Sequential Confidence Limits for the Ratio of Two Binomial Proportions

Hokwon Cho*
Department of Mathematical Sciences
University of Nevada, Las Vegas, Las Vegas, NV 89154

Z. Govindarajulu
Department of Statistics
University of Kentucky, Lexington, KY 40506

Abstract

We present a sequential method for obtaining approximate confidence limits for the ratio of two independent binomial proportions based on a slightly modified maximum likelihood estimator. Large-sample properties of the proposed sequential estimator are studied. Monte Carlo simulation is carried out in order to investigate its finite sample behavior. The proposed method is applied to a numerical example for illustration and its use.

Key words: Sequential confidence limits, ratio of two binomial proportions, modified maximum likelihood estimator.

1 Introduction

Often the ratio of two binomial proportions is of major interest or an important tool for measuring the risk ratio (Katz et al., 1978, Fleiss, 1981 and Bailey, 1987) or the relative risk (Gart, 1985) in comparative prospective studies and in biomedical experiments. The ratio or odds ratio of two binomial proportions is also related to vaccine efficacy and attributable risk (Walter, 1976), which arises frequently in epidemiological problems, e.g. cohort study involving two groups.

Among sequential methods for constructing an interval for an unknown parameter based on the fixed-sample size, Ray (1957) and Starr (1966) have studied the fixed-width confidence interval for the mean of a normal distribution. Using two-stage sampling, the analogous problem for the variance of a normal population has studied by Graybill and Connell (1964). For a paper related to ours, see Khan (1969) in which he explores a general method for determining stopping rules to obtain a fixed-width confidence interval for an unknown parameter involving possibly some unknown nuisance parameters. In this paper, we consider the sequential approach for establishing an approximate confidence limit for the ratio of two binomial proportions on the basis of a slightly modified maximum likelihood estimator (MLE).

1.1 Basic Formulation

Let $X_1, X_2, ...$ and $Y_1, Y_2, ...$ be two sequences of independent Bernoulli random variables with probabilities p_0 and p_1, respectively where $0 < p_0, p_1 < 1$. Let $\theta = p_1/p_0$. Our goal is to construct an

*Correponding author: cho@unlv.nevada.edu

interval of specified width $2d$ with confidence coefficient γ for the ratio θ. That is,

$$P\left\{\left|\hat{\theta}-\theta\right| \leq d\right\} \geq \gamma. \tag{1}$$

From samples of size n on each variable, define $R=\sum_{i=1}^{n} X_i$ and $S=\sum_{i=1}^{n} Y_i$. Then, R and S are two independent binomial random variables with parameters (n, p_0) and (n, p_1), respectively. Since there does not exist an unbiased estimator of the ratio θ, we use the estimator

$$\hat{\theta}_n = \frac{S+1/2}{R+1/2}. \tag{2}$$

Then, $\hat{\theta}_n$ is asymptotically unbiased for θ.

The likelihood (of θ and p_0) when $R=r$ and $S=s$ is given by

$$L(\theta, p_0) = \binom{n}{r}\binom{n}{s} p_0^{r+s}\left(1-p_0\right)^{n-r} \theta^s \left(1-p_0\theta\right)^{n-s}. \tag{3}$$

in which $p_1 = p_0\theta$. Hence, the log-likelihood function of θ is

$$\begin{aligned} l(\theta, p_0) = {} & \log\binom{n}{r} + \log\binom{n}{s} + (r+s)\log p_0 + (n-r)\log(1-p_0) \\ & + s\log\theta + (n-s)\log(1-p_0\theta). \end{aligned} \tag{4}$$

The parameters in Eq. (4) are estimated by the method of maximum likelihood (by means of solving two equations) and the solutions are $\hat{p}_{0,n} = R/n$ and $\hat{\theta}_{mle} = S/n\hat{p}_{0,n}$.

The information about θ is

$$I(\theta) = E\left[-\frac{\partial^2 l(\theta, p_0)}{\partial\theta^2}\right] = \frac{np_0^2}{p_1(1-p_1)}, \tag{5}$$

and the information about p_0 is given by

$$I(p_0) = E\left[-\frac{\partial^2 l(\theta, p_0)}{\partial p_0^2}\right] = \frac{n}{p_0}\left[\frac{1}{1-p_0} + \frac{\theta}{1-p_1}\right]. \tag{6}$$

Similarly, the information about θ and p_0 is

$$E\left[-\frac{\partial^2 l(\theta, p_0)}{\partial\theta\partial p_0}\right] = \frac{n(1-p_1)}{1-p_1} + \frac{n(1-p_1)p_1}{(1-p_1)^2} = \frac{n}{1-p_1}. \tag{7}$$

By combining Eqs. (5)-(7), the information matrix about (θ, p_0) (also, see Khan, 1969) is then

$$\mathbf{I}(\theta, p_0) = n\begin{bmatrix} p_0^2\left(\frac{1}{p_1(1-p_1)}\right) & \frac{1}{1-p_1} \\ \frac{1}{1-p_1} & \frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{\theta}{1-p_1}\right) \end{bmatrix}. \tag{8}$$

