5-1-2015

A study of sequential inference for the risk ratio and measure of reduction of two binomials

Zhou Wang
University of Nevada, Las Vegas, wangz13@unlv.nevada.edu

A STUDY OF SEQUENTIAL INFERENCE FOR THE RISK RATIO AND MEASURE OF REDUCTION OF TWO BINOMIALS

by

Zhou Wang

Bachelor of Statistics
University of Science and Technology of China
2009

A dissertation submitted in partial fulfillment of
the requirements for the

Doctor of Philosophy - Mathematical Sciences

Department of Mathematical Sciences
College of Sciences
The Graduate College

University of Nevada, Las Vegas
May 2015

We recommend the dissertation prepared under our supervision by

Zhou Wang

entitled

A Study of Sequential Inferences for the Risk Ratio and Measure of Reduction of Two Binomials

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy - Mathematical Sciences

Department of Mathematical Sciences

Hokwon Cho, Ph.D., Committee Chair

Malwane Ananda, Ph.D., Committee Member

Jessica Jaynes, Ph.D., Committee Member

Paulo Pinheiro, Ph.D., Graduate College Representative

Kathryn Hausbeck Korgan, Ph.D., Interim Dean of the Graduate College

May 2015

ABSTRACT

A STUDY OF SEQUENTIAL INFERENCE FOR THE RISK RATIO AND MEASURE OF REDUCTION OF TWO BINOMIALS

by

Zhou Wang

Dr. Hokwon Cho , Examination Committee Chair
Associate Professor of Statistics
University of Nevada, Las Vegas, USA

The binomial distribution is one of the most commonly and widely occurring probabilistic phenomena in our lives. Since observations from independent Bernoulli trials yield a dichotomous type, the distribution of sequences provides the basis and clue for statistical formulations of a wide variety of problems.

Occasionally, the core of biomedical studies is related to the comparison and evaluation of the risks of events or outcomes of interest in comparing populations under study. For instance, one wishes to compare two groups of subjects drawn from two independent populations. Then, two sample proportions play central roles in those comparisons. One of the most useful ways to make comparisons for the relative risk is to take a ratio, also referred to as the risk ratio. In addition, a measure of reduction of the two proportions is considered.

In this thesis, we consider sequential methods of inferences for the ratio of two independent binomial probabilities, the risk ratio, in two populations for comparison. We obtain approximate confidence intervals and optimal sample sizes for the risk ratio and measure of reduction, respectively. Since there does not exist an unbiased estimator of the risk ratio, the procedure is developed based on a slightly modified

maximum likelihood estimator. Then, we explore properties of the proposed estimator using the standard criteria, such as unbiasedness, asymptotic variance, and the normality. For further investigation, we study the first-order asymptotic expansions and large sample properties using the asymptotic results. Then, the finite sample behavior will be examined through numerical studies. Monte Carlo experiment is performed for the various scenarios of parameters of two populations.

Through illustrations, we compare the performance of the proposed methods, which is Wald-based confidence intervals, with the likelihood-ratio confidence intervals in light of length, sample sizes, and invariance. Then, we extend the proposed sequential procedure to two-stage sampling design, which has a pilot sampling stage and a stage of gathering all remaining observations if needed. The two-stage procedure is naturally a little more versatile and practical than pure sequential in terms of sample size and stopping time in many situations. Again, through numerical studies, we study the advantages and usefulness of the two- stage method as well.

Consequently, by providing more comprehensive study of dynamic sampling plans for studying the risk ratio, we hope to contribute various inferential methods to the risk ratio and related problems.

ACKNOWLEDGEMENTS

First and foremost, I would like to gratefully and sincerely thank Dr. Hokwon Cho for his guidance, understanding, patience, and most importantly, his friendship during my graduate studies at UNLV. His mentorship was paramount in providing a well rounded experience consistent my long-term career goals. He encouraged me to not only grow as a PhD. student but also as an independent thinker. At the same time, his attitudes towards research and collaboration provided me with a living example, which I will aspire to follow in my future career.

Second, I would like to thank Dr. Ananda, Dr. Jaynes and Dr. Pinheiro for serving on my dissertation committee. Their advice and comments are so valuable. I am also very grateful for all the other professors in the department of Mathematical Sciences (at the University of Nevada, Las Vegas) who have taught me statistical courses over the past five years of my pursuit of the doctoral degree.

Finally, I sincerely thank to my parents, who provide the advice and financial support. The product of this dissertation would not be possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Motivation of the Problem

In this thesis we are concerned with the risk ratio and a measure of reduction for two independent binomial variates. Binomial probability phenomena has become more and more commonly used in our lives. Taking a ratio of two binomial proportions is of major interest and provides an important tool for measuring the risk ratio or the relative risk. These measures have been studied by many researchers and frequently used in cohort studies (Katz et al., 1978 and Gart, 1985), medical and pharmaceutical problems (Koopman, 1984), and epidemiological problems (Bailey, 1987). Additionally, when we are interested in how much the risk has been reduced, a more convenient way to figure it out is to consider a measure of reduction. Then, the measure of reduction is more practical to utilize as a measurement objective and can be more useful to practitioners in comparison of two binomial proportions.

First, we study the properties of the risk ratio and measure of reduction with various types of sampling schemes focusing on sequential methods. We develop the procedures depending upon the sampling scheme. Then, we present a sequential method for constructing confidence limits based on a slightly modified maximum likelihood estimator. Monte Carlo simulation is carried out in order to investigate its finite sample behavior. Also, the proposed method is applied to a numerical example to illustrate its use.

Lachin (2000) presented various types of measures of relative risk to compare two populations and summarized their large sample distributions for testing in his book. In practice, Fagerland, Lydersen and Laake (2011) reviewed several different methods and their confidence intervals focusing on the existing difference and the ratio of proportions. We also would like to introduce some of these in the table below.

Table 1.1. Measures of Relative Risk

Parameter θ	Form	Domain	Null Value
Risk Ratio (RR)	p_1/p_0	$(0, -\infty)$	1
Risk Difference (RD)	$p_1 - p_0$	$[-1, 1]$	0
log Risk Ratio (LRR)	$\log p_1 - \log p_0$	$(-\infty, \infty)$	0
Odds Ratio (OR)	$\frac{p_1/(1-p_1)}{p_0/(1-p_0)}$	$(0, \infty)$	1
Measure of Reduction (MOR)	$\frac{p_0-p_1}{p_0}$	$(-\infty, 1]$	0

1.2 Sequential Approaches

1.2.1 Historical Background and Literature

The modern theory of sequential analysis began its march with applied motivations in response to demands for more efficient sampling inspection procedures during World War II. It first came into existence simultaneously in the United States and Great Britain. The development in large-scale survey sampling of national importance was regarded by many, including Abraham Wald, as the pioneer of sequential analysis. In Abraham Wald's 1939 paper, he first pointed out that the two central procedures of the sampling distribution form the base of statistical-theory, namely hypothesis testing and parameter estimation, are special cases of the general statistical decision-making problem.

Wald's paper renewed and synthesized many concepts of statistical theory, including loss functions, risk functions, admissible decision rules, antecedent distributions, Bayesian procedures, and minimax procedures. Making decision on the sample size efficiently was taken into consideration. Wald and his collaborators systematically developed theory and methodology of sequential tests in the early 1940s to reduce the number of sampling inspections without compromising the reliability of the terminal decisions. The developments were admirably summarized in his pioneering book, Sequential Analysis in 1947.

The well-known Neyman-Pearson lemma (1937), offers a rule of thumb for when all the data is collected and its likelihood ratio known, is one of the most important theory in statistical hypothesis testing history. However, since the error probabilities decrease as the number of observations increase, we want to characterize the minimum number of observations needed to achieve desired levels of error. Rather than fixing n ahead of time, we consider a sequential approach to testing which continues to gather samples until a confident decision can be made. This idea is attributed to Wald, inspired by Neyman and Pearson's result, where he reformulated it as a sequential analysis problem which is called the sequential probability ratio test (SPRT).

Methodologically, researchers caught on and began applying sequential analysis to solve a wide range of practical problems from inventory, reliability, life tests, quality control, designs of experiments and multiple comparisons, to name a few. In the 1960s through 1970s, researchers in clinical trials realized the relevance of emerging adaptive designs and optimal stopping rules. Clinical trials continue to be an important beneficiary of some of the basic research in sequential methodologies. The basic research in clinical trials has also enriched the area of sequential sampling designs. The development in the next two decades was mirrored admirably in Ghosh (1970). More recent theoretical developments appear in Siegmund (1985).

A number of celebrated books already exist. We have mentioned Wald (1947) before. Additionally, one will find other volumes including Bechhofer et al. (1968), Ghosh (1970), Chow et al. (1971), Gibbons et al. (1977), Gupta and Panchapakesan (1979), Govindarajulu (1981), Ghosh and Sen (1991), Mukhopadhyay and Solanky (1994), Ghosh et al. (1997), Govindarajulu (2004), Mukhopadhyay and de Silva B.M. (2009). Other articles also worth mentioning. For example, Stein (1945), Stein (1949), Anscombe (1952), Ray (1957), Robbins (1959), Chow and Robbins (1965), Woodroofe (1977), Lai and Siegmund (1977), Lai and Siegmund (1979). Two important articles emphasized the concepts of first-order and second-order efficiencies: Mukhopadhyay (1980), Ghosh and Mukhopadhyay (1981). Govindarajulu (2004) derived closed-form expressions for the effective type I error probability and the power at the specified alternative and includes codes for some selected computer programs. Cho (2007) considered a risk-efficient sequential point estimator for the ratio of two binomial proportions based on maximum likelihood estimation under squared error loss and cost proportional to the observations. Cho and Govindarajulu (2008) presented a sequential method for obtaining approximate confidence limits for the ratio of two independent binomial proportions.

1.2.2 Sequential Estimation

In contrast, sequential estimation has received scant attention. Sequential estimation refers to estimation methods in sequential analysis where the sample size is not fixed in advance. Instead, data is evaluated as it is collected, and further sampling is stopped in accordance with a pre-defined stopping rule as soon as significant results are observed. At that time, notably, Govindarajulu (1981) tried to combine sequential hypothesis testing and estimation problems. In addition, it is worth mentioning that sequential nonparametric methods have been treated by Sen (1981), Sen (1985), which contain some accounts of sequential estimation.

We notice interesting and newer applications of sequential methodologies today. This is especially so in contemporary statistical challenges in agriculture, clinical trials, data mining, finance, gene mapping, multiple comparisons and so on.

Generally, sequential methodology is known to be more efficient than a fixed-sample size method in many aspects. In some situations, sequential methodologies may be essential because no fixed sample size methodology would work or available. We believe that the theory and practice of sequential analysis should ideally move forward together as partners. To explain why sequential estimation is needed, we will take a look at the fixed-width confidence interval estimation problem in the next chapter.

Sequential analysis is also related to multistage ranking and selection methodologies, or more generally speaking, multiple comparison problems. Some advanced books devoted exclusively to the area of multiple comparisons are available. Bechhofer (1954) developed a pioneering selection methodology by advancing Steins (1945, 1949) two-stage sampling strategy. One may refer to Hochberg and Tamhane (1987) and Hsu (1996). The interface between sequential analyses and selection problems is available in the advanced book by Mukhopadhyay and Solanky (1994). For example, there is no fixed-sample size methodologies in selecting the best treatment with preassigned probability of correct selection. However, two-stage sequential methodologies can deliver.

1.3 Application of Sequential Estimation

In pharmaceutical areas, controlling clinical trials is a very important issue. There is a strong ethical and economic obligation for the researchers to analyze data periodically for evidence of efficacy and safety over the course of the trial. As compelling evidence emerges, either favoring or disfavoring the new therapy, it may become ethically or economically necessary to terminate the trial before schedule. Although periodic evaluation of data is a frequent and necessary practice in drug development, particular statistical problems of multiple testing may appear. Classical clinical trial designs do not formally provide the option for early termination. Rather, classical designs consider only fixed-sample-size trials. When data from a fixed-sample size trial are analyzed repeatedly, the true type I and type II error probabilities associated with the testing of hypotheses will be inflated above the pre-specified levels. To control the undesirable escalation of the true error probabilities, sequential methods were developed.

In the previous sections we mentioned that there are many estimating problems that cannot be solved by any fixed sample size method. However, these problems can be resolved by implementing dynamic sampling design schemes such as sequential methods or two-stage procedures:

1. Risk-efficient point estimator of exponential family.

2. Ranking and selection methodologies in deciding the best populations or subsets.

3. Constructing a fixed-width confidence interval for an unknown mean with two preassigned length $2d$ and level of confidence $1 - \alpha$ for a normal distribution.

We will discuss the third example in more details in next chapter.

CHAPTER 2

SEQUENTIAL METHOD

2.1 Introduction

In this chapter, we start with basic definitions, and study the formulation and development of the procedure, expectations and variances, asymptotic distributions of the risk ratio and measure of reduction, as well as the evaluation of the procedure.

Sequential procedures are different from other statistical procedures in sampling designs. When a researcher gathers information regarding the parameter θ, the researcher has an option of looking at a sequence of observations one at a time and decide whether to stop sampling or to continue sampling before making a decision. Thus, the total number of observations denoted by $N(> 0)$, is a random variable.

2.2 Definitions of Distributions and Measures

Definition 2.1 (Bernoulli Distribution) A random variable X is said to have a Bernoulli distribution with p if the probability mass function is given by

$$P(X = x) = p(x) = p^x(1-p)^{1-x},\ x = 0, 1$$

where $0 \leq p \leq 1$.

Definition 2.2 (Binomial Distribution) A random variable X is said to have a binomial distribution with parameters n and p if the probability mass function is given by

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, 0 < p < 1$$

where n is the number of total trials, the binomial coefficient $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, $p =$ probability of success and $1 - p =$ probability of failure. We denote this by $X \sim Bin(n, p)$. By definition,

$$E(X) = np, \ var(X) = np(1-p).$$

The binomial probability refers to the probability that a binomial experiment consisting of n trials results in exactly x successes with probability of success p in Bernoulli trials. The sum of a sequence of independent and identically distributed (i.i.d.) Bernoulli variables follows a binomial distribution.

2.3 Formulations of the Proposed Procedure

Along with the fixed-sample methods, in this thesis we consider the sequential method and propose to obtain approximate confidence intervals and corresponding optimal sample sizes for the risk ratio and the measure of reduction.

Suppose we have two samples of size n from two independent Bernoulli populations with probabilities p_0 and p_1, respectively, say X_1, X_2, ..., X_n, and Y_1, Y_2, ..., Y_n. Let us define

$$R = \sum_{i=1}^{n} X_i \quad and \quad S = \sum_{i=1}^{n} Y_i.$$

Then, $\sum_{i=1}^{n} X_i$ follows the binomial distribution with parameters (n, p_0), and $\sum_{i=1}^{n} Y_i$

follows the binomial distribution with parameters (n, p_1), respectively. That is,

$$R \sim Bin(n, p_0) \quad and \quad S \sim Bin(n, p_1)$$

.

Definition 2.3 (Risk Ratio) The risk ratio for two binomial variates is defined by:

$$\theta = \frac{p_1}{p_0}.$$

Then, the estimate of the risk ratio for the two sample proportions $\hat{p}_0$ and $\hat{p}_1$, is:

$$\hat{\theta} = \frac{\hat{p}_1}{\hat{p}_0}, \tag{2.1}$$

where $\hat{p}_0 = R/n$ and $\hat{p}_1 = S/n$.

