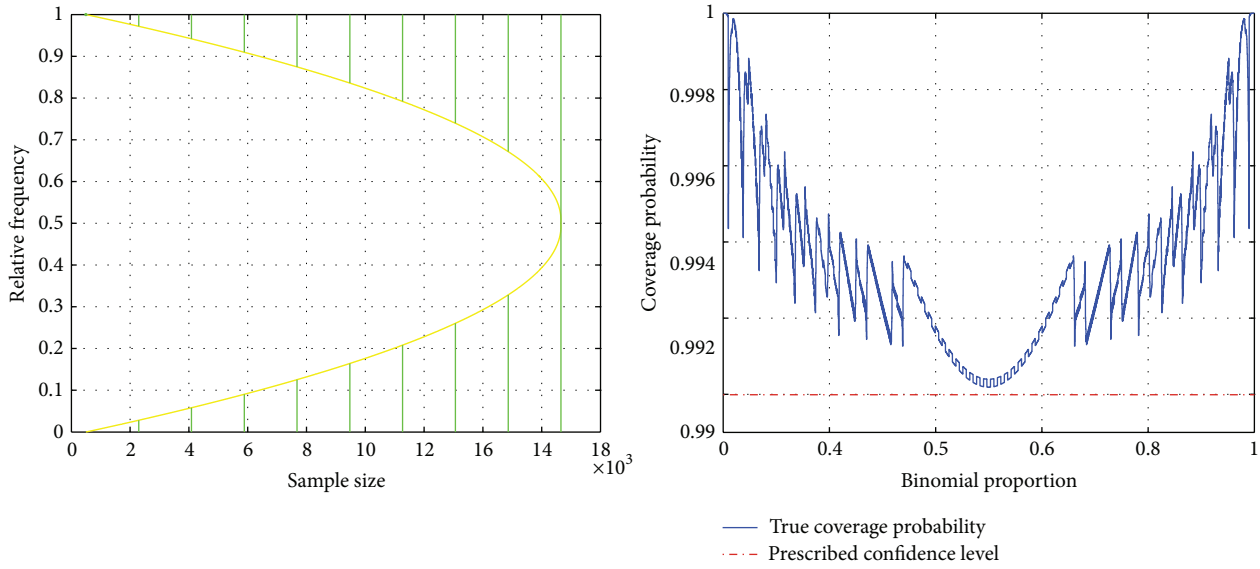

Figure 4: Double-parabolic sampling with $\epsilon = \delta = 0.01$, $s = 10$, $\rho = 3/4$, and $\zeta = 3.5753$.

of Figure 5. From Figure 5, it can be seen that a double-parabolic sampling scheme with dilation coefficient $\rho = 3/4$ has better performance in terms of average sample number and coverage probability as compared to that of the double-parabolic sampling scheme with smaller or larger values of dilation coefficient.

We have investigated the impact of confidence intervals on the performance of fully sequential sampling schemes constructed from the inclusion principle. We have observed that the stopping rule derived from Clopper-Pearson intervals generally outperforms the stopping rules derived from other types of confidence intervals. However, via appropriate choice of the dilation coefficient, the double-parabolic sampling scheme can perform uniformly better than the stopping rule derived from Clopper-Pearson intervals. To illustrate, consider the case that $\epsilon = 0.1$ and $\delta = 0.05$. For stopping rules derived from Clopper-Pearson intervals, Fishman's intervals, Wilson's intervals, and revised Wald intervals with $a = 4$, we have determined that to guarantee the prescribed confidence level $1 - \delta = 0.95$, it suffices to set the coverage tuning parameter ζ as $0.5, 1, 2.4$, and 0.37, respectively. For the stopping rule derived from Wald intervals, we have determined $\zeta = 0.77$ to ensure the confidence level, under the condition that the minimum sample size is taken as $\lceil (1/\epsilon) \ln(1/\zeta\delta) \rceil$. Recall that for the double-parabolic sampling scheme with $\rho = 3/4$, we have obtained $\zeta = 2.4$ for purpose of guaranteeing the confidence level. The average sample numbers of these sampling schemes are shown in Figure 6. From these plots, it can be seen that as compared to the stopping rule derived from Clopper-Pearson intervals, the stopping rule derived from the revised Wald intervals performs better in the region of p close to 0 or 1, but performs worse in the region of p in the middle of $(0, 1)$. The performance of stopping rules from Fishman's intervals (i.e., from Chernoff bound) and Wald intervals are obviously inferior as compared to that of the stopping rule derived from Clopper-Pearson intervals. It can be observed that the double-parabolic sampling scheme uniformly outperforms the stopping rule derived from Clopper-Pearson intervals.

6.5. Estimation with High Confidence Level. In some situations, we need to estimate a binomial proportion with a high confidence level. For example, one might want to construct a sampling scheme such that, for $\epsilon = 0.05$ and $\delta = 10^{-10}$, the resultant sequential estimator $\widehat{\mathbf{p}}$ satisfies $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} > 1 - \delta$ for any $p \in (0, 1)$. By working with the complementary coverage probability, we determined that it suffices to let the dilation coefficient $\rho = 3/4$ and the coverage tuning parameter $\zeta = 7.65$. The stopping boundary and the function of coverage probability with respect to the binomial proportion are displayed, respectively, in the left and right sides of Figure 7. As addressed in Section 5.3, it should be noted that it is impossible to obtain such a sampling scheme without working with the complementary coverage probability.

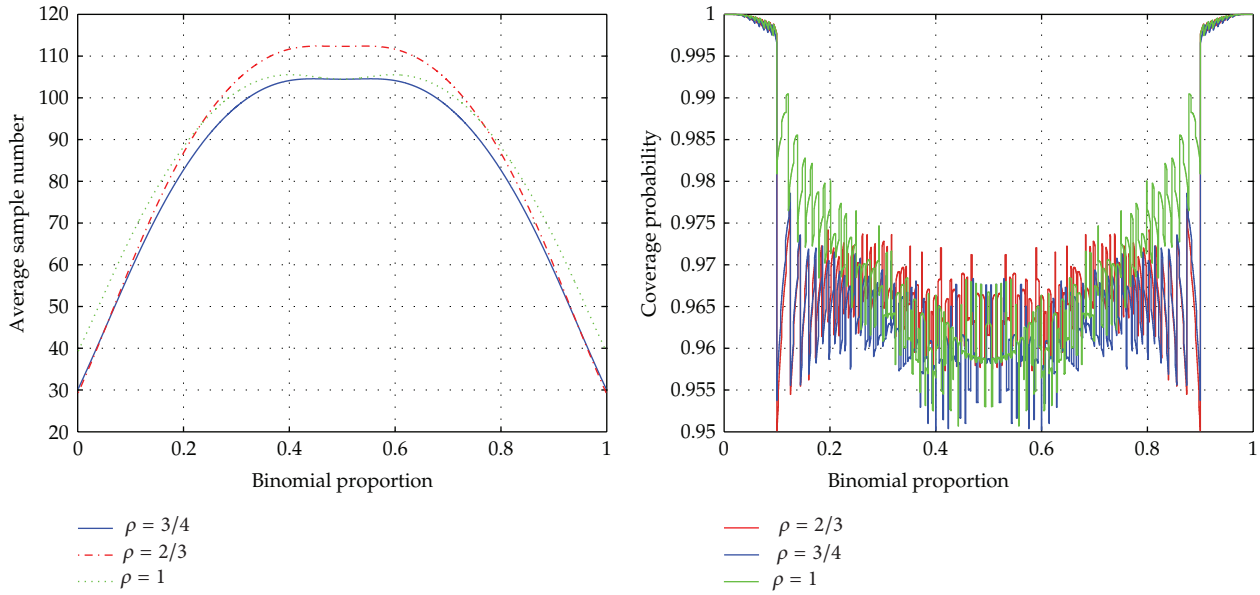

Figure 5: Double-parabolic sampling with various dilation coefficients.

