Thesis for the Degree of Master of Science

Review on the Coverage Probabilities of Confidence Intervals for a Binomial Proportion

by

Lionel Sahabo

Department of Statistics

The Graduate School

Pukyong National University

August 26, 2016

Review on the coverage probabilities of confidence intervals for a binomial proportion
(이항모수에 대한 신뢰구간들의 포함확률에 관한 고찰)

Advisor: Prof. Seongbaek Yi

by

Lionel Sahabo

A thesis submitted in partial fulfillment of the requirements
for the degree of

Master of Science

in Department of Statistics, The Graduate School,
Pukyong National University

August 26, 2016

Review on the coverage probabilities of confidence intervals for a binomial proportion

A dissertation
by
Lionel Sahabo

Approved by:

____
Ph.D. Dae-Heung Jang

____
Ph.D. Seongbaek Yi

____
Ph.D. Yaeji Lim

August 26, 2016

Table of Contents

List of Tables

List of Figures

Review on the Coverage Probabilities of Confidence Intervals
for a Binomial Proportion

Lionel Sahabo

Department of Statistics, The Graduate School,
Pukyong National University

Abstract

The interval estimation of the parameter for the probability of success p in a binomial distribution is one of the most basic and methodologically important problems in statistical practice. Since the first introduction of the Wilson interval, many modified intervals have been developed. In most elementary statistics textbooks the Wald interval is nearly universally accepted. In this study we review popular confidence intervals in terms of coverage probabilities such as the Clopper-Pearson exact interval, the Wald interval, the Wilson's score interval, the Agresti & Coull (or adjusted Wald) interval, and Bayesian credible intervals with beta priors. Their performances are evaluated using such criteria as mean coverage probability, expected length, average expected length, and mean absolute error. According to the above criteria, the interval by equal-tailed Bayesian method with a beta prior shows comparable results.

Key words : Confidence interval, Binomial proportion, Coverage probability, Bayesian credible interval.

이항모수에 대한 신뢰구간들의 포함확률에 관한 고찰

Lionel Sahabo

부경대학교 대학원 통계학과

요약

이항분포의 모수인 성공확률 p의 구간추정은 통계적 실제상황에서 가장 기본적이며 또한 방법론적으로도 매우 중요한 문제이다. Wilson에 의해 구간추정이 처음 소개된 이후로 개선된 많은 구간추정방법들이 개발되어 왔다. 대부분의 기초통계학 교재에는 Wald의 구간추정방법이 거의 모든 책에서 범용적으로 채택되어 사용되고 있다. 본 연구에서는 기존에 발표된 구간추정방법들 중 포함확률에 의거하여 Clopper-Pearsonl, Wald, Wilson, Agresti & Coull의 방법들과 Bayesian 방법들을 중심으로 고찰하였다. 또한 이러한 방법들을 비교하기 위하여 평균포함확률, 신뢰구간의 기댓값, 평균기대구간의 길이 및 절대오차의 평균값 등과 같은 평가기준을 사용하였다. 본 평가기준에 의하면 베타사전분포를 사용한 베이지안 방법에 의해 구해진 신뢰구간의 포함확률이 기존의 비베이지안방법 못지 않은 결과를 보여주었다.

Ⅰ. Introduction

The interval estimation of the parameter for the probability of success p in a binomial distribution is one of the most basic and methodologically important problems in statistical practice.

The nearly universally accepted confidence interval in most elementary statistics textbooks is the Wald confidence interval defined as

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ ,$$

where $\hat{p}$ is the sample proportion of success and z_{α} is the $100(1-\alpha)\%$ percentile of standard normal distribution. The actual coverage probabilities of the standard interval displays poor performance for p near 0 or 1. Recent articles point out that the coverage probabilities can be erratically poor even if p is not near the boundaries.

The purpose of the study is to review coverage probabilities of popular confidence intervals such as exact interval (Clopper & Pearson 1934), Wald interval (P-S Laplace 1812) with z-critical value and t-critical value as well, Wilson's score interval (Wilson 1927), Agresti & Coull (or adjusted Wald) interval (A & C 1998) and Bayesian credible interval with beta prior and Jeffrey prior (Brown, Cai and DasGupta 2001).

The Wald confidence interval is first introduced by Laplace (1812) and its coverage properties have been remarked on Gosh (1979), Blyth & Still (1983), Santner (1988), and Agresti & Coull (1998).

The Clopper-Pearson interval is based on inverting alternative two-sided tests for $H_0 : p = p_0$ and is termed exact due to the availability of endpoints reaching the desired significance level for all x. See Böhning (1994), Leemis and Trivedi (1996), and Jovanovic & Levy (1997) for the method.

Wilson's score confidence interval is also based on inverting the equal-tailed test, but uses the standard error with null value instead of estimator.

The Agresti-Coull interval is based on the Wilson's score interval, but it has simple formula like the Wald interval plugging new n and p. Thus it keeps the simplicity and catchy formula of Wald interval. This interval is first introduced in the elementary statistics textbook by Samuels & Witmer (1999).

The Bayesian confidence intervals use beta conjugate priors for the binomial likelihood. Bayesian confidence intervals with noninformative beta priors such as uniform and Jeffrey prior also perform well in a frequentist sense. For further results see Carlin and Louis (1996) and Brwon, Cai, and DasGupta (2001).

There exist many literatures about methods for constructing the confidence intervals of the binomial parameter. See Santner and Duffy (1989), Vollset (1993), Pires and Amado (2008) etc. Especially Pires and Amado list twenty methods and compared them in terms of coverage probabilities.

Many parts of this study is to reproduce major results of the existing methods and confirm the pros and cons of their coverage probabilities.

For the study we introduce some terminologies and Lemmas in the first part of chapter 2. Confidence intervals by frequentist approach are reviewed at the remaining part of chapter 2. In chapter 3 we look at interval estimator by Bayesian approach using beta conjugate priors. Their performances are compared in terms of criteria using coverage probabilities. In chapter 4 we compare the methods using some performance criteria and we summarize the pros and cons of the methods in chapter 5.

II. Non-Bayesian Confidence Intervals

2.1 Introduction

Assume a discrete random variable X has the probability function $p(x;\theta)$ of x which depends on unknown parameter θ. The problem of confidence intervals consists in ascribing to every possible values of X, e.g. x_k $(k=1,2,\ldots)$ an interval $I_k=(L(x_k),\ U(x_k))$ such that whenever we observe x_k the probability of our being correct in $\theta\in I_k$ is $P(\theta\in I_k)\geq 1-\alpha$, where α belongs to $(0,1)$ and is chosen in advance. Here are some terminologies which are used in this study:

- the confidence coefficient is $\inf_{\theta\in[0,1]} P(\theta\in I_k)$
- the coverage probability of the interval I_k is $P(\theta\in I_k)$
- the nominal coverage probability is $1-\alpha$

The bigger the actual coverage probability is than the nominal coverage probability $1-\alpha$, the corresponding interval is considered conservative. When the random variable X is of continuous type, we have $P(\theta\in I_k)=1-\alpha$ for every θ.

In this study we use some properties of the cumulative distribution function of the binomial random variable.

Lemma 2.1.1 Let $X\sim bin(n,p)$. We then are able to express the right tail summation of the binomial probabilities in terms of the cumulative distribution function (cdf) of a beta distribution as follows:

$$
\begin{aligned}
P(X \geq k) &= \sum_{j=k}^{n} \binom{n}{k} p^j (1-p)^{n-j} \\
&= \int_0^p \frac{n!}{(k-1)!(n-k)!} z^{k-1}(1-z)^{n-k} dz \qquad (2.1.1) \\
&\equiv I_p(k, n-k+1)
\end{aligned}
$$

where $I_p(a,b) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \displaystyle\int_0^p t^{a-1}(1-t)^{b-1} dt$ denotes the cdf of a beta random variable with parameters $a > 0$ and $b > 0$.

Similarly the left tail summation of binomial probabilities can be obtained as follows:

$$
P_p(X \leq k) = 1 - P_p(X \geq k+1) = 1 - I_p(k+1, n-k). \qquad (2.1.2)
$$

Proof. For the proof we refer the exercise 3.3.22 of Hogg, McKean and Craig. ❒

Lemma 2.1.2 Let us denote the cdf of binomial random variable as a function of the success proportion p, that is,

$$
P_p(X \leq k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}.
$$

The cdf is then strictly decreasing in p for $k = 0, 1, \ldots, n-1$ and $P_p(X \geq k)$ is strictly increasing in p for $k = 1, \ldots, n$.

Proof. The strictly increasing property of the upper tail sum is proved by showing that the partial derivative of $P_p(X \geq k)$ with respect to p takes positive value:

$$\begin{aligned}\frac{\partial P_p(X\geq k)}{\partial p} &= \sum_{j=k}^{n}\binom{n}{j} j p^{j-1}(1-p)^{n-j} - \sum_{j=k}^{n-1}\binom{n}{j}(n-j)p^{j-1}(1-p)^{n-j-1} \\ &= n\left\{\sum_{j=k}^{n}\binom{n-1}{j-1}p^{j-1}(1-p)^{n-j} - \sum_{j=k}^{n-1}\binom{n-1}{j}p^{j}(1-p)^{n-j-1}\right\} \\ &= k\binom{n}{k}p^{k-1}(1-p)^{n-k} > 0.\end{aligned}$$

The third equality follows from the following binomial identities

$$j\binom{n}{j} = n\binom{n-1}{j-1} \text{ and } (n-1)\binom{n}{j} = n\binom{n-1}{j}.$$

The other part, the strictly increasing property of the lower tail sum, is derived by the complement the cdf $P_p(X\geq k) = 1 - P_p(X\leq k-1)$. ❐

2.2 The Exact Interval

When we are interested in testing $H_0 : p = p_0$ vs $H_a : p < p_0$, small observed values can be used as supportive evidence against the alternative hypothesis. With level of significance α we can reject H_0 for $X \leq k(p_0,\alpha)$, where $k(p_0,\alpha)$ is selected as the largest integer k for which $P_{p_0}(X\leq k) \leq \alpha$, that is, $k(p_0,\alpha) \equiv \max\{k : P_{p_0}(X\leq k) \leq \alpha\}$. Thus we get to reject H_0 for $X > k(p_0,\alpha)$.

Sometimes the $p-$value corresponding to an observed value x of X is more informative and convenient in testing hypotheses. The $p-$value is the probability of obtaining an effect at least as extreme as the observed one under the truth of the null hypothesis. It measures how compatible sampled data are with the null hypothesis. The high $p-$values imply observed data are likely with a true null, while low $p-$values imply data in hand are unlikely with a true null. A low $p-$value thus suggests that observed data provide enough evidence that we can reject the null hypothesis for the entire population.

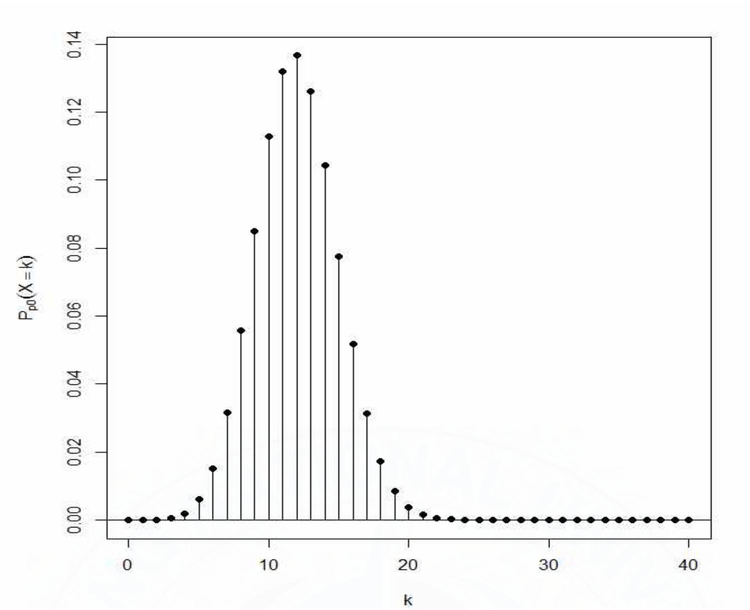
In our testing situation the $p-$value corresponding to the observed x is $P_{p_0}(X \leq x)$, denoted $p(x,p_0)$ here. The decision of rejecting or accepting H_0 is based on the $p-$value $p(x,p_0)$, that is, H_0 is rejected whenever $p(x,p_0) \leq \alpha$ and accepted when $p(x,p_0) > \alpha$.

The plots of Figure 2.1 show the probability mass function and cumulative distribution function of the binomial random variable with $n=40$, $p=0.3$. In this case the largest value x for which to reject H_0 is 6, that is

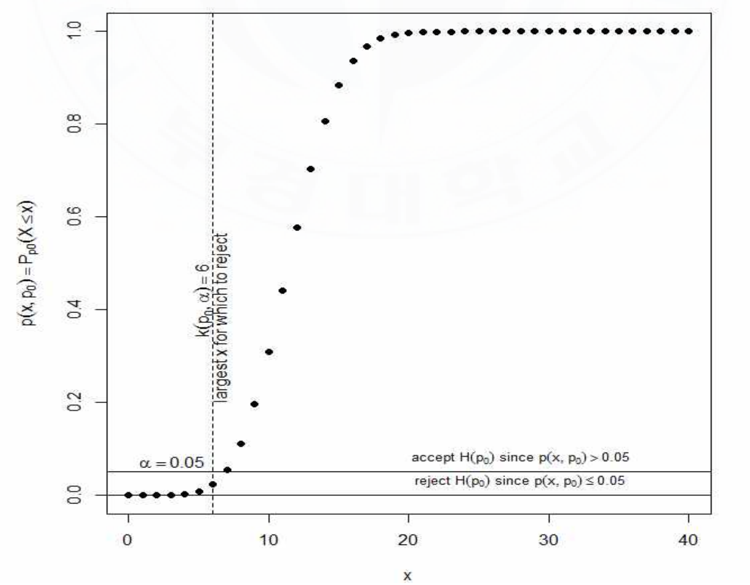
$$\max\{k: P_{0.3}(X \leq k) \leq 0.05\} = k(p_0 = 0.3, \alpha = 0.05) = 6.$$

If confidence set is defined as all values p_0 for which the observed value x of the random variable X leads into accepting H_0, we can set a relationship between hypotheses testing and interval estimation. We denote the confidence set as $I(x)$. The confidence sets can be defined for each possibly observed value $x=0,1,\ldots,n$, thus be thought of a random set $I(X)$ before having observed any x.

Figure 2.2 shows the cdf $F_p(x) = P_p(X \leq x)$ is a decreasing function of p for each x, which implies that the confidence set $I(x)$ for a specific realized value x should have the form $[0, p_{x,\alpha})$, where $p_{x,\alpha}$ is the value p solving the equation $P_p(X \leq x) = \alpha$.



(a) Probability mass function of $b(40, 0.3)$



(b) Cumulative distribution function of $b(40, 0.3)$

Figure 2.1. Binomial distribution.

Let us see in detail how to find the value $p_{x,\alpha}$ for each $x=0,1,...,n$. The definition of $I(x)$ was defined as the values of p for which the observed value x causes accepting H_0, in other words, producing $p-value>\alpha$, we can have the following expression in terms of $p-value$

$$I(x)=\{p_0 : p(x,p_0)=P_{p_0}(X\le x)>\alpha\}.$$

When $x=0,1,...,n-1$, Lemma 2.1.2 tells us that $P_{p_0}(X\le x)$ is strictly decreasing in p_0, the confidence set $I(x)$ is the collection of the values p_0 in the interval $[0,p_{u,\alpha,x})$, where $p_{u,\alpha,x}$ is the value p satisfying

$$P_p(X\le x)=\sum_{i=0}^{x}\binom{n}{i}p^i(1-p)^{n-i}=\alpha \tag{2.2.1}$$

By Lemma 2.1.1, the relationship between the cdf of binomial and beta distribution,

$$P_p(X\le x)=1-I_p(x+1,n-x)=\alpha \text{ or } I_p(x+1,n-x)=1-\alpha,$$

we can obtain the $p_{u,\alpha,x}$ in R as *qbeta*$(1-\alpha,x+1,n-x)$.