Since there does not exist an unbiased estimator of the measure θ, we define the modified $\hat{\theta}_n$ to avoid the case of undefined $\hat{\theta}_n$ when $R = 0$:

$$\hat{\theta}_n = \frac{S}{R + 1/n}. \tag{2.2}$$

Definition 2.4 (Measure of Reduction) The measure of reduction for two independent binomial variates is defined to be:

$$\rho = \frac{p_0 - p_1}{p_0} = 1 - \frac{p_1}{p_0}, -\infty < \rho \leq 1.$$

Then, the estimator for the measure of reduction for the two sample proportions $\hat{p}_0$ and $\hat{p}_1$, is:

$$\hat{\rho} = \frac{\hat{p}_0 - \hat{p}_1}{\hat{p}_0} = 1 - \frac{\hat{p}_1}{\hat{p}_0}, \tag{2.3}$$

where $\hat{p}_0 = R/n$ and $\hat{p}_1 = S/n$.

Similarly, to avoid the case of undefined $\hat{\rho}_n$ when $R = 0$:

$$\hat{\rho}_n = \frac{(R + 1/n) - S}{R + 1/n} = 1 - \frac{S}{R + 1/n}. \tag{2.4}$$

By definition, ρ is a relative figure of merit for measuring the reduction between two binomial proportions. Depending on the value of $\rho, -\infty < \rho \leq 1$, the measure of reduction has more pratical usage in comparing two populations. First, if ρ approaches one, the risk (of being infected) reduction is complete. Second, when ρ gets close to zero, this indicates that there is no risk reduction achieved. Lastly, if ρ is negative, this implies that a certain degree of reduction is made.

2.4 Asymptotic Properties of the Estimator $\hat{\theta}_n$

In this section, we study the fundamental properties based on the first two moments of the estimators $\hat{\theta}_n$ and $\hat{\rho}_n$ for further investigation.

2.4.1 Expectations

Now consider the expectation of the estimator $\hat{\theta}_n$. By definitions and independence, we have

$$\begin{aligned} E(\hat{\theta}_n) &= E\left(\frac{S}{R + 1/n}\right) \\ &= E(S)E\left(\frac{1}{R + 1/n}\right) = np_1 E\left(\frac{1}{R + 1/n}\right). \end{aligned} \tag{2.5}$$

Let $U_n = \frac{R-np_0+1/n}{np_0}$, so $\frac{1}{R+1/n} = \frac{1}{np_0}(1+U_n)^{-1}$, noting that for $U_n \neq 1$, $(1+U_n)^{-1} = 1 - U_n + (U_n)^2 - (U_n)^3 + (U_n)^4 - (U_n)^5(1+U_n)^{-1}$, we get that

$$\begin{aligned} E\left(\frac{1}{R+1/n}\right) &= E\left[\frac{1}{np_0}(1+U_n)^{-1}\right] \\ &= \frac{1}{np_0}E(1 - U_n + (U_n)^2 - (U_n)^3 + (U_n)^4 - (U_n)^5(1+U_n)^{-1}), \quad (2.6) \end{aligned}$$

$$\begin{aligned} E(U_n) &= E\left(\frac{R-np_0+1/n}{np_0}\right) = \frac{E(R)-np_0+1/n}{np_0} \\ &= \frac{1/n}{np_0} = \frac{1}{n^2p_0} \\ E(U_n)^2 &= Var(U_n) + E(U_n)^2 = Var\left(\frac{R-np_0+1/n}{np_0}\right) + E\left(U_n\right)^2 \\ &= \frac{Var(R)}{(np_0)^2} + E\left(U_n\right)^2 = \frac{1-p_0}{np_0} + \left(\frac{1}{n^2p_0}\right)^2 \end{aligned}$$

By Theorem 2 in Von Bahr (1969), if X_j is a sequence of i.i.d. random variables such that for a positive integer $k \geq 2$, $E(| X_1 |^k) < \infty$, then

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

where $\sigma^2 = Var(X_1)$ and z is a standard normal random variable. This implies that for each positive integer k

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

and

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

hence

$$E[| n^{1/2}p_0U_n |^k] = O(1) \quad (2.7)$$

By Eq.(2.7) and $k = 3$, 4 respectively, $E(U_n^3) = o(\frac{1}{n})$ and also $E(U_n^4) = o(\frac{1}{n})$.

Moreover, since $1 + U_n \geq \frac{1}{n^2 p_0}$,

hence,

$$
\begin{aligned}
E\left(\frac{1}{R+1/n}\right) &= \frac{1}{np_0}\left[1 - \frac{1}{n^2 p_0} + \frac{1-p_0}{np_0} + \left(\frac{1}{n^2 p_0}\right)^2\right] + o(n^{-2}) \\
&= \frac{1}{np_0}\left(1 + \frac{1-p_0}{np_0}\right) + o(n^{-2})
\end{aligned}
\tag{2.8}
$$

Combining Eq.(2.5) and Eq.(2.8) we get

$$
\begin{aligned}
E(\hat{\theta}_n) = np_1 E\left(\frac{1}{R+1/n}\right) &= \frac{p_1}{p_0}\left(1 + \frac{1-p_0}{np_0}\right) + o(n^{-1}) \\
= \hat{\theta} + o(n^{-1})
\end{aligned}
$$

Therefore, $\hat{\theta}_n$ is an asymptotically unbiased estimator of $\hat{\theta}$.

Now, we are able to get the expectation of the measure of reduction $\hat{\rho}$ easily,

$$
\begin{aligned}
E(\hat{\rho}_n) &= E\left(\frac{R+1/n-S}{R+1/n}\right) = E\left(1 - \frac{S}{R+1/n}\right) \\
&= 1 - E(S)E\left(\frac{1}{R+1/n}\right) = 1 - np_1 E\left(\frac{1}{R+1/n}\right) \\
&= 1 - \hat{\theta} + o(n^{-1}) = \hat{\rho} + o(n^{-1}).
\end{aligned}
\tag{2.9}
$$

Therefore, $\hat{\rho}_n$ is an asymptotically unbiased estimator of $\hat{\rho}$.

2.4.2 Asymptotic Variance

To get the variance of $\hat{\theta}$, we consider the maximum likelihood estimates of θ and p_0 and their information matrix. From the observed sample of n pairs of (X_i, Y_i), $i = 1, 2, ., n$, the likelihood function is

$$L\left(p_0, p_1\right) \propto p_0^r \left(1-p_0\right)^{n-r} p_1^s \left(1-p_1\right)^{n-s}$$

$$
\begin{aligned}
L(\theta, p_0) &\propto p_0^r (1-p_0)^{n-r} (p_0\theta)^s (1-p_0\theta)^{n-s} \\
&= p_0^{r+s} (1-p_0)^{n-r} \theta^s (1-p_0\theta)^{n-s}, \qquad (2.10)
\end{aligned}
$$

the log-likelihood function of Eq.(2.10) is then

$$
\begin{aligned}
\mathbf{l}(\theta, p_0) \propto &(r+s)\log(p_0) + (n-r)\log(1-p_0) \\
&+ s\log(\theta) + (n-s)\log(1-p_0\theta) \qquad (2.11)
\end{aligned}
$$

By setting the first derivatives to be zero, the maximum likelihood estimators (MLE) of θ and p_0 can be found:

$$\hat{\theta}_{MLE} = \frac{s}{np_0}$$

and

$$\hat{p}_0 = \frac{r}{n}.$$

It should be noted that since

$$E\left(\hat{\theta}_{MLE}\right) = E\left(\frac{s}{np_0}\right) = \frac{p_1}{p_0} = \theta,$$

the MLE of θ, $\hat{\theta}_{MLE}$ is an unbiased estimator. And so is $\hat{p}_0$, because

$$E(\hat{p}_0) = E\left(\frac{R}{n}\right) = p_0.$$

To obtain the variance of MLE of θ, we consider

$$
\begin{aligned}
\frac{\partial \mathbf{I}(\theta, p_0)}{\partial \theta} &= \frac{s}{\theta} - \frac{(n-s)p_0}{1-p_0\theta} \\
\frac{\partial \mathbf{I}(\theta, p_0)}{\partial p_0} &= \frac{r+s}{p_0} - \frac{n-r}{1-p_0} - \frac{(n-s)\theta}{1-p_0\theta}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \mathbf{I}(\theta, p_0)}{\partial \theta^2} &= -\frac{s}{\theta^2} - \frac{(n-s)\, p_0^2}{(1-p_0\theta)^2} \\
\frac{\partial^2 \mathbf{I}(\theta, p_0)}{\partial \theta \partial p_0} &= \frac{-(n-s)(1-p_0\theta) - p_0(n-s)\theta}{(1-p_0\theta)^2} = -\frac{n-s}{(1-p_0\theta)^2} \\
\frac{\partial^2 \mathbf{I}(\theta, p_0)}{\partial p_0^2} &= -\frac{r+s}{p_0^2} - \frac{n-r}{(1-p_0)^2} - \frac{(n-s)\theta^2}{(1-p_0\theta)^2}.
\end{aligned}
$$

Then, from the log-likelihood function, Fisher's information matrix about (θ, p_0) is given by

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

So,

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

Therefore, from the above equation, the asymptotic variance of $\hat{\theta}_{MLE}$ is

$$
\begin{aligned}
Var\left(\hat{\theta}_{MLE}\right) &= \frac{\theta(1-p_0)(1-p_1)}{n}\left[\frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{\theta}{1-p_1}\right)\right] \\
&= \frac{\theta(1+\theta-2\theta p_0)}{np_0}.
\end{aligned} \tag{2.12}
$$

Now, we consider the asymptotic variance of $\hat{\theta}_n = \frac{S}{R+1/n}$.

$$
Var\left(\hat{\theta}_n\right) = Var\left(\frac{S}{R+1/n}\right) = E\left(\frac{S}{R+1/n}\right)^n - \left[E\left(\frac{S}{R+1/n}\right)\right]^2.
$$

Since

$$
E\left(\frac{S}{R+1/n}\right) = \frac{p_1}{p_0}\left(1 + \frac{1-p_0}{np_0} - \frac{1}{n^2 p_0}\right) + o\left(n^{-2}\right), \tag{2.13}
$$

we only need to find $E\left(\frac{S}{R+1/n}\right)^2$.

$$E\left(\frac{S}{R+1/n}\right)^2 = E\left(S^2\right)E\left(\frac{1}{R+1/n}\right)^2$$
$$= \left[np_1(1-p_1)+n^2p_1^2\right]E\left(\frac{1}{R+1/n}\right)^2$$
$$= \left[np_1(1-p_1)+n^2p_1^2\right]\frac{1}{n^2p_0^2}E\left[(1+U_n)^{-2}\right]$$

Noting that for $U_n \neq 1$,

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

so,

$$E\left(\frac{1}{R+1/n}\right)^2$$
$$= \frac{1}{n^2p_0^2}E\left[1-2U_n+3U_n^2-4U_n^3+5U_n^4-(6U_n^5+5U_n^6)(1+U_n)^{-2}\right]. \tag{2.14}$$

In section 2.2.1, we found that

$$E(U_n) = E\left(\frac{R-np_0+1/n}{np_0}\right)$$
$$= \frac{E(R-np_0+1/n)}{np_0} = \frac{1/n}{np_0} = \frac{1}{n^2p_0}$$
$$E\left(U_n^2\right) = Var(U_n)+E(U_n)^2$$
$$= Var\left(\frac{R-np_0+1/n}{np_0}\right)+EU_n{}^2$$
$$= \frac{VarR}{(np_0)^2}+E(U_n)^2 = \frac{1-p_0}{np_0}+\left(\frac{1}{n^2p_0}\right)^2.$$

By Eq.(2.7) and k=3,4 respectively, $E\left(U_n^3\right) = o\left(n^{-1}\right)$ and $E\left(U_n^4\right) = o\left(n^{-1}\right)$. Moreover, since $1+U_n \geq \frac{1}{2np_0}$,

$$(np_0)^{-2}E\left[\mid \left(6U_n^5+5U_n^6\right)(1+U_n)^{-1}\mid\right] \leq 4E\left[6\mid U_n^5\mid +5\mid U_n^6\mid\right] = o\left(n^{-2}\right)$$

plug these into Eq.(2.14), we get

$$E\left(\frac{1}{R+1/n}\right)^2=\frac{1}{n^2p_0^2}\left[1-\frac{2}{n^2p_0}+\frac{3-3p_0}{np_0}+3\left(\frac{1}{n^2p_0}\right)^2\right]+o\left(n^{-3}\right). \qquad (2.15)$$

Combining Eq.(2.13)-(2.15)

$$\begin{aligned}
Var\left(\hat{\theta}_n\right) &= E\left(\frac{S}{R+1/n}\right)^2-\left[E\left(\frac{S}{R+1/n}\right)\right]^2\\
&= \left[np_1(1-p_1)+n^2p_1^2\right]\frac{1}{n^2p_0^2}E\left[(1+U_n)^{-2}\right]-\left[\frac{p_1}{p_0}\left(1+\frac{1-2p_0}{np_0}+\frac{1}{4n^2p_0^2}\right)\right]^2\\
&= \left[np_1(1-p_1)+n^2p_1^2\right]\frac{1}{n^2p_0^2}\left[1-\frac{2}{n^2p_0}+\frac{3-3p_0}{np_0}+3\left(\frac{1}{n^2p_0}\right)^2+o\left(n^{-1}\right)\right]\\
&- \left[\frac{p_1}{p_0}\left(1+\frac{1-p_0}{np_0}\right)+o\left(n^{-1}\right)\right]^2\\
&= \left[\frac{\theta(1-p_1)}{np_0}+\theta^2\right]\left[1+\frac{3-3p_0}{np_0}+o\left(n^{-1}\right)\right]\\
&- \theta^2\left[1+\frac{2-2p_0}{np_0}+o\left(n^{-1}\right)\right]\\
&\approx \left[\frac{\theta(1-p_1)}{np_0}+\theta^2\right]\left[1+\frac{3-3p_0}{np_0}\right]+\theta^2\frac{1-p_0}{np_0}+o\left(n^{-1}\right)\\
&= \frac{\theta(1-p_1)}{np_0}\left[1+\frac{3-3p_0}{np_0}\right]+\theta^2\frac{1-p_0}{np_0}+o\left(n^{-1}\right)\\
&= \frac{\theta(1-p_1)}{np_0}+\theta^2\frac{1-p_0}{np_0}+o\left(n^{-1}\right)\\
&\approx \frac{\theta(1+\theta-2\theta p_0)}{np_0}. \qquad (2.16)
\end{aligned}$$

From the results of Eq.(2.12) and Eq.(2.16), we see the two estimators have the same variance. Hence, we conclude that two estimators, $\hat{\theta}_{MLE}$ and $\hat{\theta}_n$ are asymptotically equivalent for large n.

Using Slutsky's theorem and for sufficiently large n, $\sqrt{n}\left(\hat{\theta}_n-\theta\right)$ converges in distribution to $N\left(0,\sigma^2\right)$ where

$$\sigma^2=\frac{\theta(1+\theta-2\theta p_0)}{p_0}.$$

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 can use the same method to find the asymptotic variance of $\hat{\rho}_{MLE}$ and $\hat{\rho}_n$.

