SMALL SAMPLE INFERENCE FOR COLLECTIONS OF BERNOULLI TRIALS

BY LU XU

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

John E. Kolassa

and approved by

New Brunswick, New Jersey

January, 2010

ABSTRACT OF THE DISSERTATION

Small Sample Inference for Collections of Bernoulli Trials

by LU XU

Dissertation Director: John E. Kolassa

This dissertation discusses two applied problems solved by saddlepoint approximation methods. The first part of this thesis concerns continuity corrected saddlepoint approximations for testing and confidence intervals for the difference of two independent binomial proportions. We propose two new continuity corrections, and compare them with the other four continuity corrections. Continuity corrections may give a more accurate approximations to the tail area of discrete random variables. To get a confidence interval for the difference of two independent binomial proportions, we proposed continuity corrections $1/2LCM(m,n)$ with the least common multiple (LCM) of the two binomial distributions sample sizes m and n considered by Xu and Kolassa[31], and $R/2LCM(m,n)$, an adjusted version of $1/2LCM(m,n)$ with R by Xu and Kolassa[32], a heuristic factor calculator from the standard error of the marginal null binomial distribution. We compare exact coverage probabilities for intervals calculated by using saddlepoint approximation with different corrections. The primary criterion for evaluating the corrections is agreement of actual with nominal coverage probabilities. Because of the ordering of the continuity corrections, the coverage probabilities will be ordered similarly. For all the cases considered with minimum expected cell size of at least 1, numerical results indicate that $R/2LCM(m,n)$ has coverage probabilities very close to

the nominal 95% and 90% even for the minimum of sample sizes as small as $5 - 9$, and $R/2LCM(m, n)$ improved uncorrected saddlepoint approximation methods moderately for the nominal 95% and 90% intervals; however, that the Yates continuity correction $(2m)^{-1} + (2n)^{-1}$ is unnecessarily conservative for 95% and 90% but reasonable for 99% intervals.

The second part of this thesis concerns the sequential likelihood ratio test using in computerized adaptive testing. We consider sequential testing techniques, including the truncated sequential probability ratio test and the Haybittle-Peto test. Both of these tests in their original forms rely on approximate normality of the signed roots of the log likelihood ratio tests, and approximate boundary crossing probabilities for discrete normal-theory random walks. Bartroff, Finkelman and Lai[5] modify these techniques by using Monte Carlo approximations to calibrate the truncation boundary. We propose a hybrid Monte Carlo-Asymptotic approach, in which we substitute an easy Monte Carlo approximation in place of boundary crossing probabilities for Brownian motions, and use asymptotic approximations for the distribution of the signed root of the likelihood ratio test statistic. We found that after selecting stopping boundaries using normal-based Monte Carlo calculations, reliance on asymptotic normality of the signed root of the log likelihood ratio statistics provided adequate control of Type I error, without recourse to more complicated Monte Carlo operations. We also observe markable improvement using Barndorff-Nielsen's r^* formula (Barndorff-Nielsen, 1991).

Acknowledgements

First, I would like to express my deepest appreciation to my adviser: Dr. John E. Kolassa and thank him for their valuable ideas, helpful comments, great support, continuous encouragement and kind patience. Without his help and advice, I can not imagine completing my dissertation. I feel very lucky to have worked with him and have benefited from his valuable experience. In addition, I would like to extend my profound gratitude to Dr. Donald R. Hoover, Dr. Minge Xie and Dr. Pamela Ohman-Strickland for their great advice and the time they dedicated to reviewing my thesis.

I would like to thank the entire Department of Statistics and Biostatistics and the Institute for Health, Health Care Policy and Aging Research at Rutgers for support in the past four years, with special thanks to Dr. Donald R. Hoover for financial support. My special thanks also go to Drs. Javier Cabrera, Richard F. Gundy, Rebecka Jornstein, William Edward Strawderman, Lawrence Shepp, David E. Tyler, Minge Xie, Cun-Hui Zhang for their wonderful lectures.

I also want to thank my supervisors of my internships from Sanofi-Aventis, Dr. Lynn Wei, Dr. Hui Quan and Dr. Lixia Jiao. When I need help, they are always there.

Finally, my most heartfelt thanks go to my dearest parents and husband, for their selfless love and support!

This research is partially supported by the National Science Foundation grant 0906569

Dedication

I dedicate my dissertation work to my family and many friends. A special feeling of gratitude to my loving parents, Yonghai Xu and Zenghua Li whose words of encouragement and push for tenacity ring in my ears, and they have never left my side and are very special.

I also dedicate this dissertation to my many friends who have supported me throughout the process. I will always appreciate all they have done, especially Juan Zhang and Jixin Li for helping me develop my technology skills.

I dedicate this work and give special thanks to my wonderful husband Yufeng Tang for being there for me throughout the entire doctorate program. You have been my best cheerleader.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Inference on the difference between two independent binomial proportions is discussed in this thesis. Table 1.1 shows the notation adopted for the comparison of two independent binomial proportions. It is assumed that the denominators m and n are fixed.

Table 1.1: Notation for comparison of the two independent proportions

Observed frequencies:

	sample	
	1	2
+	a	b
−	c	d
Total	m	n

Theoretical proportions:	Observed proportions:
$\mathrm{E}A/m = \pi_1$	$p_1 = a/m$
$\mathrm{E}B/n = \pi_2$	$p_2 = b/n$

Here A and B denote random variables of which a and b are realizations.

Reparameterization:
Parameter of interest $\theta = \pi_1 - \pi_2$
Nuisance parameter $\psi = (\pi_1 + \pi_2)/2$

For interval estimation for difference of two independent binomial proportions, Hauck[13] compare seven methods with various continuity corrections for large samples based on the normal approximation. For small samples based on the saddlepoint approximation, we propose two new continuity corrections. With the least common multiple of two sample sizes m and n notated as $LCM(m,n)$, one is $1/2LCM(m,n)$, and the other is the adjusted version $R/2LCM(m,n)$ with $R = 1 - 2\sqrt{\frac{a+b}{m+n}(1 - \frac{a+b}{m+n})}$. R is an alternative factor calculator from the standard error of the marginal null binomial distribution. We compare these two new continuity corrections together with other three corrections discussed by Hauck and Anderson[13], such as Yates correction, two corrections with maximum and minimum of two sample sizes respectively, and non-correction, by a simulation study which mimicked Newcombe[26]. To compare the continuity corrections for the testing problem, we investigate exact type I error; for the confidence interval problem, we investigate the coverage probabilities of the confidence interval of the difference of two independent binomial proportions.

Motivated by Bartroff, Finkelman and Lai[5], we also evaluate the performance of the saddlepoint approximation comparing to normal approximation to distribution of the signed square root of the likelihood ratio statistic for the problem of classifying examinees as either masters or non-masters in a given content area, known as computerized mastery testing (CMT), can be formalized by setting a cut point θ_0 and defining an examinee as a master if and only if his/her ability level θ meets or exceeds that cut point. A computerized mastery testing typically assumes a region (θ_-, θ_+) containing θ_0. The statistical hypothesis of mastery is then given by $H_0 : \theta \geq \theta_+$, while the hypothesis of non-mastery is given by $H_1 : \theta \leq \theta_-$.

1.2 The contribution of this dissertation

The thesis divided into three parts. Each part is innovative and can be reviewed separately. Meanwhile, they are closely related to each other, especially the first two parts. Integration of the first two parts can provide a satisfactory answer that achieves the goal of this study.

1.2.1 Comparing six continuity-corrected saddlepoint approximation about the difference of two binomial probabilities

The first two parts of the thesis consider tests and the confidence intervals for the difference between two binomial probabilities. I propose two innovative continuity corrections for saddlepoint approximation to the distribution of the sufficient statistic conditional on the profile estimate of nuisance parameter, which are intuitive and understandable.

The test statistic $A/m - B/n$ has support on the δ-lattice $L_\delta = \{\gamma, \gamma \pm \delta, \gamma \pm 2\delta, ...\}$. In this lattice case, there are three modification to the formula of saddlepoint approximation, we compare simulation results and provide R code for all three modifications. Though extensive research on inference concerning the difference between two independent binomial probabilities had been done in the past, these innovative continuity corrections make an important contribution.

For 90% and 95% confidence intervals for difference of two binomial proportions, the saddlepoint approximation with the new continuity correction CC_R is recommended when the minimum expected cell size (MCS) is not smaller than 1 or the nuisance parameter $\psi =$ is moderate $(.4 - .6)$ or the parameter in interest θ is large $(\theta \geq .5)$; while the confidence intervals with continuity corrections CC_S, CC_A and CC_Y are unnecessarily conservative and could not acceptable. When MCS ≤ 1, CC_S is the best choice of continuity correction, and it also the best choice for extreme $\psi \geq .1$ or $\psi \leq .9$, or small θ in $0 - .05$. And the most useful result is that for MCS > 1, CC_R has coverage probabilities very close to the nominal 95% and 90% for any range of NMIN, even for NMIN as small as $5 - 9$. For the 99% intervals, the Yates correction CC_Y is reasonable.

1.2.2 Saddlepoint approximation for the distribution of the signed root of likelihood ratio statistic in modern sequential analysis

We consider sequential testing techniques, including the truncated sequential probability ratio test and the Haybittle-Peto test. Both of these tests in their original forms rely on approximate normality of the signed roots of the log likelihood ratio tests, and approximate boundary crossing probabilities for discrete normal-theory random walks. Bartroff, Finkelman and Lai[5] modify these techniques by using Monte Carlo approximations to calibrate the truncation boundary. We propose a hybrid Monte Carlo-Asymptotic approach, in which we substitute an easy Monte Carlo approximation in place of boundary crossing probabilities for Brownian motions, and use asymptotic approximations for the distribution of the signed root of the likelihood ratio test statistic. We found that after selecting stopping boundaries using normal-based Monte Carlo calculations, reliance on asymptotic normality of the signed root of the log likelihood ratio statistics provided adequate control of Type I error, without recourse to more complicated Monte Carlo operations. We also observe markable improvement using Barndorff-Nielsen's r^* formula (Barndorff-Nielsen[1]). Therefore, the sequential likelihood ratio test with saddlepoint approximation could be recommended into the practice application.

1.3 Outline

The thesis contains five chapters. The first chapter is an introduction. The next two chapters discuss continuity-corrected saddlepoint approximation to the tail probabilities for the difference of two binomial proportions in testing and confidence interval applications. To test the difference of two independent binomial proportions, we propose a new continuity correction based on the least common multiple of two sample sizes m and n notated as $LCM(m,n)$, and compare this new continuity correction together with other three corrections discussed by Hauck and Anderson[13] and non-correction by simulation study which mimicked the one by Newcombe[26]. In the following chapter we compare the coverage probabilities of the confidence interval of the difference of two independent binomial proportions, adding a continuity correction which is an adjusted version of the one based on $LCM(m,n)$. In the fourth chapter, we discuss the saddlepoint approximation used in sequential likelihood ratio test for computerized mastery testing(CMT) and how to determine the stopping boundaries for such a test.

Chapter 2

Testing The Difference of Two Binomial Proportions: Comparison of Continuity Corrections for Saddlepoint Approximation

2.1 Introduction

We carried out a simulation study based on the methodology of Newcombe[26] to compare tests for the difference of two binomial proportions by applying different continuity corrections on saddlepoint approximation to tail probabilities. In this paper we proposed a new continuity correction based on the least common multiple of two sample sizes. By our criteria, the best test should have the exact type I error rates that are, on the whole, closest to α, but not exceeding α, where α is nominal level of significance.

We investigated the effect of the continuity corrections on tail probability approximations by checking the exact type I errors rate. The upper tail and lower tail probabilities were calculated by the Lugannani and Rice formula for the saddlepoint approximation to tail probabilities to be introduced below. Based on the simulation results, we could tell whether the test was improved by applying continuity corrections to approximation and which continuity correction was the best in different cases. Newcombe[26] used notations for the comparison of two independent binomial proportions as in Table 1.1. It was assumed that the denominators m and n are fixed. To test $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$, we had $D = p_1 - p_2$ as the point estimate of $\theta = \pi_1 - \pi_2$.

Claim: For fixed $a/m - b/n = d$, $\min[i/m - j/n | i \in 0, ..., m, j \in 0, ..., n, i/m - j/n > d] = d + 1/\text{LCM}(m,\ n)$ for m, n positive integers.

Proof. Let $v = \text{LCM}(m,\ n)$ then $v = mn/(m,n)$ with the notation (m,n) for the greatest common divisor of m and n.

First to show that $\exists\ i, j, \ni \frac{i}{m} - \frac{j}{n} = d + \frac{1}{v}$. We have

$$\begin{aligned}
\frac{i}{m} - \frac{j}{n} = d + \frac{1}{v} \quad &\Leftrightarrow \quad \frac{i}{m} - \frac{j}{n} = \frac{a}{m} - \frac{b}{n} + \frac{1}{v} \\
&\Leftrightarrow \quad \frac{i}{m} - \frac{j}{n} = \frac{a}{m} - \frac{b}{n} + \frac{(m,n)}{mn} \\
&\Leftrightarrow \quad \frac{in - jm}{mn} = \frac{an - bm + (m,n)}{mn} \\
&\Leftrightarrow \quad in - jm = an - bm + (m,n) \\
&\Leftrightarrow \quad (i-a)n - (j-b)m = (m,n)
\end{aligned}$$

By Lemma 1.3.1 (Herstein 1975), $\exists\ i, j, \ni (i-a)n - (j-b)m = (m,n)$.

