2011

Exploring Confidence Intervals in the Case of Binomial and Hypergeometric Distributions

Irene Mojica
University of Central Florida

EXPLORING CONFIDENCE INTERVALS IN THE CASE OF BINOMIAL AND HYPERGEOMETRIC DISTRIBUTIONS

by

IRENE MOJICA
B.S. The College of Staten Island, City University of New York, 1992

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Mathematics
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2011

ABSTRACT

The objective of this thesis is to examine one of the most fundamental and yet important methodologies used in statistical practice, interval estimation of the probability of success in a binomial distribution.

The textbook confidence interval for this problem is known as the Wald interval as it comes from the Wald large sample test for the binomial case. It is generally acknowledged that the actual coverage probability of the standard interval is poor for values of p near 0 or 1. Moreover, recently it has been documented that the coverage properties of the standard interval can be inconsistent even if p is not near the boundaries. For this reason, one would like to study the variety of methods for construction of confidence intervals for unknown probability p in the binomial case. The present thesis accomplishes the task by presenting several methods for constructing confidence intervals for unknown binomial probability p.

It is well known that the hypergeometric distribution is related to the binomial distribution. In particular, if the size of the population, N, is large and the number of items of interest k is such that $\frac{k}{N}$ tends to p as N grows, then the hypergeometric distribution can be approximated by the binomial distribution. Therefore, in this case, one can use the confidence intervals constructed for p in the case of the binomial distribution as a basis for construction of the confidence intervals for the unknown value k = pN. The goal of this thesis is to study this approximation and to point out several confidence intervals which are designed specifically for the hypergeometric distribution. In particular, this thesis considers several confidence intervals

which are based on estimation of a binomial proportion as well as Bayesian credible sets based on various priors.

ACKNOWLEDGMENTS

I would like to express my appreciation and gratitude to my professors, family, and friends for their support throughout my graduate studies. These last two years have been extremely challenging having undergone major surgery and months of wound care. Without the warm words of encouragement and support of all, I would not have been able to return to finish my studies.

I am tremendously thankful for the patient supervision and guidance of Dr. Pensky. Her insightfulness and constructive comments have been instrumental in the completion of this thesis.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

The objective of the present paper is to re-visit one of the most basic and methodologically important problems in statistical practice, namely, interval estimation of the probability of success in a binomial distribution. There is a textbook confidence interval for this problem that has acquired nearly universal acceptance in practice. This interval is of the form

$$\hat{p} \pm z_{\alpha/2}\, n^{-1/2}\big(\hat{p}(1-\hat{p})\big)^{1/2} \tag{1.1}$$

where $\hat{p} = \frac{X}{n}$ is the sample proportion of successes, and $z_{\alpha/2}$ is the 100 $(1 - \alpha/2)$ th percentile of the standard normal distribution. The interval is easy to present and motivate and easy to compute. The standard interval is known as the Wald interval as it comes from the Wald large sample test for the binomial case.

It is widely recognized, however, that the actual coverage probability of the standard interval is poor for p near 0 or 1. Even at the level of introductory statistics texts, the standard interval is often presented with the condition that it should be used only when

$$\min\big(np, n(1-p)\big) \geq 5\ (or\ 10).$$

Moreover, recently it has also been pointed out that the coverage properties of the standard interval can be erratically poor even if p is not near the boundaries. For this reason, one would like to study the variety of methods for construction of confidence intervals for unknown probability p in the binomial case.

Chapters 2 and 3 of the thesis accomplish the task by presenting a variety of methods for construction of the confidence intervals for unknown binomial probability p. In particular, Chapter 2 is dedicated to construction of the confidence intervals by frequentist techniques while Chapter 3 considers interval estimators based on Bayesian methodology.

It is well known that hypergeometric distribution is related to binomial distribution. In particular, if the size of population N is large and the number of items of interest k is such that

$$\frac{k}{N} \rightarrow p \text{ as } N \rightarrow \infty,$$

the hypergeometric distribution can be approximated by binomial. Therefore, in this case one can use confidence intervals constructed for p in the case of the binomial distribution as a basis for construction of the confidence intervals for the unknown value $k = pN$.

The goal of Chapter 4 is to study this approximation and also to point out several confidence intervals which are designed specifically for the hypergeometric distribution. In addition, we study several confidence intervals which are based on estimation of a binomial proportion as well as Bayesian credible sets based on various priors.

The rest of the paper is organized as follows. In Chapter 2 we consider frequentist techniques for interval estimation. In particular, we provide background information about the standard Wald interval. The following sections explain the construction of the alternative intervals that were discussed by Brown, Cai and DasGupta (2001, 2002). We explore the Wilson interval, the Agresti-Coull interval, the logit interval, the likelihood interval, the Clopper-Pearson interval and the arcsine interval.

Chapter 3 is dedicated to Bayesian techniques for construction of confidence intervals which are called “credible intervals” in this case. Section 3.1 lays down the foundations of the

Bayesian approach to statistics. Construction of Bayesian credible sets is discussed in Section 3.2. Finally, in Section 3.3, we develop a confidence interval for the binomial distribution using the noninformative Jeffreys' prior distribution.

In Chapter 4, we introduce the hypergeometric distribution and show how it can be approximated to the binomial distribution. Sections 4.3 and 4.4 show the construction of confidence intervals based on the normal approximation to hypergeometric distribution and an analog of the Wilson interval that was used for the binomial distribution. We consider the Bayesian approach for the hypergeometric distribution in Section 4.5. The priors used for designing the Bayesian confidence intervals are the binomial prior considered in Section 4.5.1 and the Polya (beta-binomial) prior, used in Section 4.5.2.

Finally, Chapter 5 brings conclusions about the construction of confidence intervals for the binomial and the hypergeometric distributions.

