J. Appl. Probab. 1–27 (2024)
*doi:*10.1017/jpr.2024.43

UNCERTAINTY QUANTIFICATION AND CONFIDENCE INTERVALS FOR NAIVE RARE-EVENT ESTIMATORS

YUANLU BAI,* ** AND
HENRY LAM,* *** *Columbia University*

Abstract

We consider the estimation of rare-event probabilities using sample proportions output by naive Monte Carlo or collected data. Unlike using variance reduction techniques, this naive estimator does not have an a priori relative efficiency guarantee. On the other hand, due to the recent surge of sophisticated rare-event problems arising in safety evaluations of intelligent systems, efficiency-guaranteed variance reduction may face implementation challenges which, coupled with the availability of computation or data collection power, motivate the use of such a naive estimator. In this paper we study the uncertainty quantification, namely the construction, coverage validity, and tightness of confidence intervals, for rare-event probabilities using only sample proportions. In addition to the known normality, Wilson, and exact intervals, we investigate and compare them with two new intervals derived from Chernoff's inequality and the Berry–Esseen theorem. Moreover, we generalize our results to the natural situation where sampling stops by reaching a target number of rare-event hits. Our findings show that the normality and Wilson intervals are not always valid, but they are close to the newly developed valid intervals in terms of half-width. In contrast, the exact interval is conservative, but safely guarantees the attainment of the nominal confidence level. Our new intervals, while being more conservative than the exact interval, provide useful insights into understanding the tightness of the considered intervals.

Keywords: Rare-event estimation; confidence interval; relative error; sample proportion

2020 Mathematics Subject Classification: Primary 62A01
Secondary 00A72

1. Introduction

We consider the problem of estimating a minuscule probability, denoted $p = \mathbb{P}(A)$, for some rare event A, using data or Monte Carlo samples. This problem, known as rare-event estimation, is of wide interest to communities such as system reliability [22, 27, 28, 38], queueing systems [9, 11, 16, 25, 30, 32, 37], and finance and insurance [3, 4, 14, 18, 20, 21, 26], where it is crucial to estimate the likelihood of events which, though unlikely, can cause catastrophic impacts.

There are multiple prominent lines of work addressing this estimation problem, depending on how information is collected. In settings where real-world data are collected, methods based on extreme value theory [15, 17, 26, 36] are often used to extrapolate distributional tails

Received 3 May 2023; accepted 29 April 2024.
* Postal address: 500 West 120th Street, New York, NY, USA.
** Email address: yb2436@columbia.edu
*** Email address: henry.lam@columbia.edu

to assist such estimation. These methods are theoretically justified and widely applicable, but their performance could be affected by intricate hyperparameter choices that affect their accuracy and challenge the reliability in uncertainty quantification [17]. In settings where A is an event described by a simulable model, Monte Carlo methods can be used, and to speed up computation variance reduction tools such as importance sampling [24, 33, 35], conditional Monte Carlo [5, 31], and multi-level splitting [6, 19, 39] are often harnessed. While variance reduction is greatly beneficial in reducing the number of Monte Carlo samples needed to estimate rare events [5, 12, 31], it is also widely known that they rely heavily on model assumptions [10, 24]. That is, to guarantee the successful performance of these techniques, we typically need to analyze the underlying model dynamics carefully to design the Monte Carlo scheme. However, recent applications, such as autonomous vehicle safety evaluation [2, 23, 29, 44, 45] and robustness evaluation of machine learning predictors [7, 42, 43], lead to rare-event estimation problems with extremely sophisticated structures that hinder the design of efficiency-guaranteed variance reduction schemes. On the other hand, with the remarkable recent surge in computational infrastructure, in some situations we can afford to run a gigantic number of simulation trials.

Motivated by the limitations of the above techniques and the potential to generate numerous samples, in this paper we focus on a more basic setting than some of the above literature, but in a sense fundamental. More precisely, we focus on the situation where all we have to estimate p is a set of independent and identically distributed (i.i.d.) Bernoulli observations $I(A)$. A natural point estimate of p is the sample proportion $\hat{p}$, i.e. given a set of Bernoulli data $I_1, \ldots, I_n$ of size n, we output $\hat{p} = (1/n)\sum_{i=1}^{n} I_i$. We are interested in understanding the statistical error in using $\hat{p}$ in the situation where *p could be very small, importantly with no lower bound on how small it could be*. Unlike the estimates given by efficiency-guaranteed variance reduction techniques, as we will explain, it is not entirely straightforward whether using simple sample proportion can give meaningful guarantees to estimating rare-event probabilities, in relation to the sample size n and the (unknown) magnitude of p. Motivated by this, our main goal in this paper is to study the construction, coverage validity, and tightness of confidence intervals (CIs) for rare-event probabilities using only the simple sample proportion estimator. The main messages from our findings are as follows: The normality and Wilson intervals are not always valid, in the sense that their actual coverage probabilities can be less than the nominal confidence level, but they are shown to be close to our two newly developed valid intervals in terms of half-width. On the other hand, the exact interval is conservative, as its coverage probability is strictly larger than the nominal confidence level and hence it is not as tight as the aforementioned two intervals, but it safely guarantees the attainment of the nominal confidence level. Our new intervals are even more conservative than the exact interval and hence not recommended in practice, but they provide useful insights in understanding the tightness of the normality and Wilson intervals.

This paper is organized as follows. Section 2 describes the problem setting and the motivating challenges. Section 3 gives an overview of the existing and new CIs, and Section 4 summarizes our main results. Then, in Sections 5 and 6, we present the details of the derivation and analyses of these intervals. After that, Section 7 reports some numerical results to visualize our comparisons. Section 8 concludes the paper with our findings and recommendations. All missing proofs can be found in the appendix.

2. Problem setting and motivation

Suppose we would like to estimate a target probability p by using information from the Bernoulli data, or equivalently $\hat{p}$. In particular, we would like to construct a CI for p that has

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

justifiable statistical guarantees. In answering this, we would also quantify the error between the point estimate $\hat{p}$ and p.

First of all, we clarify what a good CI is supposed to be. To this end, we mainly consider the *validity* of the coverage and *tightness*. Throughout this paper, we say that $[\hat{p}_{\mathrm{l}}(\alpha), \hat{p}_{\mathrm{u}}(\alpha)]$ is a valid $(1-\alpha)$-level CI if $\mathbb{P}(\hat{p}_{\mathrm{l}}(\alpha) \le p \le \hat{p}_{\mathrm{u}}(\alpha)) \ge 1-\alpha$. This notion of validity can be defined similarly for one-sided confidence bounds. On the other hand, a good CI should not be too wide; for example, in the extreme case, the trivial CI [0, 1] is valid, but it does not provide any useful information. In this paper, we quantify tightness by the 'half-width', i.e. $\hat{p}_{\mathrm{u}}(\alpha) - \hat{p}$ or $\hat{p} - \hat{p}_{\mathrm{l}}(\alpha)$ (some intervals we consider are symmetric so there is no difference between the 'upper' and 'lower' half-widths, but some intervals are not, in which case the context would make the meaning of half-width clear). Importantly, considering that p is tiny in the rare-event settings, the CI is meaningful only if the half-width is small relative to p and $\hat{p}$.

To understand the challenges, we first examine the use of a standard 'textbook' CI, and we focus on the upper confidence bound for now since the lower confidence bound can be argued analogously. More specifically, we use the following as the $(1-\alpha)$-level upper confidence bound:

$$\hat{p}^{\mathrm{CLT}} = \hat{p} + z_{1-\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{1}$$

where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of a standard normal variable. The typical way to justify (1) is a normal approximation using the central limit theorem (CLT), which entails that

$$\mathbb{P}(p \le \hat{p}^{\mathrm{CLT}}) \approx \bar{\Phi}(-z_{1-\alpha}) = 1-\alpha,$$

where we denote by $\bar{\Phi}$ (and Φ) the tail (and cumulative) distribution function of the standard normal.

To delve a little further, note that the approximation error in the previous equation is controlled by the Berry–Esseen (BE) theorem. To simplify the discussion, suppose we are in a more idealized (but unrealistic) case where we know the precise value of the variance of the Bernoulli trial, i.e. $\sigma^2 = p(1-p)$, so that we use $\hat{p} + z_{1-\alpha}\sigma/\sqrt{n}$. Then the BE theorem stipulates that

$$|\mathbb{P}(p \le \hat{p}^{\mathrm{CLT}}) - \Phi(z_{1-\alpha})| \le \frac{C\rho}{\sigma^3\sqrt{n}}, \tag{2}$$

where $\rho = E|I_i - p|^3 = p(1-p)(1-2p+2p^2)$, and C is a universal constant (≈ 0.4748). Thus, the error in (2) is bounded by

$$\frac{Cp(1-p)(1-2p+2p^2)}{p^{3/2}(1-p)^{3/2}\sqrt{n}} \le \frac{C}{\sqrt{np(1-p)}}. \tag{3}$$

The issue is that when p is tiny, np can also be tiny unless n is sufficiently big, but a priori we would not know what n is 'sufficient'. If we have used the confidence bound given by (1) where the variance σ^2 is unknown and estimated by $\hat{p}(1-\hat{p})$, a similar BE bound would ultimately conclude the same issue as revealed by (3) [34]. A straightforward idea is to use the number of successes to infer whether n is sufficiently large. Suppose we have, say, 30 'success' outcomes among n trials; then we may think that $np \approx 30$, so that from the bound in (3) the error of $\hat{p}^{\mathrm{CLT}}$ appears controlled. As another, more extreme, case, suppose we only have only one success; then we may be led to believe that $np \approx 1$, so that $\hat{p}^{\mathrm{CLT}}$ is well defined but its coverage is likely way off from $1-\alpha$. However, we note that the guess that $np \approx 30$ or $np \approx 1$ is itself based on some central limit or concentration argument, which apparently leads to circular reasoning.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

This challenge motivates us to investigate more on the validity and tightness of different CIs in order to make a suitable choice.

It is well known that a quick and implementable approach to construct a CI that is always valid regardless of n, p, or $\hat{p}$ is to utilize the fact that $n\hat{p}$ follows a binomial distribution and extract a finite-sample confidence region using this exact distribution. This is often called the Clopper–Pearson CI or the exact method [13]. Though this is computationally easy, we are interested in simpler mathematical forms that allow us to analytically study the relative half-width as well. In this regard, Wilson's interval [1] has been studied and shown to give superior empirical performance, even in the case that p is tiny, but we are not aware of any rigorous proof on its validity. In this paper, we propose two different ways of constructing CIs for p that are simultaneously valid and analytically tractable, one using Chernoff's inequality, and the other one using the BE bound. Compared to the exact CI, these two CIs have explicit forms that allow us to investigate their half-widths, and thereby understand how far the CLT or Wilson CIs are from valid CIs.

Finally, in simulation analysis and some real-data situations, it is natural to keep sampling until we observe enough successes (e.g. when the number of successes is 30) in the experiments. We also adapt the existing or newly developed CIs to this setup and investigate their performance.

3. Overview of confidence intervals

Here we briefly introduce the formulas of the CIs that we study in this paper. We consider two settings. The first one is called the 'standard' setting, where the sample size n is fixed. The other setting is when we fix the number of successes $\hat{s} = n\hat{p}$, which we call the 'targeted stopping' setting. Under each setting, we discuss three existing CIs: the *CLT CI*, *Wilson CI*, and *exact CI*. We also introduce how to construct our new *Chernoff CI* and the *BE CI* via inverting Chernoff's inequality and the BE theorem.