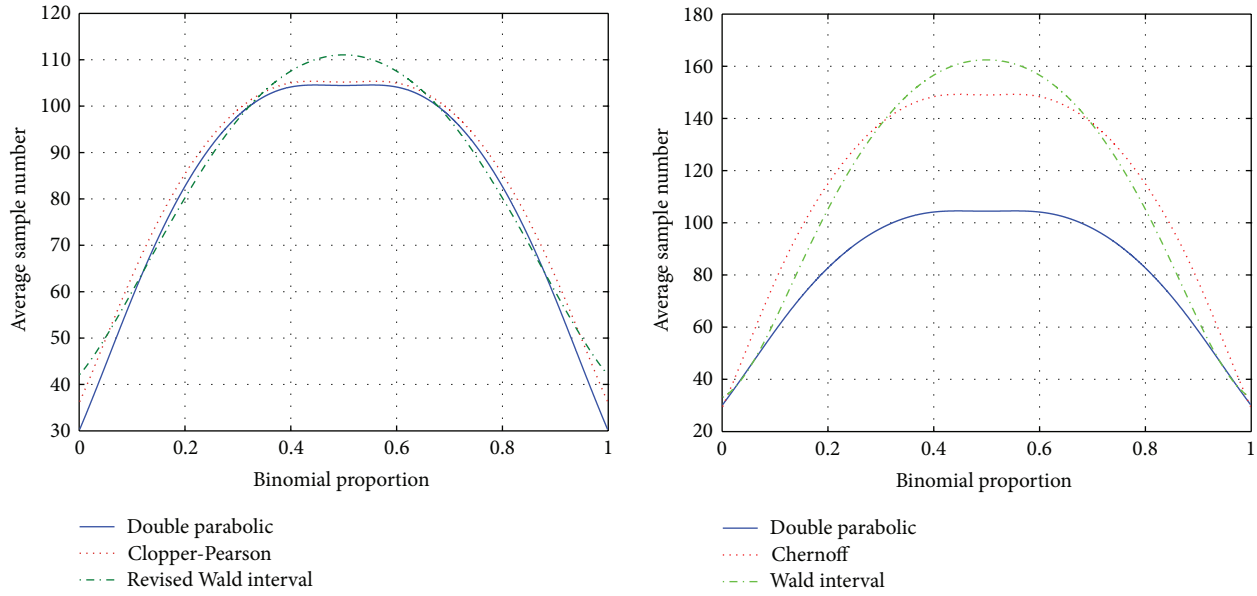

Figure 6: Comparison of average sample numbers.

7. Illustrative Examples for Clinical Trials

In this section, we shall illustrate the applications of our double-parabolic group sequential estimation method in clinical trials.

An example of our double-parabolic sampling scheme can be illustrated as follows. Assume that $\epsilon = \delta = 0.05$ is given and that the sampling procedure is expected to have 7 stages with sample sizes satisfying (8). Choosing $\rho = 3/4$, we have determined that it suffices to take $\zeta = 2.6759$ to guarantee that the coverage probability is no less than $1 - \delta = 0.95$ for all $p \in (0, 1)$. Accordingly, the sample sizes of this sampling scheme are calculated as 59, 116, 173, 231, 288, 345, and 403. This sampling scheme, with a sample path, is shown in the left side of Figure 8. In this case, the stopping rule can be equivalently described by virtue of Figure 8 as the following: continue sampling until $(\widehat{\mathbf{p}}_\ell, n_\ell)$ hit a green line at some stage. The coverage probability is shown in the right side of Figure 8.

To apply this estimation method in a clinical trial for estimating the proportion p of a binomial response with margin of error 0.05 and confidence level 95%, we can have seven groups of patients with group sizes 59, 57, 57, 58, 57, 57, and 58. In the first stage, we conduct experiment with

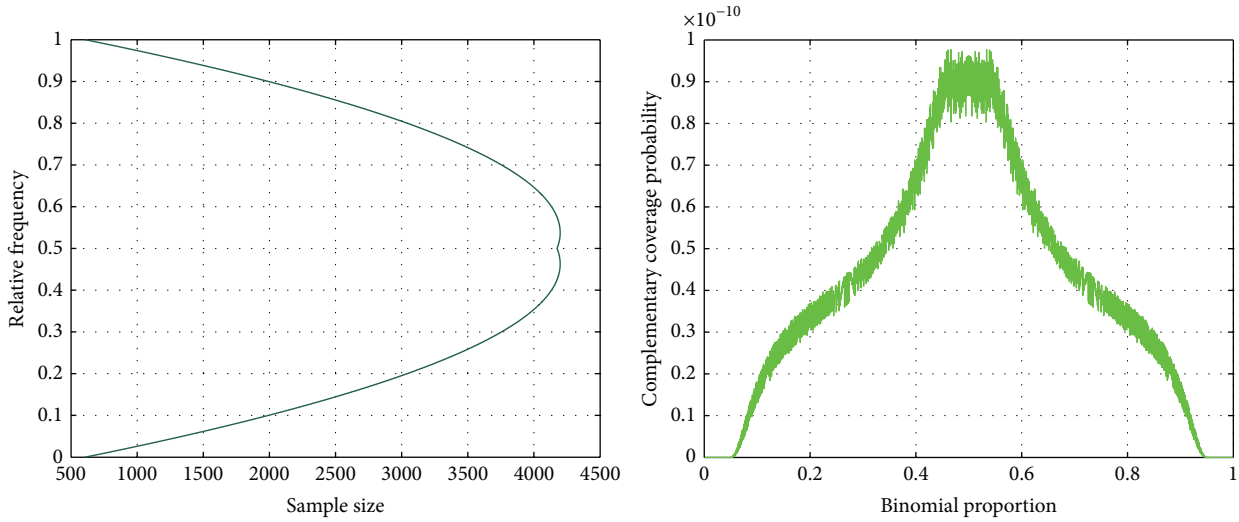

Figure 7: Double-parabolic sampling with $\epsilon = 0.05$, $\delta = 10^{-10}$, $\rho = 3/4$, and $\zeta = 7.65$.