CHAPTER TWO: CONFIDENCE INTERVALS FOR THE SUCCESS PROBABILITY IN BINOMIAL DISTRIBUTIONS

2.1 – The Standard Interval

The standard confidence interval for the estimation of the probability of success in a binomial proportion is widely accepted because of its simplicity of presentation and computation. However, it has been shown to have problems with the actual coverage probability. The justification for using this interval is based on the central limit theorem (CLT) which states that when the sample size is large, the number of successes in the binomial distribution can be approximated by a normal variable.

The standard interval, known as the Wald interval, is of the form

$$\hat{p} \pm z_{\alpha/2}\, n^{-1/2}\big(\hat{p}(1-\hat{p})\big)^{1/2} \qquad (2.1.1)$$

where $\hat{p} = \frac{X}{n}$ is the sample proportion of successes, (2.1.2)

$\hat{q} = 1 - \hat{p}$ is the proportion of failures, and

$z_{\alpha/2}$ is the 100 $(1 - \alpha/2)^{\text{th}}$ percentile of the standard normal distribution.

This interval guarantees that for any fixed $p \in (0,1)$ the coverage probability,

$$C(p,n) \to 1-\alpha \;\; as\; n \to \infty \quad \text{where } C(p,n) = P_p(p \in CI),\; 0 < p < 1.$$

The standard normal confidence interval $CI_S = \hat{p} \pm z_{\alpha/2}\, n^{-1/2}(\hat{p}\hat{q})^{1/2}$ is obtained by inverting the acceptance region of the Wald large sample normal test for a general problem. Let

θ be the generic parameter for p, $\hat{\theta}$ is the maximum likelihood estimate of θ which is $\hat{p} = \frac{X}{n}$, $\widehat{se}(\hat{\theta})$ is the estimated standard deviation of $\hat{\theta}$, $\sqrt{\frac{\hat{p}\hat{q}}{n}}$ and $\kappa = \; z_{\alpha/2}$. Replacing these into the equation

$$\left| \frac{(\hat{\theta} - \theta)}{\widehat{se}(\hat{\theta})} \right| \leq \kappa \tag{2.1.3}$$

results in the following equation:

$$\left| \frac{(\hat{p} - \theta)}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} \right| \leq \kappa$$

By solving the following equation for θ, $\hat{p} - \theta = \pm \kappa\sqrt{\frac{\hat{p}\hat{q}}{n}}$ we obtain the standard confidence interval for θ of the form:

$$\hat{p} - \kappa\sqrt{\frac{\hat{p}\hat{q}}{n}} \leq \theta \leq \hat{p} + \kappa\sqrt{\frac{\hat{p}\hat{q}}{n}} \tag{2.1.4}$$

Many textbooks present the standard interval with the provisions that the interval only be used when $\min\big(np, n(1-p)\big) \geq 5\ (or\ 10)$ because of the interval has poor coverage properties when p is near the boundaries, 0 and 1. It has also been found to have inadequate coverage even when p is not near the boundaries. The oscillatory and poor behavior of the standard interval can be attributed to the discrete nature of the binomial distribution as well as its skewness. The exact nominal confidence level cannot be achieved without a randomized procedure (Brown et al., 2001)

2.2 – The Wilson Interval

The Wilson interval is an alternate interval which uses the exact standard error $\sigma = (pq)^{1/2}\, n^{-1/2}$ instead of the estimated standard deviation $\hat{\sigma} = (\hat{p}\hat{q})^{1/2}\, n^{-1/2}$. The Wilson interval is obtained by inverting the CLT approximation to the family of equal-tailed tests of the hypothesis H_0 : $p = p_0$. If the interval includes p_0, then one accepts H_0. The Wilson interval has an actual coverage probability that is closer to the nominal value than the standard interval; however, it still oscillates when p is close to the boundaries, 0 and 1. A modification can be made to eliminate the downward spikes at 0 and 1. This interval is recommended for small n, $n \leq 40$.

The Wilson confidence interval is

$$CI_W = \frac{X + \frac{\kappa^2}{2}}{n + \kappa^2} \pm \frac{\kappa n^{1/2}}{n + \kappa^2}\left(\hat{p}\hat{q} + \frac{\kappa^2}{4n}\right)^{1/2} \qquad (2.2.1)$$

is obtained by solving inequality

$$\left|\frac{\hat{p} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\right| \leq \kappa \qquad (2.2.2)$$

which requires the quadratic formula to solve for θ. Clearing the denominator and squaring both sides yields

$$\hat{p}^2 - 2\hat{p}\theta + \theta^2 \leq \frac{\kappa^2\theta}{n} - \frac{\kappa^2\theta^2}{n} \qquad (2.2.3)$$

which can be re-grouped to obtain the quadratic equation

$$\left(1 + \frac{\kappa^2}{n}\right)\theta^2 - \left(2\hat{p} + \frac{\kappa^2}{n}\right)\theta + \hat{p}^2 \leq 0$$

Using the quadratic formula with the terms A = $1+\frac{\kappa^2}{n}$, B = $-\left(2\hat{p}+\frac{\kappa^2}{n}\right)$ and C = $\hat{p}^2$gives

$$\theta \leq \frac{2\hat{p}+\frac{\kappa^2}{n} \pm \sqrt{\left(2\hat{p}+\frac{\kappa^2}{n}\right)^2 - 4\left(1+\frac{\kappa^2}{n}\right)(\hat{p}^2)}}{2\left(1+\frac{\kappa^2}{n}\right)} \tag{2.2.4}$$

Multiplying the terms in the radicand gives $4\hat{p}^2 + \frac{4\hat{p}\kappa^2}{n} + \frac{\kappa^4}{n^2} - 4\hat{p}^2 - \frac{4\kappa^2\hat{p}^2}{n}$, grouping the terms

$$\theta \leq \frac{\frac{n}{2}\left[2\hat{p}+\frac{\kappa^2}{n} \pm \sqrt{\frac{4\hat{p}(1-\hat{p})\kappa^2}{n}+\frac{\kappa^4}{n^2}}\right]}{n+\kappa^2}$$

