12-2021

Confidence Interval for the Mean of a Beta Distribution

Sean Rangel
Stephen F Austin State University, rangelsm2@jacks.sfasu.edu

Confidence Interval for the Mean of a Beta Distribution

Confidence Interval for the Mean of a Beta Distribution.

by

Sean Rangel, B.S.

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

December 2021

Confidence Intervals for the Mean of a Beta Distribution.

by

Sean Michael Rangel, B.S.

APPROVED:

Kent Riggs, Ph.D., Co-Thesis Director

Jacob Turner, Ph.D., Co-Thesis Director

Robert Henderson, Ph.D., Committee Member

Lindsay Porter, Ph.D., Committee Member

Freddie Avant, Ph.D.
Interim Dean of Research and Graduate Studies

ABSTRACT

Statistical inference for the mean of a beta distribution has become increasingly popular in various fields of academic research. In this study, we developed a novel statistical model from likelihood-based techniques to evaluate various confidence interval techniques for the mean of a beta distribution. Simulation studies will be implemented to compare the performance of the confidence intervals. In addition to the development and study involving confidence intervals, we will also apply the confidence intervals to real biological data that was gathered by the Department of Biology at Stephen F. Austin State University and provide recommendations on the best practice.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family for their love and support through my graduate experience. Without them I do not think I would have made it through this journey. I would also like to express my appreciation to the following individuals who have helped me through this process: My thesis advisors Dr. Kent Riggs and Dr. Jacob Turner for all those long hours that they poured into me to help me succeed and for their immense patience. To my committee members, Dr. Robert Henderson and Dr. Lindsay Porter for their guidance and sage wisdom. I would also like to thank all my friends for their support. We had some good times and some bad but at the end we can say we finally did it.

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The beta distribution has a rich history in the field of statistics. It is often used to model data that are represented as proportions or percentages. The beta distribution is also reserved for representing all the possible values of a probability when the true value of the probability is unknown [6, 9]. A modified version of the beta distribution can be used to model continuous random variables that are defined on the closed interval [0,1] [6, 9].

According to the literature, very little attention has been given to the performance of asymptotic interval estimation for the mean of a beta random variable. However [3], has provided some work in this area, but their methods assume that the precision parameter is known in advance, a situation that is unlikely to occur in practice. Conversely, this study will assume the precision parameter is unknown. Confidence interval using likelihood-based estimation are quite standard and widely accepted in practice, however, [4] states that this technique is inaccurate for a large sample size. Moreover, past simulation studies have revealed that the Wald method is conservative for 95% confidence intervals [4]. A work around for this might be to employ a simple t-test since it is robust in many settings [7].

In this research, our intention is to perform simulation studies by developing a statistical model to determine appropriate scenarios and sample sizes to evaluate how various confidence interval techniques perform versus nominal coverage rates for the beta distribution. These investigations will allow numerous recommendations to be made on when an interval should and should not be applied, and offer up understanding about which methods may perform better than others.

In chapter 2, we present the formal probability density, maximum likelihood estimators (MLE) and Fisher's information for the beta distribution. In chapter 3, we present the Wald, t-interval, and bootstrap confidence intervals for the mean of a beta distribution. As an illustration of the confidence intervals, we consider an application to biological data in chapter 4. Lastly, in chapter 5, we will provide conclusions and make comments for future work.

2 Mathematical Preliminaries

In this chapter, we introduce the beta distribution, and topics related to likelihood-based estimation, such as maximum likelihood estimation and Fisher's information. Furthermore, the likelihood based applications to the beta distribution will be introduced. The results can be found in [2, 6, 9].

2.1 Beta Distribution

The formal mathematical constructs for the beta distribution can be found in [2] and they will be presented here. The probability density function for a beta random variable, X, is as follows:

$$f(X;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, 0 \leq x \leq 1, \alpha > 0, \beta > 0; \tag{2.1}$$

where the Beta function is

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

and the gamma function is

$$\Gamma(a) = \int_0^\infty w^{a-1}e^{-w}\,dw.$$

Three important parameters and their expressions are:

$$\mathrm{E[X]} = \mu = \frac{\alpha}{\alpha+\beta}$$

,

$$\mathrm{Var(X)} = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2)(1+\alpha+\beta)},$$

$$\text{Precision} = \phi = \alpha+\beta.$$

The beta distribution has two shape parameters alpha (α) and beta (β). The shape parameters, denoted by α and β, control and describe the shape of the beta distribution [6, 9]. The parameters, α and β, can be estimated by using the maximum likelihood estimation (MLE) technique [2]. The MLE technique is a common method of estimating parameters of a probability distribution. Assuming the beta distribution is a reasonable model, the MLE will find α and β values that results in a model that "best fits" a given data set.

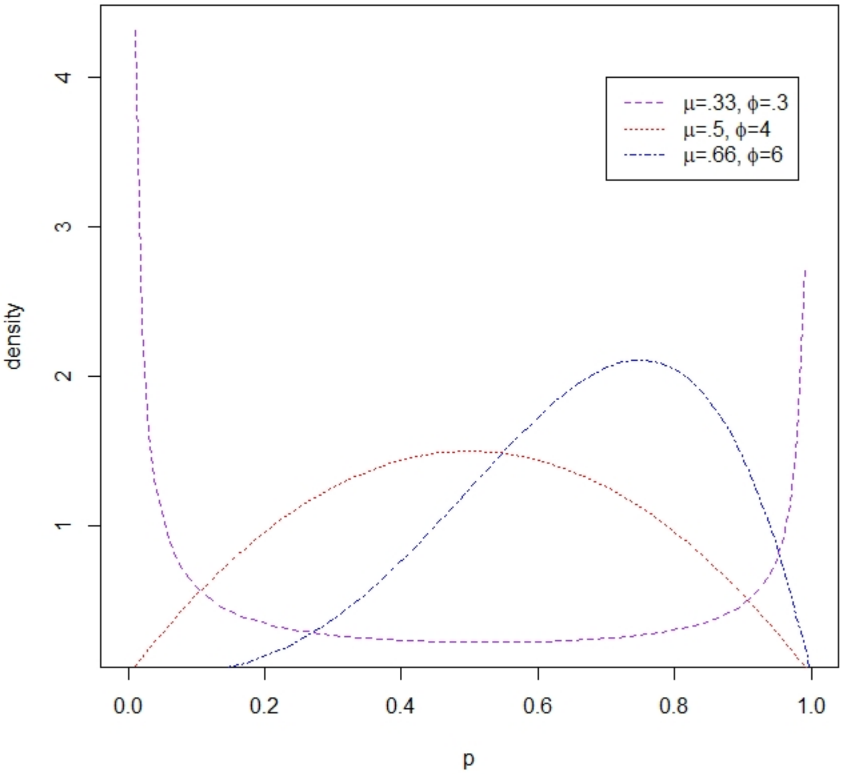

Figure 2.1: Beta Density Curves

Figure 2.1 highlights how flexible the beta density curves can be for different values of μ and ϕ. This is appealing since many data sets are often not symmetric or bell shaped. Alternatively, the beta distribution can be roughly bell and symmetric but can also be skewed and u-shaped. To demonstrate this property of the beta distribution, figure 2.1 provides the pdfs for certain values of the parameters. The pdfs of the beta can become u-shaped ($\mu = .33$, $\phi = .3$), symmetric ($\mu = .5$, $\phi = 4$), and left-skewed ($\mu = .66$, $\phi = 6$). The versatility of the beta distribution can account for many distributional shapes that one might encounter in practice.

2.2 Maximum Likelihood Estimation

In this section, the maximum likelihood estimation (MLE) method will be introduced. The method of maximum likelihood is one of the most popular methods for deriving statistical estimators. The general theory will be discussed first, then the MLE for the beta distribution will be derived. The general discussion will be for a continuous case.