3.1. Confidence intervals under the standard setting

Under the standard setting, to construct valid CIs our starting point is the following set:

$$\{0 < p < 1 \colon F(\hat{p}) \geq \alpha/2,\ F_{-}(\hat{p}) \leq 1 - \alpha/2\}, \tag{4}$$

where $F(x) = \mathbb{P}(\hat{p} \leq x)$ and $F_{-}(x) = \mathbb{P}(\hat{p} < x)$. Note that F and F_{-} depend on p. If F were continuous, we know that $\mathbb{P}(F(\hat{p}) \geq \alpha/2,\ F_{-}(\hat{p}) \leq 1 - \alpha/2) = 1 - \alpha$ since in this case $F(\hat{p}) = F_{-}(\hat{p}) \stackrel{\mathrm{d}}{=} \text{Uniform}[0, 1]$. Now we argue that $\mathbb{P}(F(\hat{p}) \geq \alpha/2,\ F_{-}(\hat{p}) \leq 1 - \alpha/2) > 1 - \alpha$ in this discrete case. Indeed, for any $\alpha \in (0, 1)$, there exist $0 \leq k, l \leq n$ such that $F((k-1)/n) < \alpha/2 \leq F(k/n)$ and $F_{-}(l/n) \leq 1 - \alpha/2 < F_{-}((l+1)/n)$. Then

$$\begin{aligned}
\mathbb{P}(F(\hat{p}) < \alpha/2 \text{ or } F_{-}(\hat{p}) > 1 - \alpha/2) &\leq \mathbb{P}(F(\hat{p}) < \alpha/2) + \mathbb{P}(F_{-}(\hat{p}) > 1 - \alpha/2) \\
&= \mathbb{P}(\hat{p} \leq (k-1)/n) + \mathbb{P}(\hat{p} \geq (l+1)/n) \\
&= F((k-1)/n) + 1 - F_{-}((l+1)/n) < \alpha.
\end{aligned}$$

Therefore, the set (4) is a valid $(1 - \alpha)$-level confidence region. From this derivation, we find that, due to the discreteness, the probability that this confidence region covers the true value p is strictly larger than the nominal confidence level $1 - \alpha$, and hence this confidence region is inevitably conservative.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

The CLT and Wilson CIs can be obtained from (4) by estimating $F(\hat{p})$ and $F_{-}(\hat{p})$ via normal approximation. As a result, these two CIs are no longer guaranteed to be valid. More specifically, using the fact that $(\hat{p}-p)/\sqrt{\hat{p}(1-\hat{p})/n} \approx N(0,1)$, we substitute

$$F(\hat{p}) \approx \Phi\left(\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}}\right), \qquad F_{-}(\hat{p}) \approx \Phi\left(\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}}\right)$$

into (4) to obtain the CLT CI.

Definition 1. (*CLT CI under the standard setting.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A, and $\hat{p}$ is the sample proportion of hitting A in n i.i.d. trials. Under this setting, the CLT CI is defined by:

$$\hat{p}_{\mathrm{u}}^{\mathrm{CLT}} = \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \qquad \hat{p}_{\mathrm{l}}^{\mathrm{CLT}} = \hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Similarly, since $(\hat{p}-p)/\sqrt{p(1-p)/n} \approx N(0,1)$ and substituting

$$F(\hat{p}) \approx \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right), \qquad F_{-}(\hat{p}) \approx \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right),$$

we get the Wilson CI.

Definition 2. (*Wilson CI under the standard setting.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A, and $\hat{p}$ is the sample proportion of hitting A in n i.i.d. trials. Under this setting, the Wilson CI is defined by:

$$\hat{p}_{\mathrm{u}}^{\mathrm{Wilson}} = \frac{1 + \left(2n\hat{p}/z_{1-\alpha/2}^2\right) + \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z_{1-\alpha/2}^2\right)}}{2\left(1 + \left(n/z_{1-\alpha/2}^2\right)\right)},$$

$$\hat{p}_{\mathrm{l}}^{\mathrm{Wilson}} = \frac{1 + \left(2n\hat{p}/z_{1-\alpha/2}^2\right) - \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z_{1-\alpha/2}^2\right)}}{2\left(1 + \left(n/z_{1-\alpha/2}^2\right)\right)}.$$

Instead of using normal approximation, the exact CI directly solves the valid confidence region (4). In fact, we know that $\hat{s} = n\hat{p} \sim \mathrm{Binomial}(n, p)$, so the functions $F(\cdot)$ and $F_{-}(\cdot)$ have exact expressions. More specifically, we have the following definition.

Definition 3. (*Exact CI under the standard setting.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A, and $\hat{p}$ is the sample proportion of hitting A in n i.i.d. trials. Under this setting, the exact CI is defined by $[\hat{p}_{\mathrm{l}}^{\mathrm{Exact}}, \hat{p}_{\mathrm{u}}^{\mathrm{Exact}}]$, where $\hat{p}_{\mathrm{u}}^{\mathrm{Exact}}$ and $\hat{p}_{\mathrm{l}}^{\mathrm{Exact}}$ are the respective solutions to

$$\sum_{k=0}^{\hat{s}} \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2, \qquad \sum_{k=\hat{s}}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2,$$

except that $\hat{p}_{\mathrm{u}}^{\mathrm{Exact}} = 1$ if $\hat{s} = n$ and $\hat{p}_{\mathrm{l}}^{\mathrm{Exact}} = 0$ if $\hat{s} = 0$.

When $0 < \hat{s} < n$, the bounds could be expressed explicitly via quantiles of the F distribution or Beta distribution, and hence are easy to compute numerically [1]. However, it is hard to

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

analyze the scale of this CI, which motivates us to further relax the confidence region (4) to get other valid CIs that are more conservative but easier to analyze.

In order to relax (4), we respectively consider using two methods: Chernoff's inequality and the BE theorem. We will only present the formulas for them here and leave the details of their development to Section 5.1.

Definition 4. (*Chernoff CI under the standard setting.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A, and $\hat{p}$ is the sample proportion of hitting A in n i.i.d. trials. Under this setting, the Chernoff CI is defined by:

$$\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}} = \hat{p} + \frac{\log(2/\alpha)}{n} + \sqrt{\frac{(\log(2/\alpha))^2}{n^2} + \frac{2\hat{p}\log(2/\alpha)}{n}},$$

$$\hat{p}_{\mathrm{l}}^{\mathrm{Chernoff}} = \hat{p} + \frac{\log(2/\alpha)}{2n} - \sqrt{\frac{(\log(2/\alpha))^2}{4n^2} + \frac{2\hat{p}\log(2/\alpha)}{n}}.$$

Definition 5. (*BE CI under the standard setting.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A, and $\hat{p}$ is the sample proportion of hitting A in n i.i.d. trials. In addition, we assume that $p < \frac{1}{2}$. Under this setting, the BE CI is solved from

$$\left\{0 < p \le \hat{p} \wedge \frac{1}{2} : \Phi\left(\frac{p-\hat{p}}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \frac{\alpha}{2}\right\}$$
$$\cup \left\{\hat{p} \le p < \frac{1}{2} : \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \frac{\alpha}{2}\right\}.$$

3.2. Confidence intervals under targeted stopping

Now we consider experiments where we keep sampling until we get n_0 successes. Under this setting, the sample size N is a random variable. More specifically, $N = N_1 + \cdots + N_{n_0}$, where $N_1, \ldots, N_{n_0}$ are i.i.d. Geometric(p) random variables, or, equivalently, $N - n_0$ follows a negative binomial distribution NB(n_0, p). Note that $N \ge n_0$. We define $F_N(x) = \mathbb{P}(N \le x)$ and $F_{N-}(x) = \mathbb{P}(N < x)$. Similar to Section 3.1, we argue that the following set is a valid $(1-\alpha)$-level confidence region for p:

$$\{0 < p < 1 : F_N(N) \ge \alpha/2,\ F_{N-}(N) \le 1 - \alpha/2\}. \tag{5}$$

Indeed, for any $\alpha \in (0, 1)$, there exist $1 \le k, l < \infty$ such that $F_N(k-1) < \alpha/2 \le F_N(k)$ and $F_{N-}(l) \le 1 - \alpha/2 < F_{N-}(l+1)$. Then

$$\begin{aligned}\mathbb{P}(F_N(N) < \alpha/2 \text{ or } F_{N-}(N) > 1-\alpha/2) &\le \mathbb{P}(F_N(N) < \alpha/2) + \mathbb{P}(F_{N-}(N) > 1-\alpha/2)\\ &= \mathbb{P}(N \le k-1) + \mathbb{P}(N \ge l+1)\\ &= F_N(k-1) + 1 - F_{N-}(l+1) < \alpha.\end{aligned}$$

By definition, the set (5) is a valid $(1-\alpha)$-level confidence region.

We could still use the CLT and Wilson CIs with $\hat{p} = n_0/N$. More specifically, we have the following definitions.

Definition 6. (*CLT CI under targeted stopping.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A. We keep sampling until we get n_0 successes and the sample size is denoted

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

by N. Under this setting, the CLT CI is defined by:

$$\hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0} = \frac{n_0}{N} + z_{1-\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}}, \qquad \hat{p}^{\mathrm{CLT}}_{\mathrm{l},n_0} = \frac{n_0}{N} - z_{1-\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}}.$$

Definition 7. (*Wilson CI under targeted stopping.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A. We keep sampling until we get n_0 successes and the sample size is denoted by N. Under this setting, the Wilson CI is defined by:

$$\hat{p}^{\mathrm{Wilson}}_{\mathrm{u},n_0} = \frac{1 + \left(2n_0/z^2_{1-\alpha/2}\right) + \sqrt{1 + \left(4n_0(N-n_0)/z^2_{1-\alpha/2}N\right)}}{2\left(1 + \left(N/z^2_{1-\alpha/2}\right)\right)},$$

$$\hat{p}^{\mathrm{Wilson}}_{\mathrm{l},n_0} = \frac{1 + \left(2n_0/z^2_{1-\alpha/2}\right) - \sqrt{1 + \left(4n_0(N-n_0)/z^2_{1-\alpha/2}N\right)}}{2\left(1 + \left(N/z^2_{1-\alpha/2}\right)\right)}.$$

Similar to the standard setting, we can directly solve (5) using the exact distribution of N.

Definition 8. (*Exact CI under targeted stopping.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A. We keep sampling until we get n_0 successes and the sample size is denoted by N. Under this setting, the exact CI is defined by $\left[\hat{p}^{\mathrm{Exact}}_{\mathrm{l},n_0}, \hat{p}^{\mathrm{Exact}}_{\mathrm{u},n_0}\right]$ where $\hat{p}^{\mathrm{Exact}}_{\mathrm{u},n_0}$ and $\hat{p}^{\mathrm{Exact}}_{\mathrm{l},n_0}$ are the respective solutions to

$$\sum_{k=0}^{N-n_0-1}\binom{k+n_0-1}{n_0-1}(1-p)^k p^{n_0} = 1-\alpha/2, \qquad \sum_{k=0}^{N-n_0}\binom{k+n_0-1}{n_0-1}(1-p)^k p^{n_0} = \alpha/2,$$

except that $\hat{p}^{\mathrm{Exact}}_{\mathrm{u},n_0} = 1$ if $N = n_0$.

While the interval is easy to compute numerically, it is not easy to analyze. Similar to the standard setting, we relax the confidence region (5) to construct valid CIs via respectively inverting Chernoff's inequality and the BE theorem. We leave the details of developing these two new CIs to Section 6.1 and only present the formulas here.

Definition 9. (*Chernoff CI under targeted stopping.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A. We keep sampling until we get n_0 successes and the sample size is denoted by N. Under this setting, the Chernoff CI is solved from

$$\left\{0 < p < 1 \colon p^{n_0}(1-p)^{N-n_0} \geq \frac{\alpha}{2}\left(\frac{n_0}{N}\right)^{n_0}\left(1-\frac{n_0}{N}\right)^{N-n_0}\right\}.$$

Definition 10. (*BE CI under targeted stopping.*) Suppose that we estimate the probability $p = \mathbb{P}(A)$ for the event A. We keep sampling until we get n_0 successes and the sample size is denoted by N. In addition, we assume that $p < \frac{1}{2}$. Under this setting, the BE CI is defined by:

$$\left\{0 < p \leq \frac{n_0}{N} \wedge \frac{1}{2} \colon \Phi\left(\frac{Np - n_0}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}\right\}$$
$$\cup \left\{\frac{n_0}{N} \leq p < \frac{1}{2} \colon \Phi\left(\frac{n_0 - Np}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}\right\},$$

where $C' = 16C$ is a universal constant.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

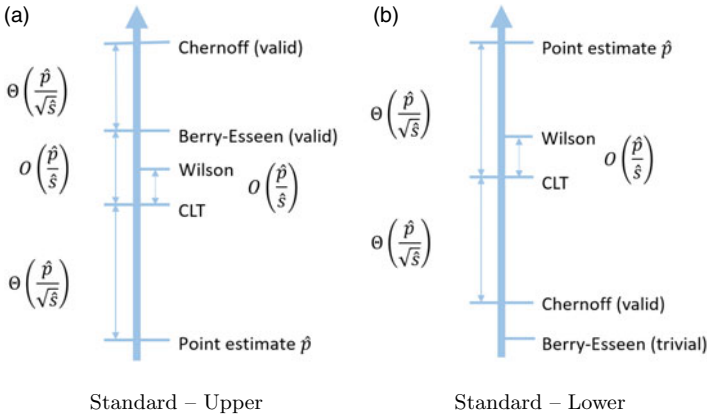

FIGURE 1. Comparisons of the positions of confidence upper and lower bounds under the standard setting. Here, 'valid' means that the CI has valid coverage in the sense that the actual coverage probability always reaches the nominal confidence level.