Simplification of the radical and the numerator

$$\theta \leq \frac{n\hat{p}+\frac{\kappa^2}{2} \pm \kappa\sqrt{n}\sqrt{\hat{p}(1-\hat{p})+\frac{\kappa^2}{4n}}}{n+\kappa^2} \tag{2.2.5}$$

and finally after substituting $\frac{X}{n}$ for $\hat{p}$ in the first term of the numerator we arrive at the Wilson confidence interval for θ of the form:

$$\frac{X+\frac{\kappa^2}{2} - \kappa\sqrt{n}\sqrt{\hat{p}(1-\hat{p})+\frac{\kappa^2}{4n}}}{n+\kappa^2} \leq \theta \leq \frac{X+\frac{\kappa^2}{2} + \kappa\sqrt{n}\sqrt{\hat{p}(1-\hat{p})+\frac{\kappa^2}{4n}}}{n+\kappa^2} \tag{2.2.6}$$

2.3 – The Agresti-Coull Interval

The Agresti-Coull interval is another alternative to the standard interval which has a similar form but a different choice for $\tilde{p}$. The center of the Wilson region is used instead of $\hat{p}$

(Agresti & Coull, 1998). Even though these intervals are centered about the same value, $\tilde{p}$, the actual coverage probability is more conservative than the Wilson interval. Due to its simple form, the Agresti-Coull interval is recommended for large values of n, n > 40. This interval may be preferred also for smaller sample sizes if the simplest form is desired (Brown et al., 2002).

To construct the Agresti-Coull confidence interval, let

$$\tilde{X} = X + \frac{\kappa^2}{2},\ \tilde{n} = n + \kappa^2,\ \tilde{p} = \frac{\tilde{X}}{\tilde{n}}$$

Using a construction similar to the standard interval we obtain the following:

$$\left| \frac{(\tilde{p} - \theta)}{\sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}}} \right| \leq \kappa \qquad (2.3.1)$$

Simplifying the last expression, we obtain the following equation

$$|\tilde{p} - \theta| \leq \kappa\sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}} \qquad (2.3.2)$$

Solving for θ, we obtain the Agresti-Coull confidence interval for θ of the form:

$$\tilde{p} + \kappa\sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}} \leq \theta \leq \tilde{p} - \kappa\sqrt{\frac{\tilde{p}\tilde{q}}{\tilde{n}}}. \qquad (2.3.3)$$

<u>2.4 – The Logit Interval</u>

The logit interval is formed by inverting the Wald-type interval for the log odds. This interval has a good coverage probability for values of p that are not close to 0 or 1. Denote

$$\lambda = g(p) = \log\left(\frac{p}{1-p}\right) \tag{2.4.1}$$

The maximum likelihood estimate, MLE of λ for 0 < X < n is

$$\hat{\lambda} = g(\hat{p}) = \log\left(\frac{\hat{p}}{1-\hat{p}}\right) = \log\left(\frac{\frac{X}{n}}{1-\frac{X}{n}}\right) = \log\left(\frac{X}{n-X}\right) \tag{2.4.2}$$

which is also known as the empirical logit transform. The variance of $\hat{\lambda}$ can be estimated by the delta method:

$$\sqrt{n}(\hat{p} - p) \sim N\big(0, p(1-p)\big) \tag{2.4.3}$$

For the functions $g(p)$, we can rewrite equation (2.4.3) as

$$\sqrt{n}\,\big(g(\hat{p}) - g(p)\big) \to N(0, p(1-p)(g'(p))^2). \tag{2.4.4}$$

We calculate the derivative of $g(p)$ as follows

$$g'(p) = [\log(p) - \log(1-p)]' = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}.$$

Substituting into the delta method equation 2.4.4 we arrive at

$$\sqrt{n}\,\big(\hat{\lambda} - \lambda\big) \sim N\left(0, p(1-p)\frac{1}{p^2(1-p)^2}\right) \tag{2.4.5}$$

which can be simplified to

$$\big(\hat{\lambda} - \lambda\big) \sim N\left(0, \frac{1}{np(1-p)}\right).$$

Replacing p with $\hat{p} = \frac{X}{n}$ we obtain

$$\big(\hat{\lambda} - \lambda\big) \sim N\left(0, \frac{1}{n\frac{X}{n}\left(1-\frac{X}{n}\right)}\right)$$

which we can simplified to

$$\left(\hat{\lambda} - \lambda\right) \sim N\left(0, \frac{n}{X(n-X)}\right). \tag{2.4.6}$$

The resulting variance is $\hat{V} = \frac{n}{X(n-X)}$. Therefore, the approximate 100 (1 – α)% confidence interval for λ is $\hat{\lambda} \pm \kappa \hat{V}^{1/2}$.

$$\log\left(\frac{X}{n-X}\right) - \kappa\sqrt{\frac{n}{X(n - X)}} \leq \lambda \leq \log\left(\frac{X}{n-X}\right) + \kappa\sqrt{\frac{n}{X(n - X)}} \tag{2.4.7}$$

We transform back to get the lower and upper limits of the confidence interval for p.

The lower limit is $\lambda_L = \log\left(\frac{p_L}{1 - p_L}\right)$.