Then to show that if $\frac{i}{m} - \frac{j}{n} > d$, then $\frac{i}{m} - \frac{j}{n} \geq d + \frac{1}{v}$.

$$\begin{aligned}
\frac{i}{m} - \frac{j}{n} > d \quad &\Leftrightarrow \quad \frac{i}{m} - \frac{j}{n} > \frac{a}{m} - \frac{b}{n} \\
&\Leftrightarrow \quad \frac{(i-a)n - (j-b)m}{mn} > 0 \\
&\Leftrightarrow \quad \frac{\frac{(i-a)n}{(m,n)} - \frac{(j-b)m}{(m,n)}}{v} > 0 \\
&\Leftrightarrow \quad \frac{(i-a)n}{(m,n)} - \frac{(j-b)m}{(m,n)} > 0
\end{aligned}$$

Since $i-a$, $j-b$, $\frac{n}{(m,n)}$ and $\frac{m}{(m,n)}$ are all integers, $\frac{(i-a)n}{(m,n)} - \frac{(j-b)m}{(m,n)}$ is an integer. Then

$$\begin{aligned}
\frac{(i-a)n}{(m,n)} - \frac{(j-b)m}{(m,n)} \geq 1 \quad &\Leftrightarrow \quad \frac{\frac{(i-a)n}{(m,n)} - \frac{(j-b)m}{(m,n)}}{v} \geq \frac{1}{v} \\
&\Leftrightarrow \quad \frac{i}{m} - \frac{j}{n} \geq \frac{a}{m} - \frac{b}{n} + \frac{1}{v}. \square
\end{aligned}$$

Based on the least common multiple of two sample sizes m and n notated as $LCM(m,n)$, we proposed a new continuity correction for saddlepoint approximation to tail probabilities, namely

$$\text{CC}_L = \{2\ LCM(m,n)\}^{-1}.$$

Besides this new continuity correction, we also considered the uncorrected form

$$\text{CC}_N = 0,$$

and three continuity corrections used by Hauck and Anderson[13],

$$\mathrm{CC}_Y = (2m)^{-1} + (2n)^{-1},$$

$$\mathrm{CC}_S = \{2\max(m,n)\}^{-1},$$

$$\mathrm{CC}_A = \{2\min(m,n)\}^{-1}.$$

Clearly, the continuity correction terms have the order

$$\mathrm{CC}_N \leq \mathrm{CC}_L \leq \mathrm{CC}_S \leq \mathrm{CC}_A \leq \mathrm{CC}_Y. \tag{2.1}$$

2.2 Justification of Approximation Method

An approximation to $P[D = d]$ is desired, based on its cumulant generating function $K(\tau)$, where

$$K_D(\tau) = m\log((1-\pi_1) + \pi_1 e^{\frac{\tau}{m}}) + n\log((1-\pi_2) + \pi_2 e^{-\frac{\tau}{n}}), \tag{2.2}$$

Let $P_\tau[D = d] = P[D = d]\exp(\tau d - K(\tau))$. Using Gaussian approximation to $P_\tau[D = d]$ by Kolassa[20]:

$$\frac{\exp(-(d-K'(\hat{\tau}))^2 K''(\hat{\tau})^{-1}/2)}{\sqrt{2\pi K''(\hat{\tau})}} \times (2CC) = \frac{1}{\sqrt{2\pi K''(\hat{\tau})}} \times (2CC), \tag{2.3}$$

where $\hat{\tau}$ is given by the solution to $K'(\hat{\tau}) = d$, implying the approximation to $P[D = d]$, $-1 \leq C_1 \leq d \leq C_2 \leq 1$:

$$P[D = d] \cong \frac{\exp(K(\hat{\tau}) - \hat{\tau}d)}{\sqrt{2\pi K''(\hat{\tau})}} \times (2CC), \tag{2.4}$$

The upper tail probability approximation is

$$P[D \geq d] \cong \sum_{f=d}^{C_2} \frac{\exp(K(\hat{\tau}) - \hat{\tau}f)}{\sqrt{2\pi K''(\hat{\tau})}} \times (2CC), \tag{2.5}$$

and by Euler-McLauren Summation Formula[14] the corrected version of the Lugannani-Rice formula [12] for $P[D \geq d]$ is

$$1 - \Phi(\hat{\omega}) + \phi(\hat{\omega})(\frac{1}{\hat{z}} - \frac{1}{\hat{\omega}}), \tag{2.6}$$

where $d^\star = d - CC$, $\hat{\tau}$ is the solution of $K'(\hat{\tau}) = d^\star$, $\hat{\omega} = \sqrt{-2(K(\hat{\tau}) - \hat{\tau}d^\star)}$, $\hat{z} = \hat{\tau}\sqrt{K''(\hat{\tau})}$, and Φ and ϕ are the distribution and density functions of the standard normal distribution. The argument is discussed in details by Kolassa[20] in section 5.5.

Similarly, the lower tail probability approximation is

$$P[D \leq d] \cong \sum_{f=C_1}^{d} \frac{\exp(K(\hat{\tau}) - \hat{\tau}f)}{\sqrt{2\pi K''(\hat{\tau})}} \times (2CC), \qquad (2.7)$$

and the corrected version of the Lugannani-Rice formula for $P[D \leq d]$ is

$$\Phi(\hat{\omega}) - \phi(\hat{\omega})(\frac{1}{\hat{z}} - \frac{1}{\hat{\omega}}), \qquad (2.8)$$

where $d^\star = d + CC$, $\hat{\tau}$ is the solution of $K'(\hat{\tau}) = d^\star$, $\hat{\omega} = \sqrt{-2(K(\hat{\tau}) - \hat{\tau}d^\star)}$ and $\hat{z} = \hat{\tau}\sqrt{K''(\hat{\tau})}$.

2.3 Simulation Study

Mimicking simulations by Newcombe[26], we compared the exact type I error rates of the five continuity corrections. The main evaluation of type I error was based on a sample of 9200 parameter space points (m, n, η, θ), with m and n between 5 and 50 inclusive. We chose a subset of 230 out of 2116 possible (m, n) pairs (Figure 2.1), including 46 diagonal pairs with $m = n$ and 92 pairs (m, n) with $m \neq n$ and the corresponding reversed pairs (n, m).

For each of $m = 5, 6, ..., 50$, two values of n were sampled, avoiding diagonal elements, duplicates and mirror-image pairs. For each of 230 (m, n) pairs, 40 pairs of (η, θ) were sampled with $\theta = \lambda\{1 - |2\eta - 1|\}$ and η and λ from $U(0, 1)$. The sampled values of θ had median 0.201, and quartiles 0.102 and 0.388. For each (m, n, η, θ) of the parameter space, the exact type I error rate was cumulated by the frequencies of all possible outcomes whose $p-$values were smaller than the nominal level of significance.

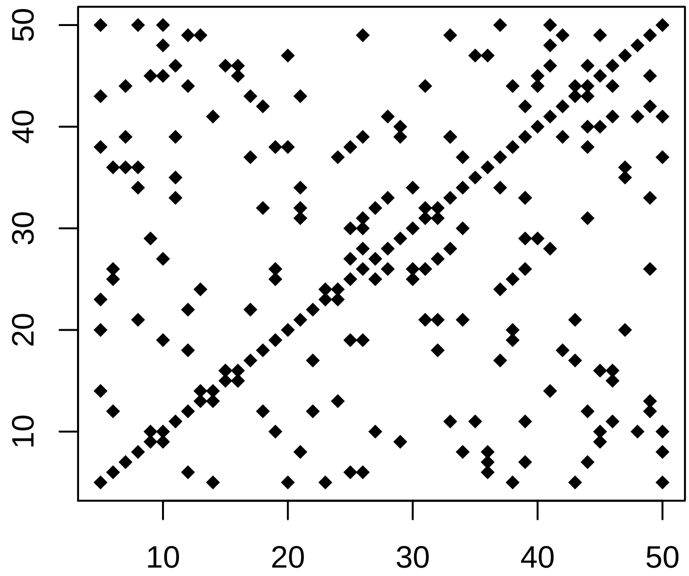

Figure 2.1: 230 (m, n) pairs chosen from 2116 possible pairs

2.4 Results

We obtained the exact type I error rates for the nominal level of significance $\alpha = 0.1$, 0.05 and 0.01 for the same set of 9200 parameter space points. We ranked performance of these continuity corrections as following: among the corrections whose exact type I error rates were lower than α, we ranked them by rate from largest to smallest. Then, among the corrections whose exact type I error rates were higher than α, we ranked them by rate from smallest to largest. In this scheme, continuity corrections leading to exact type I error rates larger than α were always ranked behind those with rates no larger than α.

Table 2.1 showed the estimated exact type I error rates and the ranks of continuity corrections for nominal $\alpha = 0.1$, 0.05 and 0.01 over the 9200 parameter space points. The average rates of uncorrected form CC_N were always higher than the nominal α levels. For $\alpha = 0.1$, all the four corrected forms had average rates below α, and the new proposed correction CC_L was the best. For $\alpha = 0.05$, CC_S was the best correction, while CC_L had an average rate a bit over α. For $\alpha = 0.01$, all the average rates were higher than α.

Table 2.2 illustrated the relation of estimated exact type I error rates for nominal

Table 2.1: Estimated exact type I error rates and ranks of the continuity corrections for nominal $\alpha = 0.1$, 0.05 and 0.01 for all 9200 points in parameter space

cc	$\alpha = 0.1$		$\alpha = 0.05$		$\alpha = 0.01$	
	rate	*rank*	rate	*rank*	rate	*rank*
CC_N	0.1067	*5*	0.0575	*5*	0.0187	*3*
CC_L	0.0956	*1*	0.0514	*4*	0.0172	*2*
CC_Y	0.0608	*4*	0.0395	*3*	0.0250	*5*
CC_S	0.0793	*2*	0.0428	*1*	0.0153	*1*
CC_A	0.0694	*3*	0.0405	*2*	0.0196	*4*

$\alpha = 0.05$ to the parameters: the lower of the two denominators m and n (min(m,n)), and the minimum expected cell frequency (min[$m\pi_1$, $m(1-\pi_1)$, $n\pi_2$, $n(1-\pi_2)$]). As expected, for each correction the average rate became lower as min(m,n) increased. All rates were higher than α for min(m,n) from 5 to 9, and CC_S was the best. But for min(m,n) from 30 to 50, the rates of all corrections except CC_N dropped below α, and CC_L beat CC_S to be the best. The pattern for minimum expected cell frequency was quite similar. CC_S was the best when the minimum expected cell frequency was below 1, and CC_L was the best when all the expected frequencies were over 5.

Table 2.2: Estimated exact type I error rates and ranks of the continuity corrections for $\alpha = 0.05$ in parameter subspaces determined by min(m,n) and minimum expected frequency

cc	min(m,n)				minimum expected frequency			
	$5-9$ (1800)		$30-50$ (2520)		$0-1$ (1503)		$5-25$ (3942)	
	rate	*rank*	rate	*rank*	rate	*rank*	rate	*rank*
CC_N	0.0665	*5*	0.0526	*5*	0.0899	*3*	0.0507	*5*
CC_L	0.0587	*3*	0.0465	*1*	0.0824	*2*	0.0453	*1*
CC_Y	0.0617	*4*	0.0304	*4*	0.1275	*5*	0.0222	*4*
CC_S	0.0523	*1*	0.0380	*2*	0.0741	*1*	0.0367	*2*
CC_A	0.0558	*2*	0.0364	*3*	0.0988	*4*	0.0310	*3*

Table 2.3 illustrated the relation of estimated exact type I error rates for nominal $\alpha = 0.05$ to the parameters η and θ. For the nuisance parameter η from 0.4 to 0.6 or the parameter of interest θ larger than 0.5, all the average rates were higher than α, and CC_S was the best. For η outside interval $(0.1, 0.9)$ or θ smaller than 0.05, the rates

of the four corrected forms dropped below α and CC_L was the best.

Table 2.3: Estimated exact type I error rates and ranks of the continuity corrections for $\alpha = 0.05$ in parameter subspaces determined by η and θ

cc	η in range				θ in range			
	≤ 0.1 or ≥ 0.9 (920)		$0.4 - 0.6$ (3220)		$0 - 0.05$ (1380)		$0.5 - 1$ (1610)	
	rate	*rank*	rate	*rank*	rate	*rank*	rate	*rank*
CC_N	0.0508	*5*	0.0695	*4*	0.0510	*5*	0.0880	*3*
CC_L	0.0410	*1*	0.0645	*2*	0.0431	*1*	0.0827	*2*
CC_Y	0.0116	*4*	0.0774	*5*	0.0160	*4*	0.1323	*5*
CC_S	0.0296	*2*	0.0562	*1*	0.0338	*2*	0.0737	*1*
CC_A	0.0198	*3*	0.0659	*3*	0.0242	*3*	0.1015	*4*

The simulation results showed that the exact type I error rates of uncorrected form CC_N were always higher than the nominal α, and in general actual rates were reduced by applying corrections to the approximation. For the cases such as larger values of m and n, larger expected cell frequencies and very small θ, the new proposed continuity correction CC_L had best performance on improving the test among the four corrections we considered in this paper.

2.5 Conclusion

We had discussed the effect of the continuity corrections on approximations to the upper tail probabilities for testing the difference of two binomial proportions. We had shown that some correction to the approximation was required to improve the test, and the uncorrected form should not be used in any case. Our new proposed continuity correction CC_L was recommended for the test with nominal 0.1 level, and it was also recommended for the test with nominal 0.05 level in any following case: the values of m and n were both large; all the expected frequencies were over 5; the nuisance parameter η was over 0.9 or below 0.1; or the parameter of interest θ was very small. For other cases we recommend CC_S in general.

Chapter 3

Continuity-corrected Saddlepoint Approximation for Confidence Interval of the Difference Between Two Binomial Proportions: Comparison of Six Continuity Corrections

3.1 Introduction

Continuity corrections may give a more accurate approximation in the tail area of discrete random variables. To get the confidence interval for the difference of two independent binomial proportions we proposed continuity corrections 1/(2 LCM) with the least common multiple (LCM) of the two sample sizes considered by Xu and Kolassa[31], and R/(2 LCM), an adjusted version of 1/(2 NLCM) with R= $1 - 2\sqrt{\frac{a+b}{m+n}(1 - \frac{a+b}{m+n})}$. By studying the coverage probabilities, we can compare tail area saddlepoint approximation with different corrections. The primary criterion for evaluating the corrections is agreement of actual with nominal coverage probabilities. Because of the ordering of the continuity corrections, the coverage probabilities will be ordered similarly. For all the cases considered with minimum expected cell size of at least 1, numerical results indicate that R/(2 LCM) has coverage probabilities very close to the nominal 95% and 90% even for the minimum of sample sizes as small as $5 - 9$, and R/(2 LCM) improved uncorrected saddlepoint approximation method moderately for the nominal 95% and 90% intervals. However, the Yates correction is unnecessarily conservative for 95% and 90% but reasonable for 99% intervals.

Hauck and Anderson[13] compared some methods for constructing confidence intervals for the difference of two binomial probabilities based on the normal approximation to the distribution of the difference of two independent sample proportions.

They concluded that a continuity correction is needed. The purpose of this chapter is to investigate the effect of the continuity corrections on saddlepoint approximations to tail probability by checking coverage probabilities. The upper tail and lower tail probabilities were calculated by the Lugannani and Rice formula for the saddlepoint approximation to tail probabilities. Again we carried out a simulation study based on the methodology of Newcombe[26] to compare the confidence intervals for the difference of two binomial proportions by applying different continuity corrections on saddlepoint approximation to tail probabilities. Based on the simulation results, we could tell whether the confidence interval was improved by applying continuity corrections to approximation and which continuity correction was the best to be applied in different cases. Newcombe[26] used notation for the comparison of two independent binomial proportions as in table 1.1. It was assumed that the denominators m and n are fixed.

Since Upton's (1982) results for the chi-squared tests found that the usual, uncorrected test was too liberal and the Yates corrected test was too conservative, we began by assuming that these general results also would hold for the saddlepoint approximation approach. We consider different choices of continuity corrections to widening the uncorrected intervals, treating the Yates correction as giving the maximum widening. We notice that for fixed $a/m - b/n$, the next value greater than it is $1/\text{LCM}(m,\ n)$. In chapter 2, we demonstrated the following claim,

Claim: For fixed $a/m - b/n = d$, $\min[i/m - j/n | i \in 0, ..., m, j \in 0, ..., n, i/m - j/n > d] = d + 1/\text{LCM}(m,\ n)$ as m, n are positive integers.

We proposed a new continuity correction for saddlepoint approximation to tail probabilities based on the the least common multiple of two sample sizes m and n notated as $\text{LCM}(m,\ n)$ (Xu and Kolassa[31])

$$\text{CC}_L = \{2\text{LCM}(m,\ n)\}^{-1},$$

and now we consider an adjusted version of it, namely,

$$\text{CC}_R = \left[1 - 2\sqrt{\frac{a+b}{m+n}(1 - \frac{a+b}{m+n})}\right] \{2\text{LCM}(m,\ n)\}^{-1},$$

the argument for this correction CC_R is primarily heuristic: From a marginal perspective, the continuity correction is most necessary for probabilities near 0 or 1, and least necessary for more moderate values, since $A/m - B/n$ is least discrete for probabilities near half.

