The Sample Size for Estimating the Binomial Parameter with a Given Margin of Error

By B.K. Ghosh and Wei-Min Huang

Lehigh University

Consider the problem of constructing an estimator $T(X_n)$ for the binomial p and determining the smallest sample size N_T such that, for specified values of (γ, δ) and for *all* $n \geq N_T$, the interval $T(X_n) \pm \delta$ contains the true value of p with a minimum probability γ. In principle, any conventional $100\gamma\%$ confidence interval for p can be adapted to solve the present problem. We show that, for small δ, all known confidence intervals effectively lead to the estimator $T_b(X_n) = (X_n + \frac{1}{2}b)(n+b)^{-1}$ with the sample size $N_b \approx (z/2\delta)^2 + \delta^{-1}$, where z is the $\frac{1}{2}(1+\gamma)$ quantile of the standard normal distribution and $b \geq 0$ is a specified constant which does not depend on n. The major purpose of this paper is to propose the estimator $T^*(X_n) = (X_n + \frac{1}{2}a_\gamma\sqrt{n})(n + a_\gamma\sqrt{n})^{-1}$ with sample size $N^* \approx \{(z/2\delta) - a_\gamma\}^2 + \delta^{-1}$, where $a_\gamma = 1$ when $\gamma \geq .917$ and $0 \leq a_\gamma < 1$ is defined in (1.10) when $\gamma < .917$. The proposed method is more efficient in the sense that $N^* < N_b$ for any $b \geq 0$. These asymptotic conclusions are shown to be quite adequate for arbitrary values of $\delta \in [.01, .10]$ and $\gamma \in [.90, .99]$ by making exact calculations for the minimum coverage probabilities of $T_b(X_n) \pm \delta$ and $T^*(X_n) \pm \delta$ as well as for N_b and N^*.

1. Introduction. An ubiquitous but rarely researched problem of practical statistics arises in the context of a binomial distribution. Suppose we can observe the number of "successes" X_n in n independent "trials", each trial having the same unknown probability of success p, $0 < p < 1$. Let $\gamma \in (0,1)$ be a specified *confidence level* and $\delta \in (0, \frac{1}{2})$ be a specified *margin of error.* Then the problem is to construct an *estimator* $T(X_n)$ for p and predetermine the (smallest) *sample size* N_T such that

$$(1.1) \qquad \inf_{0<p<1} \gamma_T(p,n) \geq \gamma \quad \text{for } all \quad n \geq N_T,$$

where

$$(1.2) \qquad \gamma_T(p,n) = P_p(|T(X_n) - p| \leq \delta)$$

is the *coverage probability* of the interval $T(X_n) \pm \delta$. The implication of (1.1) is, of course, that $T(X_n) \pm \delta$ contains the true value of p with a minimum probability γ for any $n \geq N_T$. We allow $T(X_n)$ to involve γ and, clearly, $\gamma_T(p,n)$ will involve δ while N_T will generally depend on both γ and δ. We will sometimes qualify $\gamma_T(p,n)$ and N_T with the word *exact* in order to emphasize that (1.2) and N_T are computed under the *binomial* distribution of X_n. The problem just described is encountered almost daily in opinion polls, market research, clinical trials and quality control, where one usually chooses $.90 \leq \gamma \leq .99$ and $.01 \leq \delta \leq .10$. Note that for any given $\delta \geq \frac{1}{2}$ the trivial choice $T(X_n) = \frac{1}{2}$ with $N_T = 0$ provides a solution. Clearly, if $T(X_n)$ and $T^*(X_n)$ are two competing estimators and their sample sizes satisfy $N_T > N_{T^*}$, one should prefer to use $T^*(X_n) \pm \delta$. Indeed, a main objective of this paper is to propose a $T^*(X_n)$ which is better in this sense than a wide variety of $T(X_n)$. The interval $T(X_n) \pm \delta$ for any $n \geq N_T$ is also referred to as a *fixed-width* confidence interval for p (the actual width being $\leq 2\delta$). Another version of the present problem is to determine a pair $(T'(X_n), N_{T'})$ such that, for specified $\gamma \in (0,1)$, δ_1 and $\delta_2 (0 < \delta_1 < 1 - \delta_2 < 1)$, one wishes to guarantee $\inf_p P_p(T'(X_n) - \delta_1 \leq p \leq T'(X_n) + \delta_2) \geq \gamma$ for all $n \geq N_{T'}$. The equivalence of the two problems becomes clear if one identifies $T(X_n) = T'(X_n) + \frac{1}{2}(\delta_2 - \delta_1)$ and $\delta = \frac{1}{2}(\delta_1 + \delta_2)$.

The statistical literature treats the present problem routinely as an offshoot of a conventional confidence interval for p. Suppose $[L_n, U_n]$, $n \geq n_0$, is a sequence of $100\gamma\%$ confidence intervals for p, that is, $P_p(L_n \leq p \leq U_n) \geq \gamma$ for each p and $n \geq n_0$, L_n and U_n being functions of X_n. Then the standard approach is to choose

$$T(X_n) = \tfrac{1}{2}(L_n + U_n) \tag{1.3}$$

and determine N_T so as to guarantee $U_n - L_n \leq 2\delta$ for all $X_n \in \{0, 1, ..., n\}$ and $n \geq N_T$. One encounters, of course, numerous confidence intervals for p in the literature (see [9, 3, 5, 4, 2] for references) and, in principle, one can determine N_T under each of them, at least numerically. In Section 2, we show that all known confidence intervals lead asymptotically to the estimator

$$T_b(X_n) = (X_n + \tfrac{1}{2}b)(n+b)^{-1} = p_n + b(\tfrac{1}{2} - p_n)(n+b)^{-1} \tag{1.4}$$

and the sample size

$$N_b = (z/2\delta)^2 + \delta^{-1} + 0(1), \tag{1.5}$$

where p_n and z are defined by

$$p_n = n^{-1}X_n, \quad \Phi(z) = \tfrac{1}{2}(1+\gamma), \quad \Phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-x^2/2}dx, \tag{1.6}$$

and $b \geq 0$ is a constant depending only on z. By *asymptotic* we mean throughout that $n \to \infty$ and $\delta \to 0$ in such a way that $n\delta^2$ approaches a positive constant. The result in (1.5) incorporates the notion of *correction for continuity* (i.e., the term δ^{-1}), as explained in Section 2. The effect of b on N_b actually shows up in the error term $0(1)$ of (1.5) and its contribution turns out to be negligible relative to the magnitude of δ^{-1}. In Section 2 (see also Table 3), we carry out some exact calculations for $\inf_p \gamma_b(p,n)$ and N_b for certain values of b to show that, although $\inf_p \gamma_b(p,n)$ slightly increases and N_b slightly decreases as b goes up from zero, the approximation in (1.5) without the term $0(1)$is quite adequate.

Most textbooks recommend the solution

$$T(X_n) = p_n, \ N_T \approx (z/2\delta)^2 \tag{1.7}$$

which is based on the asymptotic confidence interval

$$[\ p_n - z\{p_n(1-p_n)/n\}^{1/2}, \ p_n + z\{p_n(1-p_n)/n\}^{1/2}\ ]. \tag{1.8}$$

This solution is a cruder version of (1.4)-(1.5) when $b = 0$ in that N_T of (1.7) ignores the correction for continuity introduced in N_0 of (1.5). In Section 2 (see also Table 3), we show that, for $.90 \leq \gamma \leq .99$ and $.01 \leq \delta \leq .10$, $(z/2\delta)^2$

may considerably underestimate the exact value of N_0 while $(z/2\delta)^2 + \delta^{-1}$ often overestimates. Similar conclusions can be made for other common choices of b in (1.4) (see Tables 1 and 2). Many of these conclusions are to be expected from the recent exhaustive investigation of several standard confidence intervals made by Brown, Cai, and Dasgupta [4].