Eliminate the logarithm by rewriting the last equation as $e^{\lambda_L} = \frac{p_L}{1 - p_L}$. Then to obtain the lower limit, p_L we solve the following equation by converting the fraction:

$e^{\lambda_L} - e^{\lambda_L} p_L = p_L$. We can solve the resulting equation for the lower limit, p_L, as

$$p_L = \frac{e^{\lambda_L}}{1 + e^{\lambda_L}}. \tag{2.4.8}$$

The upper limit is obtained in a similar fashion

$$p_U = \frac{e^{\lambda_U}}{1 + e^{\lambda_U}}. \tag{2.4.9}$$

The logit confidence interval for p is of the form, $p_L \leq p \leq p_U$. :

$$\frac{e^{\lambda_l}}{1 + e^{\lambda_l}} \leq p \leq \frac{e^{\lambda_u}}{1 + e^{\lambda_u}}$$

Substituting the lower and upper limits for λ from equation (2.4.7) into the equation above results in the logit confidence interval

$$\frac{e^{\log\left(\frac{X}{n-X}\right) - \kappa\sqrt{\frac{n}{X(n-X)}}}}{1 + e^{\log\left(\frac{X}{n-X}\right) - \kappa\sqrt{\frac{n}{X(n-X)}}}} \le p \le \frac{e^{\log\left(\frac{X}{n-X}\right) + \kappa\sqrt{\frac{n}{X(n-X)}}}}{1 + e^{\log\left(\frac{X}{n-X}\right) + \kappa\sqrt{\frac{n}{X(n-X)}}}} \qquad (2.4.10)$$

Brown, Cai and DasGupta have shown that expected length of the logit interval is larger than that of the Clopper-Pearson interval (2001).

2.5 – The Anscombe Logit Interval

Anscombe (1956) suggested an alternative value of $\hat{\lambda}$ which provides a better estimate of λ:

$$\hat{\lambda} = \log\left(\frac{X + {}^{1}/_{2}}{n - X + {}^{1}/_{2}}\right) \qquad (2.5.1)$$

Using higher order series expansion of the delta method (Gart & Zweifel, 1967), one can estimate the variance of $\hat{\lambda}$ by

$$\hat{V} = \frac{(n+1)(n+2)}{n(X+1)(n-X+1)} \,. \qquad (2.5.2)$$

The new Anscombe logit interval is overall shorter than the logit confidence interval (Brown et al., 2002) Rewriting equation (2.5.1) and using $\hat{p} = \frac{X}{n}$, we obtain

$$\hat{\lambda} = \log\left(\frac{\hat{p} + {}^{1}/_{2n}}{1 - \hat{p} + {}^{1}/_{2n}}\right) \qquad (2.5.3)$$

and the estimator of the variance is

$$\hat{V} = \frac{(n+1)(n+2)}{n^3\left(\hat{p} + {}^1/_n\right)\left(1 - \hat{p} + {}^1/_n\right)} \tag{2.5.4}$$

Using similar calculations as for the logit interval (Cox & Snell, 1989), we obtain the confidence intervals for $\hat{\lambda}$ of the form

$$\log\left(\frac{X + {}^1/_2}{n - X + {}^1/_2}\right) - \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}} \le \hat{\lambda}$$

$$\le \log\left(\frac{X + {}^1/_2}{n - X + {}^1/_2}\right) - \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}} \tag{2.5.5}$$

Then with similar transformation back to *p* we obtain the lower and upper limits for *p* as shown in equations (2.4.8) and (2.4.9).

$$\frac{e^{\lambda_L}}{1 - e^{\lambda_L}} \le p \le \frac{e^{\lambda_U}}{1 - e^{\lambda_U}} \tag{2.5.6}$$

where

$$\lambda_L = \log\left(\frac{X + {}^1/_2}{n - X + {}^1/_2}\right) - \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}}$$

and

$$\lambda_U = \log\left(\frac{X + {}^1/_2}{n - X + {}^1/_2}\right) + \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}}$$

Therefore, the Anscombe logit confidence interval for p is of the form:

$$\frac{e^{\log\left(\frac{X+{}^1/_2}{n-X+{}^1/_2}\right) - \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}}}}{1 - e^{\log\left(\frac{X+{}^1/_2}{n-X+{}^1/_2}\right) - \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}}}} \le p \le \frac{e^{\log\left(\frac{X+{}^1/_2}{n-X+{}^1/_2}\right) + \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}}}}{1 - e^{\log\left(\frac{X+{}^1/_2}{n-X+{}^1/_2}\right) + \kappa\sqrt{\frac{(n+1)(n+2)}{n(X+1)(n-X+1)}}}} \tag{2.5.7}$$

2.6 – The Likelihood Ratio Interval

The likelihood ratio interval is most commonly used when constructing confidence intervals. This is accomplished by inverting the likelihood ratio test which accepts the null hypothesis H_0: $p = p_0$ if $-2 \log \Lambda_n \leq \kappa^2$, where Λ_n is the likelihood ratio

$$\Lambda_\mathrm{n} = \frac{L(p_0)}{sup_p L(p)} \tag{2.6.1}$$

Here, L is the likelihood function. In particular,

Likelihood Function of $\hat{p}$: $L(\hat{p}) = \binom{n}{x} \hat{p}^X (1-\hat{p})^{n-X}$ (2.6.2)

Likelihood Function of p_0: $L(p_0) = \binom{n}{x} {p_0}^X (1-p_0)^{n-X}$ (2.6.3)

Likelihood Ratio: $\Lambda_\mathrm{n} = \frac{L(p_0)}{sup_{\hat{p}} L(\hat{p})} = \frac{{p_0}^X (1-p_0)^{n-X}}{\hat{p}^X (1-\hat{p})^{n-X}} = \frac{{p_0}^{n\hat{p}} (1-p_0)^{n(1-\hat{p})}}{\hat{p}^{n\hat{p}} (1-\hat{p})^{n(1-\hat{p})}}$ (2.6.4)

Testing hypothesis: H_0: $p = p_0$ versus H_1: $p \neq p_0$, we obtain the acceptance rule $-2 \log \Lambda_n \leq \kappa^2$ which can be rewritten as

$$-2\, [n\hat{p}\, (\log p - \log \hat{p}) + n(1-\hat{p})(\log(1-p) - n\ \log(1-\hat{p}))] \ \leq\ \kappa^2 \tag{2.6.5}$$

Simplifying the inequality in (2.6.5), we obtain the likelihood ratio confidence interval for p. The shortcoming of this interval is that there is no closed form solution. (Brown et al., 2001)

$$p^{\,\hat{p}}(1-p)^{(1-\hat{p})} \ \leq \mathrm{e}^{\frac{-\kappa^2}{2n}}\, \hat{p}^{\hat{p}} (1-\hat{p})^{(1-\hat{p})} \tag{2.6.6}$$