2.2.1 General Theory

In a typical statistics problem, there is a random variable of interest, often denoted as X that has a probability density function (pdf) in the form of $f(x;\theta)$, where $\theta \in \Omega$ for a specified set Ω, and θ is the unknown parameter of the distribution. Now suppose, that $X_1, ..., X_n$ are independent and identically distributed (iid) random variables from a common pdf $f(x;\theta)$. The basis of many inferential procedures is done by using the following function, which is called the likelihood function:

$$L(\theta; x) = L(\theta) = \prod_{i=1}^{n} f(x_i; \theta), \tag{2.2}$$

where $x = (x_1, ..., x_n)'$ is a random vector and $\theta = (\theta_1, ..., \theta_k)'$ is the parameter vector. It is common practice and often more convenient to take the log of the likelihood function, called the log-likelihood:

$$l(\theta) = logL(\theta) = \sum_{i=1}^{n} logf(x_i;\theta). \tag{2.3}$$

Note that there is no loss of information in using $l(\theta)$ because the log is a one-to-one transformation and order preserving. Furthermore, for each observed random sample x, let the value $\hat{\theta}(x)$ be the parameter value that maximizes $l(\theta;x)$, which also maximizes $L(\theta)$.

Moreover, if the log-likelihood function is differentiable (in θ_i), the possible candidates for the MLE's are values of $(\theta_1,...,\theta_k)$ that solve the equations:

$$\frac{\partial l}{\partial \theta_i} = 0, i = 1,...,k. \tag{2.4}$$

the previous expression is often referred to as the "estimating equations".

2.2.2 MLE of the Beta Distribution.

This subsection will apply the results from 2.2.1 to the pdf of the beta distribution.

Suppose a random sample of $x_1,...,x_n$ has been collected for a random variable X from the beta distribution defined by equation 2.1. Here, we define $\theta = (\alpha,\beta)'$.

Additionally, recall the log likelihood equation, defined by equation 2.3. The log-likelihood function is a computationally convenient tool to find the MLE's of α and β. Using expression (2.1), the likelihood function of the beta distribution is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\alpha,\beta) = \prod_{i=1}^{n} \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\right] x_i^{(\alpha-1)}(1-x_i)^{(\beta-1)}. \tag{2.5}$$

Hence, the natural log-likelihood function is:

$$l(\theta) = Ln(L(\theta)) = \sum_{i=1}^{n} Ln(f(x_i;\theta)$$

which is

$$\begin{aligned} l(\theta) &= nLn(\Gamma(\alpha+\beta)) - nLn(\Gamma(\alpha)) - nLn(\Gamma(\beta)) \\ &\quad + (\alpha-1)Ln[\sum_{i=1}^{n} x_i] + (\beta-1)Ln[\sum_{i=1}^{n}(1-x_i)] \end{aligned} \tag{2.6}$$

Now that the log-likelihood function has been derived, we must take the first derivative with respect to α and β of 2.6, and we have the following result:

$\frac{\partial l}{\partial \alpha} = [n(\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)}) - n(\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}) + Ln(\sum_{i=1}^{n} x_i)]$

and

$\frac{\partial l}{\partial \beta} = [n(\frac{\Gamma'(\alpha+\beta)}{\Gamma(\alpha+\beta)}) - n(\frac{\Gamma'(\beta)}{\Gamma(\beta)}) + Ln(\sum_{i=1}^{n}(1-x_i))]$.

Now, we will set $\frac{\partial l}{\partial \alpha} = 0$ and $\frac{\partial l}{\partial \beta} = 0$ to get the estimating equations. Additionally, by the relationship $\frac{\Gamma'(a)}{\Gamma(a)} = \Psi(a) = \frac{\partial}{\partial(a)} log\Gamma(a)$, we can rewrite the estimating equations as the following:

$-\Psi(\alpha+\beta) + \Psi(\alpha) + \frac{1}{n}Ln[\sum_{i=1}^{n} x_i] = 0$

and

$-\Psi(\alpha+\beta) + \Psi(\beta) + \frac{1}{n}Ln[\sum_{i=1}^{n} 1 - x_i] = 0$.

To obtain the the MLE for α and β, we will let R studio numerically solve for the α and β estimates since no closed form solution exists. Note that if $\hat{\alpha}$ and $\hat{\beta}$ are MLE's for α and β, then by the invariance property of MLE's, the MLE's for μ and ϕ are: $\hat{\mu} = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}$ and $\hat{\phi} = \hat{\alpha} + \hat{\beta}$ [9].

2.3 Fisher's Information

This section will briefly introduce the general theory of Fisher's information and then it will be applied to the beta distribution, equation (2.1).

Let X be a random variable with pdf $f(x;\theta)$, where $\theta = (\theta_1, ..., \theta_k)' \in \Omega \subseteq \mathbb{R}^k$, where the parameter space Ω is an open interval. Fisher's information matrix is denoted as $I(\theta)$; where I_{θ_i,θ_j} where $i, j = 1, ..., k$ make up the elements of $I(\theta)$ and are defined as

$$I_{\theta i,\theta j} = -\int_{-\infty}^{\infty} \frac{\partial^2 log f(x;\theta)}{\partial\theta_i \partial\theta_j} f(x;\theta)dx = -E\left[\frac{\partial^2 log f(X;\theta)}{\partial\theta_i \partial\theta_j}\right]. \tag{2.7}$$

The diagonal entries of $I(\theta)$ provides the bounds for the variance of an unbiased estimator for the corresponding element of θ. As the information number gets larger, a smaller bound on the variance of the estimator will be produced.

2.3.1 Fisher's Information for the Beta Distribution.

Now that the theory of Fisher's information has been introduced in section 2.3, it will be applied to the pdf of the beta distribution. Consider Fisher's information matrix:

$$I(\alpha,\beta) = -E\begin{bmatrix} \frac{\partial^2 l}{\partial\alpha^2} & \frac{\partial^2 l}{\partial\alpha\partial\beta} \\ \frac{\partial^2 l}{\partial\beta\partial\alpha} & \frac{\partial^2 l}{\partial\beta^2} \end{bmatrix} = \begin{bmatrix} I_{(\alpha,\alpha)} & I_{(\alpha,\beta)} \\ I_{(\alpha,\beta)} & I_{(\beta,\beta)} \end{bmatrix}. \tag{2.8}$$

Now, if we take the second derivatives with respect to α and β, the entries for Fisher's information matrix are as follows:

$$I_{\alpha,\alpha} = -E(\frac{\partial l}{\partial\alpha^2}) = \Psi''(\alpha+\beta) - \Psi''(\alpha),$$

$$I_{(\beta,\beta)} = -E(\frac{\partial l}{\partial\beta^2}) = \Psi''(\alpha+\beta) - \Psi''(\beta),$$

and

$$I_{(\alpha,\beta)} = -E(\frac{\partial l}{\partial\alpha\partial\beta}) = -\Psi''(\alpha+\beta).$$

Therefore, Fisher's information for the beta distribution under the α and β parameterization is,

$$I(\alpha,\beta) = \begin{bmatrix} \Psi''(\alpha) - \Psi''(\alpha+\beta) & -\Psi''(\alpha+\beta) \\ -\Psi''(\alpha+\beta) & \Psi''(\beta) - \Psi''(\alpha+\beta) \end{bmatrix}. \tag{2.9}$$

To reparameterize Fisher's information in terms of μ and ϕ, we must consider that Fisher's information depends not only on the value of θ, but also on the reparameterization. For more information refer to [5]. Hence, Fisher's information is:

$$I(\mu,\phi) = B'I(\mu\phi,(1-\mu)\phi))B, \text{ where B is the Jacobian matrix.} \tag{2.10}$$

The elements for the Jacobian are defined as $b_{ij} = \frac{\partial g_i^{-1}}{\partial \theta_j}$, where $g_1(\alpha,\beta) = \frac{\alpha}{\alpha+\beta} = \mu$ and $g_2(\alpha,\beta) = \alpha+\beta = \phi$. Furthermore, note that

$$g_1^{-1}(\mu,\phi) : \alpha = \mu\phi$$

$$and$$

$$g_2^{-1}(\mu,\phi) : \beta = \phi - \mu\phi.$$

Thus, the Jacobian is:

$$B = \begin{bmatrix} \phi & \mu \\ -\phi & 1-\mu \end{bmatrix}. \tag{2.11}$$

The principle of parametrisation invariance is a valuable basis for choosing between different inferential procedures. For the verification of the result, refer to [5], page 147. So, by applying equation 2.10 and the relationship between μ and ϕ, Fisher's information for μ and ϕ is the following,

$$I(\mu,\phi) = \begin{bmatrix} \phi & -\phi \\ \mu & 1-\mu \end{bmatrix} \begin{bmatrix} \Psi''(\mu\phi) - \Psi''(\phi) & -\Psi''(\phi) \\ -\Psi(\phi) & \Psi''((1-\mu)\phi) - \Psi''(\phi) \end{bmatrix} \begin{bmatrix} \phi & \mu \\ -\phi & 1-\mu \end{bmatrix}, \tag{2.12}$$

which will simplify to the following:

$$I(\mu, \phi) = \begin{bmatrix} I_{\mu,\mu} & I_{\mu,\phi} \\ I_{\phi,\mu} & I_{\phi,\phi} \end{bmatrix}. \tag{2.13}$$

