5-2020

Using Saddlepoint Approximations and Likelihood-Based Methods to Conduct Statistical Inference for the Mean of the Beta Distribution

Bryn Brakefield
Stephen F. Austin State University, brakefiebm@jacks.sfasu.edu

Using Saddlepoint Approximations and Likelihood-Based Methods to Conduct Statistical Inference for the Mean of the Beta Distribution

Creative Commons License

USING SADDLEPOINT APPROXIMATIONS AND LIKELIHOOD-BASED METHODS TO CONDUCT STATISTICAL INFERENCE FOR THE MEAN OF THE BETA DISTRIBUTION

by

BRYN BRAKEFIELD, 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

May 2020

USING SADDLEPOINT APPROXIMATIONS AND LIKELIHOOD-BASED METHODS TO CONDUCT STATISTICAL INFERENCE FOR THE MEAN OF THE BETA DISTRIBUTION

by

BRYN BRAKEFIELD, B.S.

APPROVED:

Jacob Turner, Ph.D., Thesis Director

Gregory K. Miller, Ph.D., Committee Member

Jeremy Becnel, Ph.D., Committee Member

Emiliano Giudici, Ph.D., Committee Member

Pauline M. Sampson, Ph.D.
Dean of Research and Graduate Studies

ABSTRACT

The prevalence of conducting statistical inference for the mean of the beta distribution has been rising in various fields of academic research, such as in immunology that analyzes proportions of rare cell population subsets. For our purposes, we will address this statistical inference problem by using likelihood-based applications to hypothesis testing, along with a relatively new statistical method called saddlepoint approximations. Through simulation work, we will compare the performance of these statistical procedures and provide both the statistical and scientific communities with recommendations on best practices.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following individuals who have assisted me on this endeavor:

My thesis advisor, Dr. Jacob Turner, for all of the time you have devoted in helping me complete this research and your immense patience in looking at all of the rough drafts. I never would have realized that I wanted to continue in my education without this experience. Now we can finally say that...we have made it through Mordor!

My committee members, Dr. Gregory Miller, Dr. Jeremy Becnel, and Dr. Emiliano Giudici, for their profound guidance and suggestions.

The members of the faculty of the Department of Mathematics and Statistics at Stephen F. Austin State University for their unwavering support.

My fellow classmates, as it has been such an awarding experience going through graduate school with you all.

Don and Mara Balusek and the Department of Mathematics and Statistics at Stephen F. Austin State University for their financial support.

My friends and family for their unparalleled and unbridled support. Thank you all for keeping me sane!

CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The beta distribution is a highly flexible probability model in regards to random phenomena that are bounded above and below on a continuous interval. This distribution was first introduced by one of the founding fathers of statistical science, Karl Pearson, and was often distinguished as the Pearson Type I distribution in earlier literature. As is the case for many popular distributions, there is a plethora of parameterizations of the beta model. For statistical inference purposes, this manuscript will use the "mean/precision" parameterization.

Upon reparameterization, the probability density function (PDF) for a random variable X that follows the beta distribution is

$$f(x;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{\mu\phi-1}(1-x)^{(1-\mu)\phi-1}, \quad 0 < x < 1, \tag{1.1}$$

where $0 < \mu = \mathbb{E}(X) < 1$ is the mean of X and the variance-controlling parameter $\phi > 0$ is the precision. The variance of X is $\text{Var}(X) = \frac{\mu(1-\mu)}{\phi+1}$, which elucidates the fact that for any fixed value of μ, the variance decreases as ϕ increases [7].

As the parameters μ and ϕ vary, the beta distribution manifests different types of shapes. To illustrate the flexibility of the beta model, Figure 1.1 provides the PDFs for specific combinations of these parameters. The PDF can become the standard uniform distribution ($\mu = 0.5, \phi = 2$), be strictly decreasing ($\mu = 0.1, \phi = 10$), skewed-right ($\mu = 0.25, \phi = 10$), or symmetric ($\mu = 0.5, \phi = 20$).

This versatility of the beta distribution is applicable in a myriad of settings, such as immunology research, quality control, multivariate analysis of variance (MANOVA), and soil composition analysis (Butler [4], Gupta and Nadarajah [9]). The correspondence between this distribution and flow cytometry (FCM) data from immunology research serves as the main motivation for this manuscript.

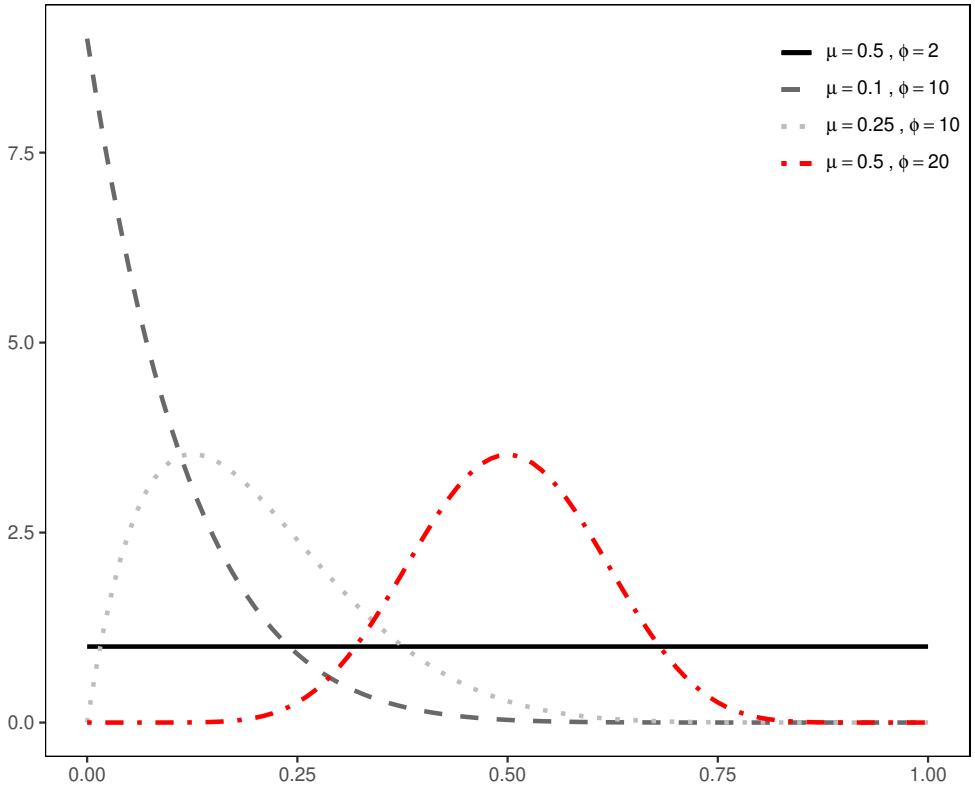

Figure 1.1: Graphs of PDFs for the Beta Distribution

FCM is an automated process for identifying, counting, and sorting through specific cells of interest from an aggregation of cells in a given sample. These cells are stained with various phosphorous fluorescence tags that bind to specific molecules, where these molecules provide information about the cells in a functional capacity, both within and on the cell itself. For example, the lymphocyte cell population has numerous sub-types, including T-cells, B-cells, and natural killers (NK), that serve different roles in the immune system's response to various perturbations, such as getting a virus, obtaining a flu vaccine, or taking a drug [13].

The primary instrument in the FCM technique is the flow cytometer. It measures the light intensity emitted from the fluorescence as each cell individually passes through the machine. This light intensity is quantified and software is then used to

identify the cell type for each individual cell. Once the cell populations have been defined on the roughly 100000 cells in the sample, the proportion, or percentage, of cells is recorded for each cell population. This process is then repeated, as many samples are needed for the study.

In accordance with the previous process, Figure 1.2 provides a panel of FCM data for four different cell populations from a study of cynomolgus macaques that were challenged with Mycobacterium tuberculosis [8]. CD20 is a marker for B-cells while the remaining three differentiate a few sub-populations of T-cells.

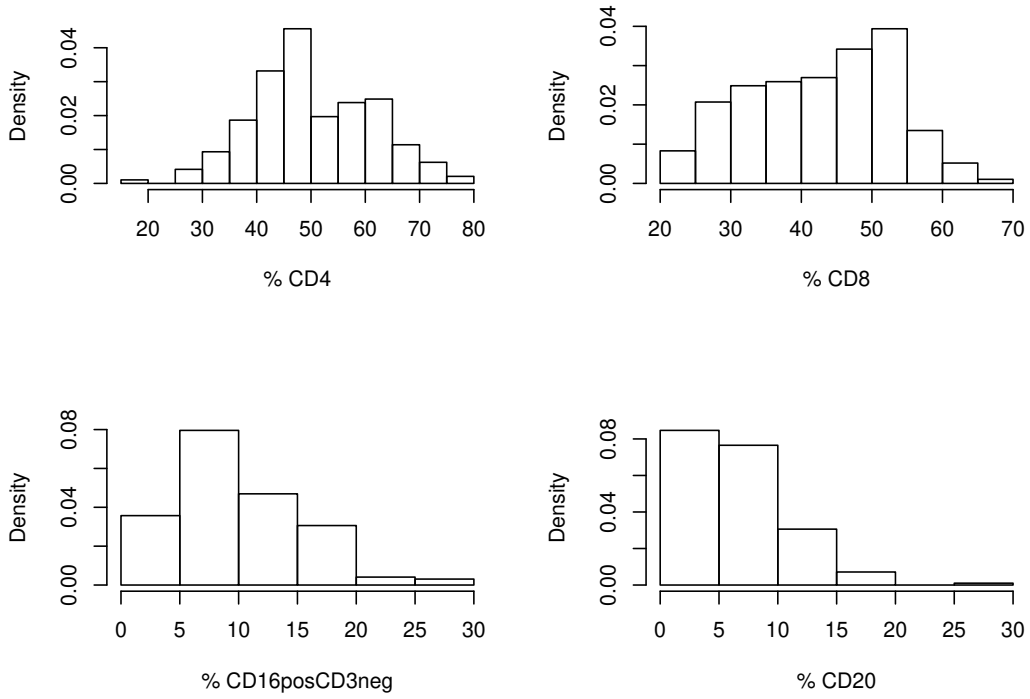

Figure 1.2: FCM Data

These histograms resemble some of the different forms of the beta distribution provided in Figure 1.1, although on the scale of $(0, 100)$. This can easily be transformed to a proportion on the scale of $(0, 1)$ through simple division. Due to this

transformation and the previous resemblance, there is a growing interest in conducting statistical inference for the mean of the beta distribution. To our knowledge, there does not exist any literature directly coinciding with this type of statistical inference problem. Therefore, this manuscript attempts to address this question of interest.

Our approach to addressing the previous question resembles standard statistical methodology. For the purposes of this manuscript, we will assume that ϕ is known and μ is the parameter of interest. While having a known value of ϕ is eminently unlikely, we are also operating with the assumption that we are able to acquire an estimate of ϕ from previous studies, such as pilot studies in immunology research. Essentially, this manuscript serves as the base case for our statistical inference problem in the same capacity as the Z-test does for conducting statistical inference for the mean of the normal distribution assuming the population standard deviation, σ, is known in advance. As for the subsequent case, where ϕ is not known in advance, the natural approach is to estimate the value of ϕ through the actual data itself. The Student's t-test used this same approach in solving the analogous problem in terms of the normal distribution, but required formulating an innovative strategy to derive the distribution of the t test statistic. As this is also true for the original case with the beta distribution where ϕ is unknown, we will not address this testing procedure in this manuscript.

Now, with the assumption of ϕ being known, we will proceed to derive different types of statistical tests using traditional large sample likelihood theory and a relatively new statistical method called saddlepoint approximations. From there, we will investigate different aspects of these tests, such as their Type I error rates and statistical power via simulations conducted with the statistical software R. In addition to hypothesis testing, we will also explore the performance of interval estimation using coverage probabilities and interval width. Through this simulation work, we will be able to obtain empirical evidence as to which tests outperform others under various

scenarios and eventually summarize our results in order to provide insight to both the statistical and scientific communities. We will also offer some examples of direct applications of our methods to real world problems, which include a FCM data set for hypothesis testing and sample size determination in terms of good study design.

2 STATISTICAL THEORY WITH ONE-PARAMETER EXPONENTIAL FAMILIES

This chapter introduces a general family of probability models known as one-parameter exponential families. These families have a vast amount of relevant, mathematical properties that make statistical inference problems computationally appeasing and straightforward to implement in practice. Throughout this chapter, we discuss the traditional statistical inference methods through maximum likelihood estimation, sufficient statistics, and confidence intervals along with the necessary connections to how these procedures are simplified in relation to exponential families.

2.1 Definition and Examples of One-Parameter Exponential Families

Let X be a random variable and suppose that the distribution of X is a member of a parametric class of probability density functions (PDFs) or probability mass functions (PMFs) with parameter ξ of the form

$$f(x;\xi) = h(x)\exp\{\theta(\xi)t(x) - c(\xi)\}, \quad x \in \mathcal{X}, \xi \in \Xi, \tag{2.1}$$

where $\mathcal{X}$ and Ξ are the sample and parameter space, respectively. Furthermore, $h(x) \geq 0$, $t(x)$, and $\mathcal{X}$ cannot depend on ξ while $\theta(\xi)$ and $c(\xi)$ cannot depend on x. Then, the parametric class $\{f(\,\cdot\,;\xi)|\xi \in \Xi\}$ is a one-parameter exponential family of distributions [4].

For an exponential family, the mapping $\xi \mapsto \theta$ is one-to-one and the notation may be simplified by working with the canonical parameterization of ξ, $\theta(\xi) = \theta$, and expressing the exponential family as

$$f(x;\theta) = h(x)\exp\{\theta t(x) - c(\theta)\}, \quad x \in \mathcal{X}, \theta \in \Theta, \tag{2.2}$$

where $c(\theta)$ is $c(\xi(\theta))$ [4].

Example 2.1. The PMF of the Poisson distribution, given as

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad 0 \leq \lambda < \infty, x = 0, 1, \ldots, \tag{2.3}$$

is of exponential family form with canonical parameter $\theta = \ln\{\lambda\}$ if we define $h(x) = I_{\{0,1,\ldots\}}(x)/x!$ with the indicator function I, $t(x) = x$, and $c(\theta) = \lambda = \exp\{\theta\}$ upon reparameterization. By substituting these functions into (2.2), we obtain

$$f(x;\theta) = \frac{I_{\{0,1,\ldots\}}(x)}{x!} \exp\{\theta x - \exp\{\theta\}\}. \tag{2.4}$$

Example 2.2. The PDF of the beta distribution with ϕ as a known constant is also of exponential family form with canonical parameter $\theta = \mu$ if we define $h(x) = \Gamma(\phi) I_{(0,1)}(x)(1-x)^{\phi-1}/x$, $t(x) = \phi \ln\{x/(1-x)\}$, and
$c(\mu) = \ln\{\Gamma(\mu\phi)\} + \ln\{\Gamma((1-\mu)\phi)\}$.

2.2 Sufficiency

The class of exponential families produces a great deal of mathematical framework in regards to statistical inference. These results are contingent on using what is called a sufficient statistic. Now, to sufficiently describe this type of statistic, we first provide a brief introduction to sufficient statistics, along with their relation to exponential families, and then an account of some of their distributional properties.

2.2.1 Sufficient Statistics

Suppose that we have a random sample $X_1, \ldots, X_n$ from a probability model whose PDF or PMF is $f(x_i;\theta)$ for each X_i and define $\boldsymbol{X} = (X_1, \ldots, X_n)^T$. Then, the joint distribution of $\boldsymbol{X}$ is defined as

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta). \tag{2.5}$$

A statistic, $T(\boldsymbol{X})$, is a sufficient statistic for the parameter θ if the conditional distribution of the sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on θ. The idea behind this definition is relatively straightforward. From a sample $\boldsymbol{X}$, a statistic should contain all of the necessary information about θ. Since the statistic typically, but not always, maps the sample from n dimensions down to a single value for one-parameter models, then this is often referred to as a data reduction strategy. Additionally, the reduction also highlights the idea that if working with two different data sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ that yield the same sufficient statistics $T(\boldsymbol{X}) = T(\boldsymbol{Y})$, the same conclusions should be made about θ.

Determining whether a statistic is sufficient is typically done by way of the Factorization Theorem [5], which states that $T(\boldsymbol{X})$ is sufficient for θ if and only if the joint density function for the random sample can be factored in the following way:

$$f(\boldsymbol{x};\theta) = g(T(\boldsymbol{x})|\theta)h(\boldsymbol{x}). \tag{2.6}$$

If the probability model in question is a member of the exponential family with canonical parameter θ, then we will show that proving $T(\boldsymbol{X}) = \sum_{i=1}^{n} t(X_i)$ is a sufficient statistic is a fairly straightforward exercise [5]:

By incorporating the functional form of exponential families into (2.5), the joint density function can be expressed as

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \left\{ \prod_{i=1}^{n} h(x_i) \right\} \exp\left\{ \theta \sum_{i=1}^{n} t(x_i) - nc(\theta) \right\}. \tag{2.7}$$

If we define $h(\boldsymbol{x}) = \prod_{i=1}^{n} h(x_i)$ and $g(T(\boldsymbol{x})|\theta) = \exp\left\{\theta \sum_{i=1}^{n} t(x_i) - nc(\theta)\right\}$, then the sufficiency of $\sum_{i=1}^{n} t(x_i)$ is clear by the factorization theorem.

2.2.2 Distribution Theory for Sufficient Statistics

Under the same conditions as in Section 2.2.1, consider the canonical sufficient statistic $T(\boldsymbol{X})$. Since $T(\boldsymbol{X})$ is a function of random variables, then $T(\boldsymbol{X})$ is also a

random variable and for notational purposes, we will denote this random variable as $T = T(\boldsymbol{X})$. The PDF or PMF of T has the form

$$f(t;\theta) = R(t)\exp\{\theta t - nc(\theta)\}, \quad t \in \mathcal{X}_T, \theta \in \Theta, \tag{2.8}$$

where $\mathcal{X}_T$ and Θ are the sample and parameter space, respectively, for some function $R(t)$ and neither $\mathcal{X}_T$ nor $R(t)$ depend on θ [10]. From the definition of exponential families, we can immediately see that the distribution of the sufficient statistic T is of exponential family form.

Obtaining the function $R(t)$ requires a multivariate transformation followed by a $(n-1)$-tuple multiple integral. In many cases, this can be done with relative ease, such as the case for the normal distribution. However, there are some cases where the integral does not have a closed form solution. This is where an alternative approach that characterizes the random variable through its moment generating function (MGF) rather than directly through its PDF or PMF comes into play.