2.7 – The Clopper-Pearson Interval

The Clopper-Pearson interval is obtained by inverting the equal-tailed binomial test rather than the normal approximation (Clopper, 1934). If X = x is observed, then the Clopper-Pearson confidence interval is

$CI_{CP} = [\ L_{CP}(x), U_{CP}(x)\ ]$

where $L_{CP}(x)$ is the solution, in p, to the inequality $P_p(X \geq x) = \frac{\alpha}{2}$, which is the $\frac{\alpha}{2}$ quantile of the beta distribution Beta (x, n – x + 1). $U_{CP}(x)$ is the solution, in p, to $P_p(X \leq x) = \frac{\alpha}{2}$, which is the $\left(1 - \frac{\alpha}{2}\right)$ quantile of a beta distribution Beta (x + 1, n – x). The Clopper-Pearson interval guarantees that the actual coverage probability is always greater than or equal to the nominal confidence level. It has been shown that this interval is very conservative and, unless n is large, the actual coverage probability is larger than 1 – α (Brown et al., 2001). Due to this fact, the Clopper-Pearson interval is not recommended unless it is desired that the coverage probability always be larger than or equal to the nominal value.

2.8 – The Arcsine Interval

The arcsine interval is obtained by using a variance stabilizing transformation to determine a function $g(\hat{p})$ with a variance that is independent of the parameter of interest, *p*.

Then through transformation we can obtain a confidence interval for p (Hogg, 2005). The interval is derived by using the delta method. Let $\lambda = g(p) = \arcsin\left(p^{1/2}\right)$ and

$\hat{\lambda} = g(\hat{p}) = \arcsin\left(\hat{p}^{1/2}\right)$

Using the delta method to estimate the variance of $g(\hat{p})$ we need to calculate the derivative of $g(\hat{p})$

$$g'(\hat{p}) = \frac{1}{2\sqrt{\hat{p}(1-\hat{p})}} \quad (2.8.1)$$

Substituting this into the delta method equation 2.4.4 we arrive at

$$\sqrt{n}\,(\hat{\lambda} - \lambda) \sim N\left(0, \frac{\hat{p}(1-\hat{p})}{n} \cdot \left(\frac{1}{2\sqrt{\hat{p}(1-\hat{p})}}\right)^2\right) \quad (2.8.2)$$

The resulting variance can be simplified to attain $\hat{V} = \frac{1}{4n}$. The approximate 100(1 – α)% confidence interval for λ is $\hat{\lambda} \pm \kappa\hat{V}^{1/2}$

$$\arcsin\left(\hat{p}^{1/2}\right) - \frac{\kappa}{2n^{1/2}} \le \lambda \le \arcsin\left(\hat{p}^{1/2}\right) + \frac{\kappa}{2n^{1/2}} \quad (2.8.3)$$

We transform this interval back to obtain the upper and lower limits of the approximate 100(1 – α)% confidence interval for p.

$$\sin^2\left(\arcsin\left(\hat{p}^{1/2}\right) - \frac{\kappa}{2n^{1/2}}\right) \le p \le \sin^2\left(\arcsin\left(\hat{p}^{1/2}\right) + \frac{\kappa}{2n^{1/2}}\right) \quad (2.8.4)$$

Anscombe presented an alternative value for $\hat{p}$ that provides a better variance stabilization $\hat{p}$ is replaced with $\check{p} = \frac{X + 3/8}{n + 3/4}$

$$2n^{1/2}\left[\arcsin\left(\check{p}^{1/2}\right) - \arcsin\left(p^{1/2}\right)\right] \to N(0,1)\ as\ n \to \infty \quad (2.8.5)$$

The approximate 100 (1 – α)% confidence interval for p

$$sin^2\left(\arcsin\left(\check{p}^{1/2}\right) - \frac{1}{2}\kappa n^{-1/2}\right) \leq p \leq sin^2\left(\arcsin\left(\check{p}^{1/2}\right) + \frac{1}{2}\kappa n^{-1/2}\right) \qquad (2.8.6)$$

The interval performs well for p not close to 0 or 1. It has downward spikes near the boundaries (Brown et al., 2001).

CHAPTER THREE: BAYESIAN CREDIBLE SETS FOR THE SUCCESS PROBABILITY OF BINOMIAL DISTRIBUTIONS

3.1 – Bayesian Approach to Statistics

To understand Bayesian approach let us begin with conditional probabilities which are used to revise the probability space based on new information. The definition of conditional probability is given by Casella and Berger as follows,

If A and B are events in S, and P(B) > 0, then the conditional probability of A given B, written P(A|B), is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{3.1.1}$$

This can be rewritten as $P(A \cap B) = P(A|B)P(B)$. Using this definition we can arrive at $P(B|A) = \frac{P(A \cap B)}{P(A)}$. Rewriting this in a similar fashion gives us $P(A \cap B) = P(B|A)P(A)$. Substituting this into the numerator of the conditional probability definition gives

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)} \tag{3.1.2}$$

an equation commonly known as Bayes' Rule after Sir Thomas Bayes. Bayes' Rule has a more general form that applies to partitions of a sample space (Casella, 2002).

Let A_1, A_2, … be a partition of the sample space and let B be any set. Then, for each I = 1, 2, …

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P\big(B\big|A_j\big)P(A_j)} \tag{3.1.3}$$

When two continuous random variables are, conditional probabilities about one variable can be created given knowledge of the other random variable using joint distributions. Let (X,Y) be a continuous bivariate random vector with joint pdf $f(x,y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the conditional pdf of Y given that X = x is the function of y

$$f(y|x) = \frac{f(x,y)}{f_X(x)} \tag{3.1.4}$$

Similarly, for any y such that $f_Y(y) > 0$, the conditional pdf of X given that Y = y is the function of x

$$f(x|y) = \frac{f(x,y)}{f_Y(y)}$$

In the Bayesian approach, θ is considered a quantity whose variation can be described by a probability distribution known as the prior distribution. This is a subjective distribution based on the individual's beliefs about the sample before the data is observed. A sample is then taken from the population indexed by θ the prior distribution is then updated using Bayes' rule with this sample information to produce the posterior distribution.