The entries for $I(\mu, \phi)$ matrix are

$$I_{\mu,\mu} = \phi^2\Psi''(\mu\phi) + \phi^2\Psi''((1-\mu)\phi)$$

$$I_{\mu,\phi} = \phi\mu\Psi''(\mu\phi) - \phi(1-\mu)\Psi''((1-\mu)\phi),$$

$$I_{\phi,\mu} = \phi\mu\Psi''(\mu\phi) - \phi(1-\mu)\Psi''((1-\mu)\phi),$$

$$and$$

$$I_{\phi,\phi} = \mu^2\Psi''(\mu\phi) + (1-\mu)^2\Psi''((1-\mu)\phi) - \Psi''(\phi).$$

Thus, the desired result of reparameterizing Fisher's information for μ and ϕ has been achieved.

2.4 Conclusion

This chapter introduced the mathematical preliminaries for the beta distribution. Additionally, the general theory of likelihood estimation and Fisher's information were introduced and then applied to the beta distribution. In the next chapter, we will introduce interval estimates and discuss how the results from this chapter can assist in the construction of confidence intervals.

3 Confidence Intervals for μ

In this chapter, the Wald, t-interval, and bootstrap intervals for the mean of a Beta distribution will be introduced. Moreover, the coverage and the width properties of the intervals will be studied and presented in a simulation study.

3.1 Confidence Interval Definitions

The first confidence interval that will be discussed is the Wald interval. The Wald interval is a basic and popular method for calculating confidence intervals for MLE's [4], and it was the main interest. Despite the popularity of the Wald interval, [4] states that the Wald interval is flawed and inaccurate for small sample sizes and larger probability values. Additionally, it has been reported that the Wald confidence intervals are very conservative meaning it will still produce interval estimates but they could potentially be wider than necessary [4].

In addition to the Wald interval, the two other intervals that were considered for estimating μ in this study were the t-interval and the bootstrap interval. The t-interval is also a popular parametric inferential technique among researchers. The use of the t-interval under the assumption of a non-normal population is based on the central limit theorem (CLT) [7, 10]. The CLT states that the mean of a random sample given a sufficiently large sample size, n, from a population with mean, μ, and variance, σ^2, is approximately normally distributed with mean, μ, and variance, $\frac{\sigma^2}{n}$, regardless of the population distribution [7, 10]. Because of the CLT, the t-interval is fairly robust to departures from normality since it is based on the sample mean and the fact that the t-distribution approaches the Normal distribution for large n. Lastly, the bootstrap interval is another inferential technique. The basis of the bootstrap interval is to build a sampling distribution for the statistic of interest through the generation of artificial samples by sampling with replacement from the original sample.

3.1.1 Wald Interval

Because $\hat{\mu}$ and $\hat{\phi}$ are MLE's, we have that

$$\begin{pmatrix} \hat{\mu} \\ \hat{\phi} \end{pmatrix} \dot{\sim} N_2 \left(\begin{pmatrix} \mu \\ \phi \end{pmatrix}, \quad \tfrac{1}{n} I^{-1}(\mu, \phi) \right),$$

where $I^{-1}(\mu, \phi)$ is a 2×2 matrix.

Also, for large sample size n and by properties of MLE's (refer to Corollary 6.4.1 in [9]), we have the following

$$\frac{\hat{\mu} - \mu}{\sqrt{\frac{1}{n} I^{-1}(\mu, \phi)_{11}}} \dot{\sim} N(0, 1),$$

.

for large n and by Slutsky's theorem [9], we have the expression:

$$\frac{\hat{\mu} - \mu}{\sqrt{\frac{1}{n} I^{-1}(\hat{\mu}, \hat{\phi})_{11}}} \dot{\sim} N(0, 1)$$

.

So, a $100(1-\alpha)\%$ a large sample confidence interval for μ is:

$$\hat{\mu} \pm Z_{\alpha/2} \sqrt{\frac{1}{n} I^{-1}(\hat{\mu}, \hat{\phi})_{11}}, \tag{3.1}$$

where $Z_{1-\frac{\alpha}{2}}$ is the upper $1 - \frac{\alpha}{2}$ quantile of a Z-distribution.

3.1.2 t-test Interval

Here, we present the the t-interval, which is robust to departures from normality because of the CLT.

Let X be a random variable and suppose $X_1, ..., X_n$ is a random sample. Also, we will let $\bar{X}$ and s^2 denote the sample mean and variance, respectively. For large n, the CLT dictates that

$$\bar{X} \dot{\sim} N(\mu, \frac{\sigma^2}{n}).$$

If the sample mean is standardized, then we have

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0,1).$$

Also, by Slutsky's theorm [9],

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \dot{\sim} N(0,1).$$

So, by the CLT and with a sufficiently large sample size, the confidence interval is:

$$\bar{X} \pm Z_{\alpha/2} \frac{s}{\sqrt{n}}.$$

However, recall that the t-interval is more robust than a Z interval, so we can rewrite the previously listed confidence interval formula as:

$$(\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}), \tag{3.2}$$

where $t_{\alpha/2}$ is the lower $\frac{\alpha}{2}$ quantile of a t_{n-1} distribution.

3.1.3 Bootstrap Interval

In this subsection, we present the non-parametric percentile bootstrap procedure, which is a resampling procedure, and it has become increasingly popular in statistical inference.

The bootstrap procedure will resample from the original sample. The sampling is done at random and with replacement and the resamples are all size n, the original sample size. Now suppose, $x' = (x_1, ..., x_n)$ denotes the original realization of the sample drawn from the pdf $f(x; \theta)$, where $\theta \in \Omega$. Also, let $\hat{\theta}$ be a point estimator for θ, and let $B \in \mathbb{Z}$ where $\mathbb{Z}$ is an integer, so let B denote the number of resamples. Also, let $\hat{\theta}_1, \hat{\theta}_2, .., \hat{\theta}_B$ be the statistic $\hat{\theta}$ evaluated on each of the B bootstrap samples. Thus, the $100(1-\alpha)\%$ percentile bootstrap confidence interval is:

$$(\hat{\theta}_{(\alpha/2)}, \hat{\theta}_{(1-\alpha/2)}); \tag{3.3}$$

where $\hat{\theta}_{(\frac{\alpha}{2})}$ and $\hat{\theta}_{(1-\frac{\alpha}{2})}$ is the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ percentiles of the $\hat{\theta}_i$'s for $i = 1, ..., B$. For more information on the percentile bootstrap confidence interval refer to page 274 in [9].

3.2 Simulation Studies

For each of the three approximate confidence intervals in this study we will evaluate the coverage probabilities for varying values of μ, ϕ, and n. Also, we will compare the precision of the three intervals by comparing their average widths. The coverage and width properties will be examined in a simulation study.

Below is pseudo code to illustrate of how the simulations were performed by the open-source statistical package R studio.

```
1. For $i$ in 1 : 10000
        a. Randomly generate a sample from the Beta distribution.
        b. Compute the MLE of the Beta distribution.
        c. Make confidence intervals for the mean of the Beta distribution.
```

d. Count how many times the true parameter value falls in the confidence interval and store result as a count.

e. Store the widths of each interval.

2. End Loop

3. Calculate the estimated coverage probabilities, average widths.

3.2.1 Results from Simulation Studies

The results from section 3.2 will be presented here. For the following discussion the reader may refer to Tables A and refer to the Graphs C in the appendix.

In the simulation studies we explored the coverage probabilities and the widths of the intervals for the Wald, t-interval, and bootstrap confidence intervals for values of $\mu = .1, ..., .9$ by .1. We will discuss three cases for the simulation studies. The first case is when $\phi = 10$ and $n = 5, 20, 50$. The second case is when $\phi = 30$ and $n = 5, 20, 50$. Lastly, the third case is when $\phi = 50$ $n = 5, 20, 50$. The values for ϕ were chosen arbitrarily.