However for $x=n$ we have $P_p(X\le n)=1$ for all p. The $p-value$ with the observed value n is thus 1 and always exceeds α. We get the confidence interval $I(n)=[0,1]$. That is $p_{u,\alpha,n}=1$.

Lemma 2.2.1 Using equation (2.2.1) we can see that the sequence $(p_{u,\alpha,x})_{x=0}^{n}$ is strictly increasing in x, that is,

$$0<p_{u,\alpha,0}<p_{u,\alpha,1}<\cdots<p_{u,\alpha,n-1}<p_{u,\alpha,n}=1$$

and we have for all $p<p_{u,\alpha,0}$

$$P_p(p < p_{u,\alpha,X}) \geq P_p(p < p_{u,\alpha,0}) = 1.$$

Figure 2.3 shows the set of $p_{u,\alpha,x}$ for $\alpha = 0.05$ and $n = 100$ together with corresponding estimates $\hat{p}(x) = x/n$. In this case the possible smallest upper bound is $p_{u,0.05,0} = qbeta(0.95, 1, 100) = 0.02951$.

Lemma 2.2.2 The confidence coefficient of the confidence set is defined as

$$\inf_{p \in [0,1]} P_p(p \in I(X)) = \inf_{p \in [0,1]} P_p(p < p_{u,\alpha,X}),$$

which is $1-\alpha$ since $P_p(p \in I(X)) = 1-\alpha$ for some p. This implies for some p

$$P_p(p \notin I(X)) = P_{p_0}(X \leq k(p_0, \alpha)) = \alpha.$$

Proof. For $x = 0, 1, \ldots, n-1$ we have $p_{u,\alpha,x}$ as the solution p of $P_p(X \leq x) = \alpha$. We thus have for $p_x = p_{u,\alpha,x}$, $x = 0, 1, \ldots, n-1$ with $k(p_x, \alpha) = x$

$$P_{p_x}(X \leq k(p_x, \alpha)) = P_{p_x}(X \leq x) = \alpha,$$

which implies that the infimum is obtained at $p = p_0, p_1, \ldots, p_{n-1}$. ❑

We now see that the coverage probability has continuity properties. For a fixed value of p, the coverage probability of a confidence set $I(X)$ is defined as the probability of the interval containing that value as follows

$$P_p(p \in I(X)) = 1 - P_p(p \notin I(X)) = 1 - P_p(X \leq k(p, \alpha)).$$

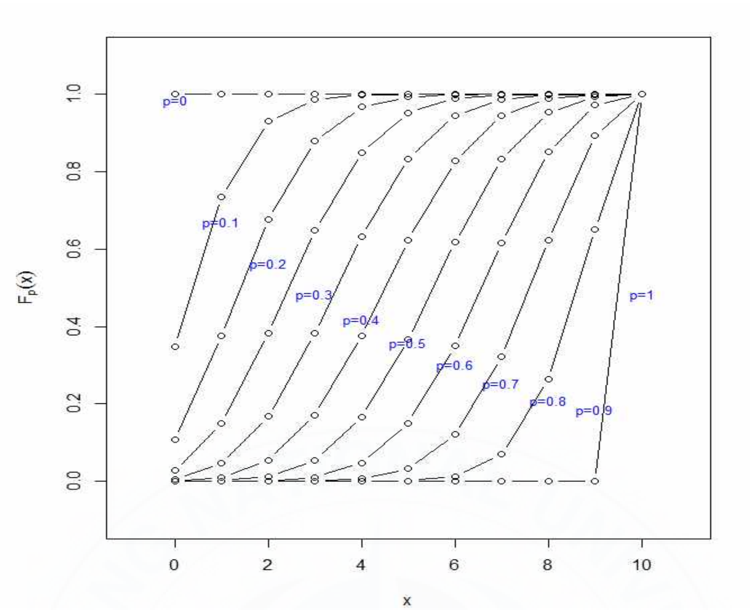


Figure 2.2. Strictly decreasing property of $F_p(x)$.

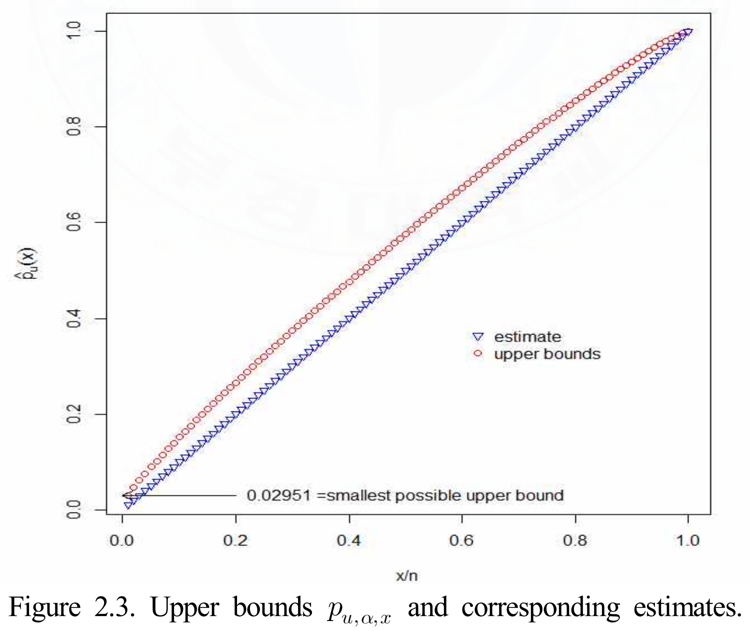


Figure 2.3. Upper bounds $p_{u,\alpha,x}$ and corresponding estimates.

Recall from Lemma 2.2.1 that $0 < p_{u,\alpha,0} < p_{u,\alpha,1} < \cdots < p_{u,\alpha,n-1} < p_{u,\alpha,n} = 1$. Let us denote $p_i = p_{u,\alpha,i}$ for $i = 0,1,\ldots,n$. For $p \in [p_{i-1}, p_i)$, the set $C = \{k : p < p_k\} = [i,n]$ does not vary, since for $j = i,\ldots,n$, we have $p < p_j$ when $p_{i-1} \le p < p_i$. Also the fact that $P_p(X \le k)$ is decreasing function of p implies that $P_p(X \in C) = P_p(X \ge i)$ increases continuously in p over the interval $[p_{i-1}, p_i)$.

When p reaches at p_i, the set C does lose the value i and $P_p(p < p_{u,\alpha,X})$ drops by $P_{p_i}(X = i)$ to

$$P_{p_i}(X \ge i+1) = 1 - P_{p_i}(X \le i) = 1 - \alpha.$$

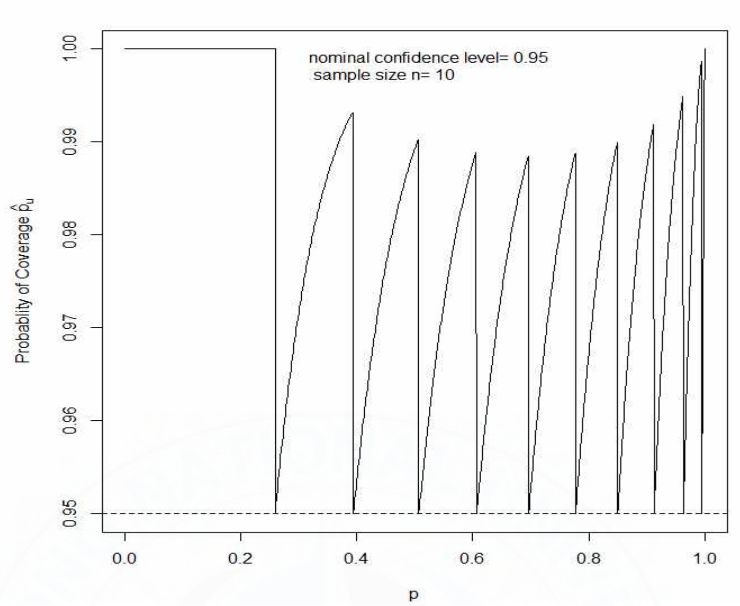
We can develop the similar procedures for the other intervals of p. Figure 2.4 displays the continuity property of the coverage probability. In R the plot can be computed by calculating

$$P(p < p_{u,\alpha,X}) = \sum_{x=0}^{n-1} I_{\{qbeta(1-\alpha, x+1, n-x) > p\}} \binom{n}{x} p^x (1-p)^{n-x} + p^n.$$

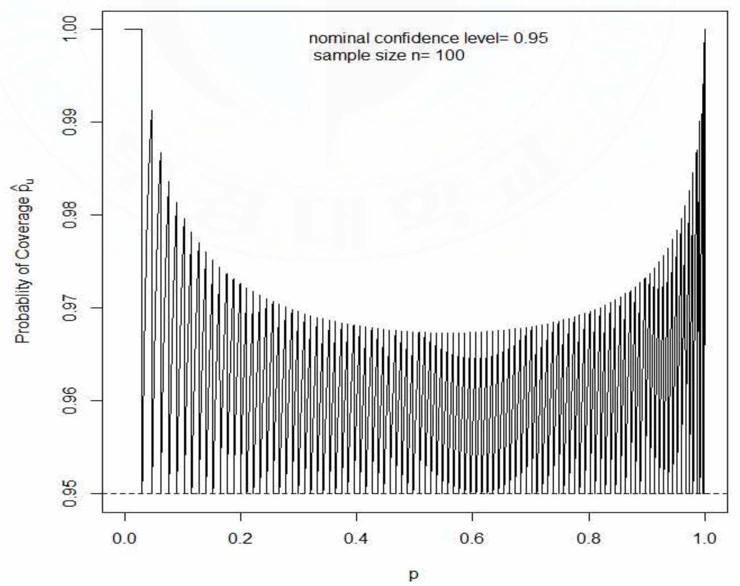
Now suppose we are interested in testing alternative hypothesis $H_a : p > p_0$. The confidence set then is given by

$$I(x) = \{p_0 : p(x, p_0) = P_{p_0}(X \ge x) > \alpha\}.$$

For $x = 1,\ldots,n$ we get the confidence set equal to $I(x) = (p_{L,\alpha,x},\ 1]$, where $p_{L,\alpha,x}$ is the value p solving $P_p(X \ge x) = \alpha$. In R we can obtain $p_{L,\alpha,x}$ as $p_{L,\alpha,x} = qbeta(\alpha, x, n-x+1)$, since Lemma 2.1.1 says $P_p(X \ge x) = I_p(x, n-x+1)$.



(a) The nominal confidence level with $n=10$



(b) The nominal confidence level with $n=100$

Figure 2.4. Coverage probability behavior of upper bound.

With $x=0$ we have $P_p(X \geq 0) = 1 > \alpha$ for all values p in $[0,1]$, we thus have $I(0) = [0,1]$. The coverage probability is given by

$$P(p > p_{L,\alpha,X}) = (1-p)^n + \sum_{x=1}^{n} I_{\{qbeta(\alpha,x,n-x+1) < p\}} \binom{n}{x} p^x (1-p)^{n-x}.$$

The continuity behavior of the coverage probability is displayed in Figure 2.5.

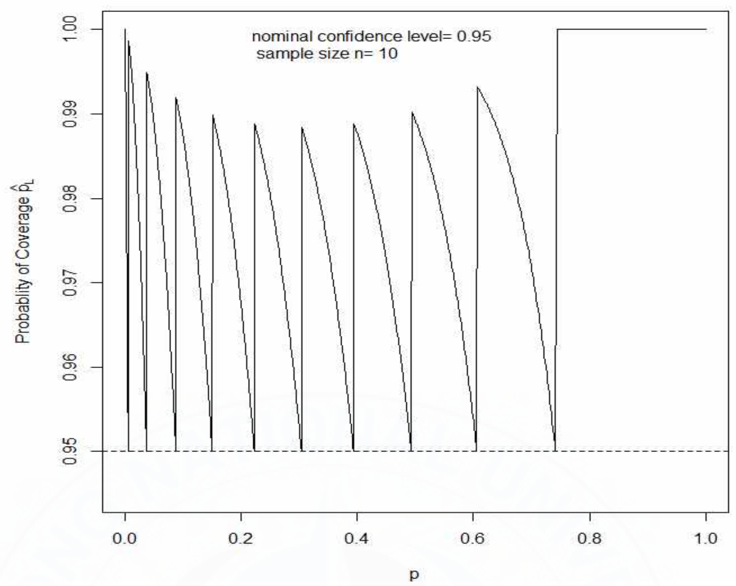
As explained till now for lower and upper bound of confidence intervals corresponding to upper tail and lower tail alternative hypotheses, we can extend the procedure to two-sided hypothesis with each confidence coefficient $1-\alpha/2$ respectively. Assuming that $P_p(p_{l,1-\alpha/2,X} < p_{u,1-\alpha/2,X}) = 1$ for any p and any $x = 0,1,\ldots,n,$ we have the desired confidence interval $I(x) = (p_{l,1-\alpha/2,x}\ ,\ p_{u,1-\alpha/2,x})$. The coverage probability for this interval is

$$\begin{aligned}
&P_p(p_{l,1-\alpha/2,X} < p < p_{u,1-\alpha/2,X}) \\
&= 1 - P_p(\{p \leq p_{l,1-\alpha/2,X}\} \cup \{p_{u,1-\alpha/2,X} \leq p\}) \\
&= 1 - P_p(\{p \leq p_{l,1-\alpha/2,X}\}) - P_p(\{p_{u,1-\alpha/2,X} \leq p\}) \\
&\geq 1 - \alpha/2 - \alpha/2 \\
&= 1 - \alpha
\end{aligned}$$

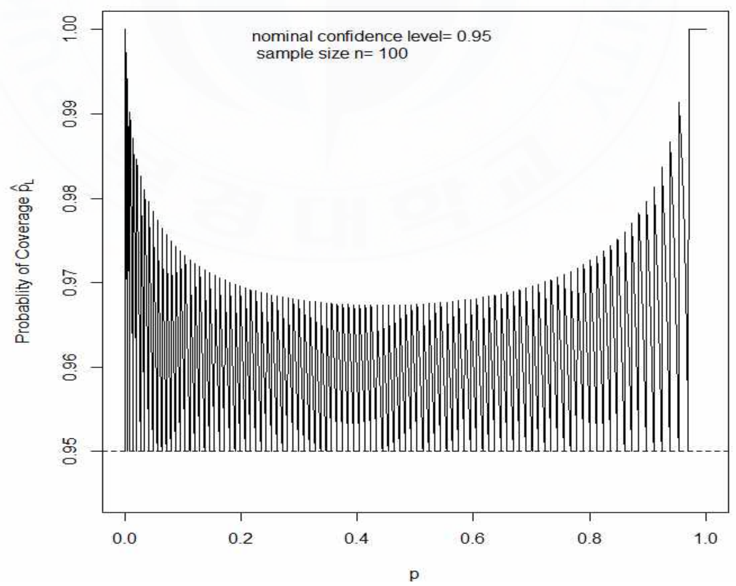
The actual coverage probability is computed in R as follows

$$\begin{aligned}
&P_p(p_{l,1-\alpha/2,X} < p_{u,1-\alpha/2,X}) \\
&= \sum_{x=1}^{n-1} I_{\{qbeta(\alpha/2,x,n-x+1) < p < qbeta(1-\alpha/2,x+1,n-x)\}} \binom{n}{x} p^x (1-p)^{n-x} \\
&\quad + (1-p)^n I_{\{qbeta(1-\alpha/2,1,n) > p\}} + p^n I_{\{qbeta(\alpha/2,n,1) < p\}}.
\end{aligned}$$

Figure 2.6 shows the behavior of the coverage probability for different sample sizes.