Consider also the Yates correction

$$CC_Y = (2m)^{-1} + (2n)^{-1},$$

and other two continuity corrections discussed by Hauck and Anderson[13],

$$CC_S = \{2\max(m,\ n)\}^{-1},$$

$$CC_A = \{2\min(m,\ n)\}^{-1}.$$

Clearly, these continuity correction terms have the order

$$CC_N \leq CC_R \leq CC_L \leq CC_S \leq CC_A \leq CC_Y. \tag{3.1}$$

Because of the ordering of the continuity corrections (3.1), the coverage probabilities will be ordered similarly. The coverage probability is defined as $P[L \leq \theta \leq U]$, where L and U are the calculated limits. Several kinds of aberrations can arise. In the extreme case, where $\hat{\theta} = a/m - b/n = +1$, it is appropriate that $U = \hat{\theta}$, likewise that $L = \hat{\theta}$ when $\hat{\theta} = -1$. In the case RZ (two zeros in the same row) the uncorrected method gives zero width interval at 0, and in the case DZ (two zeros on the same diagonal) the uncorrected method gives zero width interval at +1 (or -1).

3.2 Justification of Approximation Method

We use the saddlepoint approximation to the tail probabilities. Kolassa[20] and Butler[6] had discussed the details of the saddlepoint approximation applied to the continuous and discrete random variables. Suppose continuous random variable X has cumulative distribution functions (CDF) F and cumulant generating function (CGF) K with mean $\mu = E(X)$. The saddlepoint approximation for $F(x)$, as introduced in Lugannani and

Rice[25], is

$$\hat{F}(x) = \begin{cases} \Phi(\hat{\omega}) + \phi(\hat{\omega})(1/\hat{\omega} - 1/\hat{z}) & \text{if} \quad x \neq \mu \\ \frac{1}{2} + \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{\frac{3}{2}}} & \text{if} \quad x = \mu \end{cases} \tag{3.2}$$

where

$$\begin{aligned} \hat{\omega} &= \text{sgn}(\hat{s})\sqrt{2(\hat{s}x - K(\hat{s}))}, \\ \hat{z} &= \hat{s}\sqrt{K''(\hat{s})}, \end{aligned} \tag{3.3}$$

are functions of x and saddlepoint $\hat{s}$. Here $\hat{s}$ is the implicitly defined function of x given as the unique solution to $K'(\hat{s}) = x$. Symbols ϕ and Φ denote the standard normal density and CDF respectively and $\text{sgn}(\hat{s})$ captures the sign $\pm$ for $\hat{s}$.

The bottom expression in (3.2) defines the approximation at the mean of X, when $\hat{s} = 0$. In this case $\hat{\omega} = 0 = \hat{z}$, and the last factor in the top expression of (3.2) is undefined. As $x \to \mu$ the limiting value of the top expression is the bottom expression; Thus, the entire expression is continuous and, more generally, continuously differentiable or "smooth".

Discrete CDF approximation requires modification to the formula for the continuous CDF approximation in order to achieve the greatest accuracy. Daniels [12] introduced two such modifications. Suppose X has CDF $F(k)$ with support on the integers and mean μ. Rather than considering the CDF value $F(k)$, the right tail probability $\Pr(X \geq k)$ is approximated instead to avoid some difficult notational problems. The two continuity corrections of Daniels (1987) along with a third approximation are presented below.

Suppose $k \in I_X$, I_X is the support of k, so that the saddlepoint equation can be solved at value k. The first approximation is

$$\hat{\Pr}_1(X \geq k) = \begin{cases} 1 - \Phi(\hat{\omega}) - \phi(\hat{\omega})(1/\hat{\omega} - 1/\tilde{z}_1) & \text{if} \quad k \neq \mu \\ \frac{1}{2} - \frac{1}{\sqrt{2\pi}}\left\{\frac{K'''(0)}{6K''(0)^{\frac{3}{2}}} - \frac{1}{2\sqrt{k''(0)}}\right\} & \text{if} \quad k = \mu \end{cases} \tag{3.4}$$

where

$$\begin{aligned} \hat{\omega} &= \text{sgn}(\hat{s})\sqrt{2\{\hat{s}k - K(\hat{s})\}}, \\ \tilde{z}_1 &= \{1 - \exp(-\hat{s})\}\sqrt{K''(\hat{s})}, \end{aligned} \tag{3.5}$$

and $\hat{s}$ solves $K'(\hat{s}) = k$. The expression for $\hat{\omega}$ agrees with its counterpart (3.3) in the continuous setting whereas $\tilde{z}_1 \neq \hat{z}$. To understand the difference associated with using $\tilde{z}_1$ instead of $\hat{z}$, consider a Taylor series expansion of the leading factor in $\tilde{z}_1$ as $\{1 - \exp(-\hat{s})\} \approx \hat{s} - \hat{s}^2/2$. To first order $\tilde{z}_1 \approx \hat{z}$ from the continuous setting and the approximations are the same. However, to the second order, $\tilde{z}_1 < \hat{z}$ which implies that $\hat{\Pr}_1(X \geq k) > 1 - \hat{F}(k)$ where the latter term refers to the continuous version in (3.2) and (3.3) evaluated at k. Thus, the use of the smaller $\tilde{z}_1$ in place of $\hat{z}$ adjusts the tail probability in a direction that is consistent with a continuity correction.

Define $k^- = k - 1/2 \in I_x$ as the continuity-corrected or offset value of k. The second approximation solves the offset saddlepoint equation $K'(\tilde{s}) = k^-$ for continuity-corrected saddlepoint $\tilde{s}$. Saddlepoint $\tilde{s}$ and k^- are used to alter the inputs into the CDF approximation according to

$$\begin{aligned} \tilde{\omega}_2 &= \text{sgn}(\tilde{s})\sqrt{2(\tilde{s}k^- - K(\tilde{s}))}, \\ \tilde{z}_2 &= 2\text{sinh}(\tilde{s}/2)\sqrt{K''(\tilde{s})}. \end{aligned} \tag{3.6}$$

This leads to the continuity-corrected approximation

$$\hat{\Pr}_2(X \geq k) = \begin{cases} 1 - \Phi(\tilde{\omega}_2) - \phi(\tilde{\omega}_2)(1/\tilde{\omega}_2 - 1/\tilde{z}_2) & \text{if} \quad k^- \neq \mu \\ \frac{1}{2} - \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{\frac{3}{2}}} & \text{if} \quad k^- = \mu \end{cases} \tag{3.7}$$

The third approximation is denoted as $\hat{\Pr}_3(X \geq k)$ and uses expression (3.7) with $\tilde{\omega}_2$ as in (3.6) and $\tilde{z}_2$ replaced with

$$\tilde{z}_3 = \tilde{s}\sqrt{K''(\tilde{s})} \tag{3.8}$$

This approximation may be motivated as simply the application at k^- of the continuous version of the Lugannani and Rice approximation. Intuitively, this correction might be suggested because it is the same one used for normal approximation of the binomial. The second and third approximations are related according to the first order of the Taylor approximation $2\text{sinh}(\tilde{s}/2) \approx \tilde{s} + \tilde{s}^3/24$. Since the expansion lacks an $\tilde{s}^2$ term, $\tilde{z}_2 \approx \tilde{z}_3$ to second order so the two approximations agree to second order. While these two approximations are similar in accuracy near the mean where $\tilde{s} \approx 0$, they begin to deviate in accuracy as $|\tilde{s}|$ increases and the Taylor approximation deteriorates.

For the left tail CDF approximations for discrete distributions, the first continuity correction method will give

$$\hat{F}_{1a}(k) = \begin{cases} \Phi(\hat{\omega}_1) + \phi(\hat{\omega}_1)(1/\hat{\omega}_1 - 1/\tilde{z}_{1a}) & \text{if} \quad k < \mu \\ \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \left\{ \frac{K'''(0)}{6K''(0)^{\frac{3}{2}}} + \frac{1}{2\sqrt{k''(0)}} \right\} & \text{if} \quad k = \mu \end{cases} \tag{3.9}$$

where

$$\begin{aligned} \hat{\omega} &= -\sqrt{2(\hat{s}d - K(\hat{s}))}, \\ \tilde{z}_{1a} &= \{\exp(\hat{s}) - 1\}\sqrt{K''(\hat{s})}, \end{aligned} \tag{3.10}$$

and $\hat{s}$ solves $K'(\hat{s}) = k$. The left-tail expression $\hat{F}_{1a}(k)$ allows approximation for $F(k)$ values over $k \in (-\infty, \mu]$. An expression of the leading factor in $\tilde{z}_{1a}$ is $\{\exp(-\hat{s}) - 1\} \approx \hat{s} + \hat{s}^2/2$ and to the first order $\tilde{z}_{1a} \approx \hat{z}$ from the continuous setting. To the second order, $\tilde{z}_{1a} > \hat{z}$ which implies that $\hat{F}_{1a}(k) > \hat{F}(k)$ where the latter term refers to the continuous formula in (3.2) and (3.3) evaluated at k. Thus, the replacement of $\hat{s}$ with $\exp(-\hat{s}) - 1$ inflates the probability in a manner consistent with a continuity adjustment for the left tail.

Suppose $k^+ = k + 1/2 \in I_X$ is the left offset value of k. The second approximation is

$$\hat{F}_2(k) = \begin{cases} 1 - \Phi(\tilde{\omega}_2^+) - \phi(\tilde{\omega}_2^+)(1/\tilde{\omega}_2^+ - 1/\tilde{z}_2^+) & \text{if} \quad k^+ < \mu \\ \frac{1}{2} + \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{\frac{3}{2}}} & \text{if} \quad k^+ = \mu \end{cases} \tag{3.11}$$

where

$$\begin{aligned} \tilde{\omega}_2^+ &= -\sqrt{2(\tilde{s}^+ d^+ - K(\tilde{s}^+))} \\ \tilde{z}_2^+ &= 2\sinh(\tilde{s}^+/2)\sqrt{K''(\tilde{s}^+)} \end{aligned} \tag{3.12}$$

and $\tilde{s}^+$ is the left continuity-corrected saddlepoint defined as the unique solution to

$$K'(\tilde{s}^+) = k^+$$

The third approximation $\hat{F}_3(k)$ uses the same expression as $\hat{F}_2(k)$ but replaces $\tilde{z}_2^+$ with

$$\tilde{z}_3{}^+ = \tilde{s}^+\sqrt{K''(\tilde{s}^+)} \tag{3.13}$$

3.3 Continuity-corrected approximations on a Δ-lattice

In our study, the statistic is $D = A/m - B/n$ which has support on the Δ-lattice $L_\Delta = \{\gamma, \gamma \pm \Delta, \gamma \pm 2\Delta, ...\}$. Let consider D/Δ where $\Delta = 2CC$ with CGF $K_{\frac{D}{\Delta}}(\tau) = K_D(\frac{\tau}{\Delta}) = K_D(s)$, $s = \tau/\Delta$. where

$$\begin{aligned} K_D(s) &= K_{\frac{A}{m}-\frac{B}{n}}(s) = K_A(\tfrac{s}{m}) - K_B(\tfrac{s}{n}) \\ &= m\ln[1-(\psi+\tfrac{\theta}{2})+(\psi+\tfrac{\theta}{2})e^{\frac{s}{m}}] + n\ln[1-(\psi-\tfrac{\theta}{2})+(\psi-\tfrac{\theta}{2})e^{-\frac{s}{n}}] \end{aligned} \tag{3.14}$$

And

$$\begin{aligned} K'_{\frac{D}{\Delta}}(\tau) &= \tfrac{d}{\Delta} \Leftrightarrow K'_D(s) = d \\ K''_{\frac{D}{\Delta}}(\tau) &= \tfrac{1}{\Delta^2}K''_D(s) \end{aligned}$$

Then the first continuity corrected approximation $\hat{\Pr}_1[D \geq d|\theta, \tilde{\psi}]$ at $d \neq E(D)$ provided by the Lugannani and Rice expression has its inputs

$$\begin{aligned} \hat{\omega} &= \operatorname{sgn}(\hat{s})\sqrt{2\{\hat{s}d - K(\hat{s})\}} \\ \tilde{z}_1 &= \Delta^{-1}(1-\exp(-\Delta\hat{s}))\sqrt{K''(\hat{s})} \end{aligned} \tag{3.15}$$

and saddlepoint $\hat{s}$ solves $K'(\hat{s}) = d$.

The second continuity corrected approximation $\hat{\Pr}_2[D \geq d|\theta, \tilde{\psi}]$ provided by the Lugannani and Rice expression has inputs

$$\begin{aligned} \tilde{\omega}_2 &= \operatorname{sgn}(\tilde{s})\sqrt{2(\tilde{s}d^- - K(\tilde{s}))} \\ \tilde{z}_2 &= 2\Delta^{-1}\sinh(\tfrac{1}{2}\Delta\tilde{s})\sqrt{K''(\tilde{s})} \end{aligned} \tag{3.16}$$

and saddlepoint $\tilde{s}$ solves $K'(\tilde{s}) = d^- = d - \Delta/2$.

Both continuity corrected results approach the continuous uncorrected results as $\Delta \to 0$, since $2\Delta^{-1}\sinh(\frac{1}{2}\Delta\tilde{s}) \to \hat{s}$ as $\Delta \to 0$, with $d^- \to d$ and $\tilde{s} \to \hat{s}$.

Here $\tilde{\psi} = \psi_\theta$, the profile estimate of ψ given θ, that is the maximum likelihood estimate (MLE) of ψ conditional on the hypothesized value of θ. The likelihood function $\Lambda(\psi) = \Pr[A = a, B = b|\theta, \psi]$

$$\ln\Lambda = a\ln(\psi+\frac{\theta}{2}) + b\ln(\psi-\frac{\theta}{2}) + c\ln(1-\psi-\frac{\theta}{2}) + d\ln(1-\psi+\frac{\theta}{2}) \tag{3.17}$$

with understanding that terms corresponding to empty cells are omitted. The constraints $0 \leq \pi_i \leq 1$, i =1,2 translate into restricting evaluation of $\ln\Lambda$ only within the bounding interval $|\theta|/2 \le \psi \le 1 - |\theta|/2$.

$$\frac{\partial^2 \ln \Lambda(\psi)}{\partial \psi^2} = -\left[\frac{a}{(\psi+\frac{\theta}{2})^2} + \frac{b}{(\psi-\frac{\theta}{2})^2} + \frac{c}{(1-\psi-\frac{\theta}{2})^2} + \frac{d}{(1-\psi+\frac{\theta}{2})^2}\right] < 0$$

So with θ fixed, $\partial^2 \ln \Lambda(\psi)/\partial\psi^2$ is negative throughout the bounding interval; the log-likelihood curve is smoothly convex. By the Newton or Secant method, we get the $\tilde{\psi}$, the MLE of ψ given θ.

