# Interval Estimates for Parameters of Discrete Sampling Distributions

## Thesis

# Interval estimates for parameters of discrete sampling distributions.

By

**Maha Wael Elbakri Moustafa**

***BSc & MSc, Mansoura University, Egypt.***

**A thesis submitted for the Degree of Doctor of Philosophy in Statistics**

**School of Mathematics and Statistics, STEM**

**The Open University, UK**

**September 2020**

**Acknowledgements**

This work would never have been completed without the kind and unconditional support of a few people to whom I wish to express my sincere appreciation. First and foremost, I am extremely grateful to my main supervisor Prof. Paul Garthwaite (The Open University, UK), for suggesting the research topic, his supervision, guidance, valuable advice, encouragement, kindness, deep interest and continuous help during the preparation of this thesis. I would like also to deeply thank my co-supervisor Dr. Fadlalla Elfadaly (The Open University, UK), for his believing in me, advices, directions, his continuous willingness to help and support. Really my words can not express how I am grateful for both my supervisors. Much of the learning that has gone into this thesis would not have been possible without their guidance, meticulous comments throughout the research and writing process, especially they always made time to discuss my work with me. I would like to express my deepest gratitude to my viva examination panel, Prof. Steven Gilmour (King's College London, UK), Dr. Karen Vines and Dr. Alvaro Faria (The Open University, UK) for their valuable comments, constructive criticisms and helpful suggestions.

Also, many thanks are due to Prof. Chris Jones, Emeritus Professor (The Open University, UK), for his positive and effective comments in my mini viva and his encouragement that created a comfortable environment to discuss my work. I wish to thank Prof. Gwyneth Stallard and Dr. Heather Whitaker ( The Open University, UK), for their continuous psychological support during my PhD journey. Also, I would like to thank all members of the Statistics Group, The Open University, UK. They all helped me a lot in a very cooperative and supportive research environment that leads to continuous progress and achievement. Special gratitude to the previous PhD students, Dr. Zillur Rahman Shabuz and Dr. Kwo Chan and to the current PhD students, Mr. Ibai Aedo and Mr. Alexander Round. They formed a great academic and social atmosphere for effective work.

True gratitude and deep appreciation are due to Prof. Fatma Abdelaty, Mansoura University, Egypt, for her guidance, support and encouraging to undertake my PhD in the UK. Really, she is a role model for me in her wonderful manners, abundant

knowledge and wisdom. I am also very grateful for my dear friends Dr. Asmaa Hafez, Mansoura University, Egypt and Mrs. Amal Elsadaty for being such a sincere, supportive and helpful friends.

I am enormously grateful to my dear father, Mr. Wael Moustafa, my caring mother Olla ElHanfy, my kind sister Mrs. Nourhan Moustafa and my beloved brother Dr Mohamed Moustafa, who have always encouraged and supported me throughout my education, and for all the prayers they made for me. The last and most important people I would like to thank are my beloved husband, Dr. Ahmed Elamer, Brunel University, UK and my amazing boys Omar and Yousef Elamer. Considering that they are 9 and 7 years old, it's hard to imagine how they could contribute to a doctoral dissertation. But in fact, their bright smile and their lovely words about how I am an amazing mum even if I could not do well in my PhD journey, encourage me to do my best to be the mum who they can be proud of.

As for my husband Ahmed, I find it difficult to express my appreciation because it is so boundless. He is my most enthusiastic cheerleader; he is my best friend; and he is an amazing husband and father. He is my rock. Without his willingness to help me with our sons, this thesis would have taken even longer to complete; without his sunny optimism, I would be a much grumpier person; without his love and support, I would be lost. I am grateful to my husband not just because he has given up so much to make my career a priority in our lives, but because his believing in me long after I'd lost belief in myself, for sharing my wish to reach the goal of completing my PhD and caring enough to love me even if I never achieved it.

**Abstract**

Well recommended methods of forming confidence intervals for discrete distributions parameter give interval estimates that do not actually meet the definition of a confidence interval, in that their coverage is sometimes lower than the nominal confidence level. The methods are favored because their intervals have a shorter average length than the exact method, whose intervals really are confidence intervals. Comparison of such methods is tricky as the best method should perhaps be the one that gives the shortest intervals (on average), but when is the coverage of a method so poor it should not be classed as a means of forming confidence interval.

As the definition of a confidence interval is being flouted, a better criterion for forming interval estimates for discrete distributions parameters is needed. The aim of this thesis is to suggest a new criterion: methods that meet the criterion are said to yield *locally correct confidence intervals*. We propose a method that yields such intervals and proves that its intervals have a shorter average length than those of any other method that meets the criterion. We refer to the new estimator as the *optimal locally correct* method or just the OLC method. The thesis begins by applying the new criterion and method to the binomial parameter. Then we extend the method so as to obtain locally correct confidence intervals for parameters of the Poisson distribution and the negative binomial distribution.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Constructing a confidence interval for a scaler parameter is one of the most common analyses in statistical inference. When the interval is determined by a sample statistic $X$, we let $l_i$ and $u_i$ denote the lower limit and the upper limit of the confidence interval when $X = i$. Before the value of $x$ is observed, the confidence interval is a random quantity and the probability that it will contain the true value of the scalar parameter is referred to as its *coverage*. This coverage may depend upon the true value of the scalar parameter and will depend on the method used to form the confidence interval.

We will distinguish between two situations. Firstly, obtaining a confidence interval when sampling is from a continuous distribution, such as a normal distribution. This is relatively straightforward and typically there is one method of forming the confidence interval that is optimal. Often the method

(i) gives $(1 - 2\alpha)$ confidence intervals whose coverage is $1 - 2\alpha$, regardless of the true value of the parameter of interest.

(ii) gives intervals that are as short as possible, subject to (i) holding.

The second case is where sampling is from a discrete distribution. Sampling from binomial, Poisson and negative binomial distributions are the most common examples and are the ones considered in this thesis. The difficulty in this task stems from the discrete nature of the sample space, which leads to sharp fluctuations in the coverage probability for different values of the distribution parameter. This is illustrated in the top graph of Figure 1.1, which gives the coverage for an upper one-sided confidence interval for the binomial proportion $p$ for one method of forming confidence intervals for discrete distribution (the mid-$p$ method). The coverage is plotted as the blue line in Figure 1.1. The coverage of an upper one-sided interval is calculated as the $Pr(p < u_i) = \sum_{x=i}^{n} Pr(X = x|p)$, for $i = 0, \ldots, n$, where sampling is from

a binomial $(n, p)$ distribution. The coverage increases monotonically as $p$ increases from $u_{i-1}$ to $u_i$ but when $p$ moves from being just less than $u_i$ to being just greater than $u_i$, the coverage suddenly drops. That is, the coverage plot has spikes at $u_0, u_1, \ldots, u_n$. They occur because for any point of $p$, say $p_0$, just less than $u_i$, the coverage probability is $Pr(p < u_i) = \sum_{x=i}^{n} Pr(X = x|p = p_0)$, while for any other point of $p$, say $p*$, just greater than $u_i$, the coverage probability is $Pr(p < u_i) = \sum_{x=i+1}^{n} Pr(X = x|p = p*)$. So the difference between the coverage of the point $p_0$ just before $u_i$, and the coverage of the point $p*$ just after $u_i$, is $Pr(X = i|p = u_i)$. This probability gives the size of the spikes. It does not equal zero in the case of a discrete distribution and this happens at each $u_i$. When the sampling distribution of $x$ is continuous (rather than discrete), the probability that the upper limit exactly equals any specfied value is zero, so there are no sharp changes in coverage.

To give a specific example, we take the mid-$p$ method for a binomial proportion $p$. Suppose that we have 15 trials and that the upper one-sided limit of a 97.5 % interval for $p$ is required. When $x = 4$, the value of $u_4$ (correct to 4 decimal places) is $u_4 = 0.5253$. The coverage probability increases from $u_3$ to $u_4$, reaching the highest point of a spike at $u_4$. When $p_0$=0.5252, $Pr(p_0 < u_i) = \sum_{x=4}^{15} Pr(X = x|p = 0.5252) \simeq 0.9892959$. When $p* =$0.5254, $Pr(p* < u_i) = \sum_{x=5}^{15} Pr(X = x|p = 0.5254) \simeq 0.9607177$. As $Pr(p < u_i)$ is the coverage at $p$, the coverage drops sharply from $p_0$ to $p*$. It is clear that the drop is almost identical to $Pr(X = 4|p = 0.5253) \simeq 0.0285782$. The drop happens because there is a point mass of probability that the confidence limit is 0.5254.

Similar features arrise with the lower-tail coverage, as can be seen in the lower graph of Figure 1.1. (The lower endpoints are obtained by inversion of the upper endpoints). The coverage for the one-sided lower interval is

$Pr(p > l_i) = \sum_{x=0}^{i} Pr(X = x|p)$, for $i = 0, \ldots, n$. This coverage decreases montonically as the value of $p$ increases from $l_{i-1}$ to $l_i$, but when $p$ moves from being just less than $l_i$ to be just greater than $l_i$, the coverage suddenly increases.

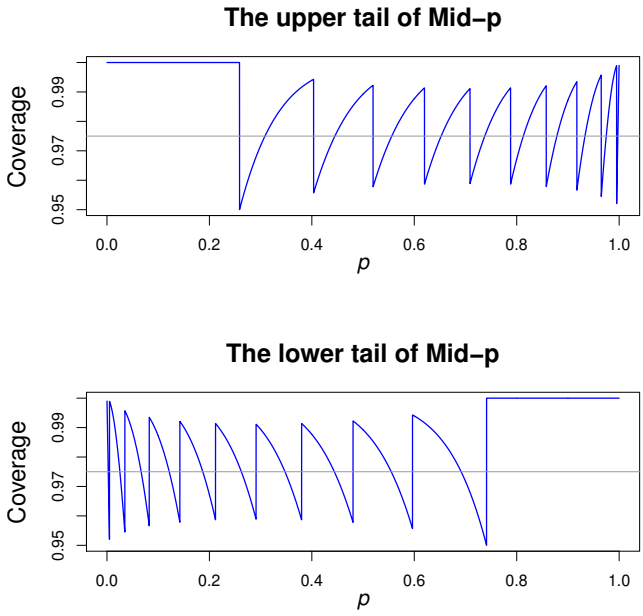


Figure 1.1: Coverage of upper and lower one-sided 97.5% confidence interval for the mid-$p$ method for a fixed sample size ($n$) of 10 and success parameter $p$. The blue lines show the coverage.

For the common discrete distributions, a good number of methods have been proposed for forming a confidence interval for the unknown parameters of the distribution. This is the case for a binomial proportion $p$, a Poisson mean $\lambda$ and a negative binomial proportion $p$. A number of methods are compared in studies by: Newcombe (1998), Swift (2009), Vollset (1993) and Brown *et al.* (2001), among many others. These studies are interested in methods that aim to form equal-tailed confidence intervals. That is, methods that obtain two-sided confidence intervals by constructing two one-sided intervals. If $(l, u)$ is a $(1-2\alpha)$ equal-tailed confidence interval, then $(l, 1)$ and $(0, u)$ are one-sided $(1-$

$\alpha$) confidence intervals. In general, when sampling from discrete distribution we cannot form an interval $(l_x, u_x)$ for which $Pr(l_x \leq p \leq u_x)$ equals exactly $(1 - 2\alpha)$. By definition (see, for example, Bickel and Doksum (1977), pages 154-155), the random interval $(l_x, u_x)$ formed by a pair of statistics $l_x, u_x$ with $l_x \leq u_x$ is a level $(1 - 2\alpha)$ or a $100(1 - 2\alpha)\%$ confidence interval for $p$ if, for all $p$,

$$Pr_p(l_x \leq p \leq u_x) \geq (1 - 2\alpha). \tag{1.1}$$

The quantity on the left of this equation is the coverage probability of the interval. If a method of forming confidence intervals satisfies (1.1) we will refer to it as a *strictly correct* method and say it gives *correct* confidence intervals. There are several methods of forming confidence interval that have been proposed that do not satisfy (1.1). We will refer to these as *approximate* methods and say they give *approximate* confidence intervals.

Strictly correct methods have been criticised as being conservative and giving intervals that are relatively wider than the approximate methods. Approximate methods typically give intervals that are narrower on average over the range of the discrete distribution parameter. However, this advantage leads to the primary disadvantage of the methods: their coverage probabilities do not satisfy (1.1) for all values of the discrete distribution parameter. It follows that these approximate methods do not actually give $(1 - 2\alpha)$ confidence intervals.

For the binomial distribution, the "gold standard" method of forming an equal-tailed confidence interval is the Clopper-Pearson method [Pearson (1924)]. This is a good example of a correct method for discrete sampling distributions. Its interval estimators meet the definition of a confidence interval but the method suffers from conservatism and in many papers it is suggested that its intervals are too wide [Newcombe (1998), Brown *et al.* (2001), Dunnigan

(2008)]. Its conservatism is illustrated in the left-hand diagram of Figure 1.2, which plots the coverage (blue lines) of its 97.5% one-sided upper intervals against $p$ for sampling from a binomial (15, $p$) distribution. In the plot, the saw-tooth pattern results from the discrete nature of the sample space and arises with any method of forming a confidence interval for discrete distribution parameters as discussed before. It is clear that the coverage is commonly above 98.5%, sometimes exceeding 99.5%. The right-hand diagram of Figure 1.2 shows the coverage of the Wald method, which is an example of an approximate method. Its intervals do not meet the definition of a confidence interval (for many values of $p$ the coverage is below 97.5%) but it gives narrower intervals. The coverage of the Wald method tends to be liberal for small values of $p$ and conservative for large values. Many other methods of forming equal-tailed confidence intervals for a binomial proportion have been proposed, for example, Wilson method [Boomsma (2005)] , Agresti-Coull method [ Brown *et al.* (2002)] and mid-p method [Berry and Armitage (1995)]. These aim to avoid the drawbacks of both Clopper-Pearson and Wald methods, but while the average lengths of their intervals are shorter than those of the Clopper-Pearson method, they only give approximate confidence intervals that do not meet the definition of a confidence interval.

This suggests that the definition of a confidence interval does not meet our needs when the sampling space is discrete. Hence a new definition of an interval estimate is needed. For the definition to be useful there should be some interval estimators that:

(i) satisfy the new definition,

(ii) give sensible intervals,

(iii) give intervals with an average length that is acceptably short.

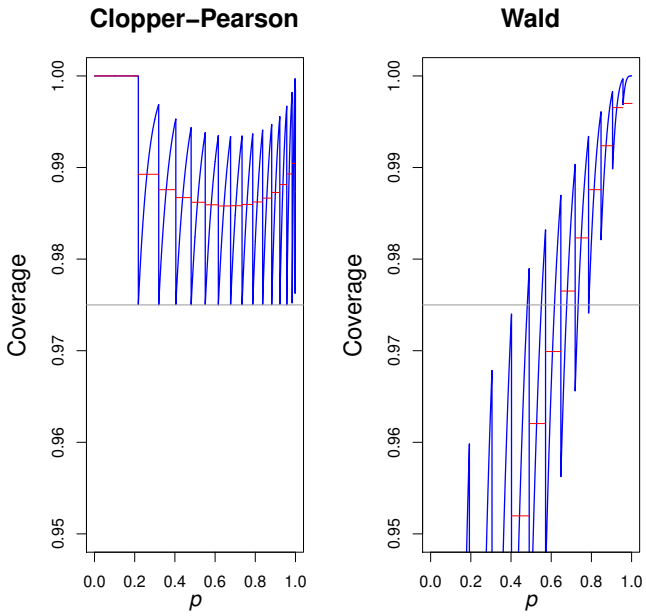


Figure 1.2: Coverage of upper one-sided 97.5% confidence interval for the Clopper-Pearson and Wald methods for a fixed sample size ($n$) of 15 and success parameter $p$. Short horizontal (red) lines show the average coverage betweeen consecutive spikes.

Given a suitable definition, attention can be restricted to methods that meet that definition and these methods can be compared on the basis of the width of the intervals. A reasonable criterion as to which is the best interval estimator would be the one with the shortest average length when averaged over the probability scale with each value of $p$ equally likely.

The main challenge is to find an appropriate definition. We propose a new definition that reflects the saw-tooth pattern of coverage that is shown in Figure 1.2. Our new definition is that, for one-sided intervals, the average coverage between any pair of consecutive spikes must equal or exceed the nominal level $(1-\alpha)$. Methods of forming intervals that meet this definition will be referred to as *locally correct* methods and the intervals will be termed *locally correct confidence* (LCC) intervals. In Figure 1.2, short horizontal (red) lines

show the average coverage between pairs of consecutive spikes. It is clear that Clopper-Pearson intervals are LCC intervals while those given by the Wald method are not. The new definition leads us to propose a new method of constructing locally correct confidence intervals for a binomial proportion. The method yields intervals with smaller average length than any other method that gives locally correct confidence intervals. We refer to the new method as the optimal locally correct method or just the OLC method. We then extend the method so as to obtain locally correct confidence intervals for parameters of the Poisson distribution and the negative binomial distribution.

This thesis consists of seven chapters. After this introductory chapter, Chapter 2 first gives a brief literature review of general methods of forming a confidence interval for discrete sampling distributions. Then we discuss some of the well-known methods of forming confidence intervals for the binomial success parameter $p$, the Poisson parameter $\lambda$ and the negative binomial success parameter $p$ and its mean $\mu$. In addition, the advantages and disadvantages of the different methods are reviewed. We also discuss the overall findings in relation to the aims of this thesis.

In Chapter 3, a precise definition of a locally correct confidence interval for the binomial success parameter $p$ is given. Then a novel interval estimator that yields locally correct confidence intervals is presented and it is proved that the new OLC estimator yields intervals with a smaller average length than any other interval estimator that yields LCC intervals. We also examine whether intervals given by the new estimator have the desirable properties (i), (ii) and (iii) that are mentioned above. For (iii), which concerns the length of the intervals, the OLC method is compared with several other methods: Clopper-Pearson, mid-p, Wilson, Wald, Agresti-Coul and Jeffreys methods.

In Chapter 4, some general results about the new interval estimator are de-

volped. These results hold for any discrete distribution with an unknown scaler parameter.

In Chapter 5, a locally correct estimator for a Poisson parameter $\lambda$ is introduced. Compared to the binomial distribution, the Poisson distribution is simpler as it has only one parameter $\lambda$, whereas the binomial has two parameters $n$ and $p$. However, a complication with the Poisson distribution that does not arise with the binomial (for fixed $n$) is that the range of the parameter $\lambda$ goes to infinity. This infinite range causes a problem in calculating OLC intervals by using the new method and we adapt the method to handle it. We also examine if the intervals given by the OLC method seem sensible and if their average length is acceptably short. For the latter, we compare the OLC method with the methods for a Poisson distribution that are most recommended: Garwood, mid-p, Wald, score and Jeffreys methods. This comparison is made a little bit awkward by the infinite range of the parameter $\lambda$ because it is impossible to compare the average lengths as they are infinite. Some previous studies determine the average width in the range 0-50 for $\lambda$. Others give more than one average width, giving an average for $\lambda$ in the range 0-2, another for $\lambda$ in the range 2-5 and a third for $\lambda$ in the range 5-50. This can make it difficult to select a "best" method, so instead we determine a weighted average width, using a weight function that gives a small weight for large values of $\lambda$ and results in a finite weighted average width. This gives a single average for the full range of $\lambda$, $0 < \lambda < \infty$.

In Chapter 6, we discuss confidence intervals for the negative binomial distribution. This distribution has two parameters $(p, r)$, where $0 < p < 1$ and $r = 1, 2, \ldots$, but $r$ is known. In applications, sometimes a confidence interval for the parameter $p$ is required, but quite often a confidence interval is required for the mean $\mu$ of the negative binomial distribution, where $\mu = r(1-p)/p$.

We use the OLC method to calculate LCC intervals for both $p$ and $\mu$. The method is compared with alternative methods and directly gives good intervals.

Concluding comments are given in Chapter 7 where some directions for future research are also considered. The material reported in Chapter 4 and appendices A to D is largely the work of my main PhD supervisor.

# Chapter 2

# Literature review

## 2.1 Introduction

Interval estimation for discrete distributions has been widely discussed in the literature for more than fifty years. This chapter aims to review the recent literature on common methods of forming a confidence interval for discrete distributions. The emphasis here is on methods of forming two-sided, equal-tailed confidence intervals, for the binomial proportion $p$, the Poisson mean $\lambda$ and the negative binomial distribution's proportion $p$ and mean $\mu$.

Methods of forming a confidence interval for the binomial proportion are reviewed in Section 2.2. Section 2.3 reviews the most commonly used interval estimators of the Poisson mean $\lambda$. Confidence interval estimators for both the negative binomial proportion $p$ and its mean $\mu$ are considered in Section 2.4. In Section 2.5 some concluding comments are given.

## 2.2 Binomial distribution confidence intervals

Constructing a confidence interval for the binomial proportion is a task that has attracted much attention in the literature. Many methods of constructing such intervals have been proposed and compared in applied studies. The methods are often classified in the literature as exact methods and approximate methods, but there are two interpretations of the term "exact". As Newcombe (1998) mentioned, a method is somtimes called an exact method if

(i) it is based on exact sampling distributions, such as binomial or Poisson, not on any asymptotic approximation.

And sometimes it is called an exact method if

(ii) it attains confidence intervals with a coverage probability equal to or

greater than the nominal level $1-\alpha$ for all values of $p$.

Both the Clopper-Pearson and mid-$p$ methods, which will be discussed in more detail in the next section, are exact in sense (i) but only the Clopper-Pearson is exact in both the senses (i) and (ii). For clarity, throughtout this thesis we will refer to a method as "exact" if it is exact in sense (i). We refer to a method as "correct" if it is exact in sense (ii). Similarly, in the literature a method is sometimes described as approximate because it is based on asymptotic approximation, and somtimes because its average is a little lower than the nominal confidence level for some parameter values. In this thesis we differentiate between the cases as follows.

(i) If the method gives confidence intervals that are based on an asymptotic approximation, we refer to it as an asymptotic method.

(ii) If the confidence intervals do not have a coverage probability that is equal to or greater than the nominal level $1-\alpha$ for all values of $p$, we refer to the method as an approximate method.

The Wald, Wilson and Agresti-Coull methods, which will be discussed in details in Section 2.2.2, are asymptotic methods while both the mid-$p$ and Jeffreys methods are exact methods, but they all are approximate methods. When the sample size is large, asymptotic methods can achieve good accuracy. The approximate confidence intervals are calculated using the normal approximation to the binomial distribution.

Most of the methods described in Sections 2.2.1 and 2.2.2 are compared in studies by Newcombe (1998), Pires and Amado (2008), Ghosh (1979) and Vollset (1993). Comparison is generally made in terms of coverage probability and average width. Both coverage probability and average width are sensible criteria as high coverage and narrow intervals are desirable qualities. In Sec-

tions 2.2.1 and 2.2.2, the methods that are most widely used in practice are discussed. We are interested in reviewing methods that give two-sided intervals with equal-tails, as equal-tailed intervals are much preferred in practice. Desirable properties in methods of forming confidence regions are described in Section 2.2.4. We briefly review methods that do not aim to give two-sided confidence intervals with equal-tails in Section 2.2.5.

### 2.2.1 Correct methods

#### 2.2.1.1 Clopper-Pearson method

Clopper and Pearson (1934) gave an exact method of forming two-sided confidence interval for a proportion $p$ by inversion of the equal-tailed binomial test of the null hypothesis $H_0 : p = p_0$ against the alternative hypothesis $H_1 : p \neq p_0$. The interval contains all values of $p$ that are not rejected by the test for a nominal confidence level $(1 - \alpha)$. Hence, if $X$ is a binomial $(n, p)$ random variable with probability density

$$f(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n,$$

then the lower limit, $l_i$, satisfies

$$\sum_{x=i}^{n} \binom{n}{x} (l_i)^x (1-l_i)^{n-x} = \alpha \tag{2.1}$$

and the upper limit, $u_i$, satisfies

$$\sum_{x=0}^{i} \binom{n}{x} (u_i)^x (1-u_i)^{n-x} = \alpha \tag{2.2}$$

except that $l_0$ is set equal to 0 and $u_n$ is set equal to 1.

Clopper-Pearson is an exact method as it is based on the exact cumulative probabilities of the binomial distribution. Of the methods we consider, it is the only one that is correct; it strictly meets the definition of a method for forming confidence intervals and satisfies equation (1.1). A number of researchers (e.g. Pearson (1924) and Brown *et al.* (2001)) showed that because of the relationship between the cumulative binomial distribution and a beta distribution, Clopper-Pearson confidence intervals can be written in terms of quantiles of the beta distribution. If $[l_{cp(x)}, u_{cp(x)}]$ is the $(1-2\alpha)$ confidence interval given by the Clopper-Pearson method, then $l_{cp(x)}$ is the $\alpha$ quantile of a Beta $(X, n-X+1)$ distribution and $u_{cp(x)}$ is the $(1-\alpha)$ quantile.

The Clopper-Pearson interval is treated as a" gold-standard" method among the methods for forming confidence interval [Leemis and Trivedi (1996), Jovanovic and Levy (1997)] because it guarantees that the coverage probability is always equal to or greater than the nominal confidence level and gives shorter intervals than any other method of forming equal-tailed intervals that has this property. It is simple computationally and has been implemented in almost all statistical software packages. However, for almost all values of $p$, the coverage probability of its intervals is larger than the nominal level, which means that it suffers from conservatism [Newcombe (1998), Brown *et al.* (2001), Dunnigan (2008)]. It is clear in the top graph of Figure 2.1, which plots the coverage of its 97.5% one-sided upper intervals against $p$ for sampling from a binomial (20, $p$) distribution, that the coverage is noticeably above 99 % almost all the time, sometimes exceeding 99.5%.

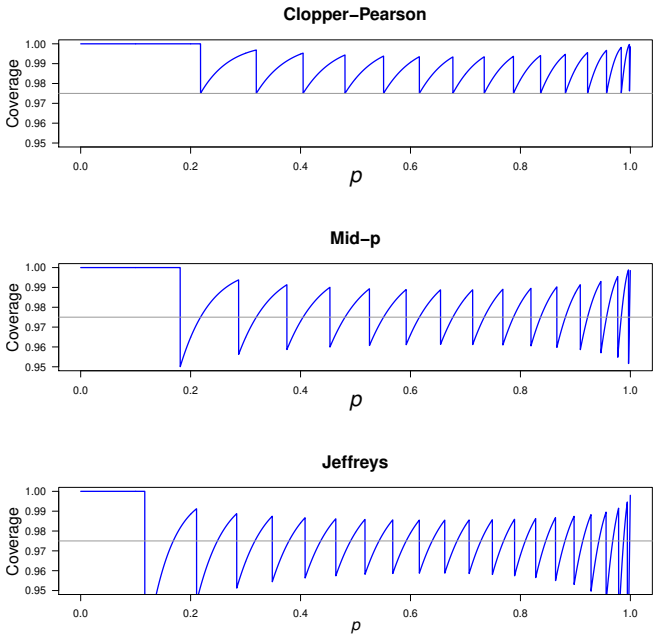


Figure 2.1: Coverage of upper one-sided 97.5% confidence interval for the Clopper-Pearson, mid-$p$ and Jeffreys methods for a fixed sample size ($n$) of 20 and success parameter $p$.

A conservative interval is wider than neccessary, so a common criticism of Clopper-Pearson intervals is that they are too wide. This is reflected in the number of approximate methods that have been proposed that give shorter intervals than the Clopper-Pearson method. Some of these methods are described in the following sub-sections.

### 2.2.2 Approximate methods

Neither the mid-$p$ nor Jeffreys methods meet the definition of the confidence interval in equation (1.1) but both are exact as they do not use asymptotic approximations. The mid-$p$ method reduces conservatism by using half of the probability of the observed result, while Jeffreys method is a Bayesian credible interval that uses a Beta $(\frac{1}{2}, \frac{1}{2})$ distribution as the prior distribution.

#### 2.2.2.1 Mid-$p$ method

To avoid the conservatism of the Clopper-Pearson method, Lancaster (1961) suggested the mid-$p$ method (Berry and Armitage (1995); Mehta and Walsh (1992)). This method is similar to the Clopper-Pearson method except that it halves the probability of the observed result. Specifically, to construct the confidence interval $(l_i, u_i)$, the lower limit is chosen to satisfy

$$Pr(X > i|l_i) + \frac{1}{2}Pr(X = i|l_i) = \alpha \tag{2.3}$$

and the upper limit satisfies

$$Pr(X < i|u_i) + \frac{1}{2}Pr(X = i|u_i) = \alpha \tag{2.4}$$

except that $l_o = 0$ when $x = 0$ and $u_n = 1$ when $x = n$, as otherwise the coverage is 0 for $p < p_l$ and $p > p_u$ (Agresti and Gottard, 2005).

The mid-$p$ method reduces the conservatism of the Clopper-Pearson method, but it no longer guarantees that the minimum coverage is at least as large as the nominal level $(1 - \alpha)$. It is clear in the middle graph of Figure 2.1, which plots the coverage of 97.5% one-sided upper interval for the mid-$p$ method for a bin (20, $p$) distribution, that the spikes spread fairly regularly around the nominal level. This method still tends to be slightly conservative, but much less than the Clopper-Pearson method. Agresti and Gottard (2007), Newcombe (1998), Agresti and Coull (1998) and Brown *et al.* (2001) recommend the mid-$p$ for practical purpose as it has good coverage (generally close to the nominal level) and good length performance.

#### 2.2.2.2 Jeffreys method

Jeffreys method is a Bayesian approach. A significant difference between this approach and the classical approach is that in the Bayesian approach the population parameter, $p$, is considered to be a random varaible while in the classical approach $p$ is considered to be a fixed unknown constant. Beta distributions are the conjugate priors for binomial distributions so it is quite common to use Beta priors for inference on $p$ (Berger, 1985). The Jeffreys intervals take a Beta (0.5,0.5) as the prior distribution. This distribution is Jeffreys' choice of noninformative prior distribution for sampling from a binomial model. Its density function is

$$f(p) = \frac{1}{\pi\sqrt{p(1-p)}}. \tag{2.5}$$

Suppose that the sample consists of $x$ successes in $n$ trails, then the posterior distribution for $p$ is Beta $(x + \frac{1}{2}, n - x + \frac{1}{2})$. The Jeffreys confidence interval is the $1-2\alpha$ equal-tailed credible interval given by this posterior distribution, except for setting $l_0 = 0$ and $u_n = 1$ (Brown *et al.*, 2001). Thus the $100(1 - 2\alpha)\%$ equal-tailed Jeffreys intervals are defined as

$$l_i = \text{Beta}_{\alpha}(i + 1/2, n - i + 1/2) \tag{2.6}$$

and

$$u_i = \text{Beta}_{1-\alpha}(i + 1/2, n - i + 1/2) \tag{2.7}$$

where $\text{Beta}_{\alpha}(i + 1/2, n - i + 1/2)$ is the $\alpha$ quantile and $\text{Beta}_{1-\alpha}(i + 1/2, n - i + 1/2)$ is the $1-\alpha$ quatile of the Beta distribution, Beta $(i+1/2, n-i+1/2)$ for $i = 1, \ldots, n-1$.

As mentioned earlier, the endpoints of the Clopper-Pearson interval are the $\alpha$ and $1-\alpha$ quantiles of the Beta $(X, n-X+1)$ distribution. It is pointed out in Brown *et al.* (2001) that Jeffreys intervals are always within Clopper-Pearson intervals, so it mitigates the conservativeness of the Clopper-Pearson interval.

The coverage of 97.5% upper-tail intervals for the Jeffreys method is given in the lowest graph in Figure 2.1. Its coverage is very similar to that of mid-$p$ intervals over most of the parameter space [0,1]. As noted in Brown et al., (2001, p.110):

> " Jeffreys confidence interval has an appealing connection to the mid-$p$ corrected version of the Clopper-Pearson "exact and correct" intervals. These are very similar to Jeffreys confidence interval, over most of the range, and have similar appealing properties."

Brown *et al.* (2001) recommend the Jeffreys method as a serious and credible candidate for practical use. However, it has the undesirable result that its coverage has a fairly deep spike near $p$=0 and $p$=1. They proposed a modification to the Jeffreys method to avoid the unfortunate downward spikes in the coverage near 0 and 1, but in this thesis we restrict our attention to the original Jeffreys method.

### 2.2.3 Asymptotic methods

For large sample sizes ($n$), Wald, Wilson and Agresti-Coull methods are commonly discussed in the literature. These methods use an asymptotic approximation to form the confidence interval, instead of using the discrete distribution. They are approximate methods and do not meet the definition of a confidence interval that the coverage probability should be equal to or greater than the nominal level, $1 - 2\alpha$. However the methods generally give shorter confidence intervals than the Clopper-Pearson method.

#### 2.2.3.1 Wald method

The most widely used asymptotic method of forming a confidence interval for the binomial proportion $p$ is the Wald method. This is the method taught

in introductory statistics courses. Agresti and Coull (1998) mention that this standard confidence interval results from inverting the Wald large sample test for $p$; that is, the interval is the set of $p_0$ values for which $H_0 : p = p_0$ is not rejected at the $2\alpha$ significance level in favour of $H_1 : p \neq p_0$, when using the test statistic $z = (\hat{p} - p_0)/\sqrt{\hat{p}(1-\hat{p})/n}$. Historically, this is surely one of the first confidence intervals proposed for any parameter (Laplace 1812, p.283). The endpoints of Wald confidence intervals are defined as:

$$l_i = \hat{p} - z_{1-\alpha}\sqrt{\hat{p}(1-\hat{p})/n} \tag{2.8}$$

and

$$u_i = \hat{p} + z_{1-\alpha}\sqrt{\hat{p}(1-\hat{p})/n} \tag{2.9}$$

where $\hat{p}$ is the point estimate of $p$, $\hat{p} = i/n$ and $z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution.

In practice, the Wald method is commonly used because it gives intervals that are easy to present and simple to compute. However, it produces intervals that are too narrow when samples are small. Also, its coverage is usually far below the nominal confidence level even for large sample sizes, especially when $p$ is near to 0 and 1 [e.g. Blyth and Still (1983), Vollset (1993)]. This seems clear in Figure 2.2, as the coverage of Wald intervals tends to be liberal for small values of $p$ and conservative for large values, both when the sample size is small (top graph, $n$=20) and when it is large (lower graph, $n$=200). Also, the lower graph shows that the coverage has very slow convergence to the nominal level as $n$ grows.

Vollset (1993) mentioned that there are two modifications of the Wald method that have been proposed to enhance the coverage performance. The simpler one is a continuity correction that gives the Wald continuity correct interval. The other one uses a different continuity correction and is given by

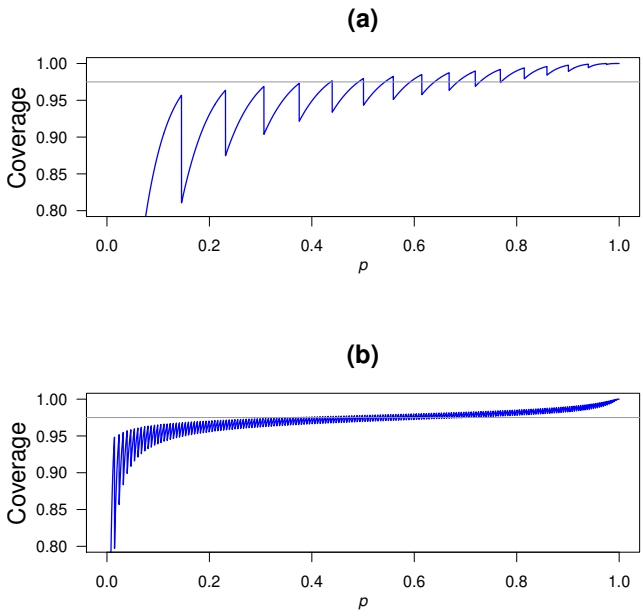


Figure 2.2: Coverage of upper one-sided 97.5% confidence interval for the Wald method for the success parameter $p$ for (a) fixed sample size ($n$) of 20 and (b) fixed sample size ($n$) of 200.

Blyth and Still (1983). The advantages and disadvantages of these corrections are minor and we only examine the standard Wald method in this thesis.

#### 2.2.3.2 Wilson (score) method

Wilson (1927) introduced an improved confidence interval that is similar to the Wald method. The Wilson interval is based on the inversion of the score test for $p$, so it is also known as the score interval. Whereas the Wald test is based on the log-likelihood at the maximum likelihood estimate ($\hat{p}$), the score test is based on the log-likelihood at the null-hypothesis value of the parameter ($p_0$) (Agresti (1996); Agresti and Coull (1998)). The interval endpoints are

$$p = \frac{1}{1 + z_{1-\alpha}^2/n}(\hat{p} + z_{1-\alpha}^2/2n \pm z_{1-\alpha}\sqrt{\hat{p}\hat{q}/n + z_{1-\alpha}^2/4n^2}). \tag{2.10}$$

The score interval has favourable coverage and length properties relative to the standard Wald interval. Brown *et al.* (2001) and Newcombe (1998) compared Wald intervals with score intervals in terms of coverage probability

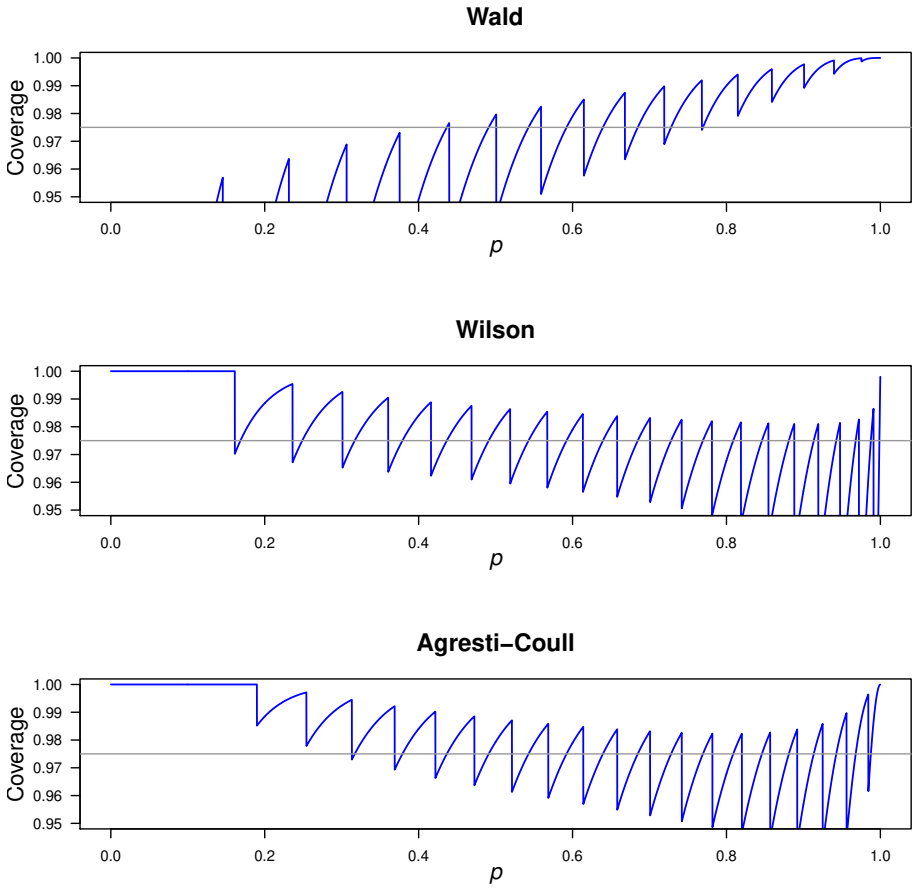


Figure 2.3: Coverage of upper one-sided 97.5% confidence interval for the Wald, Wilson and Agresti-Coull methods for a fixed sample size ($n$) of 20 and sucess parameter $p$.

and length. They found that the score intervals give a coverage probability that is close to the nominal level and the intervals are acceptably short. They recommend it over the Wald interval for all confidence intervals and sample sizes. A plot of the coverage of 97.5% upper-tail intervals against $p$ for a binomial (20,$p$) distribution was given in the middle graph of Figure 2.3. The coverage is a little conservative for small values of $p$ and quite liberal for values of $p$ near 1, which means that it does not meet the definition of a confidence interval. There is also a continuity corrected version of the score (SCC) interval that is considered in Vollset (1993), but we will only examine the original score interval.

#### 2.2.3.3 Agresti-Coull method

The Agresti-Coull (AC) method is a comparatively new way of forming a confidence interval. It was suggested by Agresti and Coull (1998) as a simple adjustment to the Wald method that appreciably improves the coverage of the 95% confidence intervals. The adjustment is to add two"successes" and two "failures" to the sample and then use the formula for the Wald interval, equation (2.8) and (2.9). Putting $\widetilde{p} = (i+2)/(n+4)$, the endpoints of its confidence interval are

$$\widetilde{p} \pm z_{1-\alpha}\sqrt{\frac{\widetilde{p}(1-\widetilde{p})}{n+4}}. \tag{2.11}$$

The AC interval has a reasonable minimum coverage probability on average but its coverage is often less than the nominal level, so it also does not meet the definition of the confidence interval and only gives approximate confidence intervals. Figure 2.3 illustrates that this simple adjustment to the standard Wald method reverses its tendancy to be liberal for small values of $p$ and conservative for large values of $p$. In general, Agresti-Coull intervals are a little more conservative than those of the score method. Both Agresti-Coull and score intervals are centred on almost the same midpoint, $\widetilde{p} = (i+2)/(n+4)$,

and an Agresti-Coull interval can never be shorter than the corresponding score interval - an Agresti-Coull interval always contains the corresponding score interval, (Agresti and Coull, 1998). Brown *et al.* (2001) recommend Agresti-Coull intervals as a simple method of forming confidence intervals for a binomial proportion when the sample size exceeds 40.

### 2.2.4 Desirable properties of confidence intervals

The following properties have been proposed in the literature as desirable attributes in a method of forming a confidence interval for the binomial proportion [see, for example, Blyth and Still (1983) and Schilling and Doi (2014)].

**Property 1**. *Interval valued.* A confidence region should be an interval and not a collection of disjoint intervals.

The remaining properties assume that the confidence region is a two-tail interval. When $X = x$, the sample size is $n$ and the confidence interval is $1-2\alpha$. Denote this interval as $(l(x, n, \alpha), u(x, n, \alpha))$.

**Property 2**. *Equivariance.* As a binomail distribution is invariant under the transformation $X \to n - x; p \to 1 - p$, a confidence interval should also be invariant under this transformation. That is, if $x$ generates the confidence interval $[l(x, n, \alpha), u(x, n, \alpha)]$, then $n - x$ yields the confidence interval $[1 - u(n - x, n, \alpha), 1 - l(n - x, n, \alpha)]$ for $x = 0, \ldots, n$.

**Property 3**. *Monotonicity in $x$.* For fixed $n$ and $\alpha$, the endpoints should be increasing in $x$. This requires $l(x + 1, n, \alpha) > l(x, n, \alpha)$ and $u(x + 1, n, \alpha) > u(x, n, \alpha)$. For example, when $n = 20$ and $x = 7$, both the upper and lower endpoint should be greater than their corresponding values when $n = 20$ and $x = 6$.

**Property 4**. *Monotonicity in n.* For given $x$ and $\alpha$, there are two possibilities: (a) If an additional trial resulted in success, both limits of the confidence interval should increase. This requires $l(x+1, n+1, \alpha) > l(x, n, \alpha)$ and $u(x+1, n+1, \alpha) > u(x, n, \alpha)$. For example when $n$=21 and $x = 7$, the endpoints should be greater than their corresponding endpoints when $n = 20$ and $x = 6$. (b) If an additional trial resulted in failure, $x$ is fixed, the lower endpoint should be non- increasing in $n$ and the upper endpoint should be decreasing in $n$. This requires $l(x, n+1, \alpha) \leq l(x, n, \alpha)$ and $u(x, n+1, \alpha) < u(x, n, \alpha)$. For example when $n = 21$ and $x = 6$, the lower endpoint should be less than or equal to their corresponding values when $n = 20$ and $x = 6$ and the upper endpoint should be less than their corresponding values when $n = 20$ and $x = 6$.

**Property 5**. *Nesting.* If two confidence intervals have different confidence levels then, for any given $n$ and $x$, the interval with the higher confidence level should contain the interval with the lower confidence level. Suppose we have two confidence levels $1 - \alpha_1$ and $1 - \alpha_2$ with $\alpha_1 < \alpha_2$. Then this requires $(l(x, n, \alpha_2), u(x, n, \alpha_2)) \in (l(x, n, \alpha_1), u(x, n, \alpha_1))$. For this to occur for all confidence levels, as the level increases the lower limit for each $x$ must be non-increasing and the upper limit must be non-decreasing.

All the methods described earlier in this chapter give confidence intervals without any disjoint points and hence have property 1. They also have property 2, the equivariance property, and property 3 the montonicity in $x$ property. However, while the Clopper-Pearson, mid-$p$ and Jeffreys methods have property 4, monotonicity in $n$, the Wald, Wilson and Agresti-Coull methods do not have this property. The methods all have property 5.

### 2.2.5 Other methods: Randomized confidence intervals and unequal tails

To avoid the conservativeness of the Clopper-Pearson intervals, Stevens (1950) suggested a randomized confidence interval. The method randomly generates a value, $v$ say, from a uniform $u(0,1)$ distribution and adds this to $i$, the observed number of success. The confidence interval is then determined from the value of $i+v$. As this is the value of a random variable that has a continuous distribution, problem that arise with discrete distribution are avoided. Specifically, the upper point is the value of $p$ that satisfies $(1-v)(Pr(x<i|p))+v(Pr(x\leq i|p))=\alpha$ and the lower endpoint satisfies $(1-v)(Pr(x\geq i|p))+v(Pr(x>i|p))=\alpha$. These days statisticians regard randomized inference as a tool for the mathematical convenience of achieving exactly the confidence level with discrete data, but they do not consider actually implementing it in practice, (Agresti and Gottard, 2005). Stevens (1950) stated,

> "We suppose that most people will find repugnant the idea of adding yet another random element to a result which is already subject to the errors of random sampling. But what one is really doing is to eliminate one uncertainty by introducing a new one. The uncertainty which is eliminated is that of the true probability that the parameter lies within the calculated interval. It is because this uncertainty is eliminated that we no longer have to keep 'on the safe side' and can, therefore, reduce the width of the interval."

Other intervals have been designed to improve either coverage or length of the Clopper-Pearson interval, by inverting two-sided tests that do not need to be equal-tailed. Sterne (1954) proposed a confidence interval that inverts the ex-

act binomial test whose acceptance region includes the most probable values of the binomial variable. It does this by including the most probable one, then the next most probable value, and so on until the total probability reaches the desired confidence level. Although this approach results in nested intervals that are shorter than the Clopper-Pearson interval, it sometimes produces two separate intervals rather than one connected interval. Crow (1956) noticed the problem of the Sterne interval and corrected it. But this modification of Crow's interval did not enhance the performance of the interval. To improve performance, Blyth and Still (1983) started from the Sterne interval and, using a complicated method, constructed an interval that met further monotonic and smooth conditions. Casella (1986) proposed a refinement procedure to the Blyth and Still procedure, called the Blyth-Still-Casella interval. The method is guaranteed to give the shortest exact interval, but has the strange property that its intervals are not nested. Blaker (2000) proposed an exact nested interval that is always contained in the Clopper-Pearson interval, but it is wider than the Blyth-Still-Casella interval (Zhao, 2005).