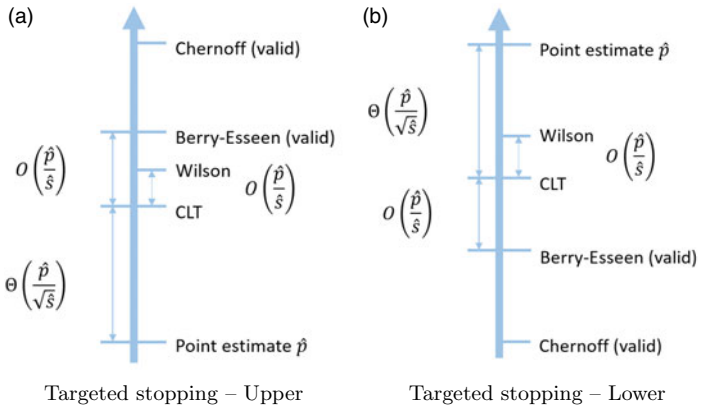

FIGURE 2. Comparisons of the positions of confidence upper and lower bounds under the targeted stopping setting. Here, 'valid' means that the CI has valid coverage in the sense that the actual coverage probability always reaches the nominal confidence level.

4. Summary of the main results

As explained in Section 2, when we compare different CIs we mainly consider the validity in terms of coverage probability, and tightness in terms of half-width. In terms of validity, the existing CLT and Wilson CIs do not possess guarantees, while the exact CI, and our new Chernoff CI and BE CI, are valid by construction. The half-widths of the CIs, which will be analyzed in detail in Sections 5.2 and 6.2, are summarized in Figures 1 and 2. In particular, these figures illustrate comparisons of these CIs in terms of upper and lower bound (the exact CI is not included since it is hard to analyze its magnitude). For instance, from (1), we clearly see that the half-width of the CLT CI scales at the same order as $\sqrt{\hat{p}/n} = \hat{p}/\sqrt{\hat{s}}$, where $\hat{s} = n\hat{p}$ is the number of positive outcomes. By expressing $\sqrt{\hat{p}/n}$ as $\hat{p}/\sqrt{\hat{s}}$ here, it is easier to see how

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

the half-width scales relative to $\hat{p}$. That is, the relative half-width is of order $1/\sqrt{\hat{s}}$. Note that in these figures, under the standard setting, for $f(n, \hat{p})$, $g(n, \hat{p}) \geq 0$, we write $f = O(g)$ if there exist $N_0, p_0, M > 0$, which do not depend on p or $\hat{p}$, such that, for any $n\hat{p} > N_0$ and $\hat{p} < p_0$, $f \leq Mg$; we write $f = \Theta(g)$ if there exist $N_0, p_0, M_1, M_2 > 0$, which do not depend on p or $\hat{p}$, such that, for any $n\hat{p} > N_0$ and $\hat{p} < p_0$, $M_1 g \leq f \leq M_2 g$. Under the targeted stopping setting, $O(\cdot)$ and $\Theta(\cdot)$ are defined similarly by replacing $n\hat{p}$ with n_0. The notations $O(\cdot)$ and $\Theta(\cdot)$ will be used throughout the rest of this paper.

More concretely, Tables 1 and 2 summarize the formulas, scales, pros, and cons of each CI under the standard and targeted stopping settings respectively. The key findings can be summarized as follows:

- The CLT CI is the 'textbook' normality interval and thus very intuitive, but its coverage probability can be far below the nominal level. However, in terms of the half-width, except for the lower bound in the standard setting, the difference between the CLT bound and the valid BE bound is of order $\hat{p}/\hat{s}$, so the relative difference with respect to $\hat{p}$ is of order $1/\hat{s}$, which is of higher order in $\hat{s}$ than its relative half-width. This can be viewed as a relatively small price of validity paid to make the CLT bound correct. For the lower bound in the standard setting, the BE bound is trivial, so we cannot come to a similar conclusion. However, in this case the difference between the CLT bound and the valid Chernoff bound is of order $\hat{p}/\sqrt{\hat{s}}$, the same order as the half-width, which shows that the CLT bound has roughly the correct magnitude.
- In practice, the Wilson CI has satisfactory performance, in the sense that it is relatively tight while the coverage probability is usually close to the nominal confidence level. The difference between the Wilson and CLT bounds is of order $\hat{p}/\hat{s}$, which is of higher order in $\hat{s}$ than the half-width. As a result, the conclusions for the CLT CI regarding the difference from the valid BE and Chernoff CIs still hold for the Wilson CI.
- The exact CI is, as aforementioned, inevitably conservative, in the sense that its coverage probability is strictly higher than the nominal level. However, it is the tightest among the valid CIs, so it is recommended when we want the nominal confidence level to be guaranteed. The Chernoff and BE CIs are valid but extremely conservative. They are not recommended for use in practice, but their analytical forms help us gain useful insights on the CLT and Wilson CIs. That is, now we learn that the CLT and Wilson CIs, although not always valid, are relatively close to these two valid CIs as mentioned in the first bullet point.

5. Developments under the standard setting

We present in detail the construction of the new Chernoff and BE CIs that endows their validity (Section 5.1). Then we analyze the half-widths of all the CIs discussed here (Section 5.2).

5.1. Derivation of new confidence intervals

5.1.1. *Chernoff's CI.* Now we present our first approach to construct a valid CI for p by relaxing (4). By Chernoff's inequality, we have

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

TABLE 1. Summary of the CIs in the standard setting ($I_1, \ldots, I_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$, $\hat{p} = (1/n)\sum_{i=1}^{n} I_i$, $\hat{s} = n\hat{p}$).

CLT	$\hat{p} \pm z_{1-\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
Scale	$\hat{p}_{\text{u}}^{\text{CLT}} - \hat{p} = \hat{p} - \hat{p}_{\text{l}}^{\text{CLT}} = \Theta(\hat{p}/\sqrt{\hat{s}})$
Pros	Follows from the intuitive 'textbook' formula
Cons	Not always valid, especially when np is not sufficiently large
Wilson	$\dfrac{1 + \left(2n\hat{p}/z_{1-\alpha/2}^2\right) \pm \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z_{1-\alpha/2}^2\right)}}{2\left(1 + \left(n/z_{1-\alpha/2}^2\right)\right)}$
Scale	$\left\lvert\hat{p}_{\text{u}}^{\text{Wilson}} - \hat{p}_{\text{u}}^{\text{CLT}}\right\rvert = O(\hat{p}/\hat{s})$, $\left\lvert\hat{p}_{\text{l}}^{\text{Wilson}} - \hat{p}_{\text{l}}^{\text{CLT}}\right\rvert = O(\hat{p}/\hat{s})$
Pros	Tight; the coverage probability is usually close to $1-\alpha$
Cons	Not always valid; lacks theoretical error control
Exact	Solutions to $\sum_{k=0}^{\hat{s}} \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2$, $\sum_{k=\hat{s}}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \alpha/2$ except that $\hat{p}_{\text{u}}^{\text{Exact}} = 1$ if $\hat{s} = n$ and $\hat{p}_{\text{l}}^{\text{Exact}} = 0$ if $\hat{s} = 0$
Pros	Always valid; tighter than other valid CIs
Cons	Conservative; hard to analyze
Chernoff	$\hat{p} + \dfrac{\log(2/\alpha)}{n} + \sqrt{\dfrac{(\log(2/\alpha))^2}{n^2} + \dfrac{2\hat{p}\log(2/\alpha)}{n}}$, $\hat{p} + \dfrac{\log(2/\alpha)}{2n} - \sqrt{\dfrac{(\log(2/\alpha))^2}{4n^2} + \dfrac{2\hat{p}\log(2/\alpha)}{n}}$
Scale	$\hat{p}_{\text{u}}^{\text{Chernoff}} - \hat{p}_{\text{u}}^{\text{CLT}} = \Theta(\hat{p}/\sqrt{\hat{s}})$, $\hat{p}_{\text{l}}^{\text{CLT}} - \hat{p}_{\text{l}}^{\text{Chernoff}} = \Theta(\hat{p}/\sqrt{\hat{s}})$
Pros	Always valid; helps us understand the relative error of $\hat{p}$
Cons	Extremely conservative
BE	$\left\{0 < p \le \hat{p} \wedge \dfrac{1}{2} : \Phi\left(\dfrac{p-\hat{p}}{\sqrt{p(1-p)/n}}\right) + \dfrac{C}{\sqrt{np(1-p)}} \ge \dfrac{\alpha}{2}\right\} \cup \left\{\hat{p} \le p < \dfrac{1}{2} : \Phi\left(\dfrac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right) + \dfrac{C}{\sqrt{np(1-p)}} \ge \dfrac{\alpha}{2}\right\}$ where C is the universal constant in the BE theorem
Scale	$\left\lvert\hat{p}_{\text{u}}^{\text{BE}} - \hat{p}_{\text{u}}^{\text{CLT}}\right\rvert = O(\hat{p}/\hat{s})$
Pros	Always valid; helps us understand the error of the CLT upper bound
Cons	Extremely conservative; trivial lower bound

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

TABLE 2. Summary of the CIs in the targeted stopping setting ($N_1, \dots, N_{n_0} \overset{\text{i.i.d.}}{\sim} \text{Geometric}(p)$, $N = \sum_{i=1}^{n_0} N_i$, $\hat{p} = n_0/N$).

CLT	$\frac{n_0}{N} \pm z_{1-\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}}$
Scale	$\hat{p}^{\text{CLT}}_{\text{u},n_0} - \hat{p} = \hat{p} - \hat{p}^{\text{CLT}}_{\text{l},n_0} = \Theta(\sqrt{n_0}/N)$
Pros	Follows from the intuitive 'textbook' formula
Cons	Not always valid
Wilson	$\frac{1 + \left(2n_0/z^2_{1-\alpha/2}\right) \pm \sqrt{1 + \left(4n_0(N-n_0)/z^2_{1-\alpha/2}N\right)}}{2\left(1 + \left(N/z^2_{1-\alpha/2}\right)\right)}$
Scale	$\left\lvert\hat{p}^{\text{Wilson}}_{\text{u},n_0} - \hat{p}^{\text{CLT}}_{\text{u},n_0}\right\rvert = O(1/N)$, $\left\lvert\hat{p}^{\text{Wilson}}_{\text{l},n_0} - \hat{p}^{\text{CLT}}_{\text{l},n_0}\right\rvert = O(1/N)$
Pros	Tight; the coverage probability is usually close to $1-\alpha$
Cons	Not always valid
Exact	Solutions to $\sum_{k=0}^{N-n_0-1} \binom{k+n_0-1}{n_0-1}(1-p)^k p^{n_0} = 1-\alpha/2$, $\sum_{k=0}^{N-n_0} \binom{k+n_0-1}{n_0-1}(1-p)^k p^{n_0} = \alpha/2$ except that $\hat{p}^{\text{Exact}}_{\text{u},n_0} = 1$ if $N = n_0$
Pros	Always valid; tighter than other valid CIs
Cons	Conservative; hard to analyze
Chernoff	$\left\{0 < p < 1 : p^{n_0}(1-p)^{N-n_0} \geq \frac{\alpha}{2}\left(\frac{n_0}{N}\right)^{n_0}\left(1 - \frac{n_0}{N}\right)^{N-n_0}\right\}$
Pros	Always valid
Cons	Extremely conservative; hard to analyze
BE	$\left\{0 < p \leq \frac{n_0}{N} \wedge \frac{1}{2} : \Phi\left(\frac{Np - n_0}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}\right\}$ $\cup \left\{\frac{n_0}{N} \leq p < \frac{1}{2} : \Phi\left(\frac{n_0 - Np}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}\right\}$ where $C' = 16C$ is a universal constant
Scale	$\hat{p}^{\text{BE}}_{\text{u},n_0} - \hat{p}^{\text{CLT}}_{\text{u},n_0} = O(1/N)$, $\hat{p}^{\text{CLT}}_{\text{l},n_0} - \hat{p}^{\text{BE}}_{\text{l},n_0} = O(1/N)$
Pros	Aways valid; helps us understand the error of $\hat{p}$ and the CLT CI
Cons	Extremely conservative; trivial for small n_0

$$\mathbb{P}(\hat{p} \leq (1-\delta)p) \leq \exp\left(-\frac{\delta^2}{2}np\right), \qquad 0 < \delta < 1,$$

$$\mathbb{P}(\hat{p} \geq (1+\delta)p) \leq \exp\left(-\frac{\delta^2}{2+\delta}np\right), \qquad \delta > 0.$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