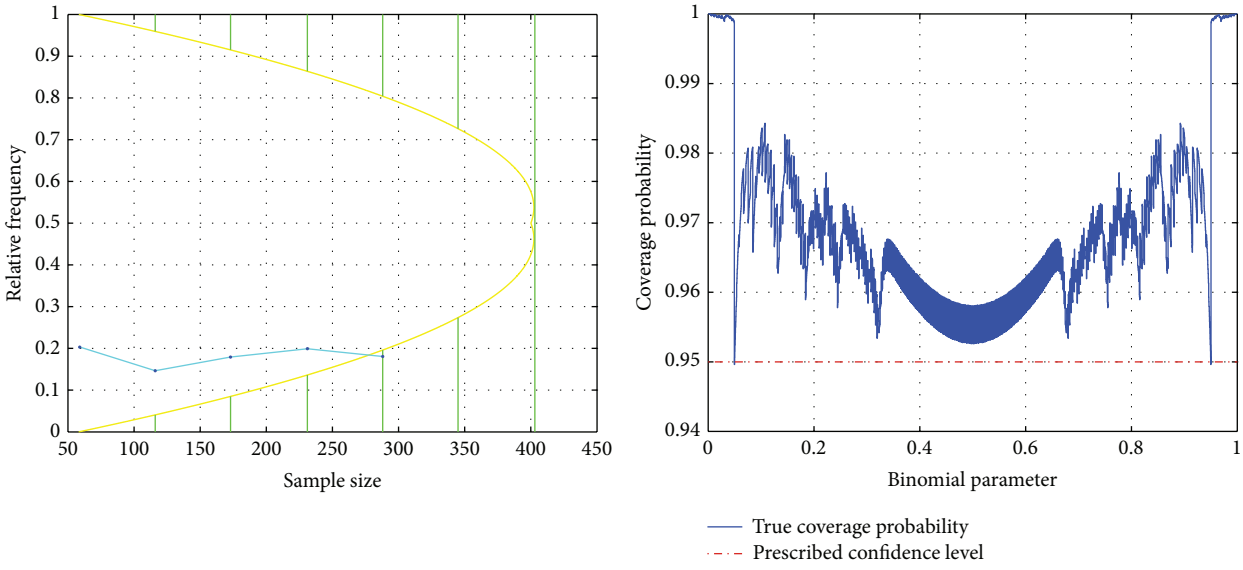

Figure 8: Double-parabolic sampling with $\epsilon = \delta = 0.05$, $s = 7$, $\rho = 3/4$, and $\zeta = 2.6759$.

the 59 patients of the first group. We observe the relative frequency of response and record it as $\widehat{\mathbf{p}}_1$. Suppose that there are 12 patients having positive responses, then the relative frequency at the first stage is $\widehat{\mathbf{p}}_1 = 12/59 = 0.2034$. With the values of $(\widehat{\mathbf{p}}_1, n_1) = (0.2034, 59)$, we check if the stopping rule is satisfied. This is equivalent to see if the point $(\widehat{\mathbf{p}}_1, n_1)$ hits a green line at the first stage. For such value of $(\widehat{\mathbf{p}}_1, n_1)$, it can be seen that the stopping condition is not fulfilled. So, we need to conduct the second stage of experiment with the 57 patients of the second group. We observe the response of these 57 patients. Suppose we observe that 5 patients among this group have positive responses. Then, we add 5 with 12, the number of positive responses before the second stage, to obtain 17 positive responses among $n_2 = 59 + 57 = 116$ patients. So, at the second stage, we get the relative frequency $\widehat{\mathbf{p}}_2 = 17/116 = 0.1466$. Since the stopping rule is not satisfied with the values of $(\widehat{\mathbf{p}}_2, n_2) = (0.1466, 116)$, we need to conduct the third stage of experiment with the 57 patients of the third group. Suppose we observe that 14 patients among this group have positive responses. Then, we add 14 with 17, the number of positive responses before the third stage, to get 31 positive responses among $n_3 = 59 + 57 + 57 = 173$ patients. So, at the third stage, we get the relative frequency $\widehat{\mathbf{p}}_3 = 31/173 = 0.1792$. Since the stopping rule is not satisfied with the values of $(\widehat{\mathbf{p}}_3, n_3) = (0.1792, 173)$, we need to conduct the fourth stage of experiment with the 58 patients of the fourth group. Suppose we observe that 15 patients among this group have positive responses. Then, we add 15 with 31, the number of positive responses before the fourth stage, to get 46 positive responses among $n_4 = 59 + 57 + 57 + 58 = 231$ patients. So, at the fourth stage, we get the relative frequency

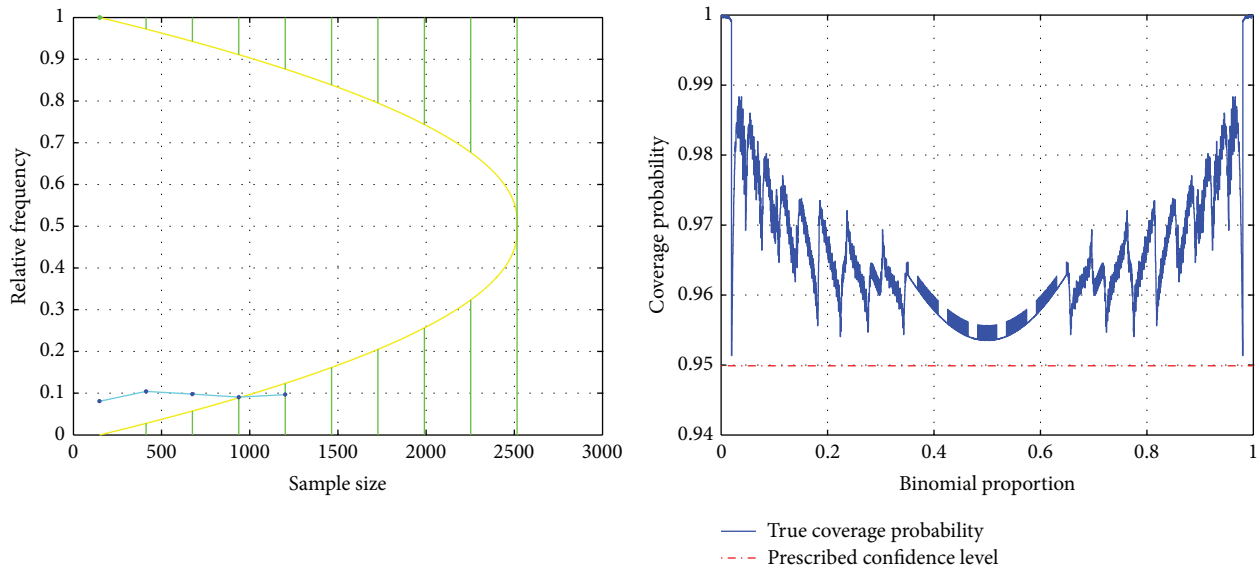

Figure 9: Double-parabolic sampling with $\epsilon = 0.02$, $\delta = 0.05$, $s = 10$, $\rho = 3/4$, and $\zeta = 2.6725$.

$\widehat{\mathbf{p}}_4 = 46/231 = 0.1991$. Since the stopping rule is not satisfied with the values of $(\widehat{\mathbf{p}}_4, n_4) = (0.1991, 231)$, we need to conduct the fifth stage of experiment with the 57 patients of the fifth group. Suppose we observe that 6 patients among this group have positive responses. Then, we add 6 with 46, the number of positive responses before the fifth stage, to get 52 positive responses among $n_5 = 59 + 57 + 57 + 58 + 57 = 288$ patients. So, at the fifth stage, we get the relative frequency $\widehat{\mathbf{p}}_5 = 52/288 = 0.1806$. It can be seen that the stopping rule is satisfied with the values of $(\widehat{\mathbf{p}}_5, n_5) = (0.1806, 288)$. Therefore, we can terminate the sampling experiment and take $\widehat{\mathbf{p}} = 52/288 = 0.1806$ as an estimate of the proportion of the whole population having positive responses. With a 95% confidence level, one can believe that the difference between the true value of p and its estimate $\widehat{\mathbf{p}} = 0.1806$ is less than 0.05.