There are other ways to handle the gap problem of Sterne's procedure. One of them is given by Reiczigel (2003), who simply fill the gaps by introducing a computer-intensive level-adjustment procedure to improve the Clopper-Pearson method, but the method seems too hard to apply. Also, Schilling and Doi (2014) presented an alternative strategy that avoids gaps in Sterne's procedure and still produces a strict length minimizing procedure that maximizes the coverage. It is called the LCO method, and manages to combine length minimization with maximal coverage but, of course it gives unequal-tailed intervals. Decrouez and Hall (2014) proposed a method of splitting the original sample size $n$ into two parts, $n_1$ and $n_2 = n - n_1$. Usually the saw-tooth pattern of the coverage probability is largely changed and removed, though

this depends on the average of the means of these two subsamples.

Park and Leemis (2017) devised a confidence interval for $p$ based on the actual coverage function that combines several exciting approximate confidence intervals. They combine five methods of forming confidence intervals that are part of most statistical packages: the Clopper-Pearson, Wilson, Jeffryes, Agresti-Coull methods and a method that uses the arcsine transformation (Anscombe, 1956). They found that the new combined method gave an ensemble confidence interval that does not act as a good interval estimate when $n$ is large, but does well for small $n$, for example, $n$=10.

## 2.3 Poisson distribution confidence intervals

After the binomial distribution, the discrete distribution that has attracted most attention in the literature is the Poisson distribution. A Poisson distribution is simpler than the binomial distribtuion, as the latter is characterized by the values of two parameters, $n$ and $p$, whereas a Poisson has only one parameter, the mean which we denote by $\lambda$. Estimation of the Poisson mean $\lambda$ is required in a wide variety of phenomena that deal with counts of rare events, especially in biomedical and epidemiology applications. There are fewer different methods of forming a confidence interval for a Poisson mean $\lambda$ than for a binomial proportion $p$. For only a few of these, which will be described in the following sections, is much known about the methods. As with the binomial distribution, the methods are often classified in the literature as exact methods or approximate methods. Here the same interpretations will be used as in the binomial section.

The data consist of a single observation of the random variable $X$ that follow

the Poisson distribution with mean $\lambda$, where $\lambda$ is a positive real number, so

$$Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \tag{2.12}$$

for $x = 0, 1, 2, \ldots$. A $(1 - \alpha)$ equal-tailed confidence interval is required and $(\lambda_l, \lambda_u)$ is the interval that is constructed. [If more than one observation is taken from a Poisson, the sum of the observations follows a Poisson, so procedures to form a confidence interval based on the sum and are essentially unchanged.]

The Garwood and mid-$p$ methods, which will be discussed in more detail in the next section, are exact in sense (i) but only the Garwood is exact in both senses (i) and (ii). The Wald and score methods, which will be discussed in details in Section 2.3.3, are asymptotic methods while both the mid-$p$ and Jeffreys methods are exact methods. Apart from the Garwood method, they are all approximate methods.

Most of the methods that will be described in Sections 2.3.1, 2.3.2 and 2.3.3 are compared in studies by Barker (2002), Byrne and Kabaila (2005), Swift (2009), Patil and Kulkarni (2012) and Nadarajah *et al.* (2015). Because high coverage and narrow intervals are desirable criteria, the comparison between methods is based on coverage probability and average width. In Sections 2.3.1, 2.3.2 and 2.3.3, the methods that are most widely used in practice are discussed. We are interested in reviewing methods that give two-sided intervals with equal-tails, as equal-tailed intervals are much preferred in practice. Desirable properties in methods of forming confidence regions are considered in Section 2.3.4. We briefly review methods that do not aim to give two-sided confidence intervals with unequal-tails in Section 2.4.5.

### 2.3.1 Correct methods

#### 2.3.1.1 Garwood method

A commonly used method for constructing two-sided confidence interval for the Poisson mean $\lambda$ is Garwood (1936). It is based on inverting an equal-tailed test for the null hypothesis $H_0 : \lambda = \lambda_0$. The Garwood method is an exact method as its intervals are constructed using the exact cumulative probabilities of the Poisson distribution (Patil and Kulkarni, 2012).

With the Garwood method, when $x = i$ the lower limit is given by the value of $\lambda_l$ such that

$$Pr(X \geq i|\lambda = \lambda_l) = \sum_{x=0}^{i}(e^{-\lambda}\lambda^x)/x! = \alpha \tag{2.13}$$

and the upper limit is given by the value of $\lambda_u$ such that

$$Pr(X \leq i|\lambda = \lambda_u) = \sum_{x=i}^{\infty}(e^{-\lambda}\lambda^x)/x! = \alpha. \tag{2.14}$$

In 1936, a problem with using this approach was the difficulty in computing the cumulative Poisson probability. Garwood used the relationship

$$\sum_{x=0}^{i}(e^{-\lambda}\lambda^x)/x! = Pr\{\chi_v^2 > 2\lambda\}, \tag{2.15}$$

where $\chi_v^2$ is a random variable having a chi-sqaure distribution with $v = 2(1 + x)$ degrees of freedom (Sahai and Kurshid, 1993). Thus, we can solve the equations

$$Pr(\chi_{2x}^2 > 2\lambda_l) = 1 - \alpha \tag{2.16}$$

and

$$Pr(\chi_{2(x+1)}^2 > 2\lambda_u) = \alpha \tag{2.17}$$

to get the Garwood confidence limits,

$$\lambda_l = \frac{1}{2}\chi_{2x,\alpha}^2 \tag{2.18}$$

and

$$\lambda_u = \frac{1}{2}\chi^2_{2(x+1),1-\alpha}. \tag{2.19}$$

The Garwood method is an exact method, and it is the only method among the methods considered here that is correct. It is defined as " correct " because it strictly meets the defintion of a method for forming confidence intervals, satisfying equation (1.1).

Its underlaying rationale is similar to that of the Clopper-Pearson method, the gold standard method for the binomial case. Hence as one would expect, the Garwood method has intervals whose coverage probability, for almost all values of $\lambda$, is larger than the nominal level. Thus, it suffers from conservatism [Garwood, 1936, Sahai and Kurshid,1993, Barke, 2002].

Figure 2.4 plots the coverage of its 97.5% one-sided upper interval in the the top graph and its lower interval in the lower graph against $\lambda$ for $x = 0, 1, 2, \ldots, 20$. The coverage, for both upper and lower tails, is substantially above 98.5% almost all the time, sometimes exceeding 99%.

Thus, although the Garwood method achieves the definition of a confidence interval it is conservative and gives intervals that are wider than necessary. This has led to alternative methods being suggested that are not exact in sense (i), but which give shorter intervals than the Garwood method. Some of these methods are described in the following subsections.

### 2.3.2 Approximate methods

As in the binomial case, neither the mid-$p$ nor Jeffreys methods meet the definition of a confidence interval in equation (1.1), but they are exact in sense (ii) as they do not use asymptotic approximations. The mid-$p$ method reduces conservatism by using half of the probability of the observed result,

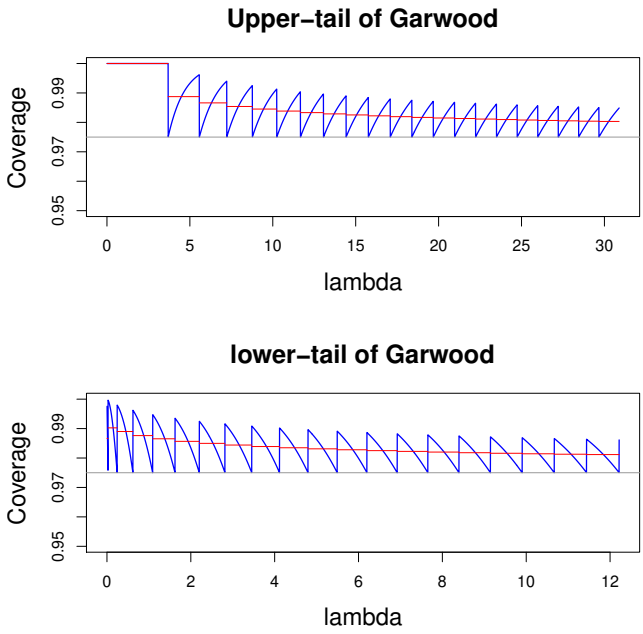


Figure 2.4: Coverage of (a) upper one-sided 97.5% confidence interval in the top graph, (b) lower one-sided 97.5% confidence interval in the lower graph for the Garwood method for $x = 0, 1, 2, \ldots, 20$.

while Jeffreys method forms a Bayesian credible interval starting with a non-informative conjugate prior for the Poisson distribution.

#### 2.3.2.1 Mid-$p$ method

The mid-$p$ that was used to form confidence intervals for the binomial distribution can, with slight modification, be used for a Poisson distribution. Lancaster (1961) suggested it as a means of avoiding the conservatism of the Garwood method (Cohen and Yang, 1994; Swift, 2009). As with the binomial distribution, it halves the probability of the observed result and in other respects is similar to the Garwood method (which has the rationale of the Clopper-Pearson method). Thus, to construct the confidence interval $[l_x, u_x]$, the lower limit is chosen to satisfy

$$Pr(X > x; \lambda) + \frac{1}{2} Pr(X = x; \lambda) = \alpha \tag{2.20}$$

and the upper limit satisfies

$$Pr(X < x; \lambda) + \frac{1}{2}Pr(X = x; \lambda) = \alpha. \tag{2.21}$$

(Equations (2.20) and (2.21) are essentially the same as equations (2.3) and (2.4).)

The mid-$p$ method reduces the conservatism of the Garwood method, but it no longer guarantees that the minimum coverage is at least as large as the nominal level $(1 - \alpha)$. The left-hand diagrams in Figure 2.5 and Figure 2.6 plot the coverage of the 97.5% one-sided upper interval and lower interval, rescpectively, for the mid-$p$ method. The spikes in the plots are spread fairly regularly around the nominal level. This method still tends to be slightly conservative, but much less than the Garwood method. Cohen and Yang (1994) and Swift (2009) recommended the mid-$p$ method for practical purposes as it has good coverage (generally close to the nominal level) and good length performance.

#### 2.3.2.2 Jeffreys method

Brown *et al.* (2003) presented Bayesian credible intervals constructed from the non-informative Jeffreys prior, which is proportional to $\lambda^{-1/2}$. The posterior distribution of $\lambda$ is $\lambda|x \sim \text{Gamma}(x + \frac{1}{2}, 1)$ which is proper. Therefore endpoints of the $100(1 - 2\alpha)\%$ equal-tailed Jeffreys interval are given by

$$l_x = Gamma(\alpha, x + \frac{1}{2}) \tag{2.22}$$

and

$$u_x = Gamma(1 - \alpha, x + \frac{1}{2}). \tag{2.23}$$

Thus $l_x$ is the $\alpha$ quantile and $u_x$ is the $1-\alpha$ quantile of the Gamma distribution, Gamma $(x + \frac{1}{2})$.

Jeffreys method gives intervals that are more liberal than the mid-$p$ method. This can be seen in the right-hand diagrams in both Figures 2.5 and 2.6.

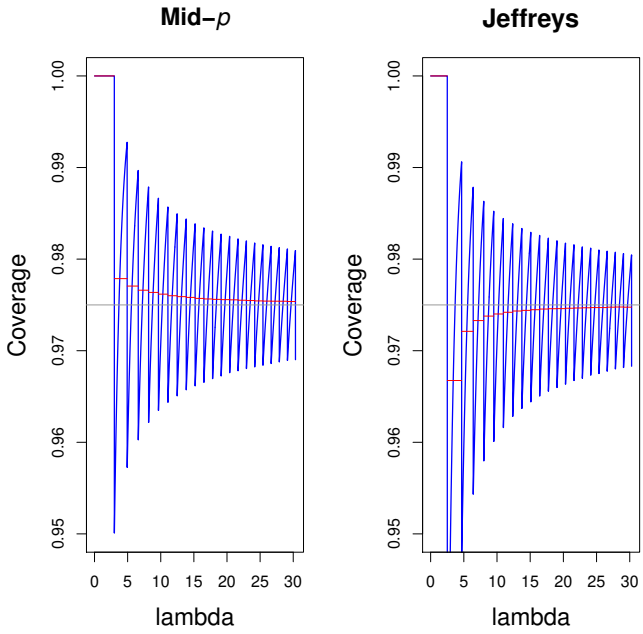


Figure 2.5: Coverage of upper one-sided 97.5% confidence intervals for the mid-$p$ and Jeffreys methods for $x = 0, 1, 2, \ldots, 20$.

In particular, Jeffreys intervals have a lower coverage for small values of $\lambda$. Brown *et al.* (2003) recommended the Jeffreys method as a good and credible candidate for practical use because of its better length properties. However it has the undesirable results that its coverage can be quite low for small values of $\lambda$.

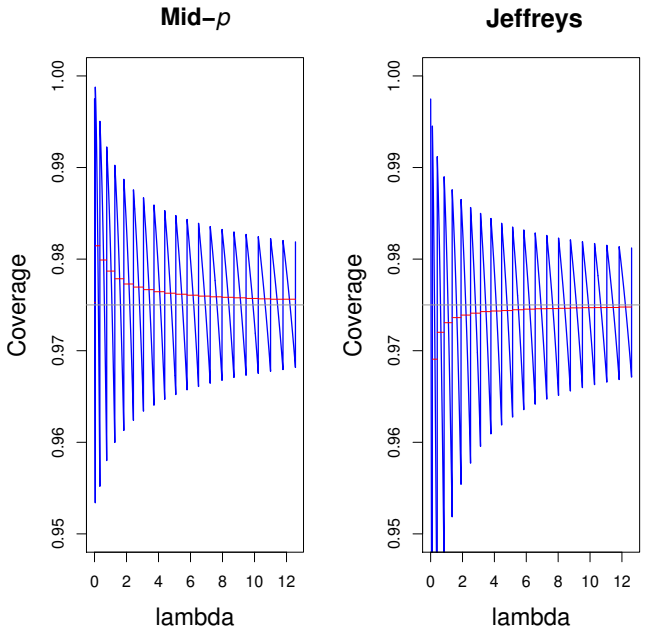


Figure 2.6: Coverage of lower one-sided 97.5% confidence intervals for the Mid-$p$ and Jeffreys methods for $x = 0, 1, 2, \ldots, 20$.

### 2.3.3 Asymptotic methods

A strand of literature suggests using the Wald and score methods, as in the binomial distribution, e.g. (Brown *et al.*, 2003), (Patil and Kulkarni, 2012). These methods are a good choice for large values of $\lambda$, where they use an asymptotic approximation to form the confidence interval. Although the methods do not meet the definition of the confidence interval, they generally give shorter confidence intervals than the Garwood method.

#### 2.3.3.1 Wald method

The simplest and most widely used method of obtaining the confidence interval for a Poisson mean, as in the binomial case, is the Wald method. It is introduced in a large number of introductory statistics courses. The idea of the Wald method is to use the normal approximation to the Poisson distribution. Its interval limits are given by Barker (2002) as

$$l_x = x - Z_{1-\alpha}\sqrt{x} \tag{2.24}$$

and

$$u_x = x + Z_{1-\alpha}\sqrt{x}, \tag{2.25}$$

where $Z_{1-\alpha}$ is the $(1-\alpha)100\%$ percentile of the standard normal distribution (Liu, 2012).

In practice, the Wald method is commonly used because it gives intervals that are easy to present and simple to compute. However, it produces intervals that are too narrow when the value of $\lambda$ is small. Also, its upper-tail coverage usually tends to be far below the nominal confidence level at small values of $\lambda$ and close to the nominal level at large values of $\lambda$. Its lower-tail coverage follows a similar pattern but in the opposite direction, as it tends to be far above the nominal level for small values of $\lambda$ and close to the nominal level at large values of $\lambda$. This can be seen clearly in the top graphs of Figures 2.7 and 2.8.

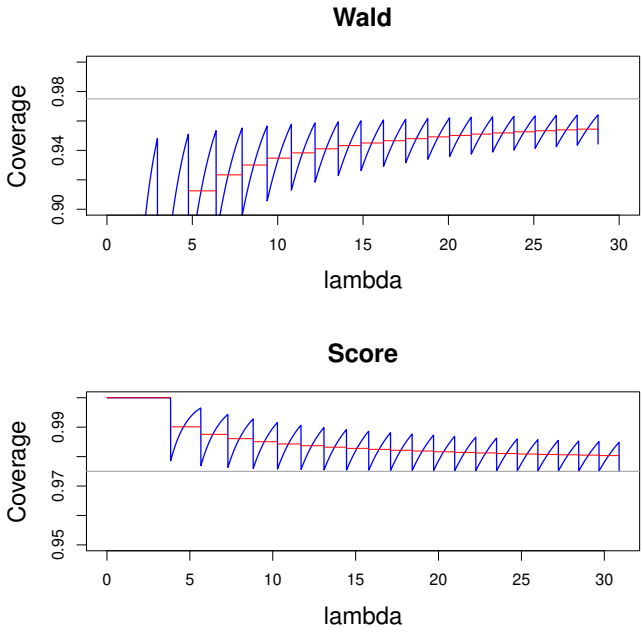


Figure 2.7: Coverage of upper one-sided 97.5% confidence intervals for the Wald and score methods for $x = 0, 1, 2, \ldots, 20$.

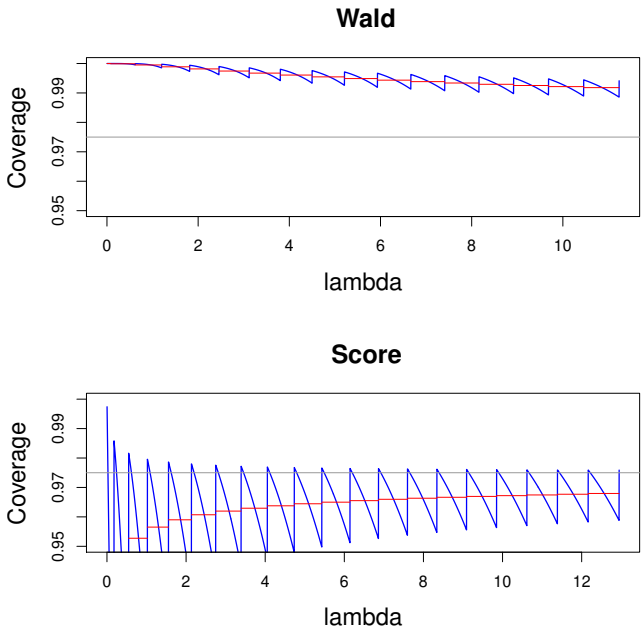


Figure 2.8: Coverage of lower one-sided 97.5% confidence intervals for the Wald and score methods for $x = 0, 1, 2, \ldots, 20$.

Patil and Kulkarni (2012) mentioned that there are modifications to the Wald method that have been proposed to enhance the coverage performance. A simple modification is a continuity correction that gives the Wald continuity corrected interval. Other modifications to the Wald method are given by Barker (2002) and Khamkong (2012). The advantages and disadvantages of these modifications are minor and we will only be interested in the standard Wald method in this thesis.

#### 2.3.3.2 Score method

The score method for the Poisson distribution is another method that uses a normal approximation. Like the score method for the binomial distribution, its interval is formed by inverting Rao's hypothesis test for equality (Rao, 1973), which is now $H_0 : \lambda = \lambda_0$. The null hypothesis $H_0$ is not rejected on the basis of Rao's score test if and only if $\lambda_0$ is in the confidence interval. Barker (2002) gives the endpoints of the score confidence interval as

$$x + \frac{z^2_{1-\alpha}}{2} \pm \sqrt{xz^2_{1-\alpha} + \frac{z^4_{1-\alpha}}{4}}. \tag{2.26}$$

The interval has favourable coverage and length properties in comparison with the Wald interval. Barker (2002), Brown *et al.* (2003) and Patil and Rajarshi (2010) compared score intervals with Wald intervals in terms of coverage probability and length. They found that the score intervals give a coverage probability that is close to the nominal level. Unlike the Wald method, the upper-tail coverage of the score method tends to be a little conservative for small values of $\lambda$ and gets closer to the nominal level as the value of $\lambda$ is increased. The coverage of its lower-tail interval is somewhat below the nominal level. That means the score method does not meet the definition of a confidence interval. The coverage of both the 97.5% upper and lower tail intervals given by the score method are plotted in the lower graphs of Figures 2.7 and 2.8, respectively.

Although the score method provides a major improvement in coverage compared to the Wald method, it suffers from giving intervals that are a little too long (Brown *et al.*, 2003). Because it has sharp downward and upward spikes for small means, Guan (2011) proposed a method that moves the score interval to the left a little (about 0.04 units) to solve this problem. They called the method the moved score method. However, we will only examine the original score interval.

### 2.3.4 Desirable properties of confidence intervals

As noted in Section 2.2.4, a number of desirable properties in methods of forming a confidence interval have been proposed. The properties assume that the confidence region is a two-tail interval. When $X = x$ and the confidence in-

terval is $1-2\alpha$, denote this interval as $(l(x,\alpha), u(x,\alpha))$.

All the methods described earlier in this section give confidence regions that are single intervals and not a collection of disjoint intervals and so they have Property 1. They also have Property 2, the monotonicity in $x$ property. That is $l(x+1,\alpha) > l(x,\alpha)$ and $u(x+1,\alpha) > u(x,\alpha)$. They also have the Property 3, the nesting property: if $\alpha_1 < \alpha_2$, then $(l(x,\alpha_2), u(x,\alpha_2)) \in (l(x,\alpha_1), u(x,\alpha_1))$.

### 2.3.5 Other methods

Other methods have been designed to improve either coverage or length of the Garwood interval, by inverting two-sided tests that do not need to be equal-tailed. That means they do not achieve the strong condition that $P_l = P_u = \alpha$ but achieve $P_l + P_u = 2\alpha$, where $P_l = Pr(X \geq x|\lambda = l_x)$ and $P_u = Pr(X \leq x|\lambda = u_x)$. To avoid the conservativeness of the Garwood intervals, other researchers such as Sterne (1954), Crow and Gardner (1959), Casella and Robert (1989), Kabaila and Byrne (2001) tried to shorten the intervals as much as possible by using non-central confidence intervals. Sterne (1954) used the same idea that is used with the binomial distribution, forming a confidence interval by inverting the exact Poisson test with an acceptance region that includes the most probable values of the Poisson variable. Thus the interval icludes the most probable value, then the next one and so on until the total probability reaches the desired nominal level (Swift, 2009). Crow and Gardner (1959) considered values of $\lambda$ from smallest to largest and, for each value of $\lambda$, they found $y$ and $z$ values that would satisfy $P_{z,y}(\lambda) \geq 1-\alpha$, where $P_{z,y}(\lambda) = Pr(y \geq X \geq z|\lambda)$ (Byrne and Kabaila, 2005).

Casella and Robert (1989) gave a refinement method that works with an input set of $1-\alpha$ confidence limits. The initial input set could be Garwood inter-

vals, for example. The main point of the refinement method is to decrease upper endpoints and increase lower endpoints until any movement in one of them reduces the coverage probability to less than the nominal confidence level (Swift, 2009). This method is considered to be a generalization of the Crow and Gardner (1959) method.

Unlike Crow and Gardner (1959) and Casella and Robert (1989), Kabaila and Byrne (2001) provided an algorithm that does not require refinement of an existing confidence interval. It is a direct method as it yields endpoints of the confidence interval that are strictly increasing functions of the observed variable. But the interval cannot be shortened without the coverage probability falling below $(1-\alpha)$. There is also Blaker (2000), who provided exact $(1-\alpha)$ confidence intervals. He improved the Garwood intervals by using an acceptability function. Garwood's lower limits are increased while its upper limits are decreased until the acceptability function reaches $(1-\alpha)$. Thus, the resulting intervals are a smaller subset nested within the Garwood intervals (Swift, 2009).

Holladay (2014) tried to find the optimal confidence interval for the Poisson mean $\lambda$ and introduced three new methods that are considered as optimal methods according to the "Inability to be shortened property" of Kabaila and Byrne (Byrne and Kabaila, 2005). These methods are modifications of the Sterne method, the least cardinality percentage method and the modified Crow and Gardner method. To derive one of these methods, he follows a different strategy to what would usually be done. He chooses the ideal coverage probability function first and then find the confidence intervals that would be given by this function. To achieve this, he created a specialized coverage probability function through an exhaustive graphical examination of all Poisson probability functions for a set of consecutive values. Then, confi-

dence intervals for $\lambda$ with any desired confidence level can be formed for all possible values of an observed event $x$. After comparing these methods with other common methods, he mentioned that no method is better than all others concerning coverage and length. However, these methods deserve serious consideration (Holladay, 2014).

Schilling and Holladay (2017) adapted the approach of Schilling and Doi (2014) for the binomial distribution to make it suitable for the Poisson case. They provided an alternative criterion, which is the minimal cardinality property, for comparing the length performance of Poisson confidence procedures. Then they identified an optimal minimal cardinality procedure depending on this criterion and compared their method with the method of Crow and Gardner (1959) and the modified Crow and Gardner methods (González *et al.*, 2020).

## 2.4 Negative binomial distribution confidence intervals

Our research extends to a third discrete distribution, which is the negative binomial distribution. The negative binomial distribution concerns the number of Bernoulli trials that must occur in order to have a predetermined number of successes. Galloway (1839) was the first to present the negative binomial distribution as a probability distribution function, as given in equation (2.27). We can see that the negative binomial distribution is related to the binomial distribution and generalizes the geometric distribution. The negative binomial distribution has two parameters $(p, r)$, which is almost the same as a binomial distribution but with one important difference; a binomial distribution counts the number of successes in a fixed number of Bernoulli trials. With the bino-

mial distribution the possible values of $x$ are $0, 1, 2, \ldots, n$, where $n$ is finite. In converse, the negative binomial distribution is concerned with the number of Bernoulli trials, $y$ where $y = r + x$, where $r$ is the number of successes and $x$ is the number of failures that must occur until we have $r$ successes, so the values of $y$ are $r, r + 1, r + 2, \ldots$, with no upper limit. The geometric distribution is a special case of negative binomial, where the trials are stopped at the first success $r = 1$.

A negative binomial distribution arises, most commonly, as the probability distribution for the number of failures $(x)$ that will occur before the rth success is observed, when the probability of success on each trial is fixed at $p$ and $y = r + x$ is the number of trials. Then

$$Pr_p[X = x|r, p] = \binom{x + r - 1}{r - 1} p^r (1-p)^x, \quad x = 0, 1, 2, \ldots, \quad r > 0, \quad 0 < p < 1, \tag{2.27}$$

This distribution has two parameters and is denoted as negative binomial $(r, p)$. The mean of negative binomial is

$$\mu = \frac{r(1 - p)}{p} \tag{2.28}$$

and its variance is

$$\sigma^2 = \frac{r(1 - p)}{p^2}. \tag{2.29}$$

Only a few different methods of forming a confidence interval for a negative binomial parameter $p$ have been studied. Tian *et al.* (2009) compare seven methods, Choi (2015) compares four methods and Young (2014) examines eight methods. We are interested in the five most common methods of calculating the confidence interval for the negative binomial proportion $p$, and these will be discussed in detail in the following sections. As well as $p$, we are also interested in calculating confidence intervals for the negative binomial mean $\mu$, a task that has attracted some attention. Kabaila and Byrne (2001)

present an algorithm for finding a $1-\alpha$ confidence interval for $\mu$. Also, Arefi *et al.* (2009) present some common methods of calculating the confidence interval for it.

As with binomial and Poisson distributions, the methods are often classified in the literature as exact and approximate methods. Here these terms will have the same interpretations as they were given in both the binomial and Poisson sections. Both the Casella and McCulloch and mid-$p$ methods, which will be discussed in details in the next section, are exact in sense (i) but only the Casella and McCulloch method is exact in both senses (i) and (ii). The Wald and score methods, which will be discussed in details in Section 2.4.3, are asymptotic methods while both the mid-$p$ and Jeffreys methods are exact methods. Apart from the Casella and McCulloch method, they are all approximate methods.

Most of the methods, which will be described in Sections 2.4.1, 2.4.2 and 2.4.3, are compared in terms of coverage probability and average width. We are interested in reviewing the methods that give two-sided intervals with equal-tails, as that is the focus of this thesis. In Section 2.4.4, desirable properties in methods of forming confidence regions are considered. Then, we briefly review some other methods of constructing confidence intervals in Section 2.4.5. These aim to improve either the coverage or length of intervals.

### 2.4.1 Correct methods

#### 2.4.1.1 The Casella and McCulloch method

As mentioned earlier, the most well-known exact confidence interval method was presented by Clopper-Pearson (1934) for the binomial parameter $p$. Then Garwood (1936) used this method to develop an exact confidence procedure

for the parameter of a Poisson distribution, $\lambda$. The method has also been adopted by Casella and McCulloch (1984) to derive confidence interval limits for the parameter $p$ of a negative binomial $(r, p)$ distribution. The method guarantees a coverage probability of at least $1 - \alpha$ for every value of $p$. The $(1 - \alpha)100\%$ exact interval is defined as $(l_x, u_x)$, where

$$l_x = \frac{1}{1 + (\frac{x+1}{r})F_{2(x+1),2r,\alpha}} \tag{2.30}$$

and

$$u_x = \frac{\frac{r}{x}F_{2r,2x,\alpha}}{1 + (\frac{r}{x})F_{2x,2r,\alpha}}, \tag{2.31}$$

where $r$ is the number of success, $x$ is the number of failures and $F_{df_1,df_2,q}$ is the $qth$ quantile of an $F$ distribution with degrees of freedom $df_1$ and $df_2$.

The Casella and McCulloch method is an exact method, and it is correct, as it strictly meets the definition of a method for forming confidence intervals, satisfing equation (1.1). It is the only method among the methods considered here that is correct.

As with the Clopper-Pearson method in the binomial case and the Garwood method in Poisson case, the Casella and McCulloch method has intervals whose coverage probability, for almost all values of $p$, is larger than the nominal level. Thus it suffers from conservatism (Casella and McCulloch, 1984; Liu, 2012).

Figure 2.9 plots the coverage of its 97.5% one-sided upper-tail interval in the top graph and its lower interval in the lower graph against $p$ for $x$=20. It is clear that the coverage, for both the upper and lower tails, is above 98.5% for almost all values of $p$ and exceeds 99% for a few values of $p$. Hence, although it meets the definition of a confidence interval, it has conservative intervals which are usually wider than necessary. This leads to alternative methods being suggested that give shorter intervals than the Casella and McCulloch

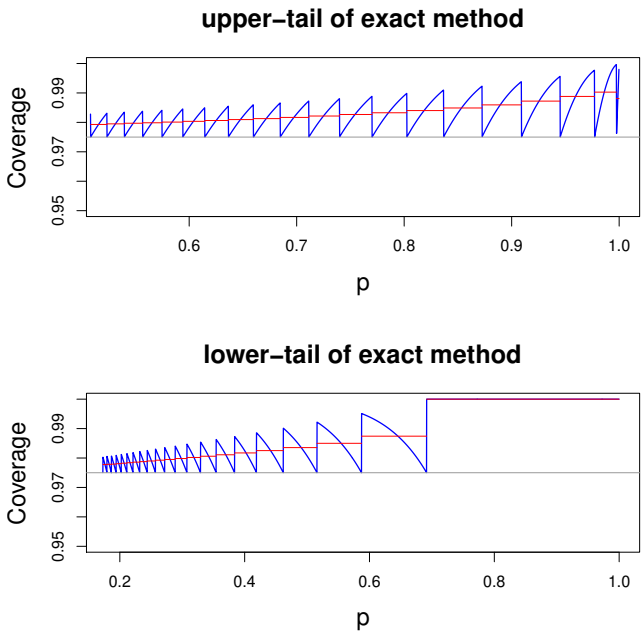


Figure 2.9: Coverage of upper and lower one-sided 97.5% confidence intervals for the negative binomial proportion $p$ for the exact method when $x$=20

method. Some of these methods are described in the following subsections.

As we mentioned earlier, we also are interested in calculating the confidence interval for the mean of the negative binomial, $\mu$. Most researchers, e.g. Liu (2012), calculate it by using the monotonic transformation from $p$ to $\mu$, $\mu = r(1-p)/p$, calculating the endpoints of the interval for $\mu$ from the endpoints of the interval for $p$. Figure 2.10 plots the coverage of its 97.5% one-sided upper-tail interval in the top graph and its lower interval in the lower graph against $\mu$ for $x$=20. Although these intervals are as conservative as the intervals constructed for $p$, they have the opposite pattern to the intervals of $p$, because large values of $p$ correspond to small values of $\mu$, and vice-versa. For both upper and lower tails, the coverage is substantially above 98.5% almost all the time and sometimes exceeds 99%. Thus, the intervals suffer from conservatism and they are wider than necessary.

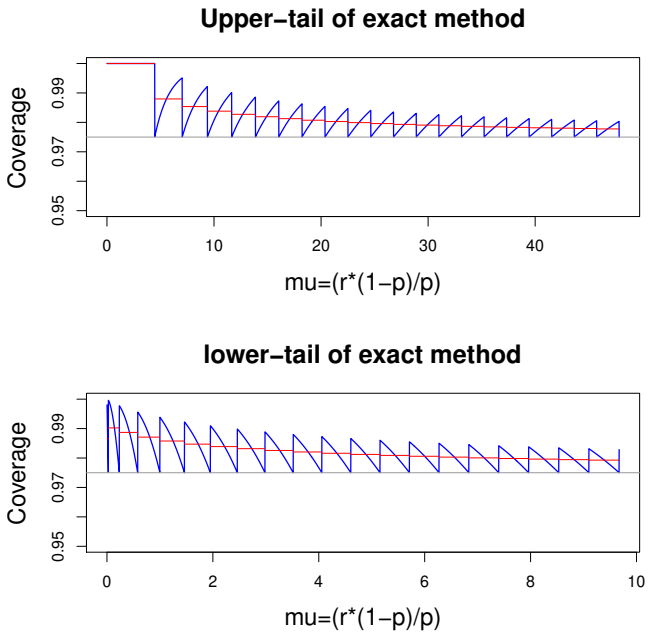


Figure 2.10: Coverage of upper and lower one-sided 97.5% confidence intervals for the negative binomial mean $\mu$ for the exact method when $x$=20

### 2.4.2 Approximate methods

As in both the binomial and Poisson cases, mid-$p$ and Jeffreys methods do not meet the definition of a confidence interval in equation (1.1), but they are exact in sense (ii) as they do not use asymptotic approximations. The mid-$p$ method reduces the conservatism of the exact method by using half of the probability of the observed result, while Jeffreys method is constructed as a Bayesian credible interval, based on the Bayesian posterior distribution of $p$, $p|x \sim Beta(r + \frac{1}{2}, x + \frac{1}{2})$, using the prior $p \propto p^{-1}(1-p)^{-1/2}$ (Cai, 2005).

#### 2.4.2.1 Mid-$p$ method

Although the mid-$p$ method is a very common method in both the binomial and Poisson cases, we find very limited work, e.g. Hepworth (2013), in which it is used to construct confidence intervals for a negative binomial distribution. So, we calculated confidence intervasl using the mid-$p$ method, using the

bisection method to search for the lower points that satisfy

$$Pr(X > x|p) + \frac{1}{2}Pr(X = x|p) = \alpha \tag{2.32}$$

and to search for the upper points that satisfy

$$Pr(X < x|p) + \frac{1}{2}Pr(X = x|p) = \alpha. \tag{2.33}$$

The mid-$p$ method reduces the conservatism of the Casella and McCulloch method, but it no longer guarantees that the minimum coverage is at least as large as the nominal level $(1-\alpha)$. The top graph of Figures 2.11 and 2.12 plot the coverage of the 97.5% one-sided upper and lower interval, respectively, for the mid-$p$ method. The spikes in the plots are spread fairly regularly around the nominal level. The coverage of its upper tail still tends to be slightly conservative at large values of $p$, but to a much lesser extent than with the exact method. In contrast, the coverage of its lower tail tends to be below the nominal level for large values of $p$. However, it is one of the desirable methods for practical purposes as it has good coverage ( generally close to the nominal level) and good length performance.

Confidence intervals of the mid-$p$ method for the negative binomial mean $\mu$ are constructed by using the transformation from $p$ to $\mu$, that was discussed in Section 2.4.1.1. The top graphs in Figures 2.13 and 2.14 plot the coverage of the 97.5% one-sided upper and lower interval of the negative binomial mean $\mu$, respectively, for the mid-$p$ method. The coverage is more variable for small values of $\mu$ when it is also a little liberal. However, it does not seem to be excessively liberal, so the mid-$p$ is one of the desirable methods for practical purposes in the case of the negative binomial mean $\mu$.

#### 2.4.2.2 Jeffreys method

The Jeffreys interval is constructed from the Bayesian posterior distribution

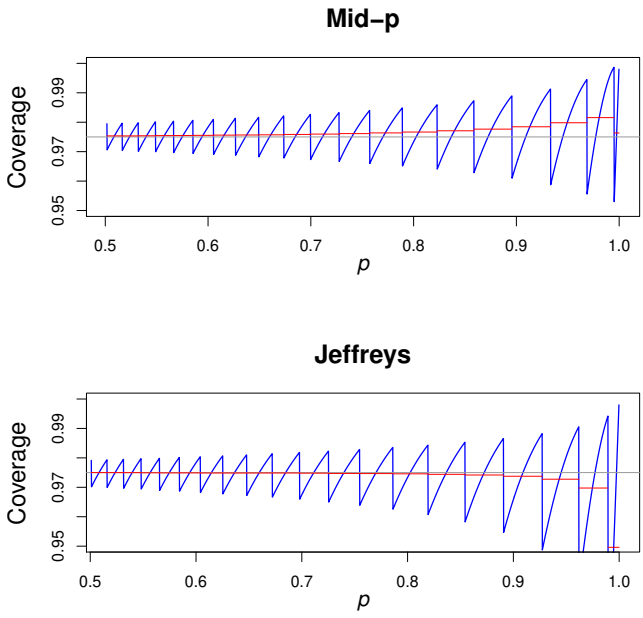


Figure 2.11: Coverage of upper one-sided 97.5% confidence intervals for the negative binomial proportion $p$ for the mid-$p$ and Jeffreys methods when $x$=20

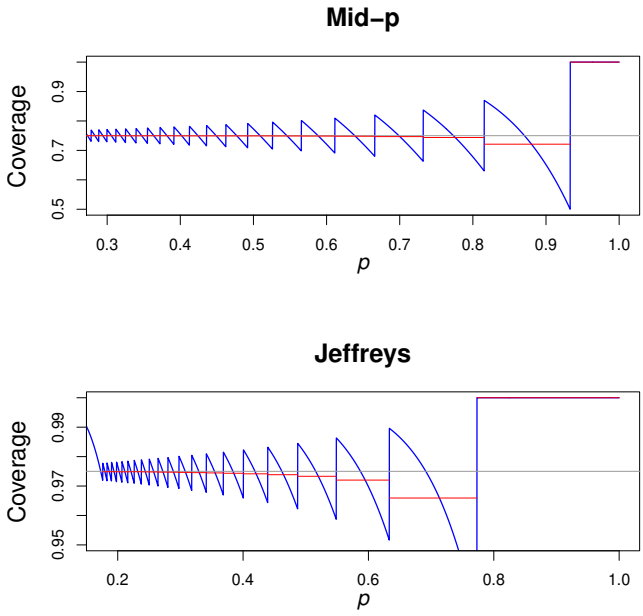


Figure 2.12: Coverage of lower one-sided 97.5% confidence intervals for the negative binomial proportion $p$ for the mid-$p$ and Jeffreys methods when $x$=20

with prior $p \propto p^{-1}(1-p)^{-1/2}$. The posterior distribution of $p$ is $p|x \sim$ Beta$(r, x+\frac{1}{2})$ (Brown *et al.*, 2003). Hence, the upper and lower limit of the Jeffreys interval for $p$ are respectively

$$l_x = Beta(\alpha, x + 1/2, r) \tag{2.34}$$

and

$$u_x = Beta(1-\alpha, x + 1/2, r). \tag{2.35}$$

The limit $l_x$ is the $\alpha$ quantile and $u_x$ is the $1-\alpha$ quantile of the beta distribution $Beta(r, x+\frac{1}{2})$. Jeffreys method gives intervals that are more liberal than the mid-$p$ method. This is clear in the lower graphs of Figures 2.11 and 2.12. In particular, Jeffreys upper-tail intervals have a low coverage for large values of $p$ while its coverage is close to the nominal level for small values of $p$. Jeffreys method is recommended as a good alternative method for practical use because it has better length properties (Cai, 2005). However, its coverage can be quite low for large values of $p$.

Jeffreys method retains these characteristics in its confidence intervals for the negative binomial mean $\mu$. This can be seen in the lower graphs of Figures 2.13 and 2.14, where coverages of the 97.5% one-sided upper and lower intervals are plotted against the negative binomial mean $\mu$. Jeffreys upper-tail intervals have a low coverage for small values of $\mu$ while its coverage is close to the nominal level for large values of $\mu$.

### 2.4.3 Asymptotic methods

#### 2.4.3.1 Wald method

The Wald method is the simplest and very commonly used method in the literature for constructing a confidence interval for the negative binomial proportion $p$. It is used most widely in practical statistical analysis and econometrics

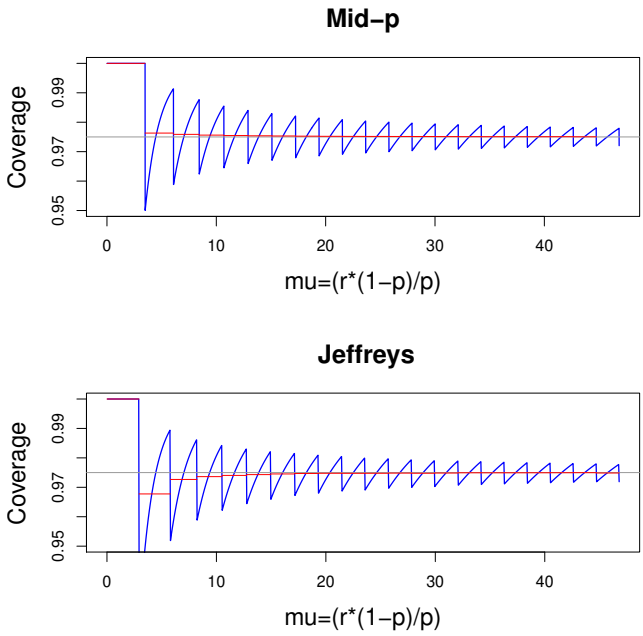


Figure 2.13: Coverage of upper one-sided 97.5% confidence intervals for the negative binomial mean $\mu$ for the mid-$p$ and Jeffreys methods when $x = 20$

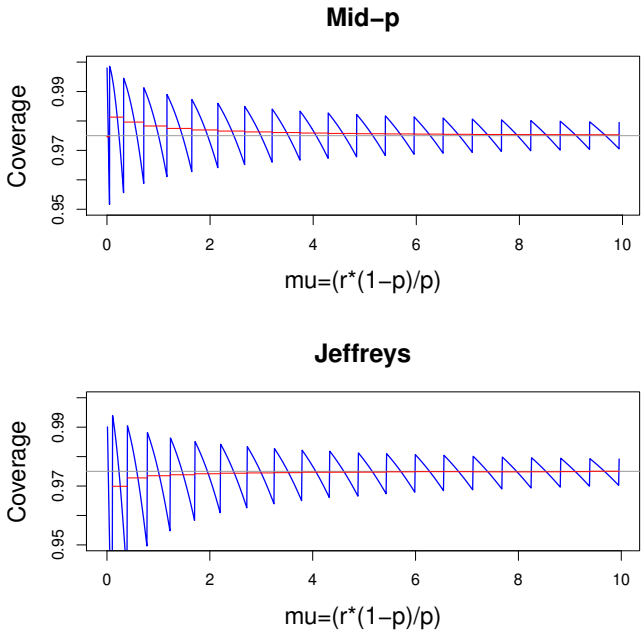


Figure 2.14: Coverage of lower one-sided 97.5% confidence intervals for the negative binomial mean $\mu$ for the mid-$p$ and Jeffreys methods when $x = 20$

research. Its confidence interval formula is obtained by using the standardization of the maximum likelihood estimator (MLE) of $p$, $\hat{p} = r/x + r$, where $x$ is the total number of failures before obtaining the predetermined number of successes $r$ (Lui, 2004). A 100 $(1-\alpha)$ % confidence interval for $p$ is given by $l_x, u_x$ where

$$l_x = \max[\hat{p} - z_{1-\alpha}\sqrt{\frac{\hat{p}^2(1-\hat{p})}{r}}, 0] \tag{2.36}$$

and

$$u_x = \min[\hat{p} + z_{1-\alpha}\sqrt{\frac{\hat{p}^2(1-\hat{p})}{r}}, 1]. \tag{2.37}$$

where $z_{1-\alpha}$ is the $(1-\alpha)100\%$ percentile of the standard normal distribution. The Wald method is commonly used because it gives intervals that are simple to compute. However its upper-tail coverage usually tends to be very conservative for large values of $p$ and close to the nominal level for small vales of $p$. Its lower-tail coverage tends to oscillate far below the nominal level for large values of $p$ and is close to the nominal level for small values of $p$. This can be seen clearly in the top graph of Figures 2.15 and 2.16, where the coverage of its 97.5% one-sided upper and lower intervals is plotted against the negative binomial proprtion $p$.

Unlike the mid-$p$ method and Jeffreys method, confidence intervals of the Wald method of the negative binomial mean $\mu$ are calculated directly by using its own formula. So, the coverage probability of its confidence interval for the upper tail is far below the nominal level for most values of $\mu$. This is shown in the top graph of Figure 2.17 where the coverage of its 97.5% one-sided upper interval is plotted against $\mu$.