The likelihood function for (ρ, p_0) is

$$\begin{aligned} L(\rho, p_0) &\propto p_0^r (1-p_0)^{n-r} [p_0 (1-\rho)]^s [1 - p_0 (1-\rho)]^{n-s} \\ &= p_0^{r+s} (1-p_0)^{n-r} (1-\rho)^s [1 - p_0 (1-\rho)]^{n-s}, \end{aligned} \tag{2.17}$$

the log-likelihood function of Eq.(2.17) is then

$$\begin{aligned} \mathbf{l}(\rho, p_0) &\propto (r+s) \log(p_0) + (n-r) \log(1-p_0) \\ &= s \log(1-\rho) + (n-s) \log[1 - p_0 (1-\rho)]. \end{aligned} \tag{2.18}$$

Solve the maximum likelihood estimators(MLE) of ρ and p_0 by setting the first derivatives to be zero

$$\hat{\rho}_{MLE} = 1 - \frac{s}{n p_0}$$

and

$$\hat{p}_0 = \frac{r}{n}.$$

The measure of reduction ρ is an induced measure from the risk ratio $\theta = p_1/p_0$. The MLE of ρ, $\hat{\rho}_{MLE}$ is asymptotic unbiased estimator of ρ. To get the variance of the MLE of ρ, consider

$$
\begin{aligned}
\frac{\partial \mathbf{I}\left(\rho, p_0\right)}{\partial \rho} &= \frac{-s}{1-\rho} + \frac{\left(n-s\right) p_0}{1-p_0+p_0\rho} \\
\frac{\partial \mathbf{I}\left(\rho, p_0\right)}{\partial p_0} &= \frac{r-s}{p_0} - \frac{n-r}{1-p_0} - \frac{\left(n-s\right)\left(1-\rho\right)}{1-p_0+p_0\rho} \\
\frac{\partial^2 \mathbf{I}\left(\rho, p_0\right)}{\partial \rho^2} &= -\frac{s}{\left(1-\rho\right)^2} - \frac{\left(n-s\right) p_0^2}{\left(1-p_0+p_0\rho\right)^2} \\
\frac{\partial^2 \mathbf{I}\left(\rho, p_0\right)}{\partial \rho \partial p_0} &= \frac{-(n-s)\left(1-p_0+p_0\rho\right) - p_0(n-s)(1-\rho)}{\left(1-p_0+p_0\rho\right)^2} \\
&= \frac{n-s}{\left(1-p_0+p_0\rho\right)^2} \\
\frac{\partial^2 \mathbf{I}\left(\rho, p_0\right)}{\partial p_0^2} &= -\frac{r-s}{p0^2} - \frac{n-r}{\left(1-p_0\right)^2} - \frac{(n-s)(1-\rho)^2}{\left(1-p_0+p_0\rho\right)^2}.
\end{aligned}
$$

Then, from the log-likelihood function, Fisher's information matrix about (ρ, p_0) is given by

$$
\begin{aligned}
\mathbf{I}(\rho, p_0) &= \left[\begin{array}{cc} E\left(-\frac{\partial^2 \mathbf{I}(\rho,p_0)}{\partial \rho^2}\right) & E\left(-\frac{\partial^2 \mathbf{I}(\rho,p_0)}{\partial \rho \partial p_0}\right) \\ E\left(-\frac{\partial^2 \mathbf{I}(\rho,p_0)}{\partial \rho \partial p_0}\right) & E\left(-\frac{\partial^2 \mathbf{I}(\rho,p_0)}{\partial p_0^2}\right) \end{array}\right] \\
&= n\left[\begin{array}{cc} \frac{p_0^2}{p_1(1-p_1)} & \frac{1}{1-p_1} \\ \frac{1}{1-p_1} & \frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{1-\rho}{1-p_1}\right) \end{array}\right].
\end{aligned}
$$

So,

$$
\mathbf{I}(\rho, p_0) = \frac{(1-\rho)(1-p_0)(1-p_1)}{n}\left[\begin{array}{cc} \frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{1-\rho}{1-p_1}\right) & -\frac{1}{1-p_1} \\ -\frac{1}{1-p_1} & \frac{p_0^2}{p_1(1-p_1)} \end{array}\right].
$$

Therefore, from the above equation, the asymptotic variance of $\hat{\rho}_{MLE}$ is

$$
\begin{aligned}
Var\left(\hat{\rho}_{MLE}\right) &= \frac{(1-\rho)(1-p_0)(1-p_1)}{n}\left[\frac{1}{p_0}\left(\frac{1}{1-p_0} + \frac{1-\rho}{1-p_1}\right)\right] \\
&= \frac{(1-\rho)(2-\rho+2\rho p_0-2p_0)}{np_0}. \qquad (2.19)
\end{aligned}
$$

The asymptotic variance of $\hat{\rho}_n$ can be simply found from Eq.(2.16)

$$\begin{aligned} Var\left(\hat{\rho}_n\right) &= Var\left(1-\hat{\theta}_n\right) = Var\left(\hat{\theta}_n\right) \\ &\approx \frac{\theta(1+\theta-2\theta p_0)}{np_0} \\ &= \frac{(1-\rho)[(2-\rho)-2p_0(1-\rho)]}{np_0} \\ &= \frac{(1-\rho)(2-\rho+2\rho p_0-2p_0)}{np_0}. \end{aligned} \tag{2.20}$$

Hence, $\hat{\rho}_{MLE}$ and $\hat{\rho}_n$ are also asymptotically equivalent for large n. Similarly, $\sqrt{n}\left(\hat{\rho}_n-\rho\right)$ converges in distribution to $N\left(0,\sigma^2\right)$ where

$$\sigma^2 = \frac{(1-\rho)(2-\rho+2\rho p_0-2p_0)}{p_0}.$$

2.5 Procedure and the Stopping Rule

Our goal is to develop the procedure and to construct an interval of specified width $2d$ with confidence coefficient $1-\alpha$ for the risk ratio θ, and measure of reduction ρ. That is,

$$P\left\{\mid \hat{\theta}-\theta \mid\leq d\right\} \geq 1-\alpha. \tag{2.21}$$

Apparently, one can make an inference of θ using a statistic T_n from a random sample of fixed size $n, (x_1, x_2, ..., x_n)$, which is referred to as the fixed-sample size method in contrast to any dynamic sampling plans. Therefore, in sequential sampling to infer θ, we need to consider a pair (N, T_N) where N is called the random sampling time.

To explore the rationale of adopting the sequential strategy, let's take a look at a fixed-width confidence interval problem with an unknown mean, which was mentioned in Section 1.3.

Consider a random sample of size n ($n > 2$), $(x_1, x_2, ..., x_n)$ from a Normal population with parameters μ and σ^2, assuming that both μ and σ^2 are unknown.

Suppose one wishes to construct a $(1-\alpha)100\%$ confidence interval I for μ with length $2d$ and the probability of the interval $P_{\mu,\sigma}(\mu \in I) \geq 1-\alpha$, where $d \geq 0$ and $0 < \alpha < 1$ are preassigned. However, Dantzig (1940) showed that the fixed-width confidence interval problem cannot be solved by any fixed-sample size method.

The Interval $I_n = [\bar{X}_n - d, \bar{X}_n + d]$ has a probability

$$\begin{aligned}
&P_{\mu,\sigma}(\mu \in I) = 2\Phi\left(\frac{d}{\sigma/\sqrt{n}}\right) - 1 \\
\Leftrightarrow &2\Phi\left(\frac{d}{\sigma/\sqrt{n}}\right) - 1 \geq 1 - \alpha = 2\Phi(\xi) - 1. \\
\Leftrightarrow &\frac{d}{\sigma/\sqrt{n}} = \frac{\sqrt{n}d}{\sigma} \geq \xi. \\
\Leftrightarrow &n \geq \frac{\xi^2\sigma^2}{d^2}
\end{aligned}$$

where $d = t_{n-1}\frac{S_n}{\sqrt{n}}$ and $\xi = \frac{d\sqrt{n}}{\sigma}$. In conclusion, n is the smallest integer $\geq \frac{a^2\sigma^2}{d^2}$ = say n^*, where n^* is the optimal fixed-sample size required to construct I_n for μ if σ had been known. Since σ is unknown, this can not be achieved.

In this case, we use the sample variance S_n^2 replacing σ^2, so the stopping rule can be stated as: $N = N(d) =$ smallest integer n ($> m$), where m is the initial sample size, such that, $n \geq a^2 S_n^2/d^2$.

Now, we will apply the above method to analyze the risk ratio and the measure of reduction for two binomial variates. First, we need to determine the optimal sample size n that satisfies

$$P\{\mid \hat{\theta} - \theta \mid \leq d\} = P\{\sqrt{n} \mid \hat{\theta} - \theta \mid /\sigma \leq d\sqrt{n}/\sigma\} \geq 1 - \alpha,$$

since $\sqrt{n}(\hat{\theta}_n - \theta) \sim N(0, \sigma^2)$ so,

$$2\Phi(d\sqrt{n}/\sigma) - 1 \geq 1 - \alpha, \tag{2.22}$$

where $\Phi(x)$ is the CDF of a standard normal distribution.

Then, Eq.(2.22) is equivalent to:

$$d\sqrt{n}/\sigma \geq z_{(2-\alpha)/2} \equiv z, \tag{2.23}$$

for specified d (> 0) where $\Phi(z_{(2-\alpha)/2}) = (2 - \alpha)/2$.

Consequently, we have

$$n \geq (z\sigma/d)^2. \tag{2.24}$$

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,

$$n^* = [(z\sigma/d)^2] + 1 \tag{2.25}$$

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

Recall that $\sigma^2 = \theta(1 + \theta - 2\theta p_0)/p_0$, since both θ and p_0 are unknown, we are not able to determine the optimal fixed-sample size. But sequentially, we could come up with the following stopping rule: Stop sampling at observation

$$N = \inf_n\{n \geq m : n \geq z^2\hat{\sigma}_n^2/d^2\} \tag{2.26}$$

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

Consequently, a $(1-\alpha)100\%$ confidence interval with length $2d$ for θ is given by

$$(\hat{\theta}_N - d, \hat{\theta}_N + d) \tag{2.27}$$

2.6 Properties and Evaluation of the Procedure

In this section we study the desirable properties of the stopping rule we proposed. One of the most primary properties is concerned about the stopping time. Because, in the sequential method the sample has to be formulated at a certain stage. Otherwise, the proposed procedure is meaningless. Second aspect is the properties about the fact how much the proposed procedure is achieved toward the inferential goals. These properties are called the (asymptotic) consistency and (asymptotic) efficiency of the proposed procedure, respectively.

Definition 2.5 (Asymptotic Consistency) An estimator $\hat{\theta}_N$ of θ is said to be asymptotically consistent if, for any preassigned significant level α, $\lim_{d\to 0} P\{|\hat{\theta}_N - \theta| \leqslant d\} \geq 1-\alpha$.

Definition 2.6 (Efficiency) Under the above set up, $\hat{\theta}_N$ is said to be asymptotically efficient if $\lim_{d\to 0} E(N)/n^* = 1$.

2.6.1 Finite Sure Termination

The following result establishes the finite sure termination holds for the proposed sequential procedure.

Theorem 2.1. Let N denote the stopping time associated with the proposed procedure. Then, $P(N=\infty)=0$.

Proof. Using the stopping rule in Eq. (2.21)

$$P(N=\infty)=\lim_{n\to\infty}P(N>n)\leq\lim_{n\to\infty}P(n\leq z^2\hat{p}_n^2/d^2)=0$$

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

2.6.2 First Order Asymptotic

To evaluate the proposed procedure, we study the asymptotic behavior of the procedure when the sample size is sufficiently large. Therefore, since the random stopping time N is a function of d, one can have large enough n by letting d gets small.

In order to fit the desirable criteria, the stopping rule N in Eq. (2.26) can be written as follow:

$$N=\inf_n\left\{n\geq m: n\geq\frac{z^2}{d^2}\frac{\hat{\theta}(1+\hat{\theta}-2\hat{\theta}\hat{p}_0)}{\hat{p}_0}\right\} \tag{2.28}$$

let

$$f(n)=n,$$

$$W_n=\frac{\hat{\theta}(1+\hat{\theta}-2\hat{\theta}\hat{p}_0)}{\theta(1+\theta-2\theta p_0)}\frac{p_0}{\hat{p}_0},$$

$$t = (t/d)^2\theta(1 + \theta - 2\theta p_0)/p_0.$$

Then, Eq. (2.28) takes the form:

$$N = N(t) = min_n\{n \geq m : W_n \leq f(n)/t\}. \tag{2.29}$$

Hence, W_n is a sequence of random variables such that W_n is positive (a.s.) and converges a.s to 1 as n approaches infinity, because $\hat{p}_{0,n} \to p_0$ (a.s.) and $\lim_{n\to\infty} \hat{\theta}_n/\theta = 1$.

Furthermore, we see 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 invoking the results of Chow and Robbins (1965), the first-order asymptotic for the properties of the proposed sequential procedure are obtained as follows:

Theorem 2.2. When d goes to zero, we have

(i) $N/n^* = 1$ a.s.,

(ii) $P\{|\hat{\theta}_N - \theta| \leqslant d\} \geq 1 - \alpha$

(iii) $E(N)/n^* = 1$.

Proof. (i) $\lim_{d\to 0} N = \infty$ a.s., $\lim_{d\to 0} E(N) = \infty$, since from the definition of N, $\lim_{d\to 0} N \geq \lim_{d\to 0} z^2\hat{\sigma}_n^2/d^2$ a.s.. Then, from $N = \inf_n\{n \geq m : n \geq z^2\hat{\sigma}_n^2/d^2\}$ we have $N - 1 \leq z^2\hat{\sigma}_{N-1}^2/d^2$,

therefore,

$$\frac{z^2\sigma_N^2/d^2}{z^2\sigma^2/d^2} \leq \frac{N}{n^*} \leq \frac{z^2\sigma_{N-1}^2/d^2 + 1}{z^2\sigma^2/d^2}. \tag{2.30}$$

From which it is easy to see that

$$\frac{\sigma_N^2}{\sigma^2} \leq \frac{N}{n^*} \leq \frac{d^2}{z^2\sigma^2} + \frac{\sigma_{N-1}^2}{\sigma^2}. \tag{2.31}$$

Hence,

$$\lim_{d\to\infty} \frac{\sigma_N^2}{\sigma^2} \leq \lim_{d\to\infty} \frac{N}{n^*} \leq \lim_{d\to\infty} \left(\frac{d^2}{z^2\sigma^2} + \frac{\sigma_{N-1}^2}{\sigma^2} \right) = \lim_{d\to\infty} \frac{\sigma_{N-1}^2}{\sigma^2}.$$

However, the quantities on the extremes of the inequality tend to unity. Thus, $\lim_{d\to\infty} (N/n^*) = 1$.

For (ii), this is directly from the set up of this procedure, since we construct this interval of width $2d$ with confidence coefficient $1-\alpha$ for the risk ratio θ. Mathematically, that is

$$P\left\{ \mid \hat{\theta} - \theta \mid \leq d \right\} \geq 1 - \alpha.$$

(iii), using the large derivation priciple (LDP) and the properties of Eq.(2.29), the proof can be done and we refer details directly to Cho and Govindarajulu (2008).

2.7 Numerical Studies

2.7.1 Confidence Intervals Based on the Proposed Method

Monte Carlo experimentation is carried out to investigate the finite-sample behavior of the risk ratio and measure of reduction we have devised. Selected values for p_0 and p_1 were chosen to generate the data sets consisting of sequences of binomial variables based on a predetermined fixed number of trials for each case. Two sample proportions of p_0 and p_1 are computed and the point estimator of the measure of reduction ρ is also calculated with replications of 10000. The results of the experiment are summarized in the following tables with the expected stopping time $E(N)$, starting sample size m, optimal sample sizes n^* and the coverage probability (CP) are shown with specified width d for the confidence level $1-\alpha$, $1-\alpha = 0.90$ and 0.95.

Table 2.1. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.499	1.000	.000	.10	(.900, 1.100)	.868	521.69	542
.499	.500	1.001	-.001	.20	(.801, 1.201)	.850	129.24	136
.501	.502	1.001	-.001	.30	(.701, 1.301)	.825	54.72	61
.503	.502	.997	.003	.40	(.597, 1.397)	.786	28.76	34
.499	.499	.999	.001	.50	(.499, 1.499)	.774	18.61	22
.502	.498	.993	.007	.60	(.393, 1.593)	.855	13.73	15
.498	.498	1.000	.000	.70	(.300, 1.700)	.930	11.13	12

In addition, plot of coverage probabilities and values of d (as getting smaller) are given in the following figure.