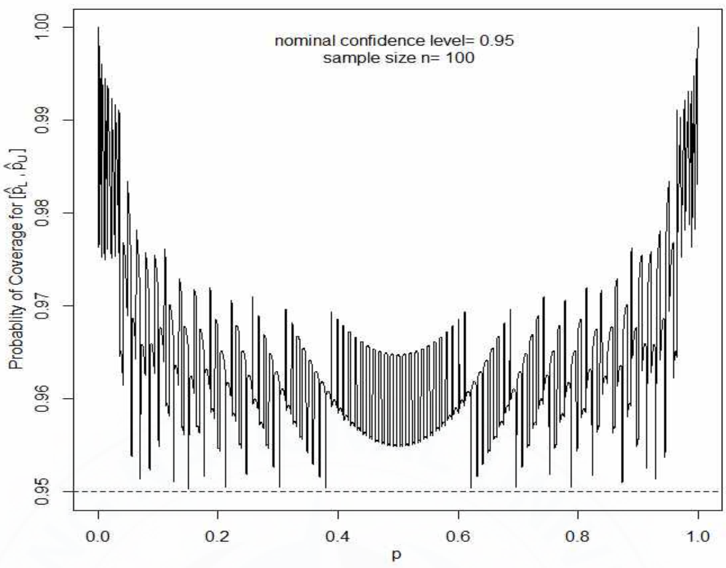


(a) The nominal confidence level 0.95 with $n=10$

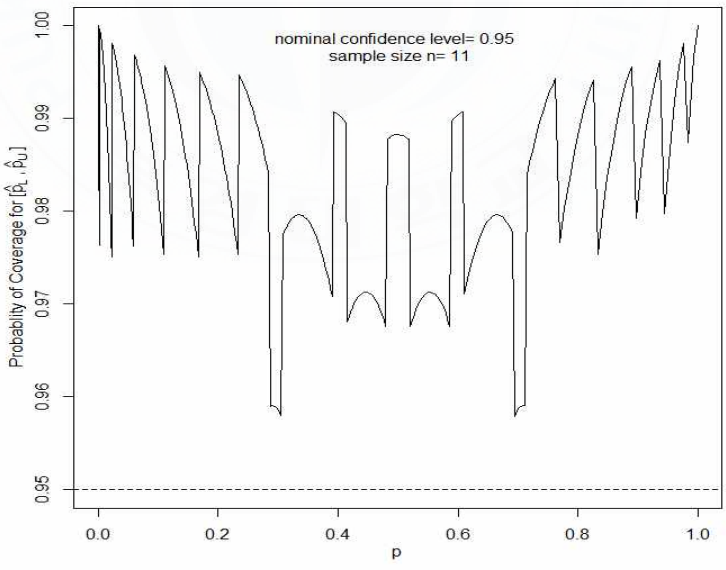


(b) The nominal confidence level 0.95 with $n=100$

Figure 2.5. Coverage probability behavior of lower bound.



(a) The nominal confidence level 0.95 with $n = 100$



(b) The nominal confidence level 0.95 with $n = 11$

Figure 2.6. Coverage probability behavior of confidence interval.

The coverage probability plots for all types of alternative hypotheses show conservatism, seemingly not reaching as far down as $1-\alpha$, 0.95 in the displayed cases. This conservative, higher than $1-\alpha$, coverage is caused by the fact that the supremum of a sum is not greater than the sum of supremum over the involved sums. For example with two-sided alternative hypothesis case, we have

$$\begin{aligned}
&\inf_p P_p\left(p_{l,1-\alpha/2,X} \le p \le p_{u,1-\alpha/2,X}\right) \\
&\quad = 1-\sup_p P_p\left(\{p < p_{l,1-\alpha/2,X}\} \cup \{p_{u,1-\alpha/2,X} < p\}\right) \\
&\quad = 1-\sup_p \left[P_p\left(\{p < p_{l,1-\alpha/2,X}\}\right) + P_p\left(\{p_{u,1-\alpha/2,X} < p\}\right)\right] \\
&\quad \ge 1-[\sup_p P_p\left(\{p < p_{l,1-\alpha/2,X}\}\right) + \sup_p P_p\left(\{p_{u,1-\alpha/2,X} < p\}\right)] \\
&\quad = 1-(\alpha/2+\alpha/2) \\
&\quad = 1-\alpha.
\end{aligned}$$

Specifically with extreme p, close to 0 or 1, the coverage probability plot shows abrupt rise to $1-\alpha/2$. This is due to the fact that for $p < p_{u,1-\alpha/2,0}$ we have $P_p\left(\{p_{u,1-\alpha/2,X} < p\}\right) = 0$, which results in

$$\begin{aligned}
&P_p\left(p_{l,1-\alpha/2,X} \le p \le p_{u,1-\alpha/2,X}\right) \\
&= 1-P_p\left(\{p < p_{l,1-\alpha/2,X}\}\right) - P_p\left(\{p_{u,1-\alpha/2,X} < p\}\right) \\
&= 1-P_p\left(\{p < p_{l,1-\alpha/2,X}\}\right) \\
&\ge 1-\alpha/2.
\end{aligned}$$

Similarly for $p > p_{l,1-\alpha/2,n}$ we have $P_p\left(\{p < p_{l,1-\alpha/2,X}\}\right) = 0$, which results in

$$\begin{aligned}
&P_p\left(p_{l,1-\alpha/2,X} \le p \le p_{u,1-\alpha/2,X}\right) \\
&= 1-P_p\left(\{p < p_{l,1-\alpha/2,X}\}\right) - P_p\left(\{p_{u,1-\alpha/2,X} < p\}\right) \\
&= 1-P_p\left(\{p_{l,1-\alpha/2,X} < p\}\right) \ge 1-\alpha/2.
\end{aligned}$$

Therefore it does not make sense to allocate the probability $\alpha/2$ to the lower bound or the upper bound, which causes a conservative evaluation for the extreme values of p.

One of many reasons why it is impossible to achieve the exact nominal confidence level is due to the discrete nature of the binomial distribution. The Clopper-Pearson's exact method always produces the actual coverage probability bigger than the nominal level. This implies the method is too conservative. We need a better method in the sense that its resulting confidence interval makes the actual coverage probability close to the nominal confidence level.

2.3 The Wald Interval

As pointed out in the last part of the exact method, we would like to have the actual coverage probability close to the nominal confidence level in developing a confidence interval estimation. However the discrete nature of the binomial distribution makes it not possible to achieve the exact nominal confidence level.

The confidence interval by Wald is based on normal approximation and has gained universal recommendation in the elementary statistics books. The interval is given by

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p}=X/n$ and $z_{\alpha/2}$ solves the equation $P(Z \leq z_{\alpha/2}) = 1-\alpha/2$. This interval is obtained through the duality of the Wald's asymptotic normal test for a general problem:

$$\left|\frac{\hat{\theta}-\theta}{\widehat{se}(\hat{\theta})}\right| \leq k$$

where $\hat{\theta}$ is the MLE of θ and $\widehat{se}(\hat{\theta})$ the estimated standard error of $\hat{\theta}$ and k is the $(1-\alpha/2)\times 100$ percentile of standard normal distribution or t distribution. In binomial case, $\theta = p$, $\hat{\theta} = X/n$ and $\widehat{se}(\hat{\theta}) = \sqrt{\hat{p}\hat{q}/n}$.

This interval estimation was first described by Pierre-Simon Laplace in his 1812 book *Théorie analytique des probabilités* (page 283). It is simple to calculate and often justified by the central limit theorem. Most students seem not to give any doubt that the larger the n, the better the asymptotic property. They believe that the conservatism of the exact method will disappear as the number n get increased. However, this is not the case. Figure 2.7 shows that there exist significant oscillation for fixed p. This swinging phenomenon is still present with increasing the n even to 1000. When we fix n and vary values of p, there exist systematic bias in the coverage probability which is displayed in Figure 2.8. This problematic unsatisfactory coverage probability, not reaching to $1-\alpha$, is bad at extreme values of p.

From this plots we can be sure that the normal approximation as a rational for Wald interval estimation can be seriously erratic when the true value of p is at boundaries such as 0 or 1.

The problematic phenomenon of coverage probability for p near 0 or 1 are cited in many popular textbooks. Brown, Cai & DasGupta (2001) list the following qualifications by examination of 11 popular textbooks:

The confidence interval may be used if

① $np,\, n(1-p) \geq 5$ (or 10);

② $np(1-p) \geq 5$ (or 10);

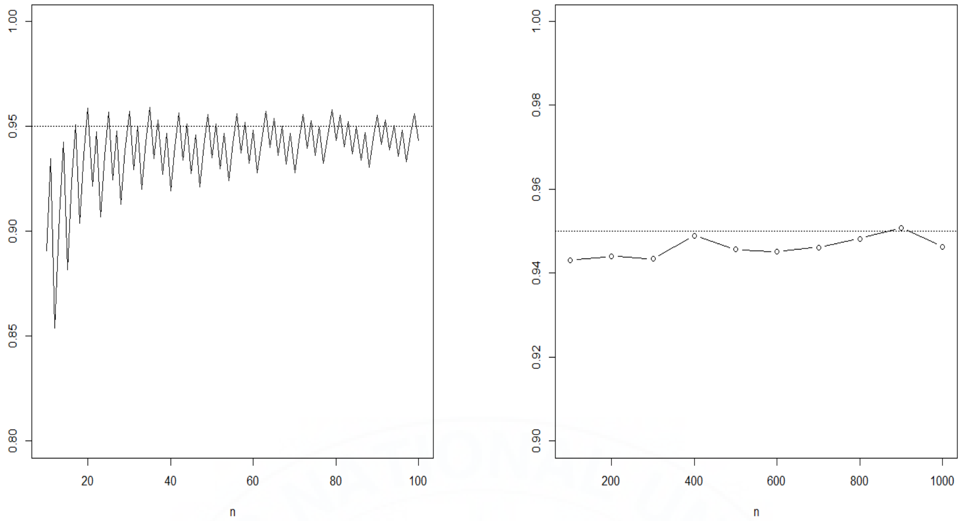
③ $n\hat{p},\, n(1-\hat{p}) \geq 5$ (or 10);

④ $\hat{p} \pm 3\sqrt{\hat{p}(1-\hat{p})}$ does not contain 0 or 1;

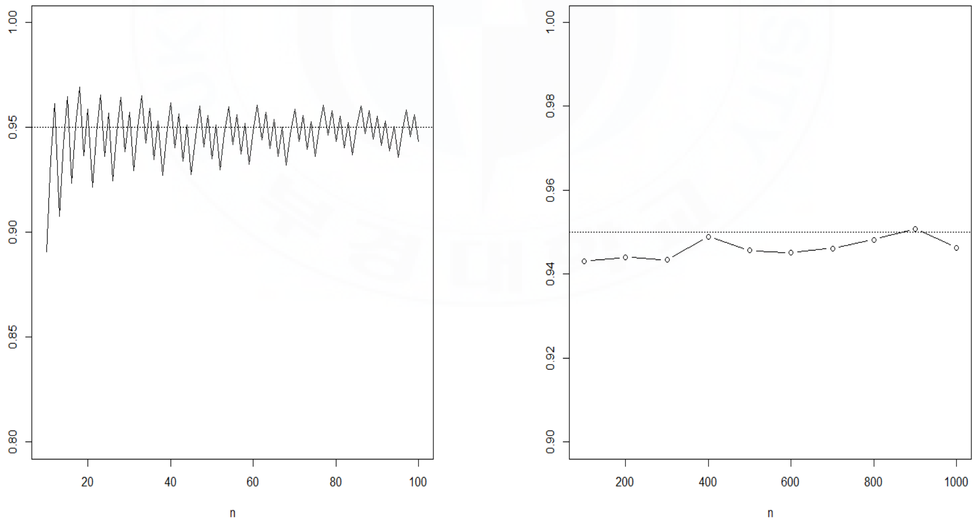
⑤ n quite large;

⑥ $n \geq 50$ unless p is very small.

However they give cautions that the above prescriptions are still defective, saying the first two conditions are not verifiable in the estimation problem, the condition ⑤ useless, ⑥ obviously misleading. Even though we can verify the conditions ③ and ④, they say the two conditions are also useless because a data-driven method is not meaningful in frequentist coverage probabilities. From this results we feel it too dangerous to use the standard Wald interval. In the next sections we review better alternatives.

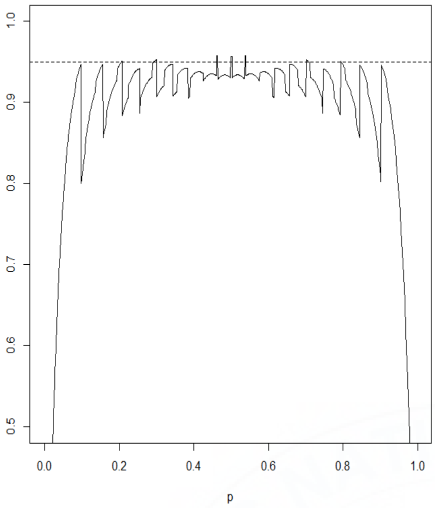


(a) Wald's interval with z critical value

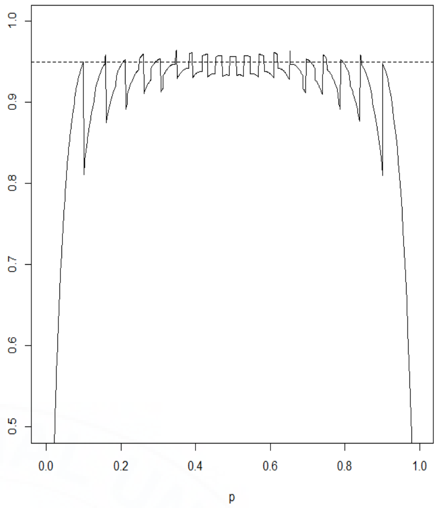


(b) Wald's interval with t critical value

Figure 2.7. Coverage probability with varying n from 10 to 1000 for $p=0.5$.

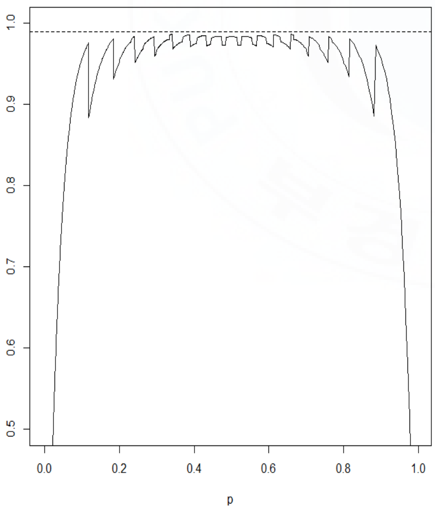


With z critical value

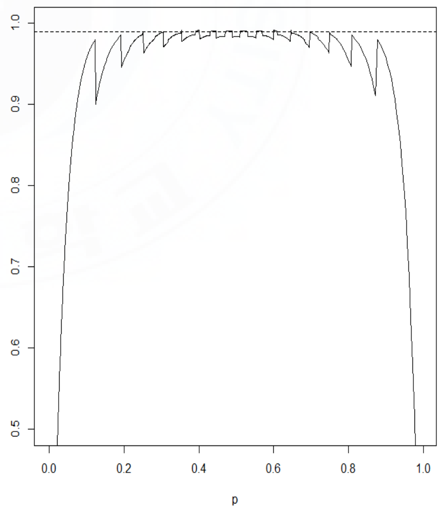


With t critical value

(a) Coverage probability with 95% level for $n = 30$ and $0 < p < 1$



With z critical value



With t critical value

(b) Coverage probability with 99% level for $n = 30$ and $0 < p < 1$

Figure 2.8. Coverage probability with varying n over 10~1000 for $p = 0.5$.