3.4 Simulation Study

Mimicking simulations by Newcombe (1995), we compared the coverage probabilities of the six continuity corrections. The main evaluation of coverage probability was based on a sample of 9200 parameter space points (m, n, ψ, θ), with m and n between 5 and 50 inclusive. We chose a subset of 230 out of 2116 possible (m, n) pairs (Figure 3.1), including 46 diagonal pairs with $m = n$ and 92 pairs (m, n) with $m \neq n$ and the corresponding reversed pairs (n, m). For each of $m = 5, 6, ..., 50$, two values of n were

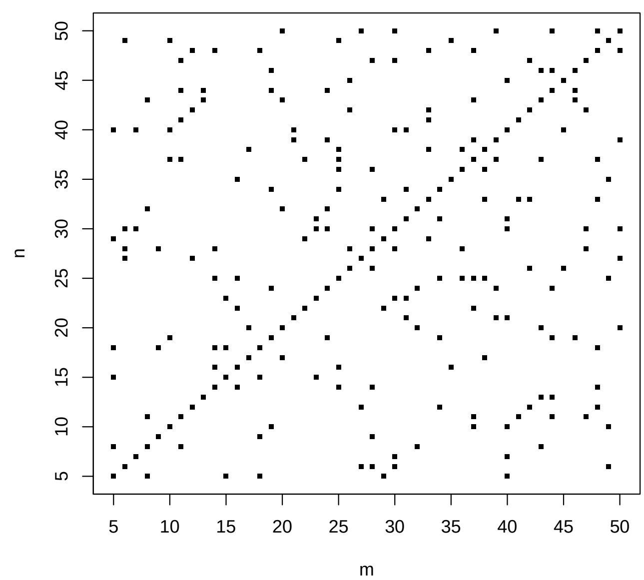

Figure 3.1: 230 (m, n) pairs chosen from 2116 possible pairs

sampled, avoiding diagonal elements, duplicates and mirror-image pairs. These were

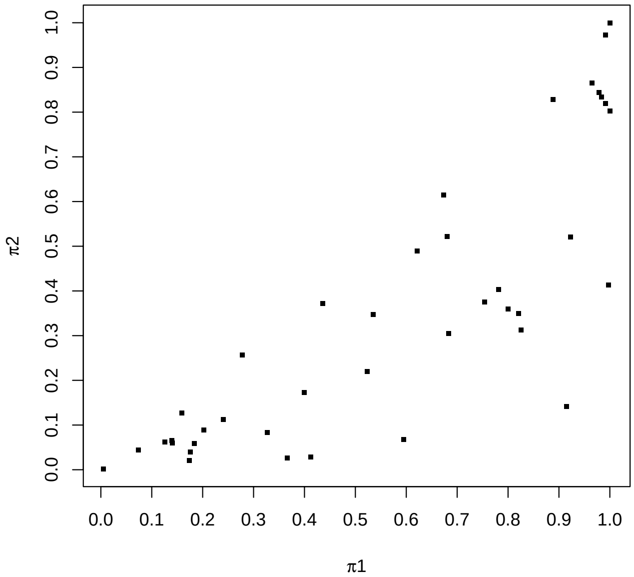

Figure 3.2: 40 (π_1, π_2) pairs 1-1 corresponding to (θ, ψ) pairs with $\pi_1 \geq \pi_2$

selected so that m and n should be uncorrelated and the distributions of $|m - n|$ and the highest common factor (HCF) of m and n should be very close to those for all 2070 off-diagonal points. For each of 230 (m, n) pairs, 40 pairs of (θ, ψ) were sampled with $\theta = \lambda\{1 - |2\psi - 1|\}$ and ψ and $\lambda \sim U(0, 1)$. The sampled values of θ had median 0.201, and quartiles 0.102 and 0.388. The corresponding (π_1, π_2) pairs are showed in Figure 3.2. For each (m, n, θ, ψ) of the parameter space, the coverage probability is obtained,

$$\text{CP} = \Sigma_{\{a,b:\ 2\min(\hat{\Pr}[D \leq d|\theta_0, \tilde{\psi}],\ \hat{\Pr}[D \geq d|\theta_0, \tilde{\psi}]) \geq \alpha\}} \Pr[A = a, B = b|\theta, \psi]$$

where α denotes the nominal significance level. And $2\min(\hat{\Pr}[D \leq d|\theta_0, \tilde{\psi}],\ \hat{\Pr}[D \geq d|\theta_0, \tilde{\psi}])$ is the p-value of the test. Our primary criterion for evaluating the different continuity corrections for saddlepoint approximation is agreement of actual with nominal coverage probabilities.

3.5 Results

The three approximation methods discussed in section 2 gave very similar results which are showed in Tables 3.1-3.15 . So the following discussion is just based on one of them, the second approximation method. From the results on the overall summary measures on the 9200 points in the parameter space with $5 \leq m \leq 50$, $5 \leq n \leq 50$, $0 < \psi < 1$, and $0 < \theta < 1 - |2\psi - 1|$ in Table 3.1, we conclude that for the nominal 90% the continuity correction CC_L has the best average coverage accuracy among the six corrections considered here, and that for the nominal 95%, the continuity correction CC_S has the best average coverage accuracy among the six corrections, and for the nominal level 99% the continuity correction CC_Y has the best average coverage accuracy among the six corrections, but this correction is more conservative than necessary for 90% and 95% intervals. 99% intervals are a different issue. It's clear that the greater the confidence required, the more conservative must be the choice of the confidence interval method or the larger the sample size. For all the nominal levels, the null correction CC_N has poor coverage accuracy.

Table 3.2 and 3.3 illustrate the relation of mean coverage probabilities of nominal 90% and 95% to parameters: the lower of the two denominators m and n (NMIN); the minimum expected cell size (MCS); the nuisance parameter ψ and the parameter of interest θ. For both nominal levels, CC_S is the best choice for small MCS 0-1, or distal $\psi \geq .1$ or $\leq .9$, or small θ 0-.05, while CC_R is the best choice for large MCS 5-25 or mesial ψ .4-.6 or large $\theta \geq .5$; and CC_L is the best choice for large NMIN 30-50. For small NMIN 5-9 CC_S and CC_A are the best for nominal 90% and 95% respectively.

Table 3.4 and 3.5 show that the estimated coverage probabilities of nominal 90% and 95% for specified values of NMIN for MCS ≥ 1. We have suppressed the relation to MCS, since in practice MCS is an unknown quantity. For 95% intervals, we conclude that the new correction CC_R has the best average coverage accuracy among the six correcions considered here and that it has the average coverage probability that is very close to the nominal level for NMIN as small as 5-9. For 90% intervals, CC_R also perform very well though CC_R ranks just behind CC_L for NMIN in $10 - 29$.

Box plots of the coverage probabilities provide the variation and spread of the coverage probabilities of the intervals from the six considered continuity corrections. Figure 3.3 and 3.4 from the second approximation method display the box plots of the coverage probabilities for 90% and 95% confidence levels respectively and for five NMIN ranges $(5-9, 10-19, 20-29, 30-39, 40-50)$. The corrections CC_Y, CC_A and CC_S can be seen to deviate the most from nominal levels. We see that CC_R have smaller spread in coverage probabilities than CC_L, and CC_R's estimated coverage probability range is generally closer to the nominal levels than any other corrections'.

3.6 Conclusion

For 90% and 95% confidence intervals for difference of two binomial proportions, the saddlepoint approximation with the new continuity correction CC_R is recommended when the minimum expected cell size (MCS) is not smaller than 1 or the nuisance parameter ψ is mesial (.4-.6) or the parameter in interest θ is large ($\theta \geq .5$); while the confidence intervals with continuity corrections CC_S, CC_A and CC_Y are unnecessarily conservative and could not acceptable. When MCS $\leq$ 1, CC_S is the best choice of continuity correction, and it also the best choice for distal $\psi \geq .1 \cup \leq .9$ or small θ $0-.05$. And the most useful result is that for MCS > 1, CC_R has coverage probabilities very close to the nominal 95% and 90% for any range of NMIN, even for NMIN as small as $5-9$. For 99% intervals, the Yates correction CC_Y is reasonable.

Table 3.1: Estimates (based on all 9200 points in parameter space) of the Coverage Probability for Nominal Significance Level $\alpha = .1$, .05 and .01 by Using the Second Approximation Method

	α		
CC	.1	.05	.01
CC_N	.8455 (5)	.8934 (6)	.9282 (6)
CC_R	.8840 (2)	.9302 (4)	.9639 (5)
CC_L	.8914 (1)	.9340 (2)	.9647 (4)
CC_S	.9209 (3)	.9558 (1)	.9799 (3)
CC_A	.9388 (4)	.9673 (3)	.9866 (2)
CC_Y	.9564 (6)	.9778 (5)	.9918 (1)

CC: continuity correction

Table 3.2: Estimates (Based on Points in Parameter Subspaces) of the Coverage Probability for Nominal Significance Level $\alpha = .1$ by Using the Second Approximation Method

	NMIN		MCS		ψ		θ	
	$5-9$	$30-50$	$0-1$	$5-25$	$\leq .1$ or $\geq .9$	$.4-.6$	$0-.05$	$.5-1$
CC	1480	2480	3208	1960	2530	2300	1380	920
CC_N	.8233 (6)	.8542 (5)	.7540 (6)	.8971 (2)	.7125 (6)	.8958 (2)	.5723 (6)	.8984 (2)
CC_R	.8767 (3)	.8880 (2)	.8601 (3)	.8975 (1)	.8457 (4)	.8961 (1)	.8102 (5)	.8998 (1)
CC_L	.8868 (2)	.8956 (1)	.8662 (2)	.9046 (3)	.8500 (3)	.9055 (3)	.8137 (4)	.9097 (3)
CC_S	.9059 (1)	.9238 (3)	.9191 (1)	.9190 (4)	.9153 (1)	.9196 (4)	.9166 (1)	.9262 (4)
CC_A	.9427 (4)	.9266 (4)	.9461 (4)	.9257 (5)	.9440 (2)	.9333 (5)	.9581 (2)	.9387 (5)
CC_Y	.9592 (5)	.9500 (6)	.9667 (5)	.9410 (6)	.9662 (5)	.9477 (6)	.9815 (3)	.9536 (6)

CC: continuity correction

Table 3.3: Estimates (Based on Points in Parameter Subspaces) of the Coverage Probability for Nominal Significance Level $\alpha = .05$ by Using the Second Approximation Method

	NMIN		MCS		ψ		θ	
	$5-9$	$30-50$	$0-1$	$5-25$	$\leq .1$ or $\geq .9$	$.4-.6$	$0-.05$	$.5-1$
CC	1480	2480	3208	1960	2530	2300	1380	920
CC_N	.8641 (6)	.9038 (6)	.7909 (6)	.9483 (2)	.7495 (6)	.9476 (2)	.6020 (6)	.9488 (2)
CC_R	.9156 (5)	.9359 (4)	.8945 (5)	.9485 (1)	.8797 (5)	.9478 (1)	.8378 (5)	.9493 (1)
CC_L	.9201 (4)	.9401 (1)	.8971 (4)	.9524 (3)	.8817 (4)	.9528 (3)	.8397 (4)	.9542 (3)
CC_S	.9345 (2)	.9609 (2)	.9427 (1)	.9605 (4)	.9382 (1)	.9607 (4)	.9359 (1)	.9635 (4)
CC_A	.9589 (1)	.9625 (3)	.9636 (2)	.9641 (5)	.9618 (2)	.9675 (5)	.9734 (2)	.9691 (5)
CC_Y	.9711 (3)	.9773 (5)	.9784 (3)	.9722 (6)	.9779 (3)	.9750 (6)	.9922 (3)	.9766 (6)

CC: continuity correction

Table 3.4: Estimates (Based on Points in Parameter Subspaces) with MCS ≥ 1 of the Coverage Probability for Nominal Significance Level $\alpha = .1$ by Using the Second Approximation Method

	NMIN				
	$5-9$	$10-19$	$20-29$	$30-39$	$40-50$
CC	641	1693	1761	1205	692
CC_N	.8915 (3)	.8931 (3)	.8950 (3)	.8960 (2)	.8967 (2)
CC_R	.8937 (1)	.8953 (2)	.8967 (2)	.8984 (1)	.9000 (1)
CC_L	.9077 (2)	.9035 (1)	.9028 (1)	.9053 (3)	.9109 (3)
CC_S	.9189 (4)	.9207 (4)	.9228 (4)	.9228 (4)	.9232 (4)
CC_A	.9522 (5)	.9408 (5)	.9323 (5)	.9272 (5)	.9240 (5)
CC_Y	.9638 (6)	.9549 (6)	.9495 (6)	.9456 (6)	.9420 (6)

CC: continuity correction

Table 3.5: Estimates (Based on Points in Parameter Subspaces) with MCS ≥ 1 of the Coverage Probability for Nominal Significance Level $\alpha = .05$ by Using the Second Approximation Method

	NMIN				
	$5-9$	$10-19$	$20-29$	$30-39$	$40-50$
CC	641	1693	1761	1205	692
CC_N	.9470 (2)	.9477 (2)	.9481 (2)	.9491 (2)	.9491 (2)
CC_R	.9483 (1)	.9489 (1)	.9491 (1)	.9503 (1)	.9508 (1)
CC_L	.9549 (3)	.9532 (3)	.9525 (3)	.9541 (3)	.9567 (3)
CC_S	.9608 (4)	.9622 (4)	.9636 (4)	.9635 (4)	.9631 (4)
CC_A	.9769 (5)	.9720 (5)	.9682 (5)	.9659 (5)	.9636 (5)
CC_Y	.9827 (6)	.9790 (6)	.9772 (6)	.9751 (6)	.9734 (6)