The MGF of T is defined by

$$M_T(s) = \mathbb{E}(e^{sT}) = \int_{-\infty}^{\infty} e^{st} f(t)\, dt \tag{2.9}$$

over values of s for which the integral converges [4]. Furthermore, we presume that $M_T(s)$ converges over an open neighborhood of 0, which we denote by (a, b), and that this interval is the largest such neighborhood of convergence. This idea stems from the fact that if the MGF exists in an open neighborhood of 0, then the distribution of T is uniquely determined [5]. That is, for two arbitrary random variables X and Y all of whose moments exist, let $F_X(x)$ and $F_Y(y)$ be two CDFs. If the MGFs exist and $M_X(u) = M_Y(u)$ for all u in some neighborhood of 0, then $F_X(v) = F_Y(v)$ for all v.

With the definition of the MGF, we can also define the cumulant generating function (CGF) of T as

$$K_T(s) = \ln\{M_T(s)\}, \quad s \in (a, b), \tag{2.10}$$

provided that $M_T(s)$ exists on the same interval [4]. Upon examining the Maclaurin series for the CGF, we are able to derive the first two moments of T, which are $K'_T(0) = \mathbb{E}(T)$ and $K''_T(0) = \text{Var}(T)$. Now, by recognizing that T is a summation of independent, identically distributed (i.i.d.) random variables, we have an approximation to the distribution of T dictated by the Central Limit Theorem. In the context of T, this theorem gives us that for sufficiently large sample sizes, the distribution of T is approximately normal with mean $\mathbb{E}(T) = K'_T(0)$ and variance $\text{Var}(T) = K''_T(0)$.

In accordance with all of the preceding properties for the sufficient statistic T regarding its MGF and CGF, suppose that our random sample is now from an exponential family under the canonical parameterization with parameter θ. Due to the functional form of exponential families, the MGF of T can now be expressed as

$$M_T(s) = \exp\{n[c(s+\theta) - c(\theta)]\}, \quad s \in \mathcal{S}_\theta, \tag{2.11}$$

where $\mathcal{S}_\theta = \{s | s+\theta \in \Theta\}$. After applying a logarithmic transformation to $M_T(s)$, we have that the CGF of T is

$$K_T(s) = n[c(s+\theta) - c(\theta)], \quad s \in \mathcal{S}_\theta. \tag{2.12}$$

Due to properties of exponential families, K_T is differentiable and the previous expression of $K_T(s)$ allows us to write the rth derivative of K_T as

$$K_T^{(r)}(s) = n\frac{d^r}{d\theta^r}c(s+\theta). \tag{2.13}$$

Thus, $\mathbb{E}(T) = K'_T(0) = nc'(\theta)$ and $\text{Var}(T) = K''_T(0) = nc''(\theta)$. The reformulated distributional statement for T derived from the Central Limit Theorem now states that the distribution of T is approximately normal with mean $\mathbb{E}(T) = nc'(\theta)$ and variance $\text{Var}(T) = nc''(\theta)$ for sufficiently large sample sizes.

Now, since the beta distribution for the case where ϕ is known is a member of the one-parameter exponential family, we can find these same properties for the

canonical sufficient statistic for μ, $T = \phi \sum_{i=1}^{n} \ln\{X_i/(1-X_i)\}$. By using that fact that $c(\mu) = \ln\{\Gamma(\mu\phi)\} + \ln\{\Gamma((1-\mu)\phi)\}$, the CGF of T is

$$\begin{aligned} K_T(s) = n[&\ln\{\Gamma((s+\mu)\phi)\} + \ln\{\Gamma((1-(s+\mu))\phi)\} - \ln\{\Gamma(\mu\phi)\} \\ &- \ln\{\Gamma((1-\mu)\phi)\}]. \end{aligned} \tag{2.14}$$

The first and second derivative of K_T are, respectively,

$$K_T'(s) = n\phi[\gamma((s+\mu)\phi) - \gamma((1-(s+\mu))\phi)], \tag{2.15}$$

where $\gamma(x) = \frac{d}{dx}\ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is known as the digamma function, and

$$K_T''(s) = n\phi^2[\gamma'((s+\mu)\phi) + \gamma'((1-(s+\mu))\phi)], \tag{2.16}$$

where $\gamma'(\cdot)$ is the trigamma function. Therefore,

$$\mathbb{E}(T) = K_T'(0) = n\phi[\gamma(\mu\phi) - \gamma((1-\mu)\phi)] \tag{2.17}$$

and

$$\mathrm{Var}(T) = K_T''(0) = n\phi^2[\gamma'(\mu\phi) + \gamma'((1-\mu)\phi)]. \tag{2.18}$$

2.3 Point Estimation with Maximum Likelihood Estimators

There is a plethora of methods for obtaining estimators of a parameter of interest θ in an exponential family. The method of maximum likelihood is an eminent technique for deriving estimators and the results are referred to as maximum likelihood estimators (MLEs). In this section, we will provide the derivation of MLEs through maximum likelihood estimation, which will exemplify its coincidence with sufficient statistics, along with the asymptotic properties of MLEs that will be used in subsequent sections to formulate hypothesis tests and their corresponding confidence intervals.

2.3.1 Maximum Likelihood Estimation

Suppose that we have a random sample $X_1, \ldots, X_n$ from an exponential family with canonical parameter θ and define $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ to be the random vector representing the unobserved data. Furthermore, suppose that we are interested in obtaining an estimate of θ based on the observed values of $\boldsymbol{X}$ denoted by $\boldsymbol{x}$. The principle idea behind maximum likelihood estimation is choosing the value of θ that aligns with the observed data points in such a way that it produces the "most likely" situation for observing said data. This "most likely" value of θ is denoted by $\hat{\theta}$ and is, in fact, the maximum likelihood estimator (MLE).

The derivation of the MLE under the previously defined conditions is contingent on the likelihood function $\mathcal{L}(\theta; \boldsymbol{x})$. The likelihood function $\mathcal{L}(\theta; \boldsymbol{x})$ is defined by

$$\mathcal{L}(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta) \tag{2.19}$$

and is essentially the joint density function of $\boldsymbol{X}$ evaluated at the observed values $\boldsymbol{x}$, which implies it is only a function of the unknown value of θ. Note that when working with the likelihood function, we reduce the notation by writing $\mathcal{L}(\theta) = \mathcal{L}(\theta; \boldsymbol{x})$. This emphasizes the fact that the values of $\boldsymbol{x}$ are known, or that the data set has been observed, and treated as a constant in the likelihood function. Now, by using the exponential family form of $f(x_i; \theta)$ in (2.2) for each $i = 1, \ldots, n$, we can express the likelihood function $\mathcal{L}(\theta)$ as

$$\mathcal{L}(\theta) = \left\{ \prod_{i=1}^{n} h(x_i) \right\} \exp\left\{ \theta \sum_{i=1}^{n} t(x_i) - nc(\theta) \right\}. \tag{2.20}$$

Since the maximum likelihood approach encapsulates the idea of obtaining the value of θ the makes the observed data the "most likely" to see, then an estimate for θ should be chosen so that $\mathcal{L}(\theta)$ is maximized. If such a value exists, then we formally define the MLE of θ, $\hat{\theta}$, to be the value of θ such that $\mathcal{L}(\hat{\theta}) \geq \mathcal{L}(\theta)$ for

all possible values of θ. It is typically mathematically advantageous to consider the log-likelihood function, $l(\theta) = \ln\{\mathcal{L}(\theta)\}$. Since the natural logarithm function is monotonically increasing, then the MLE may be obtained through maximizing the log-likelihood function. Additionally, if $l(\theta)$ is differentiable, then $\hat{\theta}$ is obtained by solving the normal equation

$$l'(\theta) = 0. \tag{2.21}$$

Applying the natural logarithmic transformation to (2.7) and then differentiating yields the following normal equation

$$nc'(\theta) = \sum_{i=1}^{n} t(X_i) = T. \tag{2.22}$$

This form of the normal equation illustrates the well known fact that the MLE $\hat{\theta}$ is a function of the data only through the sufficient statistic T. Now, there are some cases where $\hat{\theta}$ has a closed-form solution as the inverse of $c'(\theta)$ is easily obtained. However, in many situations, $\hat{\theta}$ has to be solved numerically through a root-finding algorithm.

Example 2.3. The likelihood equation in the form of (2.22) for the beta distribution is

$$n\phi[\gamma(\mu\phi) - \gamma((1-\mu)\phi)] = \phi \sum_{i=1}^{n} \ln\left\{\frac{X_i}{1-X_i}\right\}. \tag{2.23}$$

From this, we identify $T = \phi \sum_{i=1}^{n} \ln\{X_i/(1-X_i)\}$ as the canonical sufficient statistic for μ. Due to the digamma functions present in this equation, there does not exist a closed-form solution for the MLE $\hat{\mu}$. Therefore, we must obtain a numerical approximation of $\hat{\mu}$ by using numerical methods, such as Newton's method. Gupta and Nadarajah [9] elaborate on the different numerical techniques that can be used to find $\hat{\mu}$. Also, Minka [12] describes the various types of routines in terms of the dirichlet distribution, which is a multivariate generalization of the beta distribution.

2.3.2 Asymptotic Properties of MLEs

As previously discussed in Section 2.3.1, MLEs maximize the likelihood function $\mathcal{L}(\theta)$, or equivalently the log likelihood function $l(\theta)$, and they are also a function of the sufficient statistic T, which implies that MLEs are random variables before the random sample is collected. Determining their distributions, means, and variances are extremely beneficial for statistical hypothesis tests and interval estimation. While the distribution of an MLE can be difficult to obtain or is unknown in many situations, we have key asymptotic results for the sampling distribution of MLEs. Before providing those results, we first introduce an important metric used in large sample theory known as Fisher's information.

Fisher's information for a random sample of n observations, denoted by $i(\theta)$, is defined as

$$i(\theta) = \mathbb{E}\{l'(\theta; \boldsymbol{X})^2\}. \tag{2.24}$$

Computing this information measure can be quite taxing in many statistical problems and it is often helpful to express it in a different form when applicable. Under certain regularity conditions that are enumerated in [5], which exponential families possess, the above equation can be expressed as

$$i(\theta) = -\mathbb{E}\{l''(\theta; \boldsymbol{X})\}. \tag{2.25}$$

The previous expression for Fisher's information provides a more intuitive insight as to what "information" this metric actually brings to the table. The Fisher Information measure is essentially a weighted mean of the second derivative of the log-likelihood function, where the weights are given by the PDF $f(\boldsymbol{x}; \theta)$ [10]. So, in a sense, we are seeing the amount of concavity that the log-likelihood function has on average. If the concavity is quite large, then the MLEs obtained from random sampling are more consistent; that is, they do not vary as much as compared to other scenarios

where the concavity is smaller. This interpretation alludes to the fact that Fisher's information tells us something about the uncertainty, or the variability, that the likelihood function has under random sampling. Its most famous application is in the Cramér-Rao Inequality Theorem, which provides the Cramér-Rao Lower Bound (CRLB). It states that the variance for each unbiased estimator of θ cannot be smaller than $1/i(\theta)$.

With Fisher's information $i(\theta)$ defined and intuitively explained, we can now introduce the key distributional result of MLEs. This result, under sufficient regularity conditions enumerated in [5], can be summarized into the following distributional statement:

$$\sqrt{i(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1). \tag{2.26}$$

By recognizing that the quantity on the left-hand side of the arrow is simply the Z-score of the MLE for a random sample with n observations, $\hat{\theta}_n$, and this Z-score follows the standard normal distribution as $n \to \infty$, then this distributional statement indirectly states that the distribution of $\hat{\theta}_n$ converges in distribution to a normal random variable with mean θ and variance $1/i(\theta)$. This shows that $\hat{\theta}_n$ is an asymptotically unbiased estimator of θ and the CRLB is asymptotically achieved by $\hat{\theta}_n$ as $n \to \infty$, or more formally, $\hat{\theta}_n$ is asymptotically efficient [14]. Furthermore, we have by Slutsky's theorem, also provided in Young and Smith [14], that

$$\sqrt{i(\hat{\theta}_n)}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1), \tag{2.27}$$

where $i(\hat{\theta}_n)$ is the information measure evaluated at $\hat{\theta}_n$ and serves as an estimate of Fisher's information. This allows for the same large sample theory to apply, even if one estimates Fisher's information using $\hat{\theta}_n$.

At this point, we have provided the pertinent details regarding Fisher's information and the resulting distributional statements for the MLEs. Now, to help solidify these theoretical results, we present the same information, but in the context of one-parameter exponential families.

The growing theme with exponential families is that numerous technical results can be provided if one simply obtains the various components of the exponential family in question. The derivation of Fisher's Information $i(\theta)$ encapsulates this idea revolving around exponential families. For organizational purposes, we present this derivation as a theorem and provide a short proof to exemplify the relative ease of working with exponential families. Then, we demonstrate an implementation of this theorem.

Theorem 2.4. *Fisher's information for a sample of n independent, identically distributed random variables from an exponential family under the canonical parameterization is $i(\theta) = nc''(\theta)$.*

Proof. Starting with (2.25) and using (2.20), we obtain

$$\begin{aligned}
i(\theta) &= -\mathbb{E}\{l''(\theta)\} \\
&= -\int \cdots \int \left[\frac{\partial^2}{\partial\theta^2} \ln\{f(\boldsymbol{x};\theta)\}\right] f(\boldsymbol{x};\theta)\, dx_1 \ldots dx_n \\
&= -\int \cdots \int \left[\frac{\partial^2}{\partial\theta^2} \ln\left\{\prod_{i=1}^{n} h(x_i)\right\} + \theta \sum_{i=1}^{n} t(x_i) - nc(\theta)\right] f(\boldsymbol{x};\theta)\, dx_1 \ldots dx_n \\
&= -\int \cdots \int -nc''(\theta) f(\boldsymbol{x};\theta)\, dx_1 \ldots dx_n \\
&= nc''(\theta) \int \cdots \int f(\boldsymbol{x};\theta)\, dx_1 \ldots dx_n \\
&= nc''(\theta).
\end{aligned}$$

Thus, the desired result is established. □

Example 2.5. Suppose that we have a random sample of size n from the beta distribution under the "mean/precision" parameterization. We proved that the PDF is of exponential family form in Section 2.1 and with the expression for $c(\mu)$, we have that Fisher's information is

$$i(\mu) = nc''(\mu) = n\phi^2[\gamma'(\mu\phi) + \gamma'((1-\mu)\phi)]. \tag{2.28}$$

Now, by using Theorem 2.4, the distributional statement for the MLE can be expressed as

$$\sqrt{nc''(\theta)}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0,1). \tag{2.29}$$

In a similar manner as before, this statement conveys the fact that $\hat{\theta}_n$ is approximately normally distributed with mean θ and variance $1/(nc''(\theta))$ for relatively large sample sizes. The asymptotic efficiency of the MLE still allows us to have the same distributional statement with $nc''(\hat{\theta}_n)$ as the estimate for Fisher's information.

2.4 Likelihood-Based Hypothesis Tests and Interval Estimation

The previous elaboration of the asymptotic properties of MLEs is now used to demonstrate the various formulations of hypothesis tests for a parameter of interest θ and their corresponding confidence intervals. As the basis for the likelihood-based approaches, suppose we are interested in testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are two disjoint subsets of Θ satisfying $\Theta_0 \cup \Theta_1 = \Theta$. When Θ_0 only contains a single member of Θ and makes a specific statement about the parameter θ, we have the following set of hypotheses: $H_0 : \theta \in \Theta_0 = \{\theta_0\}$ $(H_0 : \theta = \theta_0)$ against $H_1 : \theta \in \Theta_1 = \{\theta \,|\, \theta \neq \theta_0\}$ $(H_1 : \theta \neq \theta_0)$, where H_1 is referred to as a two-sided alternative. There are also one-sided alternatives of the form $H_1 : \theta \in \Theta_1 = \{\theta \,|\, \theta <> \theta_0\}$ $(H_1 : \theta <> \theta_0)$, which implies that $H_0 : \theta \in \Theta_0 = \{\theta \,|\, \theta \geq\leq \theta_0\}$ $(H_0 : \theta \geq\leq \theta_0)$. Based on the inequality sign in H_1, the hypothesis tests with the previous alternatives are referred to as left and right-tailed tests, respectively. With this structure of the hypotheses under consideration, we now proceed to briefly elucidate the key hypothesis testing and interval estimation features regarding both the likelihood-ratio test and the Wald test.

2.4.1 Likelihood-Ratio Test

Letting $\mathcal{L}(\theta)$ denote the likelihood function defined in Section 2.3.1, we write

$$\mathcal{L}_0 = \sup\{\mathcal{L}(\theta)|\theta \in \Theta_0\}, \quad \mathcal{L}_1 = \sup\{\mathcal{L}(\theta)|\theta \in \Theta\}$$

and define the likelihood-ratio statistic Λ_n as

$$\Lambda_n = 2\ln\left\{\frac{\mathcal{L}_1}{\mathcal{L}_0}\right\}, \tag{2.30}$$

where the notation indicates the dependence on the sample size n. Note that the likelihood-ratio statistic is a function of $\boldsymbol{X}$ only through the sufficient statistic T and while an exact test based on Λ_n can be constructed for the likelihood-ratio test (LRT), it is often quite difficult to obtain the exact distribution of T, and thus Λ_n, under H_0. Therefore, the following asymptotic property of Λ_n alleviates this requirement.

Under the same regularity conditions that are needed for the asymptotic properties of the MLEs [5] coupled with the assumption that H_0 is true, then, as $n \to \infty$, the likelihood ratio test statistic Λ_n converges in distribution to a chi-square distribution with 1 degree of freedom, $\Lambda_n \xrightarrow{d} \chi^2_1$ [14].

With this distributional statement, an approximate size α test of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ is to reject H_0 if $\Lambda_n > \chi^2_{1,\alpha}$, where $\chi^2_{1,\alpha}$ denotes the upper-α point of the χ^2_1 distribution. (See Section 2.5.1 for the definition of the size of a test.) This rejection region alludes to the presumption that Λ_n is large if H_1 is true. As for the computation of Λ_n with the previous set of hypotheses, the likelihood-ratio test statistic can be expressed as

$$\Lambda_n = 2\ln\left\{\frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\theta_0)}\right\}, \tag{2.31}$$

where $\hat{\theta}$ is referred to as the unrestricted maximum likelihood estimator.

Upon examining (2.31), one might immediately recognize the technical simplification offered by the natural logarithmic function if the PDF associated with the

parameter of interest θ is of exponential family form,

$$\Lambda_n = 2[(\hat{\theta} - \theta_0)T - n(c(\hat{\theta}) - c(\theta_0))]. \tag{2.32}$$

Since $\hat{\theta}$ is a function of $\boldsymbol{X}$ only through T, then the previous expression explicitly shows that Λ_n also satisfies this same property.