#### 2.4.3.2 Score method

Another method that uses the normal approximation is the score method. Construction of the score interval is based on the score function $S(p)$, which

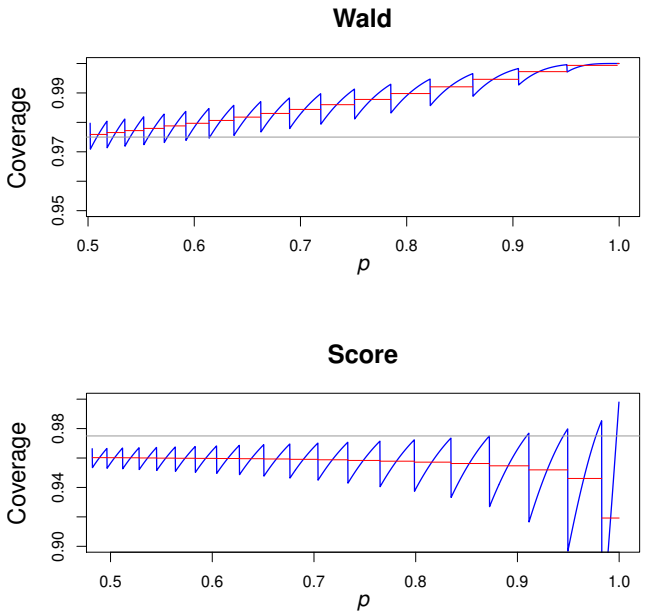


Figure 2.15: Coverage of upper one-sided 97.5% confidence intervals for the negative binomial proportion $p$ for the Wald and score methods when $x = 20$

is given by (Tian *et al.*, 2009) as

$$S(p) = \frac{r}{p} - \frac{x}{1-p}, \tag{2.38}$$

and on the Fisher information $I(p)$, which is given by

$$I(p) = \frac{r}{p^2} - \frac{x}{(1-p)^2}. \tag{2.39}$$

An approximate $(1-\alpha)100\%$ confidence interval for $p$ is $l_x, u_x$, where

$$l_x = max\left[\frac{(2(x+r)r - rZ_{1-\alpha}^2) - \sqrt{r^2Z_{1-\alpha}^4 - 4(x+r)r^2Z_{1-\alpha}^2 + 4(x+r)^2rZ_{1-\alpha}^2}}{2(x+r)^2}, 0\right] \tag{2.40}$$

and

$$u_x = min\left[\frac{(2(x+r)r - rZ_{1-\alpha}^2) + \sqrt{r^2Z_{1-\alpha}^4 - 4(x+r)r^2Z_{1-\alpha}^2 + 4(x+r)^2rZ_{1-\alpha}^2}}{2(x+r)^2}, 1\right]. \tag{2.41}$$

where $r + x$ is the total number of trials.

Compared to the Wald method, the score method has favourable coverage and length properties. The coverage of both 97.5% upper and lower tail inter-

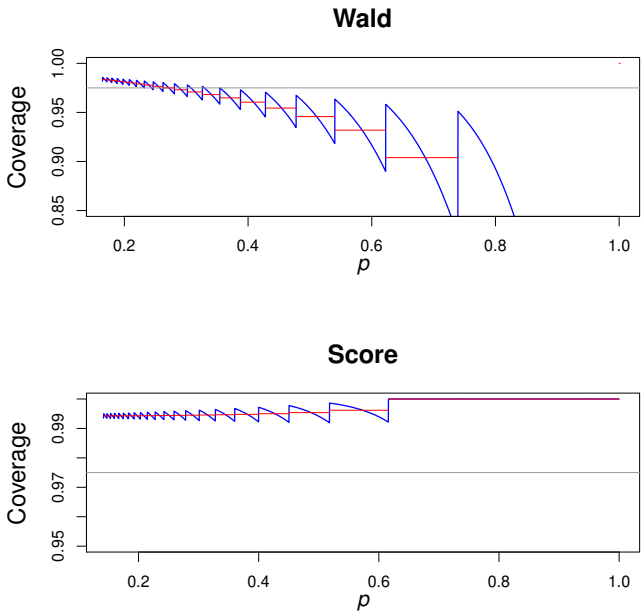


Figure 2.16: Coverage of lower one-sided 97.5% confidence intervals for the negative binomial proportion $p$ for the Wald and score methods when $x = 20$

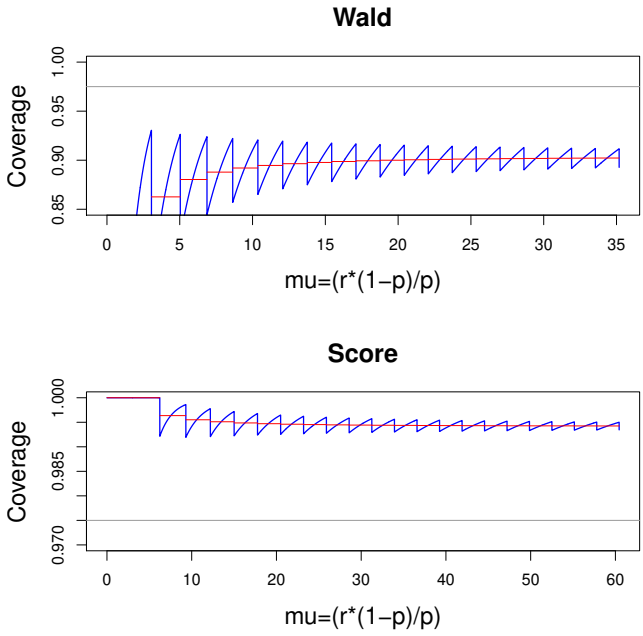


Figure 2.17: Coverage of upper one-sided 97.5% confidence intervals for the negative binomial mean $\mu$ for the Wald and score methods when $x = 20$

vals of the score method are plotted against $p$ in the lower graphs of Figures 2.13 and 2.14, respectively. Looking at these figures, it can be seen that the coverage of the upper-tail is generally somewhat below the nominal level while the lower-tail coverage tends to be very conservative for all values of $p$. That means the score method does not meet the definition of a confidence interval. Although the score method improves the coverage of its intervals relative to the coverage of Wald intervals, it suffers from giving intervals that are sometimes too long, with a coverage that tends to be very conservative for large values of $\mu$, as shown in the lower graph of Figure 2.17. These intervals are calculated also directly as it is done in the Wald method.

### 2.4.4 Desirable properties of confidence intervals

As noted in Sections 2.2.4 and 2.3.4, there are a number of desirable properties in methods of forming a confidence interval.

All the methods described earlier in this section give confidence intervals without any disjoint points and hence have Property 1. They also have Property 2, the monotonicity in $x$ property, and they also have Property 5, the nesting property.

### 2.4.5 Other methods

Using the negative binomial distribution in practical research is not as common as using the binomial or Poisson distributions. Perhaps for this reason the number of method of constructing a confidence interval for $p$ and $\mu$ is quite limited. Some methods have been proposed that aim to enhance the coverage or length of the methods described earlier. Beginning with methods that

aim to improve the performance of the Wald method, there are the Wald confidence interval with continuity correction (WCC) and the Wald confidence interval based on uniform minimum variance unbiased estimation (WUMVUE) (Young, 2014). The WCC method applies continuity correction to the original equations of the Wald method, equations (2.34) and (2.35). Whereas the WUMVUE is based on the uniform minimum variance unbiased estimation of $p$, which can be obtained by $\tilde{p} = \frac{r-1}{x-1}$, and the estimate of its variance, which can be calculated by $var(\hat{p}) = \frac{r(1-p)}{p^2}$. The WCC method tends to be very conservative due to the correction factor. The WUMVUE performs well for large values of $y = r + x$, but for small values of $y$ it tends to be well below the nominal level for $p$ near to 0 and very conservative for $p$ near to 1. There is also a likelihood ratio based confidence interval (LR), based on the likelihood ratio statistic $2[\log L(\hat{p}) - \log L(p)]$, which asymptotically follows a chi-square distribution with one degree of freedom. The LR method performs well for large values of $r + x$, but for small values its coverage is far below the nominal level (Tian *et al.*, 2009). Also, Tian *et al.* (2009) introduced a confidence interval that is based on a saddle-point approximation. It is designed to approximate the tail probability of the distribution. The method performs well, except that the oscillations of its coverage tend to be very large as $p$ gets close to 1 (Young, 2014).

## 2.5 Concluding Comments

In this chapter, some of the relevant research work has been reviewed on common methods of constructing a confidence interval for three discrete distributions: the binomial, Poisson and negative binomial distributions. The focus was on methods of forming two-sided, equal-tailed confidence intervals, for the binomial proportion $p$, Poisson mean $\lambda$ and negative binomial distribution's

proportion $p$ and mean $\mu$. All these common methods will be compared with a new method, which will be developed in the next chapters. Desirable properties in methods of forming confidence intervals have also been reviewed. In addition, some methods that give confidence interval with unequal-tails were also reviewed briefly.

# Chapter 3

# Binomial confidence interval methods

## 3.1 Introduction

Most confidence interval methods for a binomial proportion that are well-recommended in the literature do not meet the strict definition of a confidence interval. The same holds for confidence intervals for other discrete distributions that have attracted attention. Thus it seems that the strict definition of a confidence interval is not fit for purpose when the sample space is discrete. A definition of an interval estimate is required that people are willing to use. With an appropriate definition, there should be interval estimators that satisfy the properties of giving sensible intervals and giving intervals with an acceptably short average length. The definition must ideally yield an "optimal" method or, failing that, methods that have these properties. For a binomial proportion, we equate an"optimal" method to the method that gives intervals of minimum average width when the expected width of an interval is averaged over $p \in [0, 1]$.

The definition must place some restrictions on the coverage probabilities of intervals. If there is no restrictions at all on the coverage probability, the shortest interval could be taken as $(\hat{p} + 0.0)$, where $\hat{p}$ is a point estimate of $p$. These nonsensical intervals would have the minimum possible width but would have coverage probabilities of 0. Here we consider some alternative definitions of an interval estimate that place different restrictions on the coverage of intervals. So, in Section 3.2 we consider some alternative definitions of an interval estimate that place different restrictions on the coverage of intervals. In Section 3.3, we give a precise definition of a new type of interval: a locally correct confidence interval.

In Section 3.4, we present a novel interval estimator that yields locally correct confidence intervals. We refer to the new estimator as the optimal locally

correct (OLC) method. We examine whether intervals given by the OLC method seem sensible and examine whether the new estimator has properties that have been proposed in the literature as being desirable. Also, we prove that the OLC method yields intervals with a smaller average length than any other interval estimator that yields locally correct confidence (LCC ) intervals. In Section 3.5 we compare the OLC method with several methods that have been recommended for forming equal-tailed confidence intervals (as noted in the literature chapter). Concluding comments are given in Section 3.6.

## 3.2 New definitions of an interval estimate

### 3.2.1 Overall coverage restriction

We say that a method of forming interval estimates meets the *overall coverage restriction* if, for any value of $n$ and $\alpha$, it yields intervals where the average coverage over $p \in [0, 1]$ is at least $(1 - \alpha)$. We explored the optimal method under this definition (the method that gives the minimum average width while meeting the restriction imposed by definition), by examining the confidence intervals that would be formed for different values of $n$ and $x$. Common sense suggest that, for a given $n$, the endpoints of an interval should vary as x varies. However, we found that some different values of $x$ give the same endpoints, and this can happen both when $n$ is large and when it is small. For example in Figure 3.1, when $n = 10$ and $\alpha = 0.025$, the endpoints $U_1, U_2, U_3, U_4, U_5, U_6$ and $U_7$ all have almost the same value (0.7108389, 0.718388, 0.710839, 0.7108391, 0.7108394, 0.7108398 and 0.7108405) and the endpoints $U_8$, $U_9$ and $U_{10}$ have almost the same value (0.9993905, 0.9993969 and 0.9993999). These endpoints were calculated by optimization package Rsolnp in R (Ghalanos and Theussl, 2015). This seems very unsatisfactory as, intuitively, different $x$ values should

give different endpoints. Hence, this definition of an interval estimate seems inadequate and is not considered again.

### 3.2.2 Restricting average coverage in fixed intervals

Under a second definition, the range of $p$ is divided into equal subintervals, and the requirements are imposed that the average coverage over each subinterval must be greater than or equal to $(1-\alpha)$. Two variants of this definition were considered.

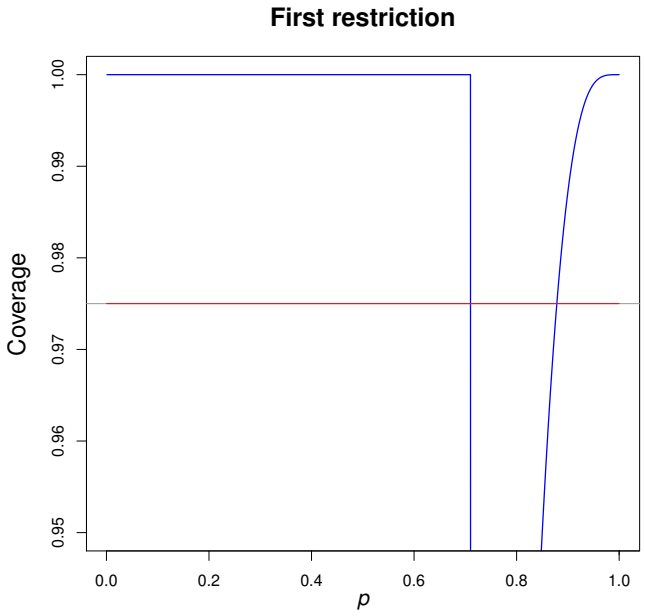


Figure 3.1: Endpoints and coverage of upper one-sided 97.5% confidence intervals for $n$ =10 and $\alpha$= 0.025 when the overall average coverage must be at least $1-\alpha$. The red horizontal line show the overall avaerage coverage over $p \in [0,1]$

***(a) No restriction on the number of spikes in each subinterval.*** The endpoints are calculated by a search method over each subinterval. We again find that, with the optimal method (shortest average width), different values of $x$ sometimes give the same endpoints. Also, when the number of

subintervals is large relative to $n$, the average coverage suffers from almost as much conservatism as with the Clopper Pearson method. This can be seen in Figure 3.2, where $n$=10, the number of subintervals=50 and $\alpha$=0.025. In the definition's favour the coverage is mostly above the nominal level, but this does not adequately compensate for the other disadvantages.

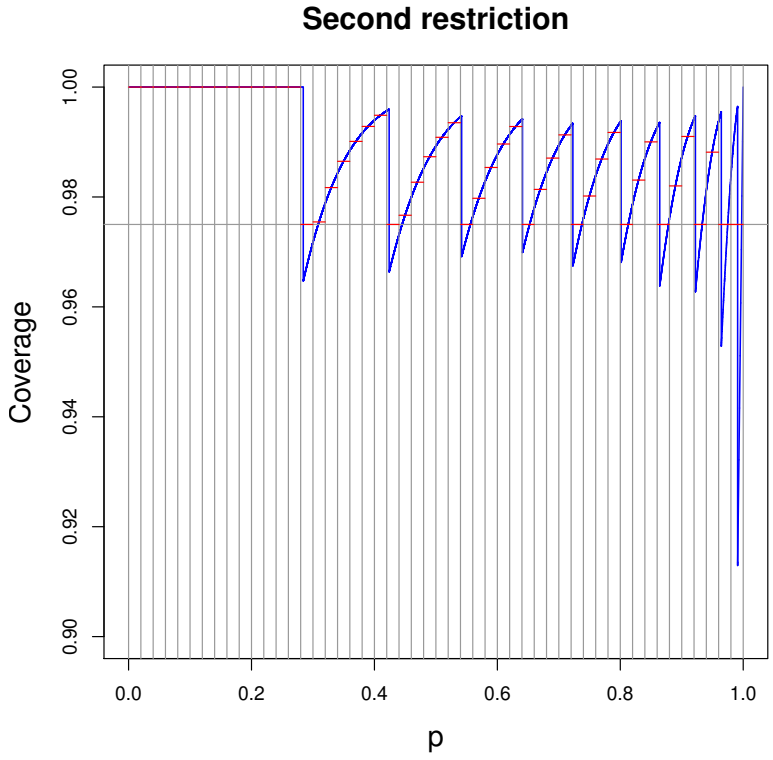


Figure 3.2: Endpoints and coverage of upper one-sided 97.5% confidence intervals for n=10 and subintervals=50 when the average coverage in each subinterval must be no less than $(1-\alpha)$. Short horizontal (red) lines show the average coverage in each subinterval

***(b) Adding the constraint that there is at most one spike in any subinterval.***

Under this restriction the number of subintervals must exceed $n-1$, there must be at most one spike in any subinterval, and the average coverage over each subinterval must be greater than or equal to $(1-\alpha)$. The endpoints

under this restriction are exactly the same as the endpoints of the subintervals themselves. This definition comes with problems, regardless of the number of subintervals that are used. When the number of intervals is small (equal to $n$ or only a little larger than $n$), there is a contradiction between obtaining one endpoint per subinterval and having an average coverage that is greater than or equal to $(1-\alpha)$ over each subinterval. This can be seen in Figure 3.3, for $n = 20$, subintervals=20 and $\alpha = 0.025$, where the positioning of spikes is too restrictive, and the average coverage does not meet the definition. Also, when the number of subintervals is very large, the intervals become as conservative as the Clopper-Pearson method. Hence, the definitions based on average coverage restrictions over fixed intervals fail to fill our needs and will not be considered further.

## 3.3 Locally correct confidence intervals

The results of the previous section indicate that a different restriction on coverage is needed. We first consider upper-tail (upper one-sided) intervals. Let $X$ denote a binomial variate based on $n$ trials with success probability $p$ and suppose an interval estimator gives $(0, u_x)$ as its upper-tail estimate for $p$ when $x$ is the observed value of $X$. For a sensible estimator,

$$0 < u_0 < u_1 < \ldots < u_n \leq 1, \tag{3.1}$$

and we assume that equation (3.1) holds. The coverage probability of the interval estimator depends on the value of $p$ and is the probability that the random interval $(0, u_x)$ contains $p$.

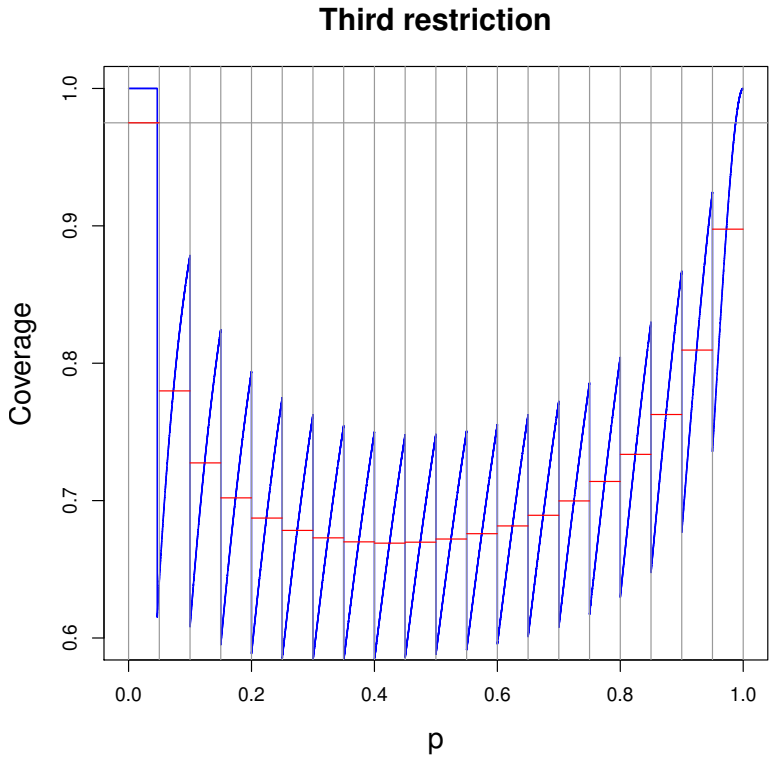


Figure 3.3: Coverage of upper one-sided 97.5% confidence intervals for 20 subintervals when $n = 20$ and the average coverage in each subinterval must be no less than $(1 - \alpha)$

We denote this coverage by $C_u(p)$. When $u_{i-1} < p \leq u_i$,

$$C_u(p) = Pr(X \geq i|p) = \sum_{x=i}^{n} \binom{n}{x} p^x (1-p)^{n-x}. \tag{3.2}$$

For $x = 0, 1, \ldots, n$, the difference between $C_u(u_x)$ and $C_u(u_x + \delta)$ does not tend to 0 as $\delta \to 0$, but equals $\binom{n}{x} u_x^x (1-u_x)^{n-x}$. This is the reason that the coverages in all previous figures have a saw-tooth appearance (spikes). The points of the spikes occur where $p$ equals $u_0, \ldots, u_n$ and the coverage drops by $\binom{n}{x} u_x^x (1-u_x)^{n-x}$ at $p = u_x$ (Garthwaite *et al.*, 2019). The question, is how should $C_u(p)$ vary with $p$ for a good interval estimator? Looking at Figure 1.1, if the nominal confidence level is $(1-\alpha)$, then $C_u(p)$ should exceed $(1-\alpha)$ when $p$ is just before a spike, as $C_u(p)$ follows a cycle with its largest values just before spikes. If the estimator is not to be very conservative, then $C_u(p)$ should be less than $(1-\alpha)$ when $p$ is just after a spike, as then $C_u(p)$ is at the lowest part of its cycle. However, the extent to which the estimator is liberal should be restricted. The restriction which we propose is that the *average coverage within each cycle* - the interval between two spikes - should exceed or equal the nominal confidence level $(1-\alpha)$. The intervals are quite small. Hence, while the coverage need not equal or exceed the nominal level at individual values of $p$, it must do so on average over quite narrow ranges of $p$. We say that an interval estimator that meets this requirement gives *locally correct confidence (LCC) intervals.*

*Definition 1.* For the upper-tail intervals, suppose an interval estimator gives $(0, u_x)$ as its upper-tail interval for $p$ when $X = x$, that $u_0, \ldots, u_n$ satisfy (3.1) and that $u_n$=1. If, for $i = 1, \ldots, n$

$$\frac{1}{u_i - u_{i-1}} \int_{p=u_{i-1}}^{u_i} C_u(p)dp \geq (1-\alpha) \tag{3.3}$$

then the interval estimator gives *upper-tail LCC intervals* with confidence level $(1-\alpha)$. We assume that $u_n$=1; otherwise the average coverage over the interval

$(u_n, 1)$ would be 0, which is not consistent with the required coverage in other intervals.

*Definition 2.* For the lower-tail intervals, suppose an interval estimator gives $(l_x, 1)$ as its lower-tail interval for $p$ when $X = x$, and that $0 = l_0 < l_1 < \ldots < l_n < 1$. Define the coverage probability, $C_l(p)$, by

$$C_l(p) = Pr(X \leq i|p) = \sum_{x=0}^{i} \binom{n}{x} p^x (1-p)^{n-x}. \tag{3.4}$$

for $l_i < p \leq l_{i+1}$. If, for $i = 1, \ldots, n-1$

$$\frac{1}{l_{i+1} - l_i} \int_{p=l_i}^{l_{i+1}} C_l(p) dp \geq (1 - \alpha) \tag{3.5}$$

then the interval estimator gives *lower-tail LCC intervals* with confidence level $(1 - \alpha)$.

Regarding two-sided equal-tail LCC interval estimators, we define it in terms of one-sided LCC intervals as follows.

*Definition 3.* Suppose that, for $x = 0, \ldots, n$, an interval estimator gives $(l_x, u_x)$ as its two-sided equal-tail intervals for $p$ when $X = x$. Then it gives equal-tail LCC intervals with confidence level $(1 - 2\alpha)$ if and only if, for $x = 0, \ldots, n$, the intervals $(l_x, 1)$, and $(0, u_x)$ are sets of one-sided lower-tail and upper-tail LCC intervals, respectively, each with confidence level $(1 - \alpha)$.

An interval estimator that gives equal-tail LCC intervals will be referred to as a *LCC interval estimator.*

## 3.4 A new interval estimator

We propose an interval estimator that uses a straightforward iterative algorithm to obtain one-sided interval estimates. By its construction, the algorithm clearly gives LCC intervals. We first give the algorithm and then give results about the average width of its intervals, which show that the method is the optimal locally correct method provided $\alpha$ is less than 0.27.

The algorithm is a sequential procedure and we give the steps for attaining one-sided upper tail intervals. First we determine $u_n$, then $u_{n-1}$ is determined given the value of $u_n$. Each $u_i$ is then determined given the value of $u_{i+1}$. This sequentionl procedure is continued until it determines $u_0$. Specifically, the steps of the algorithm are as follows.

1. Set $u_n = 1$ and put $i = n - 1$.

2. Given $u_{i+1}$, use the bisection method to search for the value $u_i$ that makes the average coverage over the interval $(u_i, u_{i+1})$ equal to $1 - \alpha$:

   (a) Put $a = 0$ and $b = u_{i+1}$.

   (b) Take $u_i^* = (a + b)/2$.

   (c) Evaluate the integral over the interval $(u_i^*, u_{i+1})$ by using incomplete beta functions to calculate the average coverage, say $m$.

3. If $m - (1 - \alpha) < 0$ put $a = u_i^*$ and go to Step 2,

   else if $m - (1 - \alpha) > \varepsilon$ , for a very small value of $\varepsilon$, put $b = u_i^*$ and go to Step 2.

   Else take $u_i = u_i^*$.

4. Repeat steps 2 and 3 for $i = n-2, n-3, \ldots, 1, 0$ to obtain $u_{n-2}, u_{n-3}, \ldots, u_1, u_0$.

From its construction, the method determines the endpoints of subintervals that have an average coverage of $1-\alpha$. A hypothetical problem that could arise is the following. After $u_i$ has been attained, it could be the case that $Pr(x \geq i-1|p = u_i)$ is less than $1-\alpha$, even though $(u_{i+1}-u_i)^{-1} \int_{u_i}^{u_{i+1}} Pr(x \geq i|p)dp = 1 - \alpha$. Then, given $u_i$, there is no $u_{i-1}$ (with $u_{i-1} \leq u_i$ ) that meets the requirement that $(u_i - u_{i-1})^{-1} \int_{u_{i-1}}^{u_i} Pr(x \geq i-1|p)dp = 1 - \alpha$. Hence the following proposition is needed to underpin the algorithm.

**Proposition 3.1**

Suppose $1 \leq n \leq 200$ and $\alpha \in \{0.001(0.001)0.27\}$. Suppose also that

$$\frac{1}{u_i - u_{i-1}} \int_{p=u_{i-1}}^{u_i} \sum_{x=i}^{n} \binom{n}{x} p^x (1-p)^{n-x} dp = 1 - \alpha \tag{3.6}$$

and $u_i > u_{i-1}$ for $i = j+1, j+2, \ldots, n; j = 0, \ldots, n-1$. Then there is a unique $u_{j-1}$ such that $u_j > u_{j-1} > 0$ and equation (3.6) holds when $i = j$.

We have been unable to prove the result in the proposition for a general $n$ and $\alpha$ as it is hard to show that $\sum_{x=i}^{n} \binom{n}{x} u_i^x (1-u_i)^{n-x} > (1-\alpha)$. But, repetitive straightforward computation showed that the result in the proposition holds if $n$ is a positive interger less than 200 and $\alpha$ equals one of the numbers $0.001, 0.002, \ldots, 0.27$. This covers the value of $n$ and $\alpha$ of practical interest. So, throughtout this chapter it is assumed that $1 \leq n \leq 200$ and $\alpha \in 0.001(0.001)0.27$.

From Proposition 3.1, for an upper-tail interval, there is always a unique $u_{i-1}$ for which equation (3.6) holds. The new interval estimator sets $(0, u_i)$ as the upper-tail interval when $X = i$ and we have that $0 < u_0 < \ldots < u_n = 1$. Then, under definition 1 the new estimator gives upper-tail LCC intervals.

Similar steps are followed to form lower-tail intervals. It begins by putting $l_0 = 0$ and then $l_1, \ldots, l_n$ are determined sequentially. Each $l_{i+1}$ is determined given the value of $l_i$. Given $l_i$, the value $l_{i+1}$, for $i = 0, \ldots, n-1$, is found that satisfies

$$\frac{1}{l_{i+1} - l_i} \int_{p=l_i}^{l_{i+1}} \sum_{x=0}^{i} \binom{n}{x} p^x (1-p)^{n-x} dp = 1 - \alpha \tag{3.7}$$

Under Definition 2 , the interval estimator gives lower-tail LCC interval, as $(l_i, 1)$ is the $1-\alpha$ lower-tail interval when $X = i$.

Two-sided intervals are obtained by combining the endpoints of one-sided upper and lower- tail intervals. Thus, when $x = i$ the new interval estimator gives $(l_i, u_i)$ as the two-sided equal-tail interval for a confidence interval of

$1 - 2\alpha$. From its construction, the new estimator is an LCC interval estimator. Different examples for the coverage of the new estimator are shown in Figure 3.4, where the coverage is plotted against $p$ for upper-tail intervals with nominal confidence levels of 97.5% and 99.5% for sample sizes 8, 20 and 50. It is clear that the coverages are evenly spread around the nominal level, so the new method gives sensible interval estimates. With the new method, the average coverage between every two consecutive spikes always equals the nominal level $(1 - \alpha)$.

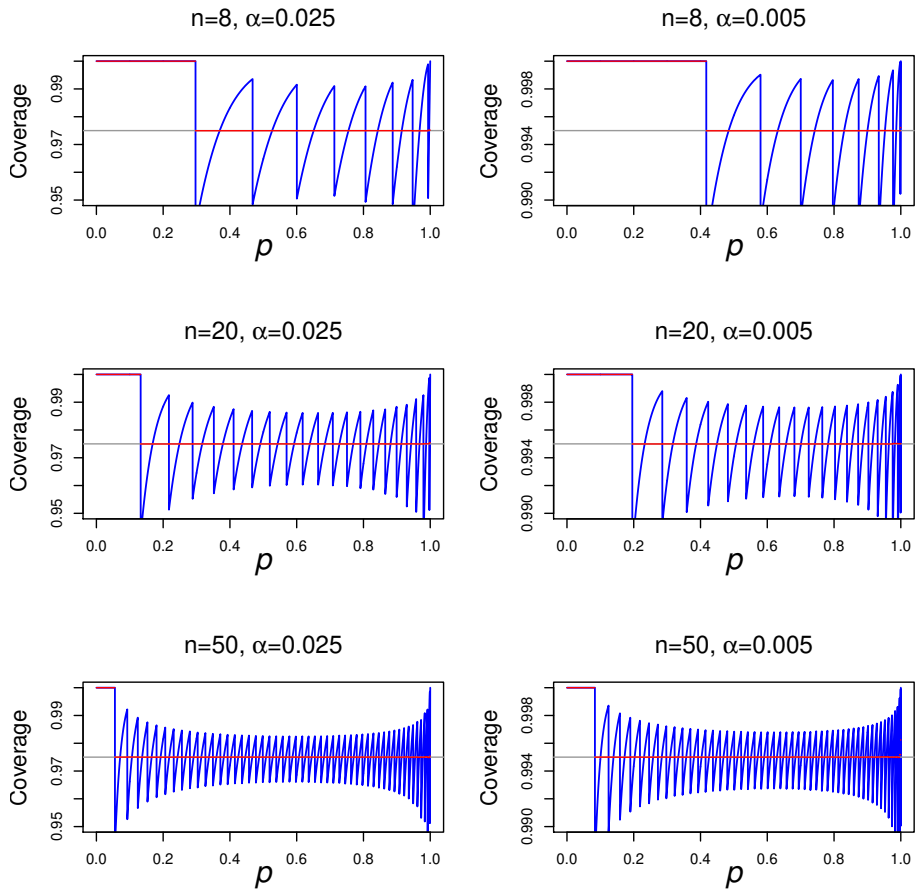


Figure 3.4: Coverage of upper-tail LCC intervals given by the new estimator for samples sizes of 8, 20 and 50, and nominal confidence levels of 97.5% and 99.5%

Moving to the most important feature of an interval estimator (after coverage), we next consider the length of intervals. If an estimator gives $(l_x, 1)$ as

its lower interval when $X = x$, then the expected length of its lower interval, $L_l(p)$, is

$$L_l(p) = \sum_{x=0}^{n}(1 - l_x)\binom{n}{x}p^x(1-p)^{n-x}. \tag{3.8}$$

Its average expected length (AEL), which is obtained by averaging over $p$, is

$$E[L_l(p)] = \int_{p=0}^{1} L_l(p)dp. \tag{3.9}$$

If an estimator gives $(0, u_x)$ as its upper interval when $X = x$, then the expected length of its upper interval, $L_u(p)$, is

$$L_u(p) = \sum_{x=0}^{n}(u_x - 0)\binom{n}{x}p^x(1-p)^{n-x}. \tag{3.10}$$

Its average expected length (AEL), (averaging over $p$), is

$$E[L_u(p)] = \int_{p=0}^{1} L_u(p)dp. \tag{3.11}$$

For two-tail intervals, the expected length of its interval, $L_n(p)$, is given by

$$L_n(p) = \sum_{x=0}^{n}(u_x - l_x)\binom{n}{x}p^x(1-p)^{n-x}. \tag{3.12}$$

and its average expected length (AEL), obtained by averaging $L_n(p)$ over $p$, is given by

$$E[L_n(p)] = \int_{p=0}^{1} L_n(p)dp. \tag{3.13}$$

We will refer to our new method of forming interval estimates as the *optimal locally correct (OLC)* method because it has the optimality property given in Proposition 3.2.

**Proposition 3.2**

Suppose $1 \leq n \leq 200$ and $\alpha \in \{0.001(0.001)0.27\}$. Then

(a) For one-tail intervals with confidence level $(1 - \alpha)$, the OLC method has the smallest AEL of any estimator that gives one-tail LCC intervals.

(b) For an equal-tail interval with confidence level $(1 - 2\alpha)$, the OLC method

has the smallest AEL of any estimator that gives equal-tail LCC intervals.

The range of $\alpha$, in Proposition 3.2, is necessarily restricted as the OLC method does not give the smallest AEL if $\alpha$ is greater than 0.27.

The proof of the proposition 3.2 involves (i) deriving the conditions in equations 3.14 and 3.15 and (ii) checking numerically that the conditions hold for the values of $n$ and $\alpha$ given in the proposition.

Let $(0, u_x^*)$ be the confidence interval given by our algorithm when $X = x$. We have that $u_0^* < \ldots < u_{x-1}^* < u_x^* < \ldots < u_n^*$, so $p$ is contained in the confidence interval if $X = x$. Let $p_x^* = (u_x^* + u_{x-1}^*)/2$ and put $\eta_x = u_x^* - p_x^*$. Also, let $p_{x-1}^{\#}$ be the value of $p$ that satisfies $g_n(x-1, p_{x-1}^{\#}) = 1 - \alpha$, where $g_n(x, p) = Pr(X \geq x|p)$. Let $\xi_i = \min(u_{x+1}^*,\ u_{x-1}^* + u_x^* - p_{x-1}^{\#})$. If $A_n(x, p_x^*, \eta_i)$ denotes the average coverage over the interval $(p_x^* - \eta_x,\ p_x^* + \eta_x)$, then our algorithm gives the shortest interval if

$$A_n(x, p_x^*, \eta_x) > A_n(x, p_x^*, \xi_x) \tag{3.14}$$

and

$$A_n'(x, p_x^*, \eta_x) < 0, \tag{3.15}$$

where $A_n'(x, p_x^*, \eta_i)$ is the differential of $A_n(x, p_x^*, \eta_x)$ with respect to $\eta_x$. For a proof of these two conditions, see Theorem 1 in Appendix A.

We conducted a computational study to examine when these conditions are satisfied. We found that they are satisfied for any sample size $n$ if $\alpha$ is less than 0.27. Also, for small values of $n$ they are satisfied for any value of $\alpha \leq 0.5$. In practice, for confidence intervals, the most commonly used values of $\alpha$ for one-sided interval are 0.005, 0.025, 0.05 and 0.1. The value $\alpha = 0.25$ is also important as it determines the upper quartile. The value $\alpha = 0.5$ is also important, of course, as it gives the median, but methods of forming confidence interval estimates will almost never set $\alpha$ greater than 0.27.

Thus, in practice the OLC method should be the preferred method of forming LCC intervals, provided it has standard properties that a well-behaved interval estimator should have.

The following properties were mentioned in Chapter 2 as desirable qualities.

**Property 1**. *Interval valued.* A confidence region should be an interval and not a collection of disjoint intervals.

The remaining properties assume that the confidence region is a two-tail interval with confidence level $1-2\alpha$, denote this interval as $(l(x,n,\alpha), u(x,n,\alpha))$ .

**Property 2**. *Equivariance.* As the binomial distribution in invariant under the transformation $X \to n-x; p \to 1-p$, confidence intervals should be invariant under this transformation. That is, if $x$ generates the confidence interval $[L(x,n,\alpha), U(x,n,\alpha)]$, then $n-x$ should yield the confidence interval $[1-u(n-x,n,\alpha), 1-l(n-x,n,\alpha)]$ for $x=0,\ldots,n$.

**Property 3**. *Monotonicity in $x$.* For fixed $n$ and $\alpha$, the endpoints should be increasing in $x$. This requires $l(x+1,n,\alpha) > l(x,n,\alpha)$ and $u(x+1,n,\alpha) > u(x,n,\alpha)$. For example, when $n=10$ and $x=5$, both the upper and lower endpoint should be greater than their corresponding values when $n=10$ and $x=4$.

**Property 4**. *Monotonicity in $n$.* For fixed $x$ and $\alpha$, the lower endpoint should be non-increasing in $n$ and the upper endpoint should be decreasing in $n$. This requires $l(x,n+1,\alpha) \leq l(x,n,\alpha)$ and $u(x,n+1,\alpha) < u(x,n,\alpha)$.

**Property 5**. *Nesting.* If two confidence intervals have different confidence levels then, for any given $n$ and $x$, the interval for the higher confidence level should contain the interval for the lower confidence level. Suppose we have two confidence levels $1-\alpha_1$ and $1-\alpha_2$ with $\alpha_1 < \alpha_2$, then this requires $(l(x,n,\alpha_2), u(x,n,\alpha_2)) \in (l(x,n,\alpha_1), u(x,n,\alpha_1))$. For this to occur for all confidence levels, as the level increases the lower limit for each $x$ must be non-

increasing and the upper limit must be non-decreasing.

Suppose $1 \leq n \leq 200$ and $\alpha \in \{0.001(0.001)0.27\}$, so that Proposition 3.1 applies. Then our OLC method will give confidence intervals without any disjoint points and hence it has Property 1. It also has Property 2, as it uses equivariance procedures to construct the lower and upper intervals. Also, interval endpoints increase as $x$ increases so it achieves the montonicity in $x$ property in Property 3. Repetitive computation has shown Properties 4 and 5, monotonicity in $n$ and nesting, are achieved when $1 \leq n \leq 200$ and $\alpha$, $\alpha_1$ and $\alpha_2$ are in $\{0.001(0.001)0.27\}$. Hence it seems clear that our OLC method meets the previous requirments for being a resonable interval estimator.

## 3.5 Comparison with other methods

In this section, we compare the OLC method performance with the following six methods of forming interval estimates, which have been described in Chapter 2: Clopper-Pearson, mid-$p$, Jeffreys, Wald, Wilson and Agresti-Coull methods. We compare them in terms of their coverage probability and expected length for sample sizes $n = 8, 20$ and 50 and for nominal confidence levels $95\%, 97.5\%$ and $99.5\%$. These sample sizes and nominal levels give a fair representation of the behaviour of intervals for other sample sizes and nominal levels.

### 3.5.1 Coverage probability

We will restrict attention to upper-tail intervals and two-tail intervals. Mentioned in Section 3, $C_u(p)$ is the coverage of an upper-tail interval estimator and is the probability that the random interval $(0, u_x)$ contains $p$. To calculate

the coverage probabilities of all methods, we define the quantity $T_u$ as

$$T_u = \frac{1}{1-u_0}\int_{p=u_o}^{1} C_u(p)\,dp, \tag{3.16}$$

and refer to it as the *truncated average coverage*, where $u_0$ is the endpoint of $x = 0$. Because coverage equals 1 for values of $p$ in the range $(0, u_0)$, we exclude the interval $(0, u_0)$ when calculating the average coverage and calculate $T_u$, as including that interval would distort the average. Instead, we give the values of both $T_u$ and $u_0$, which is more informative. Other average coverages for the values of $p > u_0$ may be calculated from $T_u$ and $u_0$. For example, for one-tail interval, the average coverage over the full range $(0, 1)$ equals$\{(1-u_0)T_u+u_0\}$ and for two-tail intervals, the average coverage over the range $(0, 1)$ equals $2\{(1-u_0)T_u + u_0\} - 1$.

In Figure 3.5, Figure 3.6 and Figure 3.7, the coverage of 95%, 97.5% and 99.5% upper tail intervals of all the methods are plotted against $p$ for sample sizes 8, 20 and 50, respectively. It is clear that the coverage of the first interval of $u_0$ equals 1, which will be excluded in forming averages. For all values of $p$ and each combination of $n$ and $\alpha$, the OLC, Clopper-Pearson and mid-$p$ methods are giving LCC intervals, i.e. the average coverage between consecutive spikes is at least (1-$\alpha$). In contrast, the Wilson, Wald, Agresti-Coull and Jeffreys methods do not give LCC intervals as the average coverage between consecutive spikes is sometimes below (1-$\alpha$). In particular, for the Wald method the average coverage is far below the nominal level for almost 70% of the spikes. To make comparison of our OLC method and other methods clearer, values of both $T_u$ and $u_0$ are given in Table 3.1. They are given for $\alpha = 0.05, 0.025$ and 0.005 and $n = 8, 20$ and 50. It is clear from Table 3.1 that the OLC method has good results. Its truncated average coverage equals the nominal confidence level of $1 - \alpha$ for all cases of $n$ and $\alpha$. In contrast, the Clopper-Pearson method suffers from conservatism, the mid-$p$ method

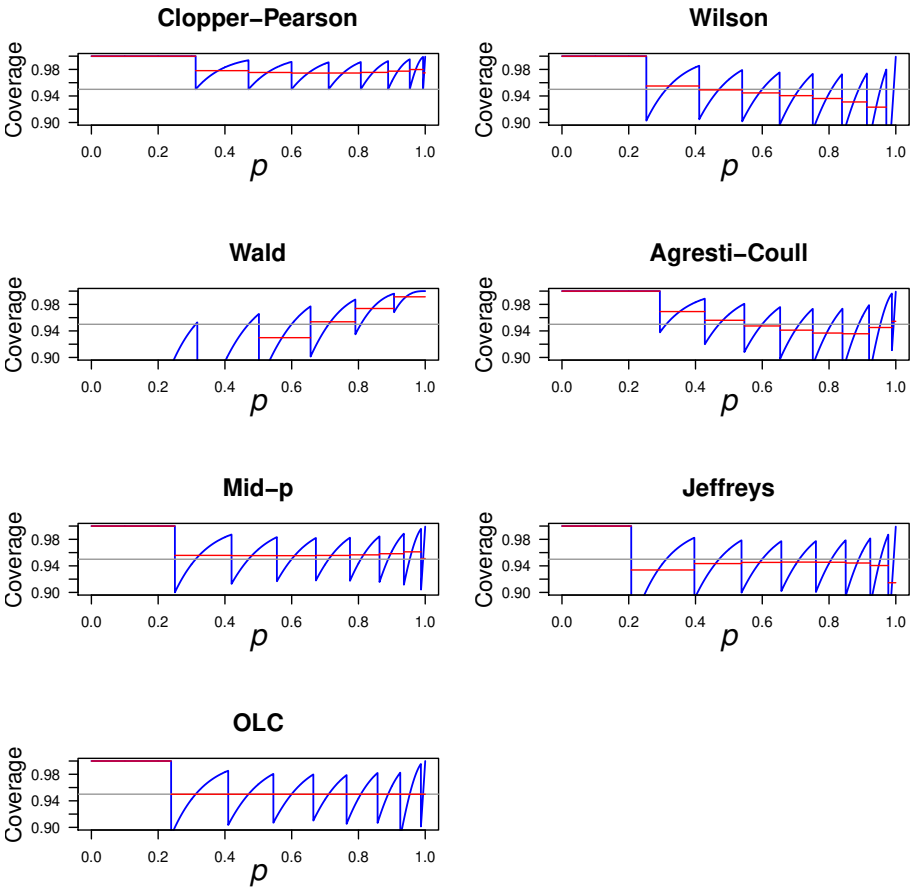


Figure 3.5: Coverage of upper-tail 95% interval estimate for the Clopper-Pearson, Wilson, Wald, Agresti-Coull, mid-$p$, Jeffreys and OLC methods, plotted against $p$ for $n = 8$. The horizontal red lines are the average coverage between consecutive spikes.

suffers a little from conservatism, and the Agresti-Coull method is generally conservative. The Wilson, Wald and Jeffreys methods are consistently liberal.

Turning to the values of $u_0$, a small value is typically desirable, so that the range over which the coverage equals 1 is small. On this basis, the Wald method always does well as its $u_0$ always equals 0. However, the coverage of the Wald method is too liberal for it to be the preferred method of forming confidence intervals. Based on the values of $u_0$, the OLC method is a little better than mid-$p$ and much better than Clopper-Pearson, Agresti-Coull and Wilson, but it is a little poorer than Jeffreys.

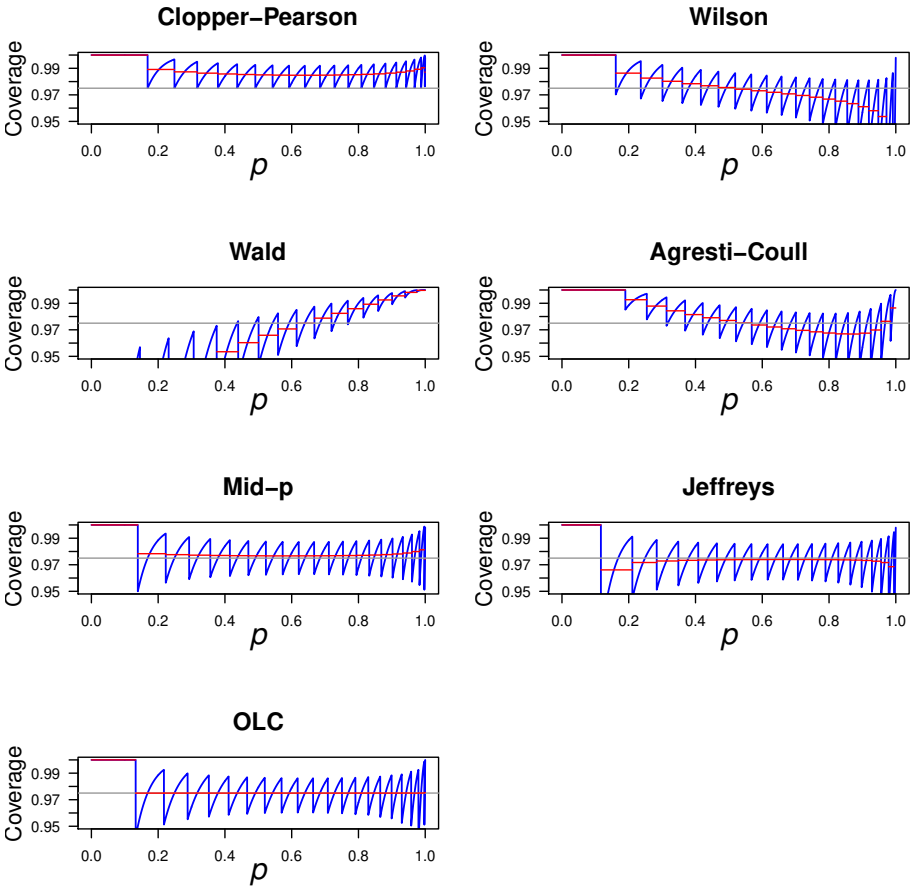


Figure 3.6: Coverage of upper-tail 97.5% interval estimate for the Clopper-Pearson, Wilson, Wald, Agresti-Coull, mid-$p$, Jeffreys and OLC methods, plotted against $p$ for $n = 20$.