CC: continuity correction

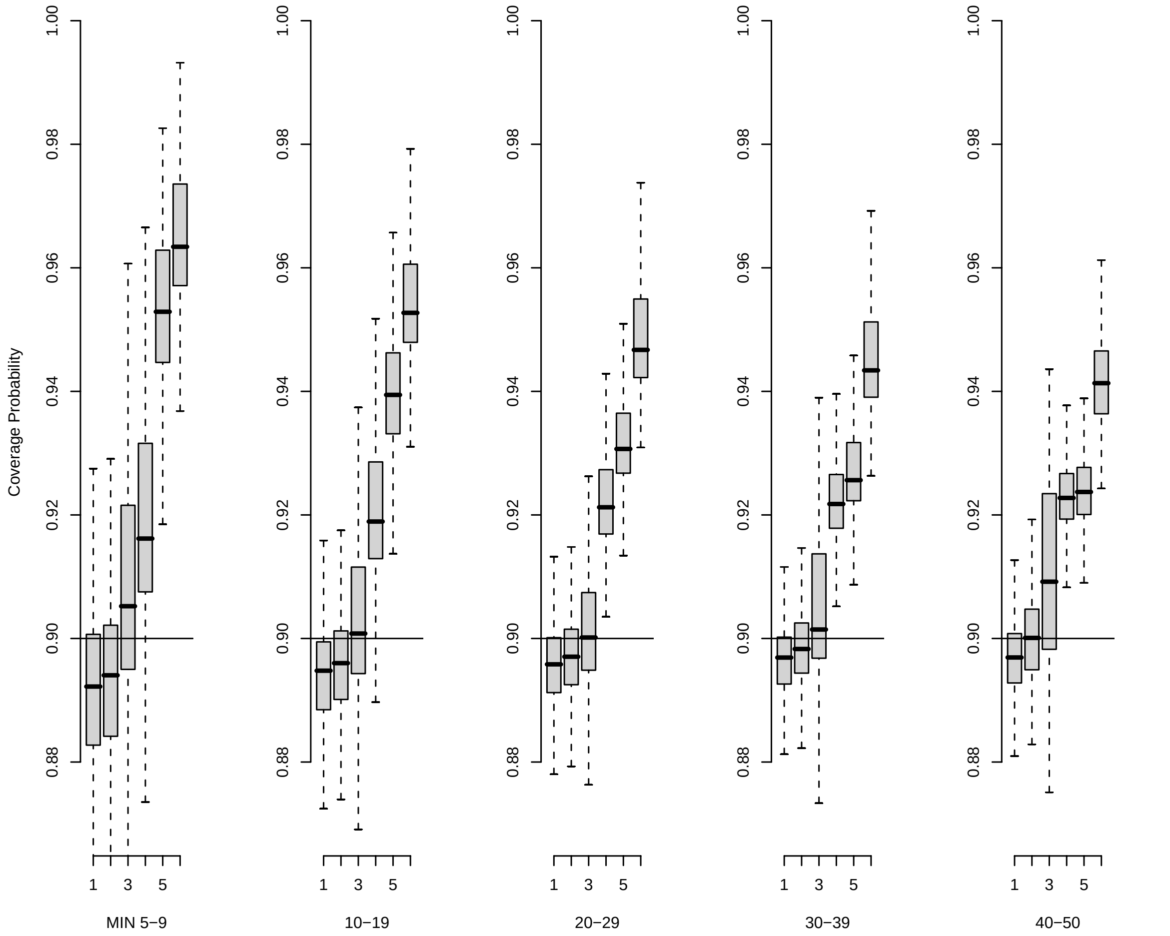

Figure 3.3: Estimates Range of the Coverage Probability of 90% Confidence Intervals by NMIN Groups for MCS $\geq$ 1 by Using the Second Approximation Method; From left to right in each plot are CC_N, CC_R, CC_L, CC_S, CC_A and CC_Y

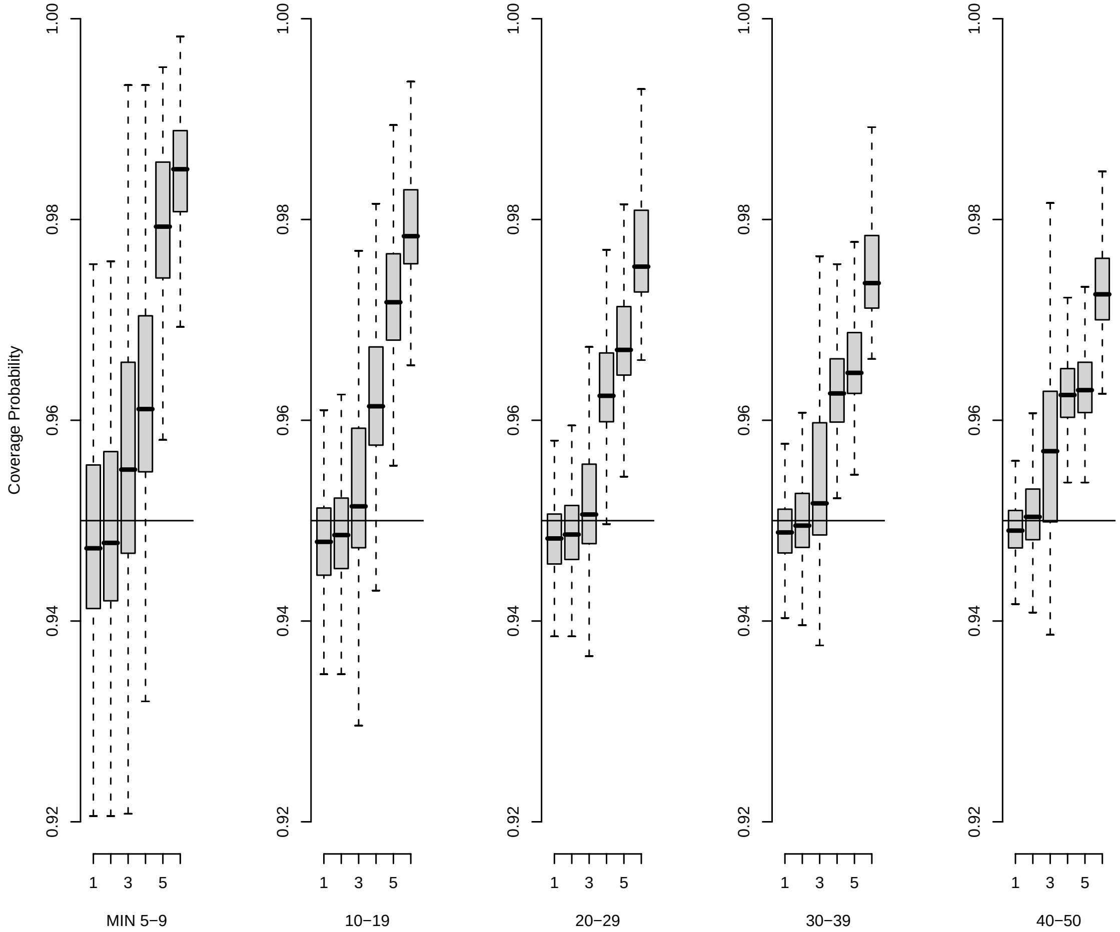

Figure 3.4: Estimates Range of the Coverage Probability of 95% Confidence Intervals by NMIN Groups for MCS $\geq$ 1 by Using the Second Approximation Method; From left to right in each plot are CC_N, CC_R, CC_L, CC_S, CC_A and CC_Y

Table 3.6: Estimates (based on all 9200 points in parameter space) of the Coverage Probability for Nominal Significance Level $\alpha = .1$, .05 and .01 by Using the First Approximation Method

	α		
CC	.1	.05	.01
CC_N	.8455 (5)	.8934 (6)	.9282 (6)
CC_R	.8840 (2)	.9303 (4)	.9639 (5)
CC_L	.8916 (1)	.9341 (2)	.9647 (4)
CC_S	.9211 (3)	.9559 (1)	.9799 (3)
CC_A	.9392 (4)	.9675 (3)	.9867 (2)
CC_Y	.9565 (6)	.9778 (5)	.9918 (1)

CC: continuity correction

Table 3.7: Estimates (Based on Points in Parameter Subspaces) of the Coverage Probability for Nominal Significance Level $\alpha = .1$ by Using the First Approximation Method

	NMIN		MCS		ψ		θ	
	$5-9$	$30-50$	$0-1$	$5-25$	$\leq .1$ or $\geq .9$	$.4-.6$	$0-.05$	$.5-1$
CC	1480	2480	3208	1960	2530	2300	1380	920
CC_N	.8233 (6)	.8542 (5)	.7540 (6)	.8971 (2)	.7125 (6)	.8958 (2)	.5723 (6)	.8984 (2)
CC_R	.8767 (3)	.8880 (2)	.8602 (3)	.8975 (1)	.8457 (4)	.8961 (1)	.8102 (5)	.8998 (1)
CC_L	.8873 (2)	.8957 (1)	.8664 (2)	.9046 (3)	.8501 (3)	.9058 (3)	.8138 (4)	.9101 (3)
CC_S	.9064 (1)	.9238 (3)	.9193 (1)	.9191 (4)	.9153 (1)	.9200 (4)	.9166 (1)	.9270 (4)
CC_A	.9440 (4)	.9267 (4)	.9464 (4)	.9259 (5)	.9441 (2)	.9341 (5)	.9580 (2)	.9399 (5)
CC_Y	.9598 (5)	.9498 (6)	.9665 (5)	.9412 (6)	.9659 (5)	.9483 (6)	.9814 (3)	.9543 (6)

CC: continuity correction

Table 3.8: Estimates (Based on Points in Parameter Subspaces) of the Coverage Probability for Nominal Significance Level $\alpha = .05$ by Using the First Approximation Method

	NMIN		MCS		ψ		θ	
	$5-9$	$30-50$	$0-1$	$5-25$	$\leq .1$ or $\geq .9$	$.4-.6$	$0-.05$	$.5-1$
CC	1480	2480	3208	1960	2530	2300	1380	920
CC_N	.8641 (6)	.9038 (6)	.7909 (6)	.9483 (2)	.7495 (6)	.9476 (2)	.6020 (6)	.9488 (2)
CC_R	.9156 (5)	.9359 (4)	.8945 (5)	.9485 (1)	.8797 (5)	.9478 (1)	.8378 (5)	.9493 (1)
CC_L	.9203 (4)	.9401 (1)	.8972 (4)	.9524 (3)	.8818 (4)	.9529 (3)	.8397 (4)	.9544 (3)
CC_S	.9348 (2)	.9610 (2)	.9429 (1)	.9606 (4)	.9383 (1)	.9609 (4)	.9359 (1)	.9638 (4)
CC_A	.9595 (1)	.9626 (3)	.9638 (2)	.9642 (5)	.9618 (2)	.9679 (5)	.9734 (2)	.9697 (5)
CC_Y	.9714 (3)	.9773 (5)	.9784 (3)	.9723 (6)	.9778 (3)	.9753 (6)	.9921 (3)	.9768 (6)

CC: continuity correction

Table 3.9: Estimates (Based on Points in Parameter Subspaces) with MCS ≥ 1 of the Coverage Probability for Nominal Significance Level $\alpha = .1$ by Using the First Approximation Method

	NMIN				
	$5-9$	$10-19$	$20-29$	$30-39$	$40-50$
CC	641	1693	1761	1205	692
CC_N	.8915 (3)	.8931 (3)	.8950 (3)	.8960 (2)	.8967 (2)
CC_R	.8938 (1)	.8954 (2)	.8967 (2)	.8983 (1)	.9000 (1)
CC_L	.9084 (2)	.9035 (1)	.9028 (1)	.9053 (3)	.9110 (3)
CC_S	.9199 (4)	.9210 (4)	.9228 (4)	.9228 (4)	.9232 (4)
CC_A	.9547 (5)	.9413 (5)	.9324 (5)	.9272 (5)	.9241 (5)
CC_Y	.9654 (6)	.9553 (6)	.9494 (6)	.9455 (6)	.9418 (6)

CC: continuity correction

Table 3.10: Estimates (Based on Points in Parameter Subspaces) with MCS ≥ 1 of the Coverage Probability for Nominal Significance Level $\alpha = .05$ by Using the First Approximation Method

	NMIN				
	$5-9$	$10-19$	$20-29$	$30-39$	$40-50$
CC	641	1693	1761	1205	692
CC_N	.9470 (2)	.9477 (2)	.9481 (2)	.9491 (2)	.9491 (2)
CC_R	.9483 (1)	.9489 (1)	.9491 (1)	.9503 (1)	.9508 (1)
CC_L	.9551 (3)	.9533 (3)	.9525 (3)	.9541 (3)	.9567 (3)
CC_S	.9612 (4)	.9623 (4)	.9637 (4)	.9635 (4)	.9631 (4)
CC_A	.9780 (5)	.9722 (5)	.9683 (5)	.9659 (5)	.9636 (5)
CC_Y	.9834 (6)	.9791 (6)	.9772 (6)	.9750 (6)	.9734 (6)

CC: continuity correction

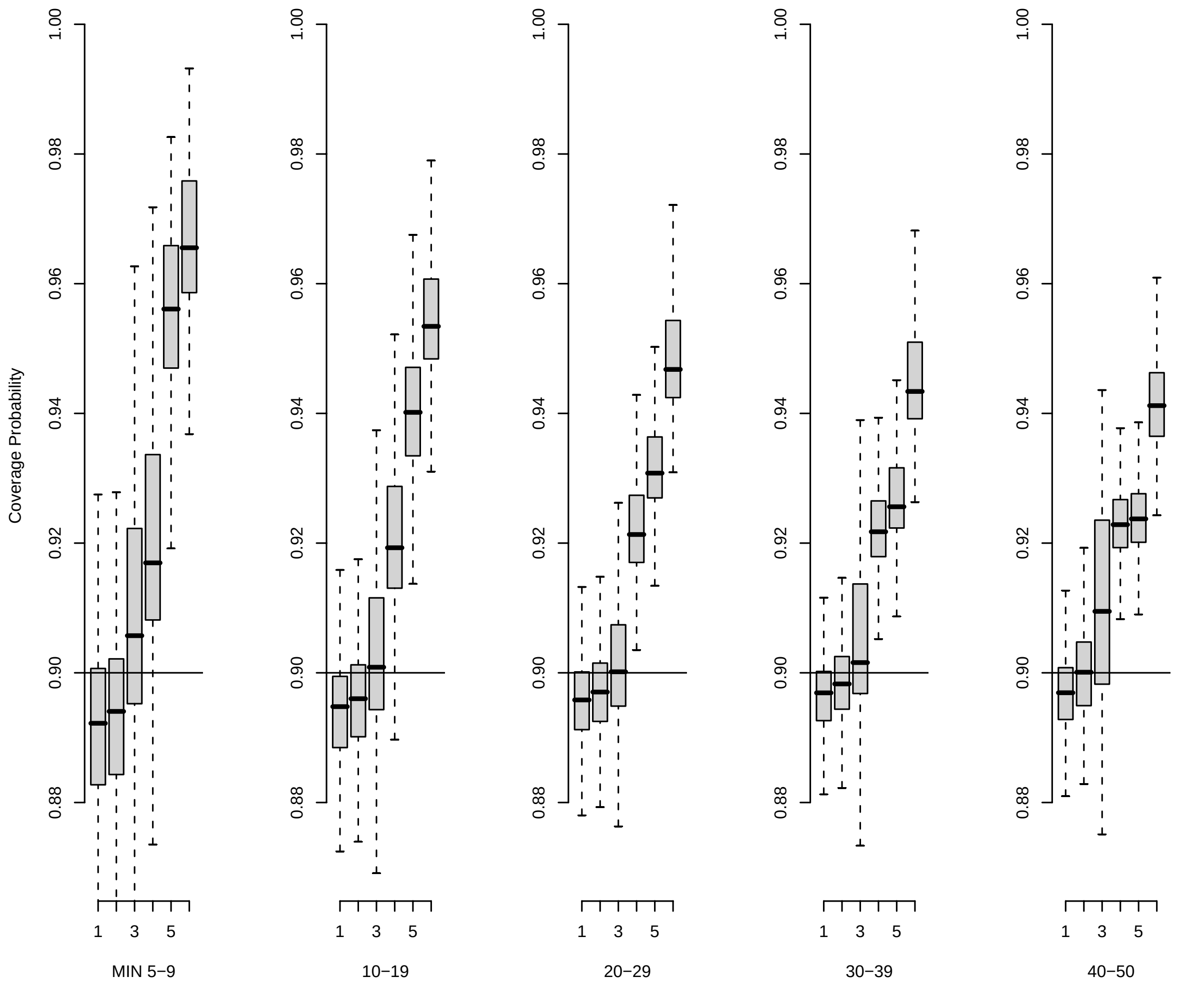

Figure 3.5: Estimates Range of the Coverage Probability of 90% Confidence Intervals by NMIN Groups for MCS ≥ 1 by Using the First Approximation Method; From left to right in each plot are CC_N, CC_R, CC_L, CC_S, CC_A and CC_Y