Hence, the inverse of $\mathbf{I}(\theta, p_0)$ is

$$\mathbf{I}^{-1}(\theta, p_0) = \frac{\theta(1-p_0)(1-p_1)}{n}\begin{bmatrix} \frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{\theta}{1-p_1}\right) & -\frac{1}{1-p_1} \\ -\frac{1}{1-p_1} & p_0^2\left(\frac{1}{p_1(1-p_1)}\right) \end{bmatrix}. \tag{9}$$

Then, from Eq. (9) the asymptotic variance of $\hat{\theta}_{mle}$ is

$$\begin{aligned} Var\left(\hat{\theta}_{mle}\right) &= \frac{\theta(1-p_0)(1-p_1)}{n}\left[\frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{\theta}{1-p_1}\right)\right] \\ &= \frac{\theta(1+\theta-2\theta p_0)}{np_0}. \end{aligned} \tag{10}$$

1.2 Asymptotic Variance of Estimator $\hat{\theta}_n$

In this section, we find the asymptotic variance of the estimator $\hat{\theta}_n = (S+1/2)/(R+1/2)$. We present the following theorem:

Theorem 1.1

$$\lim_{n\to\infty} nVar(\hat{\theta}_n) = \frac{p_1(p_0+p_1-2p_0p_1)}{p_0^3} = \frac{\theta(1+\theta-2\theta p_0)}{p_0}. \tag{11}$$

Proof. By Theorem 2 in von Bahr (1969), if $\{X_j\}$ is a sequence of i.i.d. r.v.'s such that for a positive integer $k \geq 2$, $E[|X_1|^k] < \infty$, then

$$E\left[\left(n^{-1/2}\sum_{j=1}^{n}\{X_j - E[X_j]\}\right)^k\right] \to E[(\sigma g)^k],$$

where $\sigma^2 = Var(X_1)$ and g is a r.v. with a standard normal distribution. This implies that for each positive integer

$$E\left[\left\{n^{-1/2}(R-np_0)\right\}^k\right] = O(1)$$

and

$$E\left[\left|n^{-1/2}(R-np_0)\right|^k\right] = O(1).$$

Hence,

$$E\left[\left|n^{1/2}U_n\right|^k\right] = O(1) \tag{12}$$

where $U_n = (np_0)^{-1}(R - np_0 + 1/2)$.

Using (12) and that for each $x \neq 1$, $(1+x)^{-1} = 1 - x + x^2 - x^3 + x^4 - x^5(1+x)^{-1}$, we get that

$$\begin{aligned}
E\left[\frac{n^2}{R+1/2}\right] &= np_0^{-1}E[(1+U_n)^{-1}] \\
&= np_0^{-1}E[1 - U_n + U_n^2 - U_n^3 + U_n^4 - U_n^5(1+U_n)^{-1}] \\
&= \frac{n}{p_0}\left[1 - \frac{1}{2np_0} + \frac{np_0(1-p_0)}{n^2p_0^2} + \frac{1}{4n^2p_0^2}\right] + o(1) - np_0^{-1}E[U_n^5(1+U_n)^{-1}] \\
&= \frac{n}{p_0}\left[1 + \frac{(1-2p_0)}{2np_0} + \frac{1}{4n^2p_0^2}\right] + o(1) - np_0^{-1}E[U_n^5(1+U_n)^{-1}].
\end{aligned}$$

Since $1 + U_n \geq (2np_0)^{-1}$,

$$|E\{np_0^{-1}E[U_n^5(1+U_n)^{-1}]\}| \leq np_0^{-1}E[|U_n^5(1+U_n)^{-1}] \leq 2n^2E[|U_n|^5] = o(1).$$

Hence,

$$E\left[\frac{n^2}{R+1/2}\right] = \frac{n}{p_0}\left(1 + \frac{(1-2p_0)}{2np_0} + \frac{1}{4n^2p_0^2}\right) + o(1). \tag{13}$$

Using (13) and that for each $x \neq 1$,

$$(1+x)^{-2} = 1 - 2x + 3x^2 - 4x^3 + 5x^4 - 6x^5 + 7x^6 - (8x^7 + 7x^8)(1+x)^{-2},$$

we get that

$$
\begin{aligned}
E\left[\frac{n^3}{(R+1/2)^2}\right] &= np_0^{-2}E[(1+U_n)^{-2}] \\
&= np_0^{-2}E[1-2U_n+3U_n^2-4U_n^3+5U_n^4-6U_n^5+7U_n^6 \\
&\quad -(8U_n^7+7U_n^8)(1+U_n)^{-2}] \\
&= \frac{1}{n^2p_0^2}\left[1-\frac{1}{np_0}+\frac{3np_0(1-p_0)}{n^2p_0^2}+\frac{3}{4n^2p_0^2}\right]+o(1) \\
&\quad -np_0^{-2}E[(8U_n^7+7U_n^8)(1+U_n)^{-2}] \\
&= \frac{1}{n^2p_0^2}\left[1+\frac{(2-3p_0)}{np_0}+\frac{3}{4n^2p_0^2}\right]+o(1) \\
&\quad -np_0^{-2}E[(8U_n^7+7U_n^8)(1+U_n)^{-2}]
\end{aligned}
$$