Since all known confidence intervals for p lead to the same sample size (1.5) up to order δ^{-1}, a natural question now arises: Does an estimator $T^*(X_n)$ exist which satisfies (1.1) and requires a sample size N^* which is, up to order δ^{-1}, *smaller* than $(z/2\delta)^2 + \delta^{-1}$? For an intuitive answer, observe that, up to order δ^{-1}, N_b in (1.5) does not involve b because $\inf_p \gamma_b(p, n)$ occurs at $p = \frac{1}{2}$ under the normal approximation $(X_n - np)\{np(1-p)\}^{-1/2} \to N(0,1)$ up to order $n^{-1/2}$ and $\gamma_b(\frac{1}{2}, n)$ itself happens to be independent of b (see Section 2 for details). If, however, b itself were of order $n^{1/2}$ then $\gamma_b(\frac{1}{2}, n)$ would, indeed, depend on b and one should be able to exploit this fact in order to improve upon (1.5). In Section 3, we formally establish the following result. Let

$$T^*(X_n) = (X_n + \tfrac{1}{2}a_\gamma\sqrt{n})(n + a_\gamma\sqrt{n})^{-1} = p_n + a_\gamma(\tfrac{1}{2} - p_n)(\sqrt{n} + a_\gamma)^{-1}, \tag{1.9}$$

where

$$\begin{aligned} a_\gamma &= 1 \quad \text{if} \quad \gamma_0 \le \gamma < 1 \\ &= \{4(1+z^2)^{1/2} - 4 - z^2\}^{1/2} \quad \text{if} \quad \tfrac{1}{2} < \gamma < \gamma_0 \\ &= z \quad \text{if} \quad 0 < \gamma \le \tfrac{1}{2}, \end{aligned} \tag{1.10}$$

and γ_0 is defined by $\gamma_0 = 2\Phi(\sqrt{3}) - 1 = .91673548$. Then $T^*(X_n)$ satisfies (1.1) with sample size

$$N^* = \{(z/2\delta) - a_\gamma\}^2 + \delta^{-1} + 0(1). \tag{1.11}$$

It is easily verified that the middle expression in (1.10) increases from .60818862 to 1 as γ increases from $\frac{1}{2}$ to γ_0. Since $z = .67448975$ for $\gamma = \frac{1}{2}$, a_γ is obviously discontinuous at $\gamma = \frac{1}{2}$ and the reason for this is explained in Section 3. In practical applications, one may safely use $a_\gamma = 1$ when $.9 \le \gamma < \gamma_0$ because (1.10) shows that $.997 < a_\gamma \le 1$ for all $\gamma \in [.9, \gamma_0]$. It follows from (1.5) and (1.11) that, up to order δ^{-1} and for all $\gamma \in (0,1)$, we have

$$N_b - N^* \approx a_\gamma(\delta^{-1}z - a_\gamma) \tag{1.12}$$

which represents the approximate saving in sample size if one uses $T^*(X_n) \pm \delta$ for $n \ge N^*$ instead of $T_b(X_n) \pm \delta$ for $n \ge N_b$. The quantity in (1.12) lies between 15 and 257 for all $.90 \le \gamma \le .99$ and $.01 \le \delta \le .10$. Table 3 shows a numerical comparison between N_0 and N^*, both exact and approximate. It can be seen from the table that (1.11) without the 0(1) term is remarkably close to the exact value of N^* and, in fact, the slight overestimation by (1.11) makes $T^*(X_n) \pm \delta$ for $n \ge N^*$ more trustworthy in achieving the confidence level. It is also apparent from the last column of Table 3 that the percentage savings achieved by $T^*(X_n) \pm \delta$ over $T_0(X_n) \pm \delta$ are often substantial. Finally, one can logically think of choosing some other a in (1.9) instead of $a = a_\gamma$ of (1.10). In Section 3, we show that $a = a_\gamma$

is in a sense better than most other choices of a. Note that N^* and the coverage probability $\gamma^*(p,n)$ of $T^*(X_n) \pm \delta$ are also denoted by $N^{(a_\gamma)}$ and $\gamma^{(a_\gamma)}(p,n)$ in Section 3 for reasons made clear later.

In Section 4, we give the relevant formulas for the *exact* computation of the coverage probabilities and thence exact values of N_b and N^*. They also reveal some interesting features which are obscured by the normal approximation. For instance, contrary to what the normal approximation indicates, $\gamma_b(p,n)$ and $\gamma^*(p,n)$ have a local *maximum* at $p = \frac{1}{2}$ under arbitrary $\delta > (2n)^{-1}$ and, in fact, a genuine minimum of either of them does not even exist. Moreover, $\inf_p \gamma_b(p,n)$ and $\inf_p \gamma^*(p,n)$ are not monotonic in n and, as a consequence, the exact determination of N_b and N^* becomes non-trival and rather time-consuming.

2. From Confidence Intervals to $(T_b(X_n), N_b)$. Consider the estimator $T_b(X_n)$ defined in (1.4) for any specified $b \geq 0$. Let

$$\begin{aligned}\gamma_b(p,n) &= P_p(|T_b(X_n) - p| \leq \delta) \\ &= P_p((n+b)(p-\delta) - \tfrac{1}{2}b \leq X_n \leq (n+b)(p+\delta) - \tfrac{1}{2}b) \qquad (2.1)\end{aligned}$$

denote the coverage probability of $T_b(X_n) \pm \delta$ for any given $\delta \in (0, \frac{1}{2})$, $n \geq 1$ and $p \in (0,1)$. Then the following inequalities show that $N_b = N_b(\gamma, \delta)$ satisfying (1.1) does exist for every $\gamma \in (0,1)$:

$$b(1-2\delta)(2\delta)^{-1} \leq N_b \leq [\tfrac{1}{2}\omega^2 - b + \{(\tfrac{1}{2}\omega^2 - b)^2 + b^2(\omega^2 - 1)\}^{1/2}], \qquad (2.2)$$

where $\omega = \frac{1}{2}\delta^{-1}(1-\gamma)^{-1/2} > 1$ and $[x]$ denotes the smallest integer $\geq x$. The lower inequality in (2.2) follows from the fact that $\gamma_b(p,n) = 0$ for any $p < \frac{1}{2}b(n+b)^{-1} - \delta$ and therefore $\inf_p \gamma_b(p,n) \geq \gamma > 0$ implies that n must satisfy $b(n+b)^{-1} \leq 2\delta$. The upper inequality is a consequence of

$$\text{(2.3)} \quad \gamma_b(p,n) \geq 1 - \delta^{-2}E_p(T_b(X_n) - p)^2 \geq 1 - (4\delta^2)^{-1}(n+b^2)(n+b)^{-2} \quad \text{for all } p$$

and the requirement that $\inf_p \gamma_b(p,n) \geq \gamma$.

We will now derive the approximation (1.5) for N_b as $\delta \to 0$ and, to this end, it is necessary to explain the notion of *correction for continuity* in the present context. Let Y denote a normal variable with mean np and variance $np(1-p)$. Then the well-known asymptotic result $(X_n - np)\{np(1-p)\}^{-1/2} \to N(0,1)$, as $n \to \infty$, implies that, for $A < B$, $P_p(A \leq X_n \leq B) = P(A \leq Y \leq B) + 0(n^{-1/2})$ and, under the standard correction for continuity cited in textbooks, $P_p(A \leq X_n \leq B) = P(A - \frac{1}{2} \leq Y \leq B + \frac{1}{2}) + 0(n^{-1/2})$. The error term $0(n^{-1/2})$ is a consequence of the Berry-Esséen theorem (see [12]). Although the errors in both $P(A \leq Y \leq B)$ and $P(A - \frac{1}{2} \leq Y \leq B + \frac{1}{2})$ are theoretically of the same order, it is well known that the latter is usually far more accurate in practice (see [10, p. 62]). Now, if one wants to choose A and B to *guarantee* $P_p(A \leq X_n \leq B) \geq \gamma$ for given (p, γ), then the choice coming from the solution of $P(A \leq Y \leq B) = \gamma$ may be inadequate and the choice from $P(A - \frac{1}{2} \leq Y \leq B + \frac{1}{2}) = \gamma$ may be equally or more so because one may end up with

$$P_p(A \leq X_n \leq B) < P(A \leq Y \leq B) = \gamma$$

or

$$P(A \le Y \le B) < P_p(A \le X_n \le B) < P(A - \tfrac{1}{2} \le Y \le B + \tfrac{1}{2}) = \gamma.$$

Consequently, to be safe-sided one ought to use the more conservative correction

$$P_p(A \le X_n \le B) = P(A + \tfrac{1}{2} \le Y \le B - \tfrac{1}{2}) + 0(n^{-1/2}) \tag{2.4}$$

and solve A and B from $P(A+\frac{1}{2} \le Y \le B-\frac{1}{2}) = \gamma$. We will use the representation (2.4) throughout in order to obtain approximations for $\gamma_b(p,n)$ as well as for other coverage probabilities considered in this paper. Blyth and Still [3] also used (2.4) to modify certain conventional confidence intervals for p.