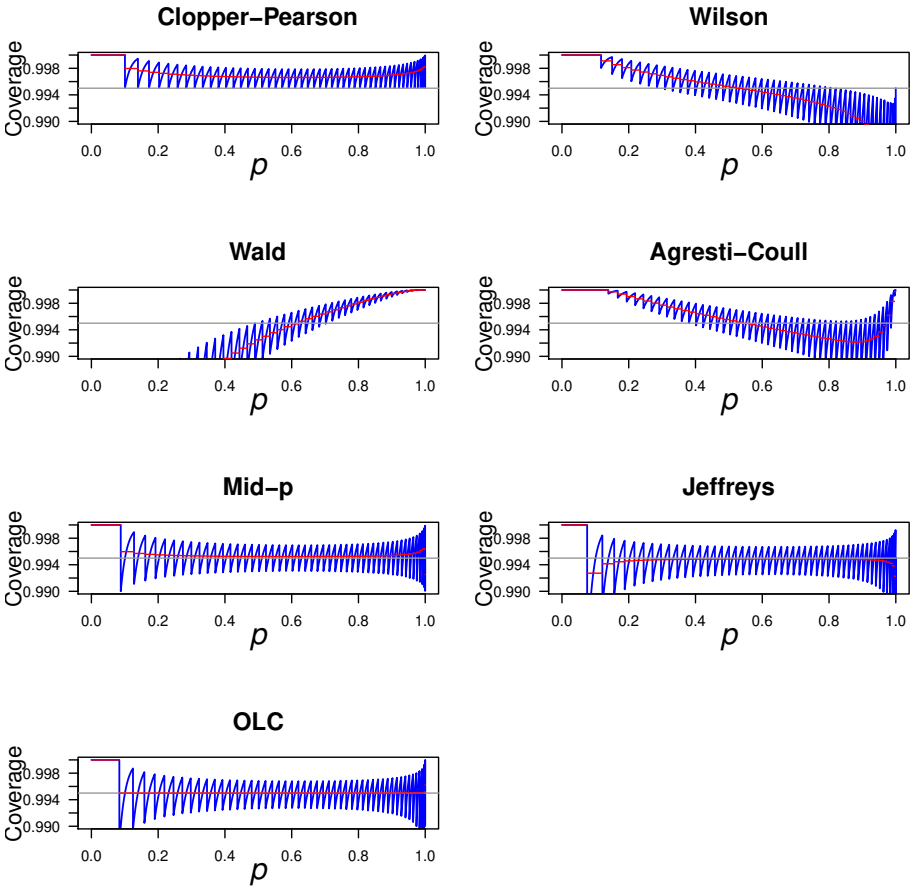


Figure 3.7: Coverage of upper-tail 99.5% interval estimate for the Clopper-Pearson, Wilson, Wald, Agresti-Coull, mid-$p$, Jeffreys and OLC methods, plotted against $p$ for $n = 50$.

As mentioned, in the literature review in Chapter 2, the mid-$p$ method does not meet the definition of a confidence interval, but it does meet our new definition of an LCC interval for $n < 200$ and $\alpha < 0.1$. Although the range of $\alpha$ for mid-$p$ method ($\alpha < 0.1$) is more limited compared to the OLC method, it does include all the confidence levels that are commonly of interest in practice. The result is giving in the following proposition

**Proposition 3.3**

For $1 \leq n \leq 200$ and $\alpha \in \{0.001(0.001)0.1\}$, the mid-$p$ method gives LCC intervals.

The results of this proposition have been verified by direct computation.

In Figure 3.5, the coverage of the 95% upper-tail interval for the mid-$p$ method for $n = 8$ is plotted against $p$ in the third graph of the left-hand side. The spikes in the plots are spaced fairly regularly and the actual coverage always crosses the nominal coverage level between consecutive spikes. In fact, a lot of researchers, e.g. Agresti and Gottard (2007), recommend mid-$p$ as an excellent method. Moreover, the mid-$p$ method gives one-tailed confidence intervals whose coverage is optimally close to the nominal level for any value of $p$. This property does not seem to have been noted before, so we specify it formally in Proposition 3.4, whose proof is given in Appendix B. This property is only given for upper-tail intervals but an equivalent result holds for lower-tail intervals.

**Proposition 3.4**

Consider the class of methods of forming upper-tail confidence intervals that (i) do not involve randomisation (i.e. confidence interval are determined by $x$ and $n$, and do not involve the value of a further hypothetical random variable), and (ii) satisfy $0 \leq u_0 \leq \ldots \leq u_n = 1$. Then the absolute error in the coverage probability, $|C_u(p) - (1 - \alpha)|$, is as small or smaller for the mid-$p$ method as

for any method in the class, for any value of $p$.

The class of methods in Proposition 3.4 includes all sensible methods of forming equal-tail confidence intervals that were discussed in Chapter 2. The property is quite strong because it relates to every value of $p$, and hence gives other properties. In comparing methods of forming confidence interval, it is common to examine the average absolute error in coverage or the root mean-square error in coverage, where averaging is over $p \sim U(0,1)$. Under either of these measures, Proposition 3.4 implies that the mid-$p$ method would be the optimal method of forming one-tail confidence intervals.

Table 3.1: Average coverage (Av.Cov) of upper-tail $1-\alpha$ intervals and smallest upper limit ($u_0$) of seven methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and sample sizes ($n$) of 8, 20 and 50.

| $\alpha$ | $n$ | statistic | Clopper Pearson | Mid-$p$ | Agresti Coull | Wilson | Wald | Jeff. | OLC |
|---|---|---|---|---|---|---|---|---|---|
| 0.05 | 8 | Av.Cov | 0.976 | 0.956 | 0.949 | 0.941 | 0.852 | 0.941 | 0.950 |
| 0.05 | 8 | $u_0$ | 0.312 | 0.250 | 0.293 | 0.253 | 0.000 | 0.208 | 0.239 |
| 0.05 | 20 | Av.Cov | 0.971 | 0.954 | 0.953 | 0.947 | 0.903 | 0.946 | 0.950 |
| 0.05 | 20 | $u_0$ | 0.139 | 0.109 | 0.141 | 0.119 | 0.000 | 0.091 | 0.105 |
| 0.05 | 50 | Av.Cov | 0.966 | 0.952 | 0.953 | 0.949 | 0.928 | 0.948 | 0.950 |
| 0.05 | 50 | $u_0$ | 0.058 | 0.045 | 0.062 | 0.051 | 0.000 | 0.038 | 0.043 |
| 0.025 | 8 | Av.Cov | 0.989 | 0.979 | 0.972 | 0.966 | 0.867 | 0.969 | 0.975 |
| 0.025 | 8 | $u_0$ | 0.369 | 0.312 | 0.372 | 0.324 | 0.000 | 0.262 | 0.297 |
| 0.025 | 20 | Av.Cov | 0.986 | 0.977 | 0.976 | 0.972 | 0.923 | 0.972 | 0.975 |
| 0.025 | 20 | $u_0$ | 0.168 | 0.139 | 0.190 | 0.161 | 0.000 | 0.117 | 0.133 |
| 0.025 | 50 | Av.Cov | 0.983 | 0.976 | 0.977 | 0.974 | 0.950 | 0.974 | 0.975 |
| 0.025 | 50 | $u_0$ | 0.071 | 0.058 | 0.085 | 0.071 | 0.000 | 0.049 | 0.056 |
| 0.005 | 8 | Av.Cov | 0.998 | 0.996 | 0.991 | 0.988 | 0.882 | 0.993 | 0.995 |
| 0.005 | 8 | $u_0$ | 0.484 | 0.438 | 0.509 | 0.453 | 0.000 | 0.379 | 0.417 |
| 0.005 | 20 | Av.Cov | 0.998 | 0.996 | 0.994 | 0.992 | 0.941 | 0.994 | 0.995 |
| 0.005 | 20 | $u_0$ | 0.233 | 0.206 | 0.289 | 0.249 | 0.000 | 0.177 | 0.196 |
| 0.005 | 50 | Av.Cov | 0.997 | 0.995 | 0.995 | 0.994 | 0.970 | 0.995 | 0.995 |
| 0.005 | 50 | $u_0$ | 0.101 | 0.088 | 0.139 | 0.117 | 0.000 | 0.075 | 0.084 |

As mentioned in the introduction, it has been argued that a good interval estimator should (i) give short intervals, and (ii) give coverage probabilites

that usually are close to the nominal level. To examine criterion (ii), Agresti and Coull (1998) calculate the root mean-square error (RMSE) of coverage probability to measure how the actual coverage probability typically varies from the nominal confidence level. They argue that a good method of forming confidence intervals should have a low value of RMSE. So, we will calculate the RMSE of each method's coverage, which is determined over the truncated range $(u_0, 1)$. This RMSE is given by

$$\text{RMSE} = \left[ \frac{1}{1-u_0} \int_{u_0}^{1} \{C_u(p) - (1-\alpha)\}^2 \, dp \right]^{1/2}, \tag{3.17}$$

where $1-\alpha$ is the nominal confidence level. Table 3.2 shows the RMSE of each method for $\alpha = 0.05, 0.025, 0.005$ and $n = 8, 20, 50$. Proposition 3.4 implies that the mid-$p$ method has the minimum possible RMSE of any method of forming a confidence interval estimate that does not use randomisation. Consequently, the mid-$p$ method has the smallest RMSE in every row of Table 2. The new method, OLC, has the second smallest RMSE in every row and is always only a little bigger, at most 20% bigger, than the mid-$p$ method. This is much better than the RMSE of the Clopper-Pearson, which is sometimes more than 45% bigger than the RMSE of the mid-$p$ method. Moreover, each of the other methods has at least one RMSE that is more than 80% bigger than the mid-$p$ method, with the Wald method often doing extremely badly. Hence, the OLC method has a very respectable RMSE, even if it is not the optimal method.

Table 3.2: Root mean-square error (RMSE) of coverage of upper-tail $1-\alpha$ intervals for seven methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and sample sizes ($n$) of 8, 20 and 50.

| $\alpha$ | $n$ | Clopper Pearson | Mid-$p$ | Agresti Coull | Wilson | Wald | Jeff. | OLC |
|---|---|---|---|---|---|---|---|---|
| 0.05 | 8 | 0.0290 | 0.0224 | 0.0267 | 0.0313 | 0.2249 | 0.0314 | 0.0242 |
| 0.05 | 20 | 0.0235 | 0.0166 | 0.0188 | 0.0213 | 0.1481 | 0.0212 | 0.0175 |
| 0.05 | 50 | 0.0180 | 0.0119 | 0.0139 | 0.0151 | 0.0972 | 0.0144 | 0.0124 |
| 0.025 | 8 | 0.0153 | 0.0121 | 0.0177 | 0.0237 | 0.2334 | 0.0186 | 0.0134 |
| 0.025 | 20 | 0.0125 | 0.0091 | 0.0118 | 0.0157 | 0.1522 | 0.0123 | 0.0097 |
| 0.025 | 50 | 0.0097 | 0.0066 | 0.0088 | 0.0110 | 0.0983 | 0.0083 | 0.0069 |
| 0.005 | 8 | 0.0033 | 0.0027 | 0.0085 | 0.0147 | 0.2407 | 0.0051 | 0.0032 |
| 0.005 | 20 | 0.0028 | 0.0021 | 0.0041 | 0.0088 | 0.1557 | 0.0032 | 0.0023 |
| 0.005 | 50 | 0.0022 | 0.0015 | 0.0028 | 0.0057 | 0.0997 | 0.0021 | 0.0017 |

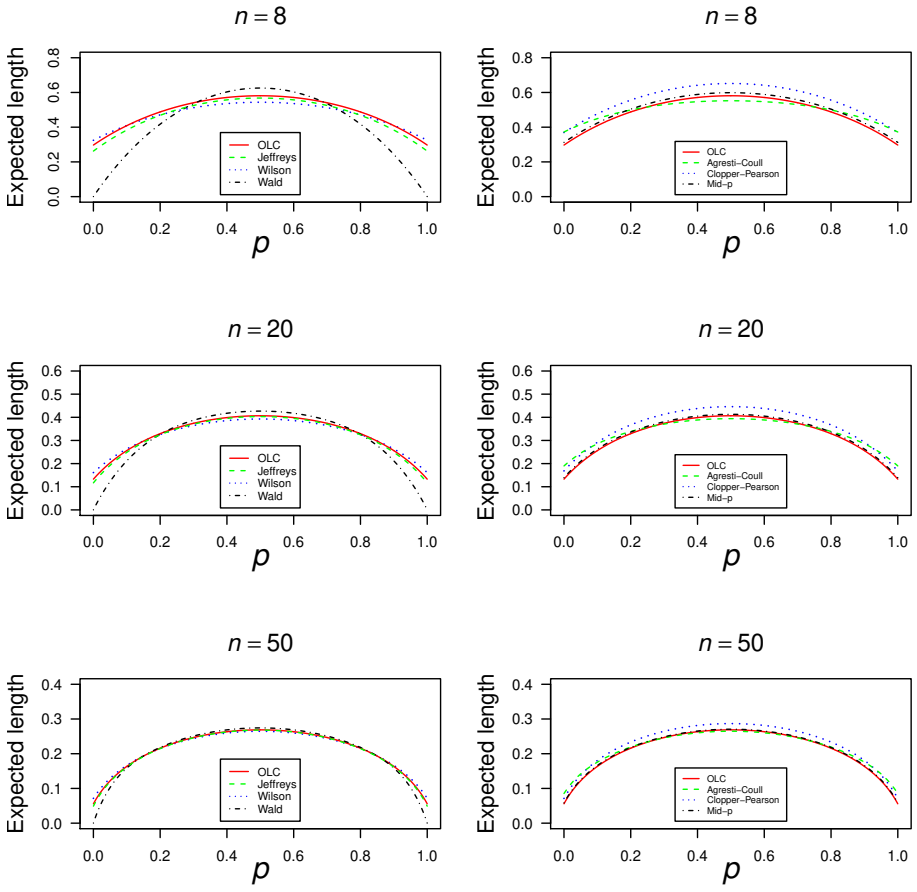


Figure 3.8: Expected lengths of two-sided 95% interval estimates for the OLC, Jeffreys, Wilson and Wald methods (left-hand panels) and the OLC, Agresti-Coull, Clopper-Pearson and mid-$p$ methods (right-hand panels), plotted against $p$ for sample sizes of 8, 20 and 50.

Turning to the length of the intervals, two-tail intervals are examined as the length of one-tailed intervals varies too much with the value of $p$. The

length of one-tail intervals is approximately proportional to $p$ for upper-tail intervals and to $1-p$ for lower-tail intervals. In Figure 3.8, the expected length of 95% two-tailed intervals are plotted against $p$ for $n = 8, 20, 50$ for the OLC, Jeffreys, Wilson and Wald methods (left-hand panels) and the OLC, Agresti-Coull, Clopper-Pearson and mid-$p$ methods (right-hand panels), where OLC method is included in all plots. For all values of $p$ and each combination of $n$ and $\alpha$, the expected lengths of the OLC, mid-$p$, Agresti-Coull, Wilson and Jeffreys intervals are all very similar, and a little smaller than the expected lengths of the Clopper-Pearson intervals. Wald intervals have a much smaller expected length than other methods when $p$ is quite large or quite small, but it only achieves this by giving coverages that are well-short of the nominal confidence level. This expected length is used to calculate the average expected length (AEL) for each method as it is defined in equation (3.13).

Table 3.3 gives the AEL of the methods considered earlier for each combination of $\alpha = 0.05, 0.025, 0.005$ and $n = 8, 20, 50$. It is clear that, apart from the Wald method and Clopper- Pearson, the AEL of the OLC method is usually similar in size to that of the other methods, and is always shorter than the mid-p method. An exception is for $\alpha = 0.005$ and $n = 8$, when its AEL is much poorer than the AEL of Wilson's method. However, Table 3.1 shows that Wilson's method gives a poor coverage for this combination of $\alpha$ and $n$ (the non-coverage in the upper tail is 0.012 rather than the nominal value of 0.005). Hence, it is reasonable to conclude that the AEL of the OLC method compares satisfactorily with that of other methods, so there is little cost in requiring intervals to be locally correct. In fact, our results and conclusion about the AEL for all compared methods, surely excepted our OLC method, are matching with the results of Schilling and Doi (2014).

Table 3.3: Average expected length (AEL) of two-tail $1-2\alpha$ intervals for seven methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and $n = 8$, 20, 50.

| $\alpha$ | $n$ | Clopper Pearson | Mid-$p$ | Agresti Coull | Wilson | Wald | Jeff. | OLC |
|---|---|---|---|---|---|---|---|---|
| 0.050 | 8 | 0.497 | 0.435 | 0.427 | 0.407 | 0.372 | 0.402 | 0.421 |
| 0.050 | 20 | 0.317 | 0.283 | 0.284 | 0.275 | 0.268 | 0.273 | 0.278 |
| 0.050 | 50 | 0.197 | 0.181 | 0.182 | 0.179 | 0.178 | 0.178 | 0.179 |
| 0.025 | 8 | 0.561 | 0.508 | 0.499 | 0.474 | 0.427 | 0.472 | 0.492 |
| 0.025 | 20 | 0.366 | 0.335 | 0.337 | 0.325 | 0.316 | 0.323 | 0.328 |
| 0.025 | 50 | 0.231 | 0.215 | 0.218 | 0.213 | 0.211 | 0.212 | 0.213 |
| 0.005 | 8 | 0.673 | 0.634 | 0.614 | 0.586 | 0.520 | 0.597 | 0.617 |
| 0.005 | 20 | 0.457 | 0.431 | 0.435 | 0.417 | 0.403 | 0.417 | 0.423 |
| 0.005 | 50 | 0.295 | 0.281 | 0.286 | 0.278 | 0.275 | 0.276 | 0.278 |

## 3.6 Concluding comments

The aim in this chapter was to find a satisfactory criterion for choosing an interval estimator for a binomial proportion. As mentioned earlier, with an appropriate criterion there should be some interval estimators that (1) satisfy the new criterion, (2) give intuitively sensible intervals, and (3) give intervals whose average length is acceptably short. As a criterion we proposed that an interval estimator should yield locally correct confidence intervals, meaning that the average coverage between any pair of consecutive spikes should at least equal the nominal confidence level. Three of the methods that were examined met this criterion: the Clopper-Pearson method, the mid-$p$ method and the new OLC method. The Clopper-Pearson method does not satisfy point (3), as the conservative length of its intervals has motivated other researchers over the years to construct many other methods of forming confidence intervals for a binomial proportion. One of these other methods is the mid-$p$ method, which has been recommended because it gives shorter intervals (Vollset, 1993; Agresti and Gottard, 1995). The mid-$p$ method also gives intuitively sensible intervals so it meets points (1)–(3) above.

Turning to the OLC method, in the examples given in Figure 2.1, the method

gave end-points that are fairly evenly spaced with coverages that are balanced around the nominal confidence level. This has also been the case in every other example we have examined. Blythe and Still (1983) and Schilling and Doi (2014) list some properties that are desirable in an interval estimator, such as equivariance and monotonicity. It was shown that the OLC method has these properties for the many combinations of $n$ and $\alpha$ that were examined through extensive computation [$1 \leq n \leq 200$ and confidence levels in $\{0.001(0.001)0.27\}$]. Hence it is reasonable to conclude that the new method OLC gives sensible intervals and meets point (2). Regarding the third point, expected length of intervals, six methods of constructing equal-tail confidence intervals were compared in Section 3.5. The six methods include the mid-$p$, Wald, Wilson, Agresti-Coull and Jeffreys methods, which have each been recommended in preference to the Clopper-Pearson method because of the lengths of their intervals. The intervals given by the OLC method had an average expected length that was shorter than Clopper-Pearson intervals and comparable to those given by other methods for all values of $n$ and $\alpha$ that were examined, except for the Wald method, which gave intervals whose coverage was often much less than the nominal confidence level. Hence the OLC appears to give intervals that are acceptably short and so it meets point (3). As there are at least two methods that meet points (1)–(3), it can be concluded that requiring intervals to be locally correct is a reasonable criterion to place on an interval estimator. Choosing between the mid-$p$ method and the OLC method is tricky because each has an optimality property. On the one hand, for any value of $p$, the coverage of one-tail intervals is as close to the nominal level as possible when intervals are determined using the mid-$p$ method (cf. Proposition 3.4). On the other hand, the average expected length of intervals is smaller with the OLC method than with any other method that gives locally

correct intervals (cf. Proposition 3.2). However, in choosing between estimators, both coverage and length of intervals are important. Hence either (i) a restriction should be placed on coverage and methods should be distinguished on the basis of interval width, or (ii) a restriction should be placed on interval width and methods distinguished on the basis of coverage. Requiring an interval to be locally correct places a restriction on coverage, so distinguishing between methods on the basis of interval length seems appropriate, in which case the OLC method is the preferred interval estimator.

The proposed algorithm for the OLC method in this chapter has made it clear that the OLC confidence intervals for binomial proportions can easily be calculated in practice for real datasets of applied importance. We give an example here for calculating an OLC confidence interval for a binomial proportion based on a dataset from an important medical application for which the Wald confidence interval is calculated in a text book.

The Open University introduces a second level module, M248, on analysing data. Activity 13 in Book B of this module (The Open University-M248, 2017, p 165) discusses a clinical trial that examined the effect of taking a low dose of penicillin for twelve months on the recurrence of leg cellulitis in patients who had previously two or more episodes of leg cellulitis. The trial contains 136 patients of which 30 patients had a recurrence of leg cellulitis during the twelve-month of treatment. In that activity, based on the given dataset, a 90% Wald confidence interval is calculated for the proportion of patients with leg cellulitis whose cellulitis recurred during the time they were taking penicillin. The 90% Wald confidence interval is reported to be (0.162, 0.280).

The proposed algorithm for the OLC confidence intervals has been used with the same dataset, for $n = 136$ and $x = 30$. This gives the 90% OLC confidence interval for the proportion of patients with leg cellulitis whose cellulitis

recurred during the time they were taking penicillin. The calculated OLC confidence interval is obtained as (0.167, 0.283). It is obvious that the endpoints of the OLC confidence interval are rather close to the corresponding endpoints of the Wald confidence interval. This is due to the fact that the normal approximation used in the Wald confidence interval is accurate enough since the sample size is large. However, the proposed OLC confidence interval is clearly shorter.

# Chapter 4

# General results for the OLC method

## 4.1 Introduction

In the previous chapter, the important task of forming an interval estimator for a binomial probability was addressed. A new definition of an interval estimator was defined and the OLC method was developed and shown to work well. The saw-tooth pattern of the coverage occurs with any sampling distribution that is discrete, so it is natural to try to extend this work without focusing on any specific discrete sampling distribution, thus obtaining results that apply more generally. The results will be used in Chapters 5 and 6, where interval estimators for the Poisson distribution and negative binomial distribution are considered.

There are two main issues to address:

1. With the binomial distribution, the OLC method starts by setting a confidence limit equal to an endpoint of the range of the parameter space when the sample statistic takes a value at the edge of its domain. When determining an upper limit it starts by setting $u_n$ equal to 1 (the upper endpoint of the range of $p$) and for a lower limit it starts by setting $l_0$ equal to 0 (the lower endpoint of the range of $p$). For some discrete sampling distributions, the domain of the sample statistic may be unbounded, when the range of the parameter space may also be unbounded. For example, the sample value of a Poisson variate can be any non-negative integer and the range of the Poisson parameter, $\lambda$, is $(0, \infty)$. The first task is to modify the OLC so that it can be used with unbounded sample spaces.

2. For the binomial distribution it was shown that the OLC method gave intervals with the smallest average width of any method that gave locally correct confidence intervals. The proof of this result is tough and will

not generalise. Instead, conditions are given under which a weaker result about average width holds. For many discrete sampling distributions, the conditions are straightforward to check though it requires repetitive computation.

These issues are examined in Sections 4.2 and 4.3. Also there is

3. As noted in Section 2.4.4, there are a number of properties that are desirable in methods of forming confidence intervals. The extent to which the OLC method generally holds these properties is examined in Section 4.4.

## 4.2 Modifying the OLC method for unbounded sample spaces

The nature of the parameter space and the random variable $X$ depends upon the sampling distribution and the parametrization. Some distributions have a parameter space and random variable with finite ranges e.g. the binomial distribution, its parameter $p \in [0, 1]$ and the observed value $x$ of its random variable takes a value from $0, 1, \ldots, n$. Other distributions have a parameter space and a random variable with infinite range e.g. the Poisson distribution; its parameter $\lambda \in [0, \infty)$ and the observed value $x$ of its random variable can be $0, 1, \ldots$. The range of both the parameter space and the random variable affect the OLC method of calculating the endpoints of the confidence interval for the parameter.

For clarification, in the case of a finite range with upper bound 1, if the value of $x$ is limited by the value of $n$, the endpoints of the confidence interval for the upper tail are calculated by setting the last endpoint $u_n = 1$. Then the method determines the previous endpoint $u_{n-1}$, which depends on the given endpoint $u_n$, and so on. So, each $u_i$ can be only calculated after $u_{i+1}$ has been

obtained. A similar strategy is used to calculate the endpoints for the lower tail. If the range of the parameter is bounded below by 0 and the smallest possible value of $x = 0$, we set the first lower endpoint $l_0 = 0$ and sequentially calculate the following endpoints $l_1, l_2, \ldots$. So, each $l_{i+1}$ is determined by $l_i$. This is explained in detail in the case of the binomial distribution in Section 3.4 in Chapter 3. In the case where the range of the random variable $x$ goes to infinity, the OLC method must be modified.

Let $\theta$ denote the parameter of interest and suppose that the range of $\theta$ is unbounded above. For definiteness, suppose an upper-tail $(1 - \alpha)$ interval for $\theta$ is to be constructed. To obtain the first confidence limit in the iteration, we select a large value of $x$, which is denoted by $N$, and need to choose a reasonable value for $u_N$, the upper endpoint of the upper-tail interval when $x = N$. This value can be calculated by using one of the classical methods of constructing confidence interval, e.g. the mid-$p$ or exact methods.

We take this upper limit as the upper limit of the interval given by the OLC method. Then we determine $u_{N-1}$ so that equation (3.1) is satisfied. Then $u_{N-2}$ is obtained using the value of $u_{N-1}$, and so on until we get $u_x$. An important question is whether this method gives a value of $u_x$ that is somewhat arbitrary. If we chose a different (large) value for $N$ would we get a different value for $u_x$, where $x < N$? Would we get a different value if we used Jeffreys method or some other method to obtain $u_N$?

Suppose $u_N, u_{N-1}, u_{N-2}, \ldots$ are the upper limits given by the OLC method when the method starts with $u_N$ as the upper limit at $x = N$, and that $u_N^*, u_{N-1}^*, u_{N-2}^*, \ldots$ are the upper limits when it starts with $u_N^*$ as the upper limit when $x = N$. Let $C_{u,i}(\theta) = Pr(X \geq i|\theta)$. When $u_N$ is the upper limit at $x = N$, $C_{u,i}(\theta)$ is the coverage probability when $\theta \in (u_{i-1}, u_i)$ and, when $u_N^*$ is the upper limit, $C_{u,i}(\theta)$ is the coverage probability when $\theta \in (u_{i-1}^*, u_i^*)$

Thus, for $i = N, N-1, N-2, \ldots$,

$$\frac{1}{u_i - u_{i-1}} \int_{\theta=u_{i-1}}^{u_i} C_{u,i}(\theta)\, d\theta = 1 - \alpha \tag{4.1}$$

and

$$\frac{1}{u_i^* - u_{i-1}^*} \int_{\theta=u_{i-1}^*}^{u_i^*} C_{u,i}(\theta)\, d\theta = 1 - \alpha. \tag{4.2}$$

The coverage probability is generally an "S"-shaped function of $\theta$ and hence concave for larger values of $\theta$. (An exception is the binomial distribution $binomial(n, p)$ when $X = n$.) For almost all sampling models in which $\theta$ has an unbounded upper limit, $C_{u,i}(\theta)$ will be a concave function of $\theta$ when $1 - \alpha > 0.7$ and $\theta$ is greater than $u_i$ and $u_i^*$. The following proposition shows that $|u_i - u_i^*| > |u_{i-1} - u_{i-1}^*|$ for $i = N, N-1, N-2, \ldots$, provided the shape of the coverage function is concave in the regions that are relevant to determining the interval endpoints. Consequently, for a given value of x the value of $u_x$ will not be arbitrary if $N$ is sufficiently large $N >> x$, provided the conditions of the proposition hold. Proof of the proposition is given in Appendix C.

**Proposition 4.1**

Suppose , $C_{u,i}(\theta)$ is a concave function of $\theta$ for $\theta > \min(u_{i-1}, u_{i-1}^*)$. Then $|u_{i-1} - u_{i-1}^*| < |u_i - u_i^*|$.

A similar approach is adopted for the lower endpoint if the value of $X$ has no lower bound. A large negative value ($-M$ say) is chosen for $X$ and the lower endpoint of a $1-\alpha$ lower-tail interval for $\theta$ is calculated using a standard method of forming confidence intervals. This endpoint is taken as $l_{-M}$ and then $l_{-M+1}$, $l_{-M+2}$, ... are constructed sequentially using the OLC method until $l_x$ is obtained.

## 4.3 Optimality properties for width of intervals

The width of interval estimates is of paramount importance – it is the reason that so many different interval estimators have been proposed. For the bino-

mial distribution it was possible to show that the OLC method gives intervals with the shortest average expected length of any interval estimator that gives locally correct confidence (LCC) intervals. Obtaining a result of this type that holds for a number of discrete distributions is not possible. However, it is possible to give conditions under which *slight* changes to the endpoints given by the OLC method will always increase the average expected length of intervals. We shall refer to this as a locally optimum property, in the same way that examining the first and second derivatives of a function can show that it is a local minimum, but does not show that it is a global minimum.

For definiteness, we again suppose an upper-tail $(1-\alpha)$ interval for $\theta$ is required. We first consider the case where $\theta$ has a finite upper bound, $\Theta$ say, and $X$ is finite, taking one of the value $0, 1, \ldots, n$. For $i = 0, \ldots, n$, let $(0, u_i)$ be the upper-tail interval given by the OLC method and define $p_i^*$ by

$$p_i^* = \int_{\theta=0}^{\Theta} Pr(X = i|\theta)\, d\theta. \tag{4.3}$$

Then, if $\theta$ is equally likely to be any value in the interval $(0, \Theta)$, the average expected length of the upper-tail interval is $\sum p_i^* u_i$.

As $u_0, \ldots, u_n$ are the endpoints given by the OLC method, they satisfy equation (4.1) where, again, $C_{u,i}(\theta^*) = Pr(X \geq i|\theta = \theta^*)$. Let $(u_0^*, \ldots, u_n^*)$ be a partition that gives locally correct confidence intervals and for which

$$|u_i^* - u_i| < \delta \qquad \text{for } i = 0, \ldots n. \tag{4.4}$$

Proposition 4.2 gives necessary and sufficient conditions for $\sum_{i=0}^{n} p_i^* u_i^* > \sum_{i=0}^{n} p_i^* u_i$ if $\delta$ is sufficiently small. That is, it gives conditions for the OLC method to yield locally correct confidence intervals with an expected average length that is a local minimum. For a specified sampling model it is usually relatively straightforward to check whether the conditions of the proposition hold for a specified value of $\alpha$.

Proof of the proposition is given in Appendix D.

**Proposition 4.2**

For $i = 1, \ldots, n$, let $h_i = C_{u,i}(u_i) - (1-\alpha)$ and $f_i = C_{u,i}(u_{i-1}) - (1-\alpha)$. For $1 \leq j \leq l \leq n$, define $\psi_{jl} = \prod_{i=j}^{l}(h_i/f_i)$ and put $\phi_l = p_l^* + \sum_{j=1}^{l} p_{j-1}^* \psi_{jl}$ for $l = 1, \ldots, n$, where $p_i > 0$ for $i = 0, \ldots, n$.

(a) Suppose $\phi_l < 0$ for some $l \in (1, \ldots, n-1)$. Then for any $\delta > 0$ there exists a partition $(u_0^*, \ldots, u_n^*)$ such that

$$\sum_{i=0}^{n} p_i^* u_i^* < \sum_{i=0}^{n} p_i^* u_i,$$

with (4.4) satisfied for $i = 0, \ldots, n.$ and

$$\frac{1}{u_i^* - u_{i-1}^*} \int_{\theta=u_{i-1}^*}^{u_i^*} C_{u,i}(\theta)\, d\theta \leq 1 - \alpha \tag{4.5}$$

for $i = 1, \ldots, n$.

(b) Suppose $\phi_l > 0$ for $i = 1, \ldots, n-1$. Then there exists $\delta > 0$ such that

$$\sum_{i=0}^{n} p_i^* u_i < \sum_{i=0}^{n} p_i^* u_i^*$$

if (4.4) holds for $i = 0, \ldots, n$ and (4.5) holds for $i = 1, \ldots, n$.

The result in Proposition 4.2 is generally of little use when the range of the parameter space has no upper bound and the domain of $X$ is $0, 1, \ldots$, with no upper bound. This is because the length of both one-tail and two-tail intervals will typically increase without bound as $X$ increases, becoming infinite, and comparing the size of two infinite quantities is tricky. Rather than compare methods on the basis of the average expected length of their intervals, it is better to form some weighted average of the expected lengths of intervals. An appropriate set of weights would give greater weight to the length of intervals when $X$ is small than when $X$ is large, and weights would decrease sufficiently quickly for the weighted average length of intervals to be finite.

Suppose again that an upper-tail $(1-\alpha)$ interval for $\theta$ is required but now the possible values of $X$ are $0, 1, \ldots$ with no upper bound. Applying the approach advocated in Section 4.2, a large value for $n$ is selected and, for $i = n, n-1, \ldots, 0$, the OLC method gives $u_i$ as the upper endpoint of its interval when $X = i$. We suppose $u_n$ is fixed but *slight* changes may be made to $u_0, \ldots, u_{n-1}$, yielding a partition $(u_0^*, \ldots, u_n^*)$ that gives locally correct confidence intervals and for which $u_n^* = u_n$ and equation (4.4) is satisfied. Then the result of Proposition 4.2 can be applied to $\sum_{i=0}^{n} p_i u_i$, where $p_0, \ldots p_n$ are now viewed as weights. This yield the following corollary.

**Corollary 4.1.**

Suppose the conditions given in Proposition 4.2 hold and, in addition, $u_n^* = u_n$. Then the result of the proposition holds for comparison of $\sum_{i=0}^{n} p_i u_i$ with $\sum_{i=0}^{n} p_i u_i^*$.

## 4.4 Desirable properties in confidence intervals

A number of desirable properties in methods of forming confidence intervals were discussed in Section 2.4.4. The generality with which OLC methods have these properties is considered below. The properties are as follows.

**Property 1**. *Interval valued.* A confidence region should be an interval and not a collection of disjoint intervals.

**Property 2**. *Monotonicity in $x$.* This monotonicity may be

- Montone increasing: For fixed $\alpha$, if the point estimator increases monotonically as $x$ increases, then the endpoints of the confidence interval should also increase monotonically. This requires $l(x+1, \alpha) > l(x, \alpha)$ and $u(x+1, \alpha) > u(x, \alpha)$.

- Monotone decreasing: For fixed $\alpha$, if the point estimator decreases monotonically as $x$ increases, then the endpoints of the confidence interval

should also decrease monotonically. This requires $l(x+1,\alpha) < l(x,\alpha)$ and $u(x+1,\alpha) < u(x,\alpha)$.

**Property 3**. *Nesting.* If two confidence intervals have different confidence levels then, for any given random variable, say $X$, the interval for the higher confidence level should contain the interval for the lower confidence level. Suppose we have two confidence levels $1-\alpha_1$ and $1-\alpha_2$ with $\alpha_1 < \alpha_2$, this requires $(l(x,\alpha_2), u(x,\alpha_2)) \in (l(x,\alpha_1), u(x,\alpha_1))$. For this to occur for all confidence levels, as the level increases the lower limit for each random variable must be non-increasing, and the upper limit must be non-decreasing.

The OLC method will not always be usable, For example, with the binomial distribution $x$ could not be used to form confidence interval if $1-\alpha$ was less than 0.4. When it is usable, by logic, from the way it forms confidence intervals, it always gives confidence intervals that are a single interval and hence the OLC method has Property 1. Similarly, its method of construction means that its intervals always meet the *monotonicity in x* property. Checking whether the OLC method has the *nesting* property is difficult. For the binomial distribution, direct computation showed that it has the property when $1 \leq n \leq 200$ and $\alpha_1$ and $\alpha_2$ are in $\{0.001(0.001)0.27\}$. This is a fine discretization, suggesting the OLC must have the property for $1 \leq n \leq 200$ and $\{0.001 \leq \alpha_1, \alpha_2 < 0.27\}$. It seems likely that this type of result about nesting will hold for many sampling models.

## 4.5 Concluding comments

This chapter has given a number of results that commonly hold for the OLC method. Examples that exploit the results and illustrate their usefulness are given in Chapters 5 and 6.

# Chapter 5

# Poisson confidence interval methods

## 5.1 Introduction

There are fewer methods of forming interval estimators for a Poisson mean than for a binomial proportion. The most important of these have been described in the Poisson section of Chapter 2. Other than the Garwood method, the methods that aim to give equal-tail intervals do not meet the strict definition of a confidence interval. As with the binomial distribution, the main obstacle facing the construction of confidence intervals for the Poisson mean ($\lambda$) is the discrete nature of the sample space, which produces spikes. This leads to wide fluctuations in the coverage probability as $\lambda$ varies. In Chapter 3, the notion of methods that give *locally correct confidence (LCC) intervals* was introduced. Methods give LCC intervals if the average coverage between spikes is at least as large as the nominal confidence level. For the binomial distribution, methods could be found which met this definition and which (1) gave sensible intervals and (2) gave intervals whose average width is acceptably short. In this chapter a method that gives LCC intervals for a Poisson mean is developed, examined and compared with other methods.

The Poisson distribution is simpler than the binomial distribution, as it has only one parameter $\lambda$ rather than the two parameters of the binomial ($n$ and $p$). However, additional challenges arise with the Poisson distribution because the range of $\lambda$ has no upper bound and because the values that $x$ can take also have no upper bound. In Section 4.2 a modification of the OLC method was suggested that is designed to handle unbounded sample spaces. The modification requires a large value for $x$ to be chosen, $N$ say, and the algorithm is started by specifying an upper interval limit for the case $x = N$. Choice of the value of $N$ will be discussed in detail.

This chapter will examine methods over the range $\lambda \in (0, 50]$, which is the

range of $\lambda$ that is considered in a number of references, e.g. Patil and Kulkarni (2012), and it covers the values of $\lambda$ that are usually of interval in practice. When comparing methods here, this range of $\lambda$ will be divided into three regions, namely $(0, 2), (2, 5)$ and $(5, 50)$. The use of weighted average expected length is also used to compare methods with a weight function that yields a finite weighted average expected length. This provides an overall measure of performance over the full range of $\lambda$.

In this chapter, we will apply the new approach of computing the confidence interval for a binomial proportion to calculate the confidence interval for the Poisson mean $\lambda$. In Section 5.2, we give a precise definition of a locally correct confidence interval that is appropriate for unbounded sample spaces. In Section 5.3, we modify the interval estimator for the binomial distribution and obtain the optimal locally correct (OLC) method for the Poisson sampling distribution. We examine whether intervals given by the OLC method seem sensible and examine whether the new estimator has properties that have been proposed in the literature as being desirable. Also, it is shown that the OLC method has an optimality property regarding the length of the intervals it yields. In Section 5.4, the OLC method is compared with several methods that have been recommended for forming equal-tailed confidence intervals (as noted in the literature chapter). Concluding comments are given in Section 5.5.

## 5.2 Locally correct confidence intervals

In this chapter, let $X$ denote a Poisson random variable with mean $\lambda$ and suppose an interval estimator gives $(0, u_x)$ as its upper-tail estimate for $\lambda$ when $x$ is the observed values of $X$. For a reasonable estimator,

$$0 < u_0 < u_1 < u_2 < \dots. \tag{5.1}$$

The coverage probability is the probability that the random interval $(0, u_x)$ contains $\lambda$, so it depends on the value of $\lambda$. We indicate this coverage probability by $C_{u,i}(\lambda)$, when $u_{i-1} < \lambda < u_i$

$$C_{u,i}(\lambda) = Pr(X \geq i|\lambda) = 1 - \sum_{x=0}^{i-1} e^{-\lambda}\lambda^x/x! \tag{5.2}$$

The points of the spikes occur where $\lambda$ equals $u_0, ....., u_\infty$ and the coverage drops by $e^{u_x}u_x^x/x!$

The focus is again on methods whose average coverage between spikes is greater than or equal to the nominal level. However the *modified OLC method*, proposed in the next section, only determines endoints $u_0, u_1, \ldots, u_N$, where $N$ is a large number chosen by the user. This leads to a slightly less stringent definition for an interval estimator to give LCC intervals.

*Definition 1.* For the upper-tail interval, suppose an interval estimator gives $(0, u_x)$ as its upper-tail interval for $\lambda$ when $X = x$ for $x = 0, \ldots, N$ and that $u_0, u_1, \ldots, u_N$ satisfy equation (5.1). If, for $i = 1, 2, \ldots, N$

$$\frac{1}{u_i - u_{i-1}} \int_{\lambda=u_{i-1}}^{u_i} C_{u,i}(\lambda)d\lambda \geq (1 - \alpha) \tag{5.3}$$

and then the interval estimator gives *upper-tail LCC* intervals with confidence level $(1 - \alpha)$ for $x \leq N$.

*Definition 2.* For the lower-tail interval, suppose an interval estimator gives $(l_x, \infty)$ as its lower-tail interval for $\lambda$ when $X = x$ for $x = 0, \ldots, N$ and that $0 = l_0 < l_1 < \ldots < l_N$. The coverage probability $C_{l,i}$ is defined as

$$C_{l,i}(\lambda) = Pr(X \leq i|\lambda) = \sum_{x=0}^{i} e^{-\lambda}\lambda^x/x! \tag{5.4}$$

for $i = 1, 2, \ldots, N$. If, for $i = 0, \ldots, N - 1$

$$\frac{1}{l_{i+1} - l_i} \int_{\lambda=l_i}^{l_{i+1}} C_{l,i}(\lambda)d\lambda \geq (1 - \alpha) \tag{5.5}$$

then the interval estimator gives *Lower-tail LCC* intervals with confidence level $(1 - \alpha)$ for $x \leq N$.

*Definition 3.* For the two-sided interval, suppose that, for $x = 0, 1, \ldots, N$, an interval estimator gives a two-sided equal-tail intervals $(l_x, u_x)$ for $\lambda$ when $X = x$. Then it gives equal-tail LCC intervals with the confidence level $(1-2\alpha)$ *for* $x \leq N$ if and only if, for $x = 0, 1, \ldots, N$, the intervals $(l_x, \infty)$ and $(0, u_x)$ are sets of one-sided lower-tail and upper-tail LCC intervals for $x \leq N$, respectively, each with confidence level $(1 - \alpha)$.

For simplicity, an interval estimator that gives equal-tail LCC intervals for $x \leq N$ will be referred as a *LCC interval estimator* unless the value of $N$ is important, when it will be referred to as an LCC interval estimator *of domain* $N$.

## 5.3 The modified OLC method

We propose an interval estimator that uses a straightforward iterative algorithm, similar to that used in the binomial case but with some differences, to obtain one-sided interval estimates. The main difference is in the algorithm for determining one-sided upper tail interval, as the infinite range of $x$ affects the way the upper endpoints are obtained. The algorithm implements the modification for unbounded sample space that was proposed in Section 4.2. The modification requires a large value, $N$, to be chosen and then the algorithm successively calculates $u_N, u_{N-1}, u_{N-2}, \ldots, u_0$. We first give the algorithm and then consider the choice of $N$. For an upper-tail interval with confidence level $1 - \alpha$, the algorithm sets $u_N$ equals to a reasonable upper point. We mentioned in Chapter 4, when the domain is finite for both $x \leq n$ and $p \leq 1$, the algorithm sets $u_n$ (the last interval) equals to 1 (as in the binomial case). But, in the case of an infinite domain for $x = 0, 1, 2 \ldots$ and $0 \leq \lambda \leq \infty$, a reasonable value must be chosen for the upper point of the last interval $u_N$.

This value can be calculated by using one of the classical methods and here the mid-$p$ method is used to calculate ($u_{N.midp}$), before $u_{N-1}, u_{N-2}, \ldots, u_0$ are determined using the same procedure used for the binomial. Specifically, the steps of the algorithm for an upper-tail interval are as follows.

1. Set $u_N = u_{N.midp}$ and put $i = N - 1$.

2. Given $u_{i+1}$, use the bisection method to find the value $u_i$ that makes the average coverage over the interval $(u_i, u_{i+1})$ equal to $1 - \alpha$.

3. Repeat step 2 for $i = N-2, N-3, \ldots, 1, 0$ to obtain $u_{N-2}, u_{N-3}, \ldots, u_1, u_0$.

From its constructions, the method determines the endpoints of subintervals that have an average coverage of $1 - \alpha$. But, after $u_i$ has been attained, it could be the case that $Pr(X \geq i - 1|\lambda)$ is less than $1 - \alpha$, even though $(u_{i+1} - u_i)^{-1} \int_{u_i}^{u_{i+1}} Pr(X \geq i|\lambda)d\lambda = 1 - \alpha$. Then, given $u_i$, there is no $u_{i-1}$ (with $u_{i-1} \leq u_i$ ) that meets the requirement that $(u_i - u_{i-1})^{-1} \int_{u_{i-1}}^{u_i} Pr(X \geq i - 1|\lambda)d\lambda = 1 - \alpha$. Hence the following proposition is needed to underpin the algorithm.

**Proposition 5.1**

Suppose $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$ , suppose also that

$$\frac{1}{u_i - u_{i-1}} \int_{\lambda=u_{i-1}}^{u_i} \left(1 - \sum_{x=0}^{i-1} \exp^{-\lambda} \lambda^x / x!\right) d\lambda = 1 - \alpha \tag{5.6}$$

and $u_i > u_{i-1}$ for $i = j + 1, j + 2, ....., 200$, where 200 is the largest value of $x$. Then there is a unique $u_{j-1}$ such that $u_j > u_{j-1} > 0$ and equation (5.6) holds when $j = i$.