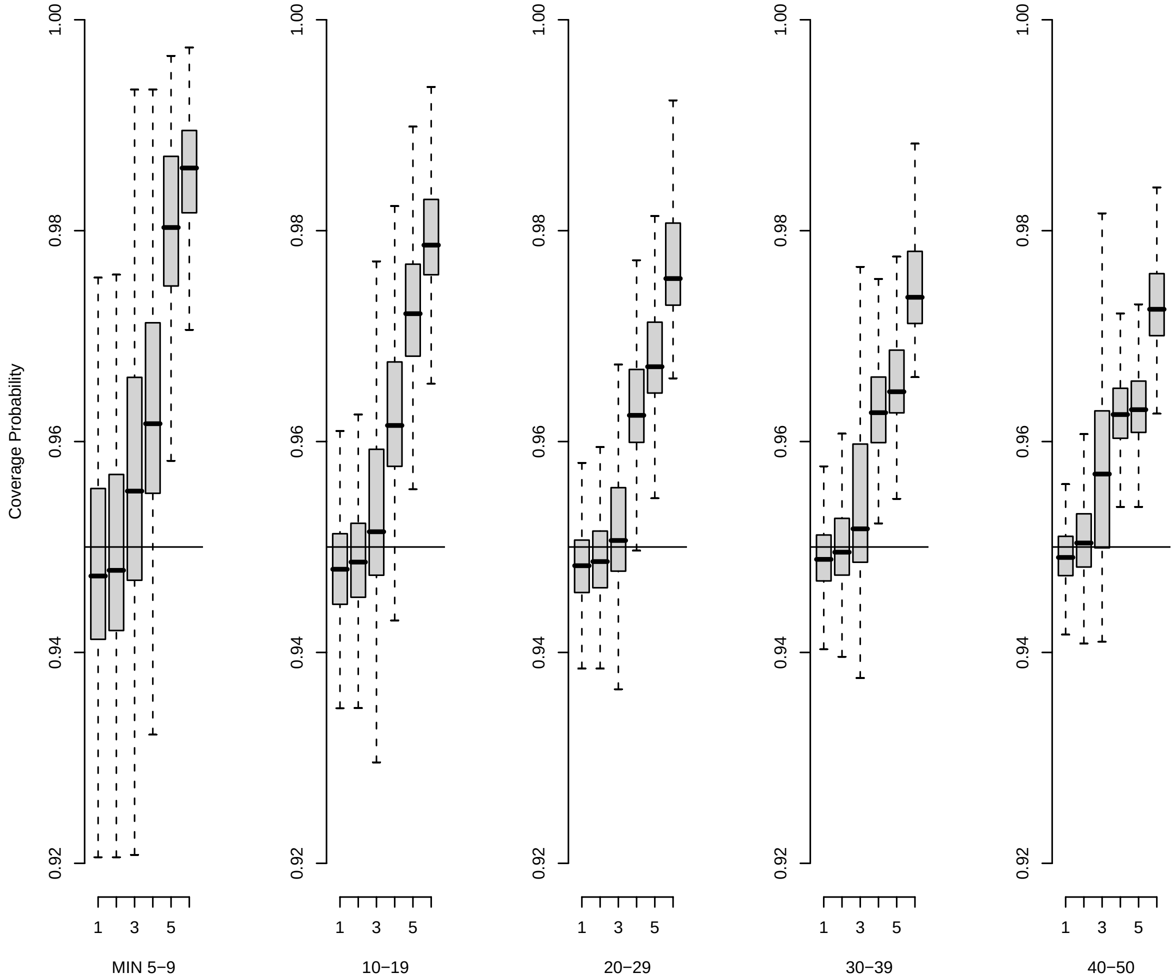

Figure 3.6: Estimates Range of the Coverage Probability of 95% Confidence Intervals by NMIN Groups for MCS ≥ 1 by Using the First Approximation Method; From left to right in each plot are CC_N, CC_R, CC_L, CC_S, CC_A and CC_Y

Table 3.11: Estimates (based on all 9200 points in parameter space) of the Coverage Probability for Nominal Significance Level α = .1, .05 and .01 by Using the Third Approximation Method

		α	
CC	.1	.05	.01
CC_N	.8455 (5)	.8934 (6)	.9282 (6)
CC_R	.8840 (2)	.9303 (4)	.9639 (5)
CC_L	.8919 (1)	.9342 (2)	.9647 (4)
CC_S	.9220 (3)	.9564 (1)	.9800 (3)
CC_A	.9411 (4)	.9685 (3)	.9869 (2)
CC_Y	.9601 (6)	.9797 (5)	.9922 (1)

CC: continuity correction

Table 3.12: Estimates (Based on Points in Parameter Subspaces) of the Coverage Probability for Nominal Significance Level $\alpha = .1$ by Using the Third Approximation Method

	NMIN		MCS		ψ		θ	
	$5-9$	$30-50$	$0-1$	$5-25$	$\leq .1$ or $\geq .9$	$.4-.6$	$0-.05$	$.5-1$
CC	1480	2480	3208	1960	2530	2300	1380	920
CC_N	.8233 (6)	.8542 (5)	.7540 (6)	.8971 (2)	.7125 (6)	.8958 (2)	.5723 (6)	.8984 (2)
CC_R	.8767 (3)	.8880 (2)	.8602 (3)	.8975 (1)	.8458 (4)	.8961 (1)	.8102 (5)	.8998 (1)
CC_L	.8874 (2)	.8961 (1)	.8667 (2)	.9048 (3)	.8504 (3)	.9060 (3)	.8139 (4)	.9103 (3)
CC_S	.9068 (1)	.9248 (3)	.9203 (1)	.9197 (4)	.9164 (1)	.9205 (4)	.9172 (1)	.9275 (4)
CC_A	.9462 (4)	.9280 (4)	.9485 (4)	.9270 (5)	.9460 (2)	.9356 (5)	.9595 (2)	.9413 (5)
CC_Y	.9630 (5)	.9531 (6)	.9696 (5)	.9438 (6)	.9690 (5)	.9515 (6)	.9836 (3)	.9579 (6)

CC: continuity correction

Table 3.13: Estimates (Based on Points in Parameter Subspaces) of the Coverage Probability for Nominal Significance Level $\alpha = .05$ by Using the Third Approximation Method

	NMIN		MCS		ψ		θ	
	$5-9$	$30-50$	$0-1$	$5-25$	$\leq .1$ or $\geq .9$	$.4-.6$	$0-.05$	$.5-1$
CC	1480	2480	3208	1960	2530	2300	1380	920
CC_N	.8641 (6)	.9038 (6)	.7909 (6)	.9483 (2)	.7495 (6)	.9476 (2)	.6020 (6)	.9488 (2)
CC_R	.9156 (5)	.9359 (4)	.8945 (5)	.9485 (1)	.8798 (5)	.9478 (1)	.8379 (5)	.9493 (1)
CC_L	.9203 (4)	.9404 (1)	.8973 (4)	.9525 (3)	.8819 (4)	.9530 (3)	.8398 (4)	.9545 (3)
CC_S	.9350 (2)	.9617 (2)	.9433 (1)	.9610 (4)	.9388 (1)	.9613 (4)	.9362 (1)	.9641 (4)
CC_A	.9606 (1)	.9633 (3)	.9646 (2)	.9649 (5)	.9627 (2)	.9689 (5)	.9739 (2)	.9706 (5)
CC_Y	.9729 (3)	.9792 (5)	.9799 (3)	.9740 (6)	.9793 (3)	.9772 (6)	.9932 (3)	.9789 (6)

CC: continuity correction

Table 3.14: Estimates (Based on Points in Parameter Subspaces) with MCS ≥ 1 of the Coverage Probability for Nominal Significance Level $\alpha = .1$ by Using the Third Approximation Method

	NMIN				
	$5-9$	$10-19$	$20-29$	$30-39$	$40-50$
CC	641	1693	1761	1205	692
CC_N	.8915 (3)	.8931 (3)	.8950 (3)	.8960 (2)	.8967 (2)
CC_R	.8937 (1)	.8954 (2)	.8968 (2)	.8984 (1)	.9001 (1)
CC_L	.9084 (2)	.9039 (1)	.9031 (1)	.9057 (3)	.9115 (3)
CC_S	.9200 (4)	.9219 (4)	.9238 (4)	.9238 (4)	.9242 (4)
CC_A	.9569 (5)	.9436 (5)	.9340 (5)	.9287 (5)	.9252 (5)
CC_Y	.9693 (6)	.9595 (6)	.9535 (6)	.9490 (6)	.9449 (6)

CC: continuity correction

Table 3.15: Estimates (Based on Points in Parameter Subspaces) with MCS ≥ 1 of the Coverage Probability for Nominal Significance Level $\alpha = .05$ by Using the Third Approximation Method

	NMIN				
	$5-9$	$10-19$	$20-29$	$30-39$	$40-50$
CC	641	1693	1761	1205	692
CC_N	.9470 (2)	.9477 (2)	.9481 (2)	.9491 (2)	.9491 (1)
CC_R	.9483 (1)	.9489 (1)	.9491 (1)	.9503 (1)	.9509 (2)
CC_L	.9552 (3)	.9534 (3)	.9527 (3)	.9543 (3)	.9571 (3)
CC_S	.9614 (4)	.9628 (4)	.9642 (4)	.9642 (4)	.9638 (4)
CC_A	.9796 (5)	.9736 (5)	.9694 (5)	.9668 (5)	.9643 (5)
CC_Y	.9855 (6)	.9814 (6)	.9794 (6)	.9771 (6)	.9751 (6)

CC: continuity correction

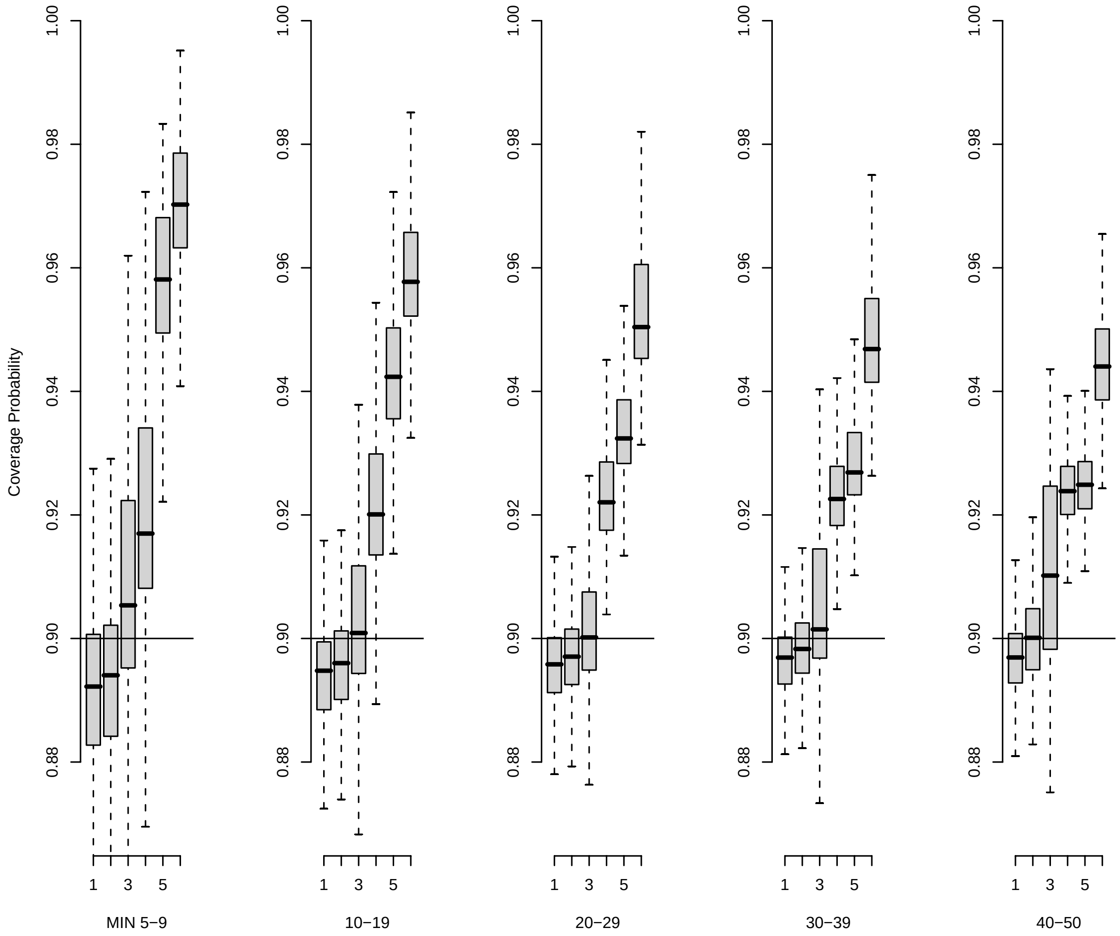

Figure 3.7: Estimates Range of the Coverage Probability of 90% Confidence Intervals by NMIN Groups for MCS ≥ 1 by Using the Third Approximation Method; From left to right in each plot are CC_N, CC_R, CC_L, CC_S, CC_A and CC_Y

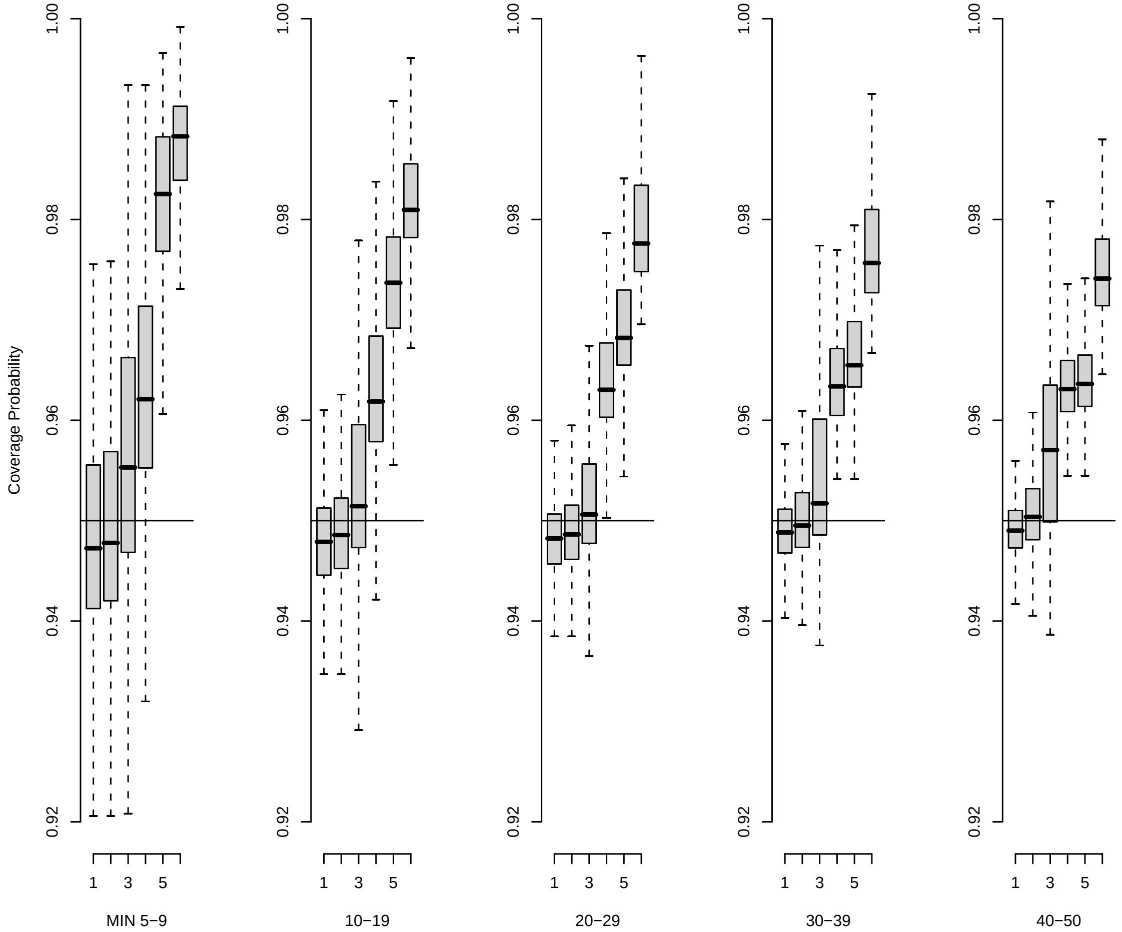

Figure 3.8: Estimates Range of the Coverage Probability of 95% Confidence Intervals by NMIN Groups for MCS ≥ 1 by Using the Third Approximation Method; From left to right in each plot are CC_N, CC_R, CC_L, CC_S, CC_A and CC_Y

Chapter 4

Using Saddlepoint Approximations in Modern Sequential Analysis and Its Applications to Computerized Adaptive Testing

4.1 Introduction

In recent years computer based tests known as the computerized adaptive tests are popularly used in the Graduate Record Examination (GRE), the Graduate Management Admission Test (GMAT) and the Test of English as a Foreign Language (TOEFL). It is an efficient alternative to paper-and-pencil tests. There have been extensive discussions for computerized adaptive testing (CAT) in the psychometric literature with various proposals: Lord (1980) , Owen (1975) , Weiss (1976, 1982), Lewis and Sheehan (1990), van der Linden and Pashley (2000), Chang and Ying (1999), Chang and Ying (2003) and Chang (2004, 2005).