2.4 The Score Interval

The Score method is based on inverting the approximately normal test called the Score test that uses the null $H_0 : p = p_0$, rather than estimated, standard error. We thus accept the null hypothesis if and only if the confidence interval contains the null parameter. The end points of the confidence intervals are the ones p_0 that solves the equations

$$\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \pm z_{\alpha/2}.$$

We have the resulting confidence interval as follows:

$$\frac{1}{(1+z_{\alpha/2}^2/n)}\left[\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n}\right) \pm z_{\alpha/2}\sqrt{\frac{\{\hat{p}(1-\hat{p}) + z_{\alpha/2}^2/4n\}}{n}}\right].$$

This confidence interval was first discussed by Edwin B. Wilson (1927) and is called as the Wilson interval in honor of him. It is sometimes termed the score confidence interval to highlight the inversion of the Score tests. Some call the interval the Wilson's score interval combining the previous two facts.

The midpoint of the Wilson's score interval look embarrassing at first, while the center of Wald's interval $\hat{p}$ is the maximum likelihood estimator. However we can figure out that the center of the score interval lies between $\hat{p}$ and 1/2 by expressing it as a weighted average of them:

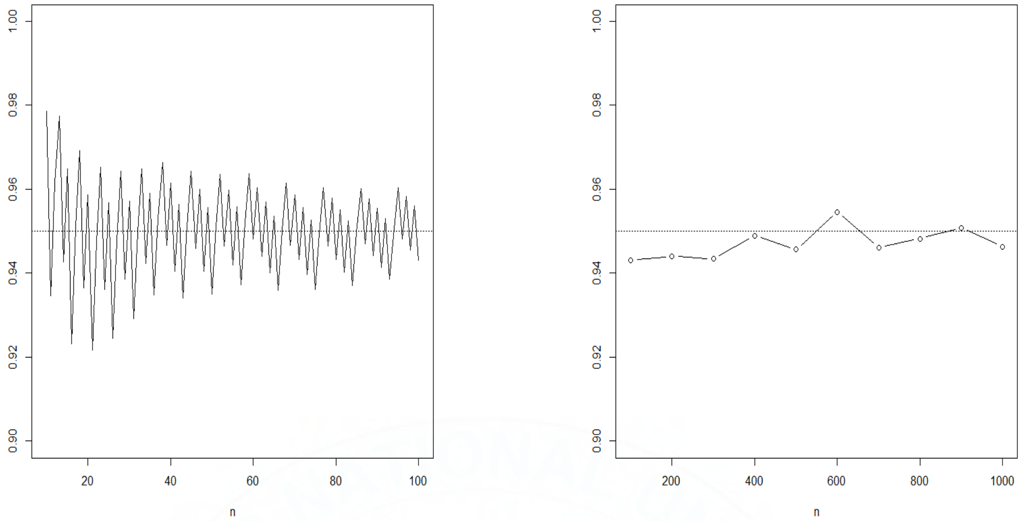
$$\frac{1}{(1+z_{\alpha/2}^2/n)}\left(\hat{p} + \frac{z_{\alpha/2}^2}{2n}\right) = \hat{p}\left(\frac{n}{n+z_{\alpha/2}^2}\right) + \frac{1}{2}\left(\frac{z_{\alpha/2}^2}{n+z_{\alpha/2}^2}\right).$$

This center point shrinks $\hat{p}$ toward 1/2 and the amount of shrinkage gets smaller as n increases.

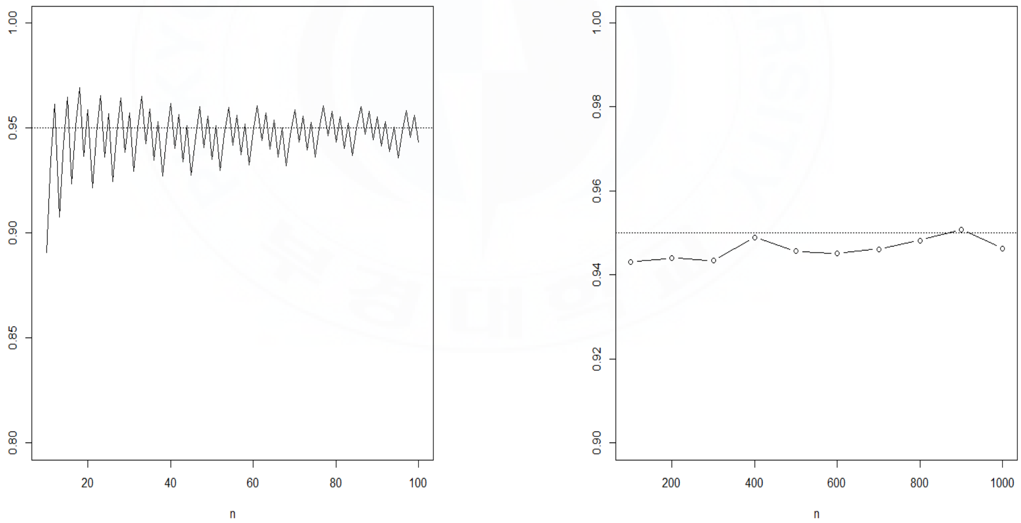
The square of the coefficient of $z_{\alpha/2}$ in the term added to and subtracted from the centered value is also of a weighted average of the variance of a sample proportion with $p=\hat{p}$ and that with $p=1/2$, replacing n with $n+z_{\alpha/2}^2$:

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

Figure 2.9 and 2.10 show the behavioral characteristics of coverage probabilities of the score interval comparing with the Wald interval with z and t critical values. The coverage probability for fixed p with varying n shows still oscillating phenomenon. However for fixed n with varying p, the systematic bias problem present in the Wald interval disappears in the Wilson's score interval.

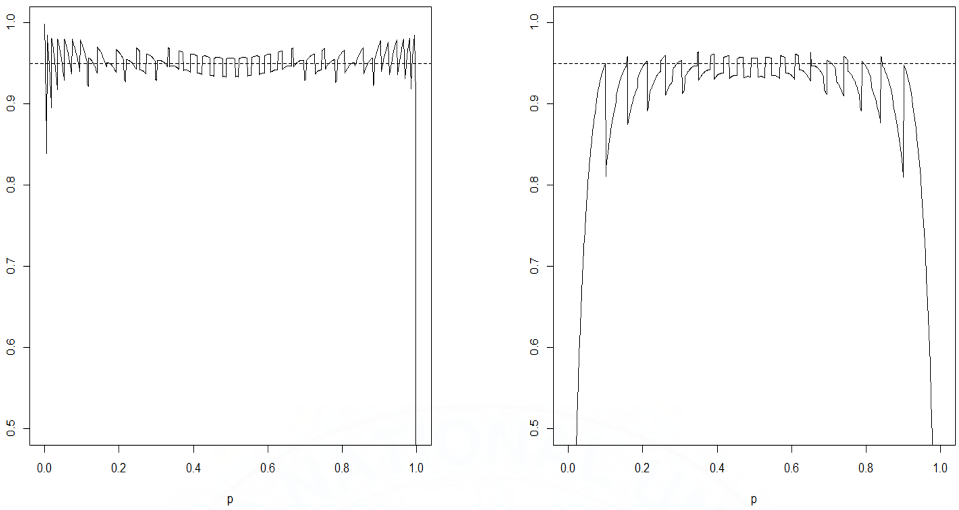


(a) Coverage Probability of Wilson's Score Interval



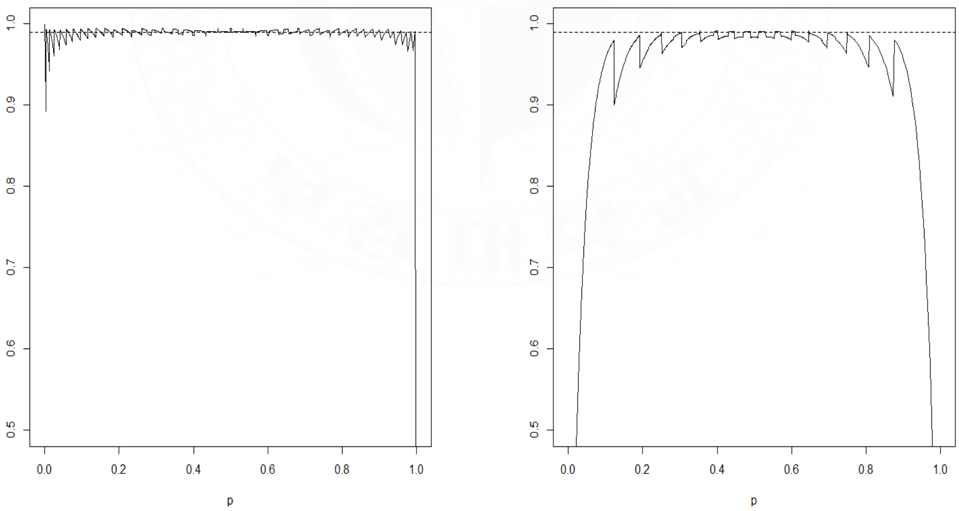
(b) Coverage Probability of Wald's Interval

Figure 2.9. Behavioral characteristics of coverage probabilities for $p = 0.5$ and $n = 10 - 1000$.



Wilson's Interval　　　　Wald's Interval with t Critical Value

(a) Coverage Probability of the Nominal 95% Confidence Intervals



Wilson's Interval　　　　Wald's Interval with t Critical Value

(b) Coverage Probability of the Nominal 99% Confidence Intervals

Figure 2.10. Behavioral characteristics of coverage probabilities for $n=30$ and $0<p<1$.

2.5 The Adjusted Wald Interval

The standard Wald interval is simple and easy to memorize, but it has poor performance. There are many instructors who recommend the score interval instead of the Wald interval. However the formula is considered to be awkward to use in elementary statistics courses. As a compromise Agresti & Coull (1998) suggests a similar form to the Wald interval with a better estimate of p than $\hat{p}=X/n$. They have an idea from the center value of the score interval and define the confidence interval for p by

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}},$$

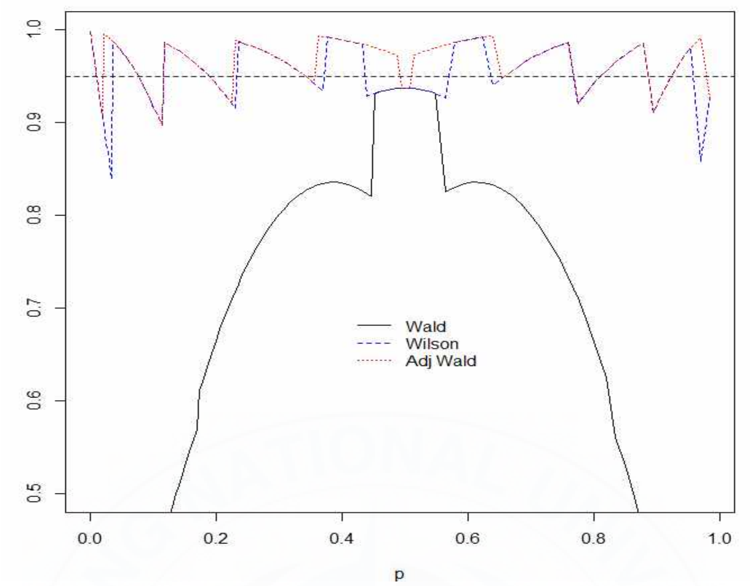
where $\tilde{p}=\tilde{X}/\tilde{n}$ with $\tilde{X}=X+z_{\alpha/2}^2$ and $\tilde{n}=n+z_{\alpha/2}^2$.

When constructing 95% confidence interval, we have $z_{\alpha/2}=1.96$ and $z^2\approx 4$. The center value of the score interval becomes

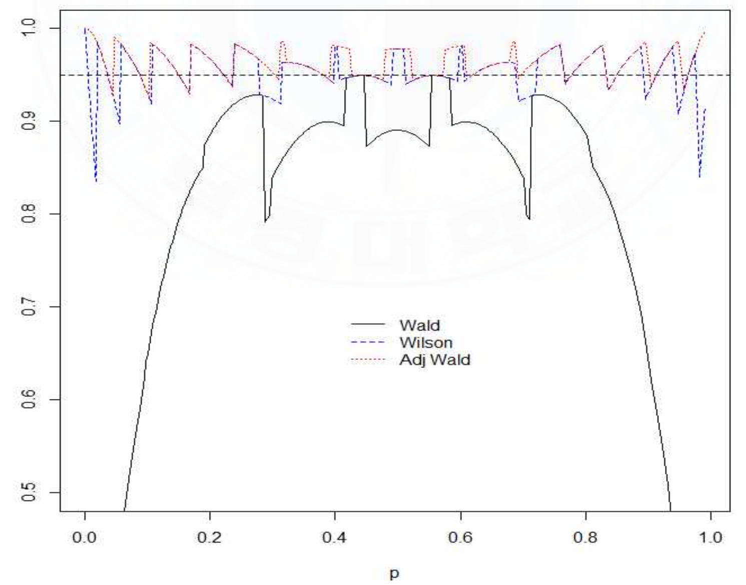
$$\tilde{p}=\frac{X+z^2/2}{n+z^2}\approx\frac{X+2}{n+4}.$$

This interval can be regarded such a version of the Wald formula as adding two successes and two failures with $n+4$ trials, which yields a point estimate $\tilde{p}=(X+2)/(n+4)$. Because of its derivation background, the interval is called an adjusted Wald interval or the Agresti-Coull interval.

Figure 2.11 shows the coverage probabilities of the adjusted Wald interval, compared with those of the Wilson's score and the Wald intervals for different n.



(a) $n=5$



(b) $n=10$

Figure 2.11. Comparison of coverage probabilities for the nominal 95% intervals.

Ⅲ. Bayesian Confidence Intervals

3.1 Bayesian Point Estimator

As an estimate of the binomial proportion p, a statistic $\hat{p}$ is computed from the observed data. We can consider the posterior mean, posterior median, or posterior mode as suitable Bayesian point estimates of p.

Suppose we use the posterior mean square (PMS) as a measure of goodness of an estimator. Then the PMS of an estimator $\hat{p}$ of the proportion p is defined as

$$\begin{aligned} \text{PMS}(\hat{p}) &= \int_0^1 (p-\hat{p})^2 g(p|x)\,dp \\ &= \int_0^1 (p-\tilde{p}+\tilde{p}+\hat{p})^2 g(p|x)\,dp \\ &= \int_0^1 (p-\tilde{p})^2 g(p|x)\,dp + \int_0^1 (\tilde{p}-\hat{p})^2 g(p|x)\,dp \\ &= Var(p|x) + (\tilde{p}-\hat{p})^2, \end{aligned}$$

where $\tilde{p} = \int_0^1 p\,g(p|x)\,dp$, i.e., the posterior mean.

Since the last term, the squared distance of the true value from the posterior mean, is always non-negative, the posterior mean is the optimum estimator of the binomial proportion. It gives us a reasonable cause of the posterior mean as the estimate and an explanation of the posterior mean being the most widely used Bayesian estimate.