For the first case when $\phi = 10$ and $n = 5, 20, 50$ the results are displayed in Tables A.1, A.3, and A.5 and Figure C.1 in the appendix. The t-test outperformed both the Wald and bootstrap confidence intervals for each sample size, and the Wald and bootstrap intervals produced similar coverage probability values to each other. For each sample size, the t-interval was near the nominal value of .95 for all values μ, which is the target value for all the confidence intervals. The Wald interval performed poorly when the sample size was small ($n = 5$), but when the sample sized increased ($n = 20, 50$), we noticed the Wald began to produce values closer to the nominal level. A similar trend was noticed for the bootstrap interval. The bootstrap interval performed poorly for when the sample size was $n = 5, 20$. The highest value for the bootstrap interval when $n = 5$ was .845, and the highest value when $n = 20$ was .933.

However, when the sample size was large ($n = 50$), the bootstrap was closer to the nominal value.

For the second case when $\phi = 30$ and $n = 5, 20, 50$, the results are displayed in Tables A.7, A.9, and A.11 and Figure C.2 in the appendix.The t-interval outperformed both the Wald and bootstrap intervals by producing coverage probabilities near the nominal value of .95 for each sample sizes, $n = 5, 20, 50$. The Wald interval under-performed when the sample size was $n = 5, 20$, but when the sample size was large ($n = 50$), the Wald interval began producing values near the nominal level of .95. The bootstrap interval under-performed by producing values well below the nominal value for in each sample size, $n = 5, 20, 50$.

For the last case when $\phi = 50$ and $n = 5, 20, 50$, the results are displayed in Tables A.13, A.15, and A.17 and Figure C.3 in the appendix. The t-test still out-performed both the Wald and bootstrap intervals by producing coverage probabilities close to the nominal value of .95 for each sample size ($n = 5, 20, 50$). The Wald interval under-performed when the sample size was $n = 5, 20$. However, like in the first two cases when the sample size was $n = 50$, the Wald interval produced coverage probabilities near the nominal value of .95. The bootstrap under performed for each of the sample sizes $n = 5, 20, 50$.

In addition to the coverage probabilities, we also examined the widths of the confidence intervals of interest in this study. In the case when $\phi = 10$ refer to Tables A.2, A.4, and A.6 in the appendix. When the sample size is small, $n = 5$, the widths for each of the confidence intervals are incredibly wide. This indicates that for this case not much information can be drawn from the intervals, and that further information is required. When the sample size increased, $n = 20, 50$, the widths of the confidence intervals become considerably smaller. However, despite the reduction in width sizes, the Wald and bootstrap intervals were not at the nominal level of .95, so even though the widths were smaller, the intervals were not producing appropriate coverage probabilities.

In the second case, when $\phi = 30$ refer to Tables A.8, A.10, and A.12 in the appendix. When the sample size is small, $n = 5$, we notice that the widths were still fairly large. Although, when the sample sized increased ($n = 20, 50$), the widths of the intervals began to reduce and the coverage probabilities were near the nominal level of .95.

For the last case when $\phi = 50$, refer to Tables A.14, A.16, and A.18. A similar pattern was noticed in the last case when $\phi = 50$. When the sample size was small the widths of the intervals were larger, but as the sample size increased, the widths became more narrow while maintaining appropriate probabilities as expected. The narrower the widths, the more precise the confidence interval is and more information is provided from the interval.

In summation, for each of the three cases, the t-interval outperformed both the Wald and bootstrap intervals for coverage, but at the cost of wider intervals. The Wald and the bootstrap interval produced interval estimates that were below the nominal level. For both methods, when size increased it produced coverage probabilities closer to the nominal level.

4 Biological Application

In this chapter, the confidence intervals introduced in chapter 3 will be applied to biological data. The data consists of counts of hemocytes using a Corning cell counter of *Amblyomma americanum* infected with *Escherichia coli* (*E. coli*). The data was collect by Miss Jacquelyn May under the supervision of Dr. Lindsay Porter in the fall of 2021.

4.1 *Amblyomma americanum* Background

Amblyomma americanum, (*A. americanum*), is an ectoparasitic arthopod that primarily feeds on vertebrates such as mammals, birds, and reptiles [1]. *A. americanum* is commonly called the lone star tick because of the distinct star-shaped spot near the posterior portion on an adult female. *A. americanum* is distributed across the south eastern United States and currently occupies 37 states [8].

Moreover, *A. americanum* is a vector of pathogens that cause diseases such as Ehrlichiosis, Tularemia, and rickettsiosis [1, 8]. Pathogens transmitted by *A. americanum* have both medical and veterinary importance. These diseases are transferred during a blood meal when the tick is feeding. The tick immune system consists of hemolymph which is comprised of hemocytes that secrete a variety of proteins to combat pathogens [1]. However, despite the measures the ticks make to combat these pathogens, they manage to evade the ticks immune system. The data that was gathered by the Miss Jacquelyn May consists of cell counts between immune-compromised ticks and non-compromised ticks to explore the impact on hemocyte response. To investigate this further, *E. coli* was used in the experiment as a model for tick infection with bacteria similar to those that are pathogenic.

4.2 Statistical Analysis

This section will discuss the analysis of the *A. americanum* cell counts by applying the confidences intervals that were introduced and developed in chapter 3. A typical total cell count is generally in the millions of cells. Moreover, the variable ”viability” indicates the ability of *A. americanum* to maintain itself or recover its potentialities from an infection. The control group in this study is the group of *A. americanum* is the non-immune-compromised ticks and the experimental group in this study is the group of *A. americanum* that is immune-compromised ticks that were infected with *Escherichia coli.* The data that was used in this is analysis is:

Viability (%)	Control Group	Experimental Group
	98.1	90.9
	93.9	83
	86.3	99
	94.5	87.3
	92.8	95.9
	94.6	96.7
	86.8	96.1
	91.9	94.3
	97.5	90.3
	80.4	92.7
	87.3	94.3
		93.2

Table 4.1: Viability percentages of *A. americanum* cell counts.

The viability (%) was calculated by:

$$\text{Precentage of Viable cells} = \frac{\text{number of viable cells}}{\text{total number of cells}} \cdot 100.$$

Hemocyte viability was calculated and used to analyze the biological data because viability provides some measure of how well hemocytes are defending the tick during the infection and also provides a suitable dataset to accomplish the goals of this thesis.

In addition to the coverage probabilities and widths of the intervals, we plotted the data sets to determine if the distribution of the data sets follow a beta distribution. Below are box plots, histograms, and QQ plots of the data sets.

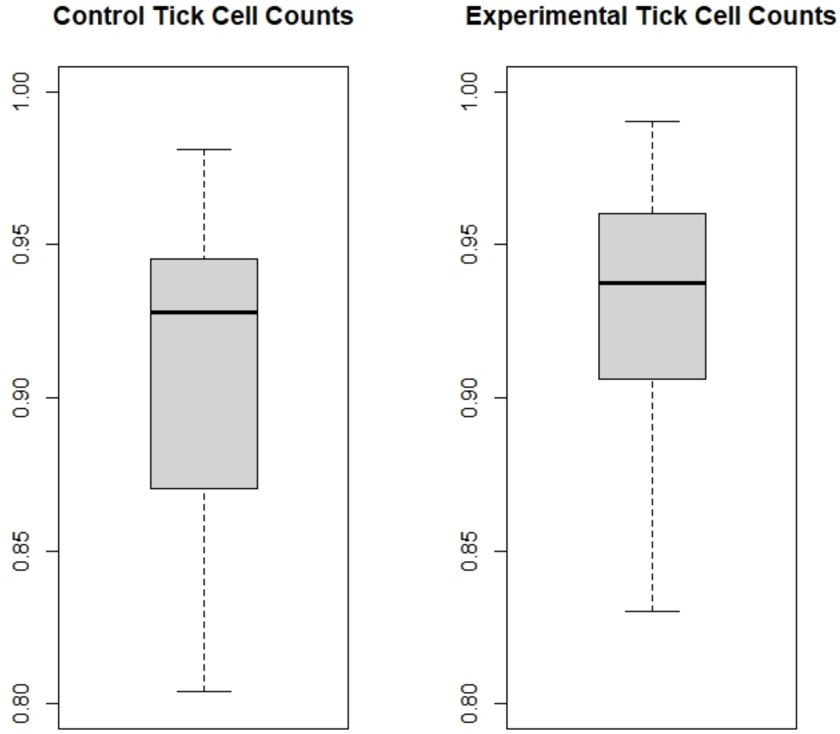

Figure 4.1: Box plots of the experimental and control *A. americanum* cell counts.