We have been unable to prove the proposition for all values of $x$ from 0 to $\infty$. It is easy to show that any $u_{j-1}$ that satisfies equation (5.6) is unique, but hard to show that $1 - \sum_{x=0}^{i-1} \exp^{-\lambda} \lambda^x / x! > (1 - \alpha)$, which is a requirement

for the existence of a $u_{j-1}$ that satisfies the equation. Instead, we choose a sufficiently large value of $N$, which will be discussed in the following section, to determine the last upper point $u_{N.midp}$. Then repetitive computation showed that the result in Proposition 5.1 holds if $x$ is a positive integer less than 200 and $\alpha$ is one of the numbers $0.001, 0.002, \ldots, 0.3$. So throughout this chapter it is assumed that $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$. From Proposition 5.1, we have that $0 < u_0 < .....u_{200} = u_{N.midp}$. Consequently, under Definition 1 the new estimator gives upper-tail LCC intervals.

Forming the lower-tail intervals is essentially the same as forming the lower-tail intervals in the case of the binomial distribution. That is because the search starts by putting $l_0 = 0$ and then $l_1, ...., l_{200}$ are determined sequentially. Given $l_i$, $l_{i+1}$ will be calculated as follows. Assume that $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$. We find the value for $l_{i+1}$ that satisfies

$$\frac{1}{l_{i+1} - l_i} \int_{\lambda=l_i}^{l_{i+1}} \sum_{x=0}^{i} e^{-\lambda} \lambda^x / x! d\lambda = 1 - \alpha \tag{5.7}$$

where $i = 0, 1, ....199$, then $(l_i, \infty)$ is the $1 - \alpha$ lower-tail interval when $X = i$ and the estimator gives lower-tail LCC intervals from Definition 2.

We obtain the two-sided intervals by combining the endpoints of one-sided intervals. Consequently, when $X = i$ the new estimator gives the two-sided equal-tail interval $(l_i, u_i)$ for a confidence level of $1 - 2\alpha$, where $(l_i, \infty)$ and $(0, u_i)$ are the lower-tail and upper-tail LCC intervals for a confidence level of $1 - \alpha$, respectively. Thus, the new estimator is an LCC interval estimator (of domain $N$).

Turning to the choice of $N$, this is allowed to depend on the observed value of $X$. Let $u_x(n)$ be the upper limit of the $1 - \alpha$ upper-tail interval when $X = x$ and the algorithm starts at $N = n$. From proposition 4.1, as $n \to \infty$ the value of $u_x(n)$ approaches some limit, say $u_x^*$. In choosing $N$, the aim is to choose a

value that is sufficiently large for $u_x(N)$ to differ from $u_x^*$ by an amount that is negligible. At the same time, it is computationally inefficient to use a value for $N$ that is unnecessarily large.

An adequate size for $N$ will depend on the value of $x$. Table 5.1 and 5.2 compare the values of upper points ($u_x(N)$) for our new OLC method at different values of $N$ and $x$. It is clear that the changes in the value of $x$ should influence the value chosen for $N$. When $x = 8$, the difference between the upper points for $N = 42$ and $N = 100$ appears from the third decimal place, while the difference between the upper points for $N = 100$ and $N = 200$ arises the fifth decimal place and difference between $N = 200$ and $N = 1000$ appears after the fifth decimal place. So setting $N$ equal to 42 will be enough when $x = 8$ as choosing $N = 100, 200$ or 1000 will only change the limit by a very small amount. When $x = 20$, the difference between the upper points for $N = 42$ and $N = 100$ appears from the third decimal place, for $N = 100$ and $N = 200$ it appears from the fifth decimal place and, for $N = 200$ and $N = 1000$ it does not appear before the sixth decimal place. This means that increasing the value of $N$ above 100 will make little difference to the values of the upper points. So, $N = 100$ will be a good choice when $x = 20$. Also, when $x = 40$, the difference between the upper points for $N = 42$ and $N = 100$ appears in the second decimal place and, for $N = 100$ and $N = 200$, the difference begins in the third decimal place, while the difference between $N = 200$ and $N = 1000$ appears after the fifth decimal place. So, $N = 200$ is clearly large enough when $x = 40$. Based on these results, setting $N$ equals to the larger of 40 and $5x$ should be a reasonable choice in practice. We refer to the new method as the modified OLC method.

Table 5.1: Upper points of the OLC method at different values of $N$ for $0 \leq x \leq 20$

| x | $N = 42$ | $N = 100$ | $N = 200$ | $N = 1000$ |
|---|---|---|---|---|
| 1 | 2.22416 | 2.22414 | 2.22413 | 2.22413 |
| 2 | 4.03284 | 4.03289 | 4.03290 | 4.03290 |
| 3 | 5.61799 | 5.61791 | 5.61791 | 5.61791 |
| 4 | 7.09690 | 7.09701 | 7.09701 | 7.09701 |
| 5 | 8.51209 | 8.51194 | 8.51194 | 8.51194 |
| 6 | 9.88264 | 9.88283 | 9.88284 | 9.88284 |
| 7 | 11.22143 | 11.22119 | 11.22119 | 11.22119 |
| 8 | 12.53394 | 12.53423 | 12.53424 | 12.53424 |
| 9 | 13.82727 | 13.82691 | 13.82691 | 13.82691 |
| 10 | 15.10226 | 15.10269 | 15.10269 | 15.10270 |
| 11 | 16.36475 | 16.36423 | 16.36422 | 16.36422 |
| 12 | 17.61286 | 17.61347 | 17.61348 | 17.61348 |
| 13 | 18.85279 | 18.85206 | 18.85205 | 18.85205 |
| 14 | 20.08033 | 20.08118 | 20.08119 | 20.08119 |
| 15 | 21.30294 | 21.30196 | 21.30194 | 21.30194 |
| 16 | 22.51400 | 22.51514 | 22.51516 | 22.51516 |
| 17 | 23.72289 | 23.72158 | 23.72156 | 23.72156 |
| 18 | 24.92023 | 24.92174 | 24.92176 | 24.92176 |
| 19 | 26.11803 | 26.11631 | 26.11628 | 26.11629 |
| 20 | 27.30361 | 27.30557 | 27.30559 | 27.30559 |

Table 5.2: Upper points of the OLC method at different values of $N$ for $21 \leq x \leq 40$

| x | $N = 42$ | $N = 100$ | $N = 200$ | $N = 1000$ |
|---|---|---|---|---|
| 21 | 28.49233 | 28.49011 | 28.49008 | 28.49008 |
| 22 | 29.66756 | 29.67007 | 29.67009 | 29.67010 |
| 23 | 30.84882 | 30.84599 | 30.84596 | 30.84596 |
| 24 | 32.01471 | 32.01789 | 32.01793 | 32.01793 |
| 25 | 33.18987 | 33.18632 | 33.18627 | 33.18627 |
| 26 | 34.34715 | 34.35113 | 34.35119 | 34.35119 |
| 27 | 35.51739 | 35.51295 | 35.51289 | 35.51289 |
| 28 | 36.66654 | 36.67149 | 36.67155 | 36.67155 |
| 29 | 37.83292 | 37.82742 | 37.82735 | 37.82735 |
| 30 | 38.97423 | 38.98034 | 38.98042 | 38.98042 |
| 31 | 40.13776 | 40.13100 | 40.13091 | 40.13091 |
| 32 | 41.27137 | 41.27884 | 41.27894 | 41.27894 |
| 33 | 42.43300 | 42.42474 | 42.42463 | 42.42463 |
| 34 | 43.55888 | 43.56797 | 43.56809 | 43.56809 |
| 35 | 44.71957 | 44.70955 | 44.70942 | 44.70942 |
| 36 | 45.83756 | 45.84856 | 45.84871 | 45.84871 |
| 37 | 46.99829 | 46.98619 | 46.98604 | 46.98604 |
| 38 | 48.10808 | 48.12132 | 48.12149 | 48.12149 |
| 39 | 49.26985 | 49.25534 | 49.25515 | 49.25515 |
| 40 | 50.37102 | 50.38686 | 50.38708 | 50.38707 |

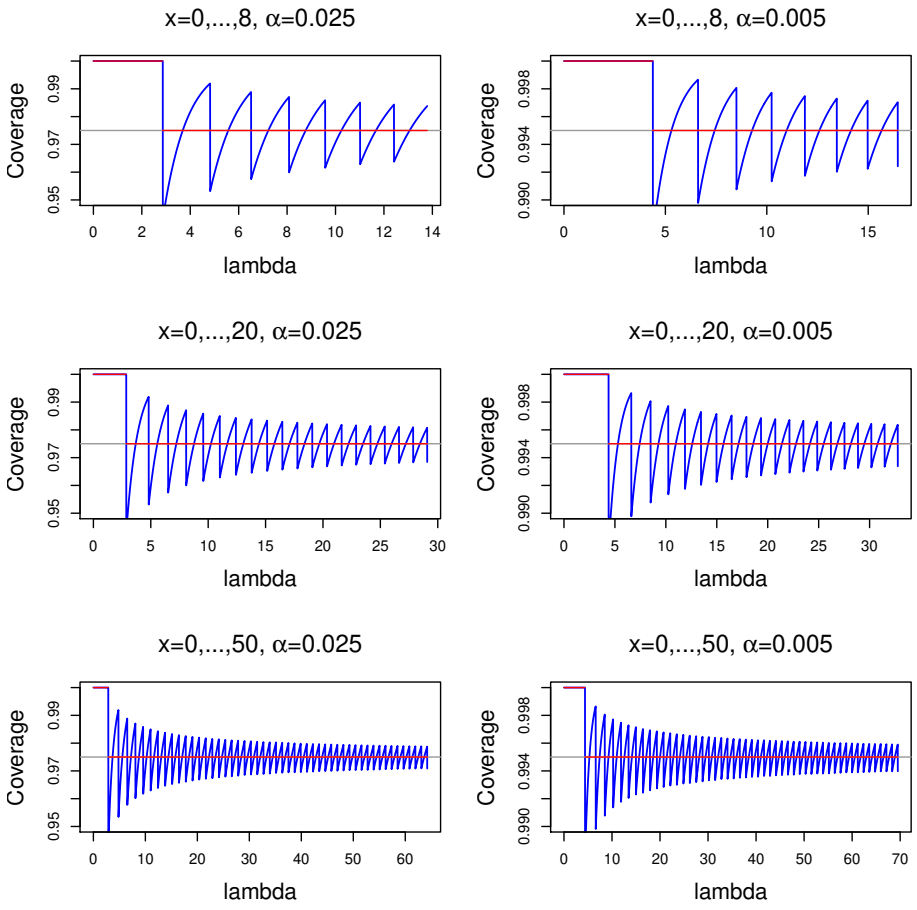


Figure 5.1: Coverage of upper-tail LCC intervals given by the modified OLC method for observed values of $0 < x < 8$, $0 < x < 20$ and $0 < x < 50$, where $N = 40, 100$ and $250$, respectively for each $x$ and the nominal confidence levels of 97.5% ($\alpha = 0.025$) and 99.5% ($\alpha = 0.005$)

To illustarate the coverage of the new estimator, examples are shown in Figures 5.1 and 5.2, which plot coverage against $\lambda$ for upper-tail intervals and lower-tail intervals, respectively, with nominal confidence levels of 97.5% and 99.5% for observed values $0 < x < 8$, $0 < x < 20$ and $0 < x < 50$. It is clear that the coverage is always evenly spread around the nominal level and appears to give sensible interval estimates. The average coverage between two consecutive spikes always equals the nominal level $(1 - \alpha)$.

Turning to a critical feature of an interval estimator, the length of its intervals, is considered next. If $(l_x, u_x)$ is an interval estimate when $X = x$,

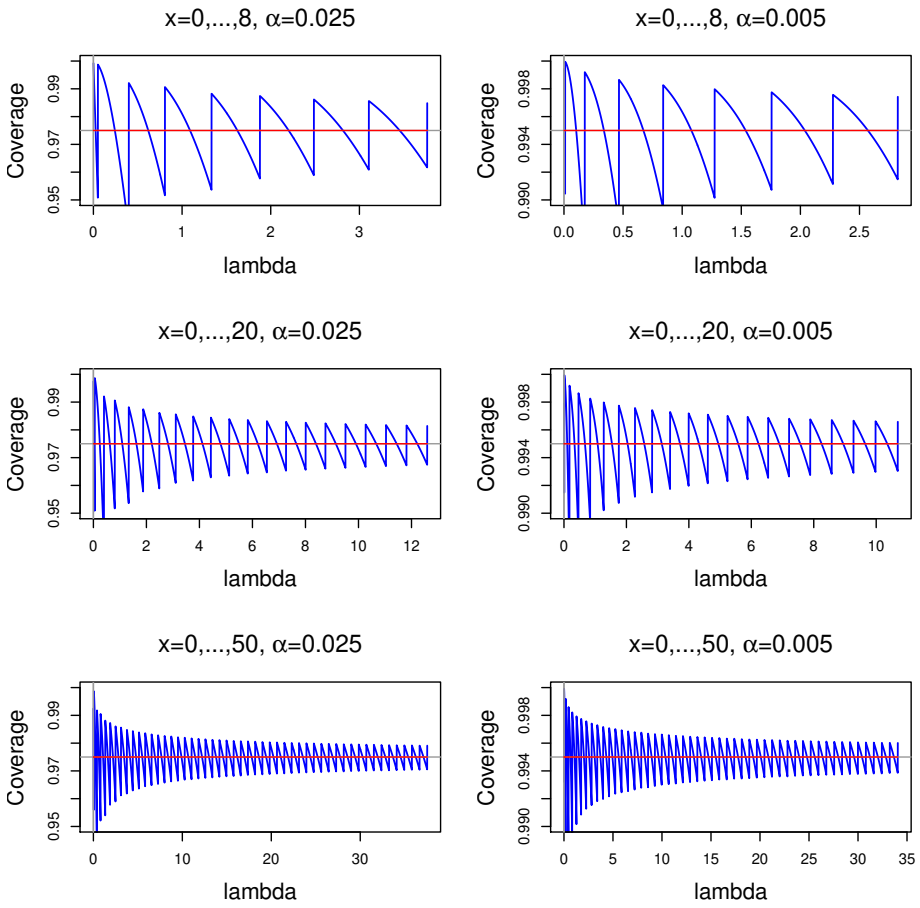


Figure 5.2: Coverage of lower-tail LCC intervals given by the modified OLC method for observed values of $0 < x < 8$, $0 < x < 20$ and $0 < x < 50$, where $N = 40, 100$ and 250, respectively for each $x$ and the nominal confidence levels of 97.5% ($\alpha = 0.025$) and 99.5% ($\alpha = 0.005$)

then the expected length of its interval, $L(\lambda)$ is given by

$$L(\lambda) = \sum_{x=0}^{\infty} (u_x - l_x) \exp^{-\lambda} \lambda^x / x! \tag{5.8}$$

and averaging over $\lambda$ gives the average expected length (AEL),

$$E(L(\lambda)) = \int_{\lambda=0}^{\infty} L(\lambda) d\lambda. \tag{5.9}$$

This definition holds for two-tail intervals and for one-tail intervals. This average expected length equals infinity for most methods of forming interval estimates, so the average expected length is of limited use. Corollary 4.1 in Chapter 4 gives conditions for a local optimality property that is suited to a sampling method for which the length of intervals increases without bound. Rather than consider the average expected length of intervals, it considers the *weighted* average expected length of intervals.

To form a set of weights in a meaningful way, we choose a function $w(\lambda)$ and put

$$p_i = \int_0^{\infty} Pr(X = i|\lambda) w(\lambda) d(\lambda) \tag{5.10}$$

for $i = 0, 1, \ldots$, where $p_i$ is the weight given to one or two-tailed intervals, $(0, u_i)$ and $(l_i, u_i)$. For mathematical tractability, $w(\lambda)$ is set equal to a gamma$(\theta, \beta)$ distribution,

$$w(\lambda) = \frac{\beta^{\theta+1}}{\Gamma(\theta+1)} \lambda^{\theta} e^{-\beta\lambda} \quad for 0 \leq \lambda < \infty. \tag{5.11}$$

As $Pr(X = i|\lambda) = e^{-\lambda}\lambda^i / i!$, for $i = 0, 1, \ldots$

$$\begin{aligned} p_i &= \int_0^{\infty} \frac{e^{-\lambda}\lambda^i}{i!} \frac{\beta^{\theta+1}}{\Gamma(\theta+1)} \lambda^{\theta} e^{-\beta\lambda} d\lambda \\ &= \frac{\Gamma(\theta+i+1)}{(\beta+1)^{\theta+i+1}} \frac{\beta^{\theta+1}}{\Gamma(\theta+1)} \int \frac{(\beta+1)^{\theta+i+1}}{\Gamma(\theta+i+1)} \lambda^{\theta+i} e^{-(\beta+1)\lambda} d\lambda \\ &= \frac{\Gamma(\theta+i+1)\beta^{\theta+1}}{(\beta+1)^{\theta+i+1}\Gamma(\theta+1)i!} \end{aligned} \tag{5.12}$$

We need to determine the weighting function by choosing the parameter values $(\theta, \beta)$ of the Gamma$(\theta, \beta)$ distribution. When comparing methods, in

the literature it is quite common to examine performance in each of a number of subintervals of the range of $\lambda$. A common choice is the intervals $(0, 2)$, $(2, 5)$ and $(5, 50)$ (Barker, 2002; Patil and Kulkarni, 2012; Lui, 2012 ). When $\theta = 1$ and $\beta = 1/4$, the weighting function $w(\lambda)$ gives almost equal weight to each of these intervals, so those are the values that will be used here. The condition for Corollary 4.1 to hold are specified in Proposition 4.2. The following steps were performed to examine whether the conditions held for the modified OLC method with $p_0, p_1, \ldots, p_N$ given by equation (5.14) and $N$ set at 200. We examined each $\alpha$ in $\{0.001(0.001)0.3\}$.

(1) Let $(0, u_i)$ be the confidence interval given by the modified OLC method when $X = i$. Determine $u_0, u_1, \ldots, u_{200}$.

(2) By definition $C_{u,i}(u_i) = 1 - \sum_{x=0}^{i-1} e^{-\lambda}\lambda^x/x!$ with $\lambda = u_i$. For $i = 0, 1, 2, \ldots, 200$, calculate $h_i = C_{u,i}(u_i) - (1-\alpha)$ and calculate $f_i = C_{u,i}(u_{i-1}) - (1 - \alpha)$ for $i = 1, 2, \ldots, 201$. Then calculate $\psi_{jl} = \prod_{i=j}^{l}(h_i/f_i)$ for $j = 1, 2, \ldots, 200$ and $l = j, j + 1, \ldots, 200$.

(3) Calculate $p_i = \frac{\Gamma(\theta+i+1)\beta^{\theta+1}}{(\beta+1)^{\theta+i+1}\Gamma(\theta+1)i!}$ for $i = 0, 1, 2, \ldots, 200$, $\theta = 1; \beta = 1/4$.

(4) Calculate $\phi_l = p_l + \sum_{j=1}^{l} \psi_{jl} p_{j-1}$ for $l = 0, 1, \ldots, 200$.

If $\phi_l$ is always positive, it means that our algorithm satisfies the conditions given in part (b) of Proposition 4.2. This was the case for each value of $\alpha$ that was considered. Thus if any of $u_0, \ldots, u_{199}$ are adjusted by a small amount to give $(u_0^*, \ldots, u_{199}^*)$, while $(u_0^*, \ldots, u_{199}^*, u_{200}^*)$ is a partition that gives locally correct confidence intervals, then the weighted average expected length of intervals is increased (i.e. $\sum p_i u_i^* > \sum p_i u_i$).

The one-tailed lower confidence intervals have an upper bound of infinity and hence are infinite. However, letting $l_i$ denote the lower limit when $X = i (i = 0, \ldots, 200)$, the weighted average $\sum p_i l_i$ was examined as follows

(1) $C_{l,i}(l_i)$ was set equal to $\sum_{x=0}^{i} e^{-\lambda}\lambda^x/x!$ with $\lambda$ set equal to $l_i$ for $i =$

$0, \ldots, 200$.

(2) Calculate $h_i = C_{l,i}(l_i) - (1-\alpha)$ for $i = 0, \ldots, 200$ and $f_i = C_{l,i}(l_{i-1}) - (1-\alpha)$ for $i = 1, 2, \ldots, 201$.

(3) As in the procedure for the upper endpoints, calculate $\psi_{jl} = \prod_{i=j}^{l}(h_i/f_i)$ for $j = 1, 2, \ldots, 200$ and $l = j, j+1, \ldots, 200$ and put $\phi_l = p_l + \sum_{j=1}^{l} \psi_{jl} p_{j-1}$ for $l = 0, 1, \ldots, 200$.

The value $\phi_0, \ldots, \phi_{200}$ was determined for equal $\alpha$ in $\{0.001(0.001)0.3\}$ and it was found that they were always negative. Hence adjusting any of $l_0, \ldots, l_{199}$ by a small amount to yield a partition $(l_0^*, \ldots, l_{199}^*, l_{200}^*)$ that gives locally correct confidence intervals will result in $\sum p_i l_i^* < \sum p_i l_i$. It follows that $\sum p_i(u_i - l_i) > \sum p_i(u_i^* - l_i^*)$ so the modified OLC method also gives two-tailed interval whose average expected length is a local minimum.

Some desirable properties in interval estimators were reviewed in Section 2.2.4. In Section 4.5, it was noted that an OLC method will always have Property 1 (a confidence region should be an interval and not a collection of disjoint intervals) and Property 2 (montonicity in $x$ ) from the way that intervals are constructed. This is also true for the modified OLC method. Whether the modified OLC method has Property 3 (the nesting property) is not self-evident and may depend on the sampling method. The property states that, if two confidence intervals have different confidence levels then, for any given $x$, the interval for the higher confidence level should contain the interval with the lower confidence level. Repetitive computation has shown that the modified OLC method has this when $1 \leq x \leq 200$ and the two confidence levels are in $\{0.001(0.001)0.3\}$. It seems clear that the modified OLC method meets the previous requirements for being a well-behaved interval estimator.

## 5.4 Comparison with other methods

In this section, we compare the OLC method's performance with the following five methods of forming interval estimates, which have been described in Chapter 2: Garwood, mid-$p$, Jeffreys, Wald and score methods. We compare them in terms of their coverage probability for $0 < \lambda < 20$ and $0 < \lambda < 50$ and expected length for $0 < \lambda < 2$, for $2 < \lambda < 5$ and $5 < \lambda < 50$ and nominal confidence level 95%, 97.5% and 99.5%. The range of $\lambda$ and these nominal levels give a fair representation of the behaviour for other ranges of $\lambda$ and other nominal levels.

### 5.4.1 Coverage probability

Attention will be restricted to upper-tail and lower-tail intervals. As mentioned in Section 3, $C_{l,i}(\lambda)$ and $C_{u,i}(\lambda)$ are the coverage of the lower-tail and an upper-tail interval estimator, respectively, and are the probability that the random interval $(l_x, \infty)$ and $(0, u_x)$ contain $\lambda$. We follow a similar structure to the binomial case for calculating the average coverage of the upper-tail. But there is a difference in that we do not calculate the average coverage for the whole range of $\lambda$. We divide $\lambda$'s range as $0 < \lambda < 20$ and $20 < \lambda < 50$ and calculate the average coverage for each range sepeartely. First, for $0 < \lambda < 20$, the quantity $T_u$ is defined as

$$T_u = \frac{1}{20 - u_0} \int_{u_0}^{20} C_{u,i}(\lambda) d\lambda \tag{5.13}$$

and refer to it as the *truncated average coverage*. In calculating $T_u$ for $0 < \lambda < 20$, the values of $\lambda$ in the range $(0, u_0)$ are excluded as the coverage equals 1 when $\lambda < u_0$. It means that the coverage for $\lambda \leq u_0$ is very different form the coverage for $\lambda > u_0$. Average coverage of an upper-tail interval over the full range $(0, 20)$ equals $\{(20 - u_0)T_u + u_0\}/20$. Second, for $20 < \lambda < 50$, the

quantity $T_u$ is defined as

$$T_u = \frac{1}{50-20}\int_{20}^{50} C_{u,i}(\lambda)d\lambda \tag{5.14}$$

A difference between the first and the second cases is that $T_u$ is calculated for the whole range of $20 < \lambda < 50$ without excluding any value of $\lambda$. That is because that the coverage never equals 1. Average coverage of an upper-tail interval over the range of $20 < \lambda < 50$ equals $T_u$.

In Figures 5.3 and 5.4, the coverage of 95% upper-tail intervals of all methods are plotted against $\lambda$ for $0 < \lambda < 20$ and $20 < \lambda < 50$, respectively. It is clear, in Figure 5.3, that the coverage of the first interval $u_0$ equals 1, which will be excluded in calculating averages unlike the coverage of the intervals in Figure 5.4 which never equals 1. For all combination of $\alpha$ and $\lambda$, the OLC, Garwood and mid-$p$ methods are giving LCC intervals, i.e. the average coverage between consecutive spikes is at least $1-\alpha$. In contrast, the Jeffreys, Wald and score methods do not give LCC intervals, as the average coverage between consecutive spikes is sometimes below $1-\alpha$. In particular, for the Wald method, the average coverage is below the nominal level for almost all values of $\lambda$.

To clarify the difference between our OLC method and other methods, values of both $T_u$ and $u_0$ ae given in Tables 5.3 and 5.4. They are given for $\alpha = 0.05, 0.025$ and 0.005 for both $0 < \lambda < 20$ and $20 < \lambda < 50$. The results show that the OLC method is better than any of the other methods. From Tables 5.3 and 5.4, the OLC method has a truncated average coverage that equals the nominal confidence level of $1-\alpha$ for all cases of $\alpha$ and $\lambda$. In contrast, the Garwood and score methods are very conservative and the mid-$p$ is slightly conservative, while, the Jeffreys and Wald methods are consistently liberal, i.e. that truncated average coverage is below the nominal level of $1-\alpha$. A small value of $u_0$ is desirable and in that respect the Wald method does exceptionally

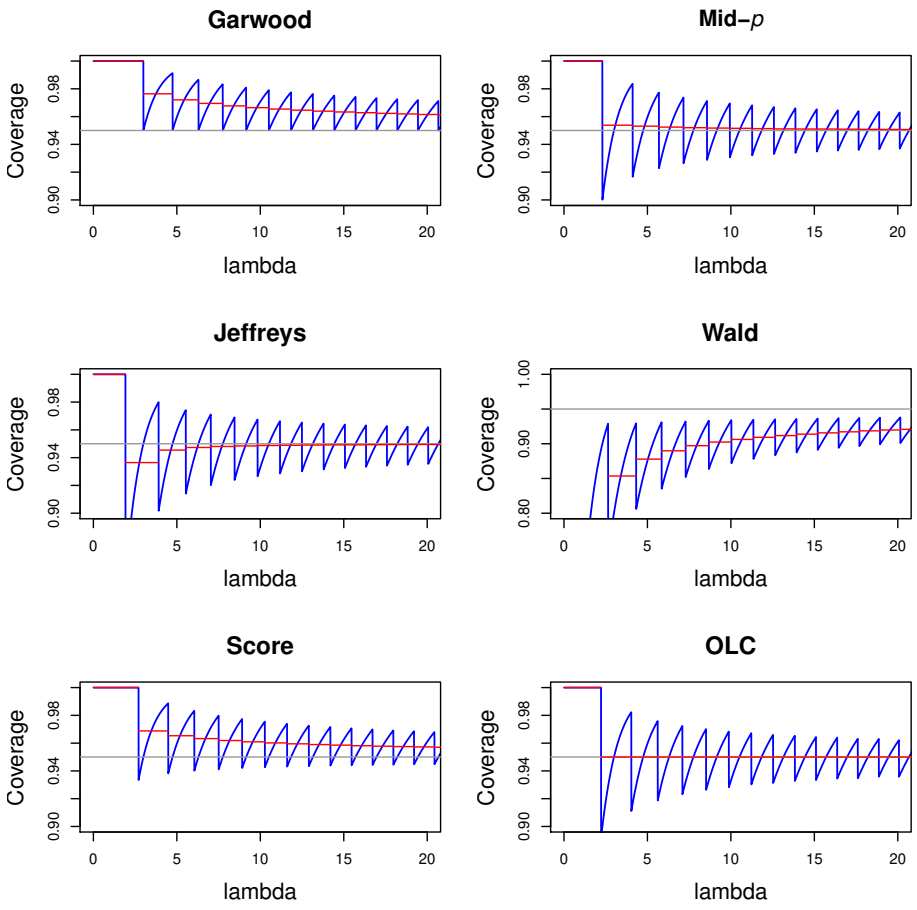


Figure 5.3: Coverage of upper-tail 95% interval estimate for Garwood, mid-$p$, Jeffreys, Wald, score and OLC methods plotted against $\lambda$, for $0 < \lambda < 20$.

well, with $u_0$ always equal to 0. However, the Wald method is not the preferred method, as its coverage is too far below the nominal level. Based on the value of $u_0$, the OLC method is a little poorer than Jeffreys, but a little better than mid-$p$ and much better than Garwood and score.

Table 5.3: Average coverage (Av.Cov) of upper tail $1-\alpha$ intervals and smallest upper limit ($u_0$) of six methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and $0 < \lambda < 20$

| $\alpha$ | statistic | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|---|
| 0.05 | $u_0$ | 2.9957 | 2.3026 | 2.7055 | 0.000 | 1.9207 | 2.2241 |
| 0.05 | Av.Cov | 0.9664 | 0.9517 | 0.9609 | 0.8673 | 0.9471 | 0.9500 |
| 0.025 | $u_0$ | 3.6889 | 2.9957 | 3.8415 | 0.000 | 2.5119 | 2.8653 |
| 0.025 | Av.Cov | 0.9841 | 0.9762 | 0.9846 | 0.8943 | 0.973 | 0.9750 |
| 0.005 | $u_0$ | 5.2983 | 4.6052 | 6.6349 | 0.000 | 3.9397 | 4.3838 |
| 0.005 | Av.Cov | 0.9972 | 0.9954 | 0.9985 | 0.9245 | 0.9943 | 0.9950 |

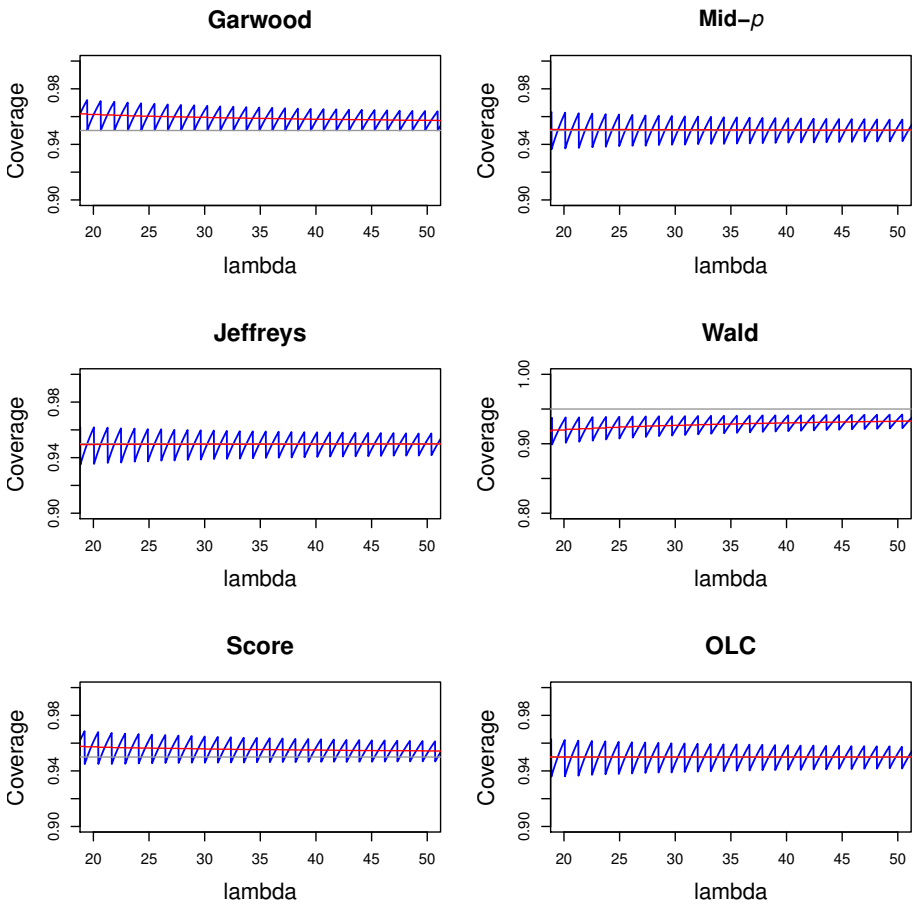


Figure 5.4: Coverage of upper-tail 95% interval estimate for Garwood, mid-$p$, Jeffreys, Wald, score and OLC methods plotted against $\lambda$, for $20 < \lambda < 50$.

Table 5.4: Average coverage (Av.Cov) of upper tail $1 - \alpha$ intervals of six methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and $20 < \lambda < 50$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.959 | 0.9504 | 0.9555 | 0.9278 | 0.9497 | 0.950 |
| 0.025 | 0.9801 | 0.9754 | 0.9801 | 0.9565 | 0.9748 | 0.975 |
| 0.005 | 0.9962 | 0.9951 | 0.9973 | 0.9851 | 0.9949 | 0.995 |

Regarding the lower-tail interval, in contrast to the upper-tail interval, we do not exclude any values of $\lambda$ in the range $(0, l_0)$ as the coverage in this range never equals 1. So, as defined in Section 5.2, $C_{l,i}(\lambda)$ is the coverage of lower-tail interval estimator and is the probability that random interval $(l_x, \infty)$ contains $\lambda$. For $0 < \lambda < 20$, we define the quantity $Av.Cov_l$ as

$$Av.Cov_L = \frac{1}{20}\int_0^{20} C_{l,i}(\lambda)d\lambda \tag{5.15}$$

and refer to it as the average coverage of the lower-tail. For $20 < \lambda < 50$, we define the quantity $Av.Cov_l$ as

$$Av.Cov_L = \frac{1}{50-20}\int_{20}^{50} C_{l,i}(\lambda)d\lambda \tag{5.16}$$

Calculating the average coverage of the lower-tail for the ranges $0 < \lambda < 20$ and $20 < \lambda < 50$ does not exclude any values of $\lambda$ because the coverage never equals 1. Tables 5.5 and 5.6 give the values of $Av.Cov_L$ for all methods described earlier at $\alpha = 0.05, 0.025$ and $0.005$ for both $0 < \lambda < 20$ and $20 < \lambda < 50$. The results show that the OLC method is better than other methods, with an average coverage that always equals the nominal level $(1-\alpha)$. In contrast, the Garwood and Wald methods are very conservative, Mid-$p$ is a bit conservative, while the score and Jeffreys methods are consistently liberal (i.e. their average coverages are below the nominal level $(1-\alpha)$). This can be seen in both Figures 5.5 and 5.6, where the coverages of 95% lower-tail intervals of all methods are plotted against $\lambda$ for $0 < \lambda < 20$ and $20 < \lambda < 50$, respectively.

Table 5.5: Average coverage (Av.Cov) of lower tail $1-\alpha$ intervals of six methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and $0 < \lambda < 20$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.9636 | 0.9520 | 0.9401 | 0.9787 | 0.9490 | 0.950 |
| 0.025 | 0.9823 | 0.9762 | 0.9651 | 0.993 | 0.9743 | 0.975 |
| 0.005 | 0.9966 | 0.9953 | 0.9891 | 0.9996 | 0.9948 | 0.995 |

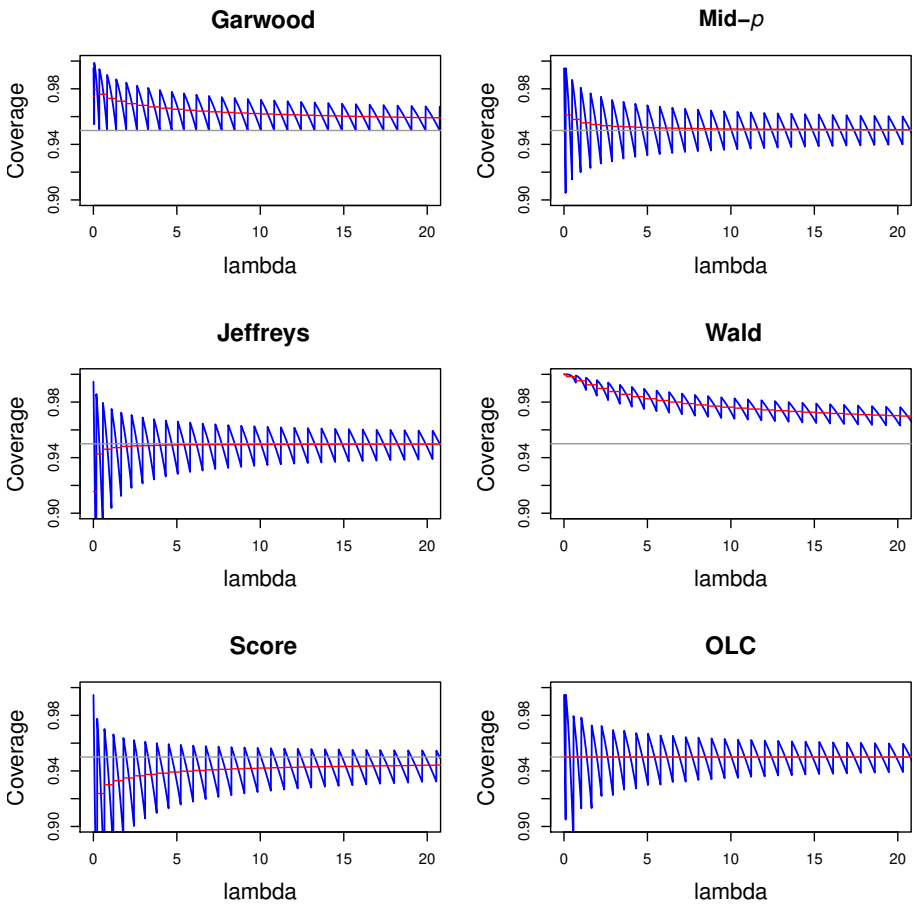


Figure 5.5: Coverage of lower-tail 95% interval estimate for Garwood, mid-$p$, Jeffreys, Wald, score and OLC methods plotted against $\lambda$, for $0 < \lambda < 20$.

Table 5.6: Average coverage (Av.Cov) of lower tail $1 - \alpha$ intervals of six methods of forming interval estimates, for $\alpha = 0.05, 0.025, 0.005$ and $20 < \lambda < 50$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.9575 | 0.9503 | 0.9453 | 0.9663 | 0.9498 | 0.9499 |
| 0.025 | 0.9791 | 0.9752 | 0.9705 | 0.9865 | 0.9749 | 0.9750 |
| 0.005 | 0.9960 | 0.9951 | 0.9926 | 0.9986 | 0.9950 | 0.9950 |

In the examples given so far, the mid-$p$ method gave LCC intervals. Direct computation showed it gives LCC intervals for $x < 200$ and $\alpha \in \{0.001(0.001)0.1\}$, giving the following result.

**Proposition 5.2.**

For $x \leq 200$ and $\alpha \in \{0.001(0.001)0.1\}$, the mid-$p$ method gives LCC intervals.

In Figures 5.3 and 5.4, the coverage of the 95% upper-tail for the mid-$p$ method

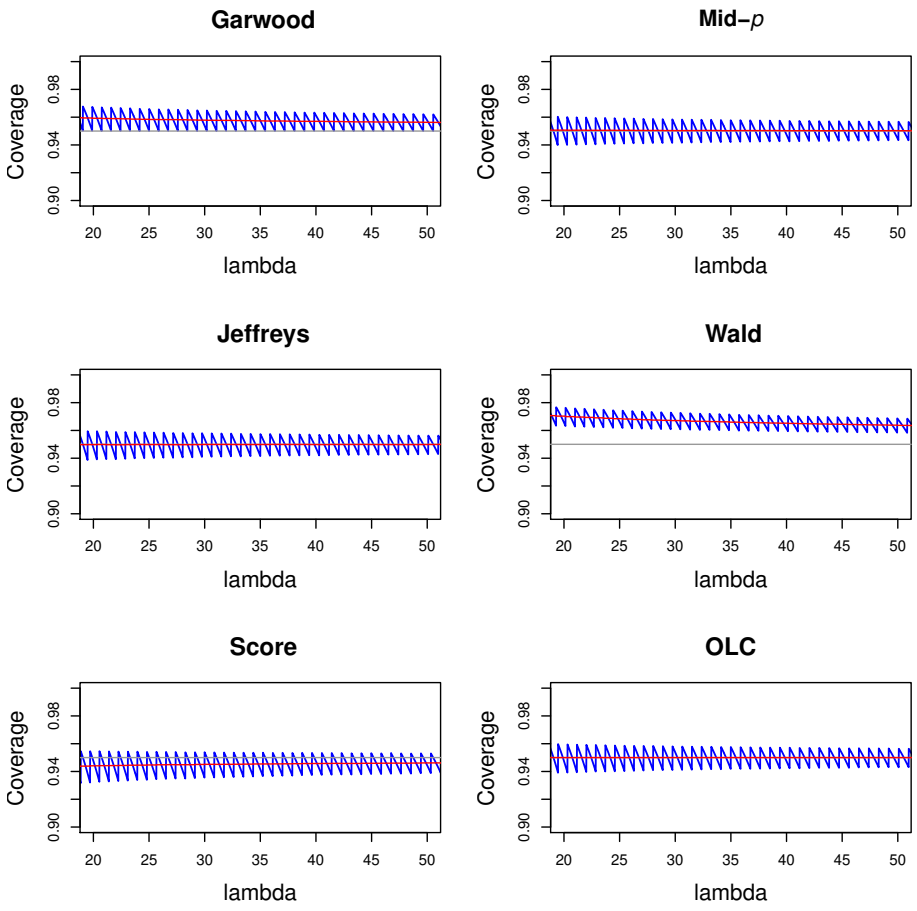


Figure 5.6: Coverage of lower-tail 95% interval estimate for Garwood, mid-$p$, Jeffreys, Wald, score and OLC methods plotted against $\lambda$, for $20 < \lambda < 50$.

is plotted against $\lambda$, in the first graph of the right-hand side, for $0 < \lambda < 20$ and $20 < \lambda < 50$, respectively. For this method, the spikes in both plots are spaced fairly regularly and its actual coverage always crosses the nominal coverage level between consecutive spikes. Moreover, the mid-$p$ method in the Poisson case has the same property as with the binomial distribution. It gives one-tailed confidence intervals whose coverage is as close to the targeted nominal level as the coverage of any method, for any value of $\lambda$. This property is given in the following proposition.

**Proposition 5.3.**

The mid-$p$ method has the smallest root-mean-sqaure error in the coverage probability, $|C_u(\lambda) - (1-\alpha)|$, among these discribed methods for any value of $\lambda$. By considering all sensible methods of forming equal-tailed confidence interval that were discussed in Section 2.3, we found that the mid-$p$ has the smallest root-mean-sqaure error, $|C_u(\lambda) - (1-\alpha)|$ or $|C_l(\lambda) - (1-\alpha)|$, among any method of forming one-tailed confidence interval.

The proof of Proposition 5.3 is given in Appendix B. As mentioned earlier, the average absoluate error in coverage or the root-mean-square error in coverage over the range of $\lambda$ is commonly used as a good measure to compare different methods. It examines how the coverage probability of the interval estimator typically varies from the nominal confidence level. So, we calculated the RMSE of each method's coverage for each range of $\lambda$.

For $u_0 < \lambda < 20$, the RMSE is defined as

$$RMSE = \left[\frac{1}{20-u_0}\int_{u_0}^{20}\{C_{u,i}(\lambda) - (1-\alpha)\}^2 d\lambda\right]^{1/2} \tag{5.17}$$

where $(1-\alpha)$ is the nominal confidence level and for $20 < \lambda < 50$, the RMSE is given by

$$RMSE = \left[\frac{1}{50-20}\int_{20}^{50}\{C_{u,i}(\lambda) - (1-\alpha)\}^2 d\lambda\right]^{1/2}. \tag{5.18}$$

The RMSE of the upper-tail for each method is given in Tables 5.7 and 5.8 at $\alpha = 0.05, 0.025$ and $0.005$ for both $0 < \lambda < 20$ and $20 < \lambda < 50$, respectively. The results of these tables emphasise the result in Proposition 5.3 as the mid-$p$ method has the smallest RMSE of any other method in every row. The OLC method has the second smallest RMSE that is only a little bigger than mid-$p$ method, at most 12%. After the OLC method, Jeffreys method almost always has the next smallest RMSE, but its RMSE can be more than 50% bigger than mid-$p$. The remaining methods have RMSEs that are at least 79% bigger for some values of $\alpha$ and $\lambda$. The RMSE of the Wald method is always very poor. Hence, although that OLC method is not the optimal method, it has a very respectable RMSE.

Table 5.7: Root mean-square error (RMSE) of coverage of upper tail $(1-\alpha)$ intervals for six methods of forming interval estimates, $\alpha = 0.05$, 0.025, 0.005 and $0 < \lambda < 20$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.0191 | 0.0133 | 0.0153 | 0.1533 | 0.0173 | 0.0138 |
| 0.025 | 0.0104 | 0.0074 | 0.0109 | 0.1575 | 0.0103 | 0.0078 |
| 0.005 | 0.0024 | 0.0018 | 0.0036 | 0.1592 | 0.0028 | 0.0020 |

Table 5.8: Root mean-square error (RMSE) of coverage of upper tail $(1-\alpha)$ intervals for six methods of forming interval estimates, $\alpha = 0.05$, 0.025, 0.005 and $20 < \lambda < 50$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.01036 | 0.00578 | 0.00772 | 0.02378 | 0.00584 | 0.0058 |
| 0.025 | 0.00584 | 0.00336 | 0.00587 | 0.01952 | 0.0034 | 0.00338 |
| 0.005 | 0.00142 | 0.00086 | 0.00235 | 0.01029 | 0.00088 | 0.00087 |

The RMSE of each method's coverage for lower-tail intervals is calculated over the full ranges of $\lambda$. Thus, for $0 < \lambda < 20$, RMSE is given by

$$RMSE = \left[\frac{1}{20}\int_0^{20} \{C_{l,i}(\lambda) - (1-\alpha)\}^2 d\lambda\right]^{1/2} \tag{5.19}$$

and for $20 < \lambda < 50$, RMSE is given by

$$RMSE = \left[\frac{1}{50-20}\int_{20}^{50} \{C_{l,i}(\lambda) - (1-\alpha)\}^2 d\lambda\right]^{1/2}. \tag{5.20}$$

Tables 5.9 and 5.10 show the RMES of the lower-tail for each method, which have the same features as the RMSE of the upper tail for all methods. The mid-$p$ always has the smallest RMSE, followed by the OLC method with an RMSE that is a little bigger than the mid-$p$. After that, the Jeffreys and Garwood methods have the next smallest RMSE, but their RMSE is typically much bigger than that of the mid-$p$. The method with the worst RMSE is Wald.