In this experiment, we only use 288 samples to obtain the estimate for p. Except the round-off error, there is no other source of error for reporting statistical accuracy, since no asymptotic approximation is involved. As compared to fixed-sample-size procedure, we achieved a substantial save of samples. To see this, one can check that using the rigorous formula (37) gives a sample size 738, which is overly conservative. From the classical approximate formula (35), the sample size is determined as 385, which has been known to be insufficient to guarantee the prescribed confidence level 95%. The exact method of [34] shows that at least 391 samples are needed. As compared to the best-fixed-sample size obtained by the method of [34], the reduction of sample sizes resulted from our double-parabolic sampling scheme is $391 - 288 = 103$. It can be seen that the fixed-sample-size procedure wastes $103/288 = 35.76\%$ samples as compared to our group sequential method, which is also an exact method. This percentage may not be serious if it were a save of a number of simulation runs. However, as the number count is for patients, the reduction of samples is important for ethical and economical reasons. Using our group sequential method, the worst-case sample size is equal to 403, which is only 12 more than the minimum sample size of fixed-sample procedure. However, a lot of samples can be saved in the average case.

As ϵ or δ become smaller, the reduction of samples is more significant. For example, let $\epsilon = 0.02$ and $\delta = 0.05$, we have a double-parabolic sample scheme with 10 stages. The sampling scheme, with a sample path, is shown in the left side of Figure 9. The coverage probability is shown in the right side of Figure 9.

8. Conclusion

In this paper, we have reviewed recent development of group sequential estimation methods for a binomial proportion. We have illustrated the inclusion principle and its applications to various stopping rules. We have introduced computational techniques in the literature, which suffice for determining parameters of stopping rules to guarantee desired confidence levels. Moreover, we have proposed a new family of sampling schemes with stopping boundary of double-parabolic shape, which are parameterized by the coverage tuning parameter and the dilation coefficient. These parameters can be determined by the exact computational techniques to reduce the sampling cost, while ensuring prescribed confidence levels. The new family of sampling schemes are extremely simple in structure and asymptotically optimal as the margin of error tends to 0. We have established analytic bounds for the distribution and expectation of the sample number at the termination of the sampling process. We have obtained parameter values via the exact computational techniques for the proposed sampling schemes such that the confidence levels are guaranteed and that the sampling schemes are generally more efficient as compared to existing ones.

Appendices

A. Proof of Theorem 1

Consider function $g(x,z) = (x-z)^2/x(1-x)$ for $x \in (0,1)$ and $z \in [0,1]$. It can be checked that $\partial g(x,z)/\partial x = (x-z)[z(1-x)+x(1-z)][x(1-x)]^{-2}$, which shows that for any fixed $z \in [0,1]$, $-g(x,z)$ is a unimodal function of $x \in (0,1)$, with a maximum attained at $x = z$. By such a property of $g(x,z)$ and the definition of Wilson's confidence intervals, we have

$$\begin{aligned}
\left\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell\right\} &= \left\{0 < \widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon\right\} \\
&= \left\{0 < \widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le \widehat{\mathbf{p}}_\ell,\ g\left(L_\ell, \widehat{\mathbf{p}}_\ell\right) = \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon\right\} \\
&= \left\{\widehat{\mathbf{p}}_\ell > \epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon\right\}, \\
\left\{\widehat{\mathbf{p}}_\ell + \epsilon \ge U_\ell\right\} &= \left\{1 > \widehat{\mathbf{p}}_\ell + \epsilon \ge U_\ell\right\} \cup \left\{\widehat{\mathbf{p}}_\ell + \epsilon \ge 1\right\} \\
&= \left\{1 > \widehat{\mathbf{p}}_\ell + \epsilon \ge U_\ell \ge \widehat{\mathbf{p}}_\ell,\ g\left(U_\ell, \widehat{\mathbf{p}}_\ell\right) = \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
\cup\left\{\widehat{\mathbf{p}}_\ell + \epsilon \ge 1\right\} &= \left\{\widehat{\mathbf{p}}_\ell < 1-\epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \ge 1 - \epsilon\right\},
\end{aligned} \tag{A.1}$$

for $\ell = 1, \ldots, s$, where we have used the fact that $\{\widehat{\mathbf{p}}_\ell > \epsilon\} \subseteq \{L_\ell > 0\}$, $\{\widehat{\mathbf{p}}_\ell < 1-\epsilon\} \subseteq \{U_\ell < 1\}$ and $0 \le L_\ell \le \widehat{\mathbf{p}}_\ell \le U_\ell \le 1$. Recall that $0 < \epsilon < 1/2$. It follows that

$$\begin{aligned}
&\left\{\widehat{\mathbf{p}}_\ell - \epsilon \le L_\ell \le U_\ell \le \widehat{\mathbf{p}}_\ell + \epsilon\right\} \\
&= \left\{\epsilon < \widehat{\mathbf{p}}_\ell < 1-\epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell},\right. \\
&\qquad \left.\frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell + \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \ge 1 - \epsilon,\ \frac{\epsilon^2}{\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\left[1-\left(\widehat{\mathbf{p}}_\ell - \epsilon\right)\right]} \ge \frac{\mathcal{Z}_{\zeta\delta}^2}{n_\ell}\right\} \\
&= \left\{\epsilon < \widehat{\mathbf{p}}_\ell < 1-\epsilon,\ \left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \le \epsilon,\ \left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\} \\
&\quad \cup \left\{\widehat{\mathbf{p}}_\ell \ge 1 - \epsilon,\ \left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\} \\
&= \left\{\left(\left|\widehat{\mathbf{p}}_\ell - \frac{1}{2}\right| - \epsilon\right)^2 \ge \frac{1}{4} - n_\ell\left(\frac{\epsilon}{\mathcal{Z}_{\zeta\delta}}\right)^2\right\},
\end{aligned} \tag{A.2}$$

for $\ell = 1, \ldots, s$. This completes the proof of the theorem.

B. Proof of Theorem 2

By the assumption that $n_s \ge (1/2\epsilon^2)\ln(1/\zeta\delta)$, we have $(1/4) + (\epsilon^2 n_s/2\ln(\zeta\delta)) \le 0$ and, consequently, $\Pr\{(|\widehat{\mathbf{p}}_s - 1/2| - \rho\epsilon)^2 \ge (1/4) + (\epsilon^2 n_s/2\ln(\zeta\delta))\} = 1$. It follows from the definition of the sampling scheme that the sampling process must stop at or before the sth stage. In other words, $\Pr\{\mathbf{l} \le s\} = 1$. This allows one to write

$$\begin{aligned}
\Pr\left\{\left|\widehat{\mathbf{p}} - p\right| \ge \epsilon \mid p\right\} &= \sum_{\ell=1}^{s} \Pr\left\{\left|\widehat{\mathbf{p}} - p\right| \ge \epsilon,\ \mathbf{l} = \ell \mid p\right\} \\
&= \sum_{\ell=1}^{s} \Pr\left\{\left|\widehat{\mathbf{p}}_\ell - p\right| \ge \epsilon,\ \mathbf{l} = \ell \mid p\right\} \\
&\le \sum_{\ell=1}^{s} \Pr\left\{\left|\widehat{\mathbf{p}}_\ell - p\right| \ge \epsilon \mid p\right\},
\end{aligned} \tag{B.1}$$

for $p \in (0,1)$. By virtue of the well-known Chernoff-Hoeffding bound [32, 33], we have