Let $\pi(\theta)$ denote the prior distribution, $f(x|\theta)$ the sampling distribution, $m(x)$ the marginal distribution of X which is $m(x) = \int f(x|\theta)\pi(\theta)d\theta$ then the posterior distribution, $\pi(\theta|x)$ is

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}. \tag{3.1.5}$$

The posterior distribution which is the conditional distribution of θ given the sample x can be used to make inferences about random quantity θ (Berger, 2003).

These inferences are influenced by the information provided in the prior. Selection of a prior distribution to represent information which can be limited or nonexistent has been researched by many authors (Dyer & Pierce, 1993). The chosen prior distribution may satisfy the prior beliefs about the data but it also carries additional information which could be invalid. To avoid these concerns, many authors have investigated ways of creating noninformative priors that do not make assumptions about the data parameters. Jeffreys' noninformative prior is one of the most popular non-informative prior distribution since it is easy to derive and to justify (Berger, 2003).

The Jeffreys' prior is the positive square root of expected Fisher information, *I(θ)*, for *θ*.

$$p(\theta) \propto [I(\theta)]^{\frac{1}{2}}$$

$$\text{where:} \quad I(\theta) = -E^{X|\theta}\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right] \tag{3.1.6}$$

For the binomial case, we can arrive at the Jeffreys' prior by calculating the second derivative of the Fisher Information:

$$p(x|\theta) \propto \theta^x(1-\theta)^{n-x}$$

$$\log p(x|\theta) \propto x\log\theta + (n-x)\log(1-\theta)$$

$$\frac{\partial}{\partial\theta}\log p(x|\theta) \propto \frac{x}{\theta} - \frac{(n-x)}{1-\theta}$$

$$\frac{\partial^2}{\partial\theta^2}\log p(x|\theta) \propto -\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2} \tag{3.1.7}$$

Now entering 3.1.7 into Fisher information and using the expected value of X as *nθ* we obtain

$$I(\theta) = -E^{X|\theta}\left[-\frac{x}{\theta^2} - \frac{(n-x)}{(1-\theta)^2}\right] = \frac{n\theta}{\theta^2} + \frac{(n-n\theta)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)}$$

The resulting Jeffreys' prior is

$$p(\theta) \propto \left[\frac{n}{\theta(1-\theta)}\right]^{\frac{1}{2}}$$

If we disregard the $n^{\frac{1}{2}}$ term we have Jeffreys' prior as $\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$ which is a beta density with parameters $\frac{1}{2}$ and $\frac{1}{2}$.

3.2 – Bayesian Interval Estimation

An interval estimate of a real-valued parameter θ based on a random sample X = (x_1, …, x_n) is a pair of functions L(x_1, …, x_n) and U(x_1, …, x_n) such that L(x) ≤ U(x) for all x. The random interval [L(X), U(X)] is the interval estimator. The coverage probability of the interval estimator [L(X), U(X)] is the probability that the random interval covers the parameter θ. The probability is referred to as the confidence level. The interval estimator and the measure of confidence are known as confidence intervals or confidence sets.

In classical statistics, the confidence interval is said to "cover the parameter'' to stress the fact that the interval is a random quantity while parameter is a fixed. The (1 – α) confidence interval [L(X), U(X)] is one of the possible realized values of the random interval. Since the parameter is a fixed quantity there is a probability of 0 or 1 that it is within the realized interval. We can then say that there is a 100(1 – α)% chance of coverage that the realized random interval covers the true parameter (Casella, 2002).

Bayesian statistics differ from classical statistics in that the parameter is treated as a random variable with a probability distribution known as the prior distribution. Bayesian claims of coverage are made with respect the posterior distribution of the parameter (Berger, 2003).

This allows us to say that the parameter is within the confidence interval with some probability, not 0 or 1.

Bayesian statistics refers to the interval estimates as credible sets to avoid any confusion with classical statistics confidence sets which are markedly different probability assessments about the parameter. If $\pi(\theta|x)$ is the posterior distribution of θ given X = x, then for any set $A \subset \Theta$, the credible probability of A is

$$P(\theta \in A|x) = \int_A \pi(\theta|x)d\theta \tag{3.2.1}$$

and A is a credible set for θ. The Bayesian credible probability reflects the experimenter's subjective beliefs about the parameter, the prior distribution, and then updated with the data to create the posterior distribution. A Bayesian claim of 100(1 – α)% coverage means that after viewing and updating the prior distribution with the data they are 100(1 – α)% sure of coverage. While in classical statistics, a claim of 100(1 – α)% coverage means repeated identical trials 100(1 – α)% of the realized confidence sets will cover the true parameter.

There can be several kinds of credible sets used. The equal-tailed credible sets are formed by breaking α equally between the upper and lower bounds giving a 1 – α credible interval. $A = \{\theta : L_\alpha \le \theta \le U_\alpha\}$ where L_α is the α^{th} quantile of $\pi(\theta|x)$ and U_α is the $1 - \alpha^{th}$ quantile of $\pi(\theta|x)$. The highest posterior density set, HPD, can also be used to create the credible set. The goal is to create the shortest credible interval for θ that satisfies the following

$$\{\theta: \pi(\theta|x) \ge k\} \ \ where \ \int_{\{\theta:\, \pi(\theta|x) \ge k\}} \pi(\theta|x)\, d\theta = 1 - \alpha \tag{3.2.2}$$

This credible set is known as HPD region since it contains values of the parameter for which the posterior density is highest. An HPD interval uses lower and upper bounds that correspond to the highest parts of the posterior distribution that contain a 1 – α area between them. The shape of the HPD region is based on the shape of the posterior distribution. If the posterior distribution is symmetric then the HPD region formed will also be symmetric. In the text of Box and Tiao, they discuss the main properties of a HPD interval. One property is that that density of every point inside the interval is greater than that of every point outside the interval (1992). The second main property is that for a given probability, 1 – α, the interval is of the shortest length.