2.4.2 Wald Test

To compute the likelihood-ratio test statistic presented in Section 2.4.1, we had to find the unrestricted maximum likelihood estimator $\hat{\theta}$. Naturally, we can use this calculation to construct another likelihood-based hypothesis test that is commonly referred to as the Wald test. This hypothesis test exhausts all of the asymptotic properties of MLEs provided in Section 2.3.2.

Under H_0, the distributional statement for the Z-score of the MLE of θ, $\hat{\theta}_n$, from (2.26) can be expressed as

$$\sqrt{i(\theta_0)}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, 1). \tag{2.33}$$

Therefore, an approximate size α test of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ is to reject H_0 if $\sqrt{i(\theta_0)} \mid \hat{\theta}_n - \theta_0 \mid > z_{\alpha/2}$, where z_β denotes the upper-β quantile of the standard normal distribution [14]. Note that for one-sided alternatives, $H_1 : \theta <> \theta_0$, the rejection regions are $\sqrt{i(\theta_0)}(\hat{\theta}_n - \theta_0) > z_\alpha$ and $\sqrt{i(\theta_0)}(\hat{\theta}_n - \theta_0) < -z_\alpha$, respectively. Another valid formulation of these tests can be made by replacing $i(\theta_0)$ with $i(\hat{\theta}_n)$ due to asymptotic efficiency of $\hat{\theta}_n$. Furthermore, to continue with the growing theme of exponential families presented in Section 2.3.2, we are, in this case, able to replace $i(\theta_0)$ and $i(\hat{\theta}_n)$ with $nc''(\theta_0)$ and $nc''(\hat{\theta}_n)$, respectively, by using Theorem 2.4.

2.4.3 Interval Estimation

In this section, we first demonstrate the direct correspondence between hypothesis testing and interval estimation through a theorem taken from Casella and Berger [5]. Then, we provide the pertinent details that parallel this result in the context of the LRT along with an abridged derivation of the Wald confidence interval that will help solidify interval estimation.

Theorem 2.6. *For each $\theta_0 \in \Theta$, let $A(\theta_0)$ be the acceptance region of a level α test (See Section 2.5.1) of $H_0 : \theta = \theta_0$ and for each $\boldsymbol{x} \in \mathcal{X}$, define a set $C(\boldsymbol{x})$ in the parameter space by*

$$C(\boldsymbol{x}) = \{\theta_0 | \boldsymbol{x} \in A(\theta_0)\}.$$

Then, the random set $C(\boldsymbol{X})$ is a $1 - \alpha$ confidence set.

Before interpreting the preceding theorem, it should be noted that the term confidence set is used deliberately as it potentially may not be an interval. However, in many cases, like the ones considered in this manuscript, the confidence sets are indeed intervals. A more intricate discussion of the other cases can be found in Casella and Berger [5].

The mathematically rigorous statement presented in Theorem 2.6 alludes to the fact that a confidence set C($\boldsymbol{X}$) provides us with a set of values for θ in which the given data set $\boldsymbol{X}$ yields a "fail to reject H_0" decision, which is contingent on the acceptance region $A(\theta)$. In a sense, confidence sets address the question of what values of the parameter of interest make the given data set most tenable.

This theorem also alludes to the general process of inverting a test statistic. In terms of the two-sided LRT, the acceptance region of this test is of the form $\Lambda_n = 2\ln\{\mathcal{L}(\hat{\theta})/\mathcal{L}(\theta_0)\} \leq \chi^2_{1,\alpha}$. Inverting the likelihood-ratio test statistic to obtain a confidence interval for θ is often achieved numerically with a root-finding algorithm. However, the inversion of the two-sided Wald test statistic has a closed-form solution

due to the symmetry of the normal distribution under H_0. Therefore, by using $i(\hat{\theta}_n)$ as an estimate for Fisher's information, the acceptance region, $\sqrt{i(\hat{\theta}_n)} \mid \hat{\theta}_n - \theta_0 \mid > z_{\alpha/2}$, can be inverted to derive the following $(1-\alpha)100\%$ confidence interval for θ:

$$\hat{\theta}_n \pm z_{\alpha/2}\sqrt{i(\hat{\theta}_n)}.$$

2.5 Sufficiency and Uniformly Most Powerful Tests

When working with exponential families, there is a plethora of literature, such as Young and Smith [14], on the support of constructing hypothesis tests using the sufficient statistic for the parameter of interest θ. The objective of this section is to expatiate the major results regarding this formulation. (We also refer the reader to Young and Smith [14] for a more exhaustive and rigorous explanation.) Now, to achieve this objective, we first provide some of the terminology and concepts from decision theory under a specific framework, which will then lead us to the definition of statistical power and the eventual construction of Uniformly Most Powerful Tests (UMP) in regards to one-parameter exponential families.

2.5.1 Statistical Power

Under the general structure of the hypotheses provided in Section 2.4, we are interested in testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$. Deriving a hypothesis test from this set of hypotheses is to obtain a test statistic $W(\boldsymbol{X})$ that is a function of the observed data $\boldsymbol{X} = \boldsymbol{x}$ and define a critical region C_α, where $0 < \alpha < 1$, that satisfies $\Pr_\theta\{W(\boldsymbol{X}) \in C_\alpha\} \leq \alpha$ for all $\theta \in \Theta_0$. This critical region is then used to determine when we reject H_0; more specifically, we reject H_0 if and only if $W(\boldsymbol{X}) \in C_\alpha$.

To be consistent with the above definitions and terminology, statistical power is the probability of rejecting H_0 when H_0 is false. This concept provides us with a basis for characterizing the power function of a test.

The power function of a test is defined to be

$$\beta(\theta) = \Pr_\theta\{W(\boldsymbol{X}) \in C_\alpha | \theta \in \Theta\}, \tag{2.34}$$

where the subscript on the probability function indicates that the resulting probabilities are a function of θ over all possible values in Θ.

When $\beta(\theta)$ is evaluated over values within Θ_1, this function produces the power of the test for some $\theta \in \Theta_1$, which is consistent with the concept of statistical power. As for the values of $\theta \in \Theta_0$, we have a corresponding concept. Under this notation, a hypothesis test with a defined $W(\boldsymbol{X})$ and C_α is a size α test, where $0 < \alpha < 1$, if $\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha$ [5]. If this property is satisfied, then $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$ and the test is also a level α test. Now, by using the definition of the power function $\beta(\theta)$ in the context of the mathematical statement pertaining to size, α essentially acts as the least upper bound for the probability of rejecting H_0 given that H_0 is true, or the probability of committing a Type I error. In a practical setting, this allows researchers to impose some control on the Type I error rate by specifying the value of α.

2.5.2 Uniformly Most Powerful Tests for One-Parameter Exponential Families

The previous definition of the power function $\beta(\theta)$ allows for the comparison of multiple tests of size α to determine which tests are "better" for a given statistical inference problem. It is quite common for one particular test to have higher power over a specific region of Θ_1, while the others have more power over a different region. It would be very appealing if one could construct a test such that its power function is uniformly greater amongst all other power functions of tests with the same size. If this property indeed exists amongst the class of all hypothesis tests of size α, then this test is said to be Uniformly Most Powerful (UMP).

Now that we have covered some the technicalities regarding finding the optimal test, we now consider one-parameter exponential families to remain consistent with the overarching theme of this chapter. According to Young and Smith [14], with details omitted, one-parameter exponential families are said to be of monotone likelihood ratio, which, in the context of this parametric class, simply means that the likelihood-ratio is a non-decreasing function of the sufficient statistic subject to an inequality based on the parameter of interest. This property facilitates the existence of a UMP test for one-sided tests. More specifically, a one-sided test of size α based on the sufficient statistic is, in fact, a UMP test.

It should be noted that two-sided tests can also be constructed with this idea of uniformity in statistical power, but there is the additional requirement of the test being unbiased. We refer the reader to Young and Smith [14] for further explanation and possible research ideas.

2.6 Discussion

Even though a hypothesis test based on the sufficient statistic for the parameter of interest is mathematically equivalent to the one with likelihood-ratio test statistic provided that the exact distributions of these test statistics are known, there are still some advantages of using the sufficient statistic rather than the likelihood-ratio test statistic in the case of one-parameter exponential families. Therefore, the purpose of this discussion is to provide insight into these advantages in accordance with some of the challenges that arise in these formulations and potential solutions to these problems.

The first advantage is that the problem of deriving the exact distribution of the likelihood-ratio statistic is typically reduced to working directly with the sufficient statistic. This can be seen through the monotone likelihood-ratio property of one-

parameter exponential families, as is mentioned in Section 2.5.2. However, there are still challenges in determining the exact distribution of the sufficient statistic, as well as with the LRT in situations where asymptotic properties of the test statistic are not necessarily maintained. One could resort to approximating the PDF of the sufficient statistic using a normal distribution. However, this test could potentially suffer in the same way as in the case of the LRT due to the fact that these approximations are only viable in large sample size situations.

To illustrate this issue in the context of the beta distribution, we simulated the sampling distribution of both the sufficient statistic, $T = \phi \sum_{i=1}^{n} \ln\{X_i/(1-X_i)\}$, and MLE, $\hat{\mu}$, by taking random samples of size 5 from a beta distribution with mean $\mu = 0.1$ and precision $\phi = 10$. The resulting histograms can be found in Figure 2.1.

Using the normal approximations for the distribution of T and $\hat{\mu}$, we overlaid their approximate density functions, which are indicated by the red-dashed lines. It is apparent that these approximations do not perform particularly well, especially in the tails of the sampling distributions where rejection regions are defined. These discrepancies can consequently influence both the actual size of a test and its statistical power.

The second advantage is that the CGF of the sufficient statistic is feasible, as depicted in Section 2.2.2. Therefore, if the CGF is identifiable, then the exact distribution is given for free and a UMP test can then be constructed. An additional strategy, which we will introduce in the subsequent chapter, is approximating PDFs by expressing the PDF solely in terms of its CGF. These approximations, referred to as saddlepoint approximations, have been shown to provide near exact approximations in a wide variety of settings, as well as for small sample sizes. We hope to utilize these approximations in developing highly accurate, approximate UMP tests.

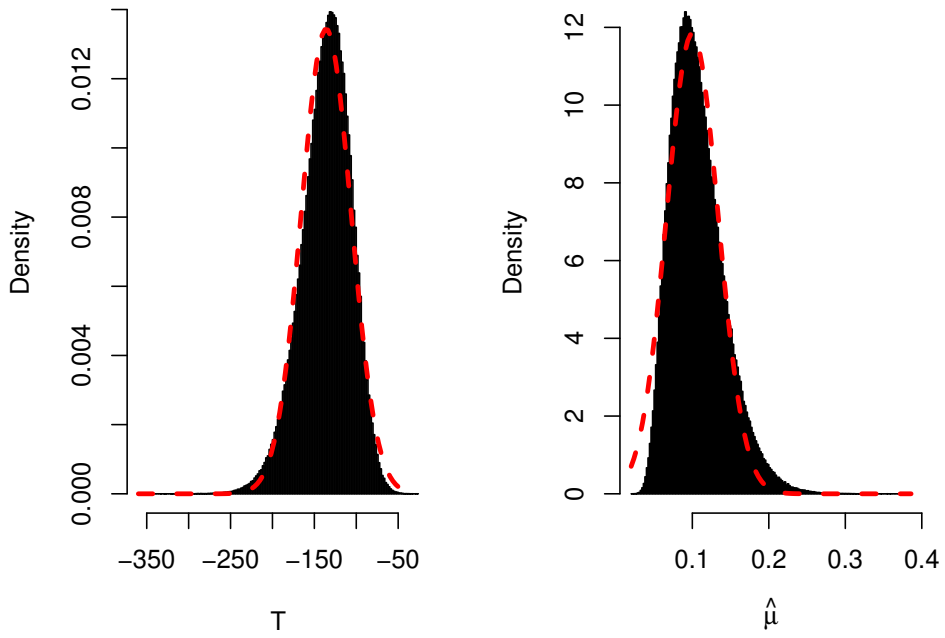

Figure 2.1: Sampling Distribution of the Sufficient Statistic T and the MLE $\hat{\mu}$ ($\mu = 0.1$, $\phi = 10$, $n = 5$)

3 SADDLEPOINT APPROXIMATIONS

As is discussed in Chapter 2, obtaining the exact PDF of a sufficient statistic for a parameter of interest can be a difficult process, even if the PDF is of exponential family form. However, we are still able to characterize the sufficient statistic through its respective MGF or CGF due to their uniqueness. This result alludes to the essence of saddlepoint approximations in that they are used to approximate a PDF from its associated MGF or CGF [6].

In this chapter, we will introduce the saddlepoint approximation for PDFs along with an example of a popular continuous distribution that emphasizes how to implement this approximation. This subtle introduction is beneficial for newcomers as the saddlepoint formula can be a bit illusive. We will then immediately proceed to present the mathematically rigorous derivation of the saddlepoint density, which will further exemplify the intricacies involved in this type of approximation, along with an illustration of the degree of accuracy of this approximation. Due to the strong correspondence between PDFs and CDFs, we will naturally conclude this chapter with a description of the saddlpoint approximation for CDFs.

3.1 Saddlepoint Approximation for PDFs

For a continuous random variable X with CGF $K_X(s)$ and unknown density $f(x)$, the saddlepoint density approximation to $f(x)$ is given as

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi K_X''(\hat{s})}} \exp\{K_X(\hat{s}) - \hat{s}x\}, \tag{3.1}$$

where $\hat{s} = \hat{s}(x)$ denotes the unique solution to the equation

$$K_X'(\hat{s}) = x \tag{3.2}$$

over the range $\hat{s} \in (a, b)$ and is an implicitly defined function of x. Equation (3.2) is referred to as the saddlepoint equation and $\hat{s}$ as the saddlepoint associated with the value of x [4]. Since $\hat{f}(x)$ is an approximation, then it is typically not a proper density function, which implies that $\int \hat{f}(x)\,dx \neq 1$. However, this approximation can be normalized in order to obtain a proper density with the appropriate choice of scale:

$$\tilde{f}(x) = c^{-1}\hat{f}(x), \tag{3.3}$$

where $c = \int \hat{f}(x)\,dx$, is a proper density on the domain of X. In many cases, c is obtained through numerical integration.

Example 3.1. A normally distributed random variable X has CGF of the form $K_X(s) = \mu s + \frac{\sigma^2 s^2}{2}$ for $s \in \mathbb{R}$. With $K_X(s)$, we can define the following components in the saddlepoint approximation:

$K_X(\hat{s}) = \mu\hat{s} + \frac{\sigma^2\hat{s}^2}{2}$;

$K'_X(\hat{s}) = \mu + \sigma^2\hat{s}$;

$K''_X(\hat{s}) = \sigma^2$.

The saddlepoint equation can then be expressed as $K'_X(\hat{s}) = \mu + \sigma^2\hat{s} = x$, which can also be explicitly written as $\hat{s} = \frac{x-\mu}{\sigma^2}$. Through algebraic manipulation, we obtain the following saddlepoint density:

$$\begin{aligned}\hat{f}(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\mu((x-\mu)/\sigma^2) + \sigma^2((x-\mu)/\sigma^2)^2/2 - ((x-\mu)/\sigma^2)x} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2},\end{aligned}$$

which is the exact reproduction of the normal density.

Remark 3.2. One thing that a first-time user should keep in mind with the saddlepoint approximation in (3.1) is the relationship between $\hat{s}$ and x, which is defined by the saddlepoint equation in (3.2). This example illustrates this relationship explicitly by showing the saddlepoint density is solely a function of x upon solving for $\hat{s}$ in the saddlepoint equation.

3.2 Derivation of the Saddlepoint Density

For consistency, we provide a derivation that parallels the one presented in [4] with slight modification. Instead of working with an average of random variables, we only work with a single random variable. (The saddlepoint approximation can be easily applied to a sum of independent and identically distributed (i.i.d.) random variables, since its CGF is readily obtained.)

Since the saddlepoint density is derived from a clever application of Laplace's approximation, then we present Laplace's approximation, which is provided in Butler [4], as a theorem to aid readers before proceeding with the actual derivation.

Theorem 3.3. *Suppose that g is a twice continuously differentiable and concave function on (c, d) with a global minimum at $\hat{x} \in (c, d)$. Then,*

$$\int_c^d e^{-g(x)}\, dx \simeq \frac{\sqrt{2\pi} e^{-g(\hat{x})}}{\sqrt{g''(\hat{x})}}. \tag{3.4}$$

3.2.1 Derivation

Let X denote a continuous random variable with CGF $K_X(s)$. Since the MGF of X is $\exp\{K_X(s)\}$, which by definition may be expressed as an integral involving the density of X, f, then

$$e^{K_X(s)} = \int_{-\infty}^{\infty} e^{sx + \ln\{f(x)\}}\, dx. \tag{3.5}$$

Letting $g(s, x) = -sx - \ln\{f(x)\}$, we have

$$e^{K_X(s)} = \int_{-\infty}^{\infty} e^{sx + \ln\{f(x)\}}\, dx \tag{3.6}$$

$$= \int_{-\infty}^{\infty} e^{-g(s,x)}\, dx. \tag{3.7}$$

Holding s fixed, Laplace's approximation for the integral in (3.7) is

$$e^{K_X(s)} \simeq \sqrt{\frac{2\pi}{g''(s, x_s)}} e^{sx_s} f(x_s), \tag{3.8}$$

where x_s minimizes $g(s,x)$ over x for a fixed s and
$g''(s,x_s) = \partial^2 g(s,x_s)/\partial x_s^2 = -\partial^2 \ln\{f(x_s)\}/\partial x_s^2 > 0$. The critical value x_s solves

$$-g'(s,x_s) = s + \frac{\partial \ln\{f(x_s)\}}{\partial x_s} = 0. \tag{3.9}$$

This implies that

$$-\frac{\partial^2 \ln\{f(x_s)\}}{\partial x_s^2} = \frac{\partial s}{\partial x_s}. \tag{3.10}$$

An approximation to $\partial \ln\{f(x_s)\}/\partial x_s$ may be determined by first solving for $\ln\{f(x_s)\}$ in (3.8) and then substituting in the expression for $g''(s,x_s)$ to obtain

$$\ln\{f(x_s)\} \simeq K_X(s) - sx_s - \frac{1}{2}\ln\left\{\frac{2\pi}{-\partial^2 \ln\{f(x_s)\}/\partial x_s^2}\right\}. \tag{3.11}$$

If the last term is assumed to be approximately constant in x_s and therefore negligible upon differentiation with respect to x_s, then

$$\frac{\partial \ln\{f(x_s)\}}{\partial x_s} \simeq K'_X(s)\frac{\partial s}{\partial x_s} - \left[s + x_s\frac{\partial s}{\partial x_s}\right] \tag{3.12}$$

$$= [K'_X(s) - x_s]\frac{\partial s}{\partial x_s} - s \tag{3.13}$$

follows from both the chain and product rules. To the degree of the approximation made in obtaining Equation (3.13),

$$\frac{\partial \ln\{f(x_s)\}}{\partial x_s} + s = 0 \quad \Longleftrightarrow \quad [K'_X(s) - x_s]\frac{\partial s}{\partial x_s} = 0.$$

Since it can be shown through (3.9) that there is a monotone increasing relationship between s and x_s, $\partial s/\partial x_s > 0$, then this implies that $K'_X(s) - x_s = 0$ or $K'_X(s) = x_s$. Therefore, s and x_s must be related through $K'_X(s) = x_s$, which is referred to as the saddlepoint equation.