It can be verified from (2.1) that, for any fixed $b \ge 0$ and all $p \in (0,1)$, $\gamma_b(p,n) \to 0$ if $n \to \infty$ and $\delta n^{1/2} \to 0$, while $\gamma_b(p,n) \to 1$ if $n \to \infty$ and $\delta n^{1/2} \to \infty$. Consequently, for the purpose of asymptotically determining the smallest n to satisfy $\gamma_b(p,n) \ge \gamma$ for any $\gamma \in (0,1)$, it is necessary and sufficient to assume that, as $n \to \infty$, $\delta n^{1/2} \to \lambda$ for some finite positive λ. It follows from (2.1) and (2.4) that

$$\gamma_b(p,n) = \tilde{\gamma}_b(p,n) + 0(n^{-1/2}), \tag{2.5}$$

where

$$\tilde{\gamma}_b(p,n) = \Phi(\frac{\beta + (q-p)\alpha}{2\sqrt{pq}}) + \Phi(\frac{\beta - (q-p)\alpha}{2\sqrt{pq}}) - 1, \tag{2.6}$$

$$q = 1-p, \quad \alpha = bn^{-1/2}, \quad \beta = 2\delta n^{1/2} - n^{-1/2}, \tag{2.7}$$

and Φ is defined in (1.5). One can show using Petrov's [12, p. 125] Theorem 14 that the error term $0(n^{-1/2})$ in (2.5) is, in fact, uniform in $p \in (0,1)$. In view of the order of approximation in (2.5) and the fact that $\Phi(A+n^{-1}) = \Phi(A) + 0(n^{-1})$, we will feel free to drop terms of order $\delta n^{-1/2}$ from the argument of Φ. Since b is assumed fixed, we may assume that $b < \min[n^{1/2}, (2n\delta - 1)3^{-1/2}]$ in which case (2.7) shows that

$$0 \le \alpha < 1 \quad \text{and} \quad \beta > \{4 - \alpha^2 - 4(1-\alpha^2)^{1/2}\}^{1/2}. \tag{2.8}$$

It follows from (2.6), (2.8) and Lemma A(i) in the Appendix that

$$\inf_{0<p<1} \tilde{\gamma}_b(p,n) = \tilde{\gamma}_b(\tfrac{1}{2}, n) = 2\Phi(2\delta n^{1/2} - n^{-1/2}) - 1. \tag{2.9}$$

Hence $\inf_p \tilde{\gamma}_b(p,n) \ge \gamma$ holds for every $\gamma \in (0,1)$ whenever n satisfies $2\delta n^{1/2} - n^{-1/2} \ge z$, that is, for all n satisfying

$$n \ge (z/4\delta)^2\{1 + (1 + 8\delta z^{-2})^{1/2}\}^2 = (z/2\delta)^2 + \delta^{-1} + 0(1), \tag{2.10}$$

which leads to (1.5). Note that, although one can formally expand the middle term in (2.10) beyond δ^{-1}, the final result for N_b cannot be expected to have accuracy beyond δ^{-1} because (2.5) itself has accuracy to order $0(n^{-1/2})$. Note also that, if the correction factor $\frac{1}{2}$ were ignored in (2.4), then the factors $n^{-1/2}$ and δ^{-1} would drop out of (2.9) and (2.10) respectively. This, in turn, implies that the normal approximation without any correction factor would overestimate the true smallest coverage probability and underestimate N_b.

We will now relate $T_b(X_n) \pm \delta$ and N_b to several well-known confidence intervals when the latter are adapted to the present problem. Three *asymptotic* confidence intervals for p are often cited in the vast literature and they are

$$p_n \pm z n^{-1/2} p_n^{1/2}(1-p_n)^{1/2}, \tag{2.11}$$

$$(1+n^{-1}z^2)^{-1}\{p_n+(2n)^{-1}z^2\} \pm (2n)^{-1}(1+n^{-1}z^2)^{-1} z\{4np_n(1-p_n)+z^2\}^{1/2}, \tag{2.12}$$

$$\sin^2(\sin^{-1} p_n^{1/2} \pm \tfrac{1}{2} n^{-1/2} z). \tag{2.13}$$

The first one is the same as (1.8). To adapt these intervals to the present problem, note that the midpoints of (2.11) and (2.12) are $T_b(X_n)$ of (1.4) with $b=0$ and $b = z^2$ respectively. Consequently, the common (up to order δ^{-1}) sample size generated by (2.11) and (2.12) is N_b in (1.5). The midpoint of (2.13) can be expressed as

$$p_n \cos(n^{-1/2}z) + \sin^2(\tfrac{1}{2}n^{-1/2}z) = T_b(X_n) + (1+p_n)0(n^{-2}) \quad \text{with} \quad b = \tfrac{1}{2}z^2. \tag{2.14}$$

The term $0(n^{-2})$ comes from the Taylor expansion of sine-cosine and does not contain p_n. It is easily verified from (1.4) and (2.4) that, for *any* fixed $b \geq 0$, the coverage probability of $T_b(X_n) + (1+p_n)0(n^{-2}) \pm \delta$ remains the same as (2.5). Hence (2.14) leads to the same sample size as in (1.5). The interval (2.13) is based on the well-known fact that $(\sin^{-1} p_n^{1/2} - \sin^{-1} p^{1/2})(4n)^{1/2} \to N(0,1)$ as $n \to \infty$. If one uses the more refined approximation (see [10, p. 65])

$$[\sin^{-1}\{(p_n + \tfrac{3}{8}n^{-1})^{1/2}(1+\tfrac{3}{4}n^{-1})^{-1/2}\} - \sin^{-1} p^{1/2}](4n+2)^{1/2} \to N(0,1), \tag{2.15}$$

one can develop a new confidence interval similar to (2.13). It is easily shown as above that the new interval leads to the estimator $T_b(X_n) + (1+p_n)0(n^{-2})$ with $b = \frac{1}{2}z^2 + \frac{3}{4}$ and therefore the same N_b as in (1.5).

One can also analyze the intervals in (2.11)–(2.13) according to the technique described after (1.3). In fact, this technique is actually used in most textbooks for (2.11) as well as by Bickel and Doksum [1, p. 161] for (2.12) and by Ghosh [9, p. 899] for (2.11)–(2.13). They all lead to $N_b = (z/2\delta)^2 + 0(\delta^{-1})$ because (2.11)–(2.13) do not incorporate any continuity correction (the factor $-z^2$ arrived at by Bickel and Doksum and by Ghosh from (2.12) for the 0(1) term in N_b is too ambitious because their coverage probability is valid only up to order $n^{-1/2}$). Blyth and Still [3] recommended continuity corrections for (2.11) and (2.12), which are their (3.3) and (2.1), along the arguments underlying (2.4). If one applies the technique described after (1.3) to their corrected intervals (3.4) and (2.4), one gets the same $T_b(X_n)$ and N_b up to order δ^{-1} as in the preceding paragraph.

There is a second family of confidence intervals for p, the so-called *exact* intervals, which are usually recommended for small values of n (e.g., [6, 7, 3, 5]). To describe the non-randomized versions of these intervals, let $\tau \in [0, 1-\gamma]$ be a suitably chosen number. Given any $n \geq 1$ and the observed value of x of X_n, one can show that the solutions $L_n = L_n(x,\tau)$ and $U_n = U_n(x,\tau)$ of the following equations

$$\sum_{k=x}^{n} \binom{n}{k} L_n^k (1-L_n)^{n-k} = 1-\gamma-\tau, \quad \sum_{k=0}^{x} \binom{n}{k} U_n^k (1-U_n)^{n-k} = \tau \tag{2.16}$$

are unique (defining $L_n(0,\tau)=0$, $U_n(n,\tau)=1$) *and* satisfy $P_p(L_n \le p \le U_n) \ge \gamma$ for every $p \in (0,1)$. Thus, $[L_n, U_n]$ constitutes a family (as τ varies) of $100\gamma\%$ confidence intervals for p and there are charts for L_n and U_n for selected values of γ, τ, n and $x \in \{0,1,...,n\}$(e.g., [11, 3, 5]). Clopper and Pearson [6] chose $\tau = \frac{1}{2}(1-\gamma)$ for convenience while other authors have suggested different choices for τ to satisfy some additional criteria (e.g., to reduce the length $U_n - L_n$). Now, if one wants to adapt $[L_n, U_n]$ for a *given* τ to our problem as explained after (1.3), one must first find $L_n(x,\tau)$ and $U_n(x,\tau)$ from (2.16) for each $n \ge 1$ and each $x \in \{0,1,...,n\}$, then verify if $U_n(x,\tau) - L_n(x,\tau) \le 2\delta$ under the specified δ, and finally determine the smallest $N(\tau)$ such that $U_n(x,\tau) - L_n(x,\tau) \le 2\delta$ for every $n \ge N(\tau)$. Evidently, such a numerical procedure to generate $T(X_n,\tau) = \frac{1}{2}\{L_n(X_n,\tau) + U_n(X_n,\tau)\}$ and $N(\tau)$ as a solution to (1.1)–(1.2) is quite daunting, especially because $N(\tau)$ will be large when $\gamma \ge .9$ and $\delta \le .1$. Moreover, the subsequent natural problem of minimizing $N(\tau)$ with respect to $\tau \in [0, 1-\gamma]$ would seem to be hopeless. Nevertheless, we will now show that, among all intervals generated by (2.16) for $0 \le \tau \le 1-\gamma$, the one that has *asymptotically* the smallest $N(\tau)$ is precisely (2.12).