$$\Pr\left\{\left|\widehat{\mathbf{p}}_\ell - p\right| \ge \epsilon \mid p\right\} \le 2\exp\left(-2n_\ell\epsilon^2\right), \tag{B.2}$$

for $\ell = 1, \ldots, s$. Making use of (B.1), (B.2), and the fact that $n_1 \ge 2\rho((1/\epsilon) - \rho)\ln(1/\zeta\delta)$ as can be seen from (30), we have

$$\begin{aligned}
&\Pr\left\{\left|\widehat{\mathbf{p}} - p\right| \ge \epsilon \mid p\right\} \\
&\le 2\sum_{\ell=1}^{s} \exp\left(-2n_\ell\epsilon^2\right) \\
&\le 2\sum_{m=n_1}^{\infty} \exp\left(-2m\epsilon^2\right) = \frac{2\exp\left(-2n_1\epsilon^2\right)}{1-\exp\left(-2\epsilon^2\right)} \\
&\le \frac{2\exp\left(-2\epsilon^2 \times 2\rho\left((1/\epsilon) - \rho\right)\ln(1/\zeta\delta)\right)}{1-\exp\left(-2\epsilon^2\right)} \\
&= \frac{2\exp\left(4\epsilon\rho(1-\rho\epsilon)\ln(\zeta\delta)\right)}{1-\exp\left(-2\epsilon^2\right)},
\end{aligned} \tag{B.3}$$

for any $p \in (0,1)$. Therefore, to guarantee that $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} \ge 1 - \delta$ for any $p \in (0,1)$, it is sufficient to choose ζ

such that $2\exp(4\epsilon\rho(1-\rho\epsilon)\ln(\zeta\delta)) \le \delta[1-\exp(-2\epsilon^2)]$. This inequality can be written as $4\epsilon\rho(1-\rho\epsilon)\ln(\zeta\delta) \le \ln(\delta/2) + \ln[1-\exp(-2\epsilon^2)]$ or, equivalently, $\zeta \le (1/\delta)\exp((\ln(\delta/2) + \ln[1-\exp(-2\epsilon^2)])/4\epsilon\rho(1-\rho\epsilon))$. The proof of the theorem is thus completed.

C. Proof of Theorem 3

First, we need to show that $\Pr\{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta)) = 1 \mid p\} = 1$ for any $p \in (0,1)$. Clearly, the sample number $\mathbf{n}$ is a random number dependent on ϵ. Note that for any $\omega \in \Omega$, the sequences $\{\overline{X}_{\mathbf{n}(\omega)}(\omega)\}_{\epsilon\in(0,1)}$ and $\{\overline{X}_{\mathbf{n}(\omega)-1}(\omega)\}_{\epsilon\in(0,1)}$ are subsets of $\{\overline{X}_m(\omega)\}_{m=1}^{\infty}$. By the strong law of large numbers, for almost every $\omega \in \Omega$, the sequence $\{\overline{X}_m(\omega)\}_{m=1}^{\infty}$ converges to p. Since every subsequence of a convergent sequence must converge, it follows that the sequences $\{\overline{X}_{\mathbf{n}(\omega)}(\omega)\}_{\epsilon\in(0,1)}$ and $\{\overline{X}_{\mathbf{n}(\omega)-1}(\omega)\}_{\epsilon\in(0,1)}$ converge to p as $\epsilon \to 0$ provided that $\mathbf{n}(\omega) \to \infty$ as $\epsilon \to 0$. Since it is certain that $\mathbf{n} \ge 2\rho((1/\epsilon) - \rho)\ln(1/\zeta\delta) \to \infty$ as $\epsilon \to 0$, we have that $\{\lim_{\epsilon\to 0}((\mathbf{n}-1)/\mathbf{n}) = 1\}$ is a sure event. It follows that $B = \{\lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}-1} = p,\ \lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}} = p,\ \lim_{\epsilon\to 0}((\mathbf{n}-1)/\mathbf{n}) = 1\}$ is an almost sure event. By the definition of the sampling scheme, we have

$$A = \left\{\left(\left|\overline{X}_{\mathbf{n}-1} - \frac{1}{2}\right| - \rho\epsilon\right)^2 < \frac{1}{4} + \frac{\epsilon^2(\mathbf{n}-1)}{2\ln(\zeta\delta)},\ \left(\left|\overline{X}_{\mathbf{n}} - \frac{1}{2}\right| - \rho\epsilon\right)^2 \ge \frac{1}{4} + \frac{\epsilon^2\mathbf{n}}{2\ln(\zeta\delta)}\right\} \tag{C.1}$$

as a sure event. Hence, $A \cap B$ is an almost sure event. Define $C = \{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta)) = 1\}$. We need to show that C is an almost sure event. For this purpose, we let $\omega \in A \cap B$ and expect to show that $\omega \in C$. As a consequence of $\omega \in A \cap B$,

$$\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} < \frac{\mathbf{n}(\omega)}{\mathbf{n}(\omega)-1} \times \frac{\left[(1/4) - \left(\left|\overline{X}_{\mathbf{n}(\omega)-1}(\omega) - (1/2)\right| - \rho\epsilon\right)^2\right]}{p(1-p)},\qquad \lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)-1}(\omega) = p,\qquad \lim_{\epsilon\to 0}\frac{\mathbf{n}(\omega)-1}{\mathbf{n}(\omega)} = 1. \tag{C.2}$$

By the continuity of the function $|x-1/2| - \rho\epsilon$ with respect to x and ϵ, we have

$$\begin{aligned}\limsup_{\epsilon\to 0}\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} &\le \lim_{\epsilon\to 0}\frac{\mathbf{n}(\omega)}{\mathbf{n}(\omega)-1} \times \frac{\left[(1/4) - \left(\left|\lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)-1}(\omega) - (1/2)\right| - \lim_{\epsilon\to 0}\rho\epsilon\right)^2\right]}{p(1-p)}\\ &= 1.\end{aligned} \tag{C.3}$$

On the other hand, as a consequence of $\omega \in A \cap B$,

$$\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} \ge \frac{\left[(1/4) - \left(\left|\overline{X}_{\mathbf{n}(\omega)}(\omega) - (1/2)\right| - \rho\epsilon\right)^2\right]}{p(1-p)},\qquad \lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)}(\omega) = p. \tag{C.4}$$

Making use of the continuity of the function $|x-1/2| - \rho\epsilon$ with respect to x and ϵ, we have

$$\begin{aligned}\liminf_{\epsilon\to 0}\frac{\mathbf{n}(\omega)}{N(p,\epsilon,\delta,\zeta)} &\ge \frac{\left[(1/4) - \left(\left|\lim_{\epsilon\to 0}\overline{X}_{\mathbf{n}(\omega)}(\omega) - (1/2)\right| - \lim_{\epsilon\to 0}\rho\epsilon\right)^2\right]}{p(1-p)}\\ &= 1.\end{aligned} \tag{C.5}$$

Combining (C.3) and (C.5) yields $\lim_{\epsilon\to 0}(\mathbf{n}(\omega)/N(p,\epsilon,\delta,\zeta)) = 1$ and thus $A \cap B \subseteq C$. This implies that C is an almost sure event and thus $\Pr\{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta)) = 1 \mid p\} = 1$ for $p \in (0,1)$.