By using (3.10) and $K''_X(s) = \partial x_s/\partial s$, we can express $g''(s,x_s)$ as

$$g''(s,x_s) = -\frac{\partial^2 \ln\{f(x_s)\}}{\partial x_s^2} = \frac{\partial s}{\partial x_s} = \left(\frac{\partial x_s}{\partial s}\right)^{-1} = (K''_X(s))^{-1}. \tag{3.14}$$

By exponentiating both sides of (3.11) and substituting in the expression for $-\partial^2 \ln\{f(x_s)\}/\partial x_s^2$ in (3.14), we obtain

$$f(x_s) \simeq \sqrt{\frac{1}{2\pi K_X''(s)}} \exp\{K_X(s) - sx_s\}, \tag{3.15}$$

where $K_X'(s) = x_s$. This is the saddlepoint density defined in (3.1) with its associated saddlepoint equation, $K_X'(s) = x_s$.

3.2.2 Degree of Accuracy

The saddlepoint approximation is often discussed in literature as being "second order" asymptotic in relation to the classical "first order" asymptotic theory used in the CLT, MLEs, and LRTs. The classical approximations are typically of order $O(n^{-1/2})$ while the saddlepoint approximations are of order $O(n^{-1})$, which is partially due to Laplace's approximation. However, there are some cases of order $O(n^{-3/2})$. This added boost in accuracy is beneficial when dealing with smaller sample sizes [4].

To provide further insight into the degree of accuracy of the saddlepoint density, we will refer to the reoccurring example of the Beta distribution and consider the PDF of the sufficient statistic for μ, $T = \phi \sum_{i=1}^{n} \ln\{X_i/(1 - X_i)\}$. If we consider the simple case when $n = 1$, then T is a function of a single Beta random variable, $T = T(X) = \phi \ln\{X/(1 - X)\}$. Furthermore, the distribution of the random variable, $X/(1 - X)$, is commonly referred to as Pearson's Type VI distribution, or the Beta Prime distribution [11]. The density function for the natural log of a Beta Prime random variable can then be easily derived through standard transformation theory, since the natural log is a strictly increasing function [5]. Therefore, the PDF of T can be expressed as

$$f(t) = \frac{\Gamma(\phi) \exp\{t/\phi\}^{\mu\phi} (1 + \exp\{t/\phi\})^{-\mu\phi - (1-\mu)\phi}}{\phi \Gamma(\mu\phi)\Gamma((1 - \mu)\phi)}, \quad t \in (-\infty, \infty).$$

Figure 3.1 provides a histogram of one million draws of T with $n = 1$ and X

following a beta distribution with $\mu = 0.1$ and $\phi = 10$. Now, to illustrate the level of accuracy of the saddlepoint approximation, we have overlaid the histogram with the exact density (solid red), the saddlepoint approximation (dot-dashed), the normalized saddlepoint density (dashed), and a Normal approximation (two-dashed) derived in Section 2.2.2 by using the mean and variance obtained through the CGF of T. Upon examination, the unnormalized saddlepoint density captures some of the shape of the theoretical curve, but not to the correct scaling. As this approximation only integrates to approximately 0.09663219, and not 1, over the domain of the simulated sampling distribution of T, this highlights the fact that saddlepoint approximations, in general, are not "true" density functions. However, they are easily normalized to become proper density functions. This is typically achieved, as is discussed in Section 3.1, by computing a normalization constant through numerical integration. By accurately scaling the saddlepoint approximation, the exact density and the normalized saddlepoint density are virtually indistinguishable.

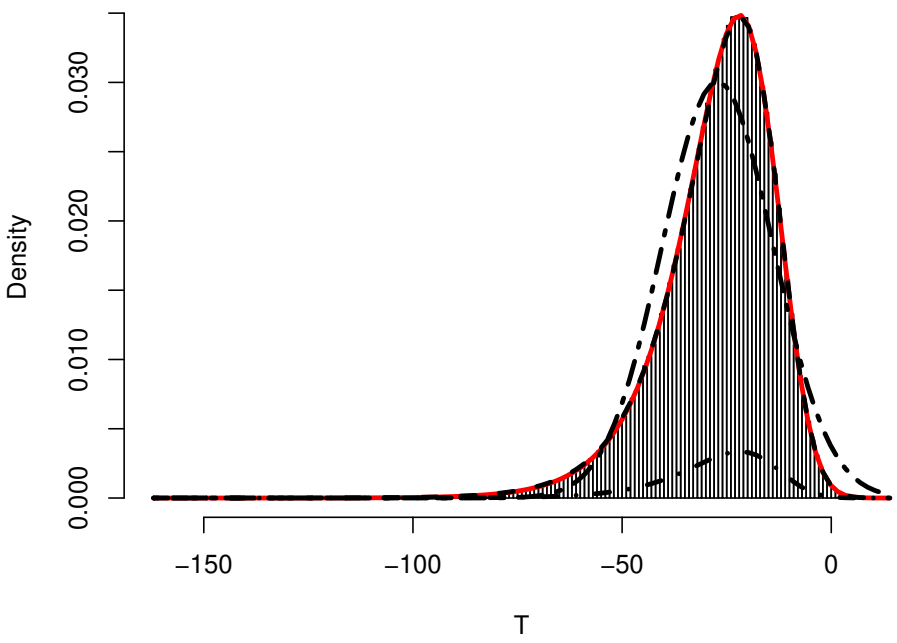

Figure 3.1: Sampling Distribution of the Sufficient Statistic T ($\mu = 0.1, \phi = 10, n = 1$) Overlaid with the Theoretical Curve (Solid Red), the Saddlepoint Approximation (Dot-Dashed), the Normalized Saddlepoint Density (Dashed), and the Normal Approximation (Two-Dashed)

While we now have some theoretical support for the exact distribution of T, we still could not obtain a closed-form distributional statement for the sum of n independent natural logarithm of Beta Prime random variables ourselves or find any insights through literature. We can approximate this distribution via the CLT for relatively large samples, but for small samples, the results of this approximation vary, which has already been demonstrated in Figure 3.1 for $n = 1$.

To further demonstrate the approximation problem, we have provided a similar graphic for the distribution of T with $n = 5$ in Figure 3.2. The normalized saddlepoint density provides a better approximation to this distribution over the Normal approximation, particularly in the tails of this distribution. This further alludes to the application of saddlepoint approximations in small sample situations, which are exceedingly prevalent in this manuscript's main motivational application of immunology research with FCM data as FCM projects are often faced with moderately low sample sizes. We will provide the details for the saddlepoint approximation for T in the next chapter.

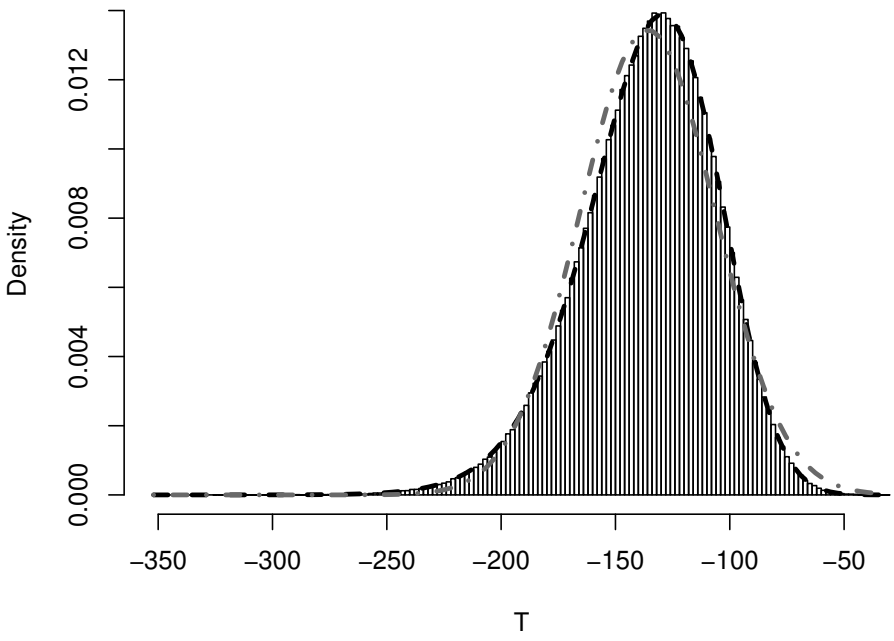

Figure 3.2: Sampling Distribution of the Sufficient Statistic T ($\mu = 0.1, \phi = 10, n = 5$) with the Normalized Saddlepoint Density (Black Dashed) and the Normal Approximation (Gray Dot-Dashed)

3.3 Lugannani and Rice Saddlepoint Approximation for CDFs

Suppose X is a continuous random variable that has CDF $F(x)$ and CGF K_X with mean $\mu = \mathbb{E}(X)$. The saddlepoint approximation for $F(x)$, $\hat{F}(x)$, provided by Lugannani and Rice is

$$\hat{F}(x) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u}) & x \neq \mu \\ \frac{1}{2} + \frac{K_X'''(0)}{6\sqrt{2\pi}K_X''(0)^{3/2}} & x = \mu \end{cases}, \tag{3.16}$$

where

$$\hat{w} = \text{sgn}(\hat{s})\sqrt{2(\hat{s}x - K_X(\hat{s}))} \text{ and } \hat{u} = \hat{s}\sqrt{K_X''(\hat{s})}$$

are functions of x and the saddlepoint $\hat{s}$ that is the unique solution to the saddlepoint equation $K_X'(s) = x$. The symbols ϕ and Φ denote the standard normal density and CDF, respectively, and $\text{sgn}(\hat{s})$ captures the sign for $\hat{s}$.

Recall from the properties of the CGF that $K_X'(0) = \mathbb{E}(X) = \mu$, which implies that $\hat{s} = 0$ when $x = \mu$. Furthermore, if $\hat{s} = 0$, then $\hat{w} = 0 = \hat{u}$. This essentially causes the second term in the expression for the first case in the piece-wise function (3.16) to be undefined. Therefore, the second case represents the limiting value as $x \to \mu$, or as $\hat{s} \to 0$. Butler provides the mathematically rigorous details for proving that $\lim_{\hat{s}\to 0} 1/\hat{w} - 1/\hat{u}$ exists and $\lim_{\hat{s}\to 0} 1/\hat{w} - 1/\hat{u} = K_X'''(0)/6K_X''(0)^{3/2}$ [4]. With this and the limiting properties of continuous functions, we obtain the expression for the second case. Thus, there is a removable singularity at $x = \mu$, which implies that $\hat{F}$ is continuously differentiable or theoretically "smooth". There is, however, the potential for numerical instability when computing $\hat{F}(x)$ for values of x relatively close to μ in software.

Remark 3.4. The derivation of the saddlepoint approximation for a CDF involves not only using a Temme approximation and the intricate properties of a particular mapping, but examining various limiting properties that require multiple Taylor series

expansions. Therefore, due to its extensiveness, we will not provide this derivation. However, an exhaustive and rigorous one can be found in [4].

4 STATISTICAL INFERENCE FOR THE MEAN OF THE BETA DISTRIBUTION

Throughout the previous two chapters, we have discussed various results from distribution theory and reviewed some of the methods for conducting statistical inference with exponential families. We used the beta distribution under the "mean/precision" parameterization, with ϕ as a known constant and μ as the parameter of interest, to illustrate many of these results. This was specifically due to the main motivation for this manuscript which is, in fact, to conduct statistical inference for the mean of the beta distribution, μ. Therefore, this chapter serves to provide a concise description of the hypothesis tests and corresponding confidence intervals for μ. To achieve this, we will first remind the readers of any pertinent details regarding the beta distribution that were provided in the previous chapters. We will then recast the hypothesis testing procedures, such as the LRT, a Z-score based test using the MLE, and an approximate UMP test via a saddlepoint approximation, under the beta distribution model. As for the approximate UMP test, we will also provide the full details of this approximation and describe the manner in which the computations were performed in software.

4.1 An Account of Key Results

The PDF of the beta distribution is of exponential family form with canonical parameter μ if we define $h(x) = \Gamma(\phi)I_{(0,1)}(x)(1-x)^{\phi-1}/x$, $t(x) = \phi \ln\{x/(1-x)\}$, and $c(\mu) = \ln\{\Gamma(\mu\phi)\} + \ln\{\Gamma((1-\mu)\phi)\}$. With reference to $t(x)$, the canonical sufficient statistic for μ based on a simple random sample is $T = \phi \sum_{i=1}^{n} \ln\{X_i/(1-X_i)\}$.

The CGF of T is

$$K_T(s) = n[\ln\{\Gamma((s+\mu)\phi)\} + \ln\{\Gamma((1-(s+\mu))\phi)\} - \ln\{\Gamma(\mu\phi)\} - \ln\{\Gamma((1-\mu)\phi)\}].$$

The MLE of μ, $\hat{\mu}$, is the unique solution to the normal equation

$$nc'(\mu) = T,$$

which is

$$n\phi[\gamma(\mu\phi) - \gamma((1-\mu)\phi)] = \phi\sum_{i=1}^{n}\ln\left\{\frac{X_i}{1-X_i}\right\},$$

and is obtained by a numerical root-finder, such as *uniroot* in R. Fisher's information is then used to obtain the asymptotic variance of $\hat{\mu}$. For a simple random sample of n observations from the beta distribution with unknown mean μ and known precision ϕ, Fisher's information is

$$i(\mu) = nc''(\mu) = n\phi^2[\gamma'(\mu\phi) + \gamma'((1-\mu)\phi)].$$

4.2 Hypothesis Testing and Interval Estimation with Respect to μ

In this section, we enumerate and elaborate on the previously mentioned hypothesis testing and interval estimation procedures in terms of our statistical inference problem.

4.2.1 LRT

In accordance with Section 2.4.1, an approximate size α test of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ for some prescribed μ_0 is to reject H_0 if $\Lambda_n > \chi^2_{1,\alpha}$, where

$$\Lambda_n = 2\left\{(\hat{\mu}-\mu_0)\left[\phi\sum_{i=1}^{n}\ln\{X_i/(1-X_i)\}\right] - n(c(\hat{\mu}) - c(\mu_0))\right\}$$

and $\chi^2_{1,\alpha}$ denotes the upper-α point of the χ^2_1 distribution.

As for the corresponding confidence interval for μ, we proceeded to invert the likelihood-ratio test statistic of the previously defined test, as is described in Section 2.4.3. This interval was achieved numerically with *uniroot* in *R.*, as there does not exists a closed-form solution.

4.2.2 Wald Test

Recall from Section 2.4.2 that under H_0, the distributional statement for the Z-score of the MLE of μ, $\hat{\mu}_n$, where the subscript is used to indicate a dependence on the sample size n, can be expressed as

$$\sqrt{i(\hat{\mu}_n)}(\hat{\mu}_n - \mu_0) \xrightarrow{d} N(0,1),$$

where $i(\hat{\mu}_n)$ is an estimate for Fisher's information, $i(\mu)$. Therefore, an approximate size α test of $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ for some prescribed μ_0 is to reject H_0 if $\sqrt{i(\hat{\mu}_n)} \mid \hat{\mu}_n - \mu_0 \mid > z_{\alpha/2}$, where z_β denotes the upper-β quantile of the standard normal distribution. Note that for one-sided alternatives, $H_1 : \mu <> \mu_0$, the rejection regions are $\sqrt{i(\hat{\mu}_n)}(\hat{\mu}_n - \mu_0) > z_\alpha$ and $\sqrt{i(\hat{\mu}_n)}(\hat{\mu}_n - \mu_0) < -z_\alpha$, respectively.

Also, recall that inversion of the two-sided Wald test statistic has a closed-form solution and the $(1-\alpha)100\%$ confidence interval for μ can thus be expressed as

$$\hat{\mu}_n \pm z_{\alpha/2}\sqrt{i(\hat{\mu}_n)}.$$

4.2.3 Approximate UMP Test Using a Saddlepoint Approximation

Since the PDF of the beta distribution is of exponential family form, then this distribution is of monotone likelihood ratio with respect to the canonical sufficient statistic for μ, $T = \phi \sum_{i=1}^{n} \ln\{X_i/(1-X_i)\}$. This implies that there exists a UMP test based on T for each set of hypotheses. More specifically, the test of $H_0 : \mu \leq \mu_0$

against $H_1 : \mu > \mu_0$ for some prescribed μ_0 is to reject H_0 if $T > t_u$, where t_u is defined to be upper-α quantile of the distribution of T, and is UMP amongst all tests of size α. Similarly, the UMP test of $H_0 : \mu \geq \mu_0$ against $H_1 : \mu < \mu_0$ amongst all tests of size α is to reject H_0 if $T < t_l$, where t_l is defined to be the lower-α quantile of the distribution of T. Note that we cease with the conventional notation regarding quantiles in favor of unique subscripts to emphasize the fact that, unlike the standard normal distribution, the distribution of T is typically not symmetric.

Even though we have now derived the UMP tests for the sets of hypotheses with one-sided alternatives, the quantiles t_l or t_u still must be determined by using the distribution of T. Specifically, if we let $F_T(t)$ denote the "true" CDF of T, then the values of t_l and t_u are determined by solving the equations $F_T(t_l) = \alpha$ and $1-F_T(t_u) = \alpha$, respectively. However, as is discussed in Section 2.6, the distribution of T is not known. Therefore, we approximate the CDF of T using the saddlepoint approximation methods discussed in Chapter 3. For consistency and completeness, we now provide the necessary information to compute both the saddlepoint approximation to the PDF and CDF of T.

With the CGF of T defined in Section 4.1, the first and second derivatives with respect to s are:

$K'_T(s) = n\phi[\gamma((s+\mu)\phi) - \gamma((1-(s+\mu))\phi)]$;

$K''_T(s) = n\phi^2[\gamma'((s+\mu)\phi) + \gamma'((1-(s+\mu))\phi)]$.