Table 5.9: Root mean-square error (RMSE) of coverage of lower tail $(1-\alpha)$ intervals for six methods of forming interval estimates, $\alpha = 0.05, 0.025, 0.005$ and $0 < \lambda < 20$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.0162 | 0.0105 | 0.0181 | 0.030 | 0.012 | 0.011 |
| 0.025 | 0.0086 | 0.0057 | 0.0151 | 0.0184 | 0.0067 | 0.006 |
| 0.005 | 0.0019 | 0.0013 | 0.0088 | 0.0046 | 0.0016 | 0.0014 |

Table 5.10: Root mean-square error (RMSE) of coverage of lower tail $(1-\alpha)$ intervals for six methods of forming interval estimates, $\alpha = 0.05, 0.025, 0.005$ and $20 < \lambda < 50$

| $\alpha$ | Garwood | Mid-p | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|
| 0.05 | 0.0086 | 0.00474 | 0.00698 | 0.01675 | 0.00478 | 0.00476 |
| 0.025 | 0.00469 | 0.00263 | 0.0055 | 0.01162 | 0.00266 | 0.00264 |
| 0.005 | 0.0011 | 0.00063 | 0.00259 | 0.00361 | 0.00064 | 0.00064 |

Turning to the length of intervals, the length of the two-tailed interval is examined for all mentioned methods. In Figure 5.7, the expected length of 95% two-tailed intervals for OLC method and other methods are plotted against $\lambda$ for $0 < \lambda < 2$, $2 < \lambda < 5$ and $5 < \lambda < 50$. OLC, Jeffreys, score and Wald methods are plotted on the left-hand panels, while OLC, mid-$p$ and Garwood methods are plotted on the right-hand panels. For all values of $\lambda$, the expected length of the OLC, mid-$p$, score and Jeffreys intervals are all very similar and a little smaller than the expected length of the Garwood interval. Wald intervals have a much smaller expected length than the other methods,

but that is because its coverage is well below the nominal confidence level.

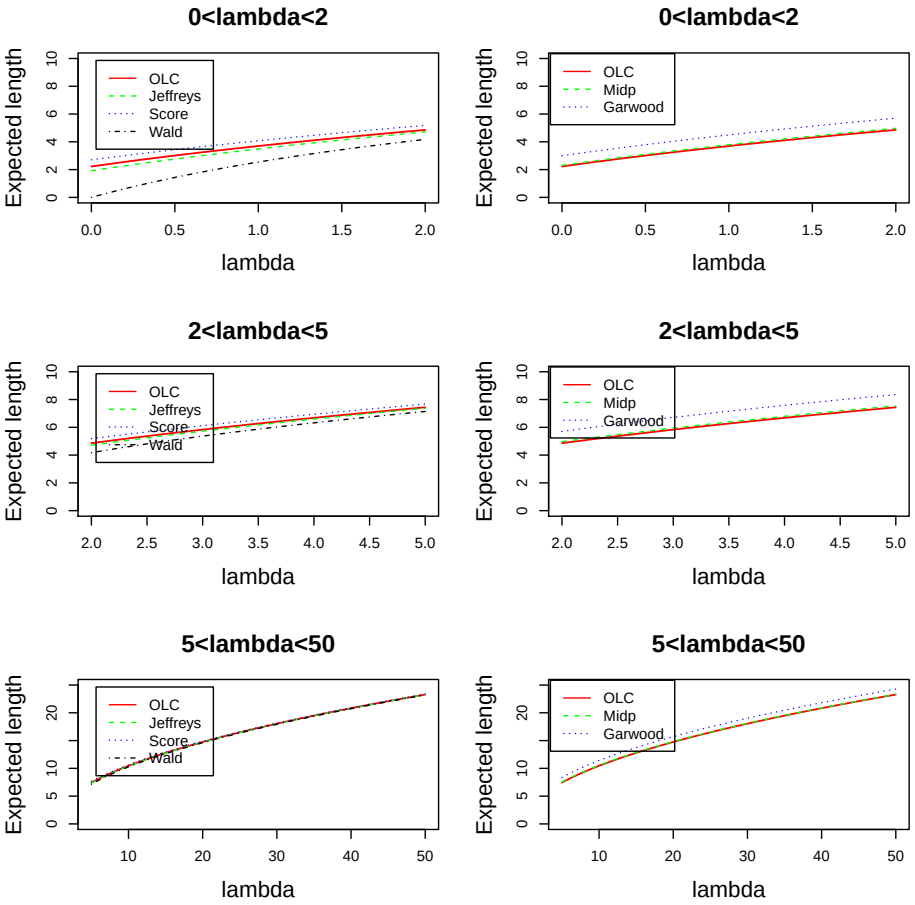


Figure 5.7: Expected length of two-sided 95% interval estimates for the OLC, Jeffreys, score and Wald methods (left-hand panels) and the OLC, Garwood and Mid-$p$ methods (right-hand panels) plotted against $\lambda$, for $0 < \lambda < 2$, $2 < \lambda < 5$ and $5 < \lambda < 50$

In comparing the length of intervals, three subintervals of the range of $\lambda$ were examined seperately: $(0, 2), (2, 5)$ and $(5, 50)$. There is clearly benefit in also having a single measure that reflects performance over the full range of $\lambda$. As noted in Section 5.3, this can be achieved by using a weight function and forming a weighted average expected length of intervals. The weight function, $w(\lambda)$ was chosen as a gamma distribution,

$$w(\lambda) = \frac{\beta^{\theta+1}}{\Gamma(\theta+1)}\lambda^{\theta}e^{-\beta\lambda}. \tag{5.21}$$

The values of $\theta$ and $\beta$ were chosen to make the probability under the gamma

curve approximately equal for the intervals $0 < \lambda < 2$, $2 < \lambda < 5$ and $5 < \lambda < 50$. This gave $\theta = 1$ and $\beta = \frac{1}{4}$. So, $w(\lambda) = \frac{1}{4}e^{-\lambda/4}$. The weighted average expected length (WAEL) of a method's intervals is given by

$$\text{WAEL}(l(\lambda)) = \frac{\int_{\lambda=0}^{\infty} \sum_{i=0}^{200} p_i l(\lambda) d\lambda}{\int_{\lambda=0}^{\infty} \frac{1}{4} e^{-\lambda/4}}, \tag{5.22}$$

where $l(\lambda) = (u_i - l_i)$ is the length of an interval.

Table 5.11: Average expected length (AEL) weighted average expected length (WAEL) of two-tail 1-2$\alpha$ intervals for six methods of forming interval estimates, $\alpha = 0.05$

| measure | $\lambda$ | Garwood | Midp | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|---|
| AEL | $0 < \lambda < 2$ | 4.449 | 3.741 | 4.034 | 2.395 | 3.432 | 3.646 |
| AEL | $2 < \lambda < 5$ | 7.117 | 6.341 | 6.504 | 5.799 | 6.129 | 6.231 |
| AEL | $5 < \lambda < 50$ | 17.688 | 16.778 | 16.837 | 16.594 | 16.69 | 16.721 |
| WAEL | $0 < \lambda < 200$ | 7.02 | 6.251 | 6.45 | 5.439 | 6.018 | 6.153 |

Table 5.12: Average expected length (AEL) weighted average expected length (WAEL) of two-tail 1-2$\alpha$ intervals for six methods of forming interval estimates, $\alpha = 0.025$

| measure | $\lambda$ | Garwood | Midp | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|---|
| AEL | $0 < \lambda < 2$ | 5.31 | 4.616 | 5.279 | 2.854 | 4.244 | 4.487 |
| AEL | $2 < \lambda < 5$ | 8.379 | 7.628 | 8.048 | 6.909 | 7.383 | 7.5 |
| AEL | $5 < \lambda < 50$ | 20.913 | 20.019 | 20.182 | 19.773 | 19.915 | 19.952 |
| WAEL | $0 < \lambda < 200$ | 8.297 | 7.548 | 8.027 | 6.481 | 7.272 | 7.427 |

Table 5.13: Average expected length (AEL) and weighted average expected length (WAEL) of two-tail 1-2$\alpha$ intervals for six methods of forming interval estimates, $\alpha = 0.005$

| measur | $\lambda$ | Garwood | Midp | Score | Wald | Jeff | OLC |
|---|---|---|---|---|---|---|---|
| AEL | $0 < \lambda < 2$ | 7.204 | 6.522 | 8.227 | 3.751 | 6.036 | 6.332 |
| AEL | $2 < \lambda < 5$ | 10.973 | 10.261 | 11.465 | 9.081 | 9.956 | 10.1 |
| AEL | $5 < \lambda < 50$ | 27.26 | 26.395 | 26.903 | 25.987 | 26.261 | 26.309 |
| WAEL | $0 < \lambda < 200$ | 10.941 | 10.221 | 11.525 | 8.518 | 9.867 | 10.057 |

Table 5.11 gives the AEL for the different ranges of $\lambda$ $[(0, 2), (2, 5)$ and $(5, 50)]$ and the WAEL for $0 < \lambda < 200$, for $\alpha = 0.05$. Tables 5.12 and 5.13 gives the equivalent information for $\alpha = 0.025$ and $\alpha = 0.005$, respectively. The Wald interval has an actual coverage that is well below the nominal coverage so that its intervals have the smallest AEL for each combination. The

Garwood interval has the longest AEL of all methods for all combinations because of its conservative coverage (except, when $\alpha$=0.005, the score method has the longest AEL for all values of $\lambda$). The score interval coverage suffers from some conservatism which underlies its AEL being the longest after the Garwood interval. Apart from these three methods (Wald, Garwood and score methods) the AEL of the OLC method is usually similar in size to the remaining methods and is always shorter than the mid-$p$ method. The weighted average expected length (WAEL) was calculated for $0 < \lambda < 200$, but it was also calculated also for $\lambda > 200$. It was found that the value of the WAEL for all methods did not change appreciably as $\lambda$ was increased, and the relative performance of the methods was unchanged. So, the range of $0 < \lambda < 200$ seems satisfactory. Comparing the WAEL with the AEL in the ranges 0-2, 2-5 and 5-50, the WAEL put the methods in the same order as AEL in every case. Thus there is good agreement between the measures so the WAEL could be used as an overall representative measure.

## 5.5 Concluding comments

The purpose of this chapter was to see whether the OLC method gave good confidence intervals for the Poisson distribution. So, our work was extended to show that the OLC method has broader application than just the binomial distribution. It was shown that the OLC method gave end-points that are fairly evenly spaced with coverages that are balanced around the nominal confidence level. Also, the OLC method had other properties that are desirable in an interval estimator, such as interval valued, monotonicity and nesting. This was examined for many combinition of $x$ and $\alpha$: $1 \leq x \leq 200$ and confidence levels is $\{0.001(0.001)0.3\}$. Regarding the expected length of intervals, for all examined values of $x$ and $\alpha$ the OLC method had the shortest

average expected length compared to all other methods of constructing equal-tail confidence intervals except for the Wald and Jeffreys methods. However, the Wald method gave intervals with coverage that was often less than the nominal confidence level. Even with the main challenge of the Poisson distribution, which is the infinite range of $\lambda$ and $x$, the estimation works well. As the OLC method had the shortest weighted average expected length for large values of $\lambda$, say $0 < \lambda < 200$, except in some cases, Jeffreys does better.

Turning to comparable methods of constructing confidence interval, there are similarites in the results for the binomial and Poisson distributions. The Garwood and Clopper-Pearson methods are the gold standard methods and give similar results: good coverage but poor length. The mid-$p$ and Jeffreys methods also performed similarly for the two distributions. The mid-$p$ method again had a coverage that is a little conservative and had the smallest RMSE among other methods, while Jeffreys method had a coverage that is sometimes far below the nominal level. The performance of OLC method had similar characteristics for both distributions but the score method performed much better for the binomial distribution than for the Poisson distribution.

# Chapter 6

# Negative binomial confidence interval methods

## 6.1 Introduction

The third sampling distribution examined in this thesis is the negative binomial distribution. Relative to both the binomial and Poisson parameters, there are comparatively few methods of forming confidence intervals for the negative binomial parameters. The most commonly used methods have been described in the negative binomial section in Chapter 2. The methods that give equal-tail intervals do not meet the strict definition of a confidence interval, except for the exact method. That is because the negative binomial sampling model has a discrete distribution, posing the same problems that arise with the binomial and Poisson distributions. The discrete nature of its sample space produces spikes, which leads to wide fluctuations in the coverage probability as the parameter of interest varies. So, the new definition of an interval estimator, based on average coverage between spikes, is applied to the negative binomial distribution. Methods that meet this definition give *locally correct confidence (LCC) intervals*. In this chapter, a method that gives LCC intervals for negative binomial parameters is developed, examined and compared with other methods.

The negative binomial distribution is similar to the binomial distribution as both of them have two parameters. The negative binomial distirbution has $(p, r)$, where $0 < p < 1$ and $r = 1, 2, \ldots$ and the binomial distribution has $(p, n)$. Usually $r$ is known and that is the case considered here. The negative binomial distribution faces a challenge similar to that faced by the Poisson distribution, as for both of them its values of $x$ are not limited, but go to infinity. However, unlike both the binomial and Poisson distributions, for the negative binomial distribution we are interested in forming confidence intervals for two different (but related) quantities: the negative binomial proportion $p$ and the

negative binomial mean $\mu$. Intervals that meet the definition of LCC intervals are constructed for both quantities. For the negative binomial proportion $p$, a method giving intervals is "optimal" if it gives an interval of minimum average width when the expected width of the interval is averaged over $p \in [0, 1]$ and for the negative binomial mean, the expected width must be mininmised when averaged over $\mu \in (0, M)$, where $M$ is sufficiently large.

In this chapter, we will apply our new approach to calculate confidence intervals for both the negative binomial proportion $p$ and mean $\mu$. So, in Section 6.2, we give a precise definition of a local confidence interval for the negative binomial proportion $p$ and modify the interval estimator for the binomial and Poisson distributions to obtain an optimal locally correct (OLC) method for the negative binomial success parameter $p$. In Section 6.3, we give the same definition of a local confidence interval for the negative binomial mean $\mu$ and modify the optimal locally correct (OLC) method for the negative binomial mean $\mu$. We examine whether the OLC method gives intervals of both $p$ and $\mu$ that seem sensible and examine whether the new estimator has properties that have been proposed in the literature as being desirable. Also, we show that the OLC method has a locally optimum property. In Section 6.4, we compare the OLC method with several methods that have been recommended for forming equal-tailed confidence intervals (which are mentioned in the literature chapter). Concluding comments are given in Section 6.5.

## 6.2 The negative binomial proportion $(p)$

Firstly, we begin with the negative binomial proportion $p$. An estimate of $p$ is monotone decreasing in $x$ unlike the situation in previous chapters, where the estimate of the parameter of interest was monotone increasing in $x$.

### 6.2.1 Locally correct confidence intervals

Let $X$ denote the number of failures before $r$ successes are observed. The success probability is $p$. Put $y = r + x$, so $y$ is the number of trials and suppose an interval estimator gives an upper-tail estimate $(0, u_x)$ for $p$. We assume that

$$0 < \ldots < u_2 < u_1 < u_0 \leq 1 \tag{6.1}$$

holds for a sensible estimator. The coverage probability of the interval estimator depends on the value of $p$, as it is the probability that the true value of the parameter $p$ is included in the interval $(0, u_x)$. When $u_{i+1} < p < u_i$, the coverage probability is given by

$$C_u(p) = Pr(X \leq i|r, p) = \sum_{x=0}^{i} \binom{x+r-1}{r-1} p^r (1-p)^x \tag{6.2}$$

The points of the spikes occur where $p$ equals $u_0, u_1, \ldots, u_\infty$. At the point $u_i$, the coverage drops by $\binom{i+r-1}{r-1} u_i^r (1-u_i)^i$ at $p = u_i$. If the average coverage between every two consecutive spikes exceeds or equals the nominal level, then the estimator gives locally correct confidence (LCC) intervals.

*Definition 1.* For the upper-tail interval, suppose an interval estimator gives $(0, u_x)$ as its upper-tail interval for $p$ when $X = x$ and that $u_0, u_1, u_2, \ldots$ satisfy equation (6.1). If for i=1,2, ....

$$\frac{1}{u_i - u_{i+1}} \int_{p=u_{i+1}}^{u_i} C_u(p) dp \geq 1 - \alpha \tag{6.3}$$

then the interval estimator gives *upper-tail LCC intervals* with confidence level $(1 - \alpha)$.

*Definition 2.* For the lower-tail intervals, suppose an interval estimator gives $(l_x, 1)$ as its lower-tail interval for $p$ when $X = x$, and that $0 < l_N < \ldots < l_2 < l_1 < l_0 < 1$ where $N$ is large. (Because of the infinite nature of the number of failures, we select a sufficiently large value for $N$). Define the coverage probability, $C_l(p)$, by

$$C_l(p) = Pr(X \geq i+1|r,p) = \sum_{x=i+1}^{\infty} \binom{x+r-1}{r-1} p^r (1-p)^x \tag{6.4}$$

for $l_{i+1} < p < l_i$. If, for $i = 1, 2, \ldots,$

$$\frac{1}{l_i - l_{i+1}} \int_{p=l_{i+1}}^{l_i} C_l(p)dp \geq 1 - \alpha \tag{6.5}$$

then the interval estimator gives *lower-tail LCC intervals* with confidence level $(1 - \alpha)$.

*Definition 3.* Suppose that, for $x = 0, 1, 2, 3, \ldots$, an interval estimator gives $(l_x, u_x)$ as its two-sided equal-tail intervals for $p$ when $X = x$. Then it gives equal-tail LCC intervals with confidence level $(1 - 2\alpha)$ for $x = 0, 1, 2, 3, \ldots$ if the intervals $(l_x, 1)$, and $(0, u_x)$ are sets of one-sided lower-tail and upper-tail LCC intervals, respectively, each with confidence level $(1 - \alpha)$. The interval estimator that gives equal-tail LCC intervals will be referred to as a *LCC interval estimator.*

### 6.2.2 The OLC method

It is supposed that the interval estimator uses the same straightforward iterative algorithm, which was used before in the binomial case, to obtain one sided interval estimates but with some differences. The main difference for an upper-tail interval is that the algorithm begins by setting $u_0 = 1$, while with the binomial distribution the algorithm starts by setting $u_n = 1$. The algorithm then sequentially determines $u_1, u_2, \ldots, u_x$, where $x$ is the observed

value of $X$. So, we determine $u_i$, then $u_{i+1}$ is determined given the value of $u_i$ and $u_{i+2}$ is determined by the calculated value of $u_{i+1}$ and so on. Specifically, the steps of the algorithm are as follows.

- Set $u_0 = 1$.
- Given $u_i$, use the bisection method to search for the value $u_{i+1}$ that makes the average coverage over the interval $(u_{i+1}, u_i)$ equal to $1 - \alpha$.
- Repeat the previous step for $i = 2, 3, 4, \ldots, x$ to obtain $u_2, u_3, u_4 \ldots, u_x$.

The ordering of these calculated upper endpoints is shown clearly in Figure 6.1, where the coverage probability is plotted against the values of our parameter of interest, $p$. The biggest upper endpoint is $u_0$ (for $x = 0$ when $u_0 = 1$), followed by $u_1$ for $x = 1$, $u_2$ for $x = 2$ and so on until we reach the value of $u_i$ that is of interest, say $u_{20}$ for $x = 20$.

From its construction, the method determines the endpoints of subintervals that have an average coverage of $1 - \alpha$. But, after $u_{i+1}$ has been attained, it could be the case that $Pr(X \leq i + 1|p = u_{i+1})$ is less than $1 - \alpha$, even though $(u_i - u_{i+1})^{-1} \int_{u_{i+1}}^{u_i} Pr(X \leq i|p)dp = 1 - \alpha$. Then, given $u_{i+1}$, there is no $u_{i+2}$ (with $u_{i+2} \leq u_{i+1}$ ) that meets the requirement that $(u_{i+1} - u_{i+2})^{-1} \int_{u_{i+2}}^{u_{i+1}} Pr(X \leq i + 1|p)dp = 1 - \alpha$. The following proposition supports the algorithm by showing that this does not happen for the values of $N$ and $\alpha$ that are of interest in practical circumstances.

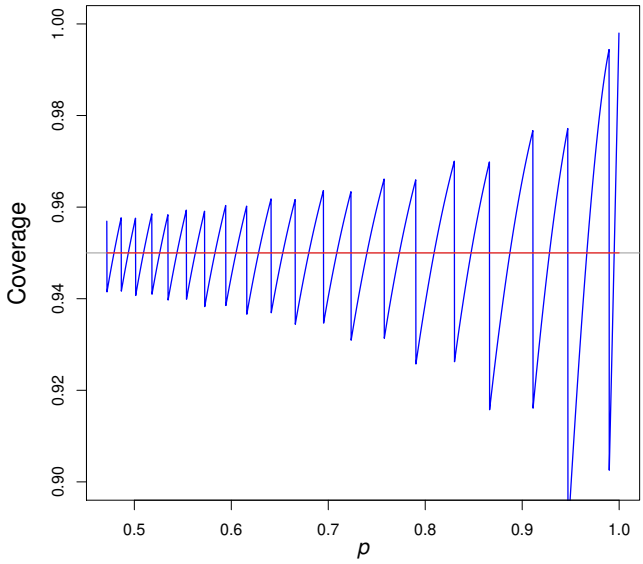


Figure 6.1: Coverage of upper one-sided 97.5% confidence interval of the new estimator for the negative binomial success parameter $p$. The spikes occur at the upper endpoint of the confidence intervals for $x = 20, 19, \ldots, 0$.

**Proposition 6.1**

Suppose $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$. Suppose also that

$$\frac{1}{u_{i+1} - u_{i+2}} \int_{p=u_{i+2}}^{u_{i+1}} \sum_{x=0}^{i+1} \binom{x+r-1}{r-1} p^r (1-p)^x dp = 1 - \alpha \qquad (6.6)$$

and $u_{i+1} > u_{i+2}$ for $i = j+1, j+2, \ldots, 200$. Then for $j \leq 200$, there is a unique $u_j$ such that $u_j > u_{j+1} > 0$ and equation (6.8) holds for $i = j - 1$.

We could not prove the proposition for all values of $x$ and $\alpha$, but by repetitive computation it was proved for the most common values of $x$ and $\alpha$ in practical research: where $x$ is a positive integer less than 200 and $\alpha$ is one of the numbers $0.001, 0.002, \ldots, 0.3$. Throughout this chapter it is assumed that $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$. From Proposition 1 and Definition 1, the new interval estimator gives upper-tail LCC intervals.

Forming the lower-tail intervals differs from forming the upper-tail intervals - its construction is very similar to finding upper end-points for the Poisson distribution. We start with a large value of $N$, calculate $l_N$ using the mid-$p$

method and set $l_N = l_{N.midp}$. Then we get $l_{N-1}$ from $l_N$ using our method of bisection search. We search from left to right. We then obtain $l_{i-1}$ from $l_i$ (or $l_i$ from $l_{i+1}$) until we get $l_0$ from $l_1$. The steps of the algorithm are as follows.

(i) Set $l_N = l_{N.midp}$.

(ii) Given $l_i$, use the bisection method to search for the value of $l_{i-1}$ that makes the average coverage over the interval $(l_i, l_{i-1})$ equal to $1 - \alpha$.

(iii) Perform step (ii) for $i = N$, then for $i = N - 1$, then for $i = N - 2, \ldots$, and finally for $i = 1$.

Figure 6.2 displays the lower endpoints of our new interval estimator, where the coverage probability is plotted against $p$ for a 97.5% confidence interval.

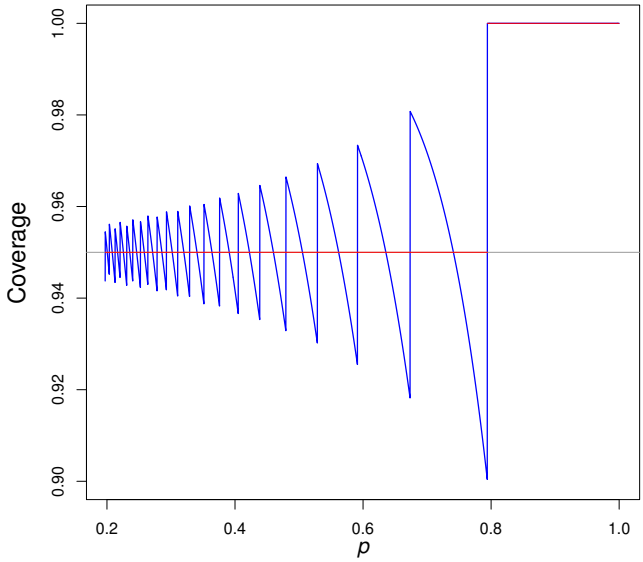


Figure 6.2: Coverage of lower one-sided 97.5% confidence interval of the new estimator for the negative binomial success parameter $p$. The spikes occur at the lower endpoint of the confidence intervals for $x = 20, 19, \ldots, 0$.

In (ii), $l_{i-1}$ is calculated to satisfy

$$\frac{1}{l_{i-1} - l_i} \int_{p=l_i}^{l_{i-1}} \left[ 1 - \sum_{x=0}^{i-1} \binom{x+r-1}{r-1} p^r (1-p)^x \right] dp = 1 - \alpha \qquad (6.7)$$

so the method gives lower-tail LCC intervals from Definition 2.

Two-sided intervals are obtained by combining the endpoints of one-sided upper and lower-tail intervals. So, the new estimator is an LCC interval estimator as it gives $(l_i, u_i)$ as the two-sided equal-tail interval for a confidence interval of $1-2\alpha$. We do not need to start with a large $N$ for the upper limit- we just set $u_0 = 1$ and then calculate $u_1, u_2, \ldots, u_x$, where $x$ is the observed value of $X$.

The number of trials does not have an upper limit and it is not feasible to evaluate the coverage probability or the interval limits for infinitely large $y = x + r$, where $r$ is fixed and $x$ has unbounded limit. This affects the calculation of the lower limit. For lower limits we select a sufficiently large value for $N$, as we did in the Poisson distribution case, and start the iterative procedure by using a standard method (mid-$p$ was used here) to attain a lower limit when $x = N$. The values of $N$ must be much larger than the observed values. As $N$ increases the lower limit (for $X = x$) changes less and less. The idea is that the differences between starting at $N$ or a value larger than $N$ has almost no effect on the lower limit when $X = x$. After comparing different values for $N$ and $x$, we found that for $0 \leq x \leq 100$ putting $N = 500$ will be sufficiently large. While for $x > 100$, putting $N = 5x$ will be enough to represent the infinite range.

The second most important feature of an interval estimator is the lengh of its intervals. We define the expected length for two-sided intervals $(l_x, u_x)$ but it also defines upper-tail intervals by setting $l_x$ equal to 0 and lower-tail intervals by setting $u_x$ equal to 1. The expected length is given by

$$L(p) = \sum_{x=0}^{N} (u_x - l_x) \binom{x+r-1}{r-1} p^r (1-p)^x \tag{6.8}$$

and its average expected length (AEL) is given by

$$E(L(p)) = \int_{p=0}^{1} L(p)dp. \tag{6.9}$$

Proposition 4.2 in Chapter 4 gives necessary and sufficient conditions for the OLC method to yield locally correct confidence intervals with average expected length that is a local minimum for methods that yield locally correct confidence intervals. In applying this proposition, there is a difference between the Poisson and negative binomial distributions, as the parameter of the negative binomial distribution, $p$, has a limited range from 0 to 1 while $\lambda$, the mean of the Pisson distribution, is unbounded. We needed to check if the conditions of the proposition held for commonly used values of $\alpha$ and $x$. The following steps were performed to examine whether the conditions held for the OLC method. Define $p_i = \int_0^1 Pr(X = i|p)dp$ for $i = 0, 1, \ldots$. We examined each $\alpha$ in $\{0.001(0.001)0.3\}$ and $x = 0, 1, \ldots$,200.

(1) Let $(0, u_i)$ be the confidence interval given by the OLC method when $X = i$. Determine $u_0, u_1, \ldots, u_{200}$.

(2) By definition $c_u(p) = \sum_{x=0}^{i} \binom{x+r-1}{r-1} p^r(1-p)^x$ with $p = u_i$. For $i = 0, 1, 2, \ldots, 200$, calculate $h_i = c_u(u_i) - (1-\alpha)$ and calculate $f_i = c_u(u_{i-1}) - (1-\alpha)$ for $i = 1, 2, \ldots, 201$. Then calculate $\psi_{jl} = \prod_{i=j}^{l}(h_i/f_i)$ for $j = 1, 2, \ldots, 200$ and $l = j, j+1, \ldots, 200$.

(3) Calculate $\phi_l = p_l + \sum_{j=1}^{l} \psi_{jl} p_{j-1}$ for $l = 0, 1, \ldots, 200$.

It was found that $\phi_l$ is always negative. This means that our algorithm satisfies the conditions given in Proposition 4.2 for each value of $\alpha$ that was considered. Thus if any of $u_0, \ldots, u_{199}$ are adjusted by a small amount to give $(u_0^*, \ldots, u_{199}^*)$, while $(u_0^*, \ldots, u_{199}^*, u_{200})$ is a partition that gives locally correct confidence intervals, then the expected length of intervals using the partition $(u_0^*, \ldots, u_{199}^*, u_{200})$ is greater than using the partition $(u_0, u_1, \ldots, u_{200})$.

A similar startegy was used for the lower tail. We examined each $\alpha$ in $\{0.001(0.001)0.3\}$ as follows

(1) $c_l(l_i)$ was set equal to $\sum_{x=i}^{\infty} \binom{x+r-1}{r-1} p^r (1-p)^x$ with $p$ set equal to $l_i$ for $i = 0, \ldots, 200$. (The iteration procedure was started at $N$=1000)

(2) Calculate $h_i = c_l(l_i) - (1-\alpha)$ for $i = 0, \ldots, 200$ and $f_i = c_l(l_{i-1}) - (1-\alpha)$ for $i = 1, 2, \ldots, 200$.

(3) Then, calculate $\psi_{jl} = \prod_{i=j}^{l} (h_i/f_i)$ for $j = 1, 2, \ldots, 200$ and $l = j, j+1, \ldots, 200$ and put $\phi_l = p_l + \sum_{j=1}^{l} \psi_{jl} p_{j-1}$ for $l = 0, 1, \ldots, 200$.

The value $\phi_l$ was found to be always positive. Hence adjusting any of $l_0, \ldots, l_{199}$ by a small amount to yield a partition $(l_0^*, \ldots, l_{199}^*, l_{200})$ that gives locally correct confidence intervals will result in $\sum p_i l_i^* < \sum p_i l_i$. Thus we can say that $\sum p_i(u_i - l_i) > \sum p_i(u_i^* - l_i^*)$ so the OLC method also gives two-tailed intervals whose average expected length are a local minimum.

Regarding the desirable properties in interval estimators that were discussed in Section 4.5, we found that the OLC method will always have Property 1, *interval valued*. Also, it has the properties of *monotone decreasing in* $x$, as the interval end-points decrease monotonically as $x$ increases. In addition it has the Property 3, *nesting*, which states that if two confidence intervals have different confidence levels then, for any given $x$, the interval for the higher confidence level should contain the interval with the lower confidence level. By achieving these properties, the OLC can be considered a well-behaved interval estimator.

## 6.3 The negative binomial mean ($\mu$)

As mentioned in the introduction, we are also interested in calculating confidence intervals for the negative binomial mean $\mu = E(X)$, where $\mu$ is the mean number of failures. There are some differences between the confidence

intervals for both $p$ and $\mu$, as $p \in [0,1]$ while $\mu \in [0,\infty]$. Also, the coverage probability function of the confidence interval of $\mu$ is monotone increasing in $x$ while the coverage probability function for the negative binomial parameter $p$ is monotone decreasing in $x$.

The lower and upper endpoints of the confidence interval for $\mu$ can be calculated by one of two approaches. By transforming the CI of $p$ or by searching directly for the endpoints, which will be mentioned in the following section. The problem with using the transformation is that the transformed in terms are not locally correct confidence intervals-the average coverage between the transformed spikes will not necessarily exceed the nomial confidence level. Using the direct approach gives intervals that are LCC intervals and so it is the method we adopt. To use the direct calculation, we search for the endpoints of the confidence interval of $\mu$ such that the average coverage between every two points equal to the nominal level $(1-\alpha)$. So, in the following work, we will follow the direct search for the endpoints of the confidence interval for $\mu$.

### 6.3.1 Locally correct confidence interval

We first consider upper-tail intervals and suppose an interval estimator gives upper-tail estimate $(0, u_x)$ for $\mu$. We assume that

$$0 < u_0 < u_1 < u_2 < \dots. \tag{6.10}$$

The coverage probability is the probability that the random interval $(0, u_x)$ contains $\mu$. So, when $u_i < \mu \leq u_{i+1}$, the coverage probability is given by

$$C_u(\mu) = Pr(X \geq i+1|\mu) = \sum_{x=i+1}^{\infty} \binom{x+r-1}{r-1} \left(\frac{r}{\mu+r}\right)^r \left(1-\left(\frac{r}{\mu+r}\right)\right)^x \tag{6.11}$$

We require the average coverage between every two spikes to exceed or equal the nominal level $(1-\alpha)$.

*Definition 4.* For the upper-tail, suppose an interval estimator gives $(0, u_x)$ as its upper-tail interval for $\mu$ when $X = x$ and that $u_0, u_1, u_2, \ldots$ satisfies equation (6.12). If for $i = 1, 2, \ldots$.

$$\frac{1}{u_i - u_{i-1}} \int_{\mu=u_{i-1}}^{u_i} C_u(\mu) d\mu \geq 1 - \alpha \tag{6.12}$$

then the interval estimator gives *upper-tail LCC intervals* with confidence interval $(1-\alpha)$.

*Definition 5.* For the lower-tail intervals, suppose an interval estimator gives $(l_x, \infty)$ as its lower-tail interval for $\mu$ when $X = x$ and that $0 = l_0 \leq l_1 \leq l_2 \leq \ldots$. The coverage probability is defined as

$$C_l(\mu) = Pr(X \leq i|\mu) = \sum_{x=0}^{i} \binom{x+r-1}{r-1} \left(\frac{r}{\mu+r}\right)^r \left(1 - \left(\frac{r}{\mu+r}\right)\right)^x \tag{6.13}$$

for $i = 0, 1, 2, \ldots$, if $l_i \leq \mu < l_{i+1}$. If for $i = 0, 1, \ldots$.

$$\frac{1}{l_{i+1} - l_i} \int_{\mu=l_i}^{l_{i+1}} C_l(\mu) d\mu \geq 1 - \alpha \tag{6.14}$$

and then the interval estimator gives *lower-tail LCC intervals* with confidence level $(1-\alpha)$.

*Definition 6.* For two sided intervals, suppose that, for $x = 0, 1, 2, \ldots$, an interval estimator gives $(l_x, u_x)$ as its two-sided equal-tails intervals for $\mu$ when $X = x$. Then it gives equal-tail LCC intervals with confidence level $(1-2\alpha)$, as the intervals $(l_x, \infty)$ and $(0, u_x)$ are sets of one-sided lower-tail and upper-tail LCC intervals, respectively, each with confidence level $(1-\alpha)$.

### 6.3.2 The OLC method

We search for the end-points of the confidence interval of $\mu$ such that the average coverage between every two consecutive spikes is equal to the nominal level $(1-\alpha)$. The interval estimator uses a similar algorithm to that used previously with the negative binomial proportion $p$ and parameters of other distributions. The steps for obtaining one-sided interval estimates for $\mu$ are almost the same as for obtaining the one-sided interval estimates for the Poisson parameter. We begin by specifying a large value for $N$ and then choosing a reasonable value of upper end-points to represent the last interval $u_N$. The Mid-$p$ method can be used to calculate this value, which is $u_{N.midp}$, and then a simple numerical search is used to sequentially determine $u_{N-1}, u_{N-2}, \ldots, u_0$. The value of $N$ is set equal to $N = 500$ for small values of $0 \leq x \leq 100$, and $N = 5x$ for $x > 100$. These values of $N$ are sufficienlty large - increasing the value of $N$ has almost no effect on the upper endpoint for $X = x$. The steps of the algorithm are as follows

1. Set $u_N$=$u_{N.midp}$.

2. Put $i = N - 1$ and, given $u_{i+1}$, use the bisection method to search for the value $u_i$ that makes the average coverage over the interval $(u_i, u_{i+1})$ equal to $1 - \alpha$:

3. Given $u_{N-1}$, repeat step 2 for $i = N-2, N-3, N-4, \ldots, 1, 0$ to obtain $u_{N-2}, u_{N-3}, u_{N-4}, \ldots, u_1, u_0$.

Figure 6.3 shows the coverage of the upper one-sided 97.5% confidence interval of the negative binomial mean $\mu$ given by our new estimator, where the coverage is plotted against $\mu$. The calculated upper points are $u_0 < u_1 < u_2 < \ldots < u_N$. As we start with $u_N$, which is calculated by using the mid-$p$ method, $u_{N.midp}$=$u_N$. Then $u_{N-1}, u_{N-2}, u_{N-3}, \ldots, u_0$ are calculated in turn.

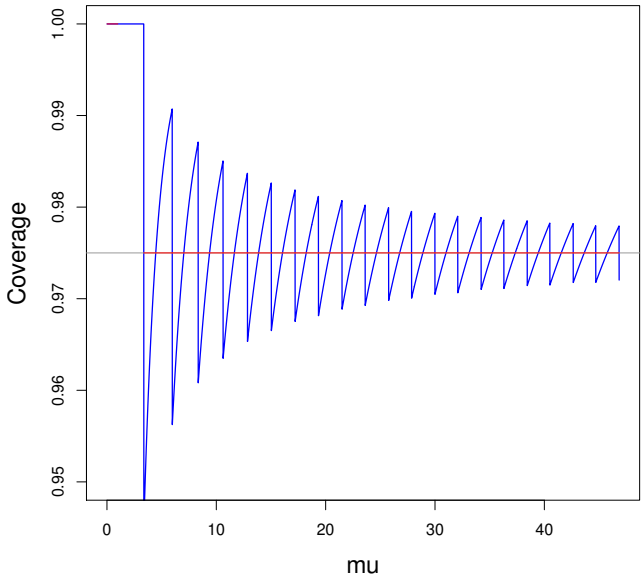


Figure 6.3: Coverage of upper one-sided 97.5% confidence interval of the new estimator for the negative binomial mean $\mu$. The spikes occur at the upper endpoints of the confidence intervals for $x = 0, 1, \ldots, 20$

The following proposition supports the algorithm for values of $x$ and $\alpha$ that are most common in practical research. It was proved by direct computation.

**Proposition 6.2**

Suppose $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$. Suppose also that

$$\frac{1}{u_i - u_{i-1}} \int_{\mu=u_{i-1}}^{u_i} \left[ 1 - \sum_{x=0}^{i-1} \binom{x+r-1}{r-1} \left( \frac{r}{\mu+r} \right)^r \left( 1 - \left( \frac{r}{\mu+r} \right) \right)^x \right] d\mu = 1-\alpha \tag{6.15}$$

and $u_i > u_{i-1}$ for $i = j+1, j+2, \ldots, 200$. Then there is a unique $u_{j-1}$ such that $u_j > u_{j-1} > 0$ and equation (6.19) holds when $j = i$. To calculate the first upper endpoint we set $N = 1000$ and put $u_{1000} = u_{N.midp}$.

From Proposition 6.2 and Definition 4, the new interval estimator for $\mu$ gives upper-tail LCC intervals.

Forming the lower-tail intervals differs from forming the upper-tail intervals. It is similar to forming the lower-tail intervals in the case of the binomial and Poisson distributions. To form the lower-tail intervals, we begin with the first interval by setting $l_0 = 0$ and then $l_1, l_2, \ldots, l_N$ are determined sequentially.

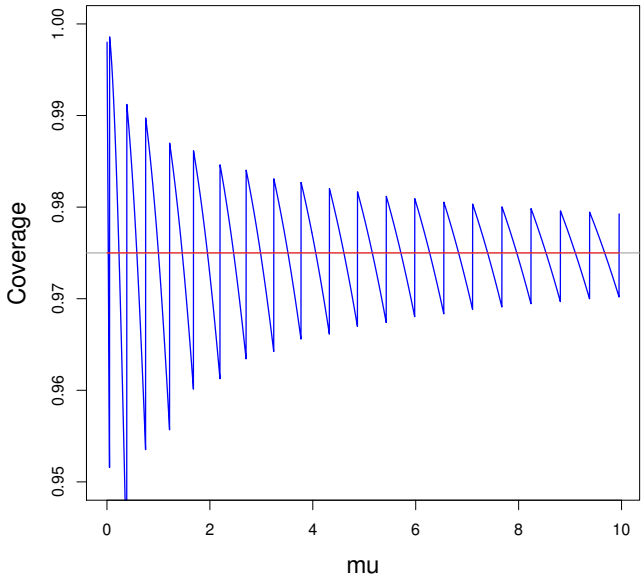


Figure 6.4: Coverage of lower one-sided 97.5% confidence interval of the new estimator for the negative binomial mean $\mu$. The spikes occur at the lower endpoints of the confidence intervals for $x = 0, 1, \ldots, 20$

So, given $l_i$ we can determine $l_{i+1}$. The coverage of the 97.5% lower-tail intervals is plotted against $\mu$ in Figure 6.4. Given $l_i$, the value of $l_{i+1}$ is calculated to satisfy

$$\frac{1}{l_{i+1} - l_i} \int_{\mu=l_i}^{l_{i+1}} \sum_{x=0}^{i} \binom{x+r-1}{r-1} \left(\frac{r}{\mu+r}\right)^r \left(1 - \left(\frac{r}{\mu+r}\right)\right)^x d\mu = 1 - \alpha \tag{6.16}$$

for $i = 1, 2, \ldots, 200$. From Definition 5, the new interval estimator for $\mu$ gives lower-tail LCC intervals.

Both upper and lower tail intervals are combined to give two-sided intervals. Thus, the new estimator for $\mu$ is an LCC interval estimator as it gives $(l_i, u_i)$ as the two-sided equal-tail interval for a confidence interval of $1 - 2\alpha$.

Moving to the length of the confidence interval of the interval estimator, the average expected length is infinite so instead we consider the weighted average expected length. As the range of $\mu$ has no upper bound, we use Corollary 4.1 in Chapter 4 and check conditions for a local optimality property that is suited to an estimator whose interval has unlimited length. The corollary considers

the weighted average expected length of intervals for a specified weighting function, $w(\mu)$ say. Put

$$p_i = \int_0^\infty Pr(X = i|\mu)w(\mu)d\mu \tag{6.17}$$

for $i = 0, 1, \ldots$. By definition, a random variable $Z$ has a beta prime $(a, b)$ distribution if its probablilty density function is

$$f(Z) = \frac{1}{B(a,b)} \frac{Z^{a-1}}{(1+Z)^{a+b}}. \tag{6.18}$$

We suppose that $w(\mu)$ has the form

$$w(\mu) = \frac{1}{r} \frac{1}{B(a,b)} \frac{(\frac{\mu}{r})^{a-1}}{(1+\frac{\mu}{r})^{a+b}}. \tag{6.19}$$

From equation (6.13), we find that

$$\begin{aligned} Pr(X = i|\mu) &= \binom{i+r-1}{r-1} \left(\frac{r}{\mu+r}\right)^r \left(1 - \left(\frac{r}{\mu+r}\right)\right)^i. \\ &= \frac{(r+i-1)!}{i!(r-1)!} \frac{(\frac{\mu}{r})^i}{(1+\frac{\mu}{r})^{r+i}} \end{aligned} \tag{6.20}$$

Then

$$\begin{aligned} p_i &= \frac{1}{B(a,b)} \frac{(r+i-1)!}{i!(r-1)!} \int_0^\infty \frac{(\frac{\mu}{r})^i}{(1+\frac{\mu}{r})^{r+i}} \frac{(\frac{\mu}{r})^{a-1}}{(1+\frac{\mu}{r})^{a+b}} \frac{1}{r} d\mu \\ &= \frac{1}{B(a,b)} \frac{(r+i-1)!}{i!(r-1)!} \int_0^\infty \frac{(\frac{\mu}{r})^{i+a-1}}{(1+\frac{\mu}{r})^{a+b+r+i}} \frac{1}{r} d\mu \\ &= \frac{1}{B(a,b)} \frac{(r+i-1)!}{i!(r-1)!} B(a+i, b+r) \\ &= \frac{(a+b-1)!}{(a-1)!(b-1)!} \frac{(r+i-1)!}{i!(r-1)!} \frac{(i+a-1)!(r+b-1)!}{(a+b+r+i-1)!}. \end{aligned} \tag{6.21}$$

To determine the weighting function $w(\mu)$, we need to choose the parameter values $(a, b)$ of the beta prime $(a, b)$ distribution. We are interested in finding values of the parameters $a$ and $b$ for which $w(\mu)$ gives more weight to small vales of $\mu$ and less weight to large values of $\mu$. After some trials, we found that the values $a = 2$ and $b = 3$ seemed suitable. Specifically, when the value of $x$ is increased above 200, the increase in the value of weighted average expected length is very small. So we will use these values throughout this section. We examine if the conditions held for the OLC method with

$p_0, p_1, \ldots, p_{200}$ given in equation (6.25) for each $\alpha$ in $\{0.001(0.001)0.3\}$ and $x$ truncated at 200.

(1) Let $(0, u_i)$ be the confidence interval given by the OLC method when $X = i$. Determine $u_0, u_1, \ldots, u_{200}$.

(2) By definition $c_u(u_i) = \sum_{x=i}^{\infty} \binom{x+r-1}{r-1} (\frac{r}{\mu+r})^r (1 - (\frac{r}{\mu+r}))^x$ with $\mu = u_i$. For $i = 0, 1, 2, \ldots, 200$, calculate $h_i = c_u(u_i) - (1 - \alpha)$ and calculate $f_i = c_u(u_{i-1}) - (1 - \alpha)$ for $i = 1, 2, \ldots, 201$. Then calculate $\psi_{jl} = \prod_{i=j}^{l} (h_i / f_i)$ for $j = 1, 2, \ldots, 200$ and $l = j, j + 1, \ldots, 200$.

(3) Calculate $p_i = \frac{(a+b-1)!}{(a-1)!(b-1)!} \frac{(r+i-1)!}{i!(r-1)!} \frac{(i+a-1)!(r+b-1)!}{(a+b+r+i-1)!}$ for $i = 0, 1, 2, \ldots, 200$, $a = 2; b = 3$.

(4) Calculate $\phi_l = p_l + \sum_{j=1}^{l} \psi_{jl} p_{j-1}$ for $l = 0, 1, \ldots, 200$.

If $\phi_l$ is always positive, then our algorithm gives the locally shortest interval. The computational study which was used considered each value of $\alpha$ from 0.001 to 0.3 and $\phi_l$ was always positive. Thus if any of $u_0, \ldots, u_{199}$ are adjusted by a small amount to give $(u_0^*, \ldots, u_{199}^*)$, while $(u_0^*, \ldots, u_{199}^*, u_{200}^*)$ is a partition that gives locally correct confidence intervals, then the weighted average expected length of intervals is increased (i.e. $\sum p_i u_i^* > \sum p_i u_i$).