Since $1+U_n \geq (2np_0)^{-1}$,

$$
\left|np_0^{-2}E[(8U_n^7+7U_n^8)(1+U_n)^{-2}]\right| \leq 4n^3E[(8|U_n|^7+7|U_n|^8)] = o(1).
$$

Hence,

$$
E\left[\frac{n^3}{(R+1/2)^2}\right] = \frac{n}{p_0^2}\left[1+\frac{(2-3p_0)}{np_0}+\frac{3}{4n^2p_0^2}\right]+o(1). \tag{14}
$$

From Eqs. (13) and (14), we obtain

$$
\begin{aligned}
Var\{(R+1/2)^{-1}\} &= \frac{1}{n^2p_0^2}\left[1+\frac{2-3p_0}{np_0}+O(n^{-2})\right]-\frac{1}{n^2p_0^2}\left[1+\frac{1-2p_0}{np_0}+O(n^{-2})\right] \\
&= \frac{1}{n^2p_0^2}\left[\frac{1-p_0}{np_0}+O(n^{-2})\right] \\
&= \frac{(1-p_0)}{n^3p_0^3}\left[1+\frac{1}{np_0}+O(n^{-2})\right]. \qquad (15)
\end{aligned}
$$

Next, if the variables V and W are independent, one can easily show that

$$
Var(VW) = [Var(V)][Var(W)]+E^2(V)Var(W)+E^2(W)Var(V).
$$

Now setting $V = S+1/2$ and $W = (R+1/2)^{-1}$, we have

$$
\begin{aligned}
nVar(\hat{\theta}_n) &= \frac{p_1(1-p_1)(1-p_0)}{np_0^3}\left[1+O(n^{-1})\right] \\
&\quad +\left(p_1+\frac{1}{2n}\right)^2\frac{(1-p_0)}{p_0^3}\left[1+\frac{1}{np_0}+O(n^{-2})\right] \\
&\quad +\frac{p_1(1-p_1)}{p_0^2}\left[1+\frac{1-2p_0}{np_0}+O(n^{-2})\right] \\
&= \frac{p_1^2(1-p_0)}{p_0^3}+\frac{p_1(1-p_1)}{p_0^2}+O(n^{-1}) \\
&= \frac{p_1(p_0+p_1-2p_0p_1)}{p_0^3}+O(n^{-1}).
\end{aligned}
$$

□

Hence, we infer that the asymptotic variance of $\hat{\theta}_n$ is the same as that of the MLE.

1.3 Asymptotic Normality of $\hat{\theta}_n$

From the ratio given in Eq. (2), we have

$$\begin{aligned}\sqrt{n}\left(\frac{S+1/2}{R+1/2}-\theta\right) &= \sqrt{n}\left[\frac{S-np_1}{R+1/2}-\frac{\theta(R+1/2)-(np_1+1/2)}{R+1/2}\right] \\ &= \sqrt{n}\left[\frac{\hat{p}_{1,n}-p_1}{\hat{p}_{0,n}+1/2n}-\frac{\theta\left(\hat{p}_{0,n}+1/2n\right)-(p_1+1/2n)}{\hat{p}_{0,n}+1/2n}\right] \\ &= \sqrt{n}\left(\frac{\hat{p}_{1,n}-p_1}{p_0}\right)-\theta\left(\frac{\hat{p}_{0,n}-p_0}{p_0}\right)\sqrt{n}+o_p(1), \end{aligned} \tag{16}$$

where $\hat{p}_{0,n} = R/n$ and $\hat{p}_{1,n} = S/n$. For sufficiently large n and due to Slutsky's theorem, it follows from (16) that

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right) \overset{d}{\simeq} N\left(0, \frac{\theta\left(1-p_1\right)+\theta^2\left(1-p_0\right)}{p_0}\right) \equiv N\left(0,\sigma^2\right)$$

where $\sigma^2 = \theta\left(1+\theta-2\theta p_0\right)/p_0$. This asymptotic variance agrees with the one for the MLE. Hence, the estimator $\hat{\theta}_n$ is asymptotically efficient. (See Section 6.3 in Lehmann and Casella, 1998.)

Now we wish to determine n that satisfies

$$P\left\{\left|\hat{\theta}-\theta\right| \leq d\right\} = P\left\{\sqrt{n}\left|\hat{\theta}-\theta\right|/\sigma \leq d\sqrt{n}/\sigma\right\} \geq \gamma.$$

Thus,

$$2\Phi\left(d\sqrt{n}/\sigma\right)-1 \geq \gamma,$$

or

$$d\sqrt{n}/\sigma \geq z_{(1+\gamma)/2} = z \text{ (say)}, \tag{17}$$

for specified $d\,(>0)$ where $\Phi\left(z_{(1+\gamma)/2}\right) = (1+\gamma)/2$.