By substituting these functions into (3.1), we obtain the following saddlepoint density for T:

$$\hat{f}_T(t) = \frac{1}{\sqrt{2\pi K''_T(\hat{s})}} \exp\{K_T(\hat{s}) - \hat{s}t\},$$

where $\hat{s}$ solves the saddlepoint equation

$$\begin{aligned} K'_T(\hat{s}) &= n\phi[\gamma((\hat{s}+\mu)\phi) - \gamma((1-(\hat{s}+\mu))\phi)] \\ &= t. \end{aligned}$$

Due to the natural logarithm of the gamma functions present in the saddlepoint equation, there does not exist a closed-form solution to $\hat{s}$. This can be numerically approximated in a similar manner as computing $\hat{\mu}$; however, if one has already computed $\hat{\mu}$, then we can actually obtain the value of $\hat{s}$ for free. Upon further examination of the saddlepoint equation, it can verified that that $K_T'(\hat{s}) = t$ is related to the normal equation defined in Section 4.1 by the way of $K_T'(\hat{s}) = nc'(\hat{s} + \mu) = t$. Since we know that $\hat{\mu}$ solves the normal equation, then we can express $\hat{s}$ as $\hat{\mu} - \mu$. (We refer the reader to Butler [4] for more details.) This particular relationship was computationally efficient when we programmed all of the statistical inference components in software, as the MLE had already been computed for the more traditional testing procedures.

Now, by substituting the properties specified in Section 4.1, along with $K_T'''(0) = n\phi^3[\gamma''(\mu\phi) - \gamma''((1-\mu)\phi)]$, into (3.16), we obtain the following saddlepoint approximation for the CDF of T:

$$\hat{F}_T(t) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})(1/\hat{w} - 1/\hat{u}) & t \neq \mathbb{E}(T) \\ \frac{1}{2} + \frac{K_T'''(0)}{6\sqrt{2\pi}K_T''(0)^{3/2}} & t = \mathbb{E}(T) \end{cases},$$

where

$$\hat{w} = \text{sgn}(\hat{s})\sqrt{2(\hat{s}t - K_T(\hat{s}))} \text{ and } \hat{u} = \hat{s}\sqrt{K_T''(\hat{s})}$$

are functions of t and the saddlepoint solution $\hat{s}$.

With the saddlepoint CDF, we can directly derive an approximate UMP test of $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$. Let T_{obs} denote the observed sufficient statistic for μ from a given data set. The decision rule associated with this test is to reject H_0 if $T_{obs} > t_u$, where t_u satisfies $1 - \hat{F}_T(t_u) = \alpha$. Similarly, the approximate UMP test of $H_0 : \mu \geq \mu_0$ against $H_1 : \mu < \mu_0$ is to reject H_0 if $T_{obs} < t_l$, where t_l satisfies $\hat{F}(t_l) = \alpha$. Since $\hat{F}_T(t)$ is a monotonically increasing function, then both t_l and t_u can be determined numerically.

Perhaps a more efficient approach to the previously defined tests is the p-value approach to hypothesis testing, as this requires computing probabilities directly through $\hat{F}_T(t)$ without the need for inversion. Succinctly, the p-value, in the context of these tests, is defined to be the probability of observing a sufficient statistic T as extreme or more extreme than the observed sufficient statistic T_{obs}, given that the null hypothesis is true. Therefore, for the test of $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$, the p-value is obtained by calculating $1 - \hat{F}_T(T_{obs})$ with $\mu = \mu_0$. Similarly, with a left-sided alternative, $H_1 : \mu < \mu_0$, the p-value is $\hat{F}_T(T_{obs})$ with $\mu = \mu_0$. The resulting decision for either test would be to reject H_0 if the respective p-value is less than α.

Now that we have elaborated on tests with one-sided alternatives, we now consider two-sided tests of the form $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Recall from Section 2.5.2 that two-sided tests can be constructed to exhibit uniformity in statistical power, as in the case of one-sided tests. However, there is an additional requirement of the test being unbiased. Note that the implementation of this property in software was a bit more rigorous due to the asymmetry of the distribution of T and there were computational difficulties caused by the numerical instability of a derivative. To circumvent this, we adopted the "double" p-value approach to this test. The rejection region of this size α test is determined in a similar manner as the one-sided tests by selecting t_l and t_u such that their respective tail probabilities are each $\alpha/2$, which implies that the p-vales are obtained by doubling the smallest one-sided p-value.

As for the confidence interval for μ, this is obtained through the inversion of the previously defined two-sided hypothesis test. In accordance with Theorem 2.6, for each $\mu_0 \in (0, 1)$, define $A(\mu_0) = \{T_{obs} | t_l < T_{obs} < t_u\}$, where T_{obs} is the observed sufficient statistic for μ given a data set, to be the acceptance region of this size α test. Note that both t_l and t_u depend on the value of μ_0. Now, for each $T_{obs} \in (-\infty, \infty)$, the corresponding $1 - \alpha$ confidence set is the set $C(T_{obs}) = \{\mu_0 | T_{obs} \in A(\mu_0)\}$. This implies that the lower and upper limits of the $(1 - \alpha)100\%$ confidence interval for μ

are determined by essentially finding the smallest and largest values of μ such that T_{obs} falls in the acceptance region. As there does not exist a closed-form solution to this interval, we used numerical routines to acquire these values.

Remark 4.1. Recall from Section 3.3 that $\hat{F}_T(t)$ has a removable discontinuity at $t = \mathbb{E}(T)$, as this implied that $\hat{w} = 0 = \hat{u}$. Theoretically, this does not present a problem. However, there is numerical instability when computing $\hat{F}_T(t)$ for values of t relatively close to $\mathbb{E}(T)$ in software, such as R. In order to alleviate this instability, we employed a computationally equivalent approximation for the saddlepoint CDF of T and used a specific mapping. We use Figure 4.1 to demonstrate this approach that was developed by Jacob Turner, Ph.D, for the purposes of this manuscript.

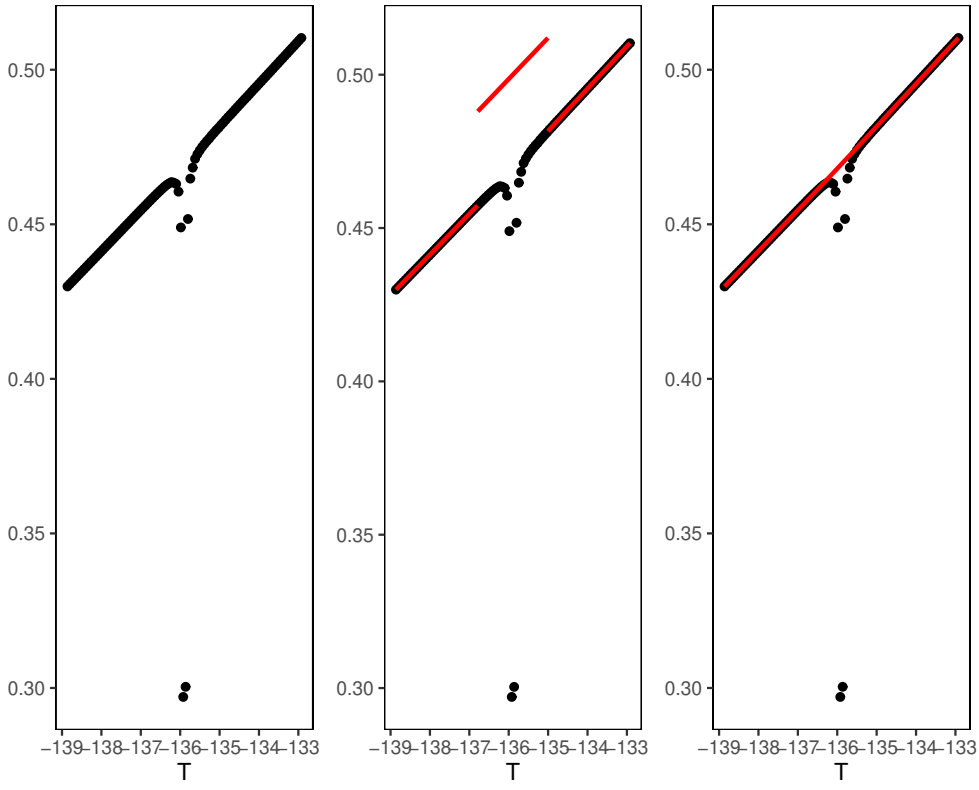

Figure 4.1: Transformation Associated with $\hat{F}_T(t)$ (Black Dotted) for $(\mu = 0.1, \phi = 10, n = 5)$

The far-left plot of $\hat{F}_T(t)$, for the case where $\mu = 0.1$, $\phi = 10$, and $n = 5$, in Figure 4.1 illustrates the major concern when implementing $\hat{F}_T(t)$ in R. As we numerically evaluate $\hat{F}_T(t)$ at values close to $\mathbb{E}(T) = 5(10)[\gamma(0.1(10)) - \gamma((1-0.1)10)] \approx -135.8929$, we can see that it starts to become unstable and lose its monotonicity. This is problematic when numerically solving for the quantiles, t_l or t_u, of the rejection regions associated with our approximate hypothesis tests and obtaining the limits of the confidence interval for μ. To circumvent this issue and provide an approximation for the CDF of T that preserves its monotonicity in software, we took the following ad hoc approach.

An alternative form of the saddlepoint approximation for the CDF of T is the r^* approximation (Barndorff-Nielsen [1], [2], [3]). The saddlepoint CDF of T can be expressed as $\hat{F}_T(t) \approx \Phi(r^*)$, where Φ is the standard normal CDF and $r^* = \hat{w} - (1/\hat{w}) \ln\{\hat{w}/\hat{u}\}$. The numerical differences between the original saddlepoint CDF of T and the r^* version are negligible, as they are related by Taylor series expansion, and they typically coincide with each other in terms of accuracy.

With this approximation, we now need to identify an interval where the numerical instability occurs. Essentially, $\hat{F}_T(t)$ is unstable for values of t that correspond to values of $\hat{w}$ in $-0.03 \leq \hat{w} \leq 0.03$. (Note that this interval is conservative as the instability occurs in a much tighter neighborhood about $\hat{w} = 0$.) After determining this interval, we then modified the implementation of $\hat{F}_T(t)$ in R by checking if t produced a $\hat{w}$ in this problematic interval. If this was the case, then we approximated $\hat{F}_T(t)$ using $\Phi(\hat{w})$, as $\hat{w}$ is the first term in r^* and the instability is caused by the second term due to the fact that $\hat{w} \approx 0 \approx \hat{u}$. This essentially created a "shifted segment" of the approximation over the problematic interval, as is indicated by the shifted solid red line in the middle graph of Figure 4.1. However, this segment can be shifted and scaled to the endpoints of $\hat{F}_T(t)$, since we can directly obtain the values of t that correspond to $\hat{w} = -0.03$ and $\hat{w} = 0.03$. The final result of this mapping

produced the solid red line in the far-right graph of Figure 4.1, which shows that the adjusted approximation for the CDF of T is now strictly monotonic, although not continuous.

This ad hoc approach has served us well in performing the necessary computations involved in hypothesis testing and constructing confidence intervals for μ in R, as we will see in our simulation studies provided in the next chapter.

5 SIMULATION STUDIES

The purpose of this chapter is to empirically provide further insight into the different aspects of the hypothesis tests, along with their corresponding confidence intervals, for the mean of the beta distribution, μ, discussed in Chapter 4. These tests include the likelihood-based tests, such as the asymptotic LRT and the Wald test, and the saddlepoint approximation approach to the tests derived from the sufficient statistic for μ. Even though the LRT and the Wald test are asymptotically equivalent, it is still practical to examine the discrepancies, if any exist, between these tests for different sample sizes. In addition to the tests specified previously, we will also consider the one-sample Student's t-test as a benchmark due to its prevalence in applied settings. As there are an infinite number of possible scenarios to generate the preliminary estimates required for each test, we chose to present the simulation results that exhibited significant discrepancies between the tests in terms of their estimated Type I error rates, simulated power curves, or the estimated coverage probabilities for their corresponding confidence intervals.

5.1 Scenarios

Recall from Section 2.5.1 that in the context of our statistical inference problem, a test is of size α if $\sup_{\mu \in M_0} \beta(\mu) = \alpha$, where M_0 is the set of values of μ under H_0 and $\beta(\mu)$ is the power function of the test. The previous mathematical statement conveys that fact that α represents the least upper bound of $\beta(\mu)$ for $\mu \in M_0$. Now, for hypotheses of the following form: $H_0 : \mu \in M_0 = \{\mu_0\}$ $(H_0 : \mu = \mu_0)$ against $H_1 : \mu \neq \mu_0$ (two-sided tests), the supremum over M_0 is achieved at the prescribed value, μ_0, as M_0 only consists of a single point. This is also true for the hypotheses

of the form: $H_0 : \mu \in M_0 = \{\mu | \mu \geq\leq \mu_0\}$ ($H_0 : \mu \geq\leq \mu_0$) against $H_1 : \mu <> \mu_0$ (Left and right-tailed tests, respectively) due to the fact that $\beta(\mu)$ is a strictly decreasing and increasing function over M_0 for left and right-tailed tests, respectively.

Now, by using the definition of the power function presented in Section 2.5.1, α acts as the least upper bound for the probability of committing a Type I error, or the probability of rejecting a true null hypothesis. For our purposes, we will impose some control on the Type I error rate and set $\alpha = 0.05$. Essentially, we want the tests under consideration to maintain this rate; more specifically, have this rate be equal to 0.05 at μ_0, as was stated previously in the context of power functions. Through simulation work, we will be able to acquire empirical evidence as to whether or not this rate is being sufficiently controlled for each test.

As for the Type I error rate simulations, we will consider the previously stated sets of hypotheses, where $\mu_0 = 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$, with $\phi = 2, 10$ and $n = 5, 10, 25$. Under each scenario, we conducted 10000 simulations to obtain an estimate of the Type I error rate for each test. Then, we proceeded to provide graphical representations of these rates based on the value of ϕ. A more detailed description of the figures will be presented later in this chapter.

An alternative approach for assessing the control of the Type I error rate for a test is through the construction of its power function and we can graphically indicate if the supremum over M_0, α, is actually achieved at μ_0, or if the Type I error rate is being controlled. As this assessment only applies to values of μ under H_0, the behavior of $\beta(\mu)$ across the values in H_1 is of some interest. Essentially, we want $\beta(\mu)$ to exhibit higher values of power over the alternative region, as this would indicate a higher probability of correctly rejecting H_0.

Now, to generate a simulated power curve for each test, we first considered testing $H_0 : \mu \geq\leq 0.1$ against $H_1 : \mu <> 0.1$ at $\alpha = 0.05$ with $\phi = 2, 10$ and $n = 5, 10, 25$. (We will expound upon the choice of the prescribed value of μ_0 later on in this chap-

ter.) Under each scenario, we conducted 10000 simulations for each μ in a specified interval, $[0.05, 0.3]$ for right-tailed tests and$[0.05, 0.25]$ for left-tailed tests, to obtain an estimate of the power and then plotted the power estimates against their respective values for μ.

As we have enumerated the scenarios associated with different aspects of two-sided tests, we will naturally want to consider the behavior of their corresponding confidence intervals for μ. Recall from Theorem 2.6 that a confidence interval is obtained by inverting the acceptance region of a two-sided test of size α for a given data set. As a result, the confidence interval provides a set of values for μ in which the data set yields a "fail to reject H_0" decision. With this, we define the coverage probability of a confidence interval for μ as the probability that the random interval covers the true parameter, μ [5]. Since the Type I error rate is specified at 0.05, the coverage probability for each confidence interval for μ should be maintained at 0.95.

Now, as for the technicalities of the coverage probability simulations, we did not consider scenarios that had both values of μ_0 relatively close to 0 and low precision values due to computational difficulties. To our knowledge, these difficulties arose in the practical implementation of our code. For instance, we experienced problems that appeared to stem from the degenerate samples generated in the simulation studies, which essentially caused numerical instability in the root-finding algorithms provided in R. However, for the time being, we believe that a majority of these situations do not reflect real world phenomenon. As part of future research, we plan on attempting to develop an approach that will circumvent some of these complications.

With this in mind, we specifically set $\mu_0 = 0.2, 0.3, 0.4, 0.5$ and $\phi = 10$. We also varied the sample size, $n = 5, 10, 25$, and then proceeded to conduct 10000 simulations under each scenario in order to obtain an estimate for the coverage probability of a specified confidence interval for μ.

5.2 Type I Error Rate Simulations

For a fixed ϕ, the estimated Type I error rates across varying sample sizes for each test are displayed in a plot, where the x-axis represents the values of μ_0 and the y-axis represents the estimated Type I error rates. Each plot has a solid-black horizontal line at 0.05 along with simulation error bounds, which are indicated by black-dashed horizontal lines. These lower and upper limits are determined by using the fact that the decisions associated with a hypothesis test based on a generated sample can be viewed as a Bernoulli trial with a probability of success of 0.05 under H_0. We can then use the binomial distribution to construct a confidence interval for this proportion. Specifically, $0.05 \pm 1.959964\sqrt{0.05(1-0.05)/10000} \approx [0.0457, 0.0543]$. After these bounds have been obtained, an estimated Type I error rate that falls outside this region is deemed as not controlled.

5.2.1 Right-Tailed Tests

As depicted in the far-left panel of Figure 5.1, the estimated Type I error rate for the test derived from the sufficient statistic for μ that used saddlepoint approximations, which recall from Section 4.2.3 is an approximate UMP test, appears to be controlled for low precision, $\phi = 2$, and across all sample sizes. This essentially provides empirical evidence substantiating the theory provided in Section 2.5.2 regarding the optimality of tests based on the sufficient statistic for μ.

The middle panel in Figure 5.1 corresponds to the likelihood-based Wald test, which uses the asymptotic properties of the MLE for μ, $\hat{\mu}$. The apparent deflation in the estimated Type I error rate for values of μ_0 relatively close to 0 is attributed to the fact that the previously mentioned asymptotic properties of $\hat{\mu}$ are not necessarily maintained under these scenarios due to the asymmetry of the sampling distribution of $\hat{\mu}$. However, this error rate does become more controlled for larger sample sizes, as

large sample theory would suggest.

As for the far-right panel in Figure 5.1, the estimated Type I error rates for the t-test are deflated, which is similar to the case of the Wald test, but with a different degree of severity. Recall from Chapter 1 that ϕ is the variance-controlling parameter and the variance of the beta distribution increases as ϕ decreases. Therefore, these scenarios are indicative of high variability. Also, recall that the beta distribution does not exhibit symmetry for values of μ_0 relatively close to 0. Therefore, the coupling of these two specifications alludes to severe departures from normality and as the t-test is not robust to these types of departures, the estimated Type I error rates are indeed not controlled.