3.3 – The Jeffreys Interval

The Jeffreys prior interval is another alternative interval chosen by Brown, Cai and DasGupta (2001). Beta priors are commonly used to make inferences on p, because the family of beta distributions is the standard conjugate prior family for binomial distributions. In general, if $X \sim Bin\,(n,p)$ and p has a prior distribution of $Beta(\alpha_1, \alpha_2)$, with the density function,

$$g(p) = \frac{p^{\alpha_1 - 1}(1-p)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} \tag{3.3.1}$$

then the corresponding posterior distribution of p will also be a beta distribution

$$Beta\,(X + \alpha_1, n - X + \alpha_2) \tag{3.3.2}$$

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

$$\left[B\left({}^{\alpha}/_{2}\,; X + \alpha_1, n - X + \alpha_2\right), B\left(1 - {}^{\alpha}/_{2}\,; X + \alpha_1, n - X + \alpha_2\right)\right] \tag{3.3.3}$$

where $B\,(\alpha;\, m_1, m_2)$ denotes the α quantile of a Beta (m_1, m_2) distribution.

A non-informative prior is preferred when creating the credible interval as it does not influence the interval it treats all values of θ the same. Continuous non-informative prior are often improper (Hogg, 2005). The non-informative prior has an advantage because it remains invariant under transformation of the parameters.

The prior chosen was the non-informative Jeffreys' prior, Beta $\left(\frac{1}{2},\frac{1}{2}\right)$ which was derived in section 3.3. The corresponding density function (3.3.1) for the Jeffreys' prior

$$g(p) = \frac{p^{-1/2}(1-p)^{-1/2}}{B\left(\frac{1}{2},\ \frac{1}{2}\right)}$$

can be simplified using the fact that

Beta $\left({}^{1}/_{2}, {}^{1}/_{2}\right) = \frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)} = \pi$ to form the density function

$$g(p) = \pi^{-1}p^{-1/2}(1-p)^{-1/2}. \tag{3.3.4}$$

The posterior pdf of p is obtained using Bayes formula

$$P(p|X=x) = \frac{P(X=x|p)g(p)}{\int P(X=x|p)g(p)dp} \tag{3.3.5}$$

Substituting the binomial probability function and density function into the above gives

$$P(p|X) = \frac{\binom{n}{x}p^x(1-p)^{n-x} \cdot \frac{1}{\pi}p^{-1/2}(1-p)^{-1/2}}{\int \binom{n}{x}p^x(1-p)^{n-x} \cdot \frac{1}{\pi}p^{-1/2}(1-p)^{-1/2}\ dp} \tag{3.3.6}$$

After simplification we can arrive at

$$P(p|X) = \frac{p^{x-1/2}(1-p)^{n-x-1/2}}{\int p^{x-1/2}(1-p)^{n-x-1/2}\ dp} \tag{3.3.7}$$

Notice the denominator of equation 3.3.7 is a beta function. Therefore, the posterior pdf of *p* is of the form:

$$\frac{p^{x-1/2}(1-p)^{n-x-1/2}}{B\left(x+1/2,\ n-x+1/2\right)} \tag{3.3.8}$$

The lower and upper limits, $[L_J(x), U_J(x)]$, of the $100(1-\alpha)\%$ equal-tailed Jeffreys prior interval are created by selecting the center of the interval to have $(1-\alpha)$ of the area with the two sides having $\alpha/2$ of the posterior probability.

An adjustment is made to the lower limit, $L_J(0) = 0$, and a similar adjustment to the upper limit, $U_J(n) = 1$, to avoid the intervals poor behavior at the boundaries.

$$L_J(x) = B\left(\alpha/2; X+1/2, n-X+1/2\right) \tag{3.3.9}$$

and at x = 0 the lower limit is $L_J(0) = 0$

The upper limit of the $100(1-\alpha)\%$ equal-tailed Jeffreys prior interval is given as

$$U_J(x) = B\left(1-\alpha/2; X+1/2, n-X+1/2\right) \tag{3.3.10}$$

and at x = n the upper limit is $U_J(n) = 1$

The actual endpoints of the Jeffreys interval have to be numerically computed. Brown, Cai and DasGupta(2001) provide a table listing the limits for the Jeffreys interval for the values $7 \leq n \leq 30$. The coverage of the Jeffreys interval is similar to the Wilson interval. It however still has two steep downward spikes near 0 and 1.

For this reason Brown, Cai and DasGupta (2001) suggested a modified Jeffreys interval to eliminate the downward spikes near 0 and 1, caused by $U_j(0)$ being too small and $L_J(n)$ being too large, new limits were chosen for the Jeffreys confidence interval.

$U_{M\text{-}J}(0) = p_l$ and $L_{M\text{-}J}(n) = 1 - p_l$

where p_l satisfies $(1 - p_l)^n = {}^{\alpha}/_{2}$ which can be rewritten as $p_l = 1 - \left({}^{\alpha}/_{2}\right)^{1/n}$

$L_{M\text{-}J}(1) = 0$, otherwise $L_{M\text{-}J} = L_J$

$U_{M\text{-}J}(n - 1) = 1$, otherwise $U_{M\text{-}J} = U_J$

CHAPTER FOUR: CONFIDENCE INTERVALS IN THE CASE OF HYPERGEOMETRIC DISTRIBUTION

4.1 – The Hypergeometric Distribution

The hypergeometric distribution is the discrete distribution which can best be described using the urn scheme (Casella, 2002). Suppose the urn is filled with N amount of identical marbles of two different colors for example k are white and $N - k$ are yellow. We randomly select n marbles from the urn (an example of sampling without replacement). We want to know what is the probability that there is x amount of white marbles in the sample of n marbles. The hypergeometric probability mass function is as follows

$$P(X = x \,|N, k, n) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad (4.1.1)$$

where N represents the total size of the population

k represents the number of white marbles in the population

$N - k$ represents the number of yellow marbles in the population

n represents the sample size

x represents the number of white marbles in the sample.