3.2 Bayesian Confidence Intervals

In Bayesian world, we would like to identify intervals of the parameter space which seem to contain the true value of the parameter. After observing data, we can construct an interval $(p_L(x),\, p_U(x))$ in a way that the interval has high probability of containing the true parameter.

The coverage probability of Bayesian approach can be defined by using the posterior probability. An interval $(p_L(x),p_U(x))$ with endpoints computed after observing data x, has coverage probability of $(1-\alpha)$ for p if the posterior probability satisfies

$$P(p \in (p_L(x),\, p_U(x)) \mid X=x) = 1-\alpha.$$

While the frequentist interprets the coverage probability of the interval as the probability of covering the true value before the data are observed, Bayesian describes it as the probability of locating the true value of p in the interval after observing the data. These intervals are often called credible intervals, to be distinct from frequentist confidence intervals. We consider a few main conventions for choosing two endpoints of the interval.

3.2.1 Equal-Tailed Interval

The simplest and easiest method to construct a Bayesian confidence interval is to find quantiles of posterior probability density with equal tails. A $100(1-\alpha)\%$ confidence intervals is obtained by finding two numbers $q_{\alpha/2}$ and $q_{1-\alpha/2}$ of posterior distribution satisfying

$$P(p > q_{\alpha/2} \mid X = x) = \alpha/2 \text{ and } P(p < q_{1-\alpha/2} \mid X = x) = \alpha/2,$$

that is, the $1-\alpha/2$ and $\alpha/2$ posterior quantiles respectively.

This implies $P(q_{1-\alpha/2} < p < q_{\alpha/2} \mid X = x) = 1-\alpha$. Thus the desired $100(1-\alpha)\%$ equal-tailed Bayesian credible interval of p is $(q_{1-\alpha/2}, q_{\alpha/2})$.

3.2.2 Highest Posterior Density Interval

When we obtain a credible interval in terms of equal-tailed property, there exist some points outside the interval with higher probability than some points inside the interval. This suggests the existence of the shortest possible interval $H(x)$ covering $100(1-\alpha)\%$ of the posterior mass or density as follows:

① $P(p \in H(X) \mid X = x) = 1-\alpha$

② If $p_1 \in H(x)$, and $p_2 \notin H(x)$, then $P(p_1 \mid X = x) > P(p_2 \mid X = x)$.

All points in $H(x)$ have higher posterior probability than points outside it. The interval $H(x)$ is termed the highest posterior density (HPD) confidence interval. Unfortunately, an HPD interval can not be one interval if the posterior density has more than one peaks.

3.2.3 Approximate Interval

When the prior distribution of the binomial proportion p follows a U$(0,1)$, i.e, beta$(1, 1)$, the posterior distribution of p after observing x successes is beta $(x+1, n-x+1)$. The posterior density has the form $f(p|x) \propto p^x (1-p)^{n-x}$, and its log transformation yields $L(p) = \log f(p|x) = k + x\log p + (n-x)\log(1-p)$.

Taking partial derivatives we can obtain maximum likelihood estimator of p as follows:

$$\frac{\partial L(p)}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0 \quad \Rightarrow \hat{p}_{MLE} = \frac{x}{n}$$

$$\frac{\partial^2 L(p)}{\partial p^2} = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2} \Rightarrow -\frac{\partial^2 L(p)}{\partial p^2}\bigg|_{p=\hat{p}} = -\frac{n}{\hat{p}(1-\hat{p})}$$

This yields an asymptotic normal distribution about $\hat{p}$:

$$\frac{\sqrt{n}(\hat{p}-p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{d} N(0,1)$$

Applying the approximation we can obtain a $100(1-\alpha)\%$ confidence interval as follows:

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

3.3 Coverage Probability with Priors

3.3.1 Beta prior

When X follows $bin(n,p)$, its standard conjugate priors are beta distributions. Suppose the prior distribution of the binomial proportion p

follows beta(a, b), then the posterior distribution of p for the observed data x is beta$(a+x, b+n-x)$.

The $100(1-\alpha)$% equal-tailed Bayesian credible interval for p is given by

$$\left[q_{1-\alpha/2, a+x, b+n-x},\ q_{\alpha/2, a+x, b+n-x}\right],$$

where $q_{\alpha,m,n}$ denotes the $(1-\alpha)$ quantile of a beta(m,n) distribution. Thus 95% equal-tailed Bayesian confidence interval can be constructed by finding the 2.5th and 97.5th percentiles. In R they are computed by $qbeta(0.025, a+x, b+n-x)$ and $qbeta(0.975, a+x, b+n-x)$ respectively.

To obtain the 95% HPD credible interval we need to find two quantiles with which the interval has the lower and upper endpoints and its outside total area is $1-\alpha$. However it is not easy to compute them mathematically. Thus we approximate the beta$(a+x, b+n-x)$ posterior distribution by the normal distribution with the same mean and variance, i.e,

$$\frac{p-m}{s} \xrightarrow{d} N(0,1),$$

where $m = \dfrac{a+x}{a+b+n}$ and $s^2 = \dfrac{(a+x)(b+n-x)}{(a+b+n)^2(a+b+n+1)}$. The resulting 95% approximate credible interval is given by $(m-1.96\times s,\ m+1.96\times s)$.

3.3.2 Jeffrey Prior

The Jeffrey prior for p is defined in terms of the Fisher information;

$$f_J(p) \propto I(p)^{1/2},$$

where the Fisher information $I(p)$ is given by

$$I(p) = -E_p\left\{\frac{\partial^2 \log f(X|p)}{\partial p^2}\right\}.$$

When $X \sim bin(n,p)$, we have

$$\frac{\partial^2}{\partial p^2} log f(x|p) = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}$$

and $E_p(X) = np$ yields

$$I(p) = -E_p\left\{\frac{\partial^2 \log f(X|p)}{\partial p^2}\right\} = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)}.$$

Therefore $f_J(p) \propto I(p)^{1/2} \propto p^{-1/2}(1-p)^{-1/2}$, which implies the Jeffrey prior is beta$(1/2, 1/2)$ with the density function

$$f_J(p) = \pi^{-1} p^{-1/2}(1-p)^{-1/2} I_{(0,1)}(p).$$

The $100(1-\alpha)$% equal-tailed Bayesian credible interval is defined as

$$I_J = [p_{L,J}(x), p_{U,J}(x)],$$

where $p_{L,J}(x) = \begin{cases} 0 & x = 0 \\ q_{1-\alpha/2, x+1/2, n-x+1/2} & otherwise \end{cases}$

and

$$p_{U,J}(x) = \begin{cases} 1 & x = n \\ q_{\alpha/2, x+1/2, n-x+1/2} & otherwise \end{cases}.$$

For $x = 0$, we have the lower limit 0 to avoid the result that the coverage probability goes to zero as p approaches 0. Similarly by the same reason the upper limit for $x = n$ is also modified to 1.

The endpoints of the credible interval with Jeffrey prior is easily computed using R:

$$p_{L,J}(x) = qbeta(\alpha/2, x+1/2, n-x+1/2)$$

and $p_{U,J}(x) = qbeta(1-\alpha/2, x+1/2, n-x+1/2)$.

Using the HPD intervals instead of equal-tailed approach, we would have a better Bayesian solution with shorter interval for the same confidence level. However the HPD intervals are not easy to compute mathematically and are approximated by normal approximation. With the Jeffrey prior, the approximate $100(1-\alpha)\%$ credible interval is $(m - z_{\alpha/2}s,\ m + z_{\alpha/2}s)$,

where $m = \dfrac{x+1/2}{n+1}$, $s^2 = \dfrac{(x+1/2)(n-x+1/2)}{(n+1)^2(n+2)}$.

The plots in Figure 3.1 display the coverage probabilities of equal-tailed interval and approximate confidence interval with the Jeffrey prior. Even though the approximate intervals are easy to compute, they are biased than the equal-tailed intervals.

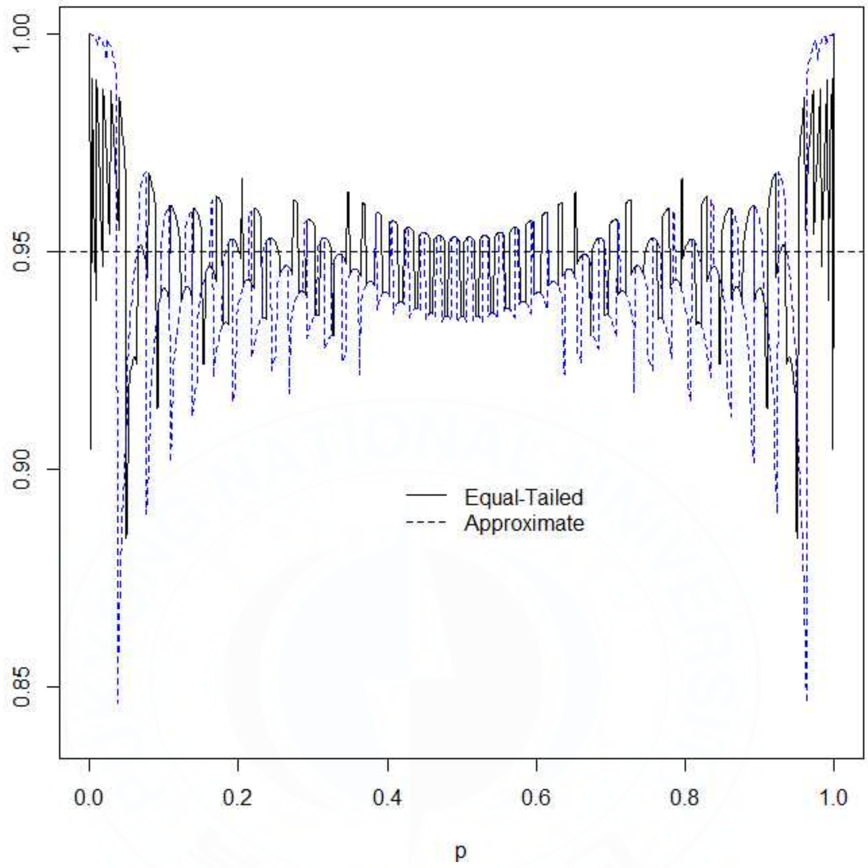


Figure 3.1. Coverage probabilities of equal-tailed and approximate intervals for $n = 50$.

Ⅳ. Measures of Performance

In order to choose good confidence intervals, we need useful measures of the performance for various interval methods. Here we consider such performance criteria as the mean coverage probability, the expected length, the average expected length and the mean absolute error.

4.1 Mean Coverage Probability

For any confidence interval method for estimating p, the actual coverage probability at a fixed value of p is defined

$$C(p,n)=\sum_{x=0}^{n} I(x,p)\binom{n}{x}p^{x}(1-p)^{n-x}$$

where $I(x,p)=1$ if p is in the interval when $X=x$ and equals 0 otherwise. The confidence coefficient is defined as the infimum of coverage probabilities over all possible values of p. The definition is a theoretical concept about the worst possible performance, thus it is more practical to use average performance instead.

As a summary measure of performance the mean coverage probability, obtained by averaging the coverage probabilities over all the possible values of p, is defined as

$$MC(n)=\int_{0}^{1} C(p,n)g(p)dp$$

where $g(p)$ is the density for p.

Table 4.1 and Figure 4.1 show the mean coverage probability with root mean squared error for the exact, score, Wald, adjusted Wald (A&C) intervals, and two Bayesian intervals. Among the considered confidence intervals the mean coverage probability of Wilson's score interval is closest to the nominal level, the Bayesian credible interval with equal-tailed approach is secondly close and the adjusted Wald (A&C) interval is thirdly close to the nominal level 0.95. However these three approaches has almost the same mean coverage probability as the nominal level.

And it is evident that the exact interval has the largest mean coverage probability, the adjusted Wald interval and Bayesian equal-tailed interval are in decreasing order. It implies that the confidence interval by those three methods show conservatism. However the Wald interval and Bayesian interval by normal approximation to posterior density have smaller coverage probability.

Table 4.1. Mean coverage probabilities with root MSE in parenthesis

n	Exact	Score	Wald	Adj. Wald	Equal Tail	Approx Bayes
5	0.989(0.0394)	0.954(0.0248)	0.867(0.1002)	0.975(0.0304)	0.957(0.0075)	0.929(0.0206)
10	0.978(0.0286)	0.952(0.0130)	0.900(0.0577)	0.962(0.0186)	0.953(0.0031)	0.933(0.0165)
15	0.974(0.0252)	0.951(0.0148)	0.916(0.0419)	0.957(0.0154)	0.952(0.0019)	0.936(0.0140)
20	0.971(0.0225)	0.951(0.0127)	0.926(0.0274)	0.955(0.0119)	0.951(0.0011)	0.938(0.0119)
40	0.966(0.0166)	0.951(0.0079)	0.938(0.0155)	0.952(0.0083)	0.950(0.0003)	0.942(0.0079)
50	0.964(0.0155)	0.95(0.0074)	0.940(0.0127)	0.951(0.0073)	0.950(0.0001)	0.943(0.0069)
60	0.963(0.0139)	0.951(0.0056)	0.942(0.0118)	0.951(0.0058)	0.950(0.0001)	0.944(0.0059)
80	0.962(0.0139)	0.95(0.0059)	0.944(0.0800)	0.951(0.0059)	0.950(0.0001)	0.945(0.0048)
100	0.961(0.0115)	0.95(0.0053)	0.945(0.0069)	0.95(0.0053)	0.950(0.0001)	0.946(0.0041)

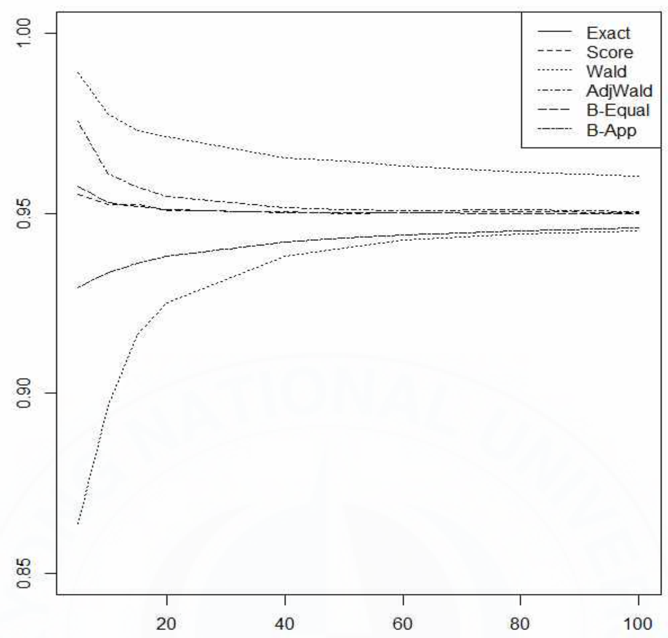


Figure 4.1. Mean coverage probabilities as a function of n.