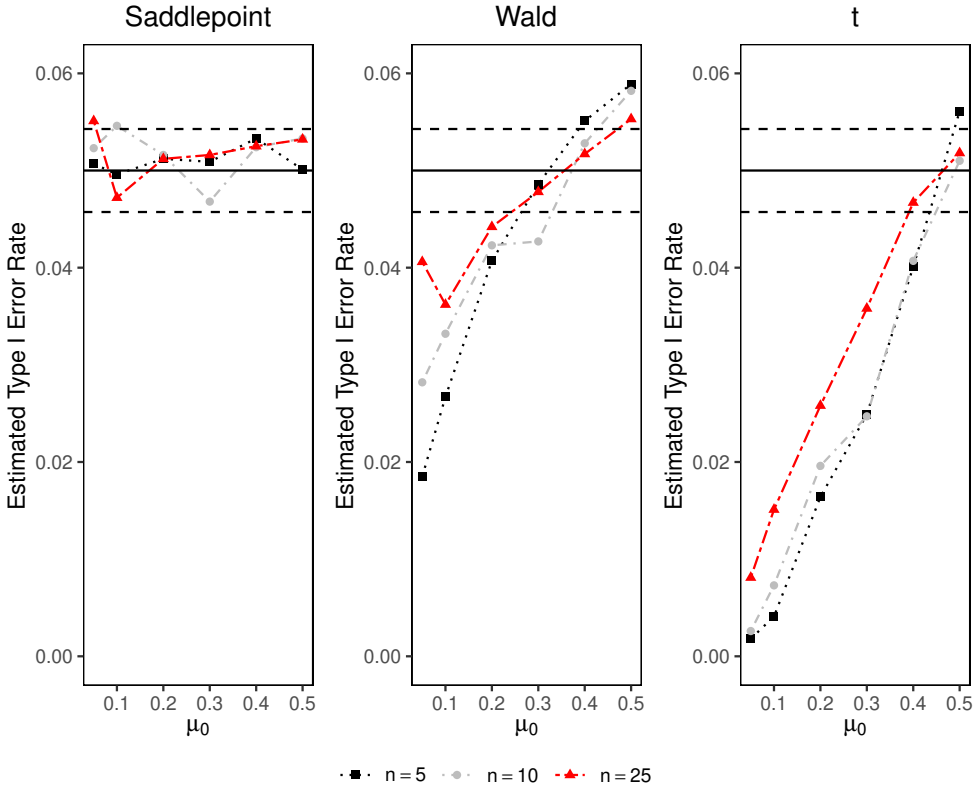

Figure 5.1: Estimated Type I Error Rates for Right-Tailed Tests ($\phi = 2$)

Now, by increasing the precision to $\phi = 10$, we obtained Figure 5.2. The approx-

imate UMP test still maintains the appropriate Type I error rate and the Wald test becomes more controlled for values of μ_0 near the boundary of 0. Even though the error rate for the t-test is still not controlled for these same values of μ_0, there is an apparent horizontal shift in the error rates across the sample sizes, which is attributed to the fact that there is less variability and skewness in the beta distribution for these scenarios.

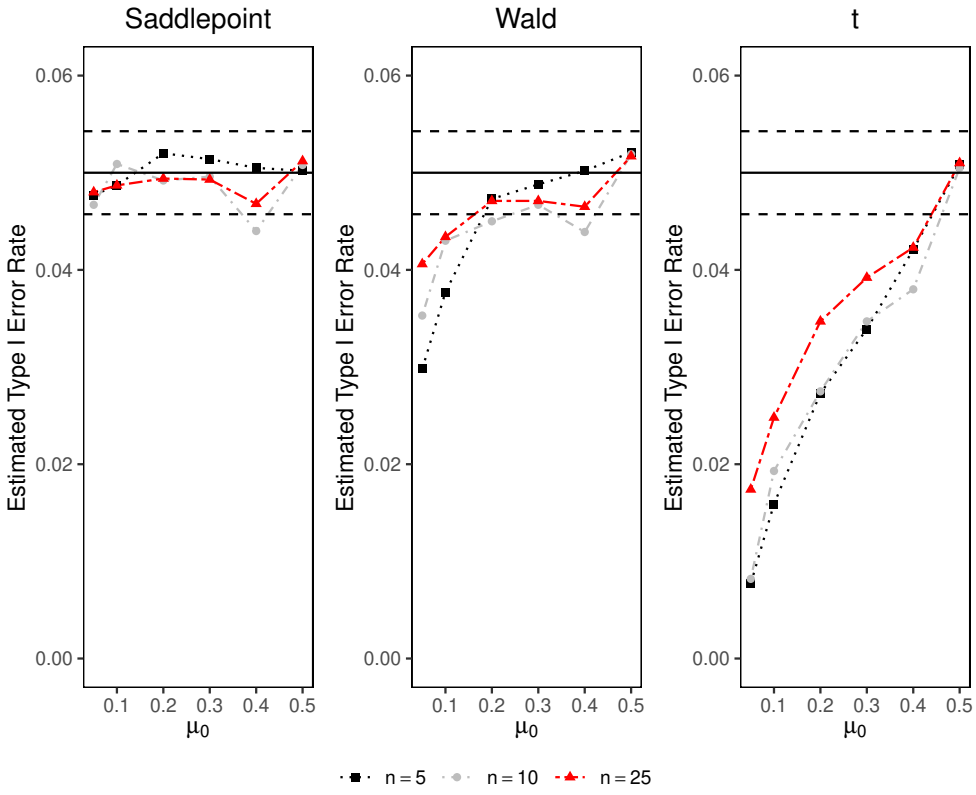

Figure 5.2: Estimated Type I Error Rates for Right-Tailed Tests ($\phi = 10$)

5.2.2 Left-Tailed Tests

A similar phenomenon exists for the estimated Type I error rates for left-tailed tests. Figure 5.3 demonstrates that for $\phi = 2$, the saddlepoint-derived approximate UMP test maintains the Type I error rate across all of the specified values of μ_0 and n. As for the Wald and t-test, their estimated Type I error rates are inflated for values of μ_0 relatively close to 0. (Recall that under the same settings, the rates for these two tests with a right-sided alternative were deflated.) Essentially, the Wald and t-test only appear to control the Type I error rate for varying neighborhoods about $\mu_0 = 0.5$. The distinction between the error rates of these two tests lies in the degree of inflation, as it is staggering for the t-test.

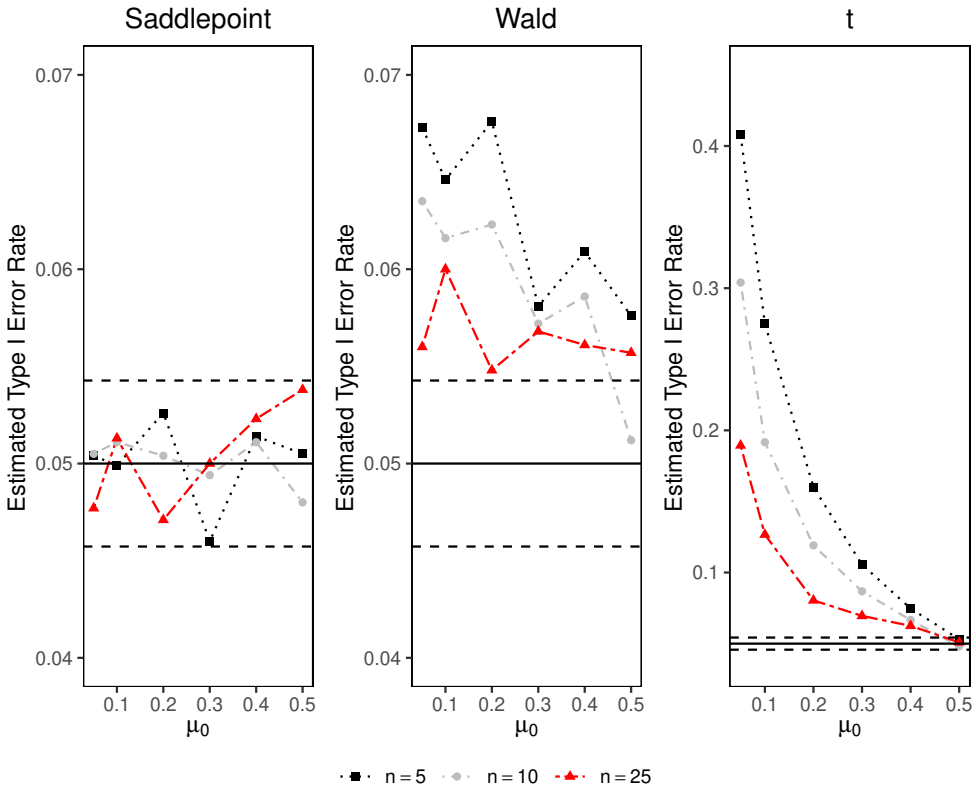

Figure 5.3: Estimated Type I Error Rates for Left-Tailed Tests ($\phi = 2$)

Now, for $\phi = 10$, Figure 5.4 provides the Type I error rate estimates for same type of test. Recall from previous simulations that higher values of ϕ abate the variance of the beta distribution, which is why there is a significant improvement in the severity of Type I error rate inflation for both the Wald and t-tests. As is to be expected, the error rate for the approximate UMP test still remains in a steady-state.

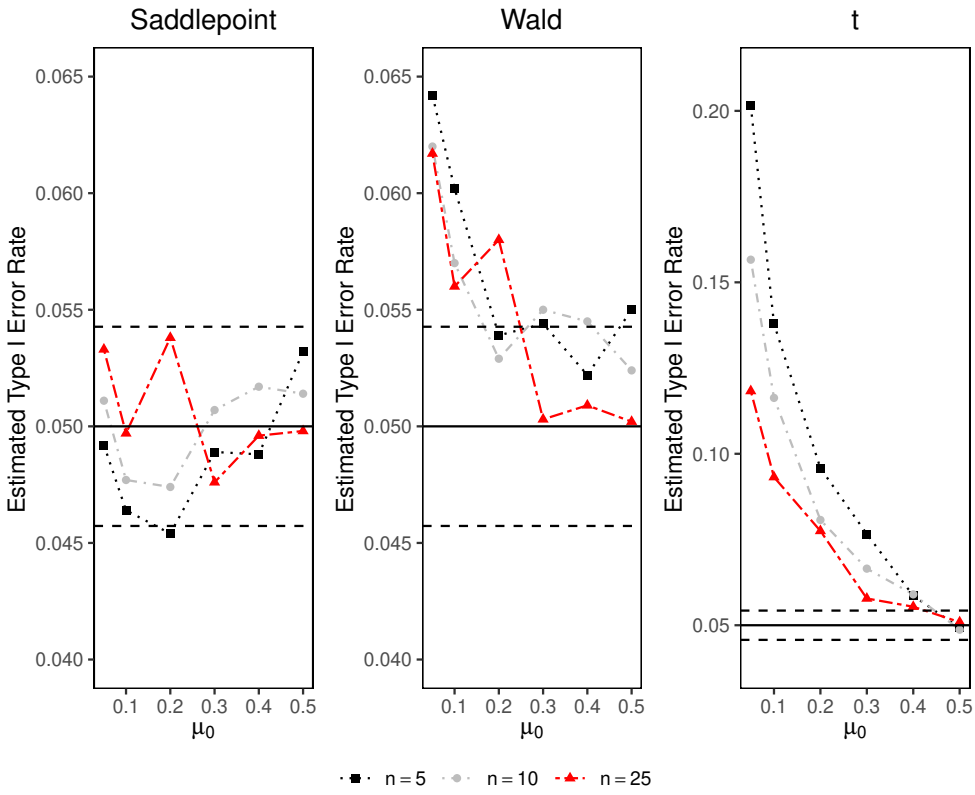

Figure 5.4: Estimated Type I Error Rates for Left-Tailed Tests ($\phi = 10$)

5.2.3 Two-Sided Tests

For two-sided testing, we also considered the asymptotic LRT in addition to the other tests. Figure 5.5 provides the Type I error rate estimates, from left to right, for the saddlepoint-derived test based on the "double" p-value approach, which was discussed in Section 4.2.3, followed by the LRT, the Wald test, and the t-test when $\phi = 2$. The most interesting observation in this simulation study is that all of the likelihood-based approaches maintained the error rate reasonably well for values of μ_0 relatively close to 0, with the LRT being slightly more conservative in some situations as compared to the Wald test. This was quite surprising to us at first given the nature

of the one-sided test results. Essentially, it appears that while each individual one-sided test performs poorly in terms of the Type I error rate, the deflation of the error rate in the right-tailed test is counter-balanced by the inflation of the error rate in the left-tailed test. As for values of μ_0 in an open neighborhood about 0.5, it appears the inflation in the error rate is caused by the slight inflation in the one-sided tests. Now, in regards to the t-test, there is still strong inflation in the Type I error rate, except now there is the inclusion of the scenario with $\mu_0 = 0.5$ and $n = 5$, which has the same line of reasoning as the Wald test.

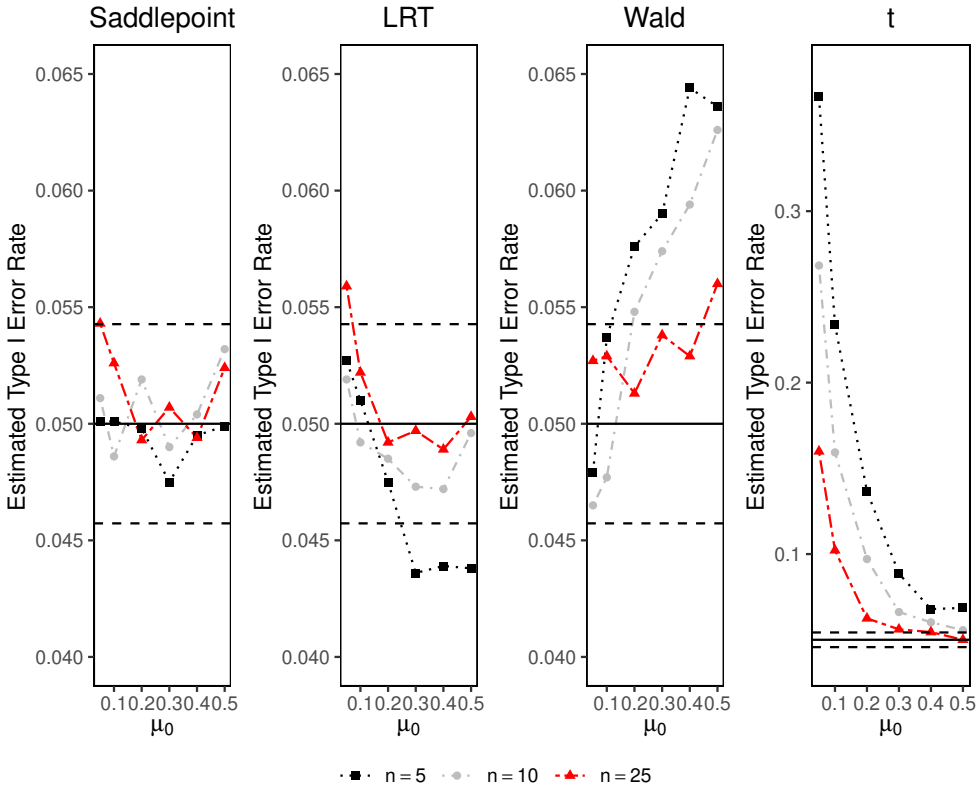

Figure 5.5: Estimated Type I Error Rates for Two-Tailed Tests ($\phi = 2$)

With $\phi = 10$, Figure 5.6 provides the simulation results. The saddlepoint-derived test, the LRT, and the Wald test all appropriately maintain the error rate, as the error rates for values of μ_0 relatively close to 0 are still counter-balnaced and relatively no

inflation in the one-sided Wald tests. The inflation in the error rate for the t-test has also decreased with respect to the scenarios with $\phi = 2$ in Figure 5.5.

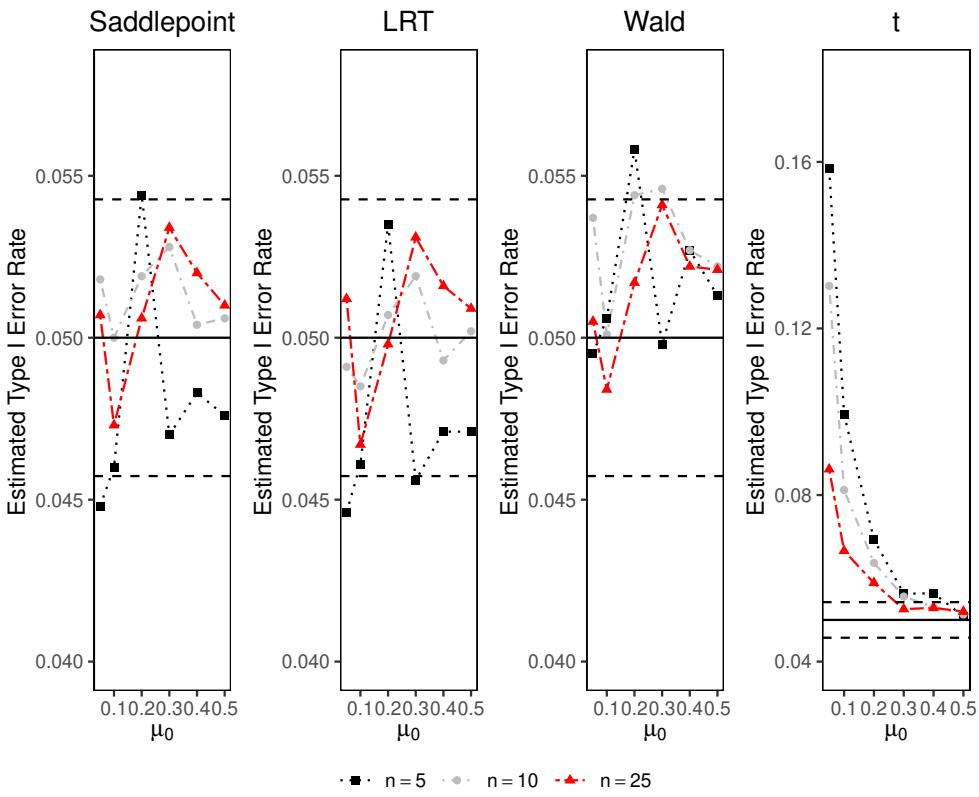

Figure 5.6: Estimated Type I Error Rates for Two-Tailed Tests ($\phi = 10$)

5.3 Power Simulations

As stated in Section 5.1, we solely focus on one-sided tests, where the prescribed value μ_0 is relatively close to 0, for the power simulations due to the fact that these tests exhibited significant discrepancies in terms of their estimated Type I error rates, while the ones for the two-sided tests were less notable.

Now, for a fixed value of ϕ, we present the simulated power curves for each test in a three-panel graph, where each panel considers a different sample size. We also use a solid-black horizontal line at 0.05 to help visualize whether or not the Type I error rate is being controlled.