In most cases, the known values are N, x and n. The unknown quantity of interest is the proportion of white marbles in the population, given as $p = \frac{k}{N}$.

The hypergeometric distribution is also widely used in reliability theory where N can be considered as the size of the population of items and k is the unknown number of defective items in the population. When a sample of size n is drawn, the objective is to make inferences on the number of defectives items k.

The mean of the hypergeometric distribution is

$$EX = \frac{nk}{N} = np. \tag{4.1.2}$$

The variance is

$$Var\ X = \frac{nk(N-k)(N-n)}{N \cdot N(N-1)} = np(1-p)\frac{N-n}{N-1}. \tag{4.1.3}$$

4.2 – Approximation of Hypergeometric Distribution by Binomial Distribution

If the population size *N* is large and *x* and *n* are very small in comparison with *N*, then the hypergeometric distribution can be approximated by the binomial distribution. To demonstrate this, let us begin by expanding the binomial coefficients of the hypergeometric distribution (4.1.1).

$$\binom{k}{x} = \frac{k!}{(k-x)!\,x!} = \frac{k(k-1)\ldots(k-x+1)(k-x)!}{(k-x)!\,x!} = \frac{k(k-1)\ldots(k-x+1)}{x!} \tag{4.2.1}$$

$$\binom{N-k}{n-x} = \frac{(N-k)!}{\big(N-k-(n-x)\big)!\,(n-x)!}$$

$$= \frac{(N-k)(N-k-1)\ldots(N-k-n+x+1)(N-k-n+x)!}{(N-k-n+x)!\,(n-x)!}$$

$$= \frac{(N-k)(N-k-1)\ldots(N-k-n+x+1)}{(n-x)!} \qquad (4.2.2)$$

$$\binom{N}{n} = \frac{N!}{(N-n)!\,n!} = \frac{N(N-1)\ldots(N-n+1)(N-n)!}{(N-n)!\,n!} = \frac{N(N-1)\ldots(N-n+1)}{n!} \qquad (4.2.3)$$

Let us rewrite the last term of the numerator in (4.2.3) as follows

$$N-n+1 = N-x+x-n+1 = N-x-(n-x-1)$$

After substituting (4.2.1),(4.2.2) and (4.2.3) into the hypergeometric distribution, (4.1.1) yields the following

$$P(X = x\,|N,k,n) = \frac{\frac{k(k-1)\ldots(k-x+1)}{x!}\,\frac{(N-k)(N-k-1)\ldots(N-k-n+x+1)}{(n-x)!}}{\frac{N(N-1)\ldots(N-n+1)}{n!}}$$

$$= \frac{n!}{x!\,(n-x)!}\,\frac{k(k-1)\ldots(k-(x-1))(N-k)(N-k-1)\ldots(N-k-(n-x-1))}{N(N-1)\ldots(N-x+1)(N-x)(N-x-1)\ldots(N-x-(n-x-1)}$$

$$= \binom{n}{x}\prod_{j=0}^{x-1}\frac{k-j}{N-j}\;\prod_{j=0}^{n-x-1}\frac{N-k-j}{N-x-j} \qquad (4.2.4)$$

If $x \ll N$ *and* $n \ll N$ then $n-x \ll N$. Also, when $j = x - 1$ or $j = n - x - 1$ the value of j can be approximated to zero since x and n are very small.

Manipulate the fractions in the equation above by dividing each term by N and then using the approximation $j/N \rightarrow 0$:

$$\frac{k-j}{N-j} = \frac{\frac{k}{N} - \frac{j}{N}}{\frac{N}{N} - \frac{j}{N}} = \frac{\frac{k}{N}}{1} = p \quad \text{for } 0 \leq j \leq x-1 \tag{4.2.5}$$

$$\frac{N-k-j}{N-x-j} = \frac{\frac{N}{N} - \frac{k}{N} - \frac{j}{N}}{\frac{N}{N} - \frac{x}{N} - \frac{j}{N}} = \frac{1-\frac{k}{N}}{1} = 1-p \quad \text{for } 0 \leq j \leq x-1 \tag{4.2.6}$$

Replacing these two simplified fractions, 4.2.5 and 4.2.6 into the equation 4.2.4 gives the following equation which can be rewritten using exponential notation to represent the binomial distribution

$$P(X = x \,|N,k,n) = \binom{n}{x} \prod_{j=0}^{x-1} \frac{k-j}{N-j} \prod_{j=0}^{n-x-1} \frac{N-k-j}{N-x-j}$$

$$= \binom{n}{x} \prod_{j=0}^{x-1} p \prod_{j=0}^{n-x-1} 1-p$$

$$= \binom{n}{x} p^x \, (1-p)^{n-x} \tag{4.2.7}$$

If the experimenter knows that the values of x and n are relatively small, $x \ll N \;\; and \;\; n \ll N,$ then the binomial distribution can be used as an estimate for the hypergeometric distribution.

4.3 – Confidence Interval Based on Normal Approximation

Using standard normal approximation to create a confidence interval based on normal approximation for the hypergeometric distribution we have $x \sim \mathcal{N}(EX, Var\ X)$ the mean and variance are given in 4.1.2 and 4.1.3

$$x \sim \mathcal{N}\left(np, np(1-p)\frac{N-n}{N-1}\right) \tag{4.3.1}$$

To create a two-sided (1 – α) confidence interval, we estimate by $\hat{p} = \frac{X}{n}$, so that

$$\frac{x - np}{\sqrt{\frac{n\hat{p}(1-\hat{p})(N-n)}{N-1}}} \approx z \quad \sim N(