Also, the weighted average expected length of lower-tail intervals was examined for $\alpha$ in $\{0.001(0.001)0.3\}$. By letting $l_i$ denote the lower limit when $X = i (i = 0, \ldots, 200)$, we followed these steps:

(1) $c_l(l_i)$ was set equal to $\sum_{x=0}^{i} \binom{x+r-1}{r-1} (\frac{r}{\mu+r})^r (1 - (\frac{r}{\mu+r}))^x$ with $\mu$ set equal to $l_i$ for $i = 0, \ldots, 200$.

(2) Calculate $h_i = c_l(l_i) - (1 - \alpha)$ for $i = 0, \ldots, 200$ and $f_i = c_l(l_{i-1}) - (1 - \alpha)$ for $i = 1, 2, \ldots, 200$.

(3) As in the procedure for the upper endpoints, calculate $\psi_{jl} = \prod_{i=j}^{l} (h_i / f_i)$ for $j = 1, 2, \ldots, 200$ and $l = j, j + 1, \ldots, 200$ and put $\phi_l = p_l + \sum_{j=1}^{l} \psi_{jl} p_{j-1}$ for $l = 0, 1, \ldots, 200$.

The value of $\phi_l$ was always negative. Hence adjusting any of $l_0, \ldots, l_{199}$ by a small amount to yield a partition $(l_0^*, \ldots, l_{199}^*, l_{200}^*)$ that gives locally correct confidence intervals will result in $\sum p_i l_i^* < \sum p_i l_i$. It follows that $\sum p_i(u_i - l_i) > \sum p_i(u_i^* - l_i^*)$ so the OLC method also gives two-tailed intervals whose average expected length is a local minimum.

We can say that the OLC method should be the preferred method of forming LCC intervals because its average expected length is a local minimum and it has the properties, discussed in Section 4.3, that a well-behaved interval estimator should have. It will always have Property 1 ( a confidence region should be an interval and not a collection of disjoint intervals). It also has both Property 2 (monotonicity in $x$, monotone increasing) and Property 3 (the nesting property). Achieving these properties was examined by repetitive computation for $1 \leq x \leq 200$ and $\alpha$ in $\{0.001(0.001)0.3\}$.

## 6.4 Comparison with other methods

In this section we compare the OLC method performance for the negative binomial proportion $p$ with the following five methods of forming confidence intervals, which have been discussed earlier in Section 2.4: exact, mid-$p$, Jeffreys, Wald and score methods. This comparison is in terms of their coverage probability and expected length for the number of failures $x < 5000$. The cases examined are where the number of successes $r = 1000, 1500$ and 2000 and nominal confidence levels of 95%, 97.5% and 99.5% are considered.

### 6.4.1 Coverage probability

We will concentrate on upper-tail intervals and lower-tail intervals as they are more informative than two-tail intervals. As mentioned earlier, $C_u(p)$ and $C_l(p)$ are the coverage of an upper-tail interval estimator and the coverage of

a lower-tail interval estimator, respectively. These are the probabilities that the random interval $(0, u_x)$ or $(l_x, 1)$ contains $p$. They are defined in equations (6.2) and (6.5).

We follow the same approach as for both the binomial and Poisson distributions to calculate the average coverage $Av.Cov_u$, where

$$Av.Cov_u = \int_0^1 C_u(p)dp. \tag{6.22}$$

We calculate the average coverage for all the above methods for 50000 values of $p$ that covered the interval $(0, 1)$.

In Figures 6.5, 6.6 and 6.7 , the coverage of 95%, 97.5% and 99.5% upper-tail intervals of all methods are plotted against $p$ for the number of success $r = 10, 30$ and 50, respectively. It is clear that the Wald, score and Jeffreys methods do not give LCC intervals, as the average coverage between consecutive spikes is sometimes below the targeted confidence level. The OLC, exact and mid-$p$ methods are giving LCC intervals as the average coverage between consecutive spikes is at least $(1 - \alpha)$. To make the comparison between our OLC method and other methods clearer, the values of the average coverage for the upper-tail of each method is given in Table 6.1. From the table, the OLC method has good average coverage as it equals the nominal confidence level for all cases of $\alpha$ and $r$. Unlike the exact method, which suffers from conservatism and the mid-$p$ method, which is a little bit conservative. The Wald method is very conservative, unlike its liberal performance for both the binomial and Poisson cases. Both the Jeffreys and score methods are consistently liberal.

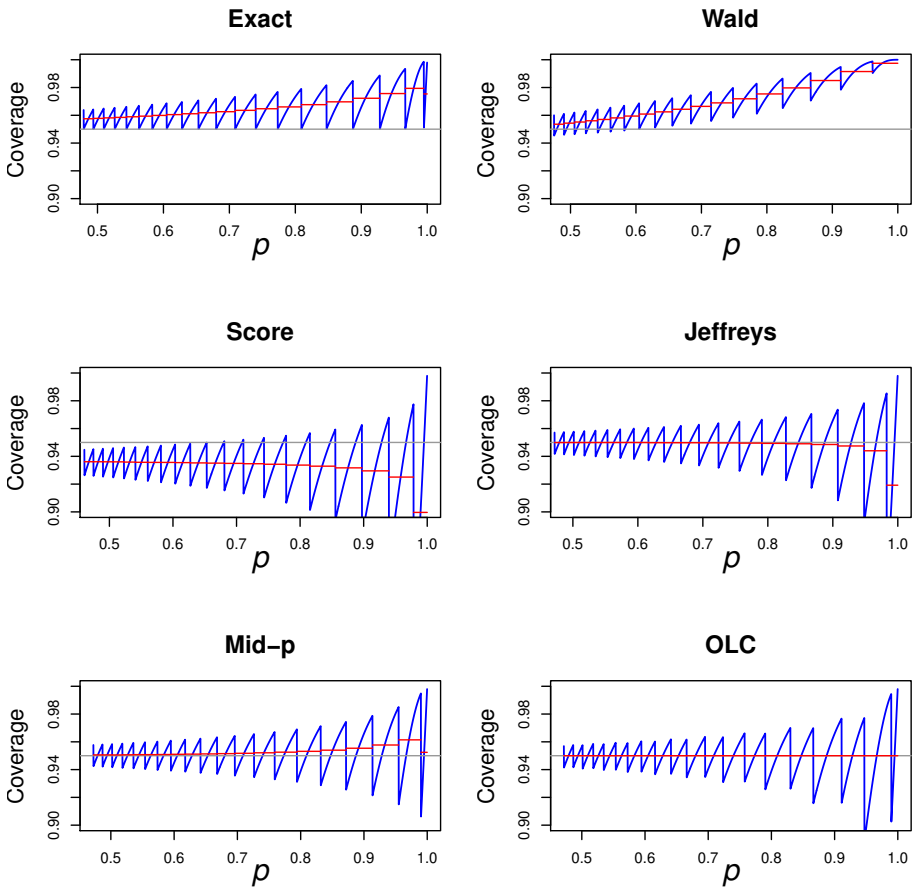


Figure 6.5: Coverage of upper-tail 95% for exact, Wald, score, Jeffreys, mid-$p$ and OLC methods plotted against $p$ for $r = 10$. The spikes occur at the upper endpoints of the confidence intervals for $x = 20, 19, \ldots, 0$.

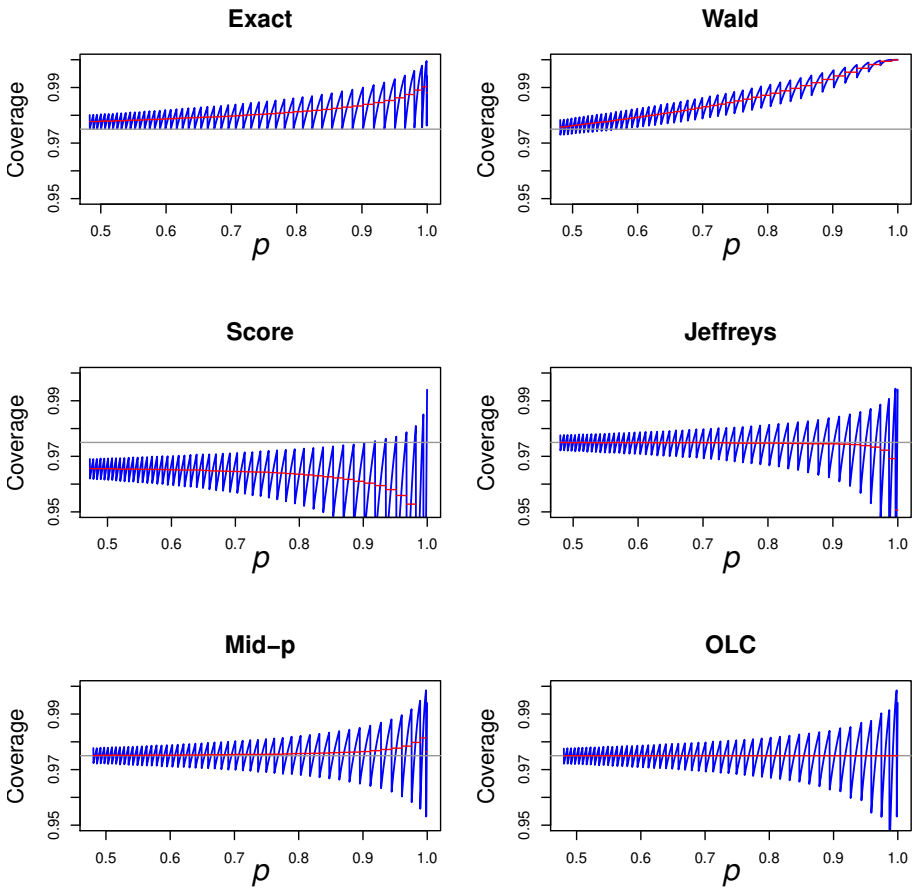


Figure 6.6: Coverage of upper-tail 97.5% for exact, Wald, score, Jeffreys, mid-$p$ and OLC methods plotted against $p$ for $r = 30$. The spikes occur at the upper endpoints of the confidence intervals for $x = 50, 49, \ldots, 0$.

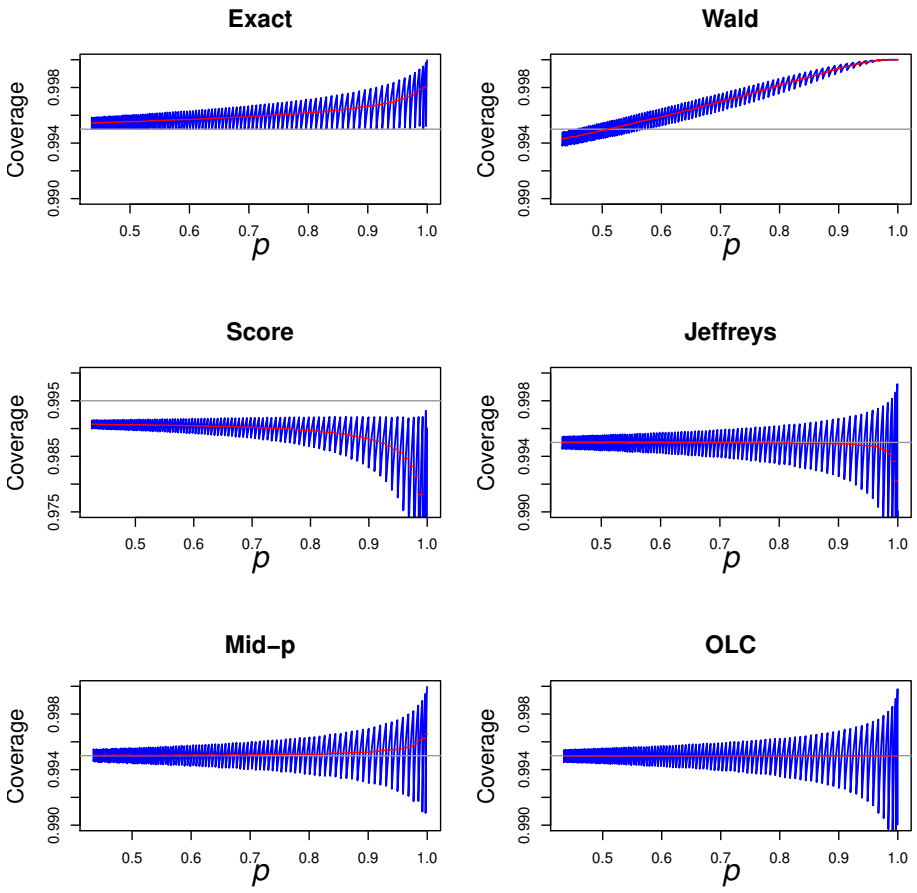


Figure 6.7: Coverage of upper-tail 99.5% for exact, Wald, score, Jeffreys, mid-$p$ and OLC methods plotted against $p$ for $r = 50$. The spikes occur at the upper endpoints of the confidence intervals for $x = 100, 99, \ldots, 0$.

Table 6.1: Average coverage (Av.Cov) of upper-tail $1-\alpha$ intervals of six methods of forming interval estimates, for $\alpha = 0.05, 0.025, 0.005$ and $r = 10, 30$ and 50.

| $\alpha$ | $r$ | Exact | Wald | Score | Jeffreys | Mid-$p$ | OLC |
|---|---|---|---|---|---|---|---|
| 0.05 | 10 | 0.96045 | 0.96029 | 0.9338839 | 0.94889 | 0.95187 | 0.95000 |
| 0.05 | 30 | 0.95842 | 0.96097 | 0.9387851 | 0.94938 | 0.95109 | 0.95000 |
| 0.05 | 50 | 0.95766 | 0.96117 | 0.940652 | 0.94953 | 0.95086 | 0.95000 |
| 0.025 | 10 | 0.98050 | 0.97875 | 0.957938 | 0.97430 | 0.97605 | 0.97500 |
| 0.025 | 30 | 0.97950 | 0.98074 | 0.963569 | 0.97461 | 0.97563 | 0.97500 |
| 0.025 | 50 | 0.97915 | 0.98140 | 0.9656062 | 0.97470 | 0.97552 | 0.97500 |
| 0.005 | 10 | 0.99620 | 0.99421 | 0.98350 | 0.99481 | 0.99525 | 0.99500 |
| 0.005 | 30 | 0.99602 | 0.99587 | 0.9879458 | 0.99489 | 0.99517 | 0.99500 |
| 0.005 | 50 | 0.99595 | 0.99640 | 0.9893693 | 0.99491 | 0.99514 | 0.99500 |

Regarding the lower-tail interval, in contrast to the upper-tail interval of the negative binomial proportion and the lower-tail interval of the Poisson case, we exclude the values of $p$ in the range $(l_0, 1)$. That is because the coverage when $p$ is in this interval $(l_0, 1)$ equals 1. So, we calulate the truncated average coverage $T_l$, which is defined as

$$T_l = \frac{1}{1 - l_0} \int_{p=l_0}^{1} C_l(p)\, dp, \tag{6.23}$$

where $l_0$ is the endpoint of $x = 0$.

Figures 6.8, 6.9 and 6.10 show the coverage of 95%, 97.5% and 99.5% lower tail intervals of all methods, plotted against $p$ for the number of successes $r = 10, 30$ and 50, respectively. It is clear that the coverage of the first interval, from the right of $l_0$, equals 1. For all tested values of $r$, $\alpha$ and for the whole range of $p$, the OLC, exact and mid-$p$ still keep on giving LCC intervals while the Wald, score and Jeffreys methods do not give LCC intervals. The comparison between all these methods are given in the Table 6.2. The $T_l$ values are given for combination of $\alpha = 0.05, 0.025$ and 0.005 and number of successes $r = 10, 30$ and 50. From the results in the table, the OLC method gives a good result for all cases of $\alpha$ and $r$. The score method becomes conservative as $\alpha$ increased and the exact method suffers also from conservatism as usual. While the mid-$p$ is a little conservative, both the Wald and Jeffreys methods are liberal.

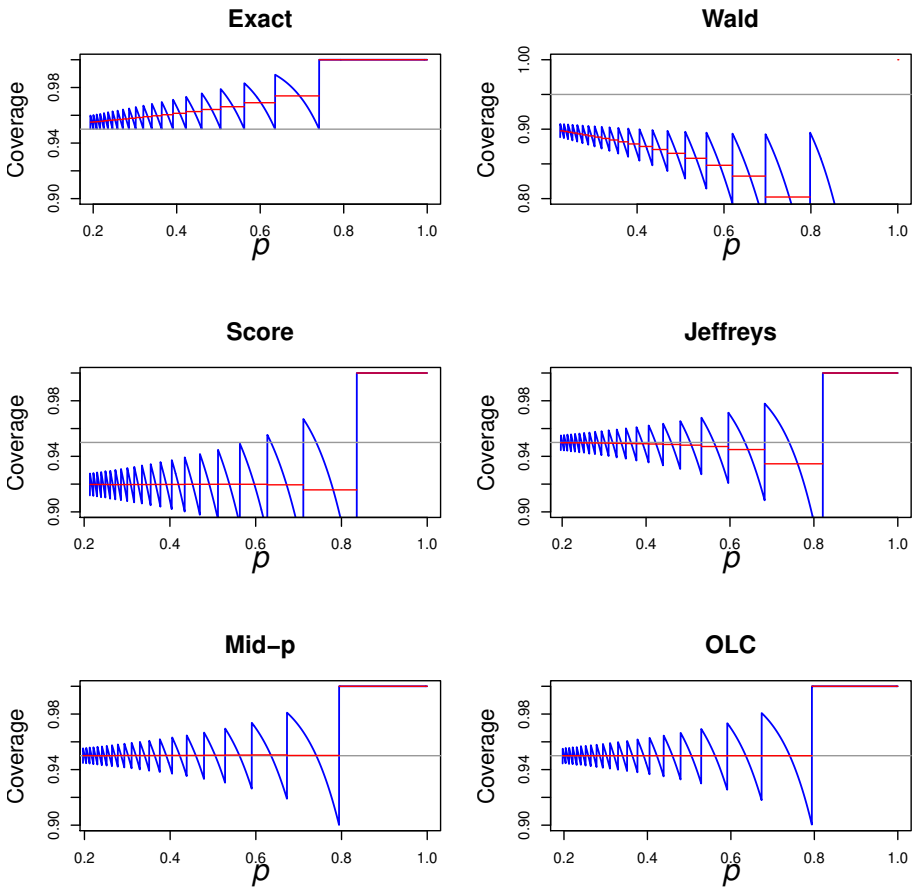


Figure 6.8: Coverage of lower-tail 95% for exact, Wald, score, Jeffreys, mid-$p$ and OLC methods plotted against $p$ for $r = 10$. The spikes occur at the lower endpoints of the confidence intervals for $x = 20, 19, \ldots, 0$.

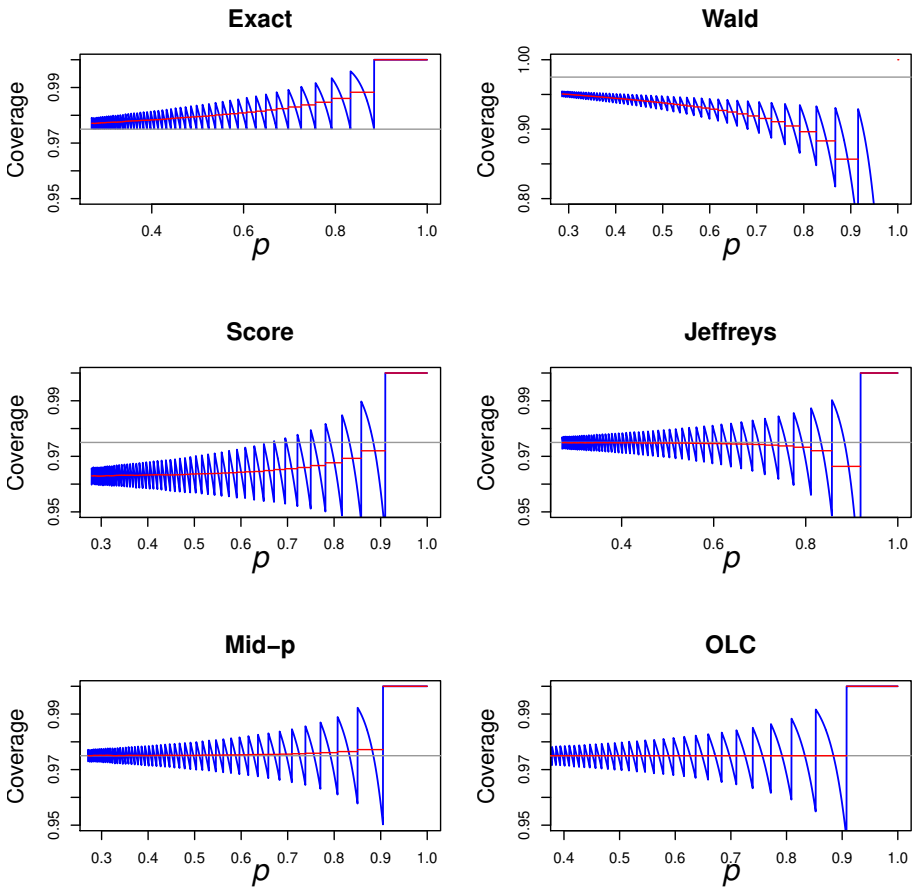


Figure 6.9: Coverage of lower-tail 97.5% for exact, Wald, score, Jeffreys, mid-$p$ and OLC methods plotted against $p$ for $r = 30$. The spikes occur at the lower endpoints of the confidence intervals for $x = 50, 49, \ldots, 0$.

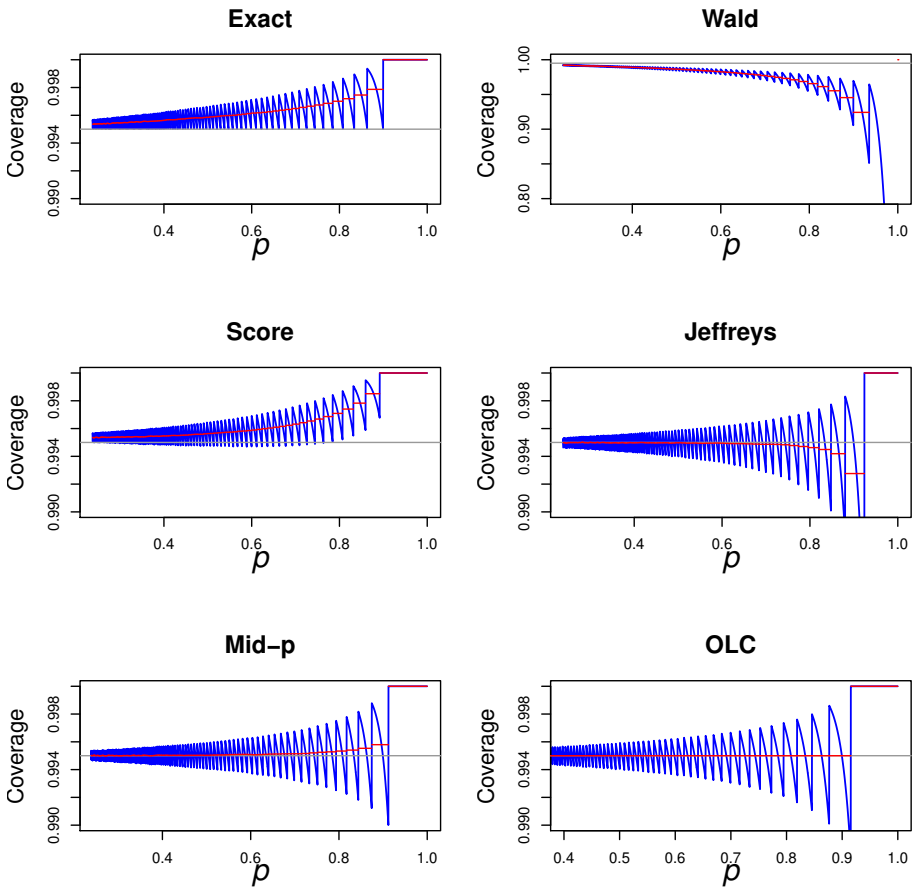


Figure 6.10: Coverage of lower-tail 99.5% for exact, Wald, score, Jeffreys, mid-$p$ and OLC methods plotted against $p$ for $r = 50$. The spikes occur at the lower endpoints of the confidence intervals for $x = 100, 99, \ldots, 0$.

Table 6.2: Average coverage (Av.Cov) of lower-tail $1-\alpha$ intervals over the range $(l_0, 1)$ for six methods of forming interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and $r = 10, 30$ and 50.

| $\alpha$ | $r$ | Exact | Wald | Score | Jeffreys | Mid-$p$ | OLC |
|---|---|---|---|---|---|---|---|
| 0.05 | 10 | 0.96166 | 0.85573 | 0.97621 | 0.94613 | 0.95020 | 0.95000 |
| 0.05 | 30 | 0.95949 | 0.90252 | 0.96476 | 0.94825 | 0.95061 | 0.95000 |
| 0.05 | 50 | 0.95852 | 0.91348 | 0.96155 | 0.94878 | 0.95056 | 0.95000 |
| 0.025 | 10 | 0.98129 | 0.87858 | 0.99485 | 0.97259 | 0.97531 | 0.97500 |
| 0.025 | 30 | 0.98020 | 0.92891 | 0.98741 | 0.97388 | 0.97547 | 0.97500 |
| 0.025 | 50 | 0.97970 | 0.94045 | 0.98495 | 0.97422 | 0.97542 | 0.97500 |
| 0.005 | 10 | 0.99642 | 0.89861 | 0.99998 | 0.99433 | 0.99514 | 0.99500 |
| 0.005 | 30 | 0.99620 | 0.95305 | 0.99920 | 0.99467 | 0.99516 | 0.99500 |
| 0.005 | 50 | 0.99610 | 0.96533 | 0.99864 | 0.99477 | 0.99514 | 0.99500 |

According to all previous examples and figures in this section, the mid-$p$ method gives LCC intervals. Direct computation showed that it has this property for $x \leq 200$ and $\alpha \in \{0.001(0.001)0.1\}$, a result recorded in the following proposition.

**Proposition 6.3**

For $x \leq 200$ and $\alpha \in \{0.001(0.001)0.1\}$, the mid-$p$ method gives LCC intervals.

Its actual coverage is illustrated in Figure 6.5, as an example. The coverage always crosses the nominal coverage level between consecutive spikes. In addition, it gives one-tailed confidence interval whose coverage is close to the targeted nominal level for any value of $p$. In fact, it has the property given in the following proposition.

**Proposition 6.4**

The mid-$p$ has the smallest root-mean-square error in coverage probability, $|C_u(p) - (1-\alpha)|$ or $|C_l(p) - (1-\alpha)|$ , among any method of forming one-tailed confidence intervals.

The proof of Proposition 6.4 is given in Appendix B.

Root mean-square error (RMSE) in coverage over the range of $p$ is considered a good measure for comparing methods. A good method of forming confi-

dence intervals should have a small RMSE, as it examines how the coverage probability of the interval estimator typically varies from the nominal confidence level. So, we calculated the RMSE of the upper-tail for each method's coverage over the whole range of $p$:

$$\text{RMSE} = \left[\int_0^1 \{C_u(p) - (1-\alpha)\}^2 \, dp\right]^{1/2}, \tag{6.24}$$

where $(1-\alpha)$ is the nominal confidence level. Table 6.3 shows the results for each method for $\alpha = 0.05$, 0.025, 0.005 and $r = 10, 30$ and 50. In accordance with Proposition 6.4, the mid-$p$ method has the smallest RMSE of any method - the mid-$p$ method has the smallest RMSE in every row of Table 6.3. The OLC method has the second smallest RMSE in every row with an increase of not more than 5% compared with the RMSE of the mid-$p$ method. This is much better than the RMSE of the other methods, as the RMSE of Jeffreys method is sometimes more than 30% bigger than the RMSE of the mid-$p$ method, while other methods have at least one RMSE that is more than 70% bigger than the mid-$p$ method. Hence, we can say that the OLC method has a very acceptable RMSE, even if it does not have the smallest RMSE.

Table 6.3: Root mean-square error (RMSE) of coverage of upper tail $(1-\alpha)$ intervals for six methods of forming interval estimates, $\alpha = 0.05$, 0.025, 0.005 and $r = 10, 30$ and 50.

| $\alpha$ | $r$ | Exact | Wald | Score | Jeffreys | Mid-$p$ | OLC |
|---|---|---|---|---|---|---|---|
| 0.05 | 10 | 0.01321 | 0.01898 | 0.01787 | 0.00902 | 0.00862 | 0.00873 |
| 0.05 | 30 | 0.01106 | 0.01408 | 0.01679 | 0.00729 | 0.00696 | 0.00706 |
| 0.05 | 50 | 0.01017 | 0.01213 | 0.01609 | 0.00658 | 0.00631 | 0.00638 |
| 0.025 | 10 | 0.00695 | 0.01806 | 0.01098 | 0.00499 | 0.00462 | 0.00473 |
| 0.025 | 30 | 0.00587 | 0.01272 | 0.01009 | 0.00405 | 0.00377 | 0.00386 |
| 0.025 | 50 | 0.00545 | 0.01077 | 0.00977 | 0.00365 | 0.00343 | 0.00351 |
| 0.005 | 10 | 0.00151 | 0.01160 | 0.00416 | 0.00118 | 0.00103 | 0.00108 |
| 0.005 | 30 | 0.00131 | 0.00751 | 0.00279 | 0.00097 | 0.00086 | 0.00090 |
| 0.005 | 50 | 0.00122 | 0.00617 | 0.00257 | 0.00089 | 0.00079 | 0.00082 |

The RMSE of the lower-tail for each method is calculated in the same way as the RMSE of the upper tail except that it is determined over the truncated range $(l_0, 1)$. This RMSE is given by

$$RMSE = \left[\frac{1}{1-l_0}\int_{l_0}^{1}\{C_l(p) - (1-\alpha)\}^2 dp\right]^{1/2}. \tag{6.25}$$

Table 6.4. shows the RMSE of lower-tail intervals for each method for $\alpha$ = 0.05, 0.025, 0.005 and $r = 10, 30$ and 50. The results of this table confirm that the mid-$p$ method has a smaller RMSE than other methods. The OLC method comes next as it has the second smallest RMSE, a little bigger than mid-$p$ method. Each of the other methods has an RMSE that is bigger, up to about 80% bigger than the RMSE of the mid-$p$ method. Lastly, the score method records the biggest RMSE. Hence, the OLC method has an acceptable RMSE relative to other methods.

Table 6.4: Root mean-square error (RMSE) of coverage of lower tail $(1-\alpha)$ intervals for six methods of forming interval estimates, $\alpha$ = 0.05, 0.025, 0.005 and $r = 10, 30$ and 50.

| $\alpha$ | $r$ | Exact | Wald | Score | Jeffreys | Mid-$p$ | OLC |
|---|---|---|---|---|---|---|---|
| 0.05 | 10 | 0.01126 | 0.05600 | 0.02228 | 0.00961 | 0.00811 | 0.00815 |
| 0.05 | 30 | 0.01107 | 0.04344 | 0.01466 | 0.00839 | 0.00737 | 0.00753 |
| 0.05 | 50 | 0.01042 | 0.03789 | 0.01214 | 0.00758 | 0.00676 | 0.00699 |
| 0.025 | 10 | 0.00583 | 0.04681 | 0.01534 | 0.00524 | 0.00426 | 0.00445 |
| 0.025 | 30 | 0.00594 | 0.03743 | 0.01147 | 0.00473 | 0.00402 | 0.00453 |
| 0.025 | 50 | 0.00564 | 0.03270 | 0.00960 | 0.00430 | 0.00372 | 0.00393 |
| 0.005 | 10 | 0.00120 | 0.02794 | 0.00286 | 0.00118 | 0.00089 | 0.00091 |
| 0.005 | 30 | 0.00132 | 0.02424 | 0.00364 | 0.00115 | 0.00092 | 0.00096 |
| 0.005 | 50 | 0.00127 | 0.02125 | 0.00333 | 0.00106 | 0.00087 | 0.00092 |

### 6.4.2 Expected length

Turning to the length of the intervals, the length of one-tailed intervals varies too much as the value of $p$ is changed. Thus, we will restrict our attention to two-tail intervals.

In Figure 6.11 , the expected length of 95% two-tailed intervals are plotted against $p$ for $r = 10, 30$ and 50 for the OLC, Jeffreys, score and Wald (left-

hand panels) and the OLC, exact and mid-$p$ methods (right-hand panels). It is clear that the expected lengths of the OLC, mid-$p$ and Jeffreys methods are all very similar, and a little smaller than the expected length of the exact method. Exact intervals have a bigger expected length than other methods whereas the expected length of the Wald method is much smaller than the expected length of the score method and other methods when $p$ is quite large or quite small.

The average expected length (AEL) for each method is calculated from its expected length using equation (6.11). Table 6.3 gives the average expected length (AEL) of all mentioned methods for each combination of $\alpha = 0.05, 0.025, 0.005$ and $r = 10, 30$ and 50. The OLC method has a longer AEL than some of the other methods, such as Jeffreys and Wald, but these methods do not give LCC intervals. Although the Wald method has the smallest AEL for each combination, it sacrifices its coverage and is very liberal. The OLC method always gives intervals with a shorter AEL than the AEL of the exact, score and mid-$p$ methods. It is noticeable that the AEL of these methods is decreased as the number of sucesses $r$ is increased.

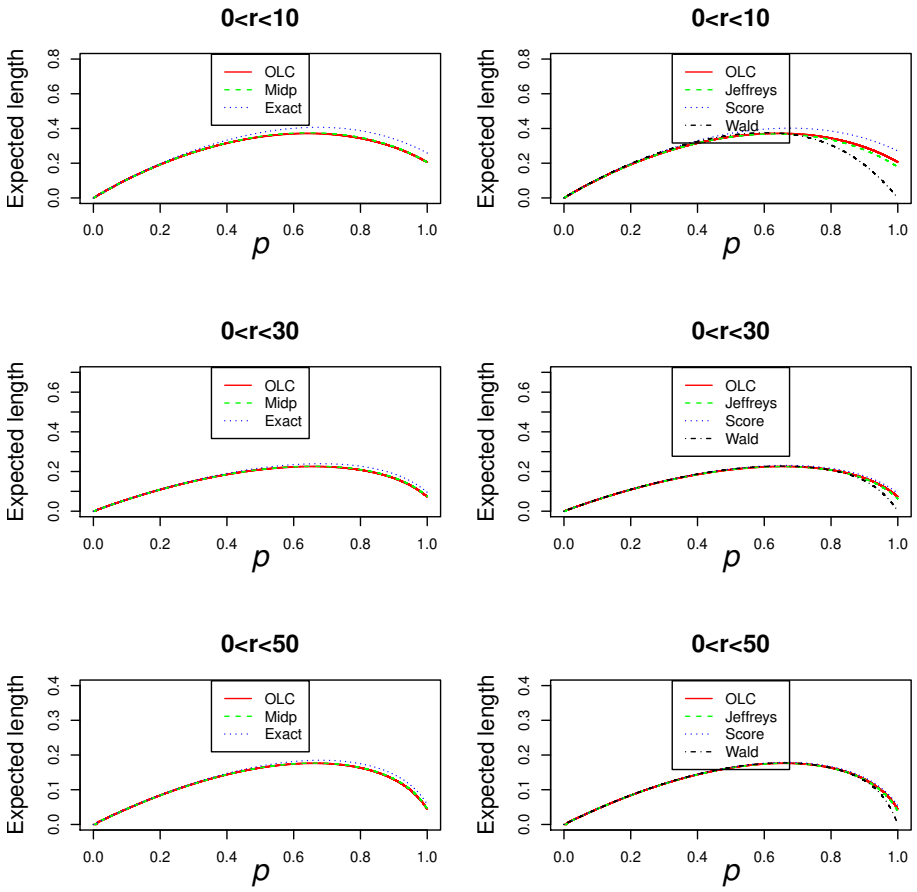


Figure 6.11: Expected length of two-sided 95% interval estimates for the OLC, Jeffreys, score and Wald (left-hand panels) and the OLC, exact and mid-$p$ methods (right-hand panels) plotted against $p$ for $r = 10, 30$ and 50.

Table 6.5: Average expected length (AEL) of two-tail $1 - 2\alpha$ intervals for six methods of forming confidence interval estimates, for $\alpha = 0.05$, 0.025, 0.005 and $r = 10, 30$ and 50.

| $\alpha$ | $r$ | Exact | Wald | Score | Jeffreys | Mid-$p$ | OLC |
|---|---|---|---|---|---|---|---|
| 0.05 | 10 | 0.296674 | 0.252030 | 0.295737 | 0.267711 | 0.274075 | 0.272390 |
| 0.05 | 30 | 0.168441 | 0.155617 | 0.163425 | 0.157826 | 0.159477 | 0.158742 |
| 0.05 | 50 | 0.129374 | 0.122044 | 0.125535 | 0.122888 | 0.123719 | 0.123306 |
| 0.025 | 10 | 0.344405 | 0.296946 | 0.364656 | 0.316708 | 0.323758 | 0.321682 |
| 0.025 | 30 | 0.198082 | 0.184711 | 0.197303 | 0.187614 | 0.189515 | 0.188643 |
| 0.025 | 50 | 0.152652 | 0.145122 | 0.150820 | 0.146218 | 0.147185 | 0.146695 |
| 0.005 | 10 | 0.432957 | 0.379669 | 0.513717 | 0.408295 | 0.415893 | 0.413390 |
| 0.005 | 30 | 0.255137 | 0.240274 | 0.266858 | 0.245043 | 0.247324 | 0.246241 |
| 0.005 | 50 | 0.197749 | 0.189633 | 0.201908 | 0.191452 | 0.192642 | 0.192025 |

## 6.5 Concluding Comments

This chapter aimed to extend application of the OLC method to the negative binomial distribution and examine wheter it gives reasonable confidence inter-

vals. Unlike other chapters, the OLC method was applied to two quantities, the negative binomial proportion $p$, and the negative binomial mean $\mu$. For both $p$ and $\mu$, the OLC method gave interval end-points that are reasonably spaced with coverage that is balanced around the nominal confidence level. In contrast to the binomial and Poisson distributions, the coverage probability of the confidence intervals of $p$ is monotone decreasing in $x$. But the coverage probability of the confidence intervals of $\mu$ is monotone increasing in $x$, exactly as with the two previous distributions. For the case of the negative binomial proportion $p$, applying the new definition of an interval estimate yields the optimal method, OLC, which gives an interval whose average expected width over $p \in [0, 1]$ is a local minimum. Also, for the negative binomial mean $\mu$, applying the same definition and including a weighting function for the length (because of the infinite nature of the mean $\mu$) resulted in an interval whose weighted average expected width over $\mu \in [0, \infty]$ is a local minimum.

There are other properties that are desirable in an interval estimator and these were examined for $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$. For both $p$ and $\mu$ we can say that the OLC method has these properties: interval-valued, monotonicity and nesting. Hence, we can conclude that the OLC method gives sensible intervals with an acceptably short average expected length.

As for the other methods of constructing a confidence interval that were examined, for the case of $p$, there are some similarities and differences in the results for the negative binomial and those for the binomial and Poisson distributions. Casella and McCulloch's method (the exact method for the negative binomial distribution) performed better in terms of average and expected length comparisons than its counterparts for the binomial and Poisson distributions, the Clopper-Pearson and Garwood methods. However, Casella and McCulloch's method is not mentioned in the literature as a gold-standard method. The

mid-$p$ method's performance was similar for the negative binomial distribution as for other distributions. It had coverage that is just a little conservative, and its RMSE is the smallest among all methods. Jeffreys method also performed similarly for the negative binomial distribution as for other distributions. It gave intervals with coverages that are far below the nominal level for some values of the parameter of interest and close to the nominal level for other values of the parameter. Wald's method again has the shortest average length but poor coverage performance, being very conservative for upper tail intervals and very liberal for lower-tail intervals. For the negative binomial mean, $\mu$, we expect that the performance of the OLC method compared to the other discussed methods will be the same as in the case of the proportion $p$. For the negative binomial mean, $\mu$, simulations were run to compare the performance of the OLC method with other methods. Results were very similar to those reported for the negative binomial proportion, $p$.

# Chapter 7

# Concluding comments and directions for future research

This chapter summarizes the main results and conclusions of the thesis. We give the main conclusions of the thesis in brief points. Then, we discuss each point beginning with the strong and weak sides of each method of forming the confidence interval for each distribution. We also discuss the purpose of proposing a new definition and the new method (OLC) briefly. Finally, some extensions for further future research are given.

## 7.1 Conclusions

From all previous chapters, we can obtain the following conclusions:

1. Obtaining a confidence interval when sampling from a discrete distribution faces special difficulties, as the discrete nature of the sample space leads to sharp fluctuations in the coverage probability, which we call spikes.

2. A number of methods have been proposed for forming a confidence interval for the parameters of interest for the binomial, Poisson and negative binomial distributions. Most of the methods did not give intervals that met the definition of a confidence interval, that the coverage probability of the interval $[l_x, u_x] \geq (1 - 2\alpha)$. Methods that meet the definition have the disadvantages of conservative coverage and poor width.

3. The Clopper-Pearson, Garwood, and Casella and McCulloch methods for the binomial, Poisson and negative binomial distributions, respectively, satisfy the definition of a confidence interval. They give strictly correct confidence intervals and are referred to as " correct" methods. The Garwood and Clopper-Pearson methods are the gold standard and give good coverage but poor width. The Casella and McCulloch method has not been referred to as a gold standard, but it did better in term of expected length in comparison to the Garwood and Clopper-Pearson methods.

4. The Mid-$p$ is a good method. Its coverage is generally a little conservative

and it has the smallest RMSE of any method.

5. Jeffreys method has a performance that is similar across the three distributions. Its coverage is often far below the nominal level.

6. The Wald and Wilson (score) methods are used with the three distributions. These methods and the Agersti-Coull methods, which is only used with the binomial distribution, do not give intervals that meet the definition of a confidence interval. They gave similar results except that the score method performs much better for the binomial distribution than the other two sampling distributions.

7. When the sampling space is discrete, the definition of a confidence interval does not meet our needs. A new definition was proposed which states that the average coverage between any pair of consecutive spikes must be equal or greater than the nominal level $(1 - \alpha)$. This led naturally to a new method, the OLC method, which constructs locally correct confidence intervals.

8. Our research began by applying the OLC method to the most common discrete distribution in the literature, the binomial distribution. It is a simple and basic case for forming optimal locally correct confidence intervals. The binomial distribution has two parameters $(p, n)$, where $p \in [0, 1]$ and $n$ is a fixed known number.

9. Applying the OLC method to the Poisson distribution differs from its application to the binomial distribution. It faces the challenge of an unbounded range for the Poisson mean $\lambda$, which is $(0, \infty)$ and the values of $x$ go to infinity. So, the OLC method was modified so that it could be used for constructing a confidence interval for the Poisson mean, and any other parameter that has infinite range.

10. Applying the OLC method to the negative binomial distribution has the same challenge as with the Poisson distribution, as the values of $x$ have an

unlimited range. However, it is similar to the binomial distribution in that it has two parameters, $(p, r)$, where $0 < p < 1$ and $r = 1, 2, \ldots$. A difference is that we are interested in forming a confidence interval for both the negative binomial proportion $p$ and the negative binomial mean $\mu$.

11. For three distributions the OLC method yields locally correct intervals of a minimum average expected length.

12. For $\lambda$ and $\mu$ it was not possible to measure the performance of the intervals over the full range, from 0 to $\infty$, because the expected length is infinite. Instead, we determine a weighting function to calculate a finite weighted average expected length. In the case of $\lambda$, the weighting function is set equal to a gamma $(\alpha, \beta)$ distribution. Whereas, in the case of $\mu$, the weighting function is set equal to a beta prime $(a, b)$ distribution, with pdf as in equation (6.22).

## 7.2 Discussion

This thesis began by searching for a satisfactory definition for a new interval estimator for a binomial proportion. We went through different restrictions in this search. First, there was the overall coverage restriction, which means that the average coverage for the intervals is at least $(1 - \alpha)$ over $p \in [0, 1]$. Second, we restricted the average coverage in fixed intervals in two cases. In case (a) we did not put any restrictions on the number of spikes in each subinterval. In case (b) we imposed the restriction that there is at most one spike in any subinterval. However, these restrictions failed to fill our needs. So, the restriction which we proposed is that the average coverage between consecutive spikes should equal or exceed the nominal confidence level $(1 - \alpha)$. We examined whether exisiting methods of forming confidence intervals met this new definition and propsed the OLC method. Three of these methods, Clopper-Pearson, mid-$p$ and the OLC method met the definition for the values

of $\alpha$ of interest in practice. Although the Clopper-Pearson method satisfies the definition and gives locally correct intervals, it gives intervals with an average expected length that is a little large. That is because its conservative coverage affects the length of its interval, making them longer than intervals given by other methods. The mid-$p$ and OLC methods both have a coverage that is close to the nominal level for any value of $p$. But from Table 3.3 and Figure 3.8, the OLC method has intervals with an average expected length that is shorter than with the mid-$p$.

All other methods, Wald, Wilson, Agresti-Coull and Jeffreys, are recommended in the literature as alternative methods to the gold standard method because the length of their intervals is a little bit shorter even if they do not meet our new definition. However, for all examined values of $n$ and $\alpha$, the OLC method gives the shortest average expected length compared to any method achieving our new definition. That is because the Wald method gives intervals whose coverage is often much less than the nominal level. Moreover, by examination for any combination of $n$ and $\alpha$ as $1 \leq n \leq 200$ and $\alpha \in \{0.001(0.001)0.27\}$, the OLC method has the properties, such as equivariance and monotonicity, that are listed by Blyth and Still (1983) and Schilling and Doi (2014) as desirable properties for a good interval estimator. Also the OLC method satisfies the new definition and gives reasonable intervals with an average expected length that is acceptably short. Hence, we can say that the OLC method is an alternative interval estimator for the binomial proportion. Then, we applied the new definition and extended the proposed method for the Poisson mean, giving a modified OLC method. Although constructing a confidence interval for the Poisson mean faces the challenge of an infinite range for the parameter $\lambda$, and $x$ values can also be unbounded, there are similarities in the results of the binomial and Poisson distributions. The modified OLC

method gave intervals whose coverage is fairly spread around the nominal confidence level. Also for all examined values of $x$ and $\alpha$, where $1 \leq x \leq 200$ and $\alpha \in \{0.001(0.001)0.3\}$, the modified OLC method has the shortest average expected length comparable to any method achieving our new definition. Also, for all examined values of $x$ and $\alpha$, the modified OLC method has the other desirable properties which are mentioned in Section 4.5, such as interval valued and nesting listed by Blyth and Still (1983) and Schilling and Doi (2014). Regarding alternative methods of forming confidence intervals, the Garwood method performs like the Clopper-Pearson method, its counterpart for the binomial distribution, as it has good coverage but with poor length. Mid-$p$ obviously has the smallest RMSE among all methods in the Poisson case, as it has the smallest RMSE for any discrete sampling distribution. However, it is a little conservative in its coverage. The performance of Jeffreys method is also the same for both distributions, with coverage that is sometimes far below the nominal confidence level. The score method performs better for the binomial distribution than for the Poisson distribution.