5.3.1 Right-Tailed Tests

In the far-left panel of Figure 5.7, we can see that for $\phi = 2$, the Type I error rate is controlled for the saddlepoint-derived approximate UMP test (0.0496), but is significantly deflated for the Wald test (0.0267) and the t-test (0.0041). Due to these discrepancies, we cannot compare the tests in terms of statistical power, as they do not all achieve a size of 0.05. However, it is still enlightening to examine the power estimates at specific points in the alternative region. For instance, when $\mu = 0.175$, the saddlepoint-derived approximate UMP test has a power estimate of 0.3069, the Wald test has 0.2081, and the t-test has 0.0485.

A similar phenomenon occurs in terms of the estimated Type I error rates for each test as n increases, which can be seen in the middle and far-right panel of Figure 5.7. The absolute difference in the power estimates, with $n = 10$ at $\mu = 0.175$, for the approximate UMP test (0.5397) and the Wald test (0.4445) are relatively small as compared to the small sample scenario, while the t-test is still significantly different (0.1473). As for $n = 25$, the power estimates at $\mu = 0.175$ are relatively high and become virtually indistinguishable for the approximate UMP and Wald tests.

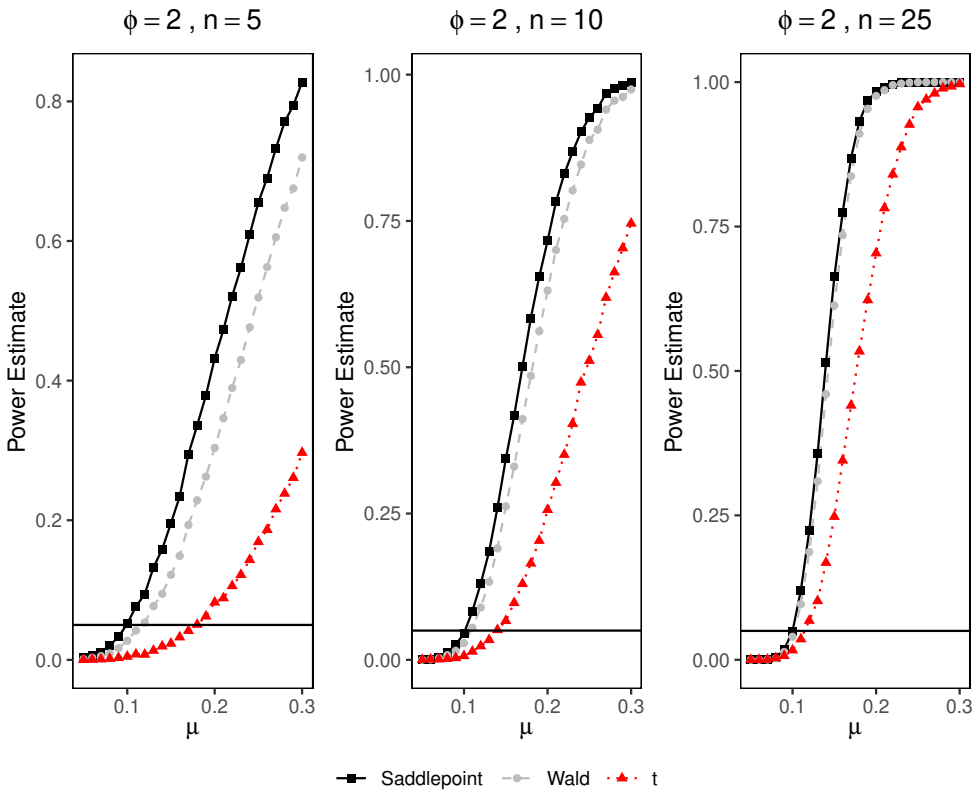

Figure 5.7: Simulated Power Curves for Right-Tailed Tests ($\mu_0 = 0.1, \phi = 2, n = 5, 10, 25$)

With $\phi = 10$ in Figure 5.8, the estimated Type I error rate for the approximate UMP test is controlled across all sample sizes, while the Wald and t-test exhibit some deflation. Therefore, the power curves are not comparable; however, we still examine the power estimates at $\mu = 0.175$ for consistency. With $n = 5$, the power estimates for the approximate UMP test, the Wald test, and the t-test are 0.5448, 0.4924, and 0.2576, respectively. If n increases to 10, then the power estimates are 0.817, 0.7897, and 0.6157, respectively. When $n = 25$, the power estimates for all three tests increase and become virtually indistinguishable. Essentially, as the sample size increases, the simulated power curves for the Wald and the t-test start to converge to the approximate UMP test.

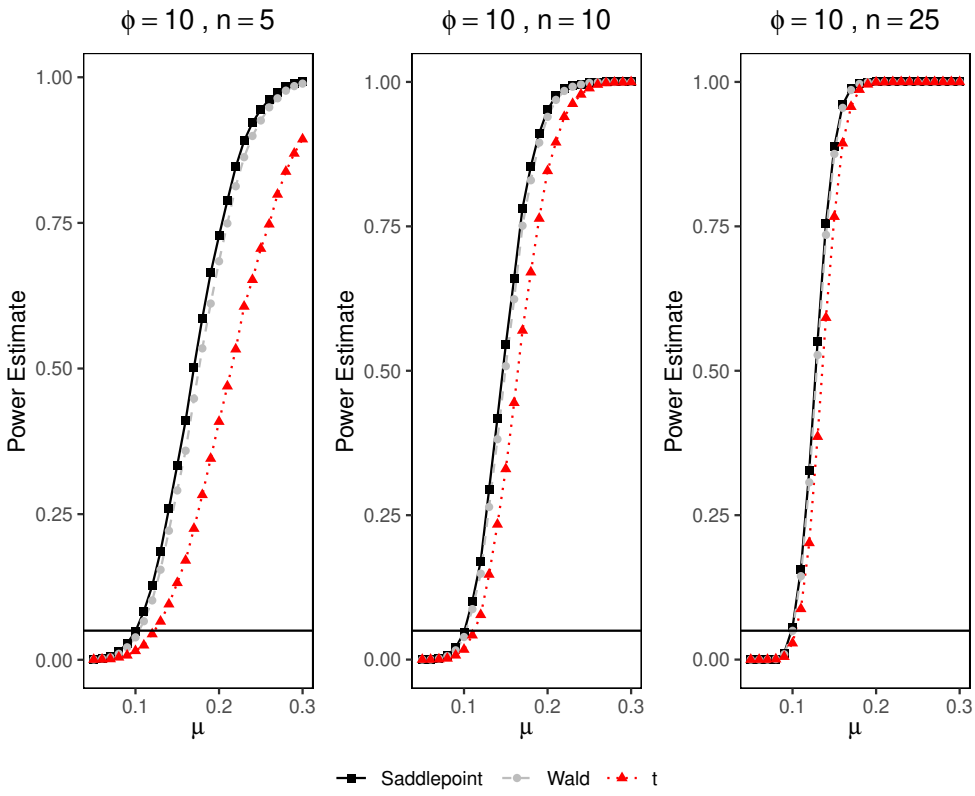

Figure 5.8: Simulated Power Curves for Right-Tailed Tests ($\mu_0 = 0.1, \phi = 10, n = 5, 10, 25$)

5.3.2 Left-Tailed Tests

As is stated in the Type I error simulations, the estimated Type I error rates for the saddlepoint-derived approximate UMP test are controlled, while the rates for the Wald and the t-test are inflated rather than deflated. With $n = 5$, as in the far-left panel of Figure 5.9, the power estimates for the approximate UMP test (0.0005) and the Wald test (0.0008) at $\mu = 0.175$ are virtually indistinguishable, while the t-test is discrepant (0.0746). By increasing n, as is demonstrated in the middle and far-right panel of Figure 5.9, the power estimate for the t-test starts to approach the estimate for the previous two tests.

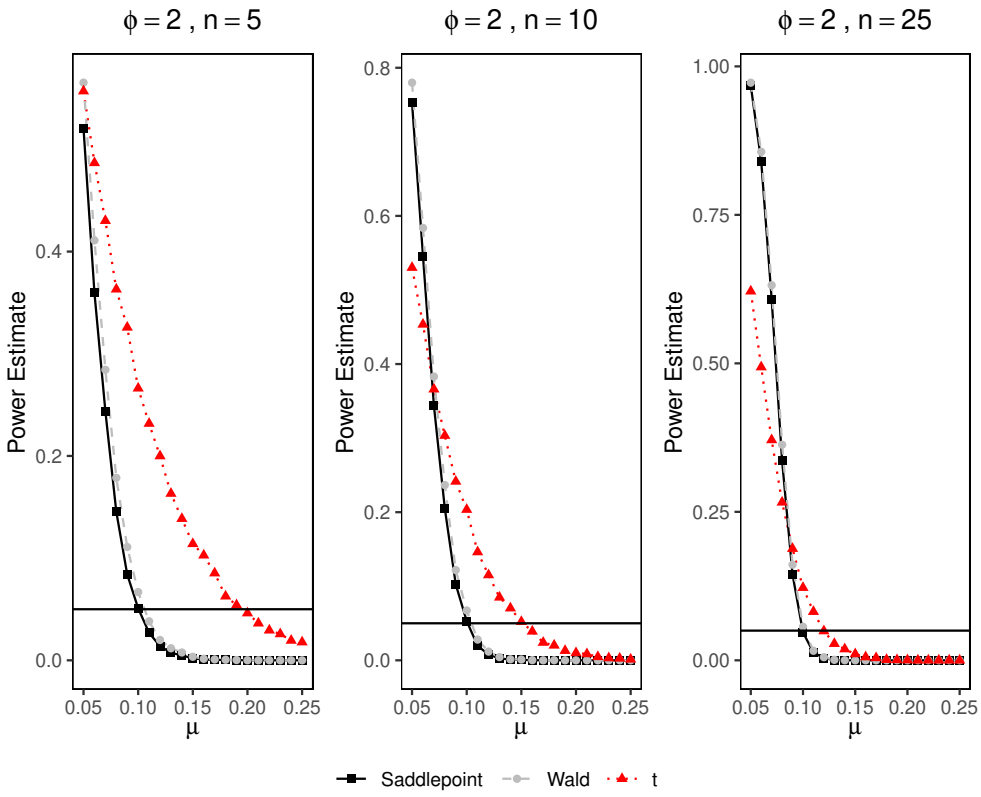

Figure 5.9: Simulated Power Curves for Left-Tailed Tests ($\mu_0 = 0.1, \phi = 2, n = 5, 10, 25$)

Now, if $\phi = 10$, then a similar story can be depicted in terms of the estimated Type I error rates and the power estimates at $\mu = 0.175$ from Figure 5.10. The power estimate for the approximate UMP test that controls the Type I error rate is virtually indistinguishable from the Wald test, which exhibits some inflation in the Type I error rate, across the values of n. As for the t-test with the significantly inflated error rate, the power estimate eventually approaches the rates of the previous two tests as n increases.

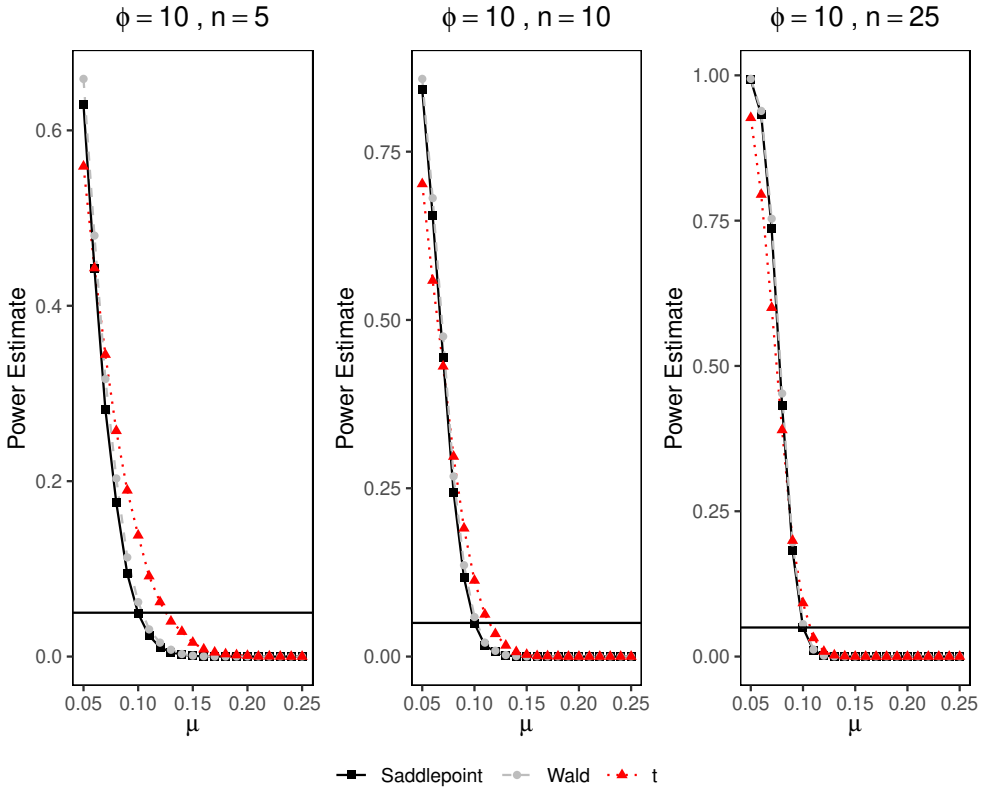

Figure 5.10: Simulated Power Curves for Left-Tailed Tests ($\mu_0 = 0.1, \phi = 10, n = 5, 10, 25$)

5.4 Coverage Probability Simulations

Since each confidence interval for μ is an inversion of the acceptance region for a two-sided test, then the results obtained in this simulation study should be consistent with the previously acquired empirical evidence for the two-sided tests. Now, in order to assess the scope of the consistency with the results, the coverage probability estimates under the scenarios specified in Section 5.1 are presented in a similar manner as in the Type I error rate simulations.

With Figure 5.11 as a reference, the confidence interval for μ in regards to the saddlepoint-derived test based on the "double" p-value approach provides appropriate coverage. Since $\phi = 10$, then the LRT and the Wald confidence intervals for μ also exhibit sufficient coverage probabilities for their respective intervals in most scenarios. The Wald confidence interval did tend to have slightly smaller coverage probabilities for values of μ_0 relatively close to 0.5 and small sample sizes. As for the the corresponding confidence interval for μ with respect to the two-sided t-test, there was not control in the coverage probabilities for small sample situations, which is consistent with Type I error simulations.

In addition to conducting coverage probability simulations, we also examined the widths of the confidence intervals for μ corresponding to each two-sided test. Figure 5.12 provides boxplots of the confidence interval widths for each method over 10000 simulations with $\mu_0 = 0.2$, $\phi = 10$, and $n = 5$. The confidence intervals for the saddlepoint, the LRT, and the Wald methods each have a consistent average width of around 0.2. These confidence intervals are also much more narrow than the t-interval, which addresses a negative property of this type of interval. The t-interval produced coverage probabilities that were lower than nominal, while simultaneously generating wider intervals as compared to the other procedures. Typically, an interval that provides this type of coverage produces more narrow intervals on average.

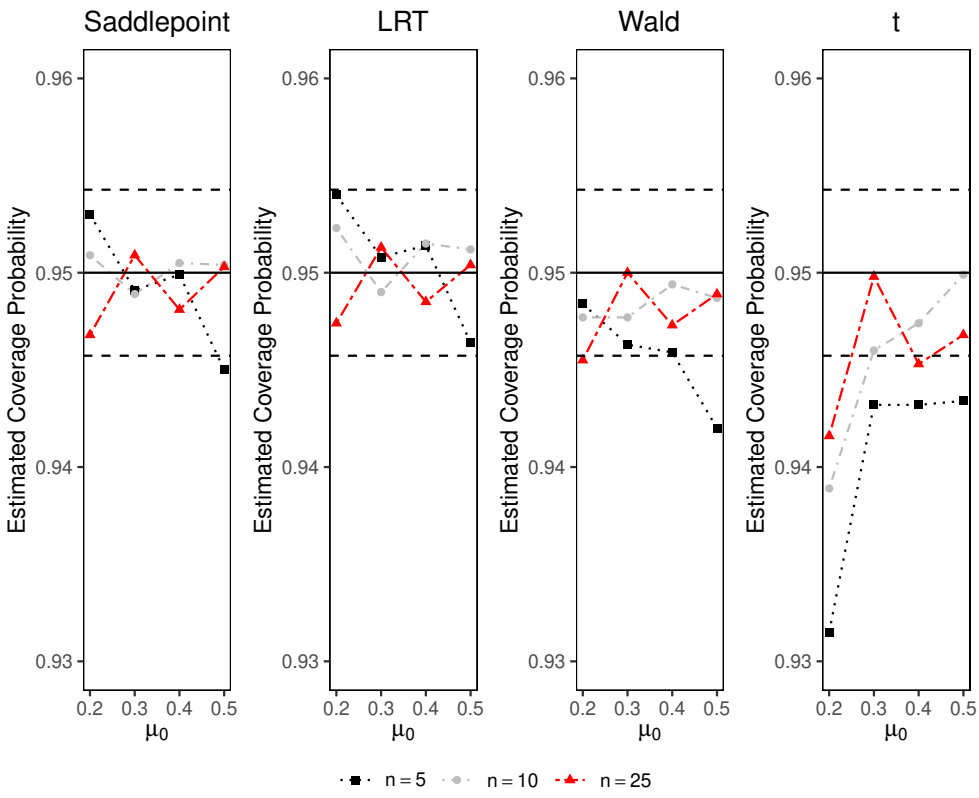

Figure 5.11: Estimated Coverage Probabilities of the Confidence Intervals for μ ($\phi = 10$)

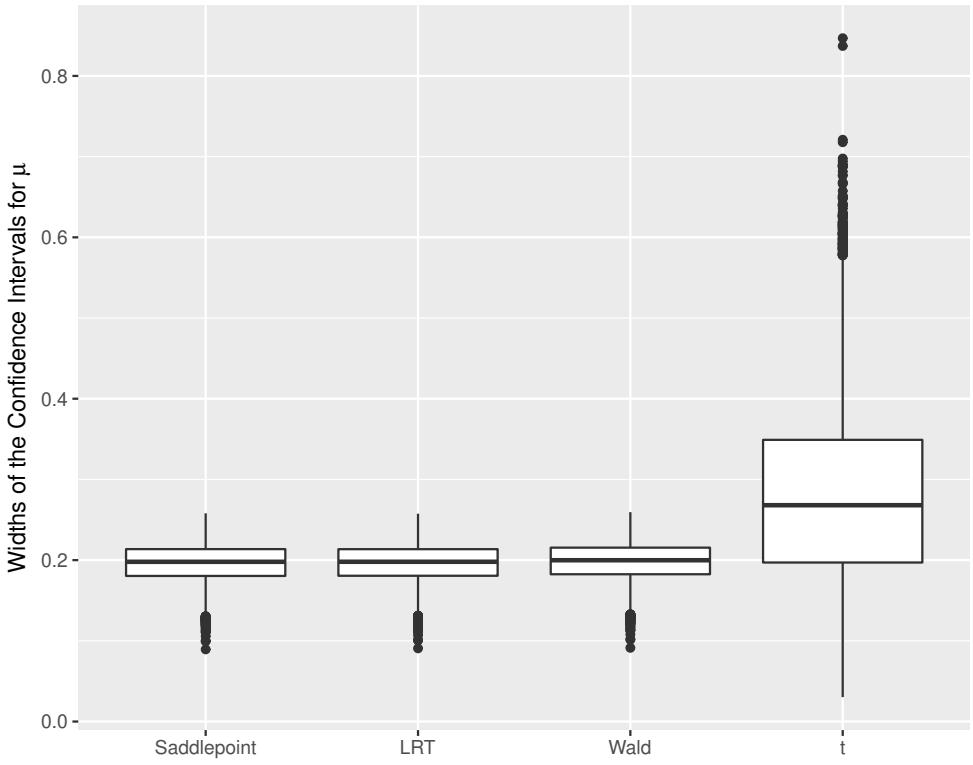

Figure 5.12: Boxplots of the Widths of the Confidence Intervals for μ ($\mu_0 = 0.2, \phi = 10, n = 5$)

5.5 Summary

The primary objective of our simulation studies was to inform the statistical community of some of the different aspects of the hypothesis tests and their corresponding confidence intervals for μ under consideration. Essentially, we were comparing the standard likelihood-based procedures to a relatively new method that incorporates the implementation of saddlepoint approximations. Our selection of these tests each had their advantages and disadvantages under specific scenarios. For values of μ_0 relatively close to 0, low precision, or small sample situations, the saddlepoint-derived approximate UMP test provided a sufficient alternative to the likelihood-based tests when working with one-sided tests, as this procedure exhibited control in the Type I error rate with equivalent or higher power estimates as compared to the procedures that roughly controlled this error rate.