Consequently, we have

$$n \geq (z\sigma/d)^2. \tag{18}$$

Hence, the optimal fixed-sample size for the procedure becomes the smallest integer n^* such that $n \leq n^* \leq n+1$, for estimating θ with specified d and z. That is,

$$n^* = \left[(z\sigma/d)^2\right]+1 \tag{19}$$

where $[\cdot]$ indicates the greatest integer function.

However, since both θ and p_0 are unknown, the goal manifestly cannot be achieved by means of a fixed-sample size procedure. So we have the following adaptive sequential rule: Stop sampling at N observations, where

$$N = \inf\left\{n \geq m : n \geq z^2\hat{\sigma}_n^2/d^2\right\} \tag{20}$$

where $m\,(\geq 2)$ is the initial sample size and $\hat{\sigma}_n^2 = \hat{\theta}_n(1+\hat{\theta}_n-2\hat{\theta}_n\hat{p}_0)/\hat{p}_0$ with $\hat{p}_0 = (R+1/2)/n$ and $\hat{p}_1 = (S+1/2)/n$.

Then, upon stopping we give the $\gamma\%$ confidence interval estimate of length $2d$ for θ as

$$\left(\hat{\theta}_N-d,\ \hat{\theta}_N+d\right). \tag{21}$$

2 Asymptotic Properties of the Procedure

In this section we study the asymptotic behavior of the proposed sequential procedure and study some properties of the stopping time N.

2.1 Finite Sure Termination

We now wish to show the fundamental property that the proposed stopping rule terminates finitely almost surely.

Theorem 2.1 *Let N be the stopping time associated with the proposed sequential procedure. Then $P\{N<\infty\}=1$.*

Proof. Using the stopping rule in Eq. (20)

$$\begin{aligned} P\{N=\infty\} &= \lim_{n\to\infty} P\{N>n\} \\ &\le \lim_{n\to\infty} P\left\{n \le z^2\hat{\sigma}_n^2/d^2\right\} = 0 \end{aligned} \tag{22}$$

since $\hat{\sigma}_n^2$ converges to σ^2 in probability as $n\to\infty$. Hence the sequential procedure terminates finitely with probability one. □

2.2 First Order Asymptotics

As a measure of evaluating the proposed procedure, we apply the desirable criteria given in Chow and Robbins (1965), which are asymptotic efficiency and consistency. That is, we examine the asymptotic behavior of the stopping time as d tends to zero.

The stopping rule given by Eq. (20) can be written as

$$N = \inf\left\{ n \ge m : \frac{n}{\theta\,(z/d)^2} \ge \frac{\hat{\theta}_n}{\theta}\left[\frac{1+\hat{\theta}_n\,(1-2\hat{p}_0)}{\hat{p}_0}\right]\right\}. \tag{23}$$

Then, Eq. (23) takes the form:

$$N = N(t) = \min\{n \ge m : Y_n \le f(n)/t\} \tag{24}$$

where

$$Y_n = \frac{\hat{\theta}_n}{\theta}\left(\frac{p_0}{\hat{p}_0}\right)\left(\frac{1+\hat{\theta}_n-2\hat{\theta}_n\hat{p}_0}{1+\theta-2\theta p_0}\right), \tag{25}$$

$$f(n) = n, \tag{26}$$

and

$$t = (z/d)^2\,\theta\,(1+\theta-2\theta p_0)\,/p_0. \tag{27}$$

Clearly, Y_n is then a sequence of random variables such that $Y_n > 0$ almost surely (a.s.), $\lim_{n\to\infty} Y_n = 1$ a.s. due to the fact that $\hat{p}_0 \to p_0$ a.s. and $\hat{\theta}_n/\theta \to 1$ a.s. as $n\to\infty$, respectively. Also we note that $\lim_{n\to\infty} f(n) = \infty$ and $\lim_{n\to\infty} f(n)/f(n-1) = 1$.

Since the stopping rule N is well-defined and non-decreasing as a function of t, by applying the results of Chow and Robbins (1965) we obtain the following first order asymptotics for the proposed sequential procedure.

Theorem 2.2

(i) $\lim_{d\to 0} N = \infty$ *a.s.*, $\lim_{d\to 0} E(N) = \infty$,

(ii) $\lim_{d\to 0} N/n^* = 1$ *a.s.*,

(iii) $\lim_{d\to 0} P\left\{\left|\hat{\theta}_N - \theta\right| \le d\right\} = \gamma$.

Proof. (i) is easily verified using the definition of N, Eq. (23) and the monotone convergence theorem.

For (ii), from (22) since $N-1 \leq z^2\sigma_{N-1}^2/d^2$, we have

$$\frac{z^2\sigma_N^2/d^2}{z^2\sigma^2/d^2} \leq \frac{N}{n^*} \leq \frac{1+z^2\sigma_{N-1}^2/d^2}{z^2\sigma^2/d^2}. \tag{28}$$

After simplifying (28) we get

$$\sigma_N^2/\sigma^2 \leq N/n^* \leq (d/z\sigma)^2 + \left(\sigma_{N-1}^2/\sigma^2\right). \tag{29}$$

Taking limits on inequalities in (29) as d goes to zero, we have

$$\lim_{d\to 0}\left(\sigma_N^2/\sigma^2\right) \leq \lim_{d\to 0}\left(N/n^*\right) \leq \lim_{d\to 0}\left(\sigma_{N-1}^2/\sigma^2\right). \tag{30}$$

However, the quantities on the extremes of the inequality tends to unity. Hence $\lim_{d\to 0}\left(N/n^*\right) = 1$.