Replacing $(1-\delta)p$ or $(1+\delta)p$ by x, we have

$$F(x) \le \exp\left\{-\left(1-\frac{x}{p}\right)^2 \frac{np}{2}\right\}, \qquad x \le p,$$

$$F_-(x) \ge 1-\exp\left\{-\frac{((x/p)-1)^2}{1+(x/p)} np\right\}, \qquad x \ge p.$$

Hence, $F(\hat{p}) \ge \alpha/2$, $F_-(\hat{p}) \le 1-\alpha/2$ implies that

$$p \ge \hat{p} \text{ and } \exp\left\{-\left(1-\frac{\hat{p}}{p}\right)^2 \frac{np}{2}\right\} \ge \frac{\alpha}{2} \quad \text{or} \quad p \le \hat{p} \text{ and } 1-\exp\left\{-\frac{((\hat{p}/p)-1)^2}{1+(\hat{p}/p)} np\right\} \le 1-\frac{\alpha}{2}.$$

Therefore,

$$\left\{0 < p \le \hat{p} \colon \exp\left\{-\frac{((\hat{p}/p)-1)^2}{1+(\hat{p}/p)} np\right\} \ge \frac{\alpha}{2}\right\} \cup \left\{\hat{p} \le p < 1 \colon \exp\left\{-\left(1-\frac{\hat{p}}{p}\right)^2 \frac{np}{2}\right\} \ge \frac{\alpha}{2}\right\}$$

is a confidence region for $\hat{p}$ with confidence level at least $1-\alpha$. Simplifying the above expression, we have

$$0 < p \le \hat{p}, \ \exp\left\{-\frac{((\hat{p}/p)-1)^2}{1+(\hat{p}/p)} np\right\} \ge \frac{\alpha}{2}$$

$$\Longrightarrow \hat{p} + \frac{\log(2/\alpha)}{2n} - \sqrt{\frac{(\log(2/\alpha))^2}{4n^2} + \frac{2\hat{p}\log(2/\alpha)}{n}} \le p \le \hat{p},$$

$$\hat{p} \le p < 1, \ \exp\left\{-\left(1-\frac{\hat{p}}{p}\right)^2 \frac{np}{2}\right\} \ge \frac{\alpha}{2}$$

$$\Longrightarrow \hat{p} \le p \le \hat{p} + \frac{\log(2/\alpha)}{n} + \sqrt{\frac{(\log(2/\alpha))^2}{n^2} + \frac{2\hat{p}\log(2/\alpha)}{n}}.$$

Hence, by taking the union, we get a valid $(1-\alpha)$-level CI for p, for any finite sample n. This id summarized in the following theorem.

Theorem 1. (Validity of Chernoff CI under the standard setting.) *The interval given by*

$$\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}} = \hat{p} + \frac{\log(2/\alpha)}{n} + \sqrt{\frac{(\log(2/\alpha))^2}{n^2} + \frac{2\hat{p}\log(2/\alpha)}{n}},$$

$$\hat{p}_{\mathrm{l}}^{\mathrm{Chernoff}} = \hat{p} + \frac{\log(2/\alpha)}{2n} - \sqrt{\frac{(\log(2/\alpha))^2}{4n^2} + \frac{2\hat{p}\log(2/\alpha)}{n}}$$

is a valid $(1-\alpha)$*-level CI for p, for any finite sample n. That is, for any n,*

$$\mathbb{P}\big(\hat{p}_{\mathrm{l}}^{\mathrm{Chernoff}} \le p \le \hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}}\big) \ge 1-\alpha.$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

5.1.2. *BE CI.* We develop another CI for p by inverting the BE theorem. Here, we assume that p is known to satisfy $p < \frac{1}{2}$ a priori (which is reasonable if we consider rare events). In this paper, we use the standard version of the BE theorem, and a potential future investigation is to consider a BE bound for the studentized statistic [40, 41].

By the BE theorem,

$$\left|\mathbb{P}\left((\hat{p}-p)\sqrt{\frac{n}{p(1-p)}} \le x\right) - \Phi(x)\right| \le \frac{C}{\sqrt{np(1-p)}},$$
$$\left|\mathbb{P}\left((p-\hat{p})\sqrt{\frac{n}{p(1-p)}} \le x\right) - \Phi(x)\right| \le \frac{C}{\sqrt{np(1-p)}},$$

where C is a universal constant. We replace x by $(\hat{p}-p)/\sqrt{p(1-p)/n}$ in the first inequality and $(p-\hat{p})/\sqrt{p(1-p)/n}$ in the second one. Then,

$$\left|F(\hat{p}) - \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right)\right| \le \frac{C}{\sqrt{np(1-p)}},$$
$$\left|1 - F_{-}(\hat{p}) - \Phi\left(\frac{p-\hat{p}}{\sqrt{p(1-p)/n}}\right)\right| \le \frac{C}{\sqrt{np(1-p)}}.$$

Hence, $F(\hat{p}) \ge \alpha/2$, $F_{-}(\hat{p}) \le 1-\alpha/2$ implies that either

$$p \ge \hat{p} \text{ and } \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \alpha/2 \qquad \text{or}$$
$$p \le \hat{p} \text{ and } \Phi\left(\frac{p-\hat{p}}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \alpha/2.$$

Thus,

$$\left\{0 < p \le \hat{p} \colon \Phi\left(\frac{p-\hat{p}}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \frac{\alpha}{2}\right\}$$
$$\cup \left\{\hat{p} \le p < 1 \colon \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \frac{\alpha}{2}\right\}$$

is a valid $(1-\alpha)$-level confidence region for p. Since we have assumed that $p < \frac{1}{2}$, the above confidence region can be shrunk further. To summarize, we have the following theorem.

Theorem 2. (Validity of BE CI under the standard setting.) *Assume that $p < \frac{1}{2}$. Then the set*

$$\left\{0 < p \le \hat{p} \wedge \frac{1}{2} \colon \Phi\left(\frac{p-\hat{p}}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \frac{\alpha}{2}\right\}$$
$$\cup \left\{\hat{p} \le p < \frac{1}{2} \colon \Phi\left(\frac{\hat{p}-p}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \ge \frac{\alpha}{2}\right\} \quad (6)$$

is a valid $(1-\alpha)$-level confidence region for p, for any finite sample n.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

5.2. Analyses of half-widths

5.2.1. *CLT CI.* Clearly, $\hat{p}_{\mathrm{u}}^{\mathrm{CLT}} - \hat{p} = \Theta(\sqrt{\hat{p}/n}) = \Theta(\hat{p}/\sqrt{\hat{s}})$ and $\hat{p} - \hat{p}_{\mathrm{l}}^{\mathrm{CLT}} = \Theta(\hat{p}/\sqrt{\hat{s}})$. As explained in Section 2, we express the half-width as $\Theta(\hat{p}/\sqrt{\hat{s}})$ instead of $\Theta(\sqrt{\hat{p}/n})$ in order to understand the magnitude of the relative half-width with respect to $\hat{p}$ more clearly.

5.2.2. *Wilson CI.* We derive the following theorem through some algebraic manipulations.

Theorem 3. (Half-width of the Wilson CI under the standard setting.)

$$\left|\hat{p}_{\mathrm{u}}^{\mathrm{Wilson}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}}\right| \le \frac{z_{1-\alpha/2}^2}{n} + \frac{z_{1-\alpha/2}^3}{2n^{3/2}}, \qquad \left|\hat{p}_{\mathrm{l}}^{\mathrm{Wilson}} - \hat{p}_{\mathrm{l}}^{\mathrm{CLT}}\right| \le \frac{z_{1-\alpha/2}^2}{n} + \frac{z_{1-\alpha/2}^3}{2n^{3/2}}.$$

Note that $1/n = \hat{p}/\hat{s}$, so the difference between the Wilson and CLT CIs is of order $O(\hat{p}/\hat{s})$, which is of higher order than $\hat{p}/\sqrt{\hat{s}}$ in $\hat{s}$. In fact, as long as $\hat{s} \ge 1$, i.e. we have at least one positive observation, then $\hat{p}/\sqrt{\hat{s}} = \sqrt{\hat{p}/n} = \sqrt{\hat{s}}/n \ge 1/n$. Since the half-width of the CLT CI is of order $\hat{p}/\sqrt{\hat{s}}$, we get that the half-width of the Wilson CI is close to the CLT CI.

5.2.3. *Chernoff CI.* When $\hat{p} = 0$, the Chernoff CI reduces to $[0, 2\log(1/\alpha)/n]$ (and in fact we can construct even tighter bounds by using the binomial distribution of $n\hat{p}$ directly in this case). On the other hand, when $\hat{p} > 0$, we can re-express using $\hat{s} = n\hat{p}$ to get

$$\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}} = \hat{p}\left(1 + \frac{\log(2/\alpha)}{\hat{s}} + \sqrt{\frac{(\log(2/\alpha))^2}{\hat{s}^2} + \frac{2\log(2/\alpha)}{\hat{s}}}\right),$$

$$\hat{p}_{\mathrm{l}}^{\mathrm{Chernoff}} = \hat{p}\left(1 + \frac{\log(2/\alpha)}{2\hat{s}} - \sqrt{\frac{(\log(2/\alpha))^2}{4\hat{s}^2} + \frac{2\log(2/\alpha)}{\hat{s}}}\right).$$

We highlight that in this case, the half-width of the Chernoff CI is of order $\Theta(\hat{p}/\sqrt{\hat{s}})$, which scales in the same order as the CLT CI. If we check the difference between this interval and the CLT interval, we find that it is of the same order as the half-width of the CLT CI. The following theorem presents the details of this claim. We will shortly contrast this result with another one presented.

Theorem 4. (Half-width of the Chernoff CI under the standard setting.)

$$\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}} \ge (\sqrt{2\log(2/\alpha)} - z_{1-\alpha/2})\sqrt{\frac{\hat{p}}{n}} + \frac{\log(2/\alpha)}{n},$$

$$\hat{p}_{\mathrm{l}}^{\mathrm{CLT}} - \hat{p}_{\mathrm{l}}^{\mathrm{Chernoff}} \ge (\sqrt{2\log(2/\alpha)} - z_{1-\alpha/2})\sqrt{\frac{\hat{p}}{n}} - \frac{\log(2/\alpha)}{2n}.$$

Note that $\sqrt{2\log(2/\alpha)} - z_{1-\alpha/2} > 0$ *for* $0 < \alpha < 1$.

We recall that $1/n = \hat{p}/\hat{s}$ is of higher order than $\sqrt{\hat{p}/\hat{s}} = \hat{p}/\sqrt{\hat{s}}$. Provided that $\sqrt{2\log(2/\alpha)} - z_{1-\alpha/2} > 0$, $\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}}$ (or $\hat{p}_{\mathrm{l}}^{\mathrm{Chernoff}} - \hat{p}_{\mathrm{l}}^{\mathrm{CLT}}$) is of no higher order than $\hat{p}_{\mathrm{u}}^{\mathrm{CLT}} - \hat{p}$ (or $\hat{p}_{\mathrm{l}}^{\mathrm{CLT}} - \hat{p}$).