Figure 4.1 provide a good indication of the spread of the data. Both of the box plots appear to be left-skewed but the control tick cell counts are heavily left-skewed. There appears to be a lot of spread in the data sets but there are no outliers present.

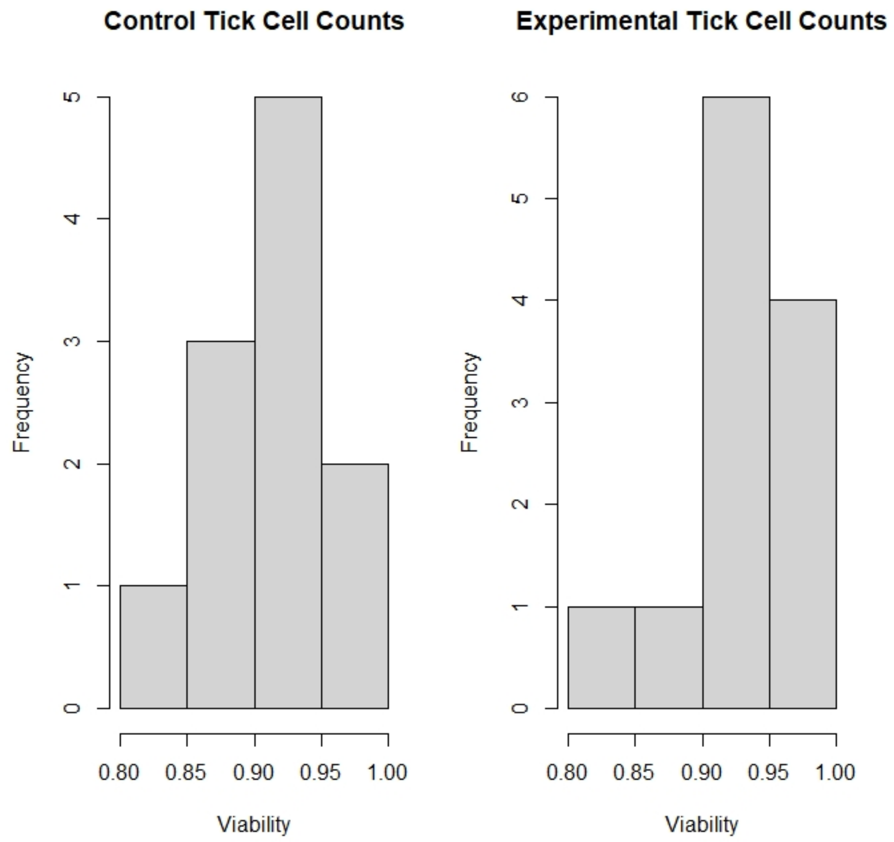

Figure 4.2: Histograms of the Experimental *A. americanum* cell counts and Control *A. americanum* cell counts

Along with the box plots, histograms provide an additional way to look at the distribution of the data sets. The histograms in Figure 4.2 emphasize the skewness in the data sets. Both of the histograms are unimodal and left-skewed with no outliers. The center of both histograms is around .90 viability, which indicates that about 90% of the hemocytes are surviving and to continue to fight the infection.

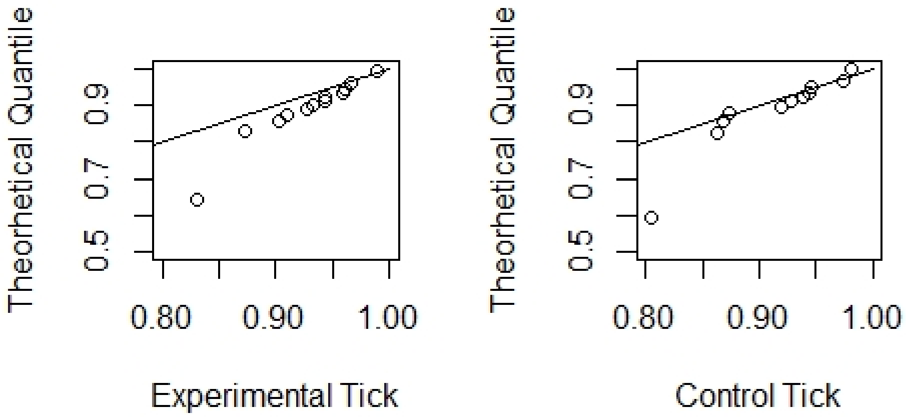

Figure 4.3: Beta distribution QQ plots of the experimental and control *A. americanum* cell counts. The rough estimates for the experimental and control tick parameter values are $\alpha = 29.7$ and $\beta = 3.3$, and $\alpha = 23.4$ and $\beta = 2.6$, respectively.

In addition to the box plots and histograms, a QQ plot provides additional evidence to determine if the data sets follow a beta distribution. Notice in the experimental QQ plot, there is a slight non-linearity of the data points. The control tick QQ plot appears more reasonable for the beta distribution. However, recall that the sample sizes for the control tick group and experimental tick group are 11 and 12, respectively. With the samples sizes being relatively small, it is difficult to ascertain a distribution assumption, but since the data is presented as percentages, the beta distribution is still appropriate in this scenario.

Below are the confidence intervals of interest in this study for the control group and experimental group of *A. americanum* cell counts with their respective lower and upper 95% bounds.

Viability of Uninfected Ticks Confidence Intervals	Lower 95%	Upper 95%
t-interval	0.876	0.949
Wald	0.879	0.942
Bootstrap	0.897	0.927

Table 4.2: The t-interval, Wald, and bootstrap confidence interval for the control group of *A. americanum*

Viability for Infected Ticks Confidence Intervals	Lower 95%	Upper 95%
t-test	0.9	0.956
Wald	0.912	0.940
Bootstrap	0.904	0.951

Table 4.3: The t-test, wald, and bootstrap confidence interval for the experimental group of *A. americanum* that was infected with *Escherichia coli.*

Upon closer examination of tables 4.2 and 4.3 we notice that the control group and experimental group confidence intervals were roughly similar. To investigate this further, we ran an additional simulation to estimate μ and ϕ for the control and experimental datasets to get a sense of which interval will perform best. The estimated μ and ϕ for the control group is .9 and 26, respectively. The estimated μ and ϕ for the experimental group is .9 and 33, respectively. Below are the coverage probabilities and widths of the confidence intervals for the estimated values of μ and ϕ for the control and experimental *A. americanum* data.

Method	Coverage Probability	Width
t-test	0.933	0.075
Wald	0.903	0.063
Bootstrap	0.893	0.063

Table 4.4: Coverage probability and confidence intervals widths when $\mu = .9$, $\phi = 26$, and $n = 11$.

Method	Coverage Probability	Width
t-test	0.94	0.063
Wald	0.911	0.055
Bootstrap	0.907	0.054

Table 4.5: Coverage probability and confidence intervals and widths when $\mu = .9$, $\phi = 33$, and $n = 12$.

The above simulation results provides more insight into which intervals are more appropriate for this data set. Notice in table 4.4 that both the t-interval and Wald interval are above .9 and the widths for the intervals are narrow, but the t-interval is closer to the nominal level of .95. So, in this case it better to use the t-interval over the other two methods. Additionally, notice in table 4.5 that each of the intervals are above .9, but again, the t-interval is closer to the nominal level so this data set would benefit more from the t-interval.

4.3 Conclusion

In this chapter, we applied the confidence intervals that were developed in chapter 3 to data that was gathered by the Department of Biology at Stephen F. Austin State University. The data consists of tick cell counts that were infected in a bacterium.

Upon closer examination, we discovered that the t-interval is the best method in this scenario.