We prove (iii) via the following lemma:

Lemma 2.1 *If N/n^* converges to one in probability, $\sqrt{n^*}\left(\hat{p}_{0,N} - p_0\right)$ has an asymptotically normal distribution with mean zero and variance $p_0\left(1-p_0\right)$.*

This follows from Anscombe's theorem (1952). From this, it follows that $\hat{p}_{0,N}$ converges to p_0 in probability. Using Slutsky theorem, we infer that

$$\sqrt{n^*}\left(\hat{\theta}_N - \theta\right) \overset{d}{\simeq} \sqrt{n^*}\left[\frac{\left(\hat{p}_{1,N} - p_1\right)}{p_0} - \frac{\theta\left(\hat{p}_{0,N} - p_0\right)}{p_0}\right]. \tag{31}$$

Then, the asymptotic normality of right-hand side in (31) follows from Anscombe's condition specialized for sums of independent and identically distributed (i.i.d.) random variables. Hence $\sqrt{n^*}\left(\hat{\theta}_N - \theta\right)$ is asymptotically $N\left(0,\sigma^2\right)$.

Thus

$$P\left\{\left|\hat{\theta}_N - \theta\right| \leq d\right\} = P\left\{\frac{\sqrt{n^*}\left|\hat{\theta}_N - \theta\right|}{\sigma} \leq \frac{d\sqrt{n^*}}{\sigma}\right\} = \gamma \tag{32}$$

as $d \to 0$. This completes the proof of Theorem 2.2. □

We use the following lemma in order to assert the asymptotic efficiency property of the sequential procedure.

Lemma 2.2 (*Lemma 2.6.3 in Dmitrienko, 1998*) *Let $\{Z_k, k \geq 1\}$ be a sequence of positive random variables and $\{m_k, k \geq 1\}$ a sequence of positive real numbers such that m_k increases with k and $Z_k/m_k \to 1$ a.s. as $k \to \infty$. Also for any $b > 0$, let*

$$T(b) = \inf\{k \geq 1 : Z_k \geq b\},\ t(b) = \inf\{k \geq 1 : m_k \geq b\} \tag{33}$$

and assume that

$$\lim_{\rho\to 1}\lim_{b\to\infty}\left[t(b\rho)/t(b)\right] = 1. \tag{34}$$

Then, as $b \to \infty$,

$$T(b)/t(b) \to 1 \text{ a.s.} \tag{35}$$

Furthermore, if, for some $\delta > 0$,

$$\sum_{k=1}^{\infty} P\left\{Z_k < \delta m_k\right\} < \infty \tag{36}$$

then, as $b \to \infty$

$$E\left[T\left(b\right)/t\left(b\right)\right] \to 1. \tag{37}$$

Theorem 2.3 (*Asymptotic Efficiency*) *Let* $0 < p_0, p_1 < 1$. *Then*

$$\lim_{d \to 0} E\left(N\right)/n^* = 1. \tag{38}$$

Proof. We apply Theorem 2.2 to

$$\begin{aligned} Z_n &:= \frac{n}{\hat{\sigma}_n^2} = \frac{n(\hat{p}_{0,n} + 1/2n)}{\hat{\theta}_n(1 + \hat{\theta}_n - 2\hat{p}_{1,n})} \\ &= \frac{n(\hat{p}_{0,n} + 1/2n)^3}{(\hat{p}_{1,n} + 1/2n)(\hat{p}_{0,n} + \hat{p}_{1,n} - 2\hat{p}_{0,n}\hat{p}_{1,n} + 1/n - (\hat{p}_{0,n} + \hat{p}_{1,n})/2n + 1/4n^2)} \end{aligned}$$

$b = z^2 d^2$ and

$$m_n := \frac{np_0^3}{p_1(p_0 + p_1 - p_0 p_1)},$$

where $\hat{p}_{0,n} = R/n$ and $\hat{p}_{1,n} = S/n$. If $0 < p_0 < 1$ and $0 < p_1 < 1$, to prove (36), we show that $\{n^{-1} Z_n\}$ satisfies a large deviation principle with speed n and rate function

$$I_Z(u) := \inf\{I_{\hat{p}_{0,n},\hat{p}_{1,n}}(s,t) : \frac{s^3}{t(s + t - 2st)} = u\}, u \geq 0,$$

where

$$\begin{aligned} I_{\hat{p}_{0,n},\hat{p}_{1,n}}(s,t) &= s \log\left(\frac{s}{p_0}\right) + (1-s)\log\left(\frac{1-s}{1-p_0}\right) \\ &\quad + t\log\left(\frac{t}{p_1}\right) + (1-t)\log\left(\frac{1-t}{1-p_1}\right), \ s,t \in [0,1]. \end{aligned}$$