5.2.4. *BE CI.* We focus on the confidence upper bound as, unfortunately, we cannot derive a non-trivial confidence lower bound from (6) since any $0 < p < \frac{1}{2}$ such that $C/\sqrt{np(1-p)} \ge \alpha/2$ is contained in this confidence region. Now we further relax (6) to develop a more explicit

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

upper bound. In particular, (6) could be relaxed to

$$\left\{0 < p < \frac{1}{2} : \Phi\left(\frac{\hat{p} - p}{\sqrt{p(1-p)/n}}\right) + \frac{C}{\sqrt{np(1-p)}} \geq \frac{\alpha}{2}\right\}.$$

In fact, for any $0 \leq \lambda \leq 1 - (4C/\sqrt{n}\alpha)$,

$$0 < p < \frac{1}{2},\ \frac{C}{\sqrt{np(1-p)}} \geq (1-\lambda)\frac{\alpha}{2} \quad \Longrightarrow \quad 0 < p \leq \frac{1 - \sqrt{1 - (16C^2/n(1-\lambda)^2\alpha^2)}}{2},$$

$$0 < p < \frac{1}{2},\ \Phi\left(\frac{\hat{p} - p}{\sqrt{p(1-p)/n}}\right) \geq \lambda\frac{\alpha}{2} \quad \Longrightarrow$$

$$0 < p \leq \frac{1 + \left(2n\hat{p}/z^2_{\lambda\alpha/2}\right) + \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)}}{2\left(1 + \left(n/z^2_{\lambda\alpha/2}\right)\right)}.$$

Therefore,

$$0 < p \leq \left(\frac{1 - \sqrt{1 - (16C^2/n(1-\lambda)^2\alpha^2)}}{2}\right) \vee \left(\frac{1 + \left(2n\hat{p}/z^2_{\lambda\alpha/2}\right) + \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)}}{2\left(1 + \left(n/z^2_{\lambda\alpha/2}\right)\right)}\right)$$

is a $(1-\alpha)$-level CI. For simplicity, we denote the two parts as U_1 and U_2. Note that λ is not necessarily deterministic. Instead, it can be dependent on the data as long as it stays within the interval $[0, 1 - (4C/\sqrt{n}\alpha)]$. In fact, we may choose λ carefully such that $U_1 \leq U_2$ is guaranteed for sufficiently large n. Specifically, the following theorem proposes another valid CI.

Theorem 5. (Relaxed BE CI under the standard setting.) *Assume that* $p < \frac{1}{2}$. *Let* $\lambda = 1 - (2\tilde{C}/\sqrt{n}\alpha)$, *where*

$$\tilde{C} = \left(\frac{C}{\sqrt{\hat{p}(1-\hat{p})}}\right) \wedge \left(\frac{u\sqrt{n}\alpha}{2}\right).$$

Here, $u < 1$ *is any constant such that* $4C^2/u^2\alpha^2 < z^2_{(1-u)\alpha/2}$. *In the case that* $\hat{p} = 0$ *or 1, naturally we set* $\tilde{C} = u\sqrt{n}\alpha/2$. *Then there exists* N_0, *which does not depend on p and* $\hat{p}$, *such that, for any* $n > N_0$,

$$\hat{p}^{\mathrm{BE}}_{\mathrm{u}} = \frac{1 + (2n\hat{p}/z^2_{\lambda\alpha/2}) + \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)}}{2\left(1 + \left(n/z^2_{\lambda\alpha/2}\right)\right)}, \qquad \hat{p}^{\mathrm{BE}}_{\mathrm{l}} = 0$$

is a valid $(1-\alpha)$*-level CI for p. In particular,* N_0 *can be chosen as*

$$\left(\frac{4C}{u\alpha}\right)^2 \vee \frac{12z^2_{(1-u)\alpha/2}C^2}{z^2_{(1-u)\alpha/2}u^2\alpha^2 - 4C^2}.$$

Actually, $\hat{p}^{\mathrm{BE}}_{\mathrm{u}}$ itself is a valid $(1-\alpha/2)$-level confidence upper bound for p, which is higher than the nominal level $1-\alpha$. The series of relaxations makes this CI more and more conservative, but we will show that the upper bound still has a similar scale to $\hat{p}^{\mathrm{CLT}}_{\mathrm{u}}$ and $\hat{p}^{\mathrm{Wilson}}_{\mathrm{u}}$.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

Namely, we can derive that $|\hat{p}_{\mathrm{u}}^{\mathrm{BE}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}}|$ is bounded by order $1/n$. In other words, though $\hat{p}_{\mathrm{u}}^{\mathrm{CLT}}$ has undesirable coverage probability in the rare-event setting, it is not 'too far' from a valid upper bound. The following theorem states this result.

Theorem 6. (Half-width of the BE CI under the standard setting.) *Assume that $p < \frac{1}{2}$, abd $\hat{p}_{\mathrm{u}}^{\mathrm{BE}}$ is as defined in Theorem 5. Then there is a constant C_0, which does not depend on p and $\hat{p}$, such that $\left|\hat{p}_{\mathrm{u}}^{\mathrm{BE}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}}\right| \le C_0/n$.*

Note that the bound in Theorem 6 can be rephrased as $\left|\hat{p}_{\mathrm{u}}^{\mathrm{BE}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}}\right| \le C_0\hat{p}/\hat{s}$. In other words, $\hat{p}_{\mathrm{u}}^{\mathrm{BE}}$ differs from $\hat{p}_{\mathrm{u}}^{\mathrm{CLT}}$ by a higher order than the half-width of the CLT CI in terms of $\hat{s}$, while all quantities scale with $\hat{p}$ in a similar manner. Compared to Theorem 4, we see in Theorem 6 that $\hat{p}_{\mathrm{u}}^{\mathrm{BE}}$ is substantially tighter than $\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}}$ when $\hat{s}$ increases, although due to the implicit constant C_0 it may not be the case for small $\hat{s}$.

6. Developments under targeted stopping

We now present our results for the targeted stopping setting, following the roadmap for the standard setting presented earlier. Namely, we first present the construction of the Chernoff and BE CIs (Section 6.1), followed by analyses of half-widths for all CIs (Section 6.2).

6.1. Derivation of new confidence intervals

To construct the new CIs under the targeted stopping setting, we again relax the confidence region (5) via Chernoff's inequality and the BE theorem. Nevertheless, now we need to deal with the distribution of N instead of $\hat{p}$ as in (4). Hence, as we show below, the specific derivations of applying Chernoff's inequality and the BE theorem differ from the standard setting.

6.1.1. *Chernoff CI.* First, we propose a Chernoff CI similar to the one in the standard setting. By Markov's inequality,

$$\mathbb{P}(N \ge x) \le \mathrm{e}^{-tx}\mathbb{E}(\mathrm{e}^{tN}) = \mathrm{e}^{-tx}\left(\frac{p\mathrm{e}^t}{1-(1-p)\mathrm{e}^t}\right)^{n_0}, \qquad 0 < t < -\log(1-p).$$

Then, for $x > n_0/p$,

$$\mathbb{P}(N \ge x) \le \min_{0<t<-\log(1-p)} \mathrm{e}^{-tx}\left(\frac{p\mathrm{e}^t}{1-(1-p)\mathrm{e}^t}\right)^{n_0} = \frac{(1-p)^{x-n_0}x^x p^{n_0}}{(x-n_0)^{x-n_0}n_0^{n_0}}.$$

Similarly,

$$\mathbb{P}(N \le x) \le \mathrm{e}^{tx}\mathbb{E}(\mathrm{e}^{-tN}) = \mathrm{e}^{tx}\left(\frac{p\mathrm{e}^{-t}}{1-(1-p)\mathrm{e}^{-t}}\right)^{n_0}, \qquad t > 0,$$

and thus, for $0 < x < n_0/p$,

$$\mathbb{P}(N \le x) \le \min_{t>0} \mathrm{e}^{tx}\left(\frac{p\mathrm{e}^{-t}}{1-(1-p)\mathrm{e}^{-t}}\right)^{n_0} = \frac{(1-p)^{x-n_0}x^x p^{n_0}}{(x-n_0)^{x-n_0}n_0^{n_0}}.$$

Therefore, $F_N(N) \ge \alpha/2$, $F_{N-}(N) \le 1-\alpha/2$ implies that

$$N \ge \frac{n_0}{p} \text{ and } \frac{(1-p)^{N-n_0}N^N p^{n_0}}{(N-n_0)^{N-n_0}n_0^{n_0}} \ge \frac{\alpha}{2} \quad \text{or} \quad N \le \frac{n_0}{p} \text{ and } \frac{(1-p)^{N-n_0}N^N p^{n_0}}{(N-n_0)^{N-n_0}n_0^{n_0}} \ge \frac{\alpha}{2}.$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

Finally, we get that

$$\left\{0 < p < 1 : \frac{(1-p)^{N-n_0} N^N p^{n_0}}{(N-n_0)^{N-n_0} n_0^{n_0}} \geq \frac{\alpha}{2}\right\}$$

is a valid $(1-\alpha)$-level confidence region for p under the targeted stopping setting. After simplification, we summarize our result in the following theorem.

Theorem 7. (Validity of the Chernoff CI under targeted stopping.) *Suppose that we keep sampling from* Bernoulli(p) *until we get n_0 successes, and the sample size is denoted by N. Then*

$$\left\{0 < p < 1 : p^{n_0}(1-p)^{N-n_0} \geq \frac{\alpha}{2}\left(\frac{n_0}{N}\right)^{n_0}\left(1-\frac{n_0}{N}\right)^{N-n_0}\right\} \tag{7}$$

is a valid $(1-\alpha)$-level confidence region for p.

It is easy to verify that $f(p) = p^{n_0}(1-p)^{N-n_0} - (\alpha/2)(n_0/N)^{n_0}(1-n_0/N)^{N-n_0}$ is increasing in $[0, n_0/N]$ and decreasing in $[n_0/N, 1]$. Moreover, we observe that $f(0) = f(1) < 0$ and $f(n_0/N) > 0$. Thus, (7) is actually an interval, and we could use the bisection method to numerically compute the bounds. Nevertheless, this CI is not as easy to study analytically as under the standard setting, so we will not include its half-width result in the following section.

6.1.2. *BE CI.* Now we apply the BE theorem again. We still assume that $p < \frac{1}{2}$ is known a priori. By the theorem, we get

$$\left|\mathbb{P}\left(\frac{N - n_0/p}{\sqrt{n_0(1-p)/p^2}} \leq x\right) - \Phi(x)\right| \leq \frac{C\rho_N}{\sigma_N^3\sqrt{n_0}}, \tag{8}$$

$$\left|\mathbb{P}\left(\frac{n_0/p - N}{\sqrt{n_0(1-p)/p^2}} \leq x\right) - \Phi(x)\right| \leq \frac{C\rho_N}{\sigma_N^3\sqrt{n_0}}, \tag{9}$$

where $\sigma_N^2 = \mathbb{E}(N_i - 1/p)^2 = (1-p)/p^2$ and $\rho_N = E|N_i - 1/p|^3$.

We need to deal with ρ_N first. In fact, we know that

$$p^3\rho_N = p^3 E\left|N_i - \frac{1}{p}\right|^3 = E|pN_i - 1|^3 \leq 1 + 3p\mathbb{E}(N_i) + 3p^2\mathbb{E}(N_i^2) + p^3\mathbb{E}(N_i^3).$$

Since $N_i \sim$ Geometric(p), we know that

$$\mathbb{E}(N_i) = \frac{1}{p}, \qquad \mathbb{E}(N_i^2) = \frac{2-p}{p^2}, \qquad \mathbb{E}(N_i^3) = \frac{p^2 - 6p + 6}{p^3},$$

and thus $p^3\rho_N \leq p^2 - 6p + 6 + 3(2-p) + 3 + 1 = p^2 - 12p + 16 \leq 16$. Hence,

$$\frac{C\rho_N}{\sigma_N^3\sqrt{n_0}} = \frac{Cp^3\rho_N}{(1-p)^{3/2}\sqrt{n_0}} \leq \frac{C'}{(1-p)^{3/2}\sqrt{n_0}},$$

where $C' = 16C$ is an absolute constant.

By setting $x = \sqrt{(p^2/n_0(1-p))}(N - (n_0/p))$ in (8) and $x = \sqrt{(p^2/n_0(1-p))}((n_0/p) - N)$ in (9), we get

$$\left|F_N(N) - \Phi\left(\frac{Np - n_0}{\sqrt{n_0(1-p)}}\right)\right| \leq \frac{C'}{\sqrt{n_0(1-p)^3}},$$

$$\left|1 - F_{N-}(N) - \Phi\left(\frac{n_0 - Np}{\sqrt{n_0(1-p)}}\right)\right| \leq \frac{C'}{\sqrt{n_0(1-p)^3}}.$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

Hence, $F_N(N) \geq \alpha/2$, $F_{N-}(N) \leq 1-\alpha/2$ implies that

$$p \geq n_0/N \text{ and } \Phi\left(\frac{n_0 - Np}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2} \qquad \text{or}$$

$$p \leq n_0/N \text{ and } \Phi\left(\frac{Np - n_0}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}.$$

Thus, we have developed a valid confidence region under this particular setting that is similar to the one in Section 5.