The problem of classifying examinees as either masters or non-masters in a given content area, known as computerized mastery testing (CMT), can be formalized by setting a cut point θ_0 and defining an examinee as a master if and only if his/her ability level θ meets or exceeds that cut point. A computerized mastery testing typically assumes a region (θ_-, θ_+) containing θ_0. The statistical hypothesis of mastery is then given by $H_0 : \theta \geq \theta_+$, while the hypothesis of non-mastery is given by $H_1 : \theta \leq \theta_-$.

Most recently by Bartroff, Finkelman and Lai (2008) had discussed different sequential data analysis approaches for CMT: the sequential probability ratio test(SPRT), the truncated sequential probability ratio test (TSPRT), modified TSPRT and modified Haybittle-Peto test.

In this paper we study the test to have no more than N items. Suppose that k

items have been presented to an examinee, yielding the independent binary responses $u_1, ...u_k$ with the value 1 if the examinee answers the ith item correctly and 0 if the examinee answers the ith item incorrectly. The likelihood of θ is

$$L_k(\theta) = \prod_{i=1}^{k} [p_i(\theta)]^{u_i} [1 - p_i(\theta)]^{1-u_i} \tag{4.1}$$

where $p_i(\theta) = P_\theta\{u_i = 1\}$ for an examinee of ability θ. Let $\widehat{\theta}_k$ denote the maximum likelihood estimator (MLE) of θ based on $u_1, ...u_k$.

We consider sequential testing techniques, including the truncated sequential probability ratio test and the Haybittle-Peto test. Both of these tests in their original forms rely on approximate normality of the signed roots of the log likelihood ratio tests, and approximate boundary crossing probabilities for discrete normal-theory random walks. Bartroff, Finkelman and Lai (2008) modified these techniques by using Monte Carlo approximations to calibrate the truncation boundary. We propose a hybrid Monte Carlo-Asymptotic approach, in which we substitute an easy Monte Carlo approximation in place of boundary crossing probabilities for Brownian motions, and use asymptotic approximations for the distribution of the signed root of the likelihood ratio test statistic. We found that after selecting stopping boundaries using normal-based Monte Carlo calculations, reliance on asymptotic normality of the signed root of the log likelihood ratio statistics provided adequate control of Type I error, without recourse to more complicated Monte Carlo operations. We also observe markable improvement using Barndorff-Nielsen's r^* formula (Barndorff-Nielsen, 1991).

The modified Haybittle-Peto test [21] involves generalized likelihood ratio (GLR) statistics $L_k(\widehat{\theta}_k)/L_k(\theta_0)$. Let $0 < \rho < 1$. For $\rho N \leq k < N$, the modified Haybittle-Peto test stops after the kth item and rejects $H_0 : \theta \geq \theta_0$ if

$$\widehat{\theta}_k < \theta_+ \qquad \text{and} \qquad \log \frac{L_k(\widehat{\theta}_k)}{L_k(\theta_+)} \geq A \tag{4.2}$$

or accepts H_0 if

$$\widehat{\theta}_k > \theta_- \qquad \text{and} \qquad \log \frac{L_k(\widehat{\theta}_k)}{L_k(\theta_-)} \geq B \tag{4.3}$$

for some constants A and B. For $k = N$ the test is always terminated, with H_0 rejected if and only if

$$\widehat{\theta}_k > \theta_+ \qquad \text{and} \qquad \log \frac{L_N(\widehat{\theta}_N)}{L_N(\theta_+)} \geq C \tag{4.4}$$

for some constant C. The thresholds A, B and C are chosen so that the false negative error rate does not exceed α and the false positive error rate, at θ_0 implied by the maximum number N of observations, is close to β. Specifically, A, B and C are chosen so that

$$P_{\theta_-}\{(4.3) \text{ occurs for some } k < N\} = \varepsilon\beta \tag{4.5}$$

$$P_{\theta_+}\{(4.2) \text{ occurs for some } k < N, \text{ (4.3) does not occur for any } j \leq k\} = \varepsilon\alpha \tag{4.6}$$

$$P_{\theta_+}\{(4.2), \text{ (4.3) do not occur for any } k < N, \text{ (4.4) occurs}\} = (1-\varepsilon)\alpha \tag{4.7}$$

for some $0 < \varepsilon < 1$. Lai and Shih(2004) have shown that values $1/3 \leq \varepsilon \leq 1/2$ work well in a variety of settings.

4.2 Methodology

In this part, first we justify the saddlepoint approximation for the curved exponential family. And discuss how we determine the stopping boundaries for the sequential likelihood ratio test.

4.2.1 Justification of Saddlepoint Approximation

The saddlepoint approximation for curved exponential family are discussed in chapter 7 and 8 of Butler (2007). Let u be the data and $x = x(u)$ the $m \times 1$ canonical sufficient statistic for an exponential family of the form

$$f(u;\theta) = \exp\{\xi_\theta^T x - c(\xi_\theta) - d(u)\}. \tag{4.8}$$

where ξ_θ is $m \times 1$ canonical parameter on the lower dimensional vector $\theta \in \Theta \subset R^p$ with $p < m$ and Θ as an open subset of R^p. The parametric class $f(;\theta) : \theta \in \Theta$ is referred to as an $(m,p)-$ curved exponential family of distributions.

Let $\mu_\theta = E(X)$, $\Sigma_\theta = Cov(X;\theta) = c^{''}(\xi_\theta)$ and $\dot{\xi}_\theta = \partial\xi_\theta/\partial\theta^T$. And differentiating $l_\theta = \ln f(u;\theta)$ is required. Then

$$\dot{\mu}_\theta = \frac{\partial\mu_\theta}{\partial\theta^T} = \Sigma_\theta\dot{\xi}_\theta$$

, and

$$\frac{\partial c(\xi_\theta)}{\partial\theta^T} = \mu_\theta^T\dot{\xi}_\theta$$

. The first derivative of l_θ with respect to θ^T is the $1\times p$ vector

$$\dot{l}_\theta^T = \frac{\partial l_\theta}{\partial\theta^T} = (x-\mu_\theta)^T\dot{\xi}_\theta \tag{4.9}$$

and $p\times p$ Hessian is

$$\ddot{l}_\theta = \frac{\partial^2 l_\theta}{\partial\theta\partial\theta^T} = -\dot{\xi}_\theta^T\Sigma_\theta\dot{\xi}_\theta + D_\theta \tag{4.10}$$

with

$$(D_\theta)_{ij} = \left\{(x-\mu_\theta)^T\frac{\partial^2\xi_\theta}{\partial\theta_i\partial\theta_j}\right\}. \tag{4.11}$$

The expected Fisher information i_θ and the observed Fisher information j_θ are

$$i_\theta = -E(\ddot{l}_\theta) = \dot{\xi}_\theta^T\Sigma_\theta\dot{\xi}_\theta = \dot{\xi}_\theta^T\dot{\mu}_\theta \tag{4.12}$$

$$j_\theta = -\ddot{l}_\theta = i_\theta - D_\theta \tag{4.13}$$

When $p=1$, (4.10) is simplified by

$$\ddot{l}_\theta = -\dot{\xi}_\theta^T\Sigma_\theta\dot{\xi}_\theta + \ddot{\xi}_\theta^T(x-\mu_\theta) \tag{4.14}$$

where $\ddot{\xi}_\theta = \partial^2\xi_\theta/\partial\theta^2$ is $m\times 1$. Likewise

$$j_\theta = i_\theta - \ddot{\xi}_\theta^T(x-\mu_\theta). \tag{4.15}$$

Let $\hat{\Theta}$ denote the MLE for θ considered as a random variable. The continuous CDF approximation for MLE $\hat{\Theta}$ given ancillary a and parameter θ is given by $\hat{\Pr}(\hat{\Theta}\le\hat{\theta}|a;\theta)$, where a is an approximate $(m-p)-$dimensional affine ancillary in the (m,p) curved exponential families. Donate the log-likelihood as $l = l(\theta;\hat{\theta},a)$. The $p\times 1$ gradient and $p\times p$ Hessian of l with respect to θ are denoted by

$$l_{\theta;} = l_{\theta;}(\theta;\hat{\theta},a) = \frac{\partial l(\theta;\hat{\theta},a)}{\partial\theta}$$

$$l_{\theta\theta;} = l_{\theta\theta;}(\theta;\hat{\theta},a) = \frac{\partial^2 l(\theta;\hat{\theta},a)}{\partial\theta\partial\theta^T} = -j_\theta$$

Partial derivatives with respect to $\hat{\theta}$ holding a fixed are denoted by

$$l_{;\hat{\theta}} = l_{;\hat{\theta}}(\theta;\hat{\theta},a) = \frac{\partial l(\theta;\hat{\theta},a)}{\partial\hat{\theta}}$$
$$l_{;\hat{\theta}\hat{\theta}} = l_{;\hat{\theta}\hat{\theta}}(\theta;\hat{\theta},a) = \frac{\partial^2 l(\theta;\hat{\theta},a)}{\partial\hat{\theta}\partial\hat{\theta}^T}$$

From Barndorff-Nielsen(1990), the CDF approximation for $\hat{\Theta}|a;\theta$ is

$$\hat{\Pr}(\hat{\Theta} \le \hat{\theta}|a;\theta) = \Phi(\hat{\omega}) + \phi(\hat{\omega})(1/\hat{\omega} - 1/\hat{z}) \quad \text{if} \quad \hat{\theta} \neq \theta \tag{4.16}$$

where

$$\hat{\omega} = \text{sgn}(\hat{\theta} - \theta)\sqrt{-2\{l(\theta;\hat{\theta},a) - l(\hat{\theta};\hat{\theta},a)\}}, \tag{4.17}$$

$$\hat{z} = J_{\hat{\theta}}^{-1/2}\{l_{;\hat{\theta}}(\theta;\hat{\theta},a) - l_{;\hat{\theta}}(\hat{\theta};\hat{\theta},a)\} \tag{4.18}$$

The factor $J^{-1/2} = -l_{\theta\theta;}(\hat{\theta};\hat{\theta},a)$ is the MLE for the Fisher information about θ. If

$$\hat{\omega}^* = \hat{\omega} + \frac{1}{\hat{\omega}} \ln\frac{\hat{\omega}}{\hat{z}} \tag{4.19}$$

then the conditional CDF approximation simply treats $\hat{\omega}^*$ as Normal$(0,1)$, or

$$\hat{\Pr}(\hat{\Theta} \le \hat{\theta}|a;\theta) \cong \Phi(\hat{\omega}^*) \tag{4.20}$$

This is known as Barndorff-Nielsen's r^* approximation, since the symbol r is often used in place of $\hat{\omega}$. Skovgaard (1996) gave an approximate expression for $\hat{z}$ denoted as $\breve{z}$

$$\breve{z} = \sqrt{j_{\hat{\theta}}} i_{\hat{\theta}}^{-1}\{(\xi_{\hat{\theta}} - \xi_{\theta})^T\Sigma_{\hat{\theta}}\dot{\xi}_{\hat{\theta}}\} = \sqrt{j_{\hat{\theta}}}(\dot{\xi}_{\hat{\theta}}^T\dot{\mu}_{\hat{\theta}})^{-1}(\xi_{\hat{\theta}} - \xi_{\theta})^T\dot{\mu}_{\hat{\theta}} \tag{4.21}$$

4.2.2 Boundary Selection

The values of A, B and C that satisfy (4.5), (4.6) and (4.7) can be determined by Monte Carlo simulation. For each simulation, first we randomly choose N samples from standard normal distribution $Y_1, Y_2, ..., Y_N$, and then get $S_k = Y_1 + ... + Y_k$, which has $N(0,k)$ distribution, $k = \rho N, ..., N$. Let $Z_k = S_k/\sqrt{k}$. The maximum and minimum of $(Z_{\rho N}, ..., Z_N)$ together with Z_N are recorded for all $M = 1,000,000$ simulations, after sorting all the maximums and minimums. The value of $-\sqrt{2A}$ is the observation $(M \times \varepsilon\alpha)^{th}$ of the sorted minimum; the $(M \times (1-\varepsilon\beta))^{th}$ maximum as the value of $\sqrt{2B}$ and the $(M \times \alpha(1-\varepsilon)/(1-\varepsilon\alpha))^{th}$ Z_N as the value of $-\sqrt{2C}$. Figure 4.1 shows the boundaries A, B and C and a sample trial of the Z_k.