5 Concluding Remarks and Future Applications

The purpose of this research was to investigate interval estimates of the mean of a beta distribution by developing a novel statistical model. Through simulation studies, we compared the performance of the Wald, t-interval, and bootstrap intervals. The t-interval performed the best in all three cases we examined, even when the μ values were close to the boundaries, which produces the most skewed distributions. This information suggests that standard built-in tools can be used for data such as the biological example that was explored in chapter 4, rather than developing an elaborate model. Furthermore, it should be noted that the simulation studies could be considered at greater breadth. For instance, when ϕ is set to values of 30 and 50, the beta distribution begins to behave more symmetrically. Of course in this situation, the t-interval will perform well given the distribution are very bell-shaped. Investigating more situations when ϕ is small with smaller sample sizes could potentially reveal additional discrepancies between the methods explored in this study.

For future work, this study did not investigate the estimation properties for the MLE of μ and ϕ. While we do not suspect that the MLE for μ to be a biased estimator, we do suspect that ϕ might be a biased estimator. This estimate is directly inserted into Fisher's information, which helps control the margin of error for the Wald interval. If such a bias is discovered, future work might be to create an unbiased estimate for for ϕ and modify the Wald interval. By modifying the Wald interval, it could potentially have better properties and possibly outperform the t-interval in some occasions. Moreover, this manuscript provides the foundation of a confidence interval for the mean of the beta distribution. Future applications of this can be extended to the Wald interval for the difference of two means, $\mu_1 - \mu_2$, or the ratio of two means, $\frac{\mu_1}{\mu_2}$, for a beta distribution.

BIBLIOGRAPHY

[1] Daniel B. Pavanelo Eliane Esteves Lanissa A. Martins Veronika Urbanova Petr Kopacek Andrea C. Fogoca, Gessica Sousa and Siriel Daffre, *Tick immune system: What is known, the interconnections, the gaps, and the challenges*, Frontiers in Immunology **12** (2021), 1–17.

[2] Saralees Nadarajah Arjun K. Gupta, *Handbook of beta distribution and its applications*, Marcel Dekker, Inc, 2004.

[3] Bryn Brakefield, *Using saddlepoint methods approximations and likelihood-based methods to conduct statistical inference for the mean of a beta distribution.*, PlumX Metrics (2020), 1–76.

[4] Keith Dunnigan, *Confidence Interval Calculation for Binomial Proportions*, Midwest SAS Users Group (2008), 1–12.

[5] R. L. Smith G. A. Young, *Essentials of statistical inference*, Cambridge University Press, 2010.

[6] Roger L. Berger George Casella, *Statistical inference*, Brooks/cole, Cengage Learning, 2002.

[7] Derek Long, *The t-test*, 2003.

[8] Marlon E. Cobos Roman Ganta Des Foley Ram K. Raghavan, A. Townsend Peterson, *Current and future distribution of the lone star tick,* Amblyomma americanum *(l.) (acari: Ixodidae) in north america*, PLOS ONE (1986), 1–13.

[9] Allen T. Craig Robert V. Hogg, Joseph W. McKean, *Introduction to mathematical statistics*, vol. 7, Pearson Education, Inc., 2013.

[10] Jong Hae Kim Sang Gyu Kwak, *Central limit theorom: the cornerstone of modern statistics*, Korean Journal of Anesthesiology (2017), 144–156.

A Coverage Tables

Below are the results for the coverage probabilities of the three intervals we were interested in this study.

Counts (ϕ=10, n=5)	t-test	Wald	Bootstrap
μ=.1	0.9	0.814	0.802
μ=.2	0.927	0.831	0.83
μ=.3	0.941	0.841	0.839
μ=.4	0.942	0.834	0.84
μ=.5	0.951	0.839	0.845
μ=.6	0.947	0.838	0.836
μ=.7	0.941	0.833	0.836
μ=.8	0.932	0.834	0.825
μ=.9	0.9	0.809	0.801

Table A.1: Overall mean count when $\phi = 10$ and when the sample size is 5.

Widths (ϕ=10, n=5)	t-test	Wald	Bootstrap
μ=.1	0.204	0.132	0.126
μ=.2	0.278	0.176	0.172
μ=.3	0.323	0.202	0.201
μ=.4	0.346	0.214	0.216
μ=.5	0.357	0.219	0.222
μ=.6	0.347	0.215	0.216
μ=.7	0.324	0.200	0.202
μ=.8	0.280	0.176	0.174
μ=.9	0.202	0.132	0.124

Table A.2: Overall mean width when $\phi = 10$ and when the sample size is 5.

Counts (ϕ=10, n=20)	t-test	Wald	Bootstrap
μ=.1	0.928	0.916	0.916
μ=.2	0.946	0.923	0.926
μ=.3	0.944	0.926	0.925
μ=.4	0.948	0.929	0.929
μ=.5	0.950	0.93	0.933
μ=.6	0.946	0.925	0.927
μ=.7	.945	0.923	0.928
μ=.8	0.939	.923	0.922
μ=.9	0.929	0.916	0.913

Table A.3: Overall mean count when $\phi = 10$ and when the sample size is 20.

Widths (ϕ=10, n=20)	t-test	Wald	Bootstrap
μ=.1	0.082	0.076	0.075
μ=.2	0.111	0.101	0.102
μ=.3	0.128	0.116	0.117
μ=.4	0.137	0.124	0.125
μ=.5	0.140	0.126	0.128
μ=.6	0.137	0.124	0.125
μ=.7	0.128	0.116	.0117
μ=.8	0.111	0.101	0.101
μ=.9	0.082	0.076	0.075

Table A.4: Overall mean width when $\phi = 10$ and when the sample size is 20.

Counts (ϕ=10, n=50)	t-test	Wald	Bootstrap
μ=.1	0.9435	0.933	0.938
μ=.2	0.946	0.940	0.939
μ=.3	0.949	0.935	0.941
μ=.4	0.952	0.941	0.944
μ=.5	0.946	0.943	0.939
μ=.6	0.951	0.943	0.945
μ=.7	0.951	0.939	0.945
μ=.8	0.947	0.937	0.941
μ=.9	0.938	0.937	0.933

Table A.5: Overall mean count when $\phi = 10$ and when the sample size is 50.

Widths (ϕ=10, n=50)	t-test	Wald	Bootstrap
μ=.1	0.051	0.049	0.049
μ=.2	0.068	0.066	0.066
μ=.3	0.078	0.075	0.076
μ=.4	0.083	0.08	0.081
μ=.5	0.085	0.082	0.083
μ=.6	0.084	0.08	.081
μ=.7	0.078	0.075	0.076
μ=.8	0.068	0.066	0.066
μ=.9	0.051	0.049	0.049

Table A.6: Overall mean width when $\phi = 10$ and when the sample size is 50.

Counts (ϕ=30, n=5)	t-test	Wald	Bootstrap
μ=.1	0.934	0.847	0.823
μ=.2	0.947	0.841	0.837
μ=.3	0.945	0.844	0.837
μ=.4	0.945	0.848	0.84
μ=.5	0.948	0.848	0.837
μ=.6	0.951	0.845	0.836
μ=.7	0.95	0.841	0.845
μ=.8	0.943	0.842	0.829
μ=.9	0.928	0.841	0.827

Table A.7: Overall mean count when $\phi = 30$ and when the sample size is 5.

Widths (ϕ=30, n=5)	t-test	Wald	Bootstrap
μ=.1	0.124	0.080	0.077
μ=.2	0.168	0.106	0.105
μ=.3	0.192	0.121	0.119
μ=.4	0.206	0.129	0.128
μ=.5	0.211	0.133	0.131
μ=.6	0.207	0.130	0.129
μ=.7	0.194	0.120	0.121
μ=.8	0.167	0.105	0.104
μ=.9	0.124	0.08	0.077

Table A.8: Overall mean width when $\phi = 30$ and when the sample size is 5 .

Counts (ϕ=30, n=20)	t-test	Wald	Bootstrap
μ=.1	0.94	0.923	0.919
μ=.2	0.951	0.930	0.932
μ=.3	0.953	0.927	0.931
μ=.4	0.951	0.928	0.929
μ=.5	0.951	0.926	0.932
μ=.6	0.952	0.924	0.93
μ=.7	0.949	0.931	0.928
μ=.8	0.948	0.928	0.927
μ=.9	0.941	0.923	0.923

Table A.9: Overall mean count when $\phi = 30$ and when the sample size is 20.