Figure 2.1. Plot of coverage probability against d with $p_0 = 0.5, p_1 = 5, \alpha = 0.05$.

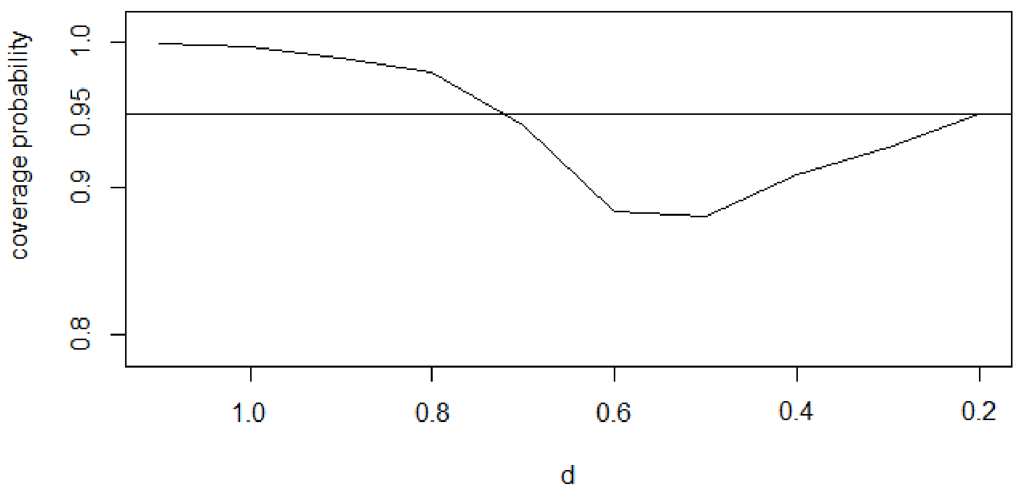


Table 2.2. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.500	1.000	.000	.10	(.900, 1.100)	.911	737.17	769
.500	.500	1.001	-.001	.20	(.801, 1.201)	.905	183.04	192
.500	.500	1.000	.000	.30	(.700, 1.300)	.885	79.65	86
.500	.500	1.000	.000	.40	(.600, 1.400)	.859	43.02	49
.499	.501	1.005	-.005	.50	(.505, 1.505)	.839	27.18	31
.498	.498	1.000	-.000	.60	(.400, 1.600)	.866	18.35	22
.500	.502	1.003	-.003	.70	(.303, 1.703)	.943	14.13	16

Table 2.3. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.500	.999	.001	.10	(.899, 1.099)	.894	536.86	541
.501	.500	.998	.002	.20	(.798, 1.198)	.881	131.83	136
.500	.498	.997	.003	.30	(.697, 1.297)	.845	56.64	61
.502	.500	.996	.004	.40	(.596, 1.396)	.851	31.12	34
.499	.498	.999	.001	.50	(.499, 1.499)	.878	21.28	22
.499	.500	1.002	-.002	.60	(.402, 1.602)	.952	16.54	16

Table 2.4. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.500	1.000	.000	.10	(.900, 1.100)	.940	765.49	769
.500	.500	1.000	.000	.20	(.800, 1.200)	.935	189.39	192
.500	.499	.997	.003	.30	(.697, 1.297)	.910	81.82	86
.500	.500	1.000	.000	.40	(.600, 1.400)	.888	44.88	48
.500	.498	.996	.004	.50	(.496, 1.496)	.901	28.78	31
.499	.500	1.002	-.002	.60	(.402, 1.602)	.947	20.98	22

Table 2.5. For $\theta = 1.2$ and $\rho = -0.2$ when $p_0 = 0.5, p_1 = 0.6$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.600	1.200	-.200	.10	(1.100, 1.300)	.879	638.92	650
.500	.601	1.201	-.201	.20	(1.001, 1.401)	.858	157.66	163
.500	.600	1.200	-.200	.30	(.900, 1.500)	.828	66.75	73
.499	.600	1.202	-.202	.40	(.802, 1.602)	.805	35.91	41
.499	.600	1.203	-.203	.50	(.703, 1.703)	.827	22.45	27
.501	.598	1.195	-.195	.60	(.595, 1.795)	.843	16.20	19
.503	.600	1.193	-.193	.70	(.493, 1.893)	.894	12.52	14
.500	.602	1.204	-.204	.80	(.404, 2.004)	.940	10.35	11

Table 2.6. For $\theta = 1.2$ and $\rho = -0.2$ when $p_0 = 0.5, p_1 = 0.6$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.600	1.200	-.200	.10	(1.100, 1.300)	.940	911.19	922
.500	.600	1.200	-.200	.20	(1.000, 1.400)	.926	225.40	231
.500	.601	1.203	-.203	.30	(.903, 1.503)	.904	98.29	103
.499	.600	1.202	-.202	.40	(.802, 1.602)	.876	52.94	58
.501	.600	1.198	-.198	.50	(.698, 1.698)	.865	32.37	37
.499	.599	1.201	-.201	.60	(.601, 1.801)	.859	21.81	26
.501	.599	1.197	-.197	.70	(.497, 1.897)	.870	16.61	19
.500	.599	1.197	-.197	.80	(.397, 1.997)	.936	13.20	15

Figure 2.2. Plot of coverage probability against d with $p_0 = 0.5, p_1 = 0.6, \alpha = 0.05$.

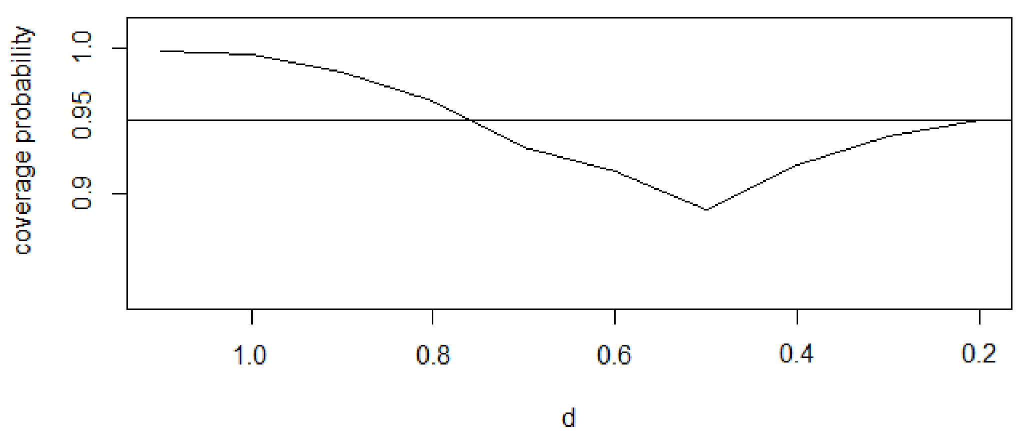


Table 2.7. For $\theta = 1.2$ and $\rho = -0.2$ when $p_0 = 0.5, p_1 = 0.6$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.600	1.200	-.200	.10	(1.100, 1.300)	.899	646.35	650
.501	.600	1.197	-.197	.20	(.997, 1.397)	.875	158.30	163
.500	.600	1.200	-.200	.30	(.900, 1.500)	.855	68.31	73
.501	.602	1.201	-.201	.40	(.801, 1.601)	.836	37.42	41
.504	.599	1.188	-.188	.50	(.688, 1.688)	.892	24.21	27
.499	.599	1.201	-.201	.60	(.601, 1.801)	.935	18.83	19
.498	.599	1.203	-.203	.70	(.503, 1.903)	.960	15.65	14

Table 2.8. For $\theta = 1.2$ and $\rho = -0.2$ when $p_0 = 0.5, p_1 = 0.6$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.600	1.200	-.200	.10	(1.100, 1.300)	.946	920.21	923
.500	.600	1.200	-.200	.20	(1.000, 1.400)	.935	228.25	231
.501	.599	1.197	-.197	.30	(.897, 1.497)	.923	98.88	103
.499	.600	1.202	-.202	.40	(.802, 1.602)	.898	54.09	58
.498	.597	1.198	-.198	.50	(.698, 1.698)	.891	34.61	37
.501	.604	1.205	-.205	.60	(.605, 1.805)	.952	24.30	26
.498	.600	1.203	-.203	.70	(.503, 1.903)	.967	19.30	19

Table 2.9. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.400	.600	1.500	-.500	.15	(1.350, 1.650)	.883	575.80	587
.401	.600	1.498	-.498	.25	(1.248, 1.748)	.862	204.97	212
.400	.600	1.501	-.501	.35	(1.151, 1.851)	.832	100.12	108
.401	.598	1.493	-.493	.45	(1.043, 1.943)	.797	57.34	65
.401	.601	1.500	-.500	.55	(.950, 2.050)	.788	36.99	44
.399	.601	1.505	-.505	.65	(.855, 2.155)	.787	26.92	32
.402	.601	1.493	-.493	.75	(.743, 2.243)	.790	20.33	24
.399	.596	1.494	-.494	.85	(.644, 2.344)	.824	16.20	19
.400	.602	1.506	-.506	.95	(.556, 2.456)	.914	13.82	15
.402	.603	1.500	-.500	1.05	(.450, 2.550)	.961	12.22	12

Figure 2.3. Plot of coverage probability against d with $p_0 = 0.4, p_1 = 0.6, \alpha = 0.05$.

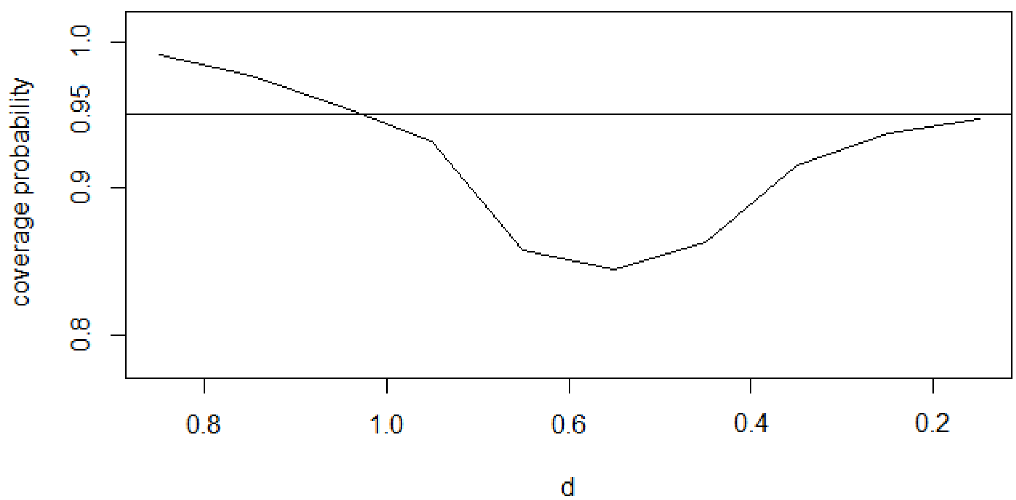


Table 2.10. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.400	.599	1.498	-.498	.15	(1.348, 1.648)	.935	817.49	832
.401	.600	1.499	-.499	.25	(1.249, 1.749)	.928	291.39	300
.400	.600	1.501	-.501	.35	(1.151, 1.851)	.904	146.01	153
.401	.600	1.497	-.497	.45	(1.047, 1.947)	.875	84.59	93
.399	.601	1.507	-.507	.55	(.957, 2.057)	.842	55.31	63
.399	.600	1.502	-.502	.65	(.852, 2.152)	.827	38.14	45
.401	.600	1.496	-.496	.75	(.746, 2.246)	.850	28.49	34
.399	.599	1.502	-.502	.85	(.652, 2.352)	.872	22.73	26
.401	.602	1.502	-.502	.95	(.552, 2.452)	.912	17.97	21
.397	.600	1.511	-.511	1.05	(.461, 2.561)	.958	15.51	18

Table 2.11. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.400	.600	1.500	-.500	.15	(1.350, 1.650)	.892	582.71	587
.400	.600	1.499	-.499	.25	(1.249, 1.749)	.881	206.56	212
.399	.600	1.501	-.501	.35	(1.151, 1.851)	.849	103.05	108
.401	.599	1.496	-.496	.45	(1.046, 1.946)	.829	59.49	65
.399	.599	1.500	-.500	.55	(.950, 2.050)	.836	40.18	44
.399	.600	1.503	-.503	.65	(.853, 2.153)	.846	29.22	32
.400	.601	1.501	-.501	.75	(.751, 2.251)	.928	23.23	24
.401	.601	1.499	-.499	.85	(.649, 2.349)	.960	19.48	19

Table 2.12. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.400	.600	1.501	-.501	.15	(1.351, 1.651)	.945	930.49	832
.400	.600	1.499	-.499	.25	(1.249, 1.749)	.929	294.42	300
.400	.600	1.501	-.501	.35	(1.151, 1.851)	.922	147.31	153
.400	.601	1.502	-.502	.45	(1.052, 1.952)	.887	87.38	93
.399	.602	1.508	-.508	.55	(.958, 2.058)	.875	57.48	63
.400	.598	1.493	-.493	.65	(.843, 2.143)	.897	40.57	45
.400	.600	1.499	-.499	.75	(.749, 2.249)	.931	30.91	34
.401	.598	1.493	-.493	.85	(.643, 2.343)	.962	25.05	26

Table 2.13. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	1.999	-.999	.20	(1.799, 2.199)	.892	611.32	619
.350	.699	1.999	-.999	.30	(1.699, 2.209)	.868	266.30	275
.349	.700	2.002	-1.002	.40	(1.602, 2.402)	.828	145.01	155
.350	.699	1.995	-.995	.50	(1.495, 2.495)	.768	86.81	99
.349	.700	2.003	-1.003	.60	(1.403, 2.603)	.757	59.96	69
.351	.701	2.000	-1.000	.70	(1.300, 2.700)	.744	42.54	51
.351	.700	1.998	-.998	.80	(1.198, 2.798)	.756	37.30	45
.349	.701	2.005	-1.005	.90	(1.305, 2.705)	.793	25.40	31
.350	.698	1.997	-.997	1.00	(.997, 2.997)	.832	21.61	25
.350	.699	1.996	-.996	1.10	(.896, 3.096)	.916	18.57	21
.351	.701	2.000	-1.000	1.20	(.800, 3.200)	.929	16.21	18
.351	.700	1.995	-.995	1.30	(.695, 3.295)	.941	14.42	15

Table 2.14. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	2.000	-1.000	.20	(1.800, 2.200)	.942	866.72	877
.350	.699	1.998	-.998	.30	(1.698, 2.208)	.936	385.22	391
.349	.700	2.003	-1.003	.40	(1.603, 2.403)	.914	212.51	221
.350	.701	2.004	-1.004	.50	(1.504, 2.504)	.869	130.81	141
.349	.700	2.003	-1.003	.60	(1.403, 2.603)	.821	85.85	98
.351	.702	2.003	-1.003	.70	(1.303, 2.703)	.804	61.85	72
.351	.700	1.998	-.998	.80	(1.198, 2.798)	.812	45.94	55
.351	.700	1.998	-.998	.90	(1.098, 2.898)	.832	36.58	44
.351	.698	1.993	-.993	1.00	(.993, 2.993)	.833	28.65	35
.351	.700	1.997	-.997	1.10	(.897, 3.097)	.921	24.33	29
.348	.702	2.018	-1.018	1.20	(.818, 3.218)	.930	21.56	25
.349	.698	1.996	-.996	1.30	(.696, 3.296)	.941	19.12	21

Figure 2.4. Plot of coverage probability against d with $p_0 = 0.35, p_1 = 0.7, \alpha = 0.05$.