The normal approximation to the binomial sums in (2.16), without any correction for continuity, leads to

$$\begin{aligned} \Phi(n^{1/2}(p_n - \ell)\ell^{-1/2}(1-\ell)^{-1/2}) &= \gamma + \tau + 0(n^{-1/2}) \quad \text{for} \quad \ell = L_n \\ &= \tau + 0(n^{-1/2}) \quad \text{for} \quad \ell = U_n. \end{aligned} \tag{2.17}$$

Define z_0 and z_1 by $\Phi(z_0) = \tau$ and $\Phi(z_1) = \gamma + \tau$, so that $-\infty < z_0 < z_1 < \infty$ for all $0 \le \tau < \gamma + \tau \le 1$. Then the solutions of (2.17), neglecting $0(n^{-1/2})$, are uniquely given by

$$\begin{aligned} L_n &= (1 + n^{-1}z_1^2)^{-1}[p_n + (2n)^{-1}z_1^2 - (2n)^{-1}z_1\{4np_n(1-p_n) + z_1^2\}^{1/2}], \\ U_n &= (1 + n^{-1}z_0^2)^{-1}[p_n + (2n)^{-1}z_0^2 - (2n)^{-1}z_0\{4np_n(1-p_n) + z_0^2\}^{1/2}], \end{aligned} \tag{2.18}$$

If $\tau = \frac{1}{2}(1-\gamma)$, then $z_1 = -z_0 = z$ and (2.18) shows that the interval $[L_n, U_n]$ is identical to (2.12). If $\tau \ne \frac{1}{2}(1-\gamma)$, then $z_0 + z_1 \ne 0$ and (2.18) yields

$$\tfrac{1}{2}(L_n + U_n) = T_b(X_n) - \tfrac{1}{2}n^{-1/2}(z_0 + z_1)\{p_n(1-p_n)\}^{1/2} + 0_p(n^{-3/2}),$$

where $T_b(X_n)$ is as in (1.4) with $b = \frac{1}{2}(z_0^2 + z_1^2)$ and the error term $0_p(n^{-3/2})$ involves p_n in such a way that $n^{3/2}0_p(n^{-3/2})$ approaches some finite $c(p)$ with probability one as $n \to \infty$. Using (2.4) and the fact that $p_n(1-p_n) \to p(1-p)$ with probability one, one finally gets

$$\begin{aligned} P_p(|\tfrac{1}{2}(L_n + U_n) - p| \le \delta) &= \Phi(\frac{\beta + (q-p)\alpha}{2\sqrt{pq}} - \tfrac{1}{2}(z_0 + z_1)) \\ &+ \Phi(\frac{\beta - (q-p)\alpha}{2\sqrt{pq}} + \tfrac{1}{2}(z_0 + z_1)) - 1 + 0(n^{-1/2}), \end{aligned}$$

where α and β are the same as in (2.7). Ignoring $0(n^{-1/2})$ as in all earlier cases, we conclude that, for any $\tau \neq \frac{1}{2}(1-\gamma)$,

$$\inf_{0<p<1} P_p(|\tfrac{1}{2}(L_n + U_n) - p| \leq \delta) \leq P_{1/2}(|\tfrac{1}{2}(L_n + U_n) - \tfrac{1}{2}| \leq \delta)$$
$$= \Phi(\beta - \tfrac{1}{2}(z_0 + z_1)) + \Phi(\beta + \tfrac{1}{2}(z_0 + z_1)) - 1$$
$$< 2\Phi(\beta) - 1 = \inf_{0<p<1} P_p(|T_{z^2}(X_n) - p| \leq \delta).$$

In the last step above, we have used the fact that $\Phi(\beta + A) + \Phi(\beta - A) < 2\Phi(\beta)$ for all $\beta > 0$ and $A \neq 0$. Consequently, $N(\tau) \geq N_{z^2} = N(\frac{1}{2}(1-\gamma))$ for all $\tau \in [0, 1-\gamma]$ so that $\tau = \frac{1}{2}(1-\gamma)$ is asymptotically the best choice in (2.16) for our purpose.

We conclude this section with some numerical comparisons among $T_b(X_n) \pm \delta$ for $b = 0, \frac{1}{2}z^2, z^2$, which were obtained from the confidence intervals in (2.11)–(2.13). Table 1 shows the exact values of the smallest coverage probabilities under certain combinations of (γ, δ, n) and these are computed using (2.1) and the formulas given in Section 4 (the last column in Table 1 arises in Section 3). The asymptotic result in (2.9) suggests that the values for all three cases should approximately equal γ. Given any pair (γ, n), we have chosen the value of δ to satisfy

$$\delta = (2n)^{-1}\{(1 + nz^2)^{1/2} + 1\}. \tag{2.19}$$

The reason for such a choice is that (1.5) implies that, if one uses $T_b(X_n)$ of (1.4) and δ of (2.19), then N_b satisfying (1.1) would be approximately the same as the n shown in Table 1. It is obvious from the table that the three cases in (2.11)–(2.13) are remarkably alike when adapted to the present problem, although their relative merits from the standpoint of confidence intervals are somewhat different. See, for example, [9, 3, 4]. We point out here that the case $b = \frac{1}{2}z^2 + \frac{3}{4}$, cited after (2.15), can be expected to show values of $\inf_p \gamma_b(p, n)$ somewhere between those of $b = \frac{1}{2}z^2$ and $b = z^2$ because $\frac{1}{2}z^2 < \frac{1}{2}z^2 + \frac{3}{4} < z^2$ for $.9 \leq \gamma \leq .99$.

Table 2 shows the exact and approximate values of N_0 and N_{z^2} for some commonly used values of (γ, δ). The exact ones are computed using (2.1) and the formulas in Section 4. Their common approximate value comes from (1.5) and the table shows this with and without the correction factor δ^{-1}. The general conclusions are quite similar to those for Table 1. Note from Table 2 that the approximation without the correction factor considerably underestimates the exact values of both N_0 and N_{z^2}. An augmented table for N_0 is given in the next section.

3. The Proposed $(T^*(X_n), N^*)$. An intrinsic and, indeed, desirable feature of any conventional confidence interval for p is that, for any fixed n, the lower and upper confidence limits approach each other as p_n approaches 0 or 1. This is evident in all intervals considered in the preceding section. However, this aspect constitutes a *drawback* when the same interval is adapted to ensure (1.1)–(1.2) because the shrinking length is achieved at the expense of widening the length around $p_n = \frac{1}{2}$ and, as noted after (1.3), it is the latter (compared with 2δ) that determines how large N_T will be. It seems, therefore, logical that in our search for a better solution

Table 1: A Comparison of the Exact Values of $\inf_p \gamma_b(p,n)$

n	γ	δ	$b=0$	$b=\frac{1}{2}z^2$	$b=z^2$	$b=a_\gamma\sqrt{n}$
100	.90	.0873945	.91094	.91125	.91137	.93274
100	.95	.1031257	.95489	.96386	.96441	.96970
100	.99	.1338885	.99100	.99332	.99501	.99528
500	.90	.0377936	.90210	.90211	.90211	.91598
500	.95	.0448375	.95100	.95573	.95579	.95983
500	.99	.0586060	.99057	.99058	.99171	.99299
1000	.90	.0265122	.90616	.90618	.90620	.91127
1000	.95	.0314938	.95013	.95362	.95364	.95695
1000	.99	.0412305	.99051	.99052	.99132	.99229
2000	.90	.0186417	.90202	.90202	.90203	.90862
2000	.95	.0221645	.95092	.95093	.95093	.95536
2000	.99	.0290497	.99051	.99052	.99052	.99159

Table 2: Some Values of N_0 *and* N_{z^2}

γ	δ	Exact N_0	Exact N_{z^2}	Approximate $b=(z/2\delta)^2$	N_b $b=(z/2\delta)^2+\delta^{-1}$
	.01	6850	6848	6764	6864
.90	.03	784	764	752	785
	.05	280	278	271	291
	.10	75	68	68	78
	.01	9700	9647	9604	9704
.95	.03	1084	1080	1068	1101
	.05	400	387	385	405
	.10	100	97	97	107
	.01	16650	16644	16588	16688
.99	.03	1867	1861	1844	1877
	.05	680	664	664	684
	.10	170	159	166	176

than (1.4)–(1.5) we should examine the family of *shrinkage* estimators

$$p_n(a) = (X_n + \tfrac{1}{2}a\sqrt{n})(n + a\sqrt{n})^{-1} = p_n + a(\tfrac{1}{2} - p_n)(\sqrt{n} + a)^{-1} \tag{3.1}$$

for constants $a \geq 0$. An attractive feature of this family is that $p_n(a)$ is strongly consistent for each a and the family includes the minimum variance unbiased estimator $p_n(0)$ of p, the minimax estimator $p_n(1)$ under quadratic loss, and the Bayes estimator $p_n(a)$ under beta-prior with parameter $\frac{1}{2}an^{1/2}$ (see [8, p. 93]). Denote the coverage probability of $p_n(a) \pm \delta$ by

$$\gamma^{(a)}(p,n) = P_p(|p_n(a) - p| \leq \delta) = \gamma_{a\sqrt{n}}(p,n) \quad \text{of (2.1).} \tag{3.2}$$