Widths (ϕ=30, n=20)	t-test	Wald	Bootstrap
μ=.1	0.049	0.045	0.045
μ=.2	0.066	0.06	0.061
μ=.3	0.0762	0.069	0.070
μ=.4	0.081	0.074	0.074
μ=.5	0.083	0.076	0.076
μ=.6	0.081	0.074	0.074
μ=.7	0.076	0.069	0.070
μ=.8	0.066	0.061	0.061
μ=.9	0.050	0.046	0.045

Table A.10: Overall mean width when $\phi = 30$ and when the sample size is 20.

Counts (ϕ=30, n=50)	t-test	Wald	Bootstrap
μ=.1	0.943	0.94	0.924
μ=.2	0.949	0.942	0.929
μ=.3	0.943	0.942	0.921
μ=.4	0.949	0.936	0.928
μ=.5	0.952	0.943	0.944
μ=.6	0.953	0.94	0.930
μ=.7	0.951	0.941	0.929
μ=.8	0.950	0.94	0.928
μ=.9	0.95	0.94	0.943

Table A.11: Overall mean count when $\phi = 30$ and when the sample size is 50.

Widths (ϕ=30, n=50)	t-test	Wald	Bootstrap
μ=.1	0.49	0.029	0.045
μ=.2	0.066	0.039	0.061
μ=.3	0.076	0.045	0.070
μ=.4	0.081	0.048	0.075
μ=.5	0.051	0.049	0.049
μ=.6	0.081	0.48	0.074
μ=.7	0.080	0.045	0.069
μ=.8	0.066	0.039	0.061
μ=.9	0.03	0.029	0.029

Table A.12: Overall mean width when $\phi = 30$ and when the sample size is 50.

Counts (ϕ=50, n=5)	t-test	Wald	Bootstrap
μ=.1	0.938	0.856	0.828
μ=.2	0.946	0.845	0.836
μ=.3	0.952	0.846	0.839
μ=.4	0.943	0.858	0.831
μ=.5	0.952	0.856	0.840
μ=.6	0.949	0.850	0.839
μ=.7	0.946	0.851	0.835
μ=.8	0.946	0.846	0.835
μ=.9	0.939	0.853	0.832

Table A.13: Overall mean count when $\phi = 50$ and when the sample size is 5.

Widths (ϕ=50, n=5)	t-test	Wald	Bootstrap
μ=.1	0.097	0.063	0.06
μ=.2	0.131	0.083	0.081
μ=.3	0.15	0.095	0.094
μ=.4	0.159	0.102	0.099
μ=.5	0.163	0.104	0.101
μ=.6	0.161	0.101	0.1
μ=.7	0.149	0.094	0.093
μ=.8	0.131	0.083	0.082
μ=.9	0.097	0.064	0.06

Table A.14: Overall mean width when $\phi = 50$ and when the sample size is 5.

Counts (ϕ=50, n=20)	t-test	Wald	Bootstrap
μ=.1	0.943	0.925	0.924
μ=.2	0.946	0.927	0.926
μ=.3	0.948	0.929	0.926
μ=.4	0.949	0.930	0.931
μ=.5	0.949	0.937	0.928
μ=.6	0.954	0.929	0.935
μ=.7	0.953	0.934	0.930
μ=.8	0.949	0.926	0.929
μ=.9	0.945	0.928	0.923

Table A.15: Overall mean count when $\phi = 50$ and when the sample size is 20.

Widths (ϕ=50, n=20)	t-test	Wald	Bootstrap
μ=.1	0.039	0.036	0.035
μ=.2	0.052	0.047	0.047
μ=.3	0.059	0.054	0.054
μ=.4	0.063	0.058	0.058
μ=.5	0.065	0.059	0.059
μ=.6	0.063	0.058	0.058
μ=.7	0.059	0.054	0.054
μ=.8	0.052	0.047	0.047
μ=.9	0.039	0.036	0.035

Table A.16: Overall mean width when $\phi = 50$ and when the sample size is 20.

Counts (ϕ=50, n=50)	t-test	Wald	Bootstrap
μ=.1	0.948	0.947	0.94
μ=.2	0.948	0.943	0.939
μ=.3	0.949	0.942	0.942
μ=.4	0.949	0.943	0.941
μ=.5	0.947	0.946	0.939
μ=.6	0.953	0.943	0.945
μ=.7	0.949	0.944	0.941
μ=.8	0.954	0.939	0.944
μ=.9	0.946	0.945	0.939

Table A.17: Overall mean count when $\phi = 50$ and when the sample size is 50.

Widths (ϕ=50, n=50)	t-test	Wald	Bootstrap
μ=.1	0.024	0.023	0.023
μ=.2	0.032	0.031	0.031
μ=.3	0.036	0.035	0.035
μ=.4	0.039	0.038	0.038
μ=.5	0.040	0.038	0.038
μ=.6	0.039	0.038	0.038
μ=.7	0.036	0.035	0.035
μ=.8	0.032	0.031	0.031
μ=.9	0.024	0.023	0.023

Table A.18: Overall mean width when $\phi = 50$ and when the sample size is 50.

B R Code

Here is the R code that was used to perform the simulations in this study.

```
library(stats4)
library(boot)
beta.mle<-function(x){logl.un<-function(mu,s){
     -1*sum(dbeta(x,shape1=mu*s,shape2=(1-mu)*s,log=TRUE))
  }
mystart<-list(mu=mean(x),s=(mean(x)*(1-mean(x)))/var(x)-1)
fit.un<-mle(logl.un,start=mystart,
method = "L-BFGS-B", lower = c(0.00001, 0.00001),upper=c(.99999,Inf))
  return(fit.un)
}

mysummary<-function(x){sderror<-sd(x)
xmax<-max(x)
result<-c(xbar,sderror,xmax)                return(result)}
mymean<-function(x,idx){
  mydat<-x[idx]
  mean(mydat)
}

#Initialize simulation parameters
mu=.9
phi=26
n=11
```

```
n.sims=10000
level=.95

#Initialize simulation result objects
count.t<-c()
count.wald<-c()
count.boot<-c()
width.t<-c()
width.wald<-c()
width.boot<-c()

#Perform Simulation for above scenario
se.results<-c()
mu.resslts<-c()
phi.results<-c()

#simulate data set
for(i in 1:n.sims)
{dat<-rbeta(n,shape1=mu*phi,shape2=(1-mu)*phi)
  ttest.ci<-t.test(dat,conf.level=level)$conf.int
  mles<-beta.mle(dat)
  tryCatch({
 },error=function(e){cat("Warning:␣Skipping␣Row",i,"\n")})

#bootstrap interval
bootstrap<-boot(dat,mymean,R=1000)
bootci<-boot.ci(boot.out=bootstrap,type="perc")$percent[4:5]
```

```
mu.hat<-coef(mles)[1]
phi.hat<-coef(mles)[2]

#Fisher's Information

a=phi.hat^2*trigamma(mu.hat*phi.hat)
+phi.hat^2*trigamma((1-mu.hat)*phi.hat)

b=phi.hat*mu.hat*trigamma(mu.hat*phi.hat)
-phi.hat*(1-mu.hat)*trigamma((1-mu.hat)*phi.hat)

d=mu.hat^2*trigamma(mu.hat*phi.hat)+(1-mu.hat)^2
*trigamma((1-mu.hat)*phi.hat)-trigamma(phi.hat)

if(a*d-b^2  >0){wald.se<-sqrt(abs(d/(a*d-b^2)))*(1/sqrt(n))}
if(a*d-b^2<0){wald.se<-sqrt(1/a)*(1/sqrt(n))}
wald.ci<-c(mu.hat-1.96*wald.se,mu.hat+1.96*wald.se)

se.results[i]<-wald.se
mu.reslts[i]<-mu.hat
phi.results[i]<-phi.hat

#Count and check if mu is inside or not.
count.t[i]<-ifelse(mu<ttest.ci[2] & mu>ttest.ci[1],1,0)
count.wald[i]<-ifelse(mu<wald.ci[2] & mu>wald.ci[1],1,0)
count.boot[i]<-ifelse(mu<bootci[2] & mu>bootci[1],1,0)
width.t[i]<-ttest.ci[2]-ttest.ci[1]
width.wald[i]<-wald.ci[2]-wald.ci[1]
width.boot[i]<-bootci[2]-bootci[1]
}

#compute coverages
counts<-cbind(count.t,count.wald,count.boot)
widths<-cbind(width.t,width.wald,width.boot)
result1<-apply(counts,2,mean)
result2<-apply(widths,2,mean)
result3<-apply(widths,2,mysummary)
result1
result2
result3
```