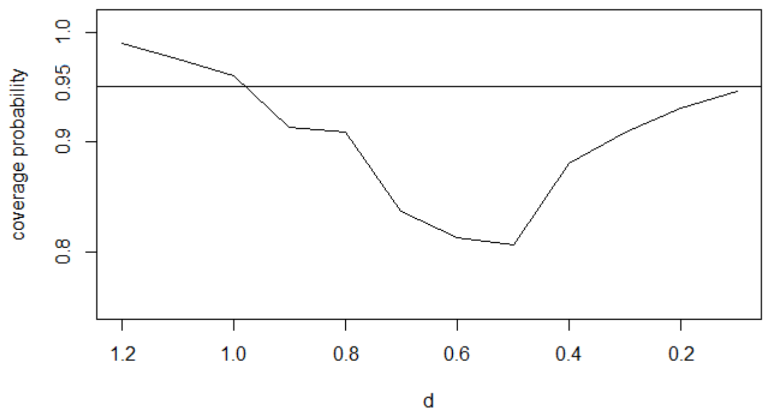


Table 2.15. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	2.001	-1.001	.20	(1.801, 2.201)	.895	616.02	619
.350	.699	1.999	-.999	.30	(1.699, 2.209)	.869	268.36	275
.349	.700	2.002	-1.002	.40	(1.602, 2.402)	.852	147.44	155
.351	.699	1.991	-.991	.50	(1.491, 2.491)	.813	90.63	99
.348	.701	2.013	-1.013	.60	(1.413, 2.613)	.826	64.54	69
.352	.702	2.003	-1.003	.70	(1.303, 2.703)	.868	46.80	51
.351	.700	1.998	-.998	.80	(1.198, 2.798)	.910	37.75	39
.349	.701	2.005	-1.005	.90	(1.305, 2.705)	.952	37.36	31

Table 2.16. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	2.000	-1.000	.20	(1.800, 2.200)	.946	873.58	877
.350	.700	1.999	-.999	.30	(1.699, 2.209)	.932	384.47	391
.349	.700	2.002	-1.002	.40	(1.602, 2.402)	.919	214.07	221
.350	.700	2.001	-1.001	.50	(1.501, 2.501)	.899	133.88	141
.349	.700	2.004	-1.004	.60	(1.404, 2.604)	.868	91.24	98
.349	.701	2.003	-1.003	.70	(1.303, 2.703)	.882	66.82	72
.350	.701	2.002	-1.002	.80	(1.202, 2.802)	.925	51.02	55
.351	.700	1.998	-.998	.90	(1.098, 2.898)	.952	41.40	44

Table 2.17. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.300	.750	2.499	-1.499	.25	(2.249, 2.749)	.886	714.59	722
.299	.750	2.502	1.502	.35	(2.152, 2.852)	.866	356.37	368
.301	.749	2.496	-1.496	.45	(2.046, 2.946)	.830	209.23	223
.299	.749	2.500	-1.500	.55	(1.950, 3.050)	.799	135.35	150
.299	.750	2.503	-1.503	.65	(1.853, 3.153)	.762	93.37	107
.300	.751	2.504	-1.504	.75	(1.754, 3.254)	.741	68.20	81
.301	.751	2.499	-1.499	.85	(1.649, 3.349)	.713	50.74	63
.298	.751	2.513	-1.513	.95	(1.563, 3.463)	.746	41.39	51
.300	.749	2.496	-1.496	1.05	(1.446, 3.546)	.761	33.64	41
.298	.749	2.509	-1.509	1.15	(1.359, 3.659)	.773	28.84	35
.301	.751	2.497	-1.497	1.25	(1.247, 3.747)	.858	25.15	30
.301	.751	2.502	-1.502	1.35	(1.152, 3.852)	.871	22.18	25
.299	.753	2.513	-1.513	1.45	(1.049, 3.949)	.890	19.56	22
.302	.751	2.495	-1.495	1.55	(.945, 4.045)	.913	17.35	19
.300	.750	2.499	-1.499	1.65	(.849, 4.149)	.960	16.29	17

Table 2.18. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.301	.750	2.498	-1.498	.25	(2.248, 2.748)	.946	1014.67	1024
.299	.750	2.504	1.504	.35	(2.154, 2.854)	.934	517.27	524
.301	.749	2.495	-1.495	.45	(2.045, 2.945)	.914	307.62	317
.299	.749	2.500	-1.500	.55	(1.950, 3.050)	.886	199.49	212
.299	.750	2.503	-1.503	.65	(1.853, 3.153)	.846	137.61	152
.300	.751	2.504	-1.504	.75	(1.754, 3.254)	.813	100.47	114
.299	.751	2.513	-1.513	.85	(1.663, 3.363)	.804	77.22	90
.300	.749	2.497	-1.497	.95	(1.547, 3.447)	.773	58.90	71
.300	.749	2.496	-1.496	1.05	(1.446, 3.546)	.774	46.43	58
.302	.753	2.506	-1.506	1.15	(1.356, 3.656)	.806	38.75	49
.301	.751	2.497	-1.497	1.25	(1.247, 3.747)	.871	33.72	41
.299	.751	2.505	-1.505	1.35	(1.155, 3.855)	.882	29.08	36
.299	.750	2.506	-1.506	1.45	(1.056, 3.956)	.876	25.76	31
.300	.750	2.499	-1.499	1.55	(.949, 4.049)	.911	23.24	27
.300	.752	2.508	-1.508	1.65	(.858, 4.158)	.965	21.46	24

Table 2.19. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.301	.750	2.496	-1.496	.25	(2.246, 2.746)	.884	711.16	722
.299	.750	2.503	1.503	.35	(2.153, 2.853)	.878	361.65	368
.303	.752	2.496	-1.496	.45	(2.046, 2.946)	.849	213.70	223
.299	.749	2.501	-1.501	.55	(1.951, 3.051)	.826	138.10	150
.302	.753	2.503	-1.503	.65	(1.853, 3.153)	.784	95.84	107
.300	.748	2.492	-1.492	.75	(1.742, 3.242)	.775	71.19	81
.301	.750	2.497	-1.497	.85	(1.647, 3.347)	.777	54.92	63
.298	.751	2.513	-1.513	.95	(1.563, 3.463)	.746	41.39	51
.299	.748	2.496	-1.496	1.05	(1.446, 3.546)	.858	37.39	41
.299	.751	2.509	-1.509	1.15	(1.359, 3.659)	.917	32.25	35
.302	.748	2.481	-1.481	1.25	(1.231, 3.731)	.925	27.44	30
.301	.751	2.503	-1.503	1.35	(1.153, 3.853)	.956	25.48	25

Table 2.20. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.300	.750	2.498	-1.498	.25	(2.248, 2.748)	.943	1018.73	1025
.301	.750	2.496	1.496	.35	(2.146, 2.846)	.937	512.81	523
.301	.749	2.494	-1.494	.45	(2.044, 2.944)	.916	308.30	317
.298	.748	2.501	-1.501	.55	(1.951, 3.051)	.887	200.36	212
.299	.750	2.503	-1.503	.65	(1.853, 3.153)	.872	141.22	152
.302	.751	2.495	-1.495	.75	(1.745, 3.245)	.844	102.70	114
.299	.752	2.511	-1.511	.85	(1.661, 3.361)	.827	79.03	89
.300	.748	2.494	-1.494	.95	(1.544, 3.444)	.849	63.65	71
.298	.751	2.510	-1.510	1.05	(1.460, 3.560)	.872	53.06	58
.301	.752	2.503	-1.503	1.15	(1.353, 3.653)	.915	43.81	49
.301	.751	2.497	-1.497	1.25	(1.247, 3.747)	.929	37.88	41
.301	.751	2.499	-1.499	1.35	(1.149, 3.849)	.962	32.76	36

Figure 2.5. Plot of coverage probability against d with $p_0 = 0.3, p_1 = 0.75, \alpha = 0.05$.

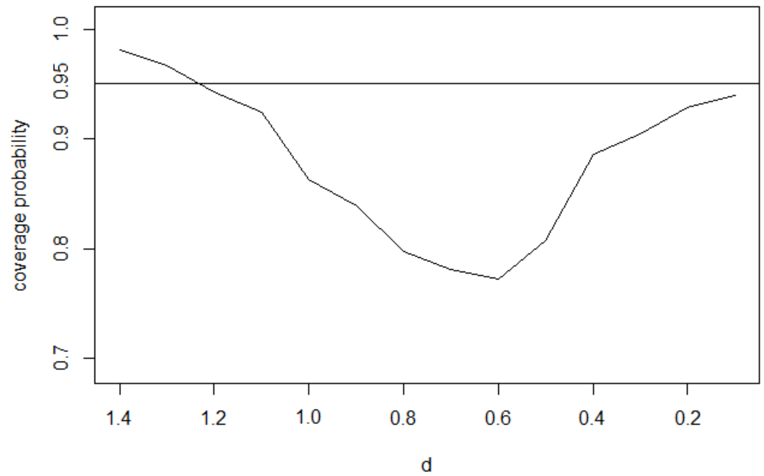


Table 2.21. For $\theta = 3.0$ and $\rho = -2.0$ when $p_0 = 0.25, p_1 = 0.75$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.250	.750	2.999	-1.999	.30	(2.699, 3.209)	.890	894.05	903
.249	.750	3.004	-2.004	.40	(2.604, 3.404)	.870	495.39	508
.250	.749	2.995	-1.995	.50	(2.495, 3.495)	.845	309.78	326
.301	.750	2.997	-1.997	.60	(2.397, 3.597)	.818	210.51	227
.251	.752	3.001	-2.001	.70	(2.301, 3.701)	.779	147.31	166
.253	.752	2.995	-1.995	.80	(2.195, 3.795)	.751	109.18	127
.249	.751	3.010	-2.010	.90	(2.010, 3.910)	.705	83.01	101
.250	.748	2.991	-1.991	1.00	(1.991, 3.991)	.700	67.38	82
.250	.751	3.006	-2.006	1.10	(1.906, 4.106)	.703	54.73	68
.251	.751	3.000	-2.000	1.20	(1.800, 4.200)	.709	46.10	57
.248	.750	3.014	-2.014	1.30	(1.714, 4.314)	.748	40.84	49
.251	.750	2.995	-1.995	1.40	(1.595, 4.395)	.751	33.45	42
.252	.750	2.982	-1.982	1.50	(1.482, 4.482)	.782	29.57	36
.251	.750	2.993	-1.993	1.60	(1.393, 4.593)	.799	27.23	32
.251	.750	2.995	-1.995	1.70	(1.295, 4.695)	.876	24.32	29
.250	.750	3.004	-2.004	1.80	(1.204, 4.804)	.915	21.94	26
.251	.751	2.992	-1.992	1.90	(1.092, 4.892)	.919	19.51	23
.251	.750	2.987	-1.987	2.00	(.987, 4.987)	.950	18.40	20

Table 2.22. For $\theta = 3.0$ and $\rho = -2.0$ when $p_0 = 0.25, p_1 = 0.75$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.250	.750	2.997	-1.997	.30	(2.697, 3.207)	.935	1272.14	1282
.250	.751	3.006	-2.006	.40	(2.606, 3.406)	.931	709.00	720
.250	.749	2.994	-1.994	.50	(2.494, 3.494)	.922	447.51	461
.302	.751	2.997	-1.997	.60	(2.397, 3.597)	.902	304.82	321
.251	.753	3.008	-2.008	.70	(2.308, 3.708)	.875	219.07	236
.253	.752	2.995	-1.995	.80	(2.195, 3.795)	.830	160.14	180
.249	.751	3.011	-2.011	.90	(2.011, 3.911)	.821	125.23	143
.250	.748	2.991	-1.991	1.00	(1.991, 3.991)	.771	96.94	116
.259	.751	3.012	-2.012	1.10	(1.912, 4.112)	.756	78.53	96
.247	.749	3.006	-2.006	1.20	(1.806, 4.206)	.749	63.62	80
.250	.750	3.001	-2.001	1.30	(1.701, 4.301)	.753	57.08	69
.252	.751	2.992	-1.995	1.40	(1.592, 4.392)	.769	47.62	59
.249	.750	3.008	-2.008	1.50	(1.508, 4.508)	.815	41.54	52
.250	.750	2.998	-1.998	1.60	(1.398, 4.598)	.836	36.68	45
.251	.750	2.995	-1.995	1.70	(1.295, 4.695)	.866	32.89	40
.250	.750	3.000	-2.000	1.80	(1.200, 4.800)	.920	30.17	36
.249	.753	3.022	-2.022	1.90	(1.122, 4.922)	.931	27.75	33
.251	.750	2.990	-1.990	2.00	(.990, 4.990)	.958	24.09	29

Figure 2.6. Plot of coverage probability against d with $p_0 = 0.25, p_1 = 0.75, \alpha = 0.05$.

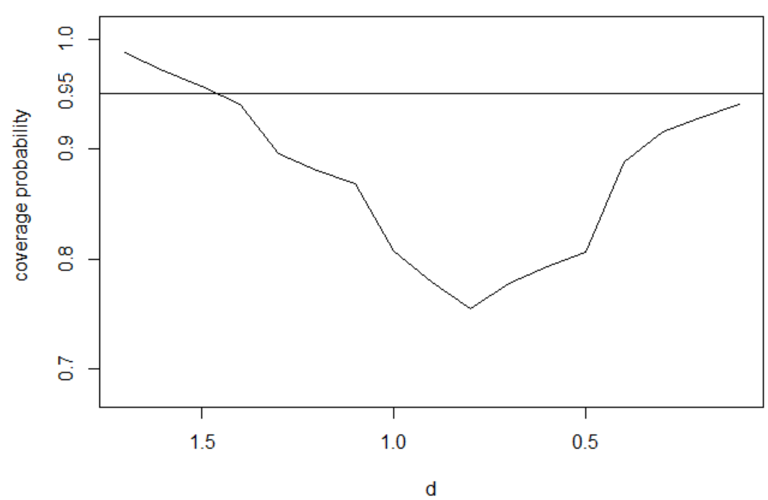


Table 2.23. For $\theta = 3.0$ and $\rho = -2.0$ when $p_0 = 0.25, p_1 = 0.75$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.250	.750	3.001	-2.001	.30	(2.701, 3.301)	.894	894.52	903
.249	.750	3.005	-2.005	.40	(2.605, 3.405)	.882	496.91	507
.250	.749	2.998	-1.998	.50	(2.498, 3.498)	.853	313.02	325
.301	.751	2.997	-1.997	.60	(2.397, 3.597)	.841	212.06	226
.251	.750	2.991	-1.991	.70	(2.291, 3.691)	.805	150.14	166
.253	.752	2.994	-1.994	.80	(2.194, 3.794)	.787	113.13	127
.251	.752	3.010	-2.010	.90	(2.010, 3.910)	.757	86.97	100
.250	.751	3.003	-2.003	1.00	(2.003, 4.003)	.747	71.04	82
.251	.751	2.996	-1.996	1.10	(1.806, 4.006)	.766	58.11	67
.250	.750	3.000	-2.000	1.20	(1.800, 4.200)	.798	50.17	57
.249	.750	3.004	-2.004	1.30	(1.704, 4.304)	.869	43.97	49
.250	.749	2.995	-1.995	1.40	(1.595, 4.395)	.883	37.91	42
.251	.750	2.989	-1.989	1.50	(1.489, 4.489)	.897	33.78	37
.248	.750	3.016	-2.016	1.60	(1.416, 4.616)	.938	31.38	32
.252	.750	2.985	-1.985	1.70	(1.285, 4.685)	.955	27.91	28