Theorem 8. (Validity of the BE CI under targeted stopping.) *Suppose that we keep sampling from* Bernoulli(p) *until we get n_0 successes, and the sample size is denoted by N. Assume that $p < \frac{1}{2}$. Then*

$$\left\{0 < p \leq \frac{n_0}{N} \wedge \frac{1}{2} : \Phi\left(\frac{Np - n_0}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}\right\}$$
$$\cup \left\{\frac{n_0}{N} \leq p < \frac{1}{2} : \Phi\left(\frac{n_0 - Np}{\sqrt{n_0(1-p)}}\right) + \frac{C'}{\sqrt{n_0(1-p)^3}} \geq \frac{\alpha}{2}\right\} \quad (10)$$

is a valid $(1-\alpha)$-level confidence region for p. Here, C' is a universal constant. In particular, we can pick $C' = 16C$, where C is the constant in the BE theorem.

Note that when n_0 is not large enough, we have $C'/\sqrt{n_0(1-p)^3} \geq \alpha/2$ anyway. That is to say, this confidence region is not really practical. However, it could still provide an insight into how close the CLT or Wilson intervals are to a valid one.

6.2. Analyses of half-widths

As mentioned before, under the targeted stopping setting, the Chernoff CI is no longer easy to analyze. Thus, in this subsection we only cover the analyses for the CLT, Wilson, and BE CIs. For the first two, as the formulas of the CIs are the same as the standard setting, we simply present the results here. For the BE CI, the main idea of the derivations is similar to the standard setting, but there are differences in the technical details. In particular, we are now able to get a non-trivial lower bound.

6.2.1. *CLT and Wilson CIs.* Under the targeted stopping setting, the formulas of the CLT and Wilson CIs are the same as under the standard setting with $\hat{p} = n_0/N$. Thus, clearly we still have that $\hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0} - \hat{p} = \hat{p} - \hat{p}^{\mathrm{CLT}}_{\mathrm{l},n_0} = \Theta(\hat{p}/\sqrt{n_0}) = \Theta(\sqrt{n_0}/N)$, and that $\left|\hat{p}^{\mathrm{Wilson}}_{\mathrm{u},n_0} - \hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0}\right| = O(1/N) = O(\hat{p}/n_0)$, $\left|\hat{p}^{\mathrm{Wilson}}_{\mathrm{l},n_0} - \hat{p}^{\mathrm{CLT}}_{\mathrm{l},n_0}\right| = O(1/N) = O(\hat{p}/n_0)$.

6.2.2. *BE CI.* Similar to Section 5, the confidence region (10) could be further relaxed. However, unlike in the standard setting, now the error term $C'/\sqrt{n_0(1-p)^3}$ could be well controlled for tiny p and as a result we are able to get a non-trivial lower bound in this case. More concretely, for any $0 < \lambda < 1$,

$$0 < p < 1,\ \frac{C'}{\sqrt{n_0(1-p)^3}} \geq (1-\lambda)\frac{\alpha}{2} \quad \Longrightarrow \quad p \geq 1 - \left(\frac{4C'}{n_0(1-\lambda)^2\alpha^2}\right)^{1/3}.$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

If we could find $0<\lambda<1$ such that $(4C'^2/n_0(1-\lambda)^2\alpha^2)^{1/3}=\frac{1}{2}$ then, for any $0<p<\frac{1}{2}$,

$$\frac{C'}{\sqrt{n_0(1-p)^3}}\le(1-\lambda)\frac{\alpha}{2}.$$

As a result, any p in (10) must satisfy

$$0<p\le\frac{n_0}{N}\wedge\frac{1}{2},\ \Phi\left(\frac{Np-n_0}{\sqrt{n_0(1-p)}}\right)\ge\lambda\frac{\alpha}{2}\quad\text{or}\quad\frac{n_0}{N}\le p<\frac{1}{2},\ \Phi\left(\frac{n_0-Np}{\sqrt{n_0(1-p)}}\right)\ge\lambda\frac{\alpha}{2}.$$

After simplification, we get

$$0<p\le\frac{n_0}{N},\ \Phi\left(\frac{Np-n_0}{\sqrt{n_0(1-p)}}\right)\ge\lambda\frac{\alpha}{2}\Longrightarrow p\ge\frac{2Nn_0-z^2_{\lambda\alpha/2}n_0-\sqrt{4z^2_{\lambda\alpha/2}Nn_0(N-n_0)+z^4_{\lambda\alpha/2}n_0^2}}{2N^2},$$

$$\frac{n_0}{N}\le p<1,\ \Phi\left(\frac{n_0-Np}{\sqrt{n_0(1-p)}}\right)\ge\lambda\frac{\alpha}{2}\Longrightarrow p\le\frac{2Nn_0-z^2_{\lambda\alpha/2}n_0+\sqrt{4z^2_{\lambda\alpha/2}Nn_0(N-n_0)+z^4_{\lambda\alpha/2}n_0^2}}{2N^2}.$$

Thus, (10) could be relaxed into a valid $(1-\alpha)$-level CI, which is defined more rigorously in the following theorem.

Theorem 9. (Relaxed BE CI under targeted stopping.) *Suppose that we keep sampling from* Bernoulli(p) *until we get n_0 successes, and the sample size is denoted by N. Assume that $p<\frac{1}{2}$. Let $\lambda=1-(4\sqrt{2}C'/\sqrt{n_0}\alpha)$. Then, for any $n_0>32C'^2/\alpha^2$,*

$$\hat{p}^{\mathrm{BE}}_{\mathrm{u},n_0}=\frac{2Nn_0-z^2_{\lambda\alpha/2}n_0+\sqrt{4z^2_{\lambda\alpha/2}Nn_0(N-n_0)+z^4_{\lambda\alpha/2}n_0^2}}{2N^2},$$

$$\hat{p}^{\mathrm{BE}}_{\mathrm{l},n_0}=\frac{2Nn_0-z^2_{\lambda\alpha/2}n_0-\sqrt{4z^2_{\lambda\alpha/2}Nn_0(N-n_0)+z^4_{\lambda\alpha/2}n_0^2}}{2N^2}$$

is a valid $(1-\alpha)$-level CI for p. Here, C' is the same as in Theorem 8.

Finally, like in the standard setting, we will compare the difference between the BE and CLT CIs.

Theorem 10. (Half-width of the BE CI under targeted stopping.) *Assume that $p<\frac{1}{2}$, and that $\hat{p}^{\mathrm{BE}}_{\mathrm{u},n_0}$ and $\hat{p}^{\mathrm{BE}}_{\mathrm{l},n_0}$ are as defined in Theorem 9. Then there is a constant C_0', which does not depend on p and N, such that $\hat{p}^{\mathrm{BE}}_{\mathrm{u},n_0}-\hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0}\le C_0'/N$ and $\hat{p}^{\mathrm{CLT}}_{\mathrm{l},n_0}-\hat{p}^{\mathrm{BE}}_{\mathrm{l},n_0}\le C_0'/N$.*

Therefore, under the targeted stopping setting, we could justify that the CLT CI is not too far from a valid one in terms of both upper and lower bounds.

7. Numerical experiments

To close the paper, we perform some numerical experiments to visualize the differences among the CIs.

7.1. Experiments under the standard setting

The true value is chosen as $p=10^{-6}$. For each of the settings $n=5/p$, $10/p$, $30/p$, $50/p$, and $100/p$ we conduct 100 000 experimental repetitions and calculate the CIs with $\alpha=0.05$.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

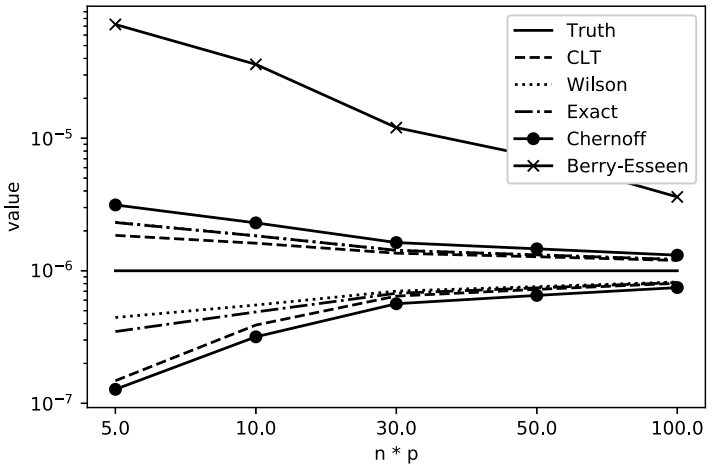

FIGURE 3. Average values of the confidence upper and lower bounds under the standard setting.

Figure 3 and Table 3 respectively present the average values of the confidence upper and lower bounds from the 100 000 repetitions and the coverage probabilities of the five CIs covered in this paper.

As analyzed in Section 5.2, when np is large the CIs scale similarly, except that BE fails to give a non-zero lower bound. While the CLT interval is closest to the truth in terms of the mean value of the upper bound, it is not reliable, especially when np is small. For instance, when $np = 5$, its coverage probability is only 0.858, which is much lower than the nominal confidence level of 0.95. The Wilson and exact CIs are quite similar, especially for the upper bound. However, we notice that the Wilson bound sometimes fails to achieve the nominal confidence level, but the error in the coverage probability is acceptable to some extent. The Chernoff and BE CIs are conservative, as expected, since they further relax the conservative confidence region (4). We would like to point out that though the BE upper bound seems to be much larger than the Chernoff, it decays much faster as np increases, which coincides with Theorems 4 and 6.

7.2. Experiments under targeted stopping

Now we consider the targeted stopping setting. We again set $p = 10^{-6}$. For each of the settings $n_0 = 5, 10, 30, 50, 100$ we conduct 100 000 experimental repetitions and calculate the CIs with $\alpha = 0.05$. Figure 4 and Table 4 respectively present the average values of the confidence upper and lower bounds from the 100 000 repetitions and the coverage probabilities of the CIs. Note that we do not include the BE CI since it is trivial due to the small n_0, as aforementioned.

The Chernoff CI is still conservative, as expected, since it further relaxes the conservative confidence region (5). We focus on comparing the other three CIs. From Figure 4 we see that, for the upper bound, the CLT bound is the closest to the truth and the Wilson bound is the farthest. For the lower bound, the Wilson bound is the closest and the CLT bound is the farthest. The exact bound falls in between. On the other hand, they all have a similar magnitude, and when n_0 is large they are all close to each other. In terms of coverage probability (Table 4), the three CIs also have similar performance when n_0 is large. In particular, in relation to the

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

TABLE 3. Coverage probabilities of the CIs under the standard setting.

np	CLT	Wilson	Exact	Chernoff	BE
5	0.869 90	0.962 52	0.979 92	0.999 30	1.0
10	0.926 01	0.962 52	0.975 42	0.997 67	1.0
30	0.931 00	0.944 76	0.955 31	0.995 60	1.0
50	0.950 98	0.945 96	0.954 51	0.996 59	1.0
100	0.945 01	0.949 20	0.954 85	0.995 75	1.0

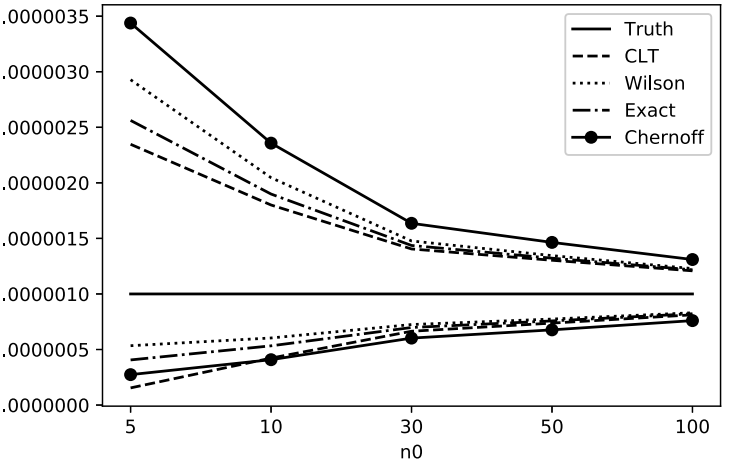

FIGURE 4. Average values of the confidence upper and lower bounds under the targeted stopping setting.

motivation in Section 2, if we can sample until we observe enough (say 30) successes, then the CLT CI is indeed reliable to use, although it is not guaranteed to be valid.