Denote the (smallest) sample size of $p_n(a) \pm \delta$ to satisfy (1.1) by $N^{(a)}$. Then, using the technique underlying (2.3), it is readily shown that, analogous to (2.2), we now have

$$a^2(1-2\delta)^2(2\delta)^{-2} \leq N^{(a)} \leq [A_a(\gamma,\delta)], \tag{3.3}$$

where

$$\begin{aligned} A_a(\gamma,\delta) &= (\omega - a)^2 \quad &\text{if } 0 \leq a \leq 1 \\ &= a^2(\omega-1)^2 \quad &\text{if } a \geq 1. \end{aligned}$$

The inequalities in (3.3) obviously guarantee the existence of the exact $N^{(a)}$ for every $a \geq 0$. Moreover, comparing the upper bound of $N^{(1)}$ with the lower bound of $N^{(a)}$ we conclude that

$$N^{(1)} < N^{(a)} \quad \text{whenever} \quad a > (1-2\delta)^{-1}\{(1-\gamma)^{-1/2} - 2\delta\}$$

which shows that a "good" choice of a cannot be "too large".

In order to avoid any confusion among different notations in this section and Section 1 observe from (1.9), (1.11), (3.1) and (3.2) that the following correspondences hold throughout

$$(T^*(X_n), N^*) \equiv (p_n(a_\gamma), N^{(a_\gamma)}) \quad \text{and} \quad (p_n, N_0) \equiv (p_n(0), N^{(0)}), \tag{3.4}$$

where a_γ is defined in (1.10). We will now prove the asymptotic result that the pair $(T^*(X_n), N^*)$ satisfies (1.1)–(1.2) for small δ. To this end, we need the following approximation, which follows from (3.2), (2.1) and (2.4),

$$\gamma^{(a)}(p,n) = \tilde{\gamma}^{(a)}(p,n) + 0(n^{-1/2}), \tag{3.5}$$

where

$$\tilde{\gamma}^{(a)}(p,n) = \Phi(\frac{\beta' + (q-p)a}{2\sqrt{pq}}) + \Phi(\frac{\beta' - (q-p)a}{2\sqrt{pq}}) - 1, \tag{3.6}$$

and

$$\beta' = 2\delta(n^{1/2} + a) - n^{-1/2}.$$

Note that $\tilde{\gamma}^{(a)}$ is symmetric about $p = \frac{1}{2}$ for all $a \geq 0$, $n > 0$, $\delta \in (0, \frac{1}{2})$ and is increasing in $n > 0$ for all $a \geq 0$, $p \in (0,1)$, $\delta \in (0, \frac{1}{2})$. If $\tilde{N}^{(a)}$ denotes the smallest number satisfying $\inf_p \tilde{\gamma}^{(a)}(p,n) \geq \gamma$ for all $n \geq \tilde{N}^{(a)}$, then $\tilde{N}^{(a)}$ is obviously the approximation we are seeking for $N^{(a)}$ when δ is small. For notational simplicity, we will use $N^{(a)}$ also for $\tilde{N}^{(a)}$ below with the understanding that the results are valid for $N^{(a)}$ only up to order δ^{-1}.

Consider first the case $\gamma \geq \gamma_0$, where γ_0 is defined in (1.10). Then $z \geq \sqrt{3}$. Choose $a = 1$ and $\beta' = z$ in (3.6), which implies that

$$n = f_1^2(z) = \{(z/2\delta) - 1\}^2 + \delta^{-1} + 0(1), \tag{3.7}$$

where

$$f_a(z) = \tfrac{1}{2}\{(z/2\delta) - a\} + [\tfrac{1}{4}\{(z/2\delta) - a\}^2 + (2\delta)^{-1}]^{1/2} \quad \text{for} \quad a \geq 0. \tag{3.8}$$

Since $a = 1$ and $\beta' \geq \sqrt{3}$ in (3.6), it follows from Lemma A(i) in the Appendix that $\tilde{\gamma}^{(1)}$ decreases from $\tilde{\gamma}^{(1)}(0, n) = 1$ to $\tilde{\gamma}^{(1)}(\frac{1}{2}, n) = 2\Phi(z) - 1 = \gamma$. Since $\tilde{\gamma}^{(1)}$ is increasing in n, we conclude that $\inf_p \tilde{\gamma}^{(1)}(p, n) \geq \gamma$ holds for all $n \geq N^{(1)}$ if and only if $\tilde{\gamma}^{(1)}(\frac{1}{2}, N^{(1)}) = \gamma$. The solution $N^{(1)}$ (i.e. N^*) of the latter is precisely the right-hand side of (3.7). Suppose next that $0 < \gamma < \gamma_0$. Choose

$$a = \{4(1 + z^2)^{1/2} - 4 - z^2\}^{1/2} \quad \text{and} \quad \beta' = z \tag{3.9}$$

in (3.6). Then it is easy to check that $0 < a < 1$ and $\beta' \geq \{4 - a^2 - 4(1 - a^2)^{1/2}\}^{1/2}$. Using Lemma A(i) as above one finds

$$N^{(a)} = f_a^2(z) = \{(z/2\delta) - a\}^2 + \delta^{-1} + 0(1). \tag{3.10}$$

This proves that $N^{(a_\gamma)}$ (i.e. N^*) is given by (1.11) when $\frac{1}{2} < \gamma < \gamma_0$. Although the pair $(p_n(a), N^{(a)})$, with a as in (3.9), satisfies (1.1)–(1.2) also for $0 < \gamma \leq \frac{1}{2}$, it is different from the pair $(T^*(X_n), N^*)$ defined in (1.9)–(1.11) and we proposed the latter on the following ground. For the case $\gamma \leq \frac{1}{2}$, one can choose $a = \beta' = z < 0.675$ in (3.6) and conclude from Lemma A(iii) that

$$N^{(z)} = f_z^2(z) = \{(z/2\delta) - z\}^2 + \delta^{-1} + 0(1) \tag{3.11}$$

which is, in fact, $N^*$of (1.11). It is easily verified that

$$\{4(1 + z^2)^{1/2} - 4 - z^2\}^{1/2} < z \quad \text{for all} \quad \gamma \in (0, \tfrac{1}{2}].$$

Consequently, (3.10) and (3.11) imply that $N^{(a)} > N^{(z)}$ for every a under (3.9) when $0 < \gamma \leq \frac{1}{2}$. Thus, the pair $(p_n(z), N^{(z)})$ is better than $(p_n(a), N^{(a)})$ under (3.9) for any $\gamma \leq \frac{1}{2}$ and this also explains the discontinuity in a_γ of (1.10) at $\gamma = \frac{1}{2}$.

One final question now remains: What is the asymptotically optimum choice of a in (3.1) that minimizes $N^{(a)}$ with respect to $a \geq 0$ for a given pair (γ, δ)? A general answer for the case $\gamma > \frac{1}{2}$ is difficult because $\min_a N^{(a)}$ will occur at some $a(\gamma, \delta)$ which depends on γ and δ in an intricate way (see Lemma A(ii) and the comments after the lemma). We provide instead a partial answer below, which also gives a justification for the special choice of a in (1.10).