Table 2.24. For $\theta = 3.0$ and $\rho = -2.0$ when $p_0 = 0.25, p_1 = 0.75$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.250	.750	2.999	-1.999	.30	(2.699, 3.209)	.948	1272.86	1281
.250	.750	3.001	-2.001	.40	(2.601, 3.401)	.937	712.18	721
.250	.749	2.994	-1.994	.50	(2.494, 3.494)	.925	452.30	461
.302	.752	2.997	-1.997	.60	(2.397, 3.597)	.917	309.72	321
.249	.752	3.011	-2.011	.70	(2.311, 3.711)	.889	221.07	235
.250	.751	3.005	-2.005	.80	(2.205, 3.805)	.854	165.01	181
.250	.748	2.991	-1.991	.90	(2.091, 3.891)	.846	128.85	143
.250	.750	2.999	-1.999	1.00	(1.999, 3.999)	.822	102.64	116
.259	.751	3.005	-2.005	1.10	(1.905, 4.105)	.822	84.85	96
.250	.749	2.998	-1.998	1.20	(1.798, 4.198)	.830	69.60	81
.250	.750	3.001	-2.001	1.30	(1.701, 4.301)	.857	60.09	69
.249	.751	3.012	-2.012	1.40	(1.612, 4.412)	.868	52.66	60
.249	.750	3.008	-2.008	1.50	(1.508, 4.508)	.901	45.08	52
.250	.749	2.996	-1.996	1.60	(1.3968, 4.596)	.933	40.69	45
.251	.749	2.988	-1.988	1.70	(1.288, 4.688)	.951	36.24	40

Table 2.25. For $\theta = 4.0$ and $\rho = -3.0$ when $p_0 = 0.2, p_1 = 0.8$, $\alpha = 0.1$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.200	.800	4.001	-3.001	.40	(3.601, 4.401)	.889	1133.98	1150
.200	.799	3.995	-2.995	.50	(3.495, 4.495)	.877	718.21	736
.201	.800	3.997	-2.997	.60	(3.397, 4.597)	.864	495.71	513
.200	.800	4.001	-3.001	.70	(3.301, 4.701)	.843	355.58	377
.200	.800	3.995	-2.995	.80	(3.195, 4.795)	.803	263.88	288
.200	.801	4.010	-3.010	.90	(3.010, 4.910)	.768	199.55	228
.202	.801	3.991	-2.991	1.00	(2.991, 4.991)	.756	158.68	185
.200	.801	4.006	-3.006	1.10	(2.906, 5.106)	.732	127.75	152
.200	.800	4.000	-3.000	1.20	(2.800, 5.200)	.718	115.80	140
.199	.801	4.014	-3.014	1.30	(2.714, 5.314)	.694	88.02	109
.201	.800	3.995	-2.995	1.40	(2.595, 5.395)	.692	74.13	94
.202	.800	3.985	-2.985	1.50	(2.485, 5.485)	.686	66.07	83
.199	.798	3.993	-2.993	1.60	(2.393, 5.593)	.682	55.37	71
.201	.800	3.995	-2.995	1.70	(2.295, 5.695)	.690	49.06	64
.200	.800	4.004	-3.004	1.80	(2.204, 5.804)	.704	45.03	57
.200	.799	3.992	-2.992	1.90	(2.092, 5.892)	.724	38.79	51
.200	.800	3.997	-2.997	2.00	(1.997, 5.997)	.734	36.73	47
.201	.800	3.987	-2.987	2.10	(1.887, 6.087)	.752	33.40	42
.199	.799	3.997	-2.997	2.20	(1.797, 6.197)	.851	31.42	39
.200	.800	4.003	-3.003	2.30	(1.703, 6.303)	.870	28.41	35
.200	.801	4.007	-3.007	2.40	(2.607, 6.407)	.879	27.60	33
.201	.802	3.990	-2.990	2.50	(1.490, 6.490)	.890	25.46	30
.200	.799	3.998	-2.98	2.60	(1.398, 6.598)	.902	24.35	28
.200	.800	4.001	-3.001	2.70	(1.301, 6.701)	.948	23.00	26

Table 2.26. For $\theta = 4.0$ and $\rho = -3.0$ when $p_0 = 0.2, p_1 = 0.8$, $\alpha = 0.05$, m=5

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.200	.800	4.000	-3.000	.40	(3.600, 4.400)	.940	1621.54	1634
.200	.800	3.999	-2.999	.50	(3.499, 4.499)	.935	1033.56	1046
.201	.800	3.996	-2.996	.60	(3.396, 4.596)	.923	711.55	727
.299	.799	4.004	-3.004	.70	(3.304, 4.704)	.913	514.61	533
.202	.801	3.991	-2.991	.80	(3.191, 4.791)	.895	392.16	410
.200	.800	4.002	-3.002	.90	(3.002, 4.902)	.866	298.28	323
.201	.799	3.991	-2.991	1.00	(2.991, 4.991)	.851	238.66	263
.198	.799	4.006	-3.006	1.10	(2.906, 5.106)	.804	186.08	216
.199	.799	4.000	-3.000	1.20	(2.800, 5.200)	.788	155.11	182
.201	.800	3.994	-2.994	1.30	(2.694, 5.294)	.787	131.72	155
.200	.800	3.998	-2.998	1.40	(2.598, 5.398)	.758	109.70	134
.202	.801	3.989	-2.989	1.50	(2.489, 5.489)	.755	94.25	116
.199	.799	3.997	-2.997	1.60	(2.397, 5.597)	.754	84.63	103
.201	.800	3.995	-2.995	1.70	(2.295, 5.695)	.726	70.79	79
.200	.801	4.004	-3.004	1.80	(2.204, 5.804)	.743	65.41	82
.199	.799	3.992	-2.992	1.90	(2.092, 5.892)	.766	57.59	73
.201	.801	3.997	-2.997	2.00	(1.997, 5.997)	.758	51.96	67
.202	.800	3.987	-2.987	2.10	(1.887, 6.087)	.763	46.17	60
.198	.799	4.008	-3.008	2.20	(1.808, 6.208)	.840	43.61	55
.200	.800	4.009	-3.009	2.30	(1.709, 6.309)	.856	37.73	50
.201	.802	4.007	-3.007	2.40	(2.607, 6.407)	.875	35.74	46
.202	.801	3.984	-2.984	2.50	(1.484, 6.484)	.892	34.07	43
.200	.799	3.995	-2.995	2.60	(1.395, 6.595)	.899	31.54	39
.200	.800	4.001	-3.001	2.70	(1.301, 6.701)	.939	30.02	37

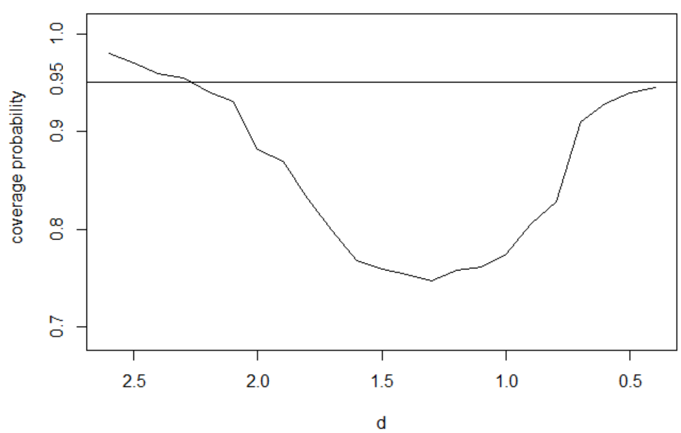
Table 2.27. For $\theta = 4.0$ and $\rho = -3.0$ when $p_0 = 0.2, p_1 = 0.8$, $\alpha = 0.1$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.200	.800	3.998	-2.998	.40	(3.598, 4.398)	.891	1138.34	1150
.200	.800	3.995	-2.995	.50	(3.495, 4.495)	.885	723.41	736
.201	.800	3.993	-2.993	.60	(3.393, 4.593)	.880	499.04	513
.200	.802	4.011	-3.011	.70	(3.311, 4.711)	.844	356.68	376
.199	.799	3.995	-2.995	.80	(3.195, 4.795)	.817	265.05	287
.200	.801	4.008	-3.008	.90	(3.008, 4.908)	.802	206.93	228
.198	.800	4.012	-3.012	1.00	(3.012, 5.012)	.777	165.13	185
.199	.801	4.009	-3.009	1.10	(2.909, 5.109)	.767	134.19	153
.200	.800	4.002	-3.002	1.20	(2.802, 5.202)	.759	110.82	128
.199	.802	4.017	-3.017	1.30	(2.717, 5.317)	.751	94.82	110
.201	.800	3.995	-2.995	1.40	(2.595, 5.395)	.751	80.63	94
.200	.799	3.995	-2.995	1.50	(2.495, 5.495)	.744	69.77	82
.199	.799	4.002	-3.002	1.60	(2.402, 5.602)	.755	62.94	72
.200	.800	3.997	-2.997	1.70	(2.297, 5.697)	.790	55.28	64
.200	.800	4.004	-3.004	1.80	(2.204, 5.804)	.822	50.29	57
.200	.799	3.996	-2.996	1.90	(2.096, 5.896)	.860	46.16	51
.201	.800	3.991	-2.991	2.00	(1.991, 5.991)	.877	42.20	46
.201	.800	3.990	-2.990	2.10	(1.890, 6.090)	.931	39.73	42
.199	.797	3.987	-2.987	2.20	(1.787, 6.187)	.937	36.15	39
.201	.801	4.001	-3.001	2.30	(1.701, 6.301)	.957	33.65	35

Table 2.28. For $\theta = 4.0$ and $\rho = -3.0$ when $p_0 = 0.2, p_1 = 0.8$, $\alpha = 0.05$, m=10

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.200	.800	3.997	-2.997	.40	(3.597, 4.397)	.939	1624.52	1634
.200	.800	3.999	-2.999	.50	(3.499, 4.499)	.936	1034.06	1046
.200	.800	4.002	-3.002	.60	(3.402, 4.602)	.927	714.89	727
.199	.799	4.004	-3.004	.70	(3.304, 4.704)	.911	515.44	533
.201	.801	3.998	-2.998	.80	(3.198, 4.798)	.909	391.36	409
.200	.800	4.002	-3.002	.90	(3.002, 4.902)	.886	302.17	322
.201	.800	3.991	-2.991	1.00	(2.991, 4.991)	.873	242.49	262
.199	.799	4.006	-3.006	1.10	(2.906, 5.106)	.853	196.22	216
.199	.800	4.011	-3.011	1.20	(2.811, 5.211)	.835	164.48	182
.201	.800	3.990	-2.990	1.30	(2.690, 5.290)	.817	136.25	155
.200	.800	3.998	-2.998	1.40	(2.598, 5.398)	.810	115.47	134
.200	.801	4.007	-3.007	1.50	(2.507, 5.507)	.810	99.57	116
.198	.801	4.021	-3.021	1.60	(2.421, 5.621)	.807	89.54	103
.201	.800	3.993	-2.993	1.70	(2.293, 5.693)	.813	77.07	91
.200	.801	4.008	-3.008	1.80	(2.208, 5.808)	.823	70.44	82
.199	.798	3.992	-2.992	1.90	(2.092, 5.892)	.865	61.72	73
.201	.801	3.997	-2.997	2.00	(1.997, 5.997)	.863	55.94	68
.200	.800	3.997	-2.997	2.10	(1.897, 6.097)	.928	51.97	60
.199	.799	4.008	-3.008	2.20	(1.808, 6.208)	.929	48.50	55
.200	.800	4.005	-3.005	2.30	(1.705, 6.305)	.957	44.26	50

Figure 2.7. Plot of coverage probability against d with $p_0 = 0.2, p_1 = 0.8, \alpha = 0.05$.



In these tables, the minimum value of d was chosen to be 10 percent of the risk ratio, and then increased by 0.1 in the following steps. (Note: For practical purposes, the size of d can be determined from the standard error of the the estimate $\hat{\theta}$.).

From Tables 2.1 to 2.28, we infer that the expected stopping time $E(N)$ monotonically increases (to infinity) as d decreases (to zero). The Monte Carlo estimates of $\hat{p_0}$ and $\hat{p_1}$ approach the true values of the parameters p_0 and p_1, respectively as the length of the interval decreases, and we also observe that as d decreases the coverage probability (CP) gets close to the nominal probability $1 - \alpha$. (This property is referred to as the *asymptotic consistency.*) Therefore, the above numerical evidence indicates that the finite sample behavior lends support to the asymptotic behavior of the proposed sequential procedure when $d \to 0$.

From Figures 2.1 to 2.7, clearly we can observe that the coverage probability starts from a level higher (in fact, close to 1.0 with large value of d) than the nominal level, and it goes down. After the coverage probability reaches its minimum value, it will eventually approach the target nominal level when d becomes small.

In fact, increasing the starting sample size m results in an increase of both stopping time and coverage probability. Accordingly, when the CP is below the nominal level, one can choose a moderate size of d which can be determined from the standard error (S.E.) of the estimate $\hat{\theta}$.

2.7.2 Wald-based CI's Versus Likelihood-based CI's

One might wish to consider a likelihood-based confidence interval which is preferable to have since it has better general performance in some aspects than a Wald-based confidence interval.

Definition 2.7 (Likelihood-based confidence interval) Suppose $(X_1, X_2, ..., X_n)$ is a random sample from a distribution having parameter θ. Let $\hat{\theta}_{MLE}$ be the maximum likelihood estimator of θ. The likelihood-based confidence interval with confidence level $1-\alpha$ is an interval $(\hat{\theta}_L^{lik}, \hat{\theta}_U^{lik})$ such that

$$P\{\hat{\theta}_L^{lik} \leqslant \hat{\theta}_{MLE} \leqslant \hat{\theta}_U^{lik}\} \geqslant 1-\alpha$$

Definition 2.8 (Parameter invariance) An interval $(\hat{\theta}_L, \hat{\theta}_U)$ is said to be parameter invariant if $P\{\hat{\theta}_L \leqslant \hat{\theta}_n \leqslant \hat{\theta}_U\} \geqslant 1-\alpha$ implies $P\{\frac{1}{\hat{\theta}_U} \leqslant \frac{1}{\hat{\theta}_n} \leqslant \frac{1}{\hat{\theta}_L}\} \geqslant 1-\alpha$.

Theorem 2.3 The likelihood confidence interval of the maximum likelihood estimator is parameter invariant.

To find the likelihood-based confidence intervals of the risk ratio, we start from the MLE of the parameter θ, then we computationally increase and also decrease θ_{MLE}, to expand the interval and get two equal heights cut off on the likelihood function, until the coverage probability approaches the confidence level $1-\alpha$. Using the optimal sample size n^*, two limits of confidence interval are found. For some of the scenarios, results are shown in the following likelihood-based CI vs. Wald-based CI tables.