C Coverage Plots

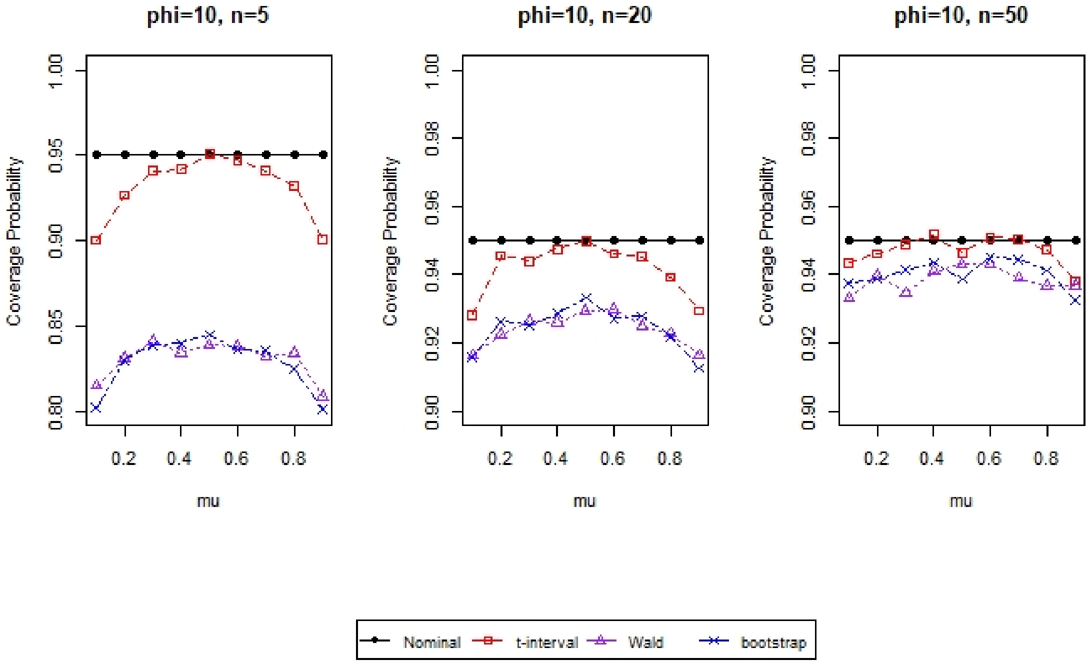

Figure C.1: Phi=10, n=5,20,50

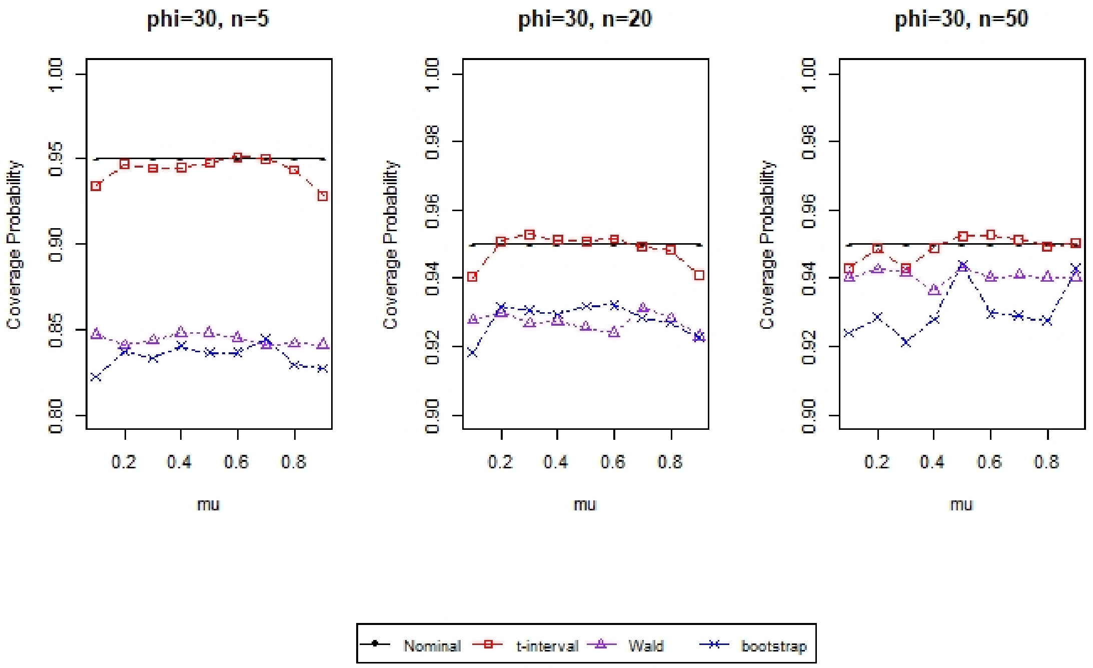

Figure C.2: Phi=50, n=5,20,50

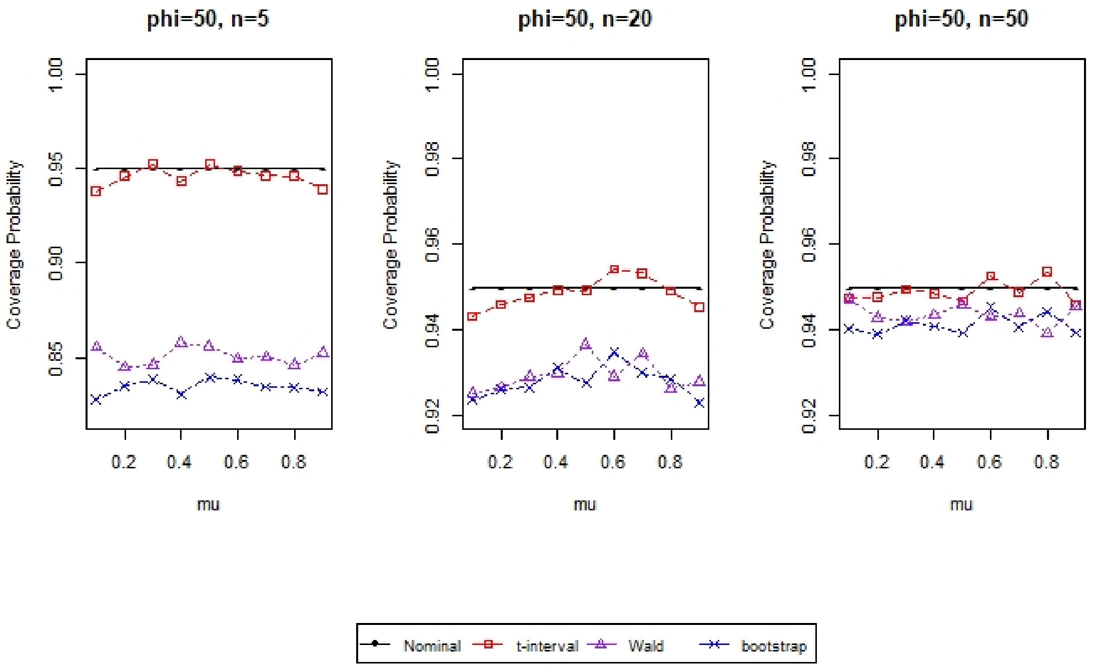

Figure C.3: Phi=50, n=5,20,50

VITA

Sean Rangel was born in Plano, Texas on the 3rd of October, 1996. He attended Concordia University Texas in Austin, Texas. There, he received a Bachelor's of Science degree in Biology in April of 2019. Later that year, he began working towards a Master's degree at Stephen F. Austin State University in Nacogdoches, Texas, USA. He is expected to graduate in December of 2021.

Permanent Address: PO Box XXX SFA Station
Nacogdoches, TX 75962

The style manual used in this thesis is A Manual For Authors of Mathematical Papers published by the American Mathematical Society.

This thesis was prepared by Sean Rangel using LaTeX.