Given any $\gamma \in (0, 1)$, $\delta \in (0, \frac{1}{2})$ and $a \geq 0$, the requirements $\tilde{\gamma}^{(a)}(0, N^{(a)}) \geq \gamma$ and $\tilde{\gamma}^{(a)}(\frac{1}{2}, N^{(a)}) \geq \gamma$ for (3.6) imply that $N^{(a)}$ must satisfy

$$\sqrt{N}^{(a)} \geq \max[f_a(a), f_a(z)], \tag{3.12}$$

where $f_a(z)$is defined in (3.8). Now, if $0 < \gamma \leq \frac{1}{2}$, it follows from (3.11) and (3.12) that

$$N^{(z)} < f_a^2(a) \text{ for every } a > z, \; N^{(z)} < f_a^2(z) \text{ for every } a < z.$$

Consequently, $\min_a N^{(a)} = N^{(z)}$ (i.e. N^*) and $a = a_\gamma$ is, indeed, the asymptotically optimum choice for a in (3.1) when $\gamma \le \frac{1}{2}$. On the other hand, if $\frac{1}{2} < \gamma < 1$, it follows from (3.7),(3.10) and (3.12) that

$$N^{(a_\gamma)} < N^{(a)} \quad \text{for} \quad a < a_\gamma \quad \text{or} \quad a > (z - 2\delta a_\gamma)(1-2\delta)^{-1},$$

and the possibility remains that $N^{(a)}$ may attain a minimum at some $a(\gamma,\delta)$ between a_γ and $(z-2\delta a_\gamma)(1-2\delta)^{-1}$. However, there are reasons to hope that in such cases both $a(\gamma,\delta) - a_\gamma$ and $N^{(a_\gamma)} - N^{(a)}$ at $a = a(\gamma,\delta)$ will be negligibly small. For instance, if $\gamma = .95$ and $\delta = .03$, then $a_\gamma = 1$ and a further numerical investigation of the equation in Lemma A(ii) shows that $a(.95,.03) = 1.0050$. Moreover, the exact values of $N^{(a)}$ at $a = 1$ and $a = 1.0050$ are both 1035, while their approximate counterparts under (3.5) are both 1037.

To summarize, the proposed $T^*(X_n)$ is defined in (1.9)-(1.10) and, for any p, the exact coverage probability of $T^*(X_n) \pm \delta$ is $\gamma^*(p,n)$, which is identical to (3.2) when $a = a_\gamma$; N^* is the smallest n for which $\inf_p \gamma^*(p,n) \le \gamma$ for all $n \ge N^*$ and a good approximation for N^* is provided in (1.11). The last column in Table 1 shows some exact values of $\inf_p \gamma^*(p,n)$ and they are based on the formulas in Section 4. Table 3 shows a detailed comparison between the exact and approximate values N_0 of $p_n \pm \delta$ and N^*. The exact ones are computed using (2.1) with $b = 0$, (3.2) with $a = a_\gamma$ and the formulas in Section 4. Note from (1.10) that $a_\gamma = 1$ for $\gamma = 0.95$ or 0.99 and $a_\gamma = 0.99718194$ for $\gamma = 0.9$ in Table 3. As noted in Table 2, N_0 will not differ much from N_b for popular choices of $b > 0$. The general conclusions one can draw from Table 3 have been summarized after (1.12).

4. Exact Coverage Probabilities and Sample Sizes. In this section, we describe the method of calculating the exact values of $\gamma_T(p,n)$ and N_T when $T(X_n)$ is of the form (1.4), (1.9) or (3.1). We may note here that there are several papers (e.g., [3, 4]) which show computer-plots of $\gamma_T(p,n)$. Such graphs are inadequate for determining the exact value of N_T that guarantees $\inf_p \gamma_T(p,n) \ge \gamma$ for all $n \ge N_T$. The p_k below are constants and should not be confused with p_n of (1.6).

Recall $\gamma_b(p,n)$ of (2.1) but assume that, for present purposes, $b \ge 0$ may depend on $n \ge 1$. Define

$$c = \text{largest integer contained in } n(\tfrac{1}{2}+\delta) + b\delta. \tag{4.1}$$

Then $n(\frac{1}{2}+\delta) + b\delta - 1 < c \le n(\frac{1}{2}+\delta) + b\delta$ by definition and it can be checked that $\gamma_b(\frac{1}{2},n) > 0$ if and only if $c \ge \frac{1}{2}n$, while $c < \frac{1}{2}n$ holds if and only if n is odd and $n + b < (2\delta)^{-1}$. For integers $k \ge 0$, let

$$p_k = \tfrac{1}{2} - \{c - n(\tfrac{1}{2}+\delta) - b\delta + k + 1\}(n+b)^{-1}, p'_k = 1 - p_k - (2k+1)(n+b)^{-1} \tag{4.2}$$

which imply that

$$\begin{aligned} &p_k - p_{k+1} = p'_k - p'_{k+1} = (n+b)^{-1} > p_k - p'_k \ge -(n+b)^{-1}, \\ &\text{and } \; p'_k \le p_k \quad \text{if and only if } \; c \le n(\tfrac{1}{2}+\delta) + b\delta - \tfrac{1}{2}. \end{aligned} \tag{4.3}$$

Table 3: Exact and Approximate Values of N_0 and N^*

γ	δ	Exact N_0	Approximate N_0 $\left(\frac{z}{2\delta}\right)^2 + \frac{1}{\delta}$	Exact N^*	Approximate N^* $\left(\frac{z}{2\delta} - a_\gamma\right)^2 + \frac{1}{\delta}$	Exact $100\times\left(1 - \frac{N^*}{N_0}\right)$
	.01	6850	6864	6669	6701	2.6
	.02	1725	1741	1656	1660	4.0
	.03	784	785	725	732	7.5
	.04	438	448	405	408	7.5
.90	.05	280	291	256	259	8.6
	.06	200	205	177	179	11.5
	.07	150	153	129	130	14.0
	.08	113	119	98	99	13.3
	.09	89	95	77	78	13.5
	.10	75	78	62	63	17.3
	.01	9700	9704	9503	9509	2.0
	.02	2425	2451	2352	2354	3.0
	.03	1084	1101	1035	1037	4.5
	.04	613	626	576	578	6.0
.95	.05	400	405	363	366	9.3
	.06	275	284	251	252	8.7
	.07	204	211	182	184	10.8
	.08	157	163	139	140	11.5
	.09	128	130	109	109	14.8
	.10	100	107	87	88	13.0
	.01	16650	16688	16422	16431	1.4
	.02	4175	4197	4062	4070	2.7
	.03	1867	1877	1792	1792	4.0
	.04	1050	1062	994	999	5.3
.99	.05	680	684	634	633	6.8
	.06	475	478	436	436	8.2
	.07	350	353	318	318	9.1
	.08	269	272	241	241	10.4
	.09	212	216	189	189	10.8
	.10	170	176	153	152	10.0

Then the right-hand side of (2.1) is equivalent to: if $c \le n(\frac{1}{2} + \delta) + b\delta - \frac{1}{2}$ then

$$\begin{aligned} \gamma_b(p,n) &= P_p(n - c - k \le X_n \le c - k) \quad \text{for} \quad p_k < p < p'_k + (n+b)^{-1} \\ &= P_p(n - c - k - 1 \le X_n \le c - k) \quad \text{for} \quad p'_k \le p \le p_k, \end{aligned} \tag{4.4}$$

and if $c > n(\frac{1}{2} + \delta) + b\delta - \frac{1}{2}$ then

$$\begin{aligned} \gamma_b(p,n) &= P_p(n-c-k \le X_n \le c-k) \text{ for } p_k' \le p \le p_k + (n+b)^{-1} \\ &= P_p(n-c-k \le X_n \le c-k-1) \text{ for } p_k < p < p_k'. \end{aligned} \tag{4.5}$$

Observe that, since $p_0 + p_0' = 1-(n+b)^{-1}$ by (4.2), the point $p = \frac{1}{2}$ is contained in the interval $(p_0, 1-p_0)$ under (4.4) and in $[p_0', 1-p_0']$ under (4.5). Since $\gamma_b(p,n) = \gamma_b(1-p,n)$, it suffices to compute $\gamma_b(p,n)$ only for $p \le \frac{1}{2}$. The largest value, m say, of k one needs to consider in the computation of $\gamma_b(p,n)$ for any $p \le \frac{1}{2}$ is determined by $p_{m+1} < 0 \le p_m$ or $p_{m+1}' < 0 \le p_m'$, whichever occurs later.

It is easy to see that γ_b is discontinuous in p at the end-points of the intervals in (4.4) and (4.5). Since the minimum of $P_p(A \le X_n \le B)$, for fixed $A \le B$, with respect to $p \in [s,t]$ occurs at $p = s$ or t, it follows that a minimum of γ_b with respect to $p \in (0,1)$ does not exist when $c \ge \frac{1}{2}n$. On the other hand, it is easily shown from (4.4) and (4.5), respectively that

$$\begin{aligned} \inf_{0<p<1} \gamma_b(p,n) &= \min_{k\ge 0}\{F_k(p_k) \wedge F_k(p_k' + (n+b)^{-1})\} \text{ if } c \le n(\tfrac{1}{2}+\delta) + b\delta - \tfrac{1}{2} \\ &= \min_{k\ge 0}\{G_k(p_k) \wedge G_k(p_k')\} \text{ if } c > n(\tfrac{1}{2}+\delta) + b\delta - \tfrac{1}{2}, \end{aligned} \tag{4.6}$$

where, for $0 < p < 1$,

$$F_k(p) = P_p(n-c-k \le X_n \le c-k), \; G_k(p) = P_p(n-c-k \le X_n \le c-k-1).$$

Given any $\delta \in (0, \frac{1}{2})$, $n \ge 1$ and $b \ge 0$, we compute the exact value of $\inf_p \gamma_b(p,n)$ as follows. First find c from (4.1) and then determine p_k and p_k' from (4.2) for $k = 0, 1, ..., m$. If $c \le n(\frac{1}{2}+\delta)+b\delta-\frac{1}{2}$, one finds the smaller one, S_k say, of $F_k(p_k)$ and $F_k(p_k'+(n+b)^{-1})$ for every k. Then (4.6) states that $\inf_p \gamma_b(p,n) = \min\{S_0, ..., S_m\}$. Similarly, one uses the second expression in (4.6) when $c > n(\frac{1}{2}+\delta)+b\delta-\frac{1}{2}$. Table 1 shows some of these values when $b = 0$, $\frac{1}{2}z^2$, $z^2, a_\gamma\sqrt{n}$.