The definition of large deviation principle (LDP) can be found in Section 1.2 in Dembo and Zeitouni (1998). By the Cramér theorem (see Section 2.2 in Dembo and Zeitouni, 1998), $\{\hat{p}_{0,n}\}$ satisfies a large deviation principle with speed n and rate function

$$I_{\hat{p}_{0,n}}(s) = s\log\left(\frac{s}{p_0}\right) + (1-s)\log\left(\frac{1-s}{1-p_0}\right), \ s \in [0,1],$$

(see Exercise 2.2.3 in Dembo and Zeitouni, 1998). $\{\hat{p}_{1,n}\}$ also satisfies the LDP. Since $\{\hat{p}_{0,n}\}$ and $\{\hat{p}_{1,n}\}$ are independent, $\{(\hat{p}_{0,n}, \hat{p}_{1,n})\}$ satisfies a large deviation principle with speed n and rate function

$$\begin{aligned} I_{\hat{p}_{0,n},\hat{p}_{1,n}}(s,t) &= I_{\hat{p}_{0,n}}(s) + I_{\hat{p}_{1,n}}(t) \\ &= s\log\left(\frac{s}{p_0}\right) + (1-s)\log\left(\frac{1-s}{1-p_0}\right) \\ &\quad + t\log\left(\frac{t}{p_1}\right) + (1-t)\log\left(\frac{1-t}{1-p_1}\right), \ s,t \in [0,1]. \end{aligned}$$

Let

$$f_n(x,y) = \frac{(x + 1/2n)^3}{(y + 1/2n)[x + y - 2xy + 1/n - (x+y)1/2n + 1/4n^2]}, \ x,y \in [0,1]$$

and let

$$f(x,y) = \frac{x^3}{y(x+y-2xy)}, \;\; x,y \in [0,1].$$

We have that if $(x_n, y_n) \to (x,y)$, then $f_n(x_n, y_n) \to f(x,y)$. So, by Theorem 2.1 in Arcones (2003), $\{n^{-1}Z_n = f_n(\hat{p}_{0,n}, \hat{p}_{1,n})\}$ satisfies a large deviation principle with speed n and rate function I_Z. This implies (38). Notice that $I_{\hat{p}_{0,n},\hat{p}_{1,n}}(s,t) = 0$, only if $(s,t) = (p_0, p_1)$ and $I_Z(u) = 0$ only if $u = p_0^3/\left[p_1(p_0 + p_1 - 2p_0p_1)\right]$. □

3 Numerical Studies

3.1 Simulation Set-up

To investigate the behavior and the performance of the stopping rule, Monte Carlo experimentation is carried out for the proposed sequential procedure. The results of the Monte Carlo simulation are summarized in the following tables, which show several choices of the parameter θ, namely $\theta = 1.0, 1.5, 2.0, 4.0$ with selected values of p_0 and p_1. Since X and Y can be interchanged when $\theta \leq 1$, we have considered only situations in which $\theta \geq 1$.

In the table, every value in each row is based on 5,000 independent replications. The expected stopping time is denoted by $E(N)$, optimal sample sizes n^* and the coverage probability (CP) are shown with specified width d with initial sample size $m = 10$ used in each experiment. The nominal level of confidence γ for the interval is 0.90 or 0.95.

Table 3.1 For $\theta = 1.0$ (when $p_0 = 0.5, p_1 = 0.5$)

	$\gamma = 90\%$				$\gamma = 95\%$			
d	$\hat{\theta}$	$E(N)$	n^*	CP	$\hat{\theta}$	$E(N)$	n^*	CP
.1	1.000	537.27	542	.897	1.000	765.09	769	.946
.2	1.000	130.41	136	.881	1.000	188.44	193	.939
.3	1.000	54.89	61	.844	1.002	80.34	86	.917
.4	1.002	29.74	34	.874	.999	42.86	49	.912

Table 3.2 For $\theta = 1.5$ (when $p_0 = 0.4, p_1 = 0.6$)

	$\gamma = 90\%$				$\gamma = 95\%$			
d	$\hat{\theta}$	$E(N)$	n^*	CP	$\hat{\theta}$	$E(N)$	n^*	CP
.15	1.499	578.86	587	.891	1.502	828.15	834	.945
.25	1.500	203.91	211	.874	1.499	291.73	300	.925
.35	1.501	98.95	108	.838	1.500	144.86	154	.911
.45	1.505	57.10	66	.823	1.500	82.93	93	.875

Table 3.3 For $\theta = 2.0$ (when $p_0 = 0.3, p_1 = 0.6$)

	$\gamma = 90\%$				$\gamma = 95\%$			
d	$\hat{\theta}$	$E(N)$	n^*	CP	$\hat{\theta}$	$E(N)$	n^*	CP
.2	2.002	804.07	813	.897	2.001	1143.39	1153	.940
.3	1.998	348.70	361	.881	1.997	499.88	512	.930
.4	1.999	189.13	203	.836	1.998	275.56	288	.915
.5	2.003	115.66	131	.801	2.000	170.82	185	.883

Table 3.4 For $\theta = 4.0$ (when $p_0 = 0.2, p_1 = 0.8$)