For the two-sided tests and their corresponding confidence intervals for μ, the choice of procedure appeared to be less concerning based on our studies of the Type I error rate. However, due to various computational difficulties discussed in Section 5.1, we were unable to extensively analyze the performance of the confidence intervals for values of μ_0 relatively close to 0, where the hypothesis tests based on saddlepoint approximations tended to perform more efficiently in small sample size situations. Therefore, we leave this for future developmental work.

Lastly, our simulations demonstrated how poorly the t-test performed under the specified scenarios of the beta model, which predominantly consisted of small sample situations coupled with highly skewed and variable distributions. As these are very scenarios we have observed in real FCM derived data sets in Chapter 1, we wanted to provide data analysts that work with this type of data, or any data with bounded support, with additional empirical evidence as to the discrepancies associated with the t-test under these scenarios.

6 APPLICATIONS

6.1 Hypothesis Testing in Immunology Research

Changes in cell compositions within the blood of humans and other mammals provide a lens into understanding the kinetics of how the immune system responds to foreign bodies. At the time of this manuscript's completion, the massive impact of the disease COVID-19 is weighing heavily on the nation. It highlights the importance of accurate and sound statistical practices to help researchers understand how viruses can impact our immune system.

In a true research system, an analysis would involve conducting experiments involving multiple groups to justify a two-sample procedure to compare population means. Although the methods discussed in this manuscript do not address this question directly, the work presented here provides the necessary information to easily transition to more complicated study designs involving beta models.

To illustrate the utility of our developed methods discussed in Chapter 4, we provide a simple modified example from a longitudinal study of cynomolgus macaques that were challenged with Mycobacterium tuberculosis. (This example is for educational purposes only and does not reflect a scientifically rigorous hypothesis.) Our data set consists of measurements taken from a blood draw. By using a flow cytometer, the percentages of CD20+ B-cells were observed 180 days after tuberculosis infection was induced. The 17 data points are listed here: (3.83, 1.59, 3.64, 1.94, 4.65, 2.83, 4.37, 2.57, 0.684, 13.5, 4.91, 4.84, 2.77, 2.2, 8.97, 3.87, 3.43). The distribution of CD20+ B-cells among healthy humans, and arguably healthy cynomolgus macaques as well, is known to be skewed-right and have an average percent of around 8%. Also, through rigorous inspections of previously derived FCM data sets amongst

various settings performed by Turner [13], a conservative value for the precision parameter associated with the beta model for CD20+ B-cell percentages was established at $\phi = 10$.

Now, suppose that a researcher is interested in testing if the average proportion of healthy macaques' CD20+ B-cell composition, which were obtained 180 days after tuberculosis infection, is less than 0.08. Formally, we will test the following set of hypotheses: $H_0 : \mu \geq 0.08$ against $H_1 : \mu < 0.08$.

The sufficient statistic for μ in the context of the data set is
$T = 10 \sum_{i=1}^{17} \ln\{X_i/(1 - X_i)\} = -568.234$. As this is a left-tailed test, the p-value using the saddlepoint CDF, $\hat{F}_T$, under H_0 is $\hat{F}_T(-568.234) = 0.273$. The MLE of μ is $\hat{\mu} = 0.072$ and the corresponding Wald Z-statistic is $Z = -0.567$. By using the standard normal CDF, Φ, the p-value is $\Phi(-0.567) = 0.285$. Lastly, the t-test produces a p-value of 0.00004.

The comparison of these three tests essentially alludes to the general trend of our simulations regarding left-tailed tests. Recall that with a modest precision parameter and sample size, the t-test had a drastically inflated Type I error rate, which explains the discrepancy between the statistically significant p-value for the t-test and the more conservative results produced by the other two procedures. Even though the simulations results indicated that there was a slight inflation in the Type I error rate for the Wald test, the saddlepoint-derived approximate UMP test and Wald test are generally more comparable.

6.2 Sample Size Logistic

To obtain grants funded by the National Institutes of Health (NIH) for FCM research projects, grant proposals must justify that the immunologists will obtain an adequate number of samples in order to increase the chances of concluding the

research hypothesis if this hypothesis is indeed true. A typical specification associated with these types of proposals is stating the sample size required to ensure that the statistical power of the hypothesis test is 0.80, or higher, for a specified value of μ that is clinically relevant under the research hypothesis.

As a motivating example, suppose that we are interested in testing $H_0 : \mu \leq 0.1$ against $H_1 : \mu > 0.1$ at $\alpha = 0.05$ with $\phi = 10$ and we want to obtain the value of the sample size n that yields a power estimate of 0.80 at $\mu = 0.12$ for both the saddlepoint-derived approximate UMP test and the t-test. We will now derive the mathematical statements needed for each power calculation.

Recall that the rejection region for the t-test under the structure of our statistical inference problem is to reject H_0 if

$$\frac{\bar{X} - \mu_0}{s_{\mu_0}/\sqrt{n}} > t_{n-1,\alpha},$$

where $s_{\mu_0} = \sqrt{\frac{\mu_0(1-\mu_0)}{\phi+1}}$ and $t_{n-1,\alpha}$ denoted the upper-α point of the t-distribution with $n-1$ degrees of freedom. With this rejection region, the power function for the t-test, $\beta_t(\mu)$, is defined as

$$\beta_t(\mu) = \Pr_\mu\left\{\frac{\bar{X} - \mu_0}{s_{\mu_0}/\sqrt{n}} > t_{n-1,\alpha} | \mu = \mu\right\},$$

and by algebraically rearranging the rejection region, $\beta_t(\mu)$ can also be expressed as

$$\beta_t(\mu) = \Pr_\mu\left\{\bar{X} > t_{n-1,\alpha}\frac{s_{\mu_0}}{\sqrt{n}} + \mu_0 | \mu = \mu\right\},$$

Now, to impose the condition that the true parameter is μ rather than μ_0, we can standardize $\bar{X}$ by using the following distributional statement: $\bar{X} \dot{\sim} N\big(\mu = \mu, s_\mu^2 = \frac{\mu(1-\mu)}{\phi+1}\big)$, which accurately reflects the true value of the parameter. After the proper standardization of $\bar{X}$ and through algebraic manipulation, we obtain

$$\beta_t(\mu) = \Pr\left\{t_{n-1} > t_{n-1,\alpha}\frac{s_{\mu_0}}{s_\mu} + (\mu_0 - \mu)\frac{\sqrt{n}}{s_\mu}\right\}. \tag{6.1}$$

As for the approximate UMP test, we reject H_0 if $T_n = \phi \sum_{i=1}^{n} \ln\{X_i/(1-X_i)\} > t_u$, where the subscript n in T_n is used to indicate a dependence on n and the critical point t_u is defined to be upper-α quantile of the distribution of T that satisfies $\Pr\{T_n > t_u|\mu = \mu_0\} = \alpha$. With this rejection region, we can define the power function for the approximate UMP test, $\beta_{T_n}(\mu)$, as

$$\beta_{T_n}(\mu) = \Pr_{\mu}\{T_n > t_u|\mu = \mu\}.$$

As algebraic manipulation is not practical in terms of this rejection region, we impose the condition that the true value of the parameter is μ, rather than μ_0, by substituting this value into the saddlepoint CDF of T_n; more specifically,

$$\beta_{T_n}(\mu) = 1 - \hat{F}_{T_n}(t_u; \mu = \mu, \phi = \phi). \tag{6.2}$$

Since we want both (6.1) and (6.2) to be 0.80 and n is the unknown value, as all of the other variables are defined for this problem, then we can use the same one-dimensional root-finding algorithm as mentioned in previous chapters. With this, we were able to determine that the required sample size is approximately 136 for the t-test and 101 for the approximate UMP test. We can visualize this discrepancy by rendering a plot of the power estimates for the two tests against a sequence of values for the sample size n, which is provided in Figure 6.1, and identify the two intersections with the solid-black horizontal line at 0.80. This is a significant discrepancy not only for this scenario, but for other prescribed power levels in Figure 6.1.

Now, in the context of funding for FCM projects, this result essentially signifies that they ensure the required statistical power for the clinically relevant value of μ with a much smaller sample size when using the approximate UMP test. A smaller sample size is advantageous for immunologists, as experiments in this area of research are often expensive.

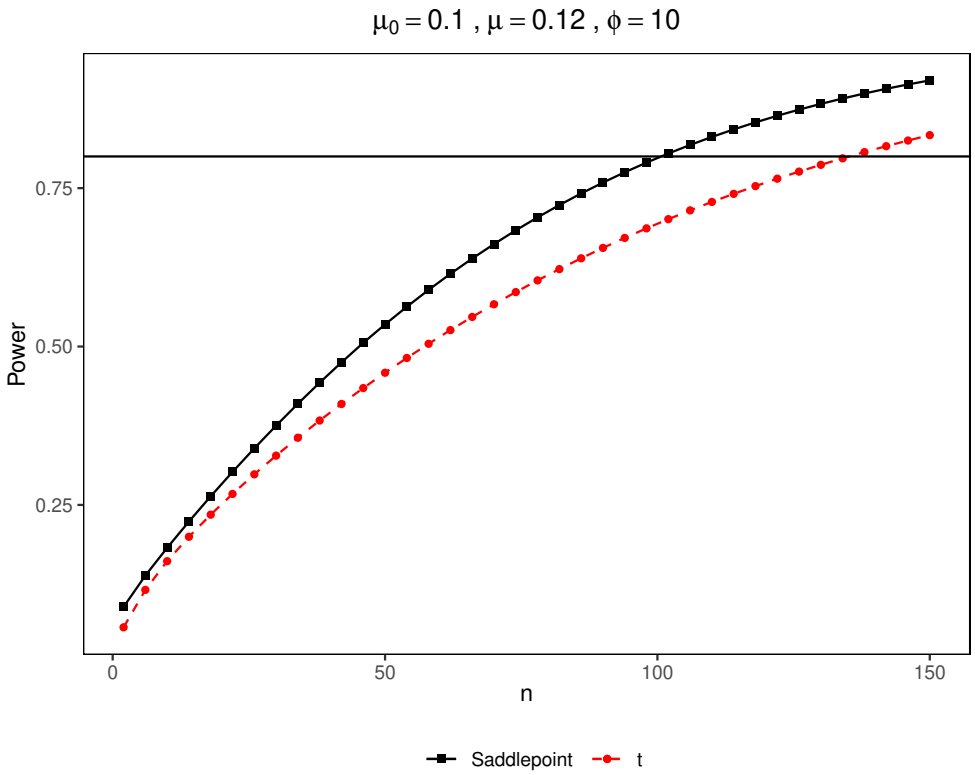

Figure 6.1: Sample Size Determination for the Approximate UMP Test and the t-Test ($\mu_0 = 0.1, \mu = 0.12, \phi = 10$)

7 CONCLUSIONS AND FUTURE RESEARCH

The purpose of this research was to expound upon the mathematics pertinent to conducting statistical inference for the mean of beta distribution based on a simple random sample and to assess the efficacy of our statistical methods. For our purposes, we formulated the standard likelihood-based procedures for hypothesis testing and confidence intervals, along with other testing procedures that were derived from the implementation of saddlepoint approximations.

Through simulation work, we were able to provide empirical evidence for various aspects of the hypothesis tests and their corresponding confidence intervals for μ under consideration. The saddlepoint-derived approximate UMP test controlled the Type I error rate across all of the versatile scenarios coinciding with FCM projects, which include values of μ_0 relatively close to 0, low precision, and small sample size situations. The one-sided asymptotic LRT and Wald tests exhibited some control in their respective error rates for sufficiently large sample sizes. When the alternative was two-sided and for values of μ_0 relatively close to 0 and low precision, the error rates from the right-tailed tests negated the error rates from the left-sided tests. As for the t-test, there was not appropriate control in the error rate, especially in situations prevalent in FCM projects.

Based on the simulation results, our overarching recommendation is to use the saddlepoint approximation method given that the saddlepoint approximations can be implemented numerically. Recall from Section 5.1 that there were difficulties in practical computations for values of μ_0 exceedingly close to 0 due to degenerate samples that caused non-finite values in preliminary estimates and thus numerical instability in root-finding algorithms. As these situations most likely do not exhibit

real world phenomenon, we do not expect this to significantly hinder conducting statistical inference in a practical setting. Furthermore, we also suggest using the either the LRT and the Wald test as an alternative, as there were some disadvantages in the settings used in immunology research. As for the Student's t-test, we do not recommend using this type of hypothesis testing procedure, especially in FCM projects.

After acquiring a sense of the scope of the applicability for each hypothesis testing procedure, we proceeded to provide a simple example using FCM data to demonstrate the utility of the hypothesis tests as seen through our simulation studies. In addition to this example, we also highlighted a significant property of the saddlepoint approximation method, which is its readily available approximate power functions. This allows for more expeditious and accurate sample size determinations using a method that not only appropriately controls the Type I error rate, but essentially gives a more realistic view of the sample sizes needed to obtain highly powerful study designs.

Now, as for future research, recall that the basis of this manuscript is contingent upon the assumption that ϕ is known in advanced, or that we can at least obtain an estimate for ϕ from previous studies. Since a known value of ϕ is highly unlikely in new areas of research, we intend to extend our testing procedures, particularly with the saddlepoint approximation method, to construct both approximate UMP and Uniformly Most Powerful Unbiased (UMPU) tests, while accounting for the nuisance parameter, ϕ. Fortunately, the beta distribution under the two-parameter model, where both μ and ϕ are unknown, is still a member of the exponential family. Therefore, marginal and joint CGFs will be easily obtained through a multivariate extension of the results provided in this manuscript. The continuation of this manuscript is currently being executed by Jacob Turner, Ph.D.

Upon completion of this future work, we will have effectively created a set of procedures for the beta distribution that is analogous to the standard Z and t-tests for

the normal distribution. From there, we will proceed to the more general two-sample comparisons and ANOVA study design settings, which will be more directly applicable in studies performed by data analysts and researchers in the FCM community.

BIBLIOGRAPHY

[1] O. E. Barndorff-Nielsen, *Inference on Full or Partial Parameters Based on the Standardized Signed Log Likelihood Ratio*, Biometrika **73** (1986), no. 2, 307–322.

[2] ———, *Approximate Interval Probabilities*, Journal of the Royal Statistical Society. Series B (Methodological) **52** (1990), no. 3, 485–496.

[3] ———, *Modified Signed Log Likelihood Ratio*, Biometrika **78** (1991), no. 3, 557–563.

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

[5] G. Casella and R.L. Berger, *Statistical Inference*, Duxbury advanced series in statistics and decision sciences, Thomson Learning, 2002.

[6] H. E. Daniels, *Saddlepoint Approximations in Statistics*, The Annals of Mathematical Statistics **25** (1954), no. 4, 631–650.

[7] Silvia Ferrari and Francisco Cribari-Neto, *Beta Regression for Modelling Rates and Proportions*, Journal of Applied Statistics **31** (2004), no. 7, 799–815.

[8] Hannah P. Gideon, Jason A. Skinner, Nicole Baldwin, JoAnne L. Flynn, and Philana Ling Lin, *Early Whole Blood Transcriptional Signatures Are Associated with Severity of Lung Inflammation in Cynomolgus Macaques with Mycobacterium tuberculosis Infection*, The Journal of Immunology **197** (2016), no. 12, 4817–4828.

[9] A.K. Gupta and S. Nadarajah, *Handbook of Beta Distribution and Its Applications*, Statistics: A Series of Textbooks and Monographs, Taylor & Francis, 2004.

[10] R.V. Hogg, J.W. McKean, and A.T. Craig, *Introduction to Mathematical Statistics*, Pearson education international, Pearson Education, 2005.

[11] N.L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Wiley series in probability and mathematical statistics: Applied probability and statistics, no. v. 2, Wiley & Sons, 1995.

[12] T. Minka, *The Dirichlet-Tree Distribution*, This unpublished paper is available online at http://research.microsoft.com/minka/papers/dirichlet/minkadirtree.pdf, 1999.

[13] Jacob Turner, *A Novel Approach to Modeling Immunology Data*, Ph.d. dissertation, Southern Methodist University, 6425 Boaz Ln, Dallas, TX 75205, May 2013, pp. 1–123.

[14] G.A. Young and R.L. Smith, *Essentials of Statistical Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2005.

VITA

Bryn Brakefield received a Bachelor's degree in Mathematics from Stephen F. Austin State University in 2017. She began her work towards a Master's degree in Mathematics at Stephen F. Austin State University in the Spring of 2018 and is expected to graduate in May 2020.

Permanent Address: PO Box 13040 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 Bryn Brakefield using LaTeX.