To determine the exact value of N_b for a given γ, one repeats the process in the preceding paragraph for successive values of n and verifies whether $\inf_p \gamma_b(p,n) \ge \gamma$. The verification may entail switching back and forth between the two expressions in (4.6). When b does not involve n, a reasonable starting value of n for this purpose is (1.4) without δ^{-1}; when $b = a_\gamma n^{1/2}$, a reasonable starting value is (1.11) without δ^{-1}. If the search leads to an n_0 satisfying $\inf_p \gamma_b(p,n_0) \ge \gamma$, it does *not* mean that $N_b = n_0$ because $\inf_p \gamma_b(p,n)$ is not monotonic in n. One needs to check whether $\inf_p \gamma_b(p,n) \ge \gamma$ also for $n = n_0+1, ..., n_1$, where n_1 is typically not too far away from n_0. This process finally generates the smallest N_b for which $\inf_p \gamma_b(p,n) \ge \gamma$ for *all* $n \ge N_b$. Table 2 shows values of N_b when $b = 0$, z^2 and Table 3 shows a few more when $b = 0$, $a_\gamma\sqrt{n}$.

To conclude, we may point out an interesting feature of γ_b. Suppose δ, n and b are such that $\gamma_b(\frac{1}{2},n) > 0$, which is true in particular if $n+b \ge (2\delta)^{-1}$. It is then easily shown from (4.4) and (4.5) that there exists a positive $\Delta = \Delta(\delta,n,b)$ such that

$$\max_{\frac{1}{2}-\Delta \le p \le \frac{1}{2}+\Delta} \gamma_b(p,n) = \gamma_b(\tfrac{1}{2},n).$$

On the other hand, it was seen in Sections 2–3 that for many values of $b \geq 0$ the normal approximation $\tilde{\gamma}_b(p,n)$ indicates that $\tilde{\gamma}_b(\frac{1}{2},n)$ is the absolute *minimum* of $\tilde{\gamma}_b(p,n)$. This anomaly disappears as $n \to \infty$ or $\delta \to 0$ for then $\Delta \to 0$.

5. Some Examples. We will now clarify certain aspects of our results by some numerical examples for the case $\gamma = .95$ and $\delta = .03$. Note that this special combination of (γ,δ) is used quite often in opinion polls, where p represents the unknown proportion of a populace "favoring" a certain product or statement.

(a) For any $b \in [0,z^2] = [0,3.84]$, the exact (minimum) sample size needed by $T_b(X_n) \pm \delta$ is N_b and it satisfies $N_{z^2} = 1080 \leq N_b \leq N_0 = 1084$. The range $[0,z^2]$ actually covers the three well-known choices $b = 0, \frac{1}{2}z^2, z^2$ (see [1, p. 161]; [9, p. 899]; [3, pp. 112, 114]). The approximation in (1.5) yields

$$N_b \approx 1068 \text{ without the correction factor } \delta^{-1}$$
$$\approx 1101 \text{ with the correction factor.}$$

The case $T_0(X_n)$ with $N_0 \approx 1068$ corresponds to (1.7), which is cited in most textbooks and, we suspect from published figures, also used by most pollsters. Since $\inf_p \gamma_0(p,1068) = .949773$ is quite close to $\gamma = .95$, one may be tempted to feel comfortable with $N_0 \approx 1068$. However, the insidious aspect of this approximation is that the coverage probability may drop even more for some $n > 1068$. In fact,

$$\min_{1068 \leq n \leq 1084} \inf_p \gamma_0(p,n) = \inf_p \gamma_0(p,1083) = .948246.$$

Consequently, if one must use the estimator $T_0(X_n)$ at all, one should fall back on $N_0 \approx 1101$ when the exact N_0 is unavailable. It maybe pointed out here that $\inf_p \gamma_0(p,1084) = .951600$ and $\inf_p \gamma_0(p,1101) = .953268$.

(b) The exact (minimum) sample size needed by the proposed $T^*(X_n) \pm \delta$ is $N^* = 1035$ and the approximation in (1.11) yields

$$N^* \approx 1003 \text{ without the correction factor } \delta^{-1}$$
$$\approx 1037 \text{ with the correction factor.}$$

Clearly, the approximation with the correction factor is remarkably accurate and Table 3 exhibits the same feature for other combinations of (γ,δ) as well. The table also shows that, under any fixed γ, a slight change in δ may drastically alter N^*, exact or approximate.

(c) $N_0 = 1084$ and $N^* = 1035$ show that one actually saves 49 observations by using $T^*(X_n) \pm \delta$ for $n = 1035$ instead of $T_0(X_n) \pm \delta$ for $n = 1084$. This saving constitutes 4.5% of N_0 and this actual percentage is slightly overestimated to 5.8% by the approximate counterparts $N_0 \approx 1101$and $N^* \approx 1037$. Table 3 (last column) shows that the actual percentage saving can increase substantially as δ increases under the same γ.

(d) Suppose one uses $n = N^* = 1035$ for both $T_b(X_n) \pm \delta$ under some $b \in [0, z^2]$ and $T^*(X_n) \pm \delta$. In practice, $T_b(X_n)$ and $T^*(X_n)$ are often rounded off to two decimal digits. It can be verified from (1.3) and (1.9) that, to two decimal digits,

$$T_b(X_{1035}) \equiv T^*(X_{1035}) \quad \text{whenever} \quad 316 \le X_{1035} \le 636.$$

This means that the reported values of $T_b(X_n) \pm \delta$ and $T^*(X_n) \pm \delta$ may turn out to be identical for a wide range of observed values of X_n around $\frac{1}{2}n$. However, since $N_b \ge 1080$ for $b \in [0, 3.84]$, one ought to keep in mind that it is the proposed $T^*(X_n) \pm \delta$ that ensures the confidence level $\gamma = .95$. In fact, exact calculations show that $\inf_p \gamma_b(p, 1035) = .946001 < .95$ when $b = 0$.

(e) Finally, suppose we observe X_{n_0} for an arbitrary n_0 and compute $T^*(X_{n_0})$ for a given γ. Then (1.11) implies that the margin of error δ in the resulting $T^*(X_{n_0})$ satisfies $n_0 \approx \{(z/2\delta) - a_\gamma\}^2 + \delta^{-1}$, which leads to

$$\text{(5.1)} \qquad \delta \approx \tfrac{1}{2}(n_0 - a_\gamma^2)^{-1}[1 - a_\gamma z + \{(1 - a_\gamma z)^2 + (n_0 - a_\gamma^2)z^2\}^{1/2}].$$

Results of opinion polls often announce the precise values of n_0 and γ but tend to round off the value of δ to two decimal digits. This custom ignores the fact that a slight change in the value of δ can drastically alter the required value of N^* to satisfy (1.1) (see Table 3). Formula (5.1) is then useful in retrieving the precise value of δ when the chosen n_0 is claimed to be N^*. For instance, if $\gamma = .95$, $n_0 = 500$ and one wants to claim $N^* = n_0$, then (5.1) yields the genuine margin of error as $\delta \approx .043$. If, however, .043 is subsequently rounded off to .04, then (1.11) shows that one would actually require $n \ge 578$ to honestly assert that $\gamma = .95$ and $\delta \approx .04$.

6. Appendix. **Lemma A.** Let $\tilde{\gamma}(p)$, $0 < p \le \frac{1}{2}$, denote the function in (2.6) for arbitrary $\beta \ge \alpha \ge 0$.

(i) If $\alpha = 0 < \beta$ or if $0 < \alpha \le 1$ and $\beta \ge \{4 - \alpha^2 - 4(1 - \alpha^2)^{1/2}\}^{1/2}$, then $\tilde{\gamma}$ strictly decreases from $\tilde{\gamma}(0) = 1$ to $\tilde{\gamma}(\frac{1}{2}) = 2\Phi(\beta) - 1$.