Next, we need to show that $\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = 2\Phi(\sqrt{2\ln(1/\zeta\delta)}) - 1$ for any $p \in (0,1)$. For simplicity of notations, let $\sigma = \sqrt{p(1-p)}$ and $a = \sqrt{2\ln(1/\zeta\delta)}$. Note that $\Pr\{|\widehat{\mathbf{p}} - p| < \epsilon \mid p\} = \Pr\{|\overline{X}_{\mathbf{n}} - p| < \epsilon \mid p\} = \Pr\{\sqrt{\mathbf{n}}|\overline{X}_{\mathbf{n}} - p|/\sigma < \epsilon\sqrt{\mathbf{n}}/\sigma\}$. Clearly, for any $\eta \in (0,a)$,

$$\begin{aligned}&\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\ &\le \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma},\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \in [a-\eta, a+\eta]\right\} + \Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \notin [a-\eta, a+\eta]\right\}\\ &\le \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < a+\eta,\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \in [a-\eta, a+\eta]\right\} + \Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \notin [a-\eta, a+\eta]\right\}\\ &\le \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}} - p\right|}{\sigma} < a+\eta\right\} + \Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma} \notin [a-\eta, a+\eta]\right\},\end{aligned} \tag{C.6}$$

$$\begin{aligned}
&\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad\geq \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma},\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\in[a-\eta,a+\eta]\right\}\\
&\quad\geq \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a-\eta,\ \frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\in[a-\eta,a+\eta]\right\}\\
&\quad\geq \Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a-\eta\right\}\\
&\qquad-\Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\notin[a-\eta,a+\eta]\right\}.
\end{aligned}\tag{C.7}$$

Recall that we have established that $\mathbf{n}/N(p,\epsilon,\delta,\zeta)\to 1$ almost surely as $\epsilon\to 0$. This implies that $(\epsilon\sqrt{\mathbf{n}}/\sigma)\to a$ and $\mathbf{n}/N(p,\epsilon,\delta,\zeta)\to 1$ in probability as ϵ tends to zero. It follows from Anscombe's random central limit theorem [35] that as ϵ tends to zero, $\sqrt{\mathbf{n}}(\overline{X}_{\mathbf{n}}-p)/\sigma$ converges in distribution to a Gaussian random variable with zero mean and unit variance. Hence, from (C.6),

$$\begin{aligned}
&\limsup_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad\leq \lim_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a+\eta\right\}\\
&\qquad+\lim_{\epsilon\to 0}\Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\notin[a-\eta,a+\eta]\right\}\\
&\quad=2\Phi(a+\eta)-1
\end{aligned}\tag{C.8}$$

and from (C.7),

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad\geq \lim_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<a-\eta\right\}\\
&\qquad-\lim_{\epsilon\to 0}\Pr\left\{\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\notin[a-\eta,a+\eta]\right\}\\
&\quad=2\Phi(a-\eta)-1.
\end{aligned}\tag{C.9}$$

Since this argument holds for arbitrarily small $\eta\in(0,a)$, it must be true that

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad=\limsup_{\epsilon\to 0}\Pr\left\{\frac{\sqrt{\mathbf{n}}\left|\overline{X}_{\mathbf{n}}-p\right|}{\sigma}<\frac{\epsilon\sqrt{\mathbf{n}}}{\sigma}\right\}\\
&\quad=2\Phi(a)-1.
\end{aligned}\tag{C.10}$$

So, $\lim_{\epsilon\to 0}\Pr\{|\widehat{\mathbf{p}}-p|<\epsilon\mid p\}=\lim_{\epsilon\to 0}\Pr\{\sqrt{\mathbf{n}}|\overline{X}_{\mathbf{n}}-p|/\sigma<\epsilon\sqrt{\mathbf{n}}/\sigma\}=2\Phi(a)-1=2\Phi(\sqrt{2\ln(1/\zeta\delta)})-1$ for any $p\in(0,1)$.

Now, we focus our attention to show that $\lim_{\epsilon\to 0}(\mathbb{E}[\mathbf{n}]/N(p,\epsilon,\delta,\zeta))=1$ for any $p\in(0,1)$. For this purpose, it suffices to show that

$$\begin{aligned}
1-\eta&\leq\liminf_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)}\\
&\leq\limsup_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)}\\
&\leq 1+\eta,\quad\forall p\in(0,1),
\end{aligned}\tag{C.11}$$

for any $\eta\in(0,1)$. For simplicity of notations, we abbreviate $N(p,\epsilon,\delta,\zeta)$ as N in the sequel. Since we have established $\Pr\{\lim_{\epsilon\to 0}(\mathbf{n}/N(p,\epsilon,\delta,\zeta))=1\}=1$, we can conclude that

$$\lim_{\epsilon\to 0}\Pr\{(1-\eta)N\leq\mathbf{n}\leq(1+\eta)N\}=1.\tag{C.12}$$

Noting that

$$\begin{aligned}
\mathbb{E}[\mathbf{n}]&=\sum_{m=0}^{\infty}m\Pr\{\mathbf{n}=m\}\\
&\geq\sum_{(1-\eta)N\leq m\leq(1+\eta)N}m\Pr\{\mathbf{n}=m\}\\
&\geq(1-\eta)N\sum_{(1-\eta)N\leq m\leq(1+\eta)N}\Pr\{\mathbf{n}=m\},
\end{aligned}\tag{C.13}$$

we have

$$\mathbb{E}[\mathbf{n}]\geq(1-\eta)N\Pr\{(1-\eta)N\leq\mathbf{n}\leq(1+\eta)N\}.\tag{C.14}$$

Combining (C.12) and (C.14) yields

$$\begin{aligned}
&\liminf_{\epsilon\to 0}\frac{\mathbb{E}[\mathbf{n}]}{N(p,\epsilon,\delta,\zeta)}\\
&\quad\geq(1-\eta)\lim_{\epsilon\to 0}\Pr\{(1-\eta)N\leq\mathbf{n}\leq(1+\eta)N\}\\
&\quad=1-\eta.
\end{aligned}\tag{C.15}$$

On the other hand, using $\mathbb{E}[\mathbf{n}]=\sum_{m=0}^{\infty}\Pr\{\mathbf{n}>m\}$, we can write

$$\begin{aligned}
\mathbb{E}[\mathbf{n}]&=\sum_{0\leq m<(1+\eta)N}\Pr\{\mathbf{n}>m\}+\sum_{m\geq(1+\eta)N}\Pr\{\mathbf{n}>m\}\\
&\leq\lceil(1+\eta)N\rceil+\sum_{m\geq(1+\eta)N}\Pr\{\mathbf{n}>m\}.
\end{aligned}\tag{C.16}$$

Since $\limsup_{\epsilon\to 0}(\lceil(1+\eta)N\rceil/N(p,\epsilon,\delta,\zeta))=1+\eta$, for the purpose of establishing $\limsup_{\epsilon\to 0}(\mathbb{E}[\mathbf{n}]/N(p,\epsilon,\delta,\zeta))\leq 1+\eta$, it remains to show that

$$\limsup_{\epsilon\to 0}\frac{\sum_{m\geq(1+\eta)N}\Pr\{\mathbf{n}>m\}}{N(p,\epsilon,\delta,\zeta)}=0.\tag{C.17}$$