4.2 Expected Length

The expected length of confidence intervals is defined for a specific p with n fixed in advance as follows:

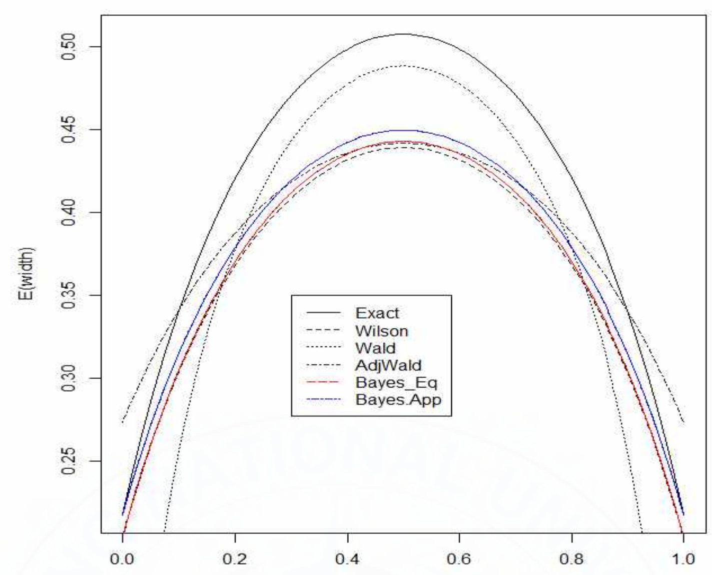
$$E_{n,p}[\text{length(CI)}] = \sum_{k=0}^{n} \left(\hat{p}_{u,\alpha,k} - \hat{p}_{l,\alpha,k}\right) \binom{n}{k} p^k (1-p)^{n-k},$$

where $\hat{p}_{u,\alpha,k}$ and $\hat{p}_{l,\alpha,k}$ are the left and the right end points of the confidence interval, respectively.

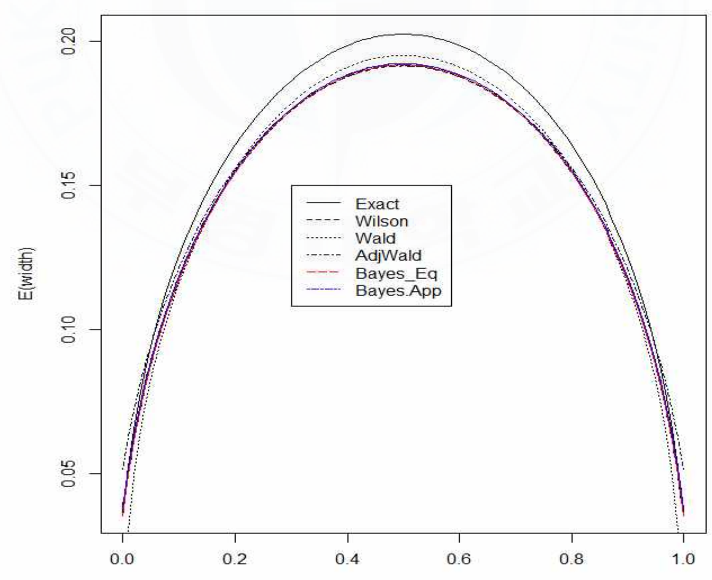
Figure 4.2 shows the expected widths for the nominal 95% exact, score, Wald, A&C and two Bayesian intervals as a function of p for $n=15$ and $n=100$. For small n, all the intervals tend to be shorter as p approaches to boundary regions. The Wald intervals are shorter than the exact intervals over the whole range of p and the Score intervals get much shorter than exact intervals. The fact that the length of the Wald intervals goes to zero as p approaches 0 or 1 implies that its interval get degenerated at $x=0$ or n. The intervals by the two Bayesian methods with equal-tailed and approximation has shorter length except for Wislon's. When $x=0$, the Wilson's score interval is $[0, 1.96^2/(n+1.96^2)]=[0,\ 0.2034]$, the exact interval $[0, 1-(0.025)^{1/n}]\approx[0,-\log(0.025)/n]=[0, 3.69/n]=[0, 0.246]$ and the adjusted Wald interval (A&C interval) is

$$\frac{z^2}{2(n+z^2)}\left(1\pm\sqrt{\frac{2n+z^2}{n+z^2}}\right)=[-0.034,\ 0.2385].$$

With large $n=100$, the exact interval is wider than the other intervals except for the boundary regions where the A&C interval is a little wider than the other intervals, though negligible. The Bayesian method yields the shortest intervals.



(a) $n = 15$



(b) $n = 100$

Figure 4.2. Comparison of the expected lengths for $n = 10$ to 100.

4.3 Average Expected Length

Suppose p follows the distribution $g(p)$. The average expected length is then defined over the whole interval $[0,1]$ of p as follows:

$$\begin{aligned} &\text{Average Expected Length} \\ &= \int_0^1 E_{n,p}[\text{length(CI)}]\, g(p)\, dp \\ &= \int_0^1 \sum_{x=0}^{n} \left(\hat{p}_{u,\alpha,x} - \hat{p}_{l,\alpha,x}\right) \binom{n}{x} p^x (1-p)^{n-x} dp, \end{aligned}$$

where $\hat{p}_{u,\alpha,x}$ and $\hat{p}_{l,\alpha,x}$ are the left and the right end points of the confidence interval, respectively.

There are in Figure 4.3 the average expected lengths of the exact, score, Wald, adjusted Wald and two Bayesian intervals for the uniform distribution of p with varying n from 10 to 100. It is clear from the plot that among the six intervals the Wald interval is the shortest, and the score interval, Bayesian with equal-tailed, Bayesian with approximation, the adjusted Wald interval and the exact interval are in the narrowness order. The tendency does not change for larger n.

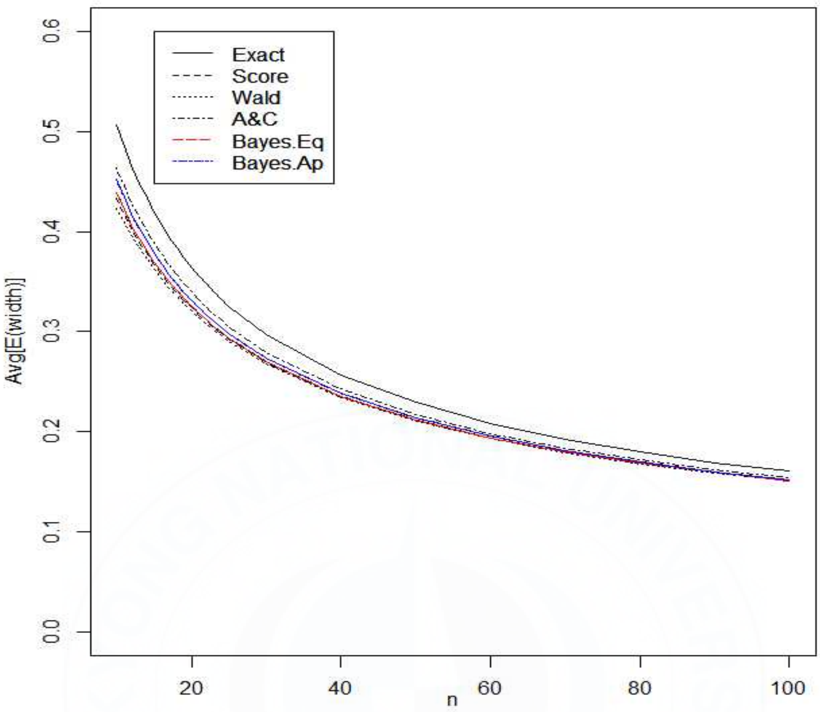


Figure 4.3. Comparison of the average expected lengths for $n=10$ to 100.

4.4 Mean Absolute Error

The last criterion we use for comparison of the alternative intervals is the mean absolute error defined as the mean of the absolute difference between the actual coverage probability and the nominal confidence level:

$$\text{Mean Absolute Error (MAE)} = \int_0^1 |C(p,n) - (1-\alpha)| g(p) dp,$$

where $C(p,n)$ is the actual coverage probability for p with distribution $g(p)$.

There are MAEs in Figure 4.4 for the exact, score, Wald, adjusted Wald, equal-tailed and approximate Bayesian intervals with the uniform distribution of p with varying n from 10 to 100. It is clear from the plot that among the four intervals the Wald and equal-tailed Bayesian interval have significantly largest errors, while the adjusted Wald interval has the smallest mean absolute errors over all the considered range of n.

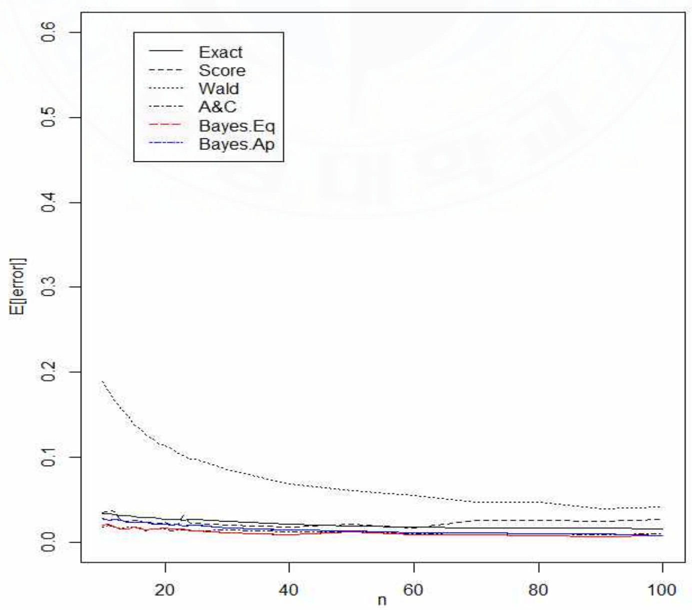


Figure 4.4. The mean absolute errors of the coverages for $n = 10 \sim 100$.

V. Conclusions

We have reviewed major existing confidence intervals of the binomial proportion parameter centered on their coverage probabilities. Since the discrete nature of the random variable, it is impossible to obtain confidence interval with the exact coverage probability. Every method studied here has its own pros and cons.

The Clopper-Pearson interval (1934) is exact for all n. The endpoints of the interval are computed using the quantiles of a beta distribution. However this interval yields that the actual coverage probability is always not less than nominal level $1-\alpha$, thus it keeps conservatism.

The Wald interval is considered standard in many elementary textbooks, due to the convenience of derivation and computation. But it shows too abnormal behavior of the coverage probability, and performed poorly especially at p close to zero or one.

The Wilson's score interval (1927) is based on inverting the score test and uses the standard error of null parameter, not its estimate. It has theoretical attraction, but is not easy to memorize the end points.

With combining the simplicity of the Wald interval and theoretical appeal of the Wilson's interval, Agresti and Coull suggest an adjusted Wald interval. Like the Wald interval, the adjusted interval has simple formula of new n and p with the property of extra addition of successes and failures.

As Bayesian intervals we reviewed the equal-tailed interval using quantiles of posterior probability density and the HPD intervals. But the HPD interval is not easy to compute, it is approximated by normal distribution.

We evaluated each method using a few performance criteria, which results in the preference of the Wilson's score, A&C and Bayesian equal-tailed intervals.

This study is not a new trial but a look-back attitude to understand their derivations and coverage probabilities. For further study topics we will make efforts to find out erratic behaviors of coverage probabilities of each confidence interval with theoretical approach.

References

Agresti, A., and Coull, B. (1998), “Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions,” *The American Statistician*, 52, 119-126.

Agresti, A., and Caffo, B. (2000), “Simple and Effective Confidence Intervals for Proportions and Differences of Proportions Result from Adding Two Successes and Two Failures,” *The American Statistician*, 54, 280-288.

Agresti, A., and Hitchcock, D. B. (2005), “Bayesian inference for categorical data analysis: A survey,” *Statistical Methods & Applications*, 14, 297-330.

Bolstad, W. M. (2004). *Introduction to Bayesian Statistics*, John Wiley & Sons, Inc..

Böhning, D. (1994), “Better Approximate Confidence Intervals for a Binomial Parameter,” *Canadian Journal of Statistics*, 22, 207-218.

Blyth, C. R., and Still, H. A. (1983), “Binomial Confidence Intervals,” *Journal of the American Statistical Association*, 78, 108-116.

Brown, L. D., Cai, T. T. and DasGupta, A. (2001), "Interval Estimation for a Binomial Proportion (with discussion)", *Statistical Science*, 16, 101-133.

Carlin, B. P., and Louis, T. A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.

Chen H. (1990), “The Accuracy of Approximate Intervals for a Binomial Parameter,” *Journal of the American Statistical Association*, 85, 514-516.

Clopper, C. J., and Pearson, E. S. (1934), “The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial,” *Biometrika*, 26, 404-413.

Ghosh, B. K. (1979), “A Comparison of Some Approximate Confidence Intervals for the Binomial Parameter,” *Journal of the American Statistical Association*, 74, 894-900.

Hogg, R. V., McKean, J. W. and Craig, A. T. (2013), *Introduction to Mathematical Statistics*, Pearson: Boston.

Jovanovic, B. D., and Levy, P. S. (1997), “A Look at the Rule of Three,” *The American Statistician*, 51, 137-139.

Laplace, P. S. (1812), *Theorie Analytique des Probabilites*, Paris: Courcier.

Leemis, L. M., and Trivedi, K. S. (1996), “A Comparison of Approximate Interval Estimators for the Bernoulli Parameter,” *The American Statistician*, 50, 63-68.

Lindley, D. V. (1965), *Introduction to Probability and Statistics from a Bayesian Viewpoint*. Part 2. Inference. Cambridge University Press: Cambridge, United Kingdom.

Neyman, J. (1934), "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection", *Journal of the Royal Statistical Society*, 97, 558-625.

Neyman, J. (1935), "On the problem of confidence intervals", *The Annals of Mathematical Statistics*, 6, 111-116.

Newcombe, R. G. (1994), “Confidence Intervals for a Binomial Proportion,” *Statistics in Medicine*, 13, 1283-1285.

Pires, A. M., and Amado, C. (2008), “Interval Estimation for a Binomial Proportion: Comparison of Twenty Methods,” *REVSTAT-Statistical Journal*, 6, 165-197.

Samuels, M. L., and Witmer, J. W. (1999), *Statistics for the Life Sciences*, 2nd ed., Prentice Hall.

Santner, T. J. (1998), "A Note on Teaching Binomial Confidence Intervals," *Teaching Statistics*, 20, 20-23.

Santner, T. J., and Duffy, D. E. (1989), *The Statistical Analysis of Discrete Data*, Berlin: Springer-Verlag.

Seber G. A. F. (2013), *Statistical Models for Proportions and Probabilities*, Springer, New York, USA.

Vollset, S. E. (1993), "Confidence Intervals for a Binomial Proportion," *Statistics in Medicine*, 12, 809-824.

Wilson, E. B. (1927), "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, 22, 209-212.