Table 2.29. W-based CI vs L-based CI, $\theta = 1$ (when $p_0 = .5, p_1 = .5$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.1	1.002	769.833	769	0.9501	(0.902, 1.102)	(0.905, 1.105)
0.2	0.999	189.365	193	0.9277	(0.799, 1.199)	(0.819, 1.221)
0.3	1.001	82.097	86	0.9084	(0.701, 1.301)	(0.739, 1.353)
0.4	1.000	44.508	49	0.8806	(0.600, 1.400)	(0.670, 1.492)

Table 2.30. W-based CI vs L-based CI, $\theta = 1.2$ (when $p_0 = .5, p_1 = .6$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.1	1.199	919.697	922	0.9506	(1.099, 1.299)	(1.108, 1.309)
0.2	1.199	228.134	231	0.9391	(0.999, 1.399)	(1.016, 1.418)
0.3	1.201	100.019	103	0.9194	(0.901, 1.501)	(0.932, 1.538)
0.4	1.200	53.209	58	0.8892	(0.800, 1.600)	(0.868, 1.705)

Table 2.31. W-based CI vs L-based CI, $\theta = 1.5$ (when $p_0 = .4, p_1 = .6$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.15	1.501	825.555	834	0.9470	(1.351, 1.651)	(1.356, 1.656)
0.25	1.501	291.797	301	0.9376	(1.251, 1.751)	(1.274, 1.779)
0.35	1.501	147.688	154	0.9158	(1.151, 1.851)	(1.214, 1.941)
0.45	1.500	84.256	94	0.8633	(1.050, 1.950)	(1.103, 2.009)

Table 2.32. W-based CI vs L-based CI, $\theta = 1.5$ (when $p_0 = .3, p_1 = .45$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.15	1.501	1357.22	1366	0.9483	(1.351, 1.651)	(1.358, 1.659)
0.25	1.500	486.307	492	0.9256	(1.250, 1.750)	(1.267, 1.768)
0.35	1.499	239.882	251	0.9087	(1.149, 1.849)	(1.197, 1.931)
0.45	1.499	142.538	152	0.8802	(1.049, 1.949)	(1.120, 2.039)

Table 2.33. W-based CI vs L-based CI, $\theta = 2$ (when $p_0 = .35, p_1 = .7$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.2	2.002	867.396	881	0.9456	(1.802, 2.202)	(1.813, 2.215)
0.3	1.998	383.572	392	0.9311	(1.698, 2.298)	(1.729, 2.334)
0.4	2.003	213.007	221	0.9087	(1.603, 2.403)	(1.660, 2.479)
0.5	2.000	130.977	142	0.8801	(1.500, 2.500)	(1.589, 2.633)

Table 2.34. W-based CI vs L-based CI, $\theta = 2.5$ (when $p_0 = .3, p_1 = .75$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.25	2.503	1018.01	1025	0.9392	(2.253, 2.753)	(2.264, 2.765)
0.35	2.503	511.997	523	0.9289	(2.153, 2.853)	(2.180, 2.885)
0.45	2.499	307.601	317	0.9055	(2.049, 2.949)	(2.107, 3.021)
0.55	2.499	197.687	212	0.8866	(1.949, 3.049)	(2.015, 3.127)

Table 2.35. W-based CI vs L-based CI, $\theta = 3$ (when $p_0 = .25, p_1 = .75$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.3	3.002	1270.55	1281	0.9398	(2.702, 3.302)	(2.720, 3.323)
0.4	2.999	707.138	721	0.9177	(2.599, 3.399)	(2.635, 3.441)
0.5	3.001	445.754	461	0.9053	(2.501, 3.501)	(2.553, 3.567)
0.6	3.000	305.139	321	0.8876	(2.400, 3.600)	(2.476, 3.698)

Table 2.36. W-based CI vs L-based CI, $\theta = 4$ (when $p_0 = .2, p_1 = .8$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.4	3.998	1613.19	1633	0.9401	(3.598, 4.398)	(3.627, 4.430)
0.5	3.999	1029.00	1045	0.9369	(3.499, 4.499)	(3.542, 4.548)
0.6	4.001	708.124	726	0.9274	(3.401, 4.601)	(3.471, 4.653)
0.7	4.002	515.877	534	0.9103	(3.302, 4.702)	(3.381, 4.796)

From the above tables, we can observe that the Wald-based confidence intervals and the likelihood-based confidence intervals are quite agreeable to each other. The likelihood-based confidence intervals are off-centered due to the fact that the Binomial distribution is skewed to the right. On the other hand, the Wald-based confidence intervals are balanced since the intervals have constructed based on the Normal approximation.

The advantage of likelihood-based confidence intervals is that the confidence intervals are invariant toward its reciprocal. We can verify the invariance by interchanging the parameters p_0 and p_1. For example,

Table 2.37. The invariance of Likelihood CI

Estimate	$E(N)$	n^*	CP	likelihood CI
$\hat{\theta}_n = 2.503$	1018.01	1025	0.9392	(2.2643, 2.7651)
$\hat{\theta}_n^{-1} = 0.399$	1019.70	1025	0.9405	(0.3616, 0.4416)

As we see that the above results satisfy that $1/2.7651 = 0.3616$ and $1/2.2643 = 0.3616$. Therefore, the likelihood-based confidence intervals are exactly invariant.

However, since the likelihood-based confidence intervals are computationally-oriented, the results are not easy to obtain analytically.

The following Tables 2.38 - 2.40 show various cases of invariance between Wald-based confidence intervals and the likelihood-based confidence intervals.

Table 2.38. The Confidence Interval of $1/\theta$; (when $p_0 = 0.4, p_1 = 0.6$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.15	1.501	825.555	834	0.9351	(1.351, 1.651)	(1.356, 1.656)
d	$1/\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.05	0.666	817.691	834	0.9248	(0.616, 0.716)	(0.619, 0.725)

Table 2.39. The Confidence Interval of $1/\theta$; (when $p_0 = 0.3, p_1 = 0.75$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.2	2.503	1018.01	1025	0.9392	(2.303, 2.703)	(2.264, 2.765)
d	$1/\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.05	0.401	994.606	1025	0.9176	(0.351, 0.451)	(0.361, 0.442)

Table 2.40. The Confidence Interval of $1/\theta$; (when $p_0 = 0.2, p_1 = 0.8$)

d	$\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.4	3.999	1613.19	1633	0.9401	(3.599, 4.399)	(3.627, 4.430)
d	$1/\hat{\theta}$	$E(N)$	n^*	CP	Wald-based CI	likelihood CI
0.025	0.2503	1466.03	1633	0.8504	(0.2253, 0.2753)	(0.226, 0.277)

From Tables 2.38 to 2.40, we can see the Wald-based confidence intervals are almost invariant. We refer this to as near-invariance. Thus, from the above numerical evidence, it is fair to say that Wald-based confidence intervals produced by the proposed procedure are as good as the likelihood-based confidence intervals in terms of length and sample sizes.

CHAPTER 3

TWO-STAGE PROCEDURE

3.1 Introduction

One can set up a sampling strategy that takes observations in two steps and proceeds to further investigation for inferences.

As we have already discussed in the previous chapter, there does not exist a fixed-sample size procedure for estimating the mean of a normal population (when the variance is unknown) with a fixed-width confidence interval and preassigned confidence level. One of the ways to overcome this problem is Stein's (1945, 1949) two-stage procedure. The first step, called Stage 1, records a pilot sample of size $m(\geq 2)$ and evaluates a statistic (e.g. sample variance), then the experimenter proceeds to the second step, Stage 2, to gather all remaining observations (if needed) for further statistical inference. We need to note that the sample size of the second sample depends upon the results from the pilot sample.

In order to diversify the sampling plan, we consider the advantage of the sequential sampling design toward the two-stage sampling procedure and inference. For example, often in many pharmaceutical studies, experimenters want to have interim or intermediate stages to gather updated information on the adequacy of the planned sample size in a study, because the experimenters are often uncertain about the assessed values of the parameters that were used initially for the calculations or obtained from some other studies. Consequently, the sample size that was initially planned does not necessarily guarantee the width of the confidence interval, but also the required power (for the testing). Therefore, it would be plausible to reestimate the required size beyond the originally planned to get the overall optimal sample sizes if it needs. This is also frequently refered to as the sample size reestimation problem in clinical trials.

We start with a procedure proposed by Stein (1945, 1949) that takes observations in two stages. Then, we extend the procedure to the two-sample case for the risk ratio and study the asymptotic properties of the procedure. Then, we will perform Monte Carlo simulations with feasible scenarios in order to investigate the finite sample behavior.

3.2 Two-Stage Procedure for One-sample Case

Let X_1, X_2, ... be a sequence of i.i.d $N(\mu, \sigma^2)$ variables with both unknown parameters. One wishes to construct a $100(1-\alpha)\%$ confidence interval with pre-specified length $2d$ and confidence level $1-\alpha$. To implement the two-stage procedure, at the initial step, we take $m(\geq 2)$ observations (X_1, X_2, ..., X_m), define the sample mean and sample variance to be $\bar{X}_m = \Sigma_{i=1}^m X_i/m$ and $s_m^2 = \Sigma_{i=1}^m (X_i - \bar{X}_n)^2/(m-1)$, respectively. Then, based on these two quantities, we need to make a decision on how many additional observations are required to satisfy the criteria of the problem.

The half-width of a $100(1-\alpha)\%$ confidence interval based on $\bar{X}_m$ is given by $t_{\alpha/2,m-1}S_m/\sqrt{m}$, where $t_{\alpha/2,m-1}$ is the $100(1-\alpha/2)\%$ point of t distribution with

$(m-1)$ degrees of freedom. Recall that the confidence interval for μ is $I_m = [\bar{X}_m - d_m, \bar{X}_m + d_m] = [\bar{X}_m - t_{\alpha/2,m-1}S_m/\sqrt{m}, \bar{X}_m + t_{\alpha/2,m-1}S_m/\sqrt{m}]$. If the half length from the sample of size m is smaller than or equal to the desired half-width d, then no further stage will be needed; otherwise the additional observations are taken so that the total sample size n is at least as large as $t^2_{\alpha/2,m-1}S^2_m/d^2_m$.

Thus, the final sample size is determined as

$$N \equiv N(d) = max\left\{m, \left[\frac{t^2_{\alpha/2,m-1}S^2_m}{d^2}\right] + 1\right\} \tag{3.1}$$

where $[k]$ denotes the smallest integer $\geq k$, Thus, easily we can see that N is finite $(N < \infty)$ with probability one.

In summary, the Stein's two-stage procedure is basically to construct a fixed-width confidence interval for μ with sample variance S^2_m that was obtained from the first-stage sample X_1, X_2, ..., X_m of size m. Then, using the sample variance S^2_m, one can estimate the required optimal fixed sample size. If $N = m$, we already have enough sample (pilot sample size m) to achieve the desirable d and no need to take any more samples. If $N > m$, then we sample X_{m+1}, X_{m+2}, ..., X_N of size $(N-m)$ at the second-stage. Then, we have X_1, X_2, ..., X_N and the interval $[\bar{X}_N \pm d]$ based on all samples of size N.

3.3 Two-Stage Procedure for Risk Ratio

As we mentioned before, the ratio of two binomial proportions is of major interest for measuring the risk ratio in comparative prospective studies and in biomedical experiments. Therefore, in this section, we move on to the two-stage procedure for the risk ratio of two binomial variates.

We consider two sequences of independent Bernoulli variates with probabilities p_0 and p_1, respectively, say X_1, X_2, ..., and Y_1, Y_2, ... Recall that in chapter two, we defined two binomial random variables $R = \sum_{i=1}^{n} X_i$ and $S = \sum_{i=1}^{n} Y_i$, where $\sum_{i=1}^{n} X_i$ follows the binomial distribution with parameters $(n,\, p_0)$, and $\sum_{i=1}^{n} Y_i$ follows the binomial distribution with parameters $(n,\, p_1)$. Based on these two sums of Bernoulli variates, we define a modified estimator for the risk ratio $\theta = p_1/p_0$, that is

$$\hat{\theta}_n = \frac{S}{R + 1/n} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i + 1/n}.$$

We state the main result (see Sec. 2.4) that $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ converges in distribution to $N\left(0, \sigma^2\right)$ for sufficiently large n with

$$\sigma^2 = \frac{\theta(1 + \theta - 2\theta p_0)}{p_0}.$$

Thus, the sequence of risk ratios asymptotically satisfies the conditions to set up the previous one-sample case when n is determined. We can take pilot samples of size m for X_i and Y_i, then we calculate the sample variance of the risk ratio $\hat{\theta}_m$.

Now, we can apply the results of stopping time given in Section 3.2 for the risk ratio:

$$N \equiv N(d) = max\left\{m, \left[\frac{t^2_{\alpha/2,m-1} S^2_m}{d^2}\right] + 1\right\}. \tag{3.2}$$

Thus, motivated by the stopping rule Eq.(3.2), we have the following proposed two-stage procedure:

Stage 1 (Pilot stage): Obtain X_1, X_2, ..., X_m and Y_1, Y_2, ..., Y_m. The integer m is called pilot sample size. If $N = m$, then we do not take sample any more (no further stage is needed) and establish a $(1-\alpha)100\%$ confidence interval, $I_m = [\bar{X}_m \pm d]$, which has width $2d$ for risk ratio θ.

Stage 2 (Sequential stage): If $N > m$, we sample X_{m+1}, X_{m+2}, ..., X_N and Y_{m+1}, Y_{m+2}, ..., Y_N. Thus, the total samples of each sequence is N. Therefore, the associated fixed-width confidence interval is given by $I_N = [\hat{\theta}_N - d, \hat{\theta}_N + d]$, based on all N samples of $X_i^{'}s$ and $Y_i^{'}s$.

3.4 Asymptotic Properties of the Two-Stage Procedure

We now study the properties of the proposed two-stage procedure that we stated in the previous section.

Theorem 3.1 (Finite sure termination). Let N be the stopping time associated with the proposed two-stage procedure. Then $P\{N < \infty\} = 1$

Proof. If $N = m$, since $m < \infty$, it is trivial. If $N > m$, using the stopping rule in Eq. (3.2)

$$P\{N = \infty\} = \lim_{n\to\infty} P(N > n) \le \lim_{n\to\infty} P\left(n \le \frac{t^2_{\alpha/2,m-1} S^2_m}{d^2}\right) = 0.$$

Hence the sequential procedure terminates finitely with probability one.

Theorem 3.2 For two-stage procedure in Eq. (3.2), we have

(i) $\quad P_{\theta,\sigma}\left(\theta \in I_N\right) \ge 1-\alpha,$

(ii) $\quad E_{\theta,\sigma}(N) \ge \frac{t^2_{\alpha/2,m-1}\sigma^2}{d^2},$

(iii) $\quad lim_{d\to 0} P_{\theta,\sigma}\left(\theta \in I_N\right) = 1-\alpha.$

Proof. From the set up of the procedure, (i) and (ii) can be easily verified. (Also, see Mukhopadhyay, N. and de Silva B.M., 2009.)

For (iii) we proceed as Theorem 2.2 in Chapter 2. In addition, it follows from Theorem 3.2, part (ii), that $E_{\theta,\sigma}(N/n^*) \geq 1$ as $d \to 0$.

3.5 Numerical Study

In the previous sections, we mentioned that the risk ratio for two binomial variates follows an asymptotic normal distribution, so we can use the stopping rule we derived in Section 3.2 to calculate optimal sample sizes. In the following tables, we used confidence level $\alpha = 0.05$ and pilot (first stage) sample size m =30 or 50. The results are summarized in the tables below, with the average sample size $E(N)$, optimal sample size n^* $(\geq m)$ and the coverage probability (CP). Note that the minimum sample size is m.