Consider functions $f(x) = (1/4) - (|x - 1/2| - \rho\epsilon)^2$ and $g(x) = x(1-x)$ for $x \in [0,1]$. Note that

$$\begin{aligned} |f(x) - g(x)| &= \left|(x-1/2)^2 - (|x-1/2| - \rho\epsilon)^2\right| \\ &= \rho\epsilon\left||2x-1| - \rho\epsilon\right| \le \rho\epsilon(1+\rho\epsilon), \end{aligned} \tag{C.18}$$

for all $x \in [0,1]$. For $p \in (0,1)$, there exists a positive number $\gamma < \min\{p, 1-p\}$ such that $|g(x) - g(p)| < (\eta/2)p(1-p)$ for any $x \in (p-\gamma, p+\gamma)$, since $g(x)$ is a continuous function of x. From now on, let $\epsilon > 0$ be sufficiently small such that $\rho\epsilon(1+\rho\epsilon) < (\eta/2)p(1-p)$. Then,

$$\begin{aligned} f(x) &\le g(x) + \rho\epsilon(1+\rho\epsilon) \\ &< g(p) + \frac{\eta}{2}p(1-p) + \rho\epsilon(1+\rho\epsilon) \\ &< (1+\eta)\,p(1-p), \end{aligned} \tag{C.19}$$

for all $x \in (p-\gamma, p+\gamma)$. This implies that

$$\begin{aligned} &\left\{\overline{X}_m \in (p-\gamma, p+\gamma)\right\} \\ &\quad\subseteq \left\{(1+\eta)\,p(1-p) \ge \frac{1}{4} - \left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2\right\} \end{aligned} \tag{C.20}$$

for all $m > 0$. Taking complementary events on both sides of (C.20) leads to

$$\begin{aligned} &\left\{(1+\eta)\,p(1-p) < \frac{1}{4} - \left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2\right\} \\ &\quad\subseteq \left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\}, \end{aligned} \tag{C.21}$$

for all $m > 0$. Since $(1+\eta)p(1-p) = ((1+\eta)N\epsilon^2/2\ln(1/\zeta\delta)) \le (m\epsilon^2/2\ln(1/\zeta\delta))$ for all $m \ge (1+\eta)N$, it follows that

$$\begin{aligned} &\left\{\frac{m\epsilon^2}{2\ln(1/\zeta\delta)} < \frac{1}{4} - \left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2\right\} \\ &\quad\subseteq \left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\}, \end{aligned} \tag{C.22}$$

for all $m \ge (1+\eta)N$. Therefore, we have shown that if ϵ is sufficiently small, then there exists a number $\gamma > 0$ such that

$$\begin{aligned} \{\mathbf{n} > m\} &\subseteq \left\{\left(\left|\overline{X}_m - \frac{1}{2}\right| - \rho\epsilon\right)^2 < \frac{1}{4} + \frac{m\epsilon^2}{2\ln(\zeta\delta)}\right\} \\ &\subseteq \left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\}, \end{aligned} \tag{C.23}$$

for all $m \ge (1+\eta)N$. Using this inclusion relationship and the Chernoff-Hoeffding bound [32, 33], we have

$$\Pr\{\mathbf{n} > m\} \le \Pr\left\{\overline{X}_m \notin (p-\gamma, p+\gamma)\right\} \le 2\exp\left(-2m\gamma^2\right), \tag{C.24}$$

for all $m \ge (1+\eta)N$ provided that $\epsilon > 0$ is sufficiently small. Letting $k = \lceil (1+\eta)N \rceil$ and using (C.24), we have

$$\begin{aligned} \sum_{m \ge (1+\eta)N} \Pr\{\mathbf{n} > m\} &= \sum_{m \ge k} \Pr\{\mathbf{n} > m\} \\ &\le \sum_{m \ge k} 2\exp\left(-2m\gamma^2\right) \\ &= \frac{2\exp\left(-2k\gamma^2\right)}{1 - \exp\left(-2\gamma^2\right)}, \end{aligned} \tag{C.25}$$

provided that ϵ is sufficiently small. Consequently,

$$\begin{aligned} &\limsup_{\epsilon \to 0} \frac{\sum_{m \ge (1+\eta)N} \Pr\{\mathbf{n} > m\}}{N(p, \epsilon, \delta, \zeta)} \\ &\quad\le \limsup_{\epsilon \to 0} \frac{2}{N} \frac{\exp\left(-2k\gamma^2\right)}{1 - \exp\left(-2\gamma^2\right)} = 0, \end{aligned} \tag{C.26}$$

since $k \to \infty$ and $N \to \infty$ as $\epsilon \to 0$. So, we have established (C.11). Since the argument holds for arbitrarily small $\eta > 0$, it must be true that $\lim_{\epsilon \to 0}(\mathbb{E}[\mathbf{n}]/N(p, \epsilon, \delta, \zeta)) = 1$ for any $p \in (0,1)$. This completes the proof of the theorem.

D. Proof of Theorem 4

Recall that $\mathbf{l}$ denotes the index of stage at the termination of the sampling process. Observing that

$$\begin{aligned} n_s - n_1 \ \Pr\{\mathbf{l} = 1\} &= n_s \Pr\{\mathbf{l} \le s\} - n_1 \Pr\{\mathbf{l} \le 1\} \\ &= \sum_{\ell=2}^{s} \left(n_\ell \Pr\{\mathbf{l} \le \ell\} - n_{\ell-1} \Pr\{\mathbf{l} < \ell\}\right) \\ &= \sum_{\ell=2}^{s} n_\ell \ \left(\Pr\{\mathbf{l} \le \ell\} - \Pr\{\mathbf{l} < \ell\}\right) \\ &\quad + \sum_{\ell=2}^{s} \left(n_\ell - n_{\ell-1}\right) \Pr\{\mathbf{l} < \ell\} \\ &= \sum_{\ell=2}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} \\ &\quad + \sum_{\ell=1}^{s-1} \left(n_{\ell+1} - n_\ell\right) \Pr\{\mathbf{l} \le \ell\}, \end{aligned} \tag{D.1}$$

we have $n_s - \sum_{\ell=1}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} = \sum_{\ell=1}^{s-1} (n_{\ell+1} - n_\ell)\Pr\{\mathbf{l} \le \ell\}$. Making use of this result and the fact $n_s = n_1 + \sum_{\ell=1}^{s-1}(n_{\ell+1} - n_\ell)$, we have

$$\begin{aligned}
\mathbb{E}[\mathbf{n}] &= \sum_{\ell=1}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} = n_s - \left(n_s - \sum_{\ell=1}^{s} n_\ell \Pr\{\mathbf{l} = \ell\} \right) \\
&= n_1 + \sum_{\ell=1}^{s-1} (n_{\ell+1} - n_\ell) - \sum_{\ell=1}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} \le \ell\} \\
&= n_1 + \sum_{\ell=1}^{\tau-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\} \\
&\quad + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\}.
\end{aligned} \tag{D.2}$$