Appendix: R-Codes

```
#--------------------------------------------------------------------------------------------------
# Figure 2.4
# Coverage Probability Behavior of Upper Bound
# The related hypothesis: Ha: p<p0
# --------------------------------------------------------------------------------------------------
Pup.cvg=function(n,alpha,p){
          # n=10; alpha=0.05;x=2
          # p=qbeta(1-alpha,x+1,n-x)
 U=matrix(0,n+1,4)
 x=0:n
 U[,1]=x
 U[,2]=qbeta(1-alpha,x+1,n-x)   # Upper bound for [0, P(alpha, x)) x=0,...,n
                                  # When x=0,...,n-1
                                  #    =P_U(1-alpha,x,n)=p*(x) is such that
                                  #    p-value=P(X<=x|p*(x))=alpha
                                  # When x=n
                                  #    P(X<=n)=1 for all p
                                  #    so let p*(n)=1
 U[,3]=(U[,2]>p) # For given x decide whether x belongs to CI i.e.
                 # p in [0,P(alpha,x)) i.e., p <p*(x), for x=0,...,n-1
 U[n+1,3]=1      # Since p*(n)=1, when x=n CI=[0,1], coverage prob=1
 U[,4]=dbinom(x,n,p)
 B=sum(U[,4][U[,3]==1])  # The coverage probability for the given p
                          # Pr{p in [0,P_U(X))}=Pr{k: p<P_U(k)}
                    # = p^n +sum_{k=0}^{n-1}P(X=k)I_{qbeta(1-a,k+1,n-k)>p}(k)
}

#------------------------------------------
# Check p*(i)=P(X<=i|pi)=alpha;
#              for i=0,...,n-1
# -----------------------------------------
alpha=0.05
B3=rep(0,n)
 for (i in 0:(n-1)){
    B3[i+1]=pbinom(i,n,qbeta(1-alpha,i+1,n-i))
}
```

```
#------------------------------------------------------------------------
# Compute the coverage probability for any p in [0,1]
#------------------------------------------------------------------------
n=10; alpha=0.05;
Np=50;                              # the no. of p's in [p*(k-1),p*(k)); k=1,...,n
whole.p=rep(1,Np*(n+1)+1)           # p[last]=p*(n)=1
whole.cvrg=rep(1,Np*(n+1)+1)        # coverage prob for p*(n)=1

for (k in 0:n){
 pstart=qbeta(1-alpha,k,n-k+1);   pend=qbeta(1-alpha,k+1,n-k)
                  # p*(n)=1; p*(i)=qbeta(1-alpha,i+1,n-i) i=0,...,n-1
                  # k=0 corresponds p=0 ie p in [0,p*(0))
 p=seq(pstart,pend,length=Np+1); p=p[-(Np+1)]
                  # Choose Np of p's in the interval [P_U(k-1), P_U(k))
                  # to calculate Coverage Prob P{p in C(X)}
 whole.p[(k*Np+1):((k+1)*Np)]=p
 for (i in 1:Np) whole.cvrg[k*Np+i]=Pup.cvg(n,alpha,p[i])
}

#---------------------------------------------
# Plot of Coverage probability
# --------------------------------------------
plot(whole.p, whole.cvrg, type="l", xlab="",ylab="")
      # ylab=expression(paste("Probablity of Coverage   ",hat(p)[U])))
abline(h=1-alpha,lty=2)

# give xlabel and ylabel using mtext
title(line=1.5,expression(
  paste("Coverage Probability Behavior of Upper Bound   ", hat(p)[U])))
mtext(expression(paste("Probablity of Coverage   ",hat(p)[U])),side=2,line=2)
mtext("p",side=1,line=2)

a1=paste("nominal confidence level=",1-alpha,"\n sample size n=",n)
text(0.5,1.0,pos=1,a1)
```

```
#---------------------------------------------------------------------------------------------------
# Figure 2.5
# Coverage Probability Behavior of Lower Bound
# The related hypothesis: Ha: p>p0
# --------------------------------------------------------------------------------------------------

Plow.cvg=function(n,alpha,p){
        #   n=10; alpha=0.05; x=n
        #   p=qbeta(alpha,x,n-x+1)
 L=matrix(0,n+1,4)
 x=0:n
 L[,1]=x
 L[,2]=qbeta(alpha,x,n-x+1)  # Lower bound for (P_L(x),1] x=0,...,n
                                # When x=1,...,n
                                #    =P_L(alpha,x,n)=p*(x) is such that
                                #    p-value=P(X>=x|p*(x))=alpha
                                # When x=0
                                #    p-value=P(X>=0|p)=1>alpha for all p
                                #    so let p*(0)=0

 L[,3]=(L[,2]<p) # For given x decide whether x belongs to CI i.e.
                   # p in (P_L(x),1] i.e., p>p*(x), for x=1,...,n
 L[1,3]=1          # Since p*(0)=0, when x=0 CI=[0,1], coverage prob=1
 L[,4]=dbinom(x,n,p)
 B=sum(L[,4][L[,3]==1])  # The coverage probability for the given p
                            # Pr{p in (P_L(X),1]}=Pr{k: p>P_L(k)}
                         # = (1-p)^n + sum_{k=1}^{n}P(X=k)I_{qbeta(a,k,n-k+1)>p}(k)
}

#--------------------------------------------------
# Check p*(i)=P(X>=i|pi)=alpha;
#                for i=1,...,n
# -------------------------------------------------
 alpha=0.05
 B3=rep(0,n)
 for (i in 1:n){
    B3[i]=1-pbinom(i-1,n,qbeta(alpha,i,n-i+1))}
```

```
#-------------------------------------------------------------------
# Compute Coverage probability for any p in [0,1]
# ------------------------------------------------------------------
n=100; alpha=0.05;
Np=30;                              # the no. of p's in (p*(k-1),p*(k)]; k=1,...,n
whole.p=rep(0,Np*(n+1)+1)        # p[1st]=p*(0)=0
whole.cvrg=rep(1,Np*(n+1)+1)   # coverage prob for p*(0)=1

for (k in 0:n){
 pstart=qbeta(alpha,k,n-k+1);   pend=qbeta(alpha,k+1,n-k)
                    # p*(i)=qbeta(alpha,i,n-i+1) i=0,...,n
                    # k=n corresponds to p in (p*(n),1]
 p=seq(pstart,pend,length=Np+1); p=p[-1]
                    # Choose Np of p's in the interval (P_U(k-1), P_U(k)]
                    # to calculate Coverage Prob P{p in C(X)}
 whole.p[(k*Np+2):((k+1)*Np+1)]=p
 for (i in 1:Np) whole.cvrg[k*Np+1+i]=Plow.cvg(n,alpha,p[i])
}

#-----------------------------------------
# Plot of Coverage probability
# ---------------------------------------
plot(whole.p, whole.cvrg, type="l", xlab="",ylab="")
    # ylab=expression(paste("Probablity of Coverage  ",hat(p)[L])))
abline(h=1-alpha,lty=2)

# give xlabel and ylabel using mtext
title(line=1.5,expression(
  paste("Coverage Probability Behavior of Lower Bound  ", hat(p)[L])))
mtext(expression(paste("Probablity of Coverage  ",hat(p)[L])),side=2,line=2)
mtext("p",side=1,line=2)

a1=paste("nominal confidence level=",1-alpha,"\n sample size n=",n)
text(0.5,1.0,pos=1,a1)
```

```
#-----------------------------------------------------------------------------------------------------------
# Figure 2.6
# Coverage Probability Behavior of Confidence Interval
# Ha: p ne p0
# i.e., both upper and lower bound
# ---------------------------------------------------------------------------------------------------------

# ------------------------------------------------------------------------
# Check P_L(alpha,x,n)<P_U(alpha,x,n) for all x=0,...,n
#          P_L(alpha,x,n)=qbeta(alpha,x,n-x+1)
#          P_U(alpha,x,n)=qbeta(1-alpha,x+1,n-x)
# ------------------------------------------------------------------------
 n=10; x=0:n; alpha=0.05
 a=qbeta(alpha,x,n-x+1);   b=qbeta(1-alpha,x+1,n-x)
 A=cbind(a,b,rep(0,11))
 A[,3]=a<b

Pboth.cvg=function(n,alpha,p){
     # n=11; alpha=0.05; x=1
     # p=qbeta(alpha,x,n-x+1)
 B=matrix(0,n+1,5)
 x=0:n
 B[,1]=x
 B[,2]=qbeta(alpha/2,x,n-x+1)       # P_L(alpha/2,x,n)
 B[,3]=qbeta(1-alpha/2,x+1,n-x)   # P_U(alpha/2,x,n)
 B[,4]=(B[,2]<p)&(B[,3]>p)
 if (qbeta(alpha/2,n,1)<p) B[n+1,4]=1
 if (qbeta(1-alpha/2,1,n)>p) B[1,4]=1
 B[,5]=dbinom(x,n,p)
 Bsum=sum(B[,5][B[,4]==1])  # The coverage probability for the given p
                                # Pr{p in (P_L(X),1]}=Pr{k: p>P_L(k)}
                          # = (1-p)^n + sum_{k=1}^{n}P(X=k)I_{qbeta(a,k,n-k+1)>p}(k)
}

#------------------------------------------------------------
# Check p*(i)=P(X>=i|pi)=alpha; for i=1,...,n
# ------------------------------------------------------------
 B3=rep(0,n)
 for (i in 1:n){
     B3[i]=1-pbinom(i-1,n,qbeta(alpha,i,n-i+1))
}
#------------------------------------------------------------------------
```

```
# Compute Coverage probability for any p in [0,1]
# -------------------------------------------------------------------
Coverage=function(n,alpha,Np){
            # n=no. of trials
            # alpha=0.05
            # Np=no. of p's in (p*(k-1),p*(k)]; k=1,...,n
   whole.p=rep(0,Np*(n+1)+1)        # p[1st]=p*(0)=0
   whole.cvrg=rep(1,Np*(n+1)+1)   # coverage prob for p*(0)=1

   for (k in 0:n){
       pstart=qbeta(alpha,k,n-k+1);   pend=qbeta(alpha,k+1,n-k)
                      # p*(i)=qbeta(alpha,i,n-i+1) i=0,...,n
                      # k=n corresponds to p in (p*(n),1]
       p=seq(pstart,pend,length=Np+1); p=p[-1]
                      # Choose Np of p's in the interval (P_U(k-1), P_U(k)]
                      # to calculate Coverage Prob P{p in C(X)}
       whole.p[(k*Np+2):((k+1)*Np+1)]=p
       for (i in 1:Np) whole.cvrg[k*Np+1+i]=Pboth.cvg(n,alpha,p[i])
    }
   return(list(p=whole.p, cov.prg=whole.cvrg))
}

#-------------------------------------------
# Plot of Coverage probability
# -----------------------------------------
n=11; alpha=0.05; Np=30
A=Coverage(n,alpha,Np)
whole.p=A$p; whole.cvrg=A$cov.prg
plot(whole.p, whole.cvrg, type="l", xlab="",ylab="", ylim=c(0.95,1.00))
abline(h=1-alpha,lty=2)

# give xlabel and ylabel using mtext
title(line=1.5,expression(
   paste("Coverage Probability of Confidence Intervals [", hat(p)[L]," , ", hat(p)[U],"]")))
mtext(expression(paste("Probablity of Coverage for [ ", hat(p)[L]," , ", hat(p)[U],"
        ]")),side=2,line=2)
mtext("p",side=1,line=2)

a1=paste("nominal confidence level=",1-alpha,"\n sample size n=",n)
text(0.5,1.0,pos=1,a1)
```

```
# ----------------------------------------------------------------------------------------------
# Figure 2.8-11
# Comparison of Coverage Probabilities for
# the Nominal 95% Intervals
# ----------------------------------------------------------------------------------------------
# Coverage probability of
#   1) Wilson's Score
#   2) Wald
#   3) Wald with t-critical
#   4) Mid-P
#   5) Continuity-corrected Score
# -------------------------------------------------------------
# n=5; alpha=0.05; p=0.05
Cvgprob.fixedP=function(n,alpha,p){
     # Coverage Probability at a fixed value of p
  x=0:n
  px=dbinom(x,n,p)
  z.crt=qnorm(1-alpha/2)
  phat=x/n

  xtilde=x+z.crt^2/2
  ntilde=n+z.crt^2
  ptilde=xtilde/ntilde
  den=1+z.crt^2/n
  wgt=n/(n+z.crt^2)
  mid.point=phat*wgt+0.5*(1-wgt)
  z.coef=phat*(1-phat)*wgt+(1/4)*(1-wgt)

  ws.width=sqrt(z.coef/(n+z.crt^2))
  wd.width=sqrt(phat*(1-phat)/n)
  adjwd.width=sqrt(ptilde*(1-ptilde)/ntilde)

  WS=WDz=WDt=AWD=matrix(0,n+1,5)
  colnames(WS)=colnames(WDz)=colnames(WDt)=colnames(AWD)=c("k",
  "L_CI","U_CI",paste("p=",p,sep=""),"P(X=k)")


 # Wilson Score CI
 WS[,1]=x
 WS[,2]=mid.point-z.crt*ws.width
 WS[,3]=mid.point+z.crt*ws.width
 WS[,4]=(WS[,2]<p)&(WS[,3]>p)
```

```
  WS[,5]=px
  Cvgprob.WS=sum(WS[,5][WS[,4]==1])   # coverage prob of Wilson's Score CI

 # Wald CI with z-critical
  WDz[,1]=x
  WDz[,2]=phat-z.crt*wd.width
  WDz[,3]=phat+z.crt*wd.width
  WDz[,4]=(WDz[,2]<p)&(WDz[,3]>p)
  WDz[,5]=px
  Cvgprob.WDz=sum(WDz[,5][WDz[,4]==1])   # coverage prob of Wald's CI

 # Wald CI with t-critical
  t.crt=qt(1-alpha/2,n-1)
  WDt[,1]=x
  WDt[,2]=phat-t.crt*wd.width
  WDt[,3]=phat+t.crt*wd.width
  WDt[,4]=(WDt[,2]<p)&(WDt[,3]>p)
  WDt[,5]=px
  Cvgprob.WDt=sum(WDt[,5][WDt[,4]==1])   # coverage prob of Wald's CI

 # Adj Wald CI
  AWD[,1]=x
  AWD[,2]=ptilde-z.crt*adjwd.width
  AWD[,3]=ptilde+z.crt*adjwd.width
  AWD[,4]=(AWD[,2]<p)&(AWD[,3]>p)
  AWD[,5]=px
  Cvgprob.AWD=sum(AWD[,5][AWD[,4]==1])   # coverage prob of Wald's CI

    return(list(WS=Cvgprob.WS, WDz=Cvgprob.WDz, WDt=Cvgprob.WDt,
  AWD=Cvgprob.AWD))
 }

#----------------------------------------------------------------------
# Coverage probabilities for all possible values of p
# which is used in plotting it with Exact CI
# --------------------------------------------------------------------
Cvgprob.allP=function(n,alpha,Np){
# Np=30
 CV=matrix(0,Np*(n+1),5)
 colnames(CV)=c("p", "Wislon","Wald.z", "Wald.t","AdjWald")

 for (k in 0:n){
```

```
    pstart=qbeta(alpha,k,n-k+1);    pend=qbeta(alpha,k+1,n-k)
                       # p*(i)=qbeta(alpha,i,n-i+1) i=0,...,n
                       # k=n corresponds to p in (p*(n),1]
    p=seq(pstart,pend,length=Np+1);p=p[-1]
                       # Choose Np of p's in the interval (P_U(k-1), P_U(k)]