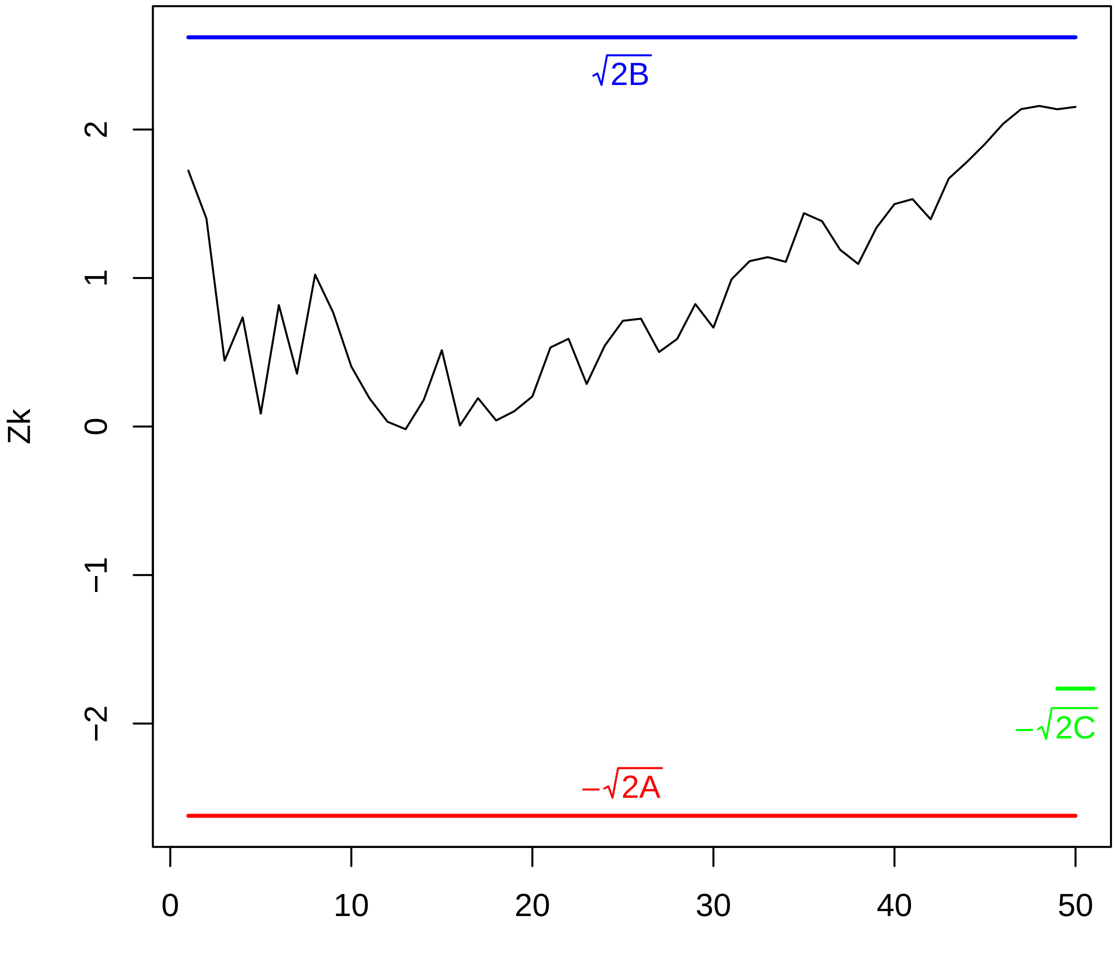

Figure 4.1: Boundaries $-\sqrt{2A}$, $\sqrt{2B}$ and $-\sqrt{2C}$ with a trial of Z_k

4.3 A Model for Responses

One of the most important statistical models used in logistic item response theory (IRT) based tests are logistic models. The item characteristic curves (ICC) of a three-parameter logistic (3-PL) model (Lord[24]) is defined as

$$P(\theta) \equiv P(U = 1|\theta; a, b, c) = c + \frac{1 - c}{1 + \exp\{-a(\theta - b)\}} \tag{4.22}$$

This models the probability of an examinee giving a correct answer with latent ability θ to a given item with parameter a, b and c. Here $U = 1$ or 0 denotes whether the item is answered correctly or incorrectly, and a, b and c are the item parameters of discrimination, difficulty and guessing, respectively. Specially, $c = 0$ makes the two parameter logistic (2-PL) model.

Let vector (a, b, c) denote an item selected from a pool with item parameters a, b and c. Suppose that there is $n - 1$ steps of an adaptive test, then for the $n - 1$ items, (a_j, b_j, c_j), $j = 1, ..., n - 1$, are administered to an examinee.

The probability that an examinee of ability θ gives the correct answer to the item j, $j \leq N$ is

$$P_j(\theta) = P(U_j = 1|\theta; a_j, b_j, c_j) = c_j + \frac{1 - c_j}{1 + \exp\{-a_j(\theta - b_j)\}} \tag{4.23}$$

N is set to 50. From $1,000,000$ simulated Markov chains, the stopping boundaries $A = B = 3.435$ and $C = 1.5565$ are chosen. Let $\alpha = \beta = 0.05$, $\varepsilon = 1/2$ and $\rho N = 10$. Bartroff, Finkelman and Lai[5] used a real item pool from the Chauncey Group International, a subsidiary of the Educational Testing Service. We generate (a_j, b_j, c_j) based on the five-number summary of Chauncey item pool parameters in Table (4.1).

Table 4.1: Five-number summary of Chauncey item pool parameters

	min	1st quartile	median	3rd quartile	max
a_j	0.289	0.683	0.862	1.074	2.372
b_j	-5.531	-1.998	-0.943	0.253	5.426
c_j	0.048	0.211	0.232	0.255	0.529

The real-life cut point associated with the item pool is $\theta_0 = -1.32$. Mimicking simulations by Lin and Spray (2000), θ_+ and θ_- are taken to be $\theta_0 \pm 0.25 = -1.57, -1.07$.

The probabilities (4.5), (4.6) and (4.7) with statistics $\hat{\omega}$ and $\hat{\omega}^*$ are showed in table 4.2

Table 4.2: The probabilities (4.5), (4.6) and (4.7) with statistics $\hat{\omega}$ and $\hat{\omega}^*$

	probabilities		
statistics	(4.5)	(4.6)	(4.7)
$\hat{\omega}$	0.03019	0.02672	0.02332
$\hat{\omega}^*$	0.02272	0.02595	0.02453

There is an obvious improvement by applying Barndorff-Nielsen's r^* formula. The three probabilities (4.5), (4.6) and (4.7) by using statistic $\hat{\omega}^*$ are all closer to the nominal $\varepsilon\beta, \varepsilon\alpha$ and $(1-\varepsilon)\alpha$ than by using the statistic $\hat{\omega}$.

Consider the case $c = 0$, the two parameter logistic (2-PL) model. We still use $\theta_0 = -1.32$, $\theta_+ = -1.07$ and $\theta_- = -1.57$. The probabilities (4.5), (4.6) and (4.7) with statistics $\hat{\omega}$ and $\hat{\omega}^*$ are showed in table 4.3

Table 4.3: The probabilities (4.5), (4.6) and (4.7) with statistics $\hat{\omega}$ and $\hat{\omega}^*$

	probabilities		
statistics	(4.5)	(4.6)	(4.7)
$\hat{\omega}$	0.0274	0.0242	0.026
$\hat{\omega}^*$	0.0268	0.0236	0.0252

The result for such a degenerated case is also showing that the method by applying Barndorff-Nielsen's r^* formula has better performance than normal used method on actual type II error and type I error the maximum cap N reached. The three probabilities (4.5) and (4.7) by using statistic $\hat{\omega}^*$ are all closer to the nominal $\varepsilon\beta$ and $(1-\varepsilon)\alpha$ than by using the statistic $\hat{\omega}$.

4.4 Conclusion

This paper shows how to choose the boundaries A, B and C for truncated sequential likelihood ratio test and the comparison of error rates using sequential GLR statistics and sequential GLR statistics in Barndorff-Nielsen's r^* formula. We find there is a markable improvement by applying Barndorff-Nielsen's r^* formula on the sequential GLR statistics.

Chapter 5

Conclusion

In this dissertation, we have discussed the comparison of various continuity corrections for saddlepoint approximation to the distribution of difference of two binomial samples. For 90% and 95% confidence intervals for difference of two binomial proportions, the saddlepoint approximation with the continuity correction

$$\mathrm{CC}_R = \left[1 - 2\sqrt{\frac{a+b}{m+n}(1-\frac{a+b}{m+n})}\right]\{2\mathrm{LCM}(m,\ n)\}^{-1}$$

is recommended when the minimum expected cell size (MCS) is not smaller than 1 or the nuisance parameter ψ is mesial $(.4-.6)$ or the parameter in interest θ is large $(\theta \geq .5)$; while the confidence intervals with continuity corrections $\mathrm{CC}_S = \{2\max(m,\ n)\}^{-1}$, $\mathrm{CC}_A = \{2\min(m,\ n)\}^{-1}$ and $\mathrm{CC}_Y = (2m)^{-1} + (2n)^{-1}$ are unnecessarily conservative and could not acceptable. When MCS ≤ 1, CC_S is the best choice of continuity correction, and it also the best choice for distal $\psi \geq .1$ or $\psi \leq .9$ or small θ in 0-.05. And the most useful result is that for MCS > 1, CC_R has coverage probabilities very close to the nominal 95% and 90% for any range of NMIN, even for NMIN as small as $5-9$. For 99% intervals, the Yates correction CC_Y is reasonable.

We also have discussed the asymptotic method with saddlepoint approximation for the modified Haybittle-Peto test. We propose a hybrid Monte Carlo-Asymptotic approach to calibrate the truncation boundary, in which we substitute an easy Monte Carlo approximation in place of boundary crossing probabilities for Brownian motions and use asymptotic approximations for the distribution of the signed root of the likelihood ratio test statistic. We found that after selecting stopping boundaries using normal-based Monte Carlo calculations, reliance on asymptotic normality of the signed root of the log likelihood ratio statistics provided adequate control of Type I error, without recourse to more complicated Monte Carlo operations. We also observe markable improvement using Barndorff-Nielsen's r^* formula.

References

[1] O. E. Barndorff-Nielsen. Modified signed log likelihood ratio. *Biometrika*, 78:557–563, 1991.

[2] O. E. Barndorff-Nielsen and D. Cox. *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, 1989.

[3] M. S. Bartlett. Approximate confidence intervals: II. more than one unknown parameter. *Biometrika*, 40:306–317, 1953.

[4] M. S. Bartlett. Approximate confidence intervals: III. a bias correction. *Biometrika*, 42:201–204, 1955.

[5] M. Bartroff, J. Finkelman and T. L. Lai. Modern sequential analysis and its applications to computerized adaptive testing. *Psychometrika*, 73:473–486, 2008.

[6] R. W. Butler. *Saddlepoint Approximations with Applications.* Cambridge Series in Statistical and Probabilistic Mathematics, 2007.

[7] H.-H. Chang and Z. Ying. a-straticed multistage computerized adaptive testing. *Appl. Psychol. Meas.*, 23:263–278, 1999.

[8] Y. Chang. Application of sequential probability ratio test to computerized criterion-referenced testing. *Sequential Analysis*, 23:45–61, 2004.

[9] Y. Chang. Application of sequential interval estimation to adaptive mastery testing. *Psychometrika*, 70:685–713, 2005.

[10] Y. Chang and Z. Ying. Sequential estimation in variable length computerized adaptive testing. *Journal of Statistical Planning and Inference*, 121:249–264, 1999.

[11] H. E. Daniels. Saddlepoint approximation in statistics. *Annals of Mathematical Statistics*, 25:630–650, 1954.

[12] H. E. Daniels. Tail probability approximations. *Internat. Statist. Rev.*, 55:37–46, 1987.

[13] W. W. Hauck and S. Anderson. A comparison of large-sample confidence interval methods for the difference of two binomial probabilities. *The American Statistician*, 40:318–322, 1986.

[14] E.V. Haynsworth and K. Goldberg. Bernoulli and euler polynomials, riemann zeta function. *Handbook of Mathematical Functions*, 1997.

[15] I. N. Herstein. *Topics in Algebra, 2nd Edition.* John Wiley & Sons, 1975.

[16] Lai D. Wang S. J. and Hsi B. P. Saddlepoint methods in testing the difference between two binomial proportions:a comparative study. *Journal of Theoretical Statistics*, 3:239–263, 1999.

[17] J. L. Jensen. The modified signed likelihood statistic and saddlepoint approximations. *Biometrika*, 79:693–703, 1992.

[18] J. L. Jensen. Asymptotic expansions at work. *Scandinavian Actuarial Journal*, 1:143–152, 1995.

[19] J. L. Jensen. *Saddlepoint Approximations.* Oxford University Press, 1995.

[20] J. E. Kolassa. *Series Approximation Methods in Statistics, 3rd Edition (Lecture Notes in Statistics #88).* Springer, New York, 2006.

[21] T.L. Lai and M.C. Shih. Power, sample size and adaptation considerations in the design of group sequential clinical trials. *Biometrika*, 91:507–528, 2004.

[22] C. Lewis and K. Sheehan. Using bayesian decision theory to design a computerized mastery test. *Appl. Psychol. Meas.*, 14:367–386, 1990.

[23] C. Lin and J. Spray. *Effects of item-selection criteria on classification testing with the sequential probability ratio test (Research Report 2000-8).* Iowa City, IA:American College Testing., 2000.

[24] F. Lord. *Applications of item response theory to practical testing problems.* Erlbaum, Hillsdale, NJ, 1980.

[25] R. Lugannani and S. Rice. saddlepoint approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, 12:475–490, 1980.

[26] R. G. Newcombe. Interval estimation for the difference between independent proportions:comparison of eleven methods. *Statistics in Medicine*, 17:873–890, 1998.

[27] R.J. Owen. A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *J. Amer. Statist. Assoc.*, 70:351–356, 1975.

[28] W.J. van der Linden and P. Pashley. Item selection and ability estimation in adaptive testing. in w.j. van der linden & c.a.w. glas (eds.). *Computerized adaptive testing: theory and practice*, pages 1–25, 2000.

[29] D.J. Weiss. Adaptive testing research in minnesota: overview, recent results, and future directions. In *In: Clark, C.L. (Ed.), Proceedings of the First Conference on Computerized Adaptive Testing, United States Civil Service Commission*, pages 24–35, Washington, DC, December 1976.

[30] D.J. Weiss. Improving measurement equality and efficiency with adaptive testing. *Appl. Psychol. Meas.*, 6:379–396, 1982.

[31] L. Xu and J. E. Kolassa. Testing the difference of two binomial proportions: comparison of continuity corrections for saddlepoint approximation. *Communications in Statistics, Theory and Methods*, 37:2213–2218, 2008.

[32] L. Xu and J. E. Kolassa. Continuity-corrected saddlepoint approximation for confidence interval of the difference between two binomial proportions: Comparison of six continuity corrections. In *Proceedings of the 65th Deming Conference*, Atlantic City, NJ, December 2009.

Vita

Lu Xu

EDUCATION

2005-2010 Ph. D. in Statistics, Rutgers University

2002-2003 M. Sc. in Statistics, University of Toronto

1998-2002 B. Sc. in Statistics, Nankai University

PROFESSIONAL EXPERIENCE

2009- Senior Statistician, GlaxoSmithKline

2008-2009 Graduate assistant, Department of Statistics, Rutgers University

2007-2008 Teaching assistant, Department of Statistics, Rutgers University

2005-2007 Fellowship, Institute for Health, Health Care Policy and Aging Research, Rutgers University

PUBLICATION

2009 Continuity-corrected Saddlepoint Approximation for Confidence Interval of the Difference Between Two Binomial Proportions: Comparison of Six Continuity Corrections. In *Proceedings of the 65th Deming Conference,* Atlantic City, NJ, December, 2009.

2008 Testing the difference of two binomial proportions: comparison of continuity corrections for saddlepoint approximation. *Communications in Statistics, Theory and Methods,* 2008, Volume 37, page 2213-2218.