By the definition of the stopping rule, we have

$$\begin{aligned}
\{\mathbf{l} > \ell\} &\subseteq \left\{ \left(\left| \widehat{\mathbf{p}}_\ell - \frac{1}{2} \right| - \rho\epsilon \right)^2 < \frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)} \right\} \\
&= \left\{ \rho\epsilon - \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} < \left| \widehat{\mathbf{p}}_\ell - \frac{1}{2} \right| \right. \\
&\qquad \left. < \rho\epsilon + \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} \right\} \\
&= \left\{ \rho\epsilon - \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} < \frac{1}{2} - \widehat{\mathbf{p}}_\ell \right. \\
&\qquad \left. < \rho\epsilon + \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}},\ \widehat{\mathbf{p}}_\ell \le \frac{1}{2} \right\} \\
&\quad \cup \left\{ \rho\epsilon - \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}} < \widehat{\mathbf{p}}_\ell - \frac{1}{2} \right. \\
&\qquad \left. < \rho\epsilon + \sqrt{\frac{1}{4} + \frac{\epsilon^2 n_\ell}{2\ln(\zeta\delta)}},\ \widehat{\mathbf{p}}_\ell > \frac{1}{2} \right\} \\
&\subseteq \left\{ a_\ell < \widehat{\mathbf{p}}_\ell < b_\ell \right\} \cup \left\{ 1 - b_\ell < \widehat{\mathbf{p}}_\ell < 1 - a_\ell \right\},
\end{aligned} \tag{D.3}$$

for $1 \le \ell < s$, where $b_\ell = (1/2) - \rho\epsilon + \sqrt{(1/4) + (\epsilon^2 n_\ell / 2\ln(\zeta\delta))}$ for $\ell = 1, \ldots, s-1$. By the assumption that ϵ and ρ are nonnegative, we have $1 - b_\ell - a_\ell = 2\rho\epsilon \ge 0$ for $\ell = 1, \ldots, s-1$. It follows from (D.3) that $\{\mathbf{l} > \ell\} \subseteq \{\widehat{\mathbf{p}}_\ell > a_\ell\}$ for $\ell = 1, \ldots, s-1$. By the definition of τ, we have $p < a_\ell$ for $\tau \le \ell < s$. Making use of this fact, the inclusion relationship $\{\mathbf{l} > \ell\} \subseteq \{\widehat{\mathbf{p}}_\ell > a_\ell\}$, $\ell = 1, \ldots, s-1$, and Chernoff-Hoeffding bound [32, 33], we have

$$\begin{aligned}
\Pr\{\mathbf{n} > n_\ell \mid p\} &= \Pr\{\mathbf{l} > \ell \mid p\} \\
&\le \Pr\{\widehat{\mathbf{p}}_\ell > a_\ell \mid p\} \le \exp\left(n_\ell \mathscr{M}(a_\ell, p)\right)
\end{aligned} \tag{D.4}$$

for $\tau \le \ell < s$. It follows from (D.2) and (D.4) that

$$\begin{aligned}
\mathbb{E}[\mathbf{n}] &\le n_1 + \sum_{\ell=1}^{\tau-1} (n_{\ell+1} - n_\ell) + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\} \\
&= n_\tau + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \Pr\{\mathbf{l} > \ell\} \\
&\le n_\tau + \sum_{\ell=\tau}^{s-1} (n_{\ell+1} - n_\ell) \exp\left(n_\ell \mathscr{M}(a_\ell, p)\right).
\end{aligned} \tag{D.5}$$

This completes the proof of the theorem.

Acknowledgment

This paper is supported in part by NIH/NCI Grants no. 1 P01 CA116676, P30 CA138292-01 and 5 P50 CA128613.

References

[1] S.-C. Chow, J. Shao, and H. Wang, *Sample Size Calculations in Clinical Research*, vol. 20 of *Chapman & Hall/CRC Biostatistics Series*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2008.

[2] C. Jennison and B. W. Turnbull, *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2000.

[3] Y. S. Chow and H. Robbins, "On the asymptotic theory of fixed-width sequential confidence intervals for the mean," *Annals of Mathematical Statistics*, vol. 36, pp. 457–462, 1965.

[4] B. K. Ghosh and P. K. Sen, *Handbook of Sequential Analysis*, vol. 118 of *Statistics: Textbooks and Monographs*, Marcel Dekker, New York, NY, USA, 1991.

[5] M. Ghosh, N. Mukhopadhyay, and P. K. Sen, *Sequential Estimation*, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1997.

[6] T. L. Lai, "Sequential analysis: some classical problems and new challenges," *Statistica Sinica*, vol. 11, no. 2, pp. 303–408, 2001.

[7] D. Siegmund, *Sequential Analysis*, Springer Series in Statistics, Springer, New York, NY, USA, 1985.

[8] L. Mendo and J. M. Hernando, "Improved sequential stopping rule for Monte Carlo simulation," *IEEE Transactions on Communications*, vol. 56, no. 11, pp. 1761–1764, 2008.

[9] M. Tanaka, "On a confidence interval of given length for the parameter of the binomial and the Poisson distributions," *Annals of the Institute of Statistical Mathematics*, vol. 13, pp. 201–215, 1961.

[10] S. Franzén, "Fixed length sequential confidence intervals for the probability of response," *Sequential Analysis*, vol. 20, no. 1-2, pp. 45–54, 2001.

[11] S. Franzén, "SPRT fixed length confidence intervals," *Communications in Statistics. Theory and Methods*, vol. 33, no. 2, pp. 305–319, 2004.

[12] A. Wald, *Sequential Analysis*, John Wiley & Sons, New York, NY, USA, 1947.

[13] J. Frey, "Fixed-width sequential confidence intervals for a proportion," *The American Statistician*, vol. 64, no. 3, pp. 242–249, 2010.

[14] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v1, September 2008.

[15] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v12, April 2009.

[16] X. Chen, "Multistage estimation of bounded-variable means," http://arxiv.org/abs/0809.4679v1, September 2008.

[17] X. Chen, "Estimating the parameters of binomial and Poisson distributions via multistage sampling," http://arxiv.org/abs/0810.0430v1, October 2008.

[18] X. Chen, "Confidence interval for the mean of a bounded random variable and its applications in point estimation," http://arxiv.org/abs/0802.3458v2, April 2009.

[19] X. Chen, "A new framework of multistage parametric inference," in *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IX*, usa, April 2010.

[20] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," *Econometrica*, vol. 28, pp. 497–520, 1960.

[21] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v16, November 2009.

[22] X. Chen, "A new framework of multistage estimation," http://arxiv.org/abs/0809.1241v4, December 2008.

[23] J. R. Schultz, F. R. Nichol, G. L. Elfring, and S. D. Weed, "Multiple stage procedures for drug screening," *Biometrics*, vol. 29, no. 2, pp. 293–300, 1973.

[24] H. Chen, "The accuracy of approximate intervals for a binomial parameter," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 514–518, 1990.

[25] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *Journal of the American Statistical Association*, vol. 22, pp. 209–212, 1927.

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

[27] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. I*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1968.

[28] G. S. Fishman, "Confidence intervals for the mean in the bounded case," *Statistics & Probability Letters*, vol. 12, no. 3, pp. 223–227, 1991.

[29] X. Chen, K. Zhou, and J. L. Aravena, "Explicit formula for constructing binomial confidence interval with guaranteed coverage probability," *Communications in Statistics. Theory and Methods*, vol. 37, no. 8–10, pp. 1173–1180, 2008.

[30] X. Chen, "Multistage estimation of bounded-variable means," http://arxiv.org/abs/0809.4679v2, October 2008.

[31] P. Massart, "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *The Annals of Probability*, vol. 18, no. 3, pp. 1269–1283, 1990.

[32] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, pp. 493–507, 1952.

[33] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

[34] X. Chen, "Exact computation of minimum sample size for estimation of binomial parameters," *Journal of Statistical Planning and Inference*, vol. 141, no. 8, pp. 2622–2632, 2011.

[35] F. J. Anscombe, "Sequential estimation," *Journal of the Royal Statistical Society. Series B*, vol. 15, pp. 1–21, 1953.