Table 3.1. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.1$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.501	1.002	-.002	.20	(.802, 1.202)	.898	185.18	136
.501	.500	0.999	.001	.30	(.699, 1.299)	.901	80.88	61
.502	.501	0.997	.003	.40	(.597, 1.397)	.956	49.71	34

Table 3.2. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.05$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.500	1.000	.000	.20	(.800, 1.200)	.938	265.91	193
.500	.499	0.999	.001	.30	(.699, 1.299)	.951	119.32	86
.500	.500	1.000	.000	.40	(.600, 1.400)	.959	67.62	49
.500	.500	1.001	.001	.50	(.501, 1.501)	.986	46.05	31

Table 3.3. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.1$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.500	1.000	.000	.10	(.900, 1.100)	.897	636.28	542
.500	.501	1.002	-.002	.20	(.802, 1.202)	.899	162.58	136
.501	.500	0.999	.001	.30	(.699, 1.299)	.901	74.10	61

Table 3.4. For $\theta = 1.0$ and $\rho = 0$ when $p_0 = 0.5, p_1 = 0.5$, $\alpha = 0.05$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.500	.500	1.000	.000	.20	(.800, 1.200)	.948	236.68	193
.500	.499	0.999	-.001	.30	(.699, 1.299)	.953	103.47	86
.500	.500	1.000	.000	.40	(.600, 1.400)	.988	64.65	50

Table 3.5. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.1$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.400	.601	1.502	-.502	.35	(1.152, 1.852)	.899	179.40	108
.400	.600	1.498	-.498	.45	(1.048, 1.948)	.902	106.62	66
.400	.601	1.501	-.501	.55	(.951, 2.051)	.939	71.99	44
.400	.599	1.499	-.499	.65	(.849, 2.149)	.971	55.62	32

Table 3.6. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.05$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.401	.599	1.497	-.497	.45	(1.047, 1.947)	.949	153.53	93
.398	.600	1.505	-.505	.55	(.955, 2.055)	.950	102.49	63
.399	.600	1.501	-.501	.65	(.851, 2.151)	.976	73.21	45
.401	.600	1.497	-.497	.75	(.747, 2.247)	.989	59.22	34

Table 3.7. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.1$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.400	.600	1.500	-.500	.35	(1.150, 1.850)	.901	138.76	108
.401	.600	1.497	-.497	.45	(1.047, 1.947)	.940	88.01	66
.399	.600	1.504	-.504	.55	(.954, 2.054)	.978	67.72	50

Table 3.8. For $\theta = 1.5$ and $\rho = -0.5$ when $p_0 = 0.4, p_1 = 0.6$, $\alpha = 0.05$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.401	.599	1.493	-.493	.45	(1.043, 1.943)	.950	121.74	93
.398	.600	1.505	-.505	.55	(.955, 2.055)	.982	87.33	63
.400	.600	1.501	-.501	.65	(.851, 2.151)	.994	67.89	50

Table 3.9. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.1$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.351	.699	1.994	-.994	.50	(1.494, 2.494)	.897	186.81	99
.351	.702	2.003	-1.003	.60	(1.403, 2.603)	.898	124.42	69
.350	.700	2.000	-1.000	.70	(1.300, 2.700)	.937	94.96	51
.351	.498	1.993	-.993	.80	(1.193, 2.793)	.961	76.32	39

Table 3.10. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.05$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	2.000	-1.000	.50	(1.500, 2.500)	.945	281.83	141
.349	.700	2.002	-1.002	.60	(1.402, 2.602)	.947	192.96	98
.349	.699	2.001	-1.001	.70	(1.301, 2.701)	.946	138.70	72
.351	.700	1.998	-.998	.80	(1.198, 2.798)	.967	127.26	55
.352	.699	1.995	-.995	.90	(1.095, 2.895)	.986	98.50	44

Table 3.11. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.1$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	1.999	-.999	.50	(1.499, 2.499)	.903	139.98	99
.351	.702	2.003	-1.003	.60	(1.403, 2.603)	.946	99.49	69
.350	.702	2.004	-1.004	.70	(1.304, 2.704)	.975	79.75	51

Table 3.12. For $\theta = 2.0$ and $\rho = -1.0$ when $p_0 = 0.35, p_1 = 0.7$, $\alpha = 0.05$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.350	.700	2.001	-1.001	.50	(1.501, 2.501)	.950	199.88	141
.351	.701	2.001	-1.001	.60	(1.401, 2.601)	.956	138.10	98
.349	.699	2.001	-1.001	.70	(1.301, 2.701)	.981	104.05	72
.351	.700	1.998	-.998	.80	(1.198, 2.798)	.990	84.53	55

Table 3.13. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.1$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.299	.750	2.501	-1.501	.65	(1.851, 3.151)	.869	228.14	107
.299	.749	2.501	-1.501	.75	(1.751, 3.251)	.889	187.10	81
.301	.750	2.497	-1.497	.85	(1.647, 3.347)	.917	179.01	63
.300	.751	2.503	-1.503	.95	(1.553, 3.453)	.948	122.00	50

Table 3.14. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.05$, m=30

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.299	.749	2.501	-1.501	.85	(1.651, 3.351)	.948	260.28	90
.300	.748	2.496	-1.496	.95	(1.546, 3.446)	.957	193.91	71
.300	.750	2.499	-1.499	1.05	(1.449, 3.549)	.974	141.43	59
.301	.752	2.504	-1.504	1.15	(1.354, 3.654)	.982	113.85	49

Table 3.15. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.1$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.300	.750	2.501	-1.501	.65	(1.851, 3.151)	.906	166.31	107
.301	.749	2.495	-1.495	.75	(1.645, 3.245)	.936	122.12	81
.301	.750	2.498	-1.498	.85	(1.648, 3.348)	.962	100.34	63
.300	.750	2.501	-1.501	.95	(1.551, 3.451)	.975	84.36	50

Table 3.16. For $\theta = 2.5$ and $\rho = -1.5$ when $p_0 = 0.3, p_1 = 0.75$, $\alpha = 0.05$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.299	.751	2.506	-1.506	.85	(1.656, 3.356)	.969	136.46	89
.300	.749	2.498	-1.498	.95	(1.548, 3.448)	.983	112.73	71
.301	.750	2.499	-1.499	1.05	(1.449, 3.549)	.991	95.47	59
.301	.751	2.502	-1.502	1.15	(1.352, 3.652)	.993	86.08	50

Table 3.17. For $\theta = 3.0$ and $\rho = -2.0$ when $p_0 = 0.25, p_1 = 0.75$, $\alpha = 0.1$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.250	.749	2.998	-1.998	1.00	(1.998, 3.998)	.941	145.90	82
.250	.750	3.001	-2.001	1.10	(1.901, 4.101)	.960	127.24	68
.251	.751	3.002	-2.002	1.20	(1.802, 4.202)	.975	107.21	57
.249	.750	3.004	-2.004	1.30	(1.704, 4.304)	.983	94.97	50

Table 3.18. For $\theta = 3.0$ and $\rho = -2.0$ when $p_0 = 0.25, p_1 = 0.75$, $\alpha = 0.05$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.250	.748	2.994	-1.994	1.00	(1.994, 3.994)	.954	202.28	116
.250	.752	3.010	-2.010	1.10	(1.910, 4.110)	.970	167.76	96
.249	.749	3.003	-2.003	1.20	(1.803, 4.203)	.981	145.06	81
.250	.750	3.001	-2.001	1.30	(1.701, 4.301)	.986	129.28	69
.252	.753	2.998	-1.998	1.40	(1.598, 4.398)	.992	112.38	59

Table 3.19. For $\theta = 4.0$ and $\rho = -3.0$ when $p_0 = 0.2, p_1 = 0.8$, $\alpha = 0.1$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.199	.800	4.004	-3.004	1.30	(2.704, 5.304)	.918	257.16	109
.199	.799	3.996	-2.996	1.40	(2.596, 5.396)	.937	218.09	94
.200	.799	3.995	-2.995	1.50	(2.495, 5.495)	.952	199.64	83
.198	.799	3.999	-2.999	1.60	(2.399, 5.599)	.963	163.78	71
.200	.800	3.996	-2.996	1.70	(2.296, 5.696)	.974	152.42	64

Table 3.20. For $\theta = 4.0$ and $\rho = -3.0$ when $p_0 = 0.2, p_1 = 0.8$, $\alpha = 0.05$, m=50

$\hat{p}_0$	$\hat{p}_1$	$\hat{\theta}$	$\hat{\rho}$	d	Confidence Limits	CP	$E(N)$	n^*
.201	.802	4.005	-3.005	1.50	(2.505, 5.505)	.954	310.00	117
.199	.798	3.993	-2.993	1.60	(2.393, 5.593)	.968	237.70	103
.201	.800	3.994	-2.994	1.70	(2.294, 5.694)	.982	229.61	91
.200	.800	4.000	-3.000	1.80	(2.200, 5.800)	.987	214.71	82
.199	.799	3.998	-2.998	1.90	(2.098, 5.898)	.993	189.49	73

From the above tables, we observe that the expected sample size is more than the required optimal sample size for the interval with the confidence level $1-\alpha$ and the half-length d , which satisfies the result in Theorem 3.2 (ii). When $d \to 0$, the random stopping time N approaches ∞ (w.p. 1) and the required optimal fixed-sample size n^* goes to infinity as well. Also, smaller pilot samples result in larger required expected sample sizes. Comparing to the numerical results from Chapter 2, we can see that the coverage probability in the two-stage procedure has improved, because it does over-sampled.

The pilot sample size m can be considered as a lower bound of the optimal sample size in the procedure. The simulation results provide substantial numerical evidence for us to conclude that the proposed two-stage procedure performs satisfactorily. In addition, in order to capture the more desirable and better properties of both sequential and two-stage procedure, one can consider the modified two-stage procedure for the risk ratio.

CHAPTER 4

CONCLUSION AND FUTURE RESEARCH

4.1 Concluding Remarks

In this dissertation, we have studied sequential methods for inference on the risk ratio of two independent binomial variates to construct confidence intervals with desirable length $2d$ and confidence level $1-\alpha$. Primarily, we proposed the sequential-sampling design as the fundamental, and we extended the sampling strategy to the two-stage procedure. The dynamic sampling methods such as sequential sampling or multi-stage sampling provide optimality of the sample size and flexibility to the experimenters to set up the plan more efficiently and effectively.

After proposing the dynamic sampling strategies, we explored their properties with finite samples and asymptotics. Also, through the Monte Carlo simulation, we verified finite sample behavior, numerical evidence and the performance of the proposed procedure. Perhaps, the proposed method could be applicable to some other measures of relative risk.

In addition, we compared the confidence interval based on the proposed method (i.e., Wald-based) with the likelihood-based confidence intervals. We can summarize that the proposed intervals are near-invariant. For more practical purposes, the two-stage method is recommendable to experimenters and researchers.

4.2 Future Direction

We now outline some future directions for feasible extensions as well as future problems that are closely related to our methods. One of the closest approaches we can consider is to look for the problem under the frame work of ranking and selection methodologies. It could be used the indifference zone approach or subset selection method for the problem. Both could be more directly decision-theoretic oriented method, because the selection rule provides the probability of correct decision/selection, $P(CS)$. Moreover, the problem can also be thought as the two-stage selection procedure as well.

Second, the risk-efficient estimation for the estimator could be plausible under the squared error loss incorporating with the cost of observations. We are able to estimate the risk of the ratio of two binomial proportions when the loss function is in the form of: $L_n = (\hat{\theta}_n - \theta)^2 + cn$, where θ is the risk ratio we defined previously and $c(> 0)$ is the known cost per unit of observations. The risk function associated with the optimal sample size could be derived by using sequential method or two-stage method.

APPENDIX

SOME R-CODES FOR THE SEQUENTIAL-BASED CONFIDENCE INTERVAL SIMULATION

```
# Program for the sequential procedure

# Number of iteration is 5000



mor <- 1-theta1/theta0

for(i in 1:5000){

x0 <- rbinom(5000,1,theta0)

x1 <- rbinom(5000,1,theta1)

for(j in 5:5000){

mor.0 <- 1-(sum(x1[1:j]))/(sum(x0[1:j])+1/j)

theta0.hat <- (sum(x0[1:j]))/j

sigma <- sqrt((1-mor.0))*

sqrt((2-mor.0+2*mor.0*theta0.hat-2*theta0.hat)/theta0.hat)

s <- j

if(s <= (qnorm(0.975)*sigma/d)^2) break}

MOR[i] <- 1-sum(x1)/(sum(x0))

N[i] <- s

mor.N <- 1-sum(x1[1:s])/(sum(x0[1:s])+1/s)

theta0.hat <- sum(x0)/5000
```

```
if(1-theta1/theta0 < mor.N+d & 1-theta1/theta0 <mor.N-d)

cp[i]=1

}

sigma <- sqrt((1-mor)*(2-mor+2*mor*theta0-2*theta0)/theta0)

n <- floor((qnorm(0.95)*sigma/d)^2)+1
```

BIBLIOGRAPHY

Anscombe, F. J. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.*, 48:600–607.

Bechhofer, R. E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.*, 25:16–39.

Bechhofer, R. E., Kiefer, J., and Sobel, M. (1968). *Sequential Identification and Ranking Proceduress*. University of Chicago Press: Chicago.

Cho, H. (2007). Sequential risk-efficient estimation for the ratio of two binomial proportions. *Journal of statistical planning and inference.*, 137:2336–2346.

Cho, H. and Govindarajulu, Z. (2008). Sequential confidence limits for the ratio of two binomial proportions. *International journal of statistics and management system.*, 3:26–42.

Chow, Y. S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.*, 36:457–462.

Chow, Y. S., Robbins, H., and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton Mifflin: Boston.

Dantzig, G. B. (1940). On the non-existence of tests of "student's" hypothesis having power functions independent of sigma. *Ann. Math. Statist.*, 11:186–192.

Ghosh, B. K. (1970). *Sequential Tests of Statistical Hypotheses*. Addison-Welsley: Reading.

Ghosh, M. and Mukhopadhyay, N. (1981). Consistency asymptotic efficiency of two-stage and sequential estimation procedures. *Sankyha Ser*, A 43:220–227.

Ghosh, M., Mukhopadhyay, N., and Sen, P. K. (1997). *Sequential Estimation*. Wiley: New York.

Ghosh, M. and Sen, P. K. (1991). Sequential confidence intervals for regression coefficients based on kendall's tau. *Calcutta Statist*, 20:23–26.

Gibbons, J. D., Olkin, I., and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology.* Wiley: New York. Reprinted 1999, SIAM: Philadelphia.

Govindarajulu, Z. (1981). *The Sequential Statistical Analysis.* American Sciences Press: Columbus.

Govindarajulu, Z. (2004). *Sequential Statistics.* World Scientific Publishing: Singapore.

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. Wiley: New York. Reprinted 2002, SIAM: Philadelphia.

Hochberg, Y. and Tamhane, A. C. (1987). *Multiple comparison methods*. New York: Wiley.

Hsu, J. C. (1996). *Multiple comparisons: theory and methods.* Chapman and Hall, London.

Lai, T. L. and Siegmund, D. (1977). A nonlinear renewal theory with applications to sequential analysis. *Ann. Statist.*, 5:946–954.

Lai, T. L. and Siegmund, D. (1979). A nonlinear renewal theory with applications to sequential analysis ii. *Ann. Statist.*, 1:60–76.

Mukhopadhyay, N. (1980). A consistent and asymptotically efficient two-stage procedure to construct fixed-width confidence intervals for the mean. *Metrika*, 27:281–284.

Mukhopadhyay, N. and de Silva B.M. (2009). *Sequential methods and their applications.* Marcel Dekker: New York.

Mukhopadhyay, N. and Solanky, T. (1994). *Multistage Selection and Ranking Procedures: Second-Order Asymptotics.* Marcel Dekker, Inc.

Ray, W. D. (1957). Sequential confidence intervals for the mean of a normal population with unknown variance. *J. Roy. Statist.*, B 19:133–143.

Robbins, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics*, The Herald Cramer:235–245.

Sen, P. K. (1981). *Sequential nonparametrics: invariance principles and statistical inference.* New York: Wiley.

Sen, P. K. (1985). Theory and applications of sequential nonparametrics. *Society for Industrial and Applied Mathematics.*, page 100.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Addison-Welsley: Reading.

Stein, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.*, 16:243–258.

Stein, C. (1949). Some problems in sequential estimation. *Econometrica*, 17:77–78.

Von Bahr, B. (1969). On the convergence of moments in the central limit theorem. *Ann. Math. Statist*, 36:808–818.

Wald, A. (1947). *Sequential Analysis*. Wiley: New York.

Woodroofe, M. (1977). Second-order approximations for sequential point and interval estimation. *Ann. Statist.*, 5:984–995.

VITA

Graduate College
University of Nevada, Las Vegas, USA

Zhou Wang

Degrees:

Bachelor of Statistics, 2009
University of Science and Technology of China

Thesis Title:

A Study of Sequential Inference for the Risk Ratio and Measure of Reduction of Two Binomials

Thesis Examination Committee:

Chairperson, Hokwon Cho , Ph.D.
Committee Member, Malwane Ananda, Ph.D.
Committee Member, Jessica Jaynes, Ph.D.
Graduate Faculty Representative, Paulo Pinheiro, Ph.D.