                       # to calculate Coverage Prob P{p in C(X)}
    CV[(k*Np+1):((k+1)*Np),1]=p
    for (i in 1:Np) {
   a=Cvgprob.fixedP(n,alpha,p[i])
   CV[k*Np+i,2:5]=c(a$WS,a$WDz,a$WDt, a$AWD)
    }
  }
  return(CV)
}

# -----------------------------------------------------------------
# Figure 2.8-10
# Behavioral characteristics of coverage probability
# for various methods
# -----------------------------------------------------------------
par(mar=c(4,3,2,1))
alpha=0.05
A=Cvgprob.allP(30,alpha,30)
n=dim(A)
A1=A[-n,]
plot(A1[,1],A1[,4],type="l", xlab="p", ylab="",ylim=c(0.5,1.0))
abline(h=1-alpha, lty=2)

# -----------------------------------------------------------------
# Figure 2.11
# Compare coverage probability of CIs
# for various methods according to sample sizes
# Exact, Wilson, Wald with z, Wald with t
# -----------------------------------------------------------------
par(mar=c(4,3,2,1))
A=Cvgprob.allP(10,0.05,30)
n=dim(A)
A1=A[-n,]
plot(A1[,1],A1[,3],type="l", xlab="p", ylab="",ylim=c(0.5,1.0))
lines(A1[,1],A1[,2],lty=2,col="blue")
lines(A1[,1],A1[,5],lty=3, col="red")
legend(0.4, 0.7, lty=1:3, bty="n", col=c("black", "blue", "red"),
```

```
    legend=c("Wald", "Wilson","Adj Wald"))
abline(h=1-alpha, lty=2)

# ----------------------------------------------------------------------
# Figure 3.1
# Bayesian Coverage Probabilities
# of Equal-tailed and approximate Intervals
# --------------------------------------------------------------------------
# Bayesian Credible Region with Equal-Tailed Approach
# for a binomial proportion p with Jeffrey prior
#    p ~ beta(a,b)
#    X|p ~ bin(n,p)
#    p|x ~ beta(a+x, b+n-x)
#    E(p|x)=(a+x)/(a+b+n)
#    Var(p|x)=(a+x)(n-x+b)/ (a+b+n)^2 (a+b+n+1)
#
Bay.cvg1p=function(alpha,a,b,n,p){
  # Coverage probability of
  # Bayesian CI with Equal-tailed quantiles
  # prior beta(a,b)
  # inputs: alpha=0.05; a=b=1/2; n=50; p=0.5

  x=0:n
  px=dbinom(x,n,p)
  Eq.plow=qbeta(alpha/2,a+x,b+n-x)
  Eq.pup=qbeta(1-alpha/2,a+x,b+n-x)
  Length.Beq=Eq.pup-Eq.plow
  EL.Beq=sum(Length.Beq*px)

  z.crt=qnorm(1-alpha/2)
  Amean=(a+x)/(a+b+n)
  Aden=(a+b+n)^2*(a+b+n+1)
  Avar=(a+x)*(n-x+b)/Aden
  Asd=sqrt(Avar)
  Apup=Amean+z.crt*Asd
  Aplow=Amean-z.crt*Asd
  Length.Bap=2*z.crt*Asd
  EL.Bap=sum(Length.Bap*px)

  A=E=matrix(0,n+1,5)
    E[,1]=x
  E[,2]=Eq.plow ; E[1,2]=0      # modification
```

```
  E[,3]=Eq.pup   ;  E[n+1,3]=1
  E[,4]=(E[,2]<=p)&(E[,3]>=p)
  E[,5]=px
  Cvgprob.E=sum(E[,5][E[,4]==1])   # coverage prob of Bayesian Equal Tailed

  A=matrix(0,n+1,5)
  A[,1]=x
  A[,2]=Aplow; #A[1,2]=0        # modification
  A[,3]=Apup;   #A[n+1,3]=1
  A[,4]=(A[,2]<=p)&(A[,3]>=p)
  A[,5]=px
  Cvgprob.A=sum(A[,5][A[,4]==1])
  return(list(Eq=Cvgprob.E, As=Cvgprob.A,EL.Beq=EL.Beq, EL.Bap=EL.Bap))
}


ELength=function(n,alpha,p){
  # Expected Length of CI at a fixed value of p
  # n is fixed in advance
  # E[Length(CI)]=Sum_{k=0}^n {length[CI(k,p)]*P(x=k)}
  x=0:n
  px=dbinom(x,n,p)
  z.crt=qnorm(1-alpha/2)

# Exact Method
  U.exact=qbeta(1-alpha/2,x+1,n-x)
  L.exact=qbeta(alpha/2,x,n-x+1)
  Length.exact=U.exact-L.exact
  EL.exact=sum(Length.exact*px)

# Wilson Method
  phat=x/n
  wgt=n/(n+z.crt^2)
  s1=phat*(1-phat)*wgt+0.25*(1-wgt)
  s2=sqrt(s1/(n+z.crt^2))
  EL.wilson=2*z.crt*sum(s2*px)

# Wald Method
  a1=sqrt(phat*(1-phat)/n)
  EL.wald=2*z.crt*sum(a1*px)

# Adjusted Wald
```

```
   xtilde=x+z.crt^2/2
   ntilde=n+z.crt^2
   ptilde=xtilde/ntilde
   aw1=sqrt(ptilde*(1-ptilde)/ntilde)
   EL.AWD=2*z.crt*sum(aw1*px)

   return(list(exact=EL.exact, wilson=EL.wilson,
   wald=EL.wald, adj=EL.AWD))
}


# Check the function Bay.cvg1p
# a=Bay.cvg1p(0.05,1/2,1/2,50,0.02)
# b=Bay.cvg1p(0.05,1/2,1/2,50,0.98)

# ------------------------------------------------------------
# Function to compute coverage probabilities
# over the interval of p [0,1] uniform
# ----------------------------------------------
Bay.cvgallp=function(alpha,a,b,n){
   np=1000
   p=seq(0.0,1.0,length=np)
   CV=matrix(0,np,3)
   for (i in 1:np) {
   cvg=Bay.cvg1p(alpha,a,b,n,p[i])
   CV[i,]=c(p[i],cvg$Eq,cvg$As)
      }
   return(CV)
}

#-------------------------------------------------------
# Figure 3.1
# Plot the coverage probability for n=50
# ----------------------------------------------------
x=Bay.cvgallp(0.05,1/2,1/2,50)
par(mar=c(4,3,2,1))
plot(x[,1],x[,2],xlab="p", ylab="",type="l", ylim=c(0.84,1.00))
abline(h=1-alpha, lty=2)
lines(x[,1],x[,3],col="blue", lty=2)
legend(0.4,0.90,lty=c(1,2),legend=c("Equal-Tailed","Approximate"),bty="n")
```

```
# -------------------------------------------------------------------------------
# Table 4.1
# Comparison of Mean coverage probability with root MSE
# for the various sample size
# 1) Bayesian Credible Region
# ----------------------------------------------------------------
nc=c(5,10,15,20,40,50,60,80,100)
MC=matrix(0,length(nc),5)
MC[,1]=nc
for (i in 1:length(nc)){
  X=Bay.cvgallp(0.05,1/2,1/2, nc[i])
  Xmean=apply(X[,2:3],2, mean)
  MC[i,2]=Xmean[1]
  MC[i,3]=sqrt(mean(X[,2]-0.95)^2)
  MC[i,4]=Xmean[2]
  MC[i,5]=sqrt(mean(X[,3]-0.95)^2)
}

#------------------------------------------------------------------------------
# 2) Mean coverage probability and Root MSE
#    for the various sample size and
#    p from uniform or beta distribution
#    for Exact, Wilson, Wald with z, Wald with t, Adj-Wald
# ----------------------------------------------------------------------------

Comp.cvg.beta=function(n, alpha, Nsim){
  CVG=matrix(0, Nsim, 6)
  # colnames(CVG)=c("p","Exact","Wilson","Wald.z", "Wald.t","Adj Wald")
  for(j in 1:Nsim){
    # p=rbeta(1,3.5, 31.5)     # beta(3.5, 3.5)
     p=rbeta(1,12,12)            # beta(12, 12)
     CVG[j,1]=p
     CVG[j,2]=Pboth.cvg(n,alpha,p)      # Exact Method
     a=Cvgprob.fixedP(n,alpha,p)        # Wilson, Wald, Adj-Wald
     CVG[j,3:6]=c(a$WS,a$WDz,a$WDt, a$AWD)
  }
  mean.cvg=apply(CVG[,2:6],2,mean)
  X=CVG[,2:6]-0.95
  RMSE.cvg=sqrt(apply(X^2,2,mean))

  return(c(mean.cvg, RMSE.cvg))
}
```

```
nsize=c(5,10,15,20,40,50,60,80,100);  alpha=0.05
Nsim=1000
ntype=length(nsize)
A=matrix(0,ntype,11)
colnames(A)=c("n","m-Ex","m-WS","m-WDz","m-WDt","m-AWD",
              "Rm-Ex","Rm-WS","Rm-WDz", "Rm-WDt", "RM-AWD")
for (k in 1:ntype){
n=nsize[k]
#A[k,]=c(n,Comp.cvg(n,alpha,Nsim))            # If Uniform
A[k,]=c(n,Comp.cvg.beta(n,alpha,Nsim))        # If Beta distn
}

# In order to have the Table 4.1
# combine MC and A

# -------------------------------------------------------------
# Figure 4.1
# Mean coverage probability as a function of n
# We need the output A and MC from Table 4.1
# -------------------------------------------------------------
par(mar=c(3,5,2,2))
nc=0.8   # point size
F4=cbind(A[,2:4],A[,6],MC[,2],MC[,4])
 colnames(F4)=c("Ex","WS","WD","AdW","B-Eq","B-Ap")
plot(nsize,F4[,1],type="l", ylim=c(0.85,1.0), xlab="n", ylab="", lty=3,cex=nc)
color=c("red","purple","green","red","purple","green")
for (k in 2:6) {
  lines(nsize,F4[,k],lty=k,cex=nc)
}
legend("topright", lty=1:6,
     legend=c("Exact", "Score", "Wald", "AdjWald","B-Equal","B-App"))


# ------------------------------------------------------------------------------------------
# Figure 4.2 Comparison of the Expected Lengths
# Actual coverage probability
# for p from uniform(0,1)
# ------------------------------------------
Comp.cvg=function(n, alpha, Nsim){
  CVG=matrix(0, Nsim, 6)
#  colnames(CVG)=c("p","Exact","Wilson","Wald.z", "Wald.t", "Adj Wald")
```

```
  for(j  in  1:Nsim){
      p=runif(1)
      CVG[j,1]=p
      CVG[j,2]=Pboth.cvg(n,alpha,p)
      a=Cvgprob.fixedP(n,alpha,p)
      CVG[j,3:6]=c(a$WS,a$WDz,a$WDt,a$AWD)
  }
  mean.cvg=apply(CVG[,2:6],2,mean)
  X=CVG[,2:6]-0.95
  RMSE.cvg=sqrt(apply(X^2,2,mean))

  return(c(mean.cvg,  RMSE.cvg))
}


# -------------------------------------------------------------------------------
# Figure  4.2-4.4
# A.  Expected  Length  of  CI
#   as  a  function  of  p  with  n  fixed  in  advance
# B.  Average  Expected  Length
#   over  p  in  [0,1]  as  a  function  of  n.
# C.  Mean  Absolute  Error
#   over  p  as  a  function  of  n
# -------------------------------------------------------------------------------

ELength=function(n,alpha,p){
  # Expected  Length  of  CI  at  a  fixed  value  of  p
  # n  is  fixed  in  advance
  # E[Length(CI)]=Sum_{k=0}^n  {length[CI(k,p)]*P(x=k)}
  x=0:n
  px=dbinom(x,n,p)
  z.crt=qnorm(1-alpha/2)

  # Exact  Method
  U.exact=qbeta(1-alpha/2,x+1,n-x)
  L.exact=qbeta(alpha/2,x,n-x+1)
  Length.exact=U.exact-L.exact
  EL.exact=sum(Length.exact*px)

   # Wilson  Method
  phat=x/n
  wgt=n/(n+z.crt^2)
```

```
   s1=phat*(1-phat)*wgt+0.25*(1-wgt)
   s2=sqrt(s1/(n+z.crt^2))
   EL.wilson=2*z.crt*sum(s2*px)

    # Wald Method
   a1=sqrt(phat*(1-phat)/n)
   EL.wald=2*z.crt*sum(a1*px)

    # Adjusted Wald
   xtilde=x+z.crt^2/2
   ntilde=n+z.crt^2
   ptilde=xtilde/ntilde
   aw1=sqrt(ptilde*(1-ptilde)/ntilde)
   EL.AWD=2*z.crt*sum(aw1*px)

   return(list(exact=EL.exact, wilson=EL.wilson,
   wald=EL.wald, adj=EL.AWD))
}

# ------------------------------------------------------------
# Figure 4.2
# Compute Expected Length of CI for a p in [0,1]
# and plot it
# ------------------------------------------
n=100; alpha=0.05;
p=seq(0,1,by=0.01); np=length(p)
A=matrix(0,np,6)
for (k in 1:np){
   a=ELength(n,alpha,p[k])
   A[k,1:4]=c(a$exact, a$wilson, a$wald, a$adj)
   b=Bay.cvg1p(alpha,1,1,n,p[k])      # U(0,1)=Beta(1,1)
   A[k,5:6]=c(b$EL.Beq, b$EL.Bap)
}

plot(p,A[,1],type="l",xlab="",ylab="E(width)")
for (k in 2:6){
   lines(p,A[,k],lty=k)
}

lines(p,A[,5],col="red")       # Bayes with equal tailed
lines(p,A[,6],col="blue")      # Bayes with approximate
```

```
legend(0.3,0.15,lty=1:6,
c("Exact","Wilson","Wald","AdjWald","Bayes_Eq","Bayes.App"),
col=c("black","black","black","black","red","blue"))

# ---------------------------------------------------
# Figure 4.3-4.4
# Compute Average Expected Length and
# Mean Absolute Error
# as a function of sample size and plot it
# --------------------------------------------------
AvgEL=function(n,alpha) {
         # Average Expected Length
  p=seq(0,1,by=0.01); np=length(p)
  A=matrix(0,np,6)
  for(k in 1:np){
  a=ELength(n,alpha,p[k])
  A[k,1:4]=c(a$exact, a$wilson, a$wald, a$adj)
  b=Bay.cvg1p(alpha,1,1,n,p[k])      # U(0,1)=Beta(1,1)
  A[k,5:6]=c(b$EL.Beq, b$EL.Bap)
  }
AL=apply(A,2,mean)
}

AbsErr=function(n,alpha,binomp){
   # Absolute Error
  p=binomp
  AE=Pboth.cvg(n,alpha,p)           # Cov. prob for exact method
  a=Cvgprob.fixedP(n,alpha,p)       # coverage probability for approximate methods
  b=Bay.cvg1p(alpha,1,1,n,p)        # Coverage Prob for Bayesian
  Abs.Error=abs(c(AE, a$WS, a$WDz, a$AWD, b$Eq, b$As)-(1-alpha))  #relative error
}

MeanAbsErr=function(n,alpha) {
    # Mean Absolute Error
  p=seq(0,1,by=0.01); np=length(p)
  A=matrix(0,np,6)
  for(k in 1:np){
  A[k,]=AbsErr(n,alpha,p[k])     # Exact, Wilson, Wald-z, Adj-WD, B-eq, B-appr
  }
AL=apply(A,2,mean)
}
```

```
nsize=c(seq(10,25),seq(30,100,by=10)); nlen=length(nsize)
B=matrix(0,nlen,6)
MAE=matrix(0,nlen,6)
for (k in 1:nlen){
   B[k,]=AvgEL(nsize[k],0.05)
   MAE[k,]=MeanAbsErr(nsize[k],0.05)
}

#--------------------------------------------
# Figure 4.3
# Plot Average Expected Length
# ------------------------------------------
plot(nsize,B[,1],type="l",ylim=c(0,0.6),xlab="", ylab="")
mtext("n",side=1,line=1.5)
mtext("Avg[E(width)]",side=2,line=2.5)
for (k in 2:6){
   lines(nsize,B[,k],lty=k)
}
lines(nsize,B[,5],col="red")
lines(nsize,B[,6],col="blue")
legend(15,0.6,lty=1:6,c("Exact", "Score", "Wald", "A&C", "Bayes.Eq", "Bayes.Ap"),
   col=c("black","black","black","black","red","blue"))


#--------------------------------------------
# Figure 4.4
# Plot Mean Absolute Error
# ------------------------------------------
plot(nsize,MAE[,1],type="l",ylim=c(0,0.6),xlab="", ylab="")
mtext("n",side=1,line=1.5)
mtext("E[|error|]",side=2,line=2.5)
for (k in 2:6){
   lines(nsize,MAE[,k],lty=k)
}
lines(nsize,MAE[,5],col="red")
lines(nsize,MAE[,6],col="blue")
legend(15,0.6,lty=1:6,c("Exact", "Score", "Wald", "A&C", "Bayes.Eq", "Bayes.Ap"),
   col=c("black","black","black","black","red","blue"))
```