	$\gamma = 90\%$				$\gamma = 95\%$			
d	$\hat{\theta}$	$E(N)$	n^*	CP	$\hat{\theta}$	$E(N)$	n^*	CP
.4	4.006	1141.54	1152	.891	4.004	1620.48	1635	.945
.5	3.997	715.14	736	.877	4.005	1030.20	1047	.936
.6	3.993	484.80	511	.857	4.002	706.34	727	.931
.7	4.005	349.98	377	.833	3.997	508.34	533	.908

From Tables 3.1 to 3.4, we infer that the expected stopping time $E(N)$ monotonically increases (to infinity) as d decreases (to zero). The Monte Carlo estimate of $\hat{\theta}$ approaches the true value of the parameter θ as the length of the interval decreases, and we also observe that as d decreases the coverage probability (CP) is getting close to the nominal probability γ. (This property is referred to as the *asymptotic consistency*.) Therefore, the above numerical evidence indicates that the small sample behavior lends support to the asymptotic behavior of the proposed sequential procedure when $d \to 0$.

In fact, increasing the starting sample size m results in increase of both $E(N)$ and CP. Accordingly, when the CP is below the nominal level, choosing a moderate size of m is a trade-off for obtaining higher coverage probability. For practical purposes, the size of d can be determined from the standard error (S.E.) of the estimate $\hat{\theta}$.

Remark 3.1 The (detailed) Monte Carlo simulation results show that the coverage probability normally starts at a level higher (in fact, close to 100% with large value of d) than the nominal level γ, and it goes down below the level γ. Then, eventually CP converges to the nominal level γ as d gets smaller (see, also pp. 42-44 in Starr, 1966). It is worthwhile to note the fact that N monotonically increases to infinity as d decreases to zero; however, Nd^2 does converge to n^* but not monotonically.

Remark 3.2 The coverage probability will be close to nominal level especially when d is small. As Tables 3.1-3.4 suggest CPs fall below the nominal levels when d gets moderately large. In order to alleviate this situation, one can introduce the *damping factor* in the stopping time given by Eq. (20). Consider the damping factor $l_n = 1 + l_0/n^{\delta}$, where $l_0 = 1, 2, \ldots (< \infty)$ and $0 < \delta \leq 1$. Then, the stopping time with the damping factor will be

$$N_D = \inf\left\{ n \geq m : n \geq (z\hat{\sigma}_n/d)^2 \, l_n \right\}.$$

After studying empirical results for various choices of l_0 and δ, we find the best values are $l_0 = 1$ and $\delta = 1/2$. Thus, we get improved CPs to nominal level γ. Since the damping factor converges to one as n gets large, the properties of first order asymptotics of the stopping time will not be affected. Table 3.5 summarizes some of the results when $l_0 = 1$ and $\delta = 1/2$ are used.

Table 3.5: N_D and damping factor: $l_n = 1 + l_0/n^{\delta}$, $m = 10$

$(\theta - d, \theta + d)$		$\gamma = 90\%$			$\gamma = 95\%$		
Value of θ and (p_0, p_1)	d	$E(N_D)$	n^*	CP	$E(N_D)$	n^*	CP
(1) $\theta = 1.0$ ($p_0 = .5, p_1 = .5$)	.40	35.06	34	.897	50.25	49	.925
(2) $\theta = 1.5$ ($p_0 = .4, p_1 = .6$)	.45	65.57	66	.854	94.19	93	.904
(3) $\theta = 2.0$ ($p_0 = .3, p_1 = .6$)	.50	129.72	131	.848	187.18	185	.912
(4) $\theta = 4.0$ ($p_0 = .2, p_1 = .8$)	.70	374.03	377	.855	533.61	533	.941

3.2 Application

Instead of observing Bernoulli variables at each stage, one can observe binomial variables based on a fixed number of trials, say k (that is, the number of trials does not change from stage to stage). Then, the procedure as well as the entire asymptotic theory goes through provided Nk tends to be large where N is now to be interpreted as the number of stages.

We consider data given in Montgomery (1985, pp. 123-127), which consist of binomial variables based on a fixed number of trials instead of a data set with a large number of Bernoulli trials.

Example 3.1 Consider a packing process of frozen orange juice concentrate packed in 6-oz cardboard. We monitor the fraction of nonconforming cans of the frozen ones and each stage sampling involves taking 50 observations. For the first three-shift period 30 stages of samples were collected. For the next three shifts after the adjustment to the process, 24 stages of samples are collected. Since two sample fractions in stages 15 and 23 in the first period were identified to be anomalous observations, we discard these two pairs of observations from both periods. Let X be the number of nonconformances in each stage during the first period, and let Y represent the number of nonconformances during the second period. Thus, the resulting 22 pairs of observations (x, y) are obtained as follows:

Sample No.	1	2	3	4	5	6	7	8	9	10	11
x	12	15	8	10	4	7	16	9	14	10	5
y	9	6	12	5	6	4	5	3	7	6	2
Sample No.	12	13	14	15	16	17	18	19	20	21	22
x	6	17	12	8	10	5	13	11	20	18	15
y	4	3	6	4	8	5	6	7	5	6	4

Let p_0 denotes the proportion of nonconforming cans during the first period and let p_1 be the proportion of nonconformances during the second period. We define the ratio $\theta = p_1/p_0$ of two nonconforming proportions p_0 and p_1. We wish to construct an interval with confidence γ for the ratio θ with width $2d$ based on the estimate $\hat{\theta}_n = \hat{p}_{1,n}/\hat{p}_{0,n}$, where n here denotes the stopping stage, and the total sample size after stopping in each period is $n \times 50$ cans. The initial sample stage $m = 1$ is taken for all procedures.