8. Conclusion

In this paper we have studied the uncertainty quantification, more precisely the construction, validity, and tightness of CIs, for rare-event probability using naive sample proportion estimators from Bernoulli data. We focused on two settings, the standard setting where the sample size is fixed, and the targeted stopping setting where the number of successes is fixed. Under each setting, we first reviewed the existing CLT, Wilson, and exact CIs. It is known that the CLT and Wilson CIs are not necessarily valid in the sense that the coverage probability can be lower than the nominal confidence level, and the exact CI is valid yet its tightness is hard to analyze. These motivated us to derive other valid CIs with more explicit expressions. More specifically, we relaxed the exact confidence region via inverting Chernoff's inequality and the BE theorem to obtain Chernoff and BE CIs, respectively. Tables 1 and 2 in Section 4 provide a comprehensive summary of our findings, and we briefly summarized our key findings in Section 4.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

TABLE 4. Coverage probabilities of the CIs under the targeted stopping setting.

n_0	CLT	Wilson	Exact	Chernoff
5	0.956 60	0.924 70	0.949 98	0.992 33
10	0.955 11	0.937 14	0.950 15	0.993 00
30	0.951 20	0.945 64	0.949 51	0.993 28
50	0.951 69	0.947 50	0.950 25	0.992 76
100	0.950 77	0.948 90	0.950 07	0.993 37

Overall, we recommend the exact CI when we want to ensure a guarantee of the nominal confidence level, otherwise we suggest using the Wilson CI, given its excellent empirical performance. Moreover, in either the standard setting or the targeted stopping setting, we have justified that the CLT CI is not far from the valid Chernoff and BE CIs, so it also exhibits reasonable tightness in terms of half-width. However, its coverage probability can deviate significantly from the nominal level when np is small. The latter two intervals are conservative and hence not recommended for use in practice, but they provide useful insights in understanding that the CLT and Wilson CIs are relatively close to these two valid CIs.

Appendix A. Proofs

Proof of Theorem 3. We get directly from the formula for $\hat{p}_{\mathrm{u}}^{\mathrm{Wilson}}$ that

$$\begin{aligned}
&\left|\hat{p}_{\mathrm{u}}^{\mathrm{Wilson}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}}\right| \\
&= \left| \frac{1 - 2\hat{p} + \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z_{1-\alpha/2}^2\right)} - 2\left(1 + \left(n/z_{1-\alpha/2}^2\right)\right) z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}}{2\left(1 + \left(n/z_{1-\alpha/2}^2\right)\right)} \right| \\
&\le \frac{|1 - 2\hat{p}| + |2z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}| + \left|\sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z_{1-\alpha/2}^2\right)} - (2n/z_{1-\alpha/2})\sqrt{\hat{p}(1-\hat{p})/n}\right|}{2n/z_{1-\alpha/2}^2}.
\end{aligned}$$

We have that $|1 - 2\hat{p}| \le 1$, $|2z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}| \le z_{1-\alpha/2}/\sqrt{n}$, and

$$\begin{aligned}
&\left| \sqrt{1 + \frac{4n\hat{p}(1-\hat{p})}{z_{1-\alpha/2}^2}} - \frac{2n}{z_{1-\alpha/2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right| \\
&\qquad = \left| \frac{1}{\sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z_{1-\alpha/2}^2\right)} + (2n/z_{1-\alpha/2})\sqrt{\hat{p}(1-\hat{p})/n}} \right| \le 1,
\end{aligned}$$

which concludes the proof for the upper bounds. The proof for the lower bound is almost the same. □

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

Proof of Theorem 4. We have

$$\hat{p}_{\mathrm{u}}^{\mathrm{Chernoff}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}} = \sqrt{\frac{2\log(2/\alpha)\hat{p}}{n} + \frac{(\log(2/\alpha))^2}{n^2}} + \frac{\log(2/\alpha)}{n} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\geq \left(\sqrt{2\log(2/\alpha)} - z_{1-\alpha/2}\right)\sqrt{\frac{\hat{p}}{n}} + \frac{\log(2/\alpha)}{n}.$$

Similarly, we could prove the inequality for the lower bounds. □

Proof of Theorem 5. Following our derivations, it suffices to show that, for the given N_0 and any $n > N_0$, $0 \leq \lambda \leq 1 - (4C/\sqrt{n}\alpha)$ and $U_1 \leq U_2$. Obviously, $2\tilde{C}/\sqrt{n}\alpha \leq u < 1$, so $\lambda > 0$. On the other side, $\lambda \leq 1 - (4C/\sqrt{n}\alpha)$ holds since $n \geq (4C/u\alpha)^2$.

Now we prove that $U_1 \leq U_2$ for $n > N_0$. Indeed, if $\tilde{C} = C/\sqrt{\hat{p}(1-\hat{p})}$, then $U_1 = \hat{p} \wedge (1-\hat{p}) \leq \hat{p}$ and we know that $U_2 \geq \hat{p}$, so $U_1 \leq U_2$. In the other case that $\tilde{C} = u\sqrt{n}\alpha/2$, we have

$$U_1 = \frac{1 - \sqrt{1 - (16C^2/nu^2\alpha^2)}}{2} = \frac{16C^2/nu^2\alpha^2}{2(1 + \sqrt{1 - (16C^2/nu^2\alpha^2)})}$$

$$\leq \frac{16C^2/nu^2\alpha^2}{2(1 + 1 - (16C^2/nu^2\alpha^2))} = \frac{4C^2}{nu^2\alpha^2 - 8C^2},$$

$$U_2 \geq \frac{1}{1 + \left(n/z^2_{(1-u)\alpha/2}\right)} = \frac{z^2_{(1-u)\alpha/2}}{z^2_{(1-u)\alpha/2} + n}.$$

Since u is chosen such that $4C^2/u^2\alpha^2 < z^2_{(1-u)\alpha/2}$ and

$$n \geq \frac{12 z^2_{(1-u)\alpha/2} C^2}{z^2_{(1-u)\alpha/2} u^2\alpha^2 - 4C^2},$$

we get

$$\frac{4C^2}{nu^2\alpha^2 - 8C^2} \leq \frac{z^2_{(1-u)\alpha/2}}{z^2_{(1-u)\alpha/2} + n},$$

and hence $U_1 \leq U_2$. Note that as $u \uparrow 1$, $4C^2/u^2\alpha^2 \to 4C^2/\alpha^2$ while $z^2_{(1-u)\alpha/2} \to \infty$, and thus such a u exists. □

Proof of Theorem 6. We have

$$\hat{p}_{\mathrm{u}}^{\mathrm{BE}} - \hat{p}_{\mathrm{u}}^{\mathrm{CLT}} = \frac{1 - 2\hat{p} + \sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)} - 2\left(1 + \left(n/z^2_{\lambda\alpha/2}\right)\right)z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}}{2\left(1 + \left(n/z^2_{\lambda\alpha/2}\right)\right)}.$$

We first deal with

$$\sqrt{1 + \frac{4n\hat{p}(1-\hat{p})}{z^2_{\lambda\alpha/2}}} - 2\left(1 + \frac{n}{z^2_{\lambda\alpha/2}}\right)z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= \frac{\left(1 - \left(8z^2_{1-\alpha/2}\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)\right) - 4z^2_{1-\alpha/2}\hat{p}(1-\hat{p})/n + \left(4n\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)\left(1 - \left(z^2_{1-\alpha/2}/z^2_{\lambda\alpha/2}\right)\right)}{\sqrt{1 + \left(4n\hat{p}(1-\hat{p})/z^2_{\lambda\alpha/2}\right)} + 2\left(1 + \left(n/z^2_{\lambda\alpha/2}\right)\right)z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}}.$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

The denominator satisfies

$$\sqrt{1+\frac{4n\hat{p}(1-\hat{p})}{z_{\lambda\alpha/2}^2}}+2\left(1+\frac{n}{z_{\lambda\alpha/2}^2}\right)z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\geq\left(\frac{2\sqrt{n\hat{p}(1-\hat{p})}(z_{1-\lambda\alpha/2}+z_{1-\alpha/2})}{z_{\lambda\alpha/2}^2}\right)\vee 1.$$

Note that $(z_{1-\lambda\alpha/2}+z_{1-\alpha/2})/z_{\lambda\alpha/2}^2$ increases with the value of λ. Since $\lambda\geq 1-u>0$, we can find a constant C_1 such that

$$\sqrt{1+\frac{4n\hat{p}(1-\hat{p})}{z_{\lambda\alpha/2}^2}}+2\left(1+\frac{n}{z_{\lambda\alpha/2}^2}\right)z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\geq(C_1\sqrt{n\hat{p}(1-\hat{p})})\vee 1.$$

Then we deal with the numerator. We know that $z_{\alpha/2}=\Phi^{-1}(\alpha/2)$ and $z_{\lambda\alpha/2}=\Phi^{-1}(\lambda\alpha/2)$. By Taylor expansion, we have

$$\frac{1}{z_{\lambda\alpha/2}^2}=\frac{1}{z_{\alpha/2}^2}-\frac{2\sqrt{2\pi}}{z_{\alpha/2}^3}\mathrm{e}^{z_{\alpha/2}^2/2}(\lambda-1)\frac{\alpha}{2}+r(\lambda).$$

Here, $r(\lambda)$ is continuous in λ and $r(\lambda)/(1-\lambda)\to 0$ as $\lambda\uparrow 1$. We also know that $1-\lambda\leq u$, and thus $|r(\lambda)/(1-\lambda)|=|(\sqrt{n}\alpha r(\lambda))/(2\tilde{C})|$ is bounded by a constant. Hence, $|\sqrt{n\hat{p}(1-\hat{p})}r(\lambda)|$ is bounded by a constant. We have

$$1-\frac{z_{1-\alpha/2}^2}{z_{\lambda\alpha/2}^2}=\frac{2\sqrt{2\pi}}{z_{1-\alpha/2}}\mathrm{e}^{z_{1-\alpha/2}^2/2}\frac{\tilde{C}}{\sqrt{n}}-z_{1-\alpha/2}^2r(\lambda).$$

Thus, the numerator satisfies

$$\left|\left(1-\frac{8z_{1-\alpha/2}^2\hat{p}(1-\hat{p})}{z_{\lambda\alpha/2}^2}\right)-\frac{4z_{1-\alpha/2}^2\hat{p}(1-\hat{p})}{n}+\frac{4n\hat{p}(1-\hat{p})}{z_{\lambda\alpha/2}^2}\left(1-\frac{z_{1-\alpha/2}^2}{z_{\lambda\alpha/2}^2}\right)\right|$$

$$\leq 1+8\hat{p}(1-\hat{p})+\frac{4z_{1-\alpha/2}^2\hat{p}(1-\hat{p})}{n}+\frac{4n\hat{p}(1-\hat{p})}{z_{\lambda\alpha/2}^2}\left(\frac{2\sqrt{2\pi}}{z_{1-\alpha/2}}\mathrm{e}^{z_{1-\alpha/2}^2/2}\frac{\tilde{C}}{\sqrt{n}}-z_{1-\alpha/2}^2r(\lambda)\right).$$

Clearly, the first three terms divided by the denominator are bounded by some constants. Now we consider the fourth term. Since $|\sqrt{n\hat{p}(1-\hat{p})}r(\lambda)|$ is bounded, we can also get that the fourth term divided by the denominator is bounded by some universal constant.