(ii) If $\alpha = 1 < \beta < \sqrt{3}$ or $\beta > \alpha > 1$, then $\tilde{\gamma}$ decreases from $\tilde{\gamma}(0) = 1$ to $\tilde{\gamma}(p_0)$ and then increases to $\tilde{\gamma}(\frac{1}{2}) = 2\Phi(\beta) - 1$, where $p_0 = p_0(\alpha, \beta)$ is the unique solution of the equation $\psi(p_0) = 0$ and

$$\text{(1)} \qquad \psi(p) = \ell n\left(\frac{\alpha - \beta(q-p)}{\alpha + \beta(q-p)}\right) + \frac{2\alpha\beta(q-p)}{1 - (q-p)^2}, \quad \tfrac{1}{2}(1 - \alpha\beta^{-1}) < p < \tfrac{1}{2}.$$

(iii) If $\alpha = \beta \ge 1$, then $\tilde{\gamma}$ strictly increases from $\tilde{\gamma}(0) = \frac{1}{2}$ to $\tilde{\gamma}(\frac{1}{2}) = 2\Phi(\alpha) - 1$. If $0 < \alpha = \beta < 1$, then $\tilde{\gamma}$ increases from $\tilde{\gamma}(0) = \frac{1}{2}$ to $\tilde{\gamma}(p_0)$ and then decreases to $\tilde{\gamma}(\frac{1}{2}) = 2\Phi(\alpha) - 1$, $p_0 = p_0(\alpha)$ being defined by $\psi(p_0) = 0$, where $\psi(p)$ is as in (1) with $\beta = \alpha$. **Proof.** Observe first that $\tilde{\gamma}(\frac{1}{2}) = 2\Phi(\beta) - 1$ for all $\beta \ge \alpha \ge 0$, $\tilde{\gamma}(0) = 1$ for $\beta > \alpha \ge 0$, and $\tilde{\gamma}(0) = \frac{1}{2}$ for $\beta = \alpha > 0$. Consider the assertions in (i) and (ii). If $\alpha = 0 < \beta$, it is obvious from (2.6) that $\tilde{\gamma}$ is strictly decreasing. If $\beta > \alpha > 0$, then (2.6) yields

$$\text{(2)} \qquad \tilde{\gamma}'(p) = K(p)\left(\frac{\alpha - \beta(q-p)}{\alpha + \beta(q-p)} - \exp\left(-\frac{2\alpha\beta(q-p)}{1 - (q-p)^2}\right)\right),$$

where

$$K(p) = \frac{\alpha + \beta(q-p)}{4(pq)^{3/2}} \Phi'(\frac{\beta - (q-p)\alpha}{2(pq)^{1/2}}) > 0.$$

Clearly, $\tilde{\gamma}'(p) < 0$ for $0 < p \leq \frac{1}{2}(1 - \alpha\beta^{-1})$ and therefore $\tilde{\gamma}$ is decreasing on this interval. It also follows from (1) and (2) that, for $\frac{1}{2}(1-\alpha\beta^{-1}) < p < \frac{1}{2}$, $\tilde{\gamma}'(p) <$ or > 0 according as $\psi(p) <$ or > 0. Now, (1) shows

$$\psi(\tfrac{1}{2}(1-\alpha\beta^{-1})) = -\infty, \quad \psi(\tfrac{1}{2}) = 0, \quad \psi'(p) = J(p)g(p),$$

where

$$J(p) = \tfrac{1}{4}\alpha\beta(pq)^{-2}\{\alpha^2 - \beta^2(q-p)^2\}^{-1} > 0,$$

$$g(p) = (\beta^2+1)(q-p)^4 + (\beta^2 - \alpha^2 - 2)(q-p)^2 + (1-\alpha^2). \tag{3}$$

If $0 < \alpha \leq 1$ and $\beta \geq (\alpha^2+2)^{1/2}$, then obviously $g(p) > 0$ and therefore $\psi(p) < 0$ for $\frac{1}{2}(1-\alpha\beta^{-1}) < p < \frac{1}{2}$, implying that $\tilde{\gamma}$ is decreasing on the same interval. If $0 < \alpha \leq 1$ and $\{4-\alpha^2-4(1-\alpha^2)^{1/2}\}^{1/2} \leq \beta < (\alpha^2+2)^{1/2}$, then again $g(p) > 0$ and $\tilde{\gamma}$ is decreasing; the lower bound for β guarantees that $g(p) > 0$ for $\frac{1}{2}(1-\alpha\beta^{-1}) < p < \frac{1}{2}$. This proves part (i) of the lemma. Next, suppose $\alpha = 1 < \beta < \sqrt{3}$ or $\beta > \alpha > 1$. In these cases, it is easily verified that $g(p) > 0$ for $\frac{1}{2}(1-\alpha\beta^{-1}) < p < p'$ and $g(p) < 0$ for $p' < p < \frac{1}{2}$, where

$$p' = \tfrac{1}{2} - \tfrac{1}{2}\{2(\beta^2+1)\}^{-1/2}[2-\beta^2+\alpha^2+\{(\beta^2+\alpha^2-4)^2+16(\alpha^2-1)\}^{1/2}]^{1/2}. \tag{4}$$

Consequently, there exists a unique p_0 satisfying $\psi(p_0) = 0$ such that $\psi(p) < 0$ for $\frac{1}{2}(1-\alpha\beta^{-1}) < p < p_0$ and $\psi(p) > 0$ for $p_0 < p < \frac{1}{2}$. Note that p_0 must lie between $\frac{1}{2}(1-\alpha\beta^{-1})$ and p' of (4) and that $p_0(\alpha, \alpha) = 0$ for any $\alpha \geq 1$. Part (ii) of the lemma thus follows. The proof of (iii) is similar; in the case of $0 < \alpha = \beta < 1$, p_0 satisfying $\psi(p_0) = 0$ lies between 0 and $\frac{1}{2} - \frac{1}{2}(1-\alpha^2)^{1/2}(1+\alpha^2)^{-1/2}$.

It can be seen that the only case left unexplored by the lemma is when $0 < \alpha < 1$ and $\alpha < \beta < \{4-\alpha^2-4(1-\alpha^2)^{1/2}\}^{1/2}$. One can show that in this case there exists a unique $\beta(\alpha)$ for any $\alpha \in (0,1)$ such that $\tilde{\gamma}$ behaves differently depending on whether $\beta < \beta(\alpha)$ or $\beta \geq \beta(\alpha)$. For $\beta \geq \beta(\alpha)$, $\tilde{\gamma}(p)$ strictly decreases in $p \in (0, \frac{1}{2})$. For $\beta < \beta(\alpha)$, $\tilde{\gamma}$ decreases from $\tilde{\gamma}(0)$ to $\tilde{\gamma}(p_0)$, then increases to $\tilde{\gamma}(p_1)$, and finally decreases to $\tilde{\gamma}(\frac{1}{2})$, where p_0 and p_1 are two possible solutions of $\psi(p) = 0$. Unfortunately, it is not possible to obtain a simple expression for $\beta(\alpha)$ in terms of α.

REFERENCES

[1] P.J. Bickel and K.A. Doksum. *Mathematical Statistics.* Holden-Day, Oakland, CA, 1977.

[2] C.R. Blyth. Approximate binomial confidence limits. *Journal of American Statistical Association*, 81:843–855, 1986.

[3] C.R. Blyth and H.A. Still. Binomial confidence intervals. *Journal of the American Statistical Association*, 78:108–116, 1983.

[4] L.D. Brown, T.T. Cai, and A. Dasgupta. Confidence intervals for a binomial proportion and asymptotic expansions. *Annals of Statistics*, 30:160–201, 2002.

[5] G. Casella. Refining binomial confidence intervals. *Canadian Journal of Statistics*, 14:113–129, 1986.

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

[7] E.L. Crow. Confidence intervals for a proportion. *Biometrika*, 43:423–435, 1956.

[8] T.S. Ferguson. *Mathematical Statistics.* Academic Press, New York, 1967.

[9] B.K. Ghosh. A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, 74:894–900, 1979.

[10] N.L. Johnson and S. Kotz. *Distributions in Statistics: Discrete Distributions.* Houghton Mifflin, Boston, 1969.

[11] E.S. Pearson and H.O. Hartley. *Biometrika Tables for Statisticians.* University Press, Cambridge, 1954.

[12] V.V. Petrov. *Sums of Independent Random Variables.* Spinger–Verlag, New York, 1975.

DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PA 18015

DEPARTMENT OF MATHEMATICS
LEHIGH UNIVERSITY
BETHLEHEM, PA 18015
wh02@lehigh.edu