(a) First, we construct a 90% confidence interval with $d = 0.3$. The proposed sequential procedure stops at sample stage $n = 3$, the estimates are $\hat{p}_0 = 0.180$, $\hat{p}_1 = 0.233$ and $\hat{\theta} = 1.296$ with standard error 0.296. Therefore, a 90% confidence interval for the risk ratio θ between two periods is given by $(0.996, 1.596)$ based on sample size $n = 150$ cardboard cans used for each period.

(b) Next, for a 95% confidence interval for the ratio $\theta = p_1/p_0$ of two nonconforming proportions p_0 and p_1 with $d = 0.2$, the sequential procedure terminates at stage $n = 6$, which yields the estimates $\hat{p}_0 = 0.140$, $\hat{p}_1 = 0.187$ and $\hat{\theta} = 1.333$ with standard error 0.249 based on 300 cans. Therefore, a 95% confidence interval for the true ratio θ with $d = 0.2$ is given by $(1.133, 1.533)$. The following table illustrates the rest of the results we obtained.

Table 3.6 The Confidence Limits of θ for Orange Juice Packed Data

Conf. level γ	d	n	$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}_n$	S.E.	Conf. limits
90%	.3	3	.180	.233	1.296	.296	(0.996,1.596)
	.2	5	.152	.196	1.289	.254	(1.089,1.489)
95%	.3	4	.160	.225	1.406	.293	(1.106,1.706)
	.2	6	.140	.187	1.333	.249	(1.133,1.533)
99%	.3	5	.152	.196	1.289	.254	(0.989,1.589)
	.2	10	.126	.210	1.667	.244	(1.467,1.867)

4 Concluding Remarks

The degenerate case in which $p_0 = p_1 = 1$ can be handled by the methods of this paper provided $\hat{\sigma}_n^2 = 1/n + \hat{\theta}_n(1 + \hat{\theta}_n - 2\hat{\theta}_n\hat{p}_0)/\hat{p}_0$. It is possible to extend the results of this paper to the case where one sample size is a multiple of the other sample size. However, this will be a topic for a future investigation.

Acknowledgment. The authors are indebted to Professor Arcones, Editor-in-chief, for making valuable and detailed suggestions which have led to improved proofs of Theorems 1.1 and 2.3. The authors also express their appreciation to the Associate Editor and a referee for their comments and helpful suggestions.

References

[1] Arcones, M. A. (2003). The large deviation principle for stochastic processes I. *Theory of Probability & Its Applications* **47** 567-583.

[2] Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600-607.

[3] Bailey, B. J. R. (1987). Confidence limits to the risk ratio. *Biometrics* **43** 201-205.

[4] Bedrick, E. J. (1987). A family of confidence intervals for the ratio of two binomial proportions. *Biometrics* **43** 993-998.

[5] Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *The Annals of Mathematical Statistics* **36** 457-462.

[6] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Second Edition. Springer, New York.

[7] Dmitrienko, A. (1998). Sequential methods in linear and generalized linear models. Doctoral Dissertation, University of Kentucky.

[8] Fleiss, J. (1981). *Statistical Methods for Rates and Proportions*, Wiley and Sons, New York.

[9] Gart, J. (1985). Approximate tests and interval estimation of the common relative risk in the combination of 2×2 tables. *Biometrika* **72** 673-677.

[10] Graybill, F. and Connell, T. (1964). Sample size required for estimating the variance width d units of the true value. *The Annals of Mathematical Statistics* **35** 438-440.

[11] Katz, D., Baptista, J., Azen, S., and Pike, M. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics* **34** 469-474.

[12] Khan, R. (1969). A general method of determining fixed-width confidence intervals. *The Annals of Mathematical Statistics* **40** 704-709.

[13] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Second edition. Springer-Verlag, New York.

[14] Montgomery, D. (1985). *Introduction to Statistical Quality Control*. Wiley and Sons, New York.

[15] Ray, D. (1957). Sequential confidence intervals for the mean of a normal distribution with unknown variance. *Journal of the Royal Statistical Society. Series B*, **19** 133-143.

[16] Starr, N. (1966). The performance of a sequential procedure for the fixed-width interval estimation of the mean. *The Annals of Mathematical Statistics* **37** 36-50.

[17] Von Bahr, B. (1965). On the Convergence of Moments in the Central Limit Theorem. *Annals of Mathematical Statistics* **36** 808-818.

[18] Walter, S. (1976). The estimation and interpretation of attributable risk in health research. *Biometrics* **32** 829-849.