Therefore, combining the above results, we know that

$$\left|1-2\hat{p}+\sqrt{1+\frac{4n\hat{p}(1-\hat{p})}{z_{\lambda\alpha/2}^2}}-2\left(1+\frac{n}{z_{\lambda\alpha/2}^2}\right)z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right|\leq C_2,$$

where C_2 is a positive constant. We also have

$$2\left(1+\frac{n}{z_{\lambda\alpha/2}^2}\right)\geq\frac{2n}{z_{(1-u)\alpha/2}^2}.$$

Hence, the error term satisfies

$$\left|\frac{1-2\hat{p}+\sqrt{1+\left(4n\hat{p}(1-\hat{p})/z_{\lambda\alpha}^2\right)}-2\left(1+\left(n/z_{\lambda\alpha}^2\right)\right)z_{1-\alpha}\sqrt{\hat{p}(1-\hat{p})/n}}{2\left(1+\left(n/z_{\lambda\alpha}^2\right)\right)}\right|\leq\frac{C_0}{n}$$

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

for some constant C_0. From the above derivations, we find that C_0 only depends on α and the choice of u. □

Proof of Theorem 9. If $n_0 > 32C'^2/\alpha^2$, then $0 < \lambda < 1$. The validness of the CI is justified by the derivations above the theorem. □

Proof of Theorem 10. We have

$$\hat{p}^{\mathrm{BE}}_{\mathrm{u},n_0} - \hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0} = -\frac{z^2_{\lambda\alpha/2}n_0}{2N^2} + \sqrt{\frac{z^2_{\lambda\alpha/2}n_0(N-n_0)}{N^3} + \frac{z^4_{\lambda\alpha/2}n_0^2}{4N^4}} - z_{1-\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}}$$

$$= \frac{(z^2_{\lambda\alpha/2} - z^2_{1-\alpha/2})n_0(N-n_0)/N^3 - \left(n_0 z^2_{\lambda\alpha/2}z_{1-\alpha/2}/N^2\right)\sqrt{n_0(N-n_0)/N^3}}{\sqrt{z^2_{\lambda\alpha/2}n_0(N-n_0)/N^3 + z^4_{\lambda\alpha/2}n_0^2/4N^4} + z^2_{\lambda\alpha/2}n_0/2N^2 + z_{1-\alpha/2}\sqrt{n_0(N-n_0)/N^3}}.$$

First, for the denominator:

$$\sqrt{\frac{z^2_{\lambda\alpha/2}n_0(N-n_0)}{N^3} + \frac{z^4_{\lambda\alpha/2}n_0^2}{4N^4}} + \frac{z^2_{\lambda\alpha/2}n_0}{2N^2} + z_{1-\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}} \geq \sqrt{\frac{z^2_{\lambda\alpha/2}n_0(N-n_0)}{N^3}}$$

$$= z_{1-\lambda\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}}.$$

Next, for the numerator:

$$\frac{(z^2_{\lambda\alpha/2} - z^2_{1-\alpha/2})n_0(N-n_0)}{N^3} - \frac{n_0 z^2_{\lambda\alpha/2}z_{1-\alpha/2}}{N^2}\sqrt{\frac{n_0(N-n_0)}{N^3}} \leq \left(1 - \frac{z^2_{1-\alpha/2}}{z^2_{\lambda\alpha/2}}\right)\frac{z^2_{\lambda\alpha/2}n_0(N-n_0)}{N^3}.$$

As mentioned in the proof of Theorem 6, we have

$$\frac{1}{z^2_{\lambda\alpha/2}} = \frac{1}{z^2_{\alpha/2}} - \frac{2\sqrt{2\pi}}{z^3_{\alpha/2}}\mathrm{e}^{z^2_{\alpha/2}/2}(\lambda-1)\frac{\alpha}{2} + r(\lambda),$$

$$1 - \frac{z^2_{1-\alpha/2}}{z^2_{\lambda\alpha/2}} = \frac{\sqrt{2\pi}}{z_{1-\alpha/2}}\mathrm{e}^{z^2_{\alpha/2}/2}\frac{C_1'}{\sqrt{n_0}}\alpha - r(\lambda)z^2_{\alpha/2},$$

where $r(\lambda)$ is continuous in λ and $r(\lambda)/(1-\lambda) \to 0$ as $\lambda \uparrow 1$. We know that $1-\lambda = C_1'/\sqrt{n_0}$ for some constant $C_1' > 0$ from the choice of λ. For $n_0 > 32C'^2/\alpha^2$, λ is bounded away from 0, and hence $|r(\lambda)/(1-\lambda)| = |\sqrt{n_0}r(\lambda)/C_1'|$ has a constant upper bound. Hence, there exists a constant $C_2' > 0$ such that

$$\left(1 - \frac{z^2_{1-\alpha/2}}{z^2_{\lambda\alpha/2}}\right)\frac{z^2_{\lambda\alpha/2}n_0(N-n_0)}{N^3} \leq \frac{C_2'}{\sqrt{n_0}}\frac{z^2_{\lambda\alpha/2}n_0(N-n_0)}{N^3}.$$

Combining the results, we get

$$\hat{p}^{\mathrm{BE}}_{\mathrm{u},n_0} - \hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0} \leq \frac{C_2'}{\sqrt{n_0}}z_{1-\lambda\alpha/2}\sqrt{\frac{n_0(N-n_0)}{N^3}} \leq \frac{C_2'z_{1-\lambda\alpha/2}}{N} \leq \frac{C_3'}{N},$$

where $C_3' > 0$ is a constant and we get the last inequality since λ has a non-zero lower bound.

Now notice that $\hat{p}^{\mathrm{CLT}}_{\mathrm{l},n_0} - \hat{p}^{\mathrm{BE}}_{\mathrm{l},n_0} = \hat{p}^{\mathrm{BE}}_{\mathrm{u},n_0} - \hat{p}^{\mathrm{CLT}}_{\mathrm{u},n_0} + z^2_{\lambda\alpha/2}n_0/N^2$ and $z^2_{\lambda\alpha/2}n_0/N^2 \leq z^2_{\lambda\alpha/2}/N = O(1/N)$. Therefore, we could find a constant C_0' such that the theorem holds. □

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press

Acknowledgements

We thank the associate editor and referees for their helpful suggestions that have greatly improved this paper. A preliminary conference version of this work has appeared in [8].

Funding information

We gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] AGRESTI, A. AND COULL, B. A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *Amer. Statistician* **52**, 119–126.
[2] ARIEF, M. *et al.* (2021). Deep probabilistic accelerated evaluation: A robust certifiable rare-event simulation methodology for black-box safety-critical systems. In *International Conference on Artificial Intelligence and Statistics*, eds A. Banerjee and K. Fukumizu. Proceedings of Machine Learning Research, pp. 595–603.
[3] ASMUSSEN, S. AND ALBRECHER, H. (2010). *Ruin Probabilities* (Adv. Ser. Statist. Sci. Appl. Pron. **14**). World Scientific, Singapore.
[4] ASMUSSEN, S. *et al.* (1985). Conjugate processes and the simulation of ruin problems. *Stoch. Process. Appl.* **20**, 213–229.
[5] ASMUSSEN, S. AND GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer, New York.
[6] AU, S.-K. AND BECK, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Prob. Eng. Mechanics* **16**, 263–277.
[7] BAI, Y., HUANG, Z., LAM, H. AND ZHAO, D. (2022). Rare-event simulation for neural network and random forest predictors. *ACM Trans. Model. Comput. Simul.* **32**, 1–33.
[8] BAI, Y. AND LAM, H. (2020). On the error of naive rare-event Monte Carlo estimator. In *2020 IEEE Winter Simulation Conf. (WSC)*, pp. 397–408.
[9] BLANCHET, J., GLYNN, P. AND LAM, H. (2009). Rare event simulation for a slotted time M/G/S model. *Queueing Systems* **63**, 33–57.
[10] BLANCHET, J. AND LAM, H. (2012). State-dependent importance sampling for rare-event simulation: An overview and recent advances. *Surv. Operat. Res. Manag. Sci.* **17**, 38–59.
[11] BLANCHET, J. AND LAM, H. (2014). Rare-event simulation for many-server queues. *Math. Operat. Res.* **39**, 1142–1178.
[12] BUCKLEW, J. (2004). *Introduction to Rare Event Simulation*. Springer, New York.
[13] CLOPPER, C. J. AND PEARSON, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26**, 404–413.
[14] COLLAMORE, J. F. (2002). Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors. *Ann. Appl. Prob.* **12**, 382–421.
[15] DAVISON, A. C. AND SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc. B* **52**, 393–425.
[16] DUPUIS, P., LEDER, K. AND WANG, H. (2009). Importance sampling for weighted-serve-the-longest-queue. *Math. Operat. Res.* **34**, 642–660.
[17] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.
[18] GLASSERMAN, P. (2004). *Monte Carlo Methods in Financial Engineering* (Stoch. Model. Appl. Prob. **53**). Springer, New York.
[19] GLASSERMAN, P., HEIDELBERGER, P., SHAHABUDDIN, P. AND ZAJIC, T. (1999). Multilevel splitting for estimating rare event probabilities. *Operat. Res.* **47**, 585–600.
[20] GLASSERMAN, P., KANG, W. AND SHAHABUDDIN, P. (2008). Fast simulation of multifactor portfolio credit risk. *Operat. Res.* **56**, 1200–1217.
[21] GLASSERMAN, P. AND LI, J. (2005). Importance sampling for portfolio credit risk. *Manag. Sci.* **51**, 1643–1656.

[22] HEIDELBERGER, P. (1995). Fast simulation of rare events in queueing and reliability models. *ACM Trans. Model. Comput. Sim.* **5**, 43–85.
[23] HUANG, Z., LAM, H., LEBLANC, D. J. AND ZHAO, D. (2017). Accelerated evaluation of automated vehicles using piecewise mixture models. *IEEE Trans. Intellig. Transport. Syst.* **19**, 2845–2855.
[24] JUNEJA, S. AND SHAHABUDDIN, P. (2006). Rare-event simulation techniques: An introduction and recent advances. In *Simulation* (Handbooks Operat. Res. Manag. Sci. **13**), eds S. G. Henderson and B. L. Nelson. NORTH HOLLAND, Amsterdam, pp. 291–350.
[25] KROESE, D. P. AND NICOLA, V. F. (1999). Efficient estimation of overflow probabilities in queues with breakdowns. *Performance Evaluation* **36**, 471–484.
[26] MCNEIL, A. J., FREY, R. AND EMBRECHTS, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
[27] NICOLA, V. F., NAKAYAMA, M. K., HEIDELBERGER, P. AND GOYAL, A. (1993). Fast simulation of highly dependable systems with general failure and repair processes. *IEEE Trans. Computers* **42**, 1440–1452.
[28] NICOLA, V. F., SHAHABUDDIN, P. AND NAKAYAMA, M. K. (2001). Techniques for fast simulation of models of highly dependable systems. *IEEE Trans. Reliab.* **50**, 246–264.
[29] O'KELLY, M. *et al.* (2018). Scalable end-to-end autonomous vehicle testing via rare-event simulation. In *Proc. 32nd Int. Conf. Neural Inf. Proc. Syst.*, eds S. Bengio *et al.* Curran Associates, Inc., Red Hook, NY, pp. 9849–9860.
[30] RIDDER, A. (2009). Importance sampling algorithms for first passage time probabilities in the infinite server queue. *Europ. J. Operat. Res.* **199**, 176–186.
[31] RUBINSTEIN, R. Y. AND KROESE, D. P. (2016). *Simulation and the Monte Carlo Method*. John Wiley, Hoboken, NJ.
[32] SADOWSKY, J. S. (1991). Large deviations theory and efficient simulation of excessive backlogs in a GI/GI/M queue. *IEEE Trans. Automatic Control* **36**, 1383–1394.
[33] SADOWSKY, J. S. AND BUCKLEW, J. A. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inf. Theory* **36**, 579–588.
[34] SHAO, Q.-M. AND WANG, Q. (2013). Self-normalized limit theorems: A survey. *Prob. Surv.* **10**, 69–93.
[35] SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann. Statist.* **4**, 673–684.
[36] SMITH, R. L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications*, ed. J. T. DE OLIVEIRA. SPRINGER, DORDRECHT, pp. 621–638.
[37] SZECHTMAN, R. AND GLYNN, P. W. (2002). Rare-event simulation for infinite server queues. In *Proc. IEEE Winter Simul. Conf.*, Vol. 1, pp. 416–423.
[38] TUFFIN, B. (2004). On numerical problems in simulations of highly reliable Markovian systems. In *First IEEE Int. Conf. Quant. Eval. Syst.*, pp. 156–164.
[39] VILLÉN-ALTAMIRANO, M. AND VILLÉN-ALTAMIRANO, J. (1994). Restart: A straightforward method for fast simulation of rare events. In *Proc. IEEE Winter Simul. Conf.*, eds J. D. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, pp. 282–289.
[40] WANG, Q. AND HALL, P. (2009). Relative errors in central limit theorems for Student's t statistic, with applications. *Statistica Sinica* **19**, 343–354.
[41] WANG, Q. AND JING, B.-Y. (1999). An exponential nonuniform Berry–Esseen bound for self-normalized sums. *Ann. Prob.* **27**, 2068–2088.
[42] WEBB, S., RAINFORTH, T., TEH, Y. W. AND KUMAR, M. P. (2018). A statistical approach to assessing neural network robustness. Preprint, arXiv:1811.07209.
[43] WENG, T.-W. *et al.* (2018). Evaluating the robustness of neural networks: An extreme value theory approach. Preprint, arXiv:1801.10578.
[44] ZHAO, D. *et al.* (2017). Accelerated evaluation of automated vehicles in car-following maneuvers. *IEEE Trans. Intellig. Transport. Syst.* **19**, 733–744.
[45] ZHAO, D. *et al.* (2016). Accelerated evaluation of automated vehicles safety in lane-change scenarios based on importance sampling techniques. *IEEE Trans. Intellig. Transport. Syst.* **18**, 595–607.

https://doi.org/10.1017/jpr.2024.43 Published online by Cambridge University Press