Our new OLC method is also applied to a third distribution, the negative binomial distribution. For this distribution, we applied the new definition and used the OLC method to construct confidence intervals for its proportion $p$ and its mean $\mu$. The OLC method gave sensible locally correct confidence intervals with an average expected length over $p \in [0, 1]$ that is acceptably short. Unlike the case with the binomial and Poisson distributions, the endpoints of the confidence interval for the negative binomial proportion $p$ are monotone decreasing in $x$. The OLC method has the other desirable properties, such as monotonicity and nesting.

Comparing the results of the negative binomial case with those of other methods of forming confidence intervals, we find that the exact method (the Casella and McCulloch method) did better in terms of average expected length than both the Clopper-Pearson and Garwood methods. While both the mid-$p$ and Jeffreys methods gave similar performance to their performance with the two other distributions. The score method performed better in the case of the binomial than Poisson and negative binomial distributions.

Constructing a confidence intervals for the negative binomial mean $\mu$ is tricky as $\mu$ has an infinite range. Unlike the case of $p$, the coverage probability of the confidence interval of $\mu$ is monotone increasing in $x$, the same as for the binomial and Poisson distributions. Because of the unbounded range of $\mu$, we used a weighting function to weight the average expected length for the intervals of $\mu$. We could not prove our definition for all values of $\alpha$, $n$ and $x$. That is because the average coverage between some spikes becomes below the nominal level for the biggest values of $\alpha$. In the case of the binomial distribution the range of values for which the OLC method can be used is $\alpha \in \{0.001(0.001)0.27\}$, and for the Poisson and negative binomial it is $\alpha \in \{0.001(0.001)0.3\}$. The value of $x$ in the binomial case is limited by the value of trials, $n$. We considered values of $n$ up to $n = 200$, which is commonly the highest value considered in the literature. In contrast, the value of $x$, in Poisson and negative binomial distributions, does not have a limited range. In this thesis it was assumed that observed sample value of $x$ is less than 200. Also, in the last two distributions, we needed to determine the value of $N$ which is very useful and important in calculating the endpoints of the intervals. This value is determined to be $N = 500$ for $0 \leq x \leq 100$, while for $x > 100$, $N = 5x$ is enough to represent the infinite range. The reason for choosing these values of $N$ is that increasing the value of $N$ above these values makes to almost

no difference to the confidence interval for $X = x$. Because of the infinite range of both $\lambda$ and $\mu$, we could not prove the definition for all their values. So, the definition and methods are examined over the range $\lambda \in (0, 50)$ and $\mu \in (0, 50)$, where these values have attracted most attention in the practical research. However, we would like to evaluate the performance of the methods, especially the average expected length of the intervals of each method, over the full range of $\lambda$ and $\mu$. So, we had to use a weighting function, which is mentioned above in point 13, to weight the average expected length. The tricky point in determining the weighting function is choosing the parameter values $(\alpha, \beta)$ of gamma$(\alpha, \beta)$ and $(a, b)$ of beta prime$(a, b)$ distribution. We tried to find the weighting function, that gives almost equal weight to each interval. The values of the parameters that achieve our purpose are $\alpha = 1$ and $\beta = 1/4$ for the gamma distribution in the case of $\lambda$ and are $a = 2$ and $b = 3$ for the beta prime distribution in the case of $\mu$.

## 7.3 Future work

Future research in forming confidence intervals for the discrete distributions may include the following points:

- More than one parameter (e.g difference between two proportions, risk ratio and odds ratio). These three measurements are used commonly in medical research to compare two treatments for a disease, for example.
  - Difference ($\delta = p_1 - p_2$) is perhaps the most direct methods of comparison between the two event probabilities. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. Many methods have been devised for computing confidence intervals for the difference between two proportions.

These include the following seven methods: Score (Farrington and Manning), Score (Miettinen and Nurminen), Score with correction for Skewness (Gat and Nam), Score (Wilson), Score with continuity correction (Wilson), Chi-Square with continuity correction (Yates) and Chi-Square (Pearson)

- Ratio ($\phi = p_1/p_2$) gives the relative change in the disease risk due to the application of the treatment. It is also direct and easy to interpret. Many methods have been devised for computing confidence intervals for the ratio of two proportions. Amongst them are the following six methods: Score (Farrington and Manning), Score (Miettinen and Nurminen), Score with correction for Skewness (Gat and Nam), Logarithm (Katz), Logarithm + 1/2 (Walter) and Fleiss.

- Odds Ratio [$\psi = (p_1/q_1)/(p_2/q_2) = (p_1q_2)/(p_2q_1)$] is a relative measure for comparing outcomes. It has a direct relationship with the regression coefficient in logistic regression. Although the odds ratio is more complicated to interpret than the ratio, it is often the parameter of choice. Many methods have been devised for computing confidence intervals for the odds ratio of two proportions. Eight of these methods are: Exact (Conditional), Score (Farrington and Manning), Score (Miettinen and Nurminen), Fleiss, Logarithm, Mantel-Haenszel, Simple and Simple + 1/2 (Agresti, 2003), (Fagerland *et al.*, 2015) and (Wang and Shan, 2015).

# Appendices

All the proofs of results are the work of my main supervisor, Prof Paul Garthwaite, the work is included in my thesis as it has not been published elsewhere.

**Appendix (A)**

**Conditions for our algorithm to give locally correct confidence interval with the shortest average length (equations (3.16) and (3.17))**

We observe the value of $X$ where $X \sim \text{bin}(n, p)$ and aim to form an upper one-tailed confidence interval for $p$ with nominal confidence level $\gamma = 1 - \alpha$. Let $(0, u_i^*)$ be the confidence interval given by our algorithm when $X = i$. As $i$ is an integer between 0 and $n$, our algorithm gives the partition $(u_0^*, \ldots, u_n^*)$ where

$$0 < u_0^* \leq \ldots \leq u_{i-1}^* \leq u_i^* \leq \ldots \leq u_n^* = 1. \tag{7.1}$$

If $p \leq u_i^*$, then $p$ is contained in the confidence interval if $X \geq i$. Let $g_n(i, p) = \Pr(X \geq i)$ when $X \sim \text{bin}(n, p)$. From our definition of a locally correct confidence interval,

$$\frac{1}{u_i^* - u_{i-1}^*} \int_{u_{i-1}^*}^{u_i^*} g_n(i, p)\, dp \geq \gamma \tag{7.2}$$

for $i = 1, \ldots, n$.

The length of our confidence interval when $X = i$ is $u_i^*$ so, given $p$, the expected length of our confidence interval is

$$\sum_{i=0}^{n} u_i^* \Pr(X = i \,|\, p). \tag{7.3}$$

When $p$ is equally likely to take any value in the interval (0, 1), $\Pr(X = i) = 1/(n+1)$ for $i = 0, \ldots, n$. From equation (7.3), the average length (when each value of $p$ is equally likely) is thus $\sum_{i=0}^{n} u_i^*/(n+1)$. We aim to show that our algorithm gives a shorter average length than any other partition that yields a set of locally correct confidence intervals.

To this end, suppose the partition $(b_0, \ldots, b_n)$ is a partition that has the minimum average length while satisfying

$$\frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} g_n(i, p)\, dp \geq \gamma \tag{7.4}$$

for $i = 1, \ldots, n$, with

$$0 < b_0 \leq \ldots \leq b_i \leq b_{i+1} \leq \ldots \leq b_n = 1. \tag{7.5}$$

We must show that

$$b_i = u_i^* \quad \text{for } i = 0, \ldots, n. \tag{7.6}$$

Further conditions are needed for equation (7.6) to necessarily hold. These conditions only involve the partition given by our algorithm and it is feasible to examine them under ranges of values for $n$ and $\alpha$ that effectively cover all integer values of $n$ and $0.00001 \leq \alpha \leq 0.27$. We prove the main result in this section (Theorem 1) by induction, showing that if $b_i = u_i^*$ for $i = n, n-1, \ldots, k+1$ $(k \geq 0)$, then it also holds for $i = k$. To start the induction, note that $1 = b_n = u_n^*$. We first prove some preparatory results that place bounds on the $b_i$.

Lemma 1. If $b_{k+1} = u_{k+1}^*$ then ( i) $b_k \geq u_k^*$ for any $k = 0, \ldots, n-1$,; (ii) $b_{k-1} \leq u_{k-1}^*$ if $k \geq 1$ ; (iii) $b_0 = u_0^*$ if $b_1 = u_1^*$; and (iv) $u_{k-1}^* + u_k^* \geq b_{k-1} + b_k$ if $k \geq 1$.

Proof. As $g_n(k+1, p)$ is a monotonic strictly decreasing function of $p$, $\int_c^{u_{k+1}^*} g_n(k+1, p)\, dp/(u_{k+1}^* - c) < \int_{u_k^*}^{u_{k+1}^*} g_n(k+1, p)\, dp/(u_{k+1}^* - u_k^*)$ for any $c < u_k^*$. Given $u_{k+1}^*$, our algorithm chooses $u_k^*$ as the value for which $\int_{u_k^*}^{u_{k+1}^*} g_n(k+1, p)\, dp/(u_{k+1}^* - u_k^*) = \gamma$. Thus $\int_c^{u_{k+1}^*} g_n(k+1, p)\, dp/(u_{k+1}^* - c) < \gamma$ for any $c < u_k^*$. By assumption, $b_{k+1} = u_{k+1}^*$, so setting $i = k+1$ in (7.4) gives result (i), that $b_k \geq u_k^*$. For (ii), note that $(b_0, \ldots, b_{k-1}, u_k^*, \ldots, u_n^*)$ would be a partition that gave locally correct confidence intervals if $b_{k-1} \geq u_{k-1}^*$.

Since $\sum_{i=0}^{k-1} b_i + \sum_{i=k}^{n} u_i^* = \sum_{i=0}^{n} b_i + (u_k^* - b_k)$ and (from (i)) $b_k \geq u_k^*$, this would contradict the assumption that $(b_0, \ldots, b_n)$ is the partition of minimum average length that satisfies (7.4), unless $b_{k-1} = u_{k-1}^*$. Consequently, $b_{k-1} \leq u_{k-1}^*$ if $b_{k+1} = u_{k+1}^*$. For (iii), suppose that $b_1 = u_1^*$. Then the partition $(u_0^*, b_1, b_2, \ldots, b_n)$ gives locally correct confidence intervals. As $(b_0, \ldots, b_n)$ is the partition that gives locally correct confidence intervals with the shortest average length, it follows that $b_0 \leq u_0^*$ so, from (i), $b_0 = u_0^*$. For (iv), from (ii) we have that $b_k - 1 \leq a_{k-1}$, so the average coverage for $p \in (b_{k-2},\, u_{k-1}^*)$ is not less than the average coverage for $p \in (b_{k-2},\, b_{k-1})$. Hence, as $(b_0, \ldots, b_n)$ is a partition that gives locally correct confidence intervals, so does the partition $(b_0, \ldots, b_{k-2}, u_{k-1}^*, u_k^*, b_{k+1}, \ldots, b_n)$. Now $(b_0, \ldots, b_n)$ is the partition that gives the *shortest* average length of locally correct confidence intervals, so we have that $b_{k-1} + b_k \leq u_{k-1}^* + u_k^*$. $\diamond$

To place a further bound on $b_{k-1}$, let $p_{k-1}^{\#}$ be the value of $p$ for which $g_n(k-1, p) = \gamma$. Analogous to equation (7.2) we have

$$\frac{1}{b_{k-1} - b_{k-2}} \int_{b_{k-2}}^{b_{k-1}} g_n(k-1, p)\, dp \geq \gamma, \tag{7.7}$$

so $b_{k-1} \geq p_{k-1}^{\#}$, since $g_n(k-1, p)$ is a monotonically increasing function of $p$. In combination with Lemma 1, this gives the following result.

Lemma 2. If $b_{k+1} = u_{k+1}^*$ and $k \geq 1$, then

$$u_k^* \leq b_k \leq \min(u_{k+1}^*, u_{k-1}^* + u_k^* - p_{k-1}^{\#}) \tag{7.8}$$

where $p_{k-1}^{\#}$ is given by $g_n(k-1, p_{k-1}^{\#}) = \gamma$.

Proof. For the lower limit, $u_k^* \leq b_k$ from part (i) of Lemma 1. For the upper limit, first $b_k \leq b_{k+1}$ so $b_k \leq u_{k+1}^*$. Also, from part (iv) of Lemma 1, $b_k \leq u_{k-1}^* + u_k^* - b_{k-1}$, so $b_k \leq u_{k-1}^* + u_k^* - p_{k-1}^{\#}$. $\diamond$

We will require a quantity $A_n(i, p^*, y)$ that is defined by:

$$A_n(i, p^*, y) = \frac{1}{2y} \int_{p^*-y}^{p^*+y} g_n(i, p)\, dp \qquad \text{for } 0 \le y \le \min(p^*,\, 1 - p^*). \tag{7.9}$$

Thus $A_n(i, p^*, y)$ is the average coverage given by $g_n(i, p)$ in the interval $(p^* - y,\, p^* + y)$. Its differential with respect to $y$ is denoted as $A'_n(i, p^*, y)$. The following lemma gives an important characteristic of $A_n(i, p^*, y)$ that underlies conditions for our algorithm to yield locally correct confidence intervals with minimum average length.

Lemma 3. Suppose that $0 < y_1 < y_2 \le \min(p^*,\, 1-p^*)$ and that $A_n(i, p^*, y_1) > A_n(i, p^*, y_2)$. Then,

$$A_n(i, p^*, y_1) > A_n(i, p^*, y) \qquad \text{for any } y \in (y_1,\, y_2] \tag{7.10}$$

if $A'_n(i, p^*, y_1) < 0$.

Proof. Differentiation of $A_n(i, p^*, y)$ with respect to $y$ gives

$$A'_n(i, p^*, y) = -\frac{1}{2y^2} \int_{p^*-y}^{p^*+y} g_n(i, p)\, dp + \frac{1}{2y}\{g_n(i, p^*+y) + g_n(i, p^*-y)\}. \tag{7.11}$$

For sufficiently small $y$, the smooth function $g_n(i, p)$ is approximately linear for $p$ in the interval $(p^* - y,\, p^* + y)$, so $A_n(i, p^*, y)$ does not vary as $y \to 0$, giving $A'_n(i, p^*, 0) = 0$.

Let $Q_n(i, p^*, y) = 2y^2\, A'_n(i, p^*, y)$. Then the differential of $Q_n(i, p^*, y)$ with respect to $y$ is:

$$\begin{aligned} Q'_n(i, p^*, y) &= -\{g_n(i, p^* + y) + g_n(i, p^* - y)\} + \{g_n(i, p^* + y) + g_n(i, p^* - y)\} \\ &\quad + y\,\{g'_n(i, p^* + y) - g'_n(i, p^* - y)\} \\ &= y\,\{g'_n(i, p^* + y) - g'_n(i, p^* - y)\}. \end{aligned} \tag{7.12}$$

Put $h(y) = g'_n(i, p^* + y)/g'_n(i, p^* - y)$. The differential of $h(y)$ with respect to $y$ is:

$$h'(y) = \psi(y)[c_1 - c_2 y^2], \tag{7.13}$$

where

$$\psi(y) = 2(p^* + y)^{i-2}(1 - p^* - y)^{n-i-1}(p^* - y)^{-i}(1 - p^* + y)^{-n+i-1}.$$

$$c_1 = p^*(1 - p^*)\{i - 1 - p^*(n - 1)\},$$

and

$$c_2 = p^*(n - 1) - n + i.$$

The behavior of $A'_n(i, p^*, y)$ depends upon the signs of $c_1$ and $c_2$.

Case 1: $c_1 \leq 0$; $c_2 > 0$ or $c_1 < 0$ ; $c_2 = 0$.

As $\psi > 0$ for all $y$, the following results follow sequentially.

(i) $h'(y) < 0$ for any $y$, so $h(y)$ is a monotonic decreasing function of $y$.

(ii) As $h(0) = 1$, it follows that $h(y) < 1$ for all $y > 0$, so $g'_n(i, p^* + y) - g'_n(i, p^* - y)$ is negative for any $y > 0$.

(iii) $Q'_n(i, p^*, y)$ is negative for $y > 0$, so $Q_n(i, p^*, y)$ is a monotonic decreasing function of $y$.

(iv) As $Q_n(i, p^*, 0) = 0$, it follows that $Q_n(i, p^*, 0)$ is negative for any $y > 0$. Thus $A'_n(i, p^*, y)$ is also negative for any $y > 0$.

(v) $A_n(i, p^*, y)$ is a monotonic decreasing function of $y$, so $A_n(i, p^*, y_1) > A_n(i, p^*, y)$ for any $y > y_1$. Thus Lemma 3 holds if $c_1 \leq 0$ and $c_2 > 0$, or if $c_1 < 0$ and $c_2 = 0$.

Case 2: $c_1 \geq 0$; $c_2 \leq 0$.

Now $h'(y) \geq 0$ for any $y$ and reasoning similar to that for Case 1 shows that $A_n(i, p^*, y_1) \leq A_n(i, p^*, y)$ for any $y > y_1$. Thus the conditions required by Lemma 3 do not hold, so there is nothing to verify.

Case 3: $c_1 < 0$; $c_2 < 0$.

The following results hold.

(i) $h'(y) < 0$ for $y < (c_1/c_2)^{1/2}$ and $h'(y) > 0$ for $y > (c_1/c_2)^{1/2}$, so $h(y)$ is a $\cup$-shaped function of $y$.

(ii) Suppose $h(y)$ is below 1 until $y = y^*$. As $h(0) = 1$, it follows that $g'_n(i, p^* + y) - g'_n(i, p^* - y)$ is negative for $0 < y < y^*$ and positive for $y > y^*$.

(iii) $Q'_n(i, p^*, y)$ is negative for $0 < y < y^*$ and positive for $y > y^*$, so $Q_n(i, p^*, y)$ is a $\cup$-shaped function of $y$.

(iv) Suppose $Q_n(i, p^*, y)$ is below 0 until $y = y^{\#}$. As $Q_n(i, p^*, 0) = 0$ and $Q_n(i, p^*, y) = 2y^2 A'_n(i, p^*, y)$, it follows that both $Q_n(i, p^*, y)$ and $A'_n(i, p^*, y)$ are negative for $0 < y < y^{\#}$ and positive for $y > y^{\#}$. Hence, $A_n(i, p^*, y)$ is a $\cup$-shaped function of $y$.

(v) If $y_2 > y_1$ and $A_n(i, p^*, y_1) > A_n(i, p^*, y_2)$, then $A_n(i, p^*, y_1) > A_n(i, p^*, y)$ for any $y \in (y_1, y_2)$. Thus Lemma 2 holds if $c_1 < 0$ and $c_2 < 0$.

Notes. In (i) it is assumed that $(c_1/c_2)^{1/2} \leq \min(p^*, 1 - p^*)$. If $(c_1/c_2)^{1/2} > \min(p^*, 1 - p^*)$, then $h'(y) < 0$ for all feasible values of $y$ and Case 1 applies. Similarly, in (ii), if $h(y)$ never reaches 1, then $Q'_n(i, p^*, y)$ is always negative and Case 1 again applies (c.f. Case 1, parts (iii)-(v)). Likewise, in (iv), if $Q_n(i, p^*, y)$ is always below 0 for any feasible values of $y$, then Case 1 applies (parts (iv) and (v)).

Case 4: $c_1 > 0$; $c_2 > 0$.

The following results hold.

(i) $h'(y) > 0$ for $y < (c_1/c_2)^{1/2}$ and $h'(y) < 0$ for $y > (c_1/c_2)^{1/2}$, so $h(y)$ is a $\cap$-shaped function of $y$.

(ii) Suppose $h(y)$ drops below 1 at $y = y^*$. As $h(0) = 1$, it follows that $g'_n(i, p^* + y) - g'_n(i, p^* - y)$ is positive for $0 < y < y^*$ and negative for $y > y^*$.

(iii) $Q'_n(i, p^*, y)$ is positive for $0 < y < y^*$ and negative for $y > y^*$, so $Q_n(i, p^*, y)$ is a $\cap$-shaped function of $y$.

(iv) Suppose $Q_n(i, p^*, y)$ drops below 0 at $y = y^{\#}$. As $Q_n(i, p^*, 0) = 0$, it follows that both $Q_n(i, p^*, y)$ and $A'_n(i, p^*, y)$ are positive for $0 < y < y^{\#}$ and negative for $y > y^{\#}$. Thus, $A_n(i, p^*, y)$ is a $\cap$-shaped function of $y$.

(v) If $A'_n(i, p^*, y_1) < 0$, then $A_n(i, p^*, y_1) > A_n(i, p^*, y)$ for any $y > y_1$. Thus Lemma 2 holds if $c_1 > 0$ and $c_2 > 0$.

Notes. In (i) it is assumed that $(c_1/c_2)^{1/2} \leq \min(p^*, 1 - p^*)$. If $(c_1/c_2)^{1/2} > \min(p^*, 1 - p^*)$, then $h'(y) > 0$ for all feasible values of $y$ and Case 2 applies. Similarly, Case 2 applies if $h(y)$ never drops below 1 in (ii) for any feasible value of $y$, or if $Q_n(i, p^*, y)$ never drops below 0 in (iii).

This completes the proof of the lemma, as Cases 1 - 4 cover all combinations of $c_1$ and $c_2$. $\diamond$

Theorem 1. For $i = 1, \ldots, n-1$, let $p_i^* = (u_i^* + u_{i-1}^*)/2$ and let $\eta_i = u_i^* - p_i^*$. Also, define $p_{i-1}^{\#}$ by $g_n(i-1, p_{i-1}^{\#}) = \gamma$ and let $\xi_i = \min(u_{i+1}^*, u_{i-1}^* + u_i^* - p_{i-1}^{\#}) - p_i^*$. Suppose

$$A_n(i, p_i^*, \eta_i) > A_n(i, p_i^*, \xi_i) \tag{7.14}$$

and

$$A'_n(i, p_i^*, \eta_i) < 0 \tag{7.15}$$

both hold for $i = 1, \ldots, n-1$. Then $b_i = u_i^*$ for $i = 0, \ldots, n$ and $(u_0^*, \ldots, u_n^*)$ is the partition that yields locally correct confidence intervals of shortest average length.

Proof. We have that $b_n = u_n^*$ and, from Lemma 1, $b_0 = u_0^*$ if $b_1 = u_1^*$. Hence we must show that $b_i = u_i^*$ for $i = 1, \ldots, n-1$. We will prove the result by induction: we assume that $b_i = u_i^*$ for $i = k+1, \ldots, n$ and will show this implies that $b_k = u_k^*$.

By assumption, $A_n(k, p_k^*, \eta_k) > A_n(k, p_k^*, \xi_k)$ and $A_n'(k, p_k^*, \eta_k) < 0$ so, from Lemma 3, $A_n(k, p_k^*, \eta_k) > A_n(k, p_k^*, y)$ for any $y \in (\eta_k, \xi_k]$. Also, $A_n(k, p_k^*, \eta_k) = (2\eta_k)^{-1} \int_{p_k^*-\eta_k}^{p_k^*+\eta_k} g_n(k, p)\, dp = (2\eta_k)^{-1} \int_{u_{k-1}^*}^{u_k^*} g_n(k, p)\, dp = \gamma$. Consequently,

$$\gamma > A_n(k, p_k^*, y) \qquad \text{for any } y \in (\eta_k, \xi_k]. \tag{7.16}$$

Put $b_k = p_k^* + \tau_k$. From part (i) of Lemma 1, $\tau_k \geq 0$, and from part (iv), $b_{k-1} \leq u_{k-1}^* + u_k^* - b_k = 2p_k^* - b_k = p_k^* - \tau_k$. Hence, as $g_n(k, p)$ is a monotonic strictly increasing function of $p$, we have

$$\gamma \leq \frac{1}{b_k - b_{k-1}} \int_{b_{k-1}}^{b_k} g_n(k, p)\, dp \leq \frac{1}{2\tau_k} \int_{p_k^*-\tau_k}^{p_k^*+\tau_k} g_n(k, p)\, dp = A_n(k, p_k^*, \tau_k).$$

From equation (7.16), it follows that $\tau_k$ is not within the interval $(\eta_k, \xi_k]$. However, $u_k^* = p_k^* + \eta_k$ so, from Lemma 2,

$$p_k^* + \eta_k \leq b_k \leq \min(u_{k+1}^*, u_{k-1}^* + u_k^* - p_{k-1}^{\#}) = p_k^* + \xi_k.$$

Thus $\eta_k \leq \tau_k \leq \xi_k$. It follows that $\tau_k = \eta_k$, so $b_k = u_k^*$. $\diamond$

**Appendix (B)**: Proof of proposition 3.4

Suppose $p = p^*$ and $0 \leq u_0 \leq \ldots \leq u_n = 1$. If $1 - \frac{1}{2}\mathrm{P}(X = 0 \,|\, p = p^*) \leq 1 - \alpha$, put $i^* = 0$. Otherwise, define $i^*$ by

$$\mathrm{P}(X \geq i^* \,|\, p = p^*) - \tfrac{1}{2}\mathrm{P}(X = i^* \,|\, p = p^*) \leq 1 - \alpha \tag{7.17}$$

and

$$\mathrm{P}(X \geq i^* - 1 \,|\, p = p^*) - \tfrac{1}{2}\mathrm{P}(X = i^* - 1 \,|\, p = p^*) > 1 - \alpha. \tag{7.18}$$

The coverage of a method of forming confidence intervals must equal $\mathrm{P}(X \geq i \,|\, p = p^*)$ for some $i$. The value midway between $\mathrm{P}(X \geq i - 1)$ and $\mathrm{P}(X \geq i)$ is $\mathrm{P}(X \geq i - 1) - \frac{1}{2}\mathrm{P}(X = i - 1)$. Similarly, $\mathrm{P}(X \geq i) - \frac{1}{2}\mathrm{P}(X = i)$ is midway between $\mathrm{P}(X \geq i)$ and $\mathrm{P}(X \geq i + 1)$. Hence, if $i^* \geq 1$, the feasible coverage that is closest to equalling $1 - \alpha$ is $\mathrm{P}(x \geq i^* \,|\, p = p^*)$. A method of forming confidence intervals achieves this coverage if $u_{i^*-1} < p^* \leq u_{i^*}$. If $i^* = 0$, the coverage closest to $1 - \alpha$ is 1, which is the coverage when $p^* \leq u_0$.

Let $\tilde{u}_i$ denote the upper limit given by the mid-$p$ method when $X = i$. From equation (13) in the paper, $\tilde{u}_i$ satisfies

$$\mathrm{P}(X \geq i \,|\, p = \tilde{u}_i) - \tfrac{1}{2}\mathrm{P}(X = i \,|\, p = \tilde{u}_i) \;=\; 1 - \alpha, \tag{7.19}$$

for $i \leq n - 1$. Suppose $i^* \leq n - 1$. Putting $i = i^*$ in (7.19) and comparison with (7.17) yields

$$\mathrm{P}(X \geq i^* \,|\, p = p^*) - \tfrac{1}{2}\mathrm{P}(X = i^* \,|\, p = p^*) \;\leq\; \mathrm{P}(X \geq i^* \,|\, p = \tilde{u}_{i^*}) - \tfrac{1}{2}\mathrm{P}(X = i^* \,|\, p = \tilde{u}_{i^*}), \tag{7.20}$$

so $p^* \leq \tilde{u}_{i^*}$. Hence the mid-$p$ method has the feasible coverage closest to $1 - \alpha$ when $i^* = 0$. As $\tilde{u}_n = 1$, we also have $p^* \leq \tilde{u}_{i^*}$ for $i^* = n$. When $i^* \geq 1$, putting $i = i^* - 1$ in (7.19) and comparison with (7.18) similarly yields

$$\mathrm{P}(X \geq i^* - 1 \,|\, p = p^*) - \tfrac{1}{2}\mathrm{P}(X = i^* - 1 \,|\, p = p^*)$$

$$> \mathrm{P}(X \geq i^* - 1 \,|\, p = \tilde{u}_{i^*-1}) - \tfrac{1}{2}\mathrm{P}(X = i^* - 1 \,|\, p = \tilde{u}_{i^*-1}), \tag{7.21}$$

so $p^* > \tilde{u}_{i^*-1}$. Thus, when $1 \leq i^* \leq n$, we have that $\tilde{u}_{i^*-1} < p^* \leq \tilde{u}_{i^*}$. Hence, for any $p^*$, the coverage of the mid-$p$ method is as close to $1-\alpha$ as the coverage of any method that meets the conditions of Proposition 3.4. $\diamond$

**Appendix (C)**: Proof of Proposition 4.1

For definiteness, suppose $u_i < u_i^*$. Then from equations (4.1) and (7.2), $u_{i-1} > u_{i-1}^*$. Figure 7.1 is a diagram of the coverage probability $C_{u,i}(\theta)$ plotted against $\theta$ for $u_{i-1}^* < \theta < u_i^*$. As the coverage function is concave, the average coverage over the interval $(u_{i-1}, u_i)$ is greater than $C_{u,i}(u_i) + C_{u,i}(u_{i-1})/2$, so from equation (4.1), $C_{u,i}(u_i) + C_{u,i}(u_{i-1})/2 < 1 - \alpha$. Using the notation in Figure 7.1, it follows that the distance from B to C exceeds the distance from E to H:

$$BC > EH \tag{7.22}$$

As the average coverage over the interval $(u_{i-1}, u_i)$ is $1 - \alpha$, the area with vertexes BIC equals the area with vertexes IEH. Similarly, the average coverage over the interval $(u_{i-1}^*, u_i^*)$ is $1 - \alpha$, so the area with vertexes AID equals the area with vertexes IFG. Consequently, area(AID)-area(BIC) = area(IFG)-area(IEH), so the area with vertexes ABCD equals the area with vertexes EFGH. From equation (7.22), $BC > EH$ and the gradient of the coverage function is greater between $u_{i-1}^*$ and $u_{i-1}$ than between $u_i$ and $u_i^*$. As area($ABCD$) equals area($EFGH$), it follows that the distance from A to B is less than the distance from H to G. That is $|u_{i-1} – u_{i-1}^*| < |u_i^* - u_i|$, as required.

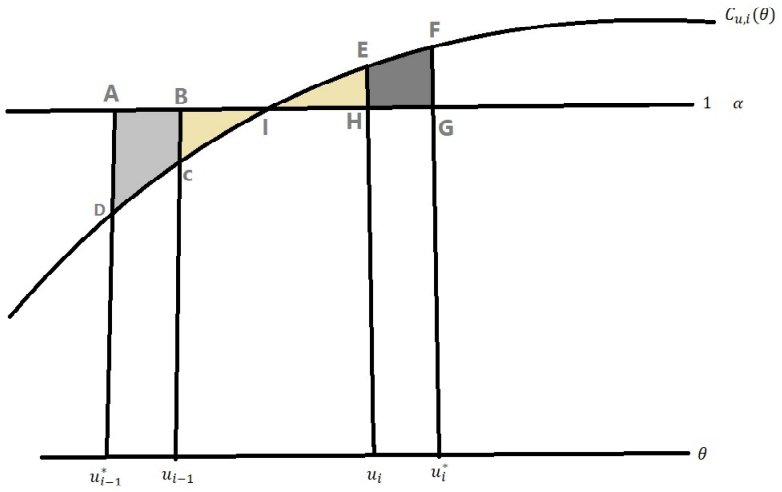


Figure 7.1: Figure for the proof of Proposition 4.1.

**Appendix (D)**: Proof of Proposition 4.2

Let $\gamma = 1 - \alpha$. As $(u_0^*, \ldots, u_n^*)$ gives locally correct confidence intervals

$$\frac{1}{u_i^* - u_{i-1}^*} \int_{\theta=u_{i-1}^*}^{u_i^*} C_{u,i}(\theta)\, d\theta \leq \gamma \tag{7.23}$$

for $i = 1, \ldots, n$. When (7.23) is an equality we say that $u_{i-1}^*$ is a *tight lower limit* – (7.23) would no longer be satisfied if $u_{i-1}^*$ was made smaller (with $u_i^*$ fixed).

Preliminary lemma

Suppose that $\delta$ is small and let $\delta_i = u_i^* - u_i$ for $i = 0, \ldots n$. If $u_{i-1}^*$ is a tight lower limit, then

$$\delta_{i-1} = \delta_i \{C_{u,i}(u_i) - \gamma\} / \{C_{u,i}(u_{i-1}) - \gamma\} + O(\delta^2). \tag{7.24}$$

Proof of lemma. For small $\delta_{i-1}$, $\delta_i$,

$$\begin{aligned} \int_{u_{i-1}^*}^{u_i^*} C_{u,i}(\theta)\, d\theta &= \int_{u_i}^{u_i+\delta_i} C_{u,i}(\theta)\, d\theta + \int_{u_{i-1}}^{u_i} C_{u,i}(\theta)\, d\theta + \int_{u_{i-1}+\delta_{i-1}}^{u_{i-1}} C_{u,i}(\theta)\, d\theta \\ &= \delta_i\, C_{u,i}(u_i) + (u_i - u_{i-1})\gamma - \delta_{i-1}\, C_{u,i}(u_{i-1}) + O(\delta^2). \end{aligned}$$

As $\int_{u_{i-1}^*}^{u_i^*} C_{u,i}(\theta)\, d\theta = (u_i^* - u_{i-1}^*)\,\gamma = \delta_i\,\gamma + (u_i - u_{i-1})\,\gamma - \delta_{i-1}\gamma$, we have

that $\delta_i\, C_{u,i}(u_i) - \delta_{i-1}\, C_{u,i}(u_{i-1}) + O(\delta^2) = \delta_i\, \gamma - \delta_{i-1}\gamma$, and the result follows.

$\diamond$

Proof of part (a) of proposition. Suppose $\phi_l < 0$ for some $l \in (1, \ldots, n-1)$. Set $u_i^* = u_i$ for $i = l+1, \ldots n$. Choose $\epsilon > 0$ such that $|\,\epsilon\psi_{jl}| < \delta$ for $j = 1, \ldots, l$ and put $u_l^* = u_l + \epsilon$. As $\epsilon > 0$, (7.23) holds for $i = l+1$. Sequentially choose $u_{l-1}^*, u_{l-2}^*, \ldots, u_0^*$ such that each is a tight lower limit. That is,

$$\frac{1}{u_i^* - u_{i-1}^*} \int_{u_{i-1}^*}^{u_i^*} C_{u,i}(\theta)\, d\theta = \gamma \tag{7.25}$$

for $i = l, l-1, \ldots, 1$. Then (7.23) holds for $i = 1, \ldots, n$. Also, from the preliminary lemma, $u_{i-1}^* - u_{i-1} = (u_i^* - u_i)h_i/f_i + O(\delta^2)$ for $i = 1, \ldots l$. Consequently,

$$\begin{aligned} u_{i-1}^* - u_{i-1} &= (u_l^* - u_l)h_l h_{l-1} \ldots h_i/(f_l f_{l-1} \ldots f_i) + O(\delta^2) \\ &= (u_l^* - u_l)\psi_{il} + O(\delta^2) \\ &= \epsilon\psi_{il} + O(\delta^2) \end{aligned}$$

for $i = 1, \ldots, l$. Hence, to order $O(\delta^2)$, $\sum_{i=0}^{l} p_i^*(u_i^* - u_i) = p_l^*\epsilon + \sum_{i=0}^{l-1} p_i^*(u_i^* - u_i) = p_l^*\epsilon + \epsilon\sum_{i=1}^{l} p_{i-1}^*\psi_{il}$, giving

$$\sum_{i=0}^{l} p_i^*(u_i^* - u_i) = \epsilon\phi_l + O(\delta^2). \tag{7.26}$$

If $\phi_l < 0$, then $\sum_{i=0}^{l} p_i^* u_i^* < \sum_{i=0}^{l} p_i^* u_i$, so $\sum_{i=0}^{n} p_i^* u_i^* < \sum_{i=0}^{n} p_i^* u_i$.

Proof of (b). Suppose $u_0^*, \ldots, u_n^*)$ is a partition that satisfies equations (4.4) and (4.5) and that $(u_0^*, \ldots, u_n^*) \neq (u_0, \ldots, u_n)$. As each of $u_0, \ldots, u_n$ is a tight lower limit, it follows that at least one of $(u_0^*, \ldots, u_n^*)$ is not. Let $q$ be the smallest value of $i$ for which $u_i^*$ is not a tight lower limit. Put $\epsilon_1 = u_q^* - u_q$. As equation (7.25) holds for $i = 1, \ldots, q$, as in equation (7.26) we have that

$$\sum_{i=0}^{q} p_i^*(u_i^* - u_i) = \epsilon_1\phi_q + O(\delta^2). \tag{7.27}$$

Sequentially choose $b_q, b_{q-1}, \ldots, b_0$ such that

$$\frac{1}{u_q^* - b_q} \int_{b_q}^{u_q^*} C_{u,i}(\theta)\, d\theta = \gamma \tag{7.28}$$

and

$$\frac{1}{b_i - b_{i-1}} \int_{b_{i-1}}^{b_i} C_{u,i}(\theta)\, d\theta = \gamma \tag{7.29}$$

for $i = q,\ q-1, \ldots, 1$. Let $\epsilon_2 = b_q - u_q$. Then analogous to equation (7.26),

$$\sum_{i=0}^{q} p_i^*(b_i - u_i) = \epsilon_2 \phi_q + O(\delta^2). \tag{7.30}$$

Since equation (7.29) holds while (7.23) is a strict inequality when $i = q+1$, it follows that $b_q < a_q^*$, and so $\epsilon_2 < \epsilon_1$.

Now consider the partition $(b_0, \ldots, b_q, u_{q+1}^*, \ldots, u_n^*)$. It gives locally correct confidence intervals and

(i) its confidence intervals have shorter expected average length than the partition $(u_o^*, \ldots, u_n^*)$ – from comparison of equations (7.27) and (7.30);

(ii) it has one more tight lower limit than $(u_o^*, \ldots, u_n^*)$. Repeating the process that gave $(b_0, \ldots, b_q, u_{q+1}^*, \ldots, u_n^*)$, we can construct a partition that has one more tight lower limit than $(b_0, \ldots, b_q, u_{q+1}^*, \ldots, u_n^*)$ and which gives confidence intervals with a shorter expected average length. This can continue until we obtain a partition whose points are all tight lower limits. But that partition is $u_0, \ldots, u_n)$. Thus $(u_0, \ldots, u_n)$ gives confidence intervals with a shorter expected average length than all preceding partitions, including $(u_0^*, \ldots, u_n^*)$. That is, $\sum_{i=0}^n p_i^* u_i < \sum_{i=0}^n p_i^* u_i^*$, as required. $\diamond$

# Bibliography

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Wiley, New Jersey, second edition.

Agresti, A. (2003). Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. *Statistical Methods in Medical Research*, **12**, 3–21.

Agresti, A. and Coull, B. A. (1998). Approximate is better than exact for interval estimation of binomial proportions. *Journal of the American Statistical Association*, **52**, 119–126.

Agresti, A. and Gottard, A. (2005). Comment: Randomized confidence intervals and the mid-P approch. *Statistical Science*, **20**, 367–371.

Agresti, A. and Gottard, A. (2007). Nonconservative exact small-sample inference for discrete data. *Computational Statistics and Data Analysis*, **51**, 6447–6458.

Anscombe, F. (1956). On estimating binomial response relations. *Biometrika*, **43**, 461–464.

Arefi, M., Borzadaran, G. M., and Vaghei, Y. (2009). Interval estimation for the means of binomial, negative binomial, and Takacs distributions. *Italian Journal of Applied Statistics*, **21**, 361–375.

Barker, L. (2002). A comparison of nine confidence intervals for a Poisson

parameter when the expected number of events is $\leq 5$. *The American Statistician*, **56**, 85–89.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York, second edition.

Berry, G. and Armitage, P. (1995). Mid-P confidence intervals: a brief review. *Journal of the Royal Statistical Society, Series D*, **44**, 417–423.

Blaker, H. (2000). Confidence curves and improved exact confidence intervals for discrete distributions. *The Canadian Journal of Statistics*, **28**, 783–798.

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

Boomsma, A. (2005). Confidence intervals for a binomial proportion. *http://www.ppsw.rug.nl/ boomsma/confbin.pdf*, **68**, 1–9.

Brown, L. D., Cai, T. T., and Dasgupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, **16**, 101–133.

Brown, L. D., Cai, T. T., and Dasgupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *The Annals of Statistics*, **30**, 160–201.

Brown, L. D., Cai, T. T., and DasGupta, A. (2003). Interval estimation in exponential families. *Statistica Sinica*, **13**, 19–49.

Byrne, J. and Kabaila, P. (2005). Comparison of Poisson confidence intervals. *Communication in Statistics-Theory and Methods*, **34**, 545–556.

Cai, T. (2005). One-sided confidence intervals in discrete ditribution. *Journal of Statistical Planning and Inference*, **131**, 63–88.

Casella, G. (1986). Refining binomial confidence intervals. *The Canadian Journal of Statistics*, **14**, 113–129.

Casella, G. and McCulloch, C. E. (1984). Confidence intervals for discrete distributions. *Biometrics Unit, Cornell University, NY*, pages 1–17.

Choi, S. A. (2015). *Confidence intervals for functions of the success parameter of a negative binomial random variable.* Master's thesis, California State University, Northridge.

Cohen, C. and Yang, S. (1994). Mid-P confidence intervals for Poisson expectations. *Statistics in Medicine*, **13**, 2189–2203.

Crow, E. L. (1956). Confidence intervals for a proportion. *Biometrika*, **43**, 423–435.

Decrouez, G. and Hall, P. (2014). Split sample methods for constructing confidence intervals for binomial and Poisson parameters. *Journal of the Royal Satistical Society, Series B*, **76**, 949–975.

Dunnigan, K. (2008). confidence interval calculation for binomial proportions. *http://www.mwsug.org/proceedings/2008/pharma/MWSUG-2008-P08.pdf*, **68**, 1–9.

Fagerland, M., Lydersen, S., and Laake, P. (2015). Recommended confidence intervals for two independent binomial proportions. *Statistical Methods In Medical Research*, **24**, 224–254.

Garthwaite, P. H., Moustafa, M. W., and Elfadaly, F. G. (2019). Locally correct confidence intervals for a binomial proportion. *Submitted for publication*, **1**, 1–2.

Ghalanos, A. and Theussl, S. (2015). Package rsolnp. CRAN Repository.

Ghosh, B. (1979). A comparison of some approximate confidence intervals for the binomial parameter. *Journal of the American Statistical Association*, **74**, 894–900.

González, J. E., Barquinero, J. F., Holladay, B. A., Di Giorgio, M., and Higueras, M. (2020). Uncertainty calculation methods in dose assessment for dicentric chromosome assay. *International Journal of Radiation Biology*, **96**, 606–613.

Guan, Y. (2011). Moved score confidence intervals for means of discrete distributions. *Open Journal of Statistics*, **1**, 81–86.

Hepworth, G. (2013). Improved estimation of proportions using inverse binomial group testing. *Journal of Agricultural, Biological and Environmental Statistics*, **18**, 102–119.

Holladay, B. A. (2014). *Optimal confidence intervals for the expectation of a Poisson random variable*. Master's thesis, California State University, Northridge.

Jovanovic, B. D. and Levy, P. S. (1997). A look at the rule of three. *The American Statistician*, **51**, 137–139.

Kabaila, P. and Byrne, J. (2001). Exact short confidence intervals from discrete data. *Australian & New Zealand Journal of Statistics*, **43**, 303–309.

Khamkong, M. (2012). Approximate confidence interval for the mean of Poisson distribution. *Open Journal of Statistics*, **2**, 204–207.

Lancaster, H. O. (1961). Significance tests in discrete distribution. *Journal of the American Statistical Association*, **56**, 223–234.

Leemis, M. L. and Trivedi, K. S. (1996). A comparison of approximate interval

estimators for the bernoulli paeameter. *The American Statistician*, **50**, 63–68.

Liu, L. (2012). *Intervals estimation for binomial proportion, Poisson mean and negative binomial mean*. Ph.D. thesis, University of Cincinnati, USA.

Lui, K. (2004). *Statistical Estimation of Epidemiological Risk*. John Wiley & Sons, New Jersey, first edition.

Mehta, C. and Walsh, S. (1992). Comparison of exact, mid-p, and mantel-haenszel confidence intervals for the common odds ratio across several 2*2 contingency tables. *The American Statistician*, **46**, 146–150.

Nadarajah, S., Alizadeh, M., and Bagheri, S. (2015). Bayesian and non-bayesian interval estimators for the Poisson mean. *REVSTAT- Statistical Journal*, **13**, 245–262.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, **17**, 857–872.

Park, H. and Leemis, L. M. (2017). Ensemble confidence intervals for binomial proportions. *Statistics in Medicine*, **38**, 3460–3475.

Patil, D. and Rajarshi, M. (2010). Comparison of confidence intervals for the Poisson mean. *Model Assisted Statistics and Applications*, **5**, 79–91.

Patil, V. and Kulkarni, H. (2012). Comparison of confidence intervals for the Poisson mean: some new aspects. *Statistical Journal*, **10**, 211–227.

Pearson, K. (1924). A note on the relationship of the incomplete b-function to the sum of the first p terms of the binomial $(a+b)^n$. *Biometrika Trust*, **16**, 202–203.

Pires, A. M. and Amado, C. (2008). Interval estimation for a binomial proportion: comparison of twenty methods. *REVSTAT- Statistical Journal*, **6**, 165–197.

Reiczigel, J. (2003). Confidence intervals for the binomial parameter: some new considerations. *Statistics in Medicine*, **22**, 611–621.

Sahai, H. and Kurshid, A. (1993). Confidence intervals for the mean of a Poisson distribution: a review. *Biometrical Journal*, **35**, 857–867.

Schilling, M. F. and Doi, J. A. (2014). A coverage probability approach to finding an optimal binomial confidence procedure. *The American Statistician*, **68**, 133–145.

Schilling, M. F. and Holladay, B. (2017). Length minimization for Possion confidence procedures. *Communications in Statistics-Theory and Methods*, **46**, 861–873.

Sterne, T. H. (1954). Some remarks on confidence or fiducial limits. *Biometrika*, **41**, 275–278.

Swift, M. (2009). Comparison of confidence intervals for a Poisson mean further considerations. *Communications in Statistics-Theory and Methods*, **38**, 748–759.

Tian, M., Tang, M., Ng, H. K. T., and Chan, P. S. (2009). A comparitive study of confidence intervals for negative binomial proportion. *Journal of Statistical Computation and Simulation*, **79**, 241–249.

Vollset, S. (1993). Confidence intervals for a binomial proportion. *Statistics in Medicine*, **12**, 809–824.

Wang, W. and Shan, G. (2015). Exact confidence intervals for the relative risk and the odds ratio. *Biometrics*, **71**, 985–995.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, **22**, 209–212.

Young, D. S. (2014). A procedure for approximate negative binomial tolerance intervals. *Journal of Statistical Computation and Simulation*, **84**, 438–450.

Zhao, S. (2005). *Statistical Inference on Binomial Proportions*. Ph.D. thesis, University of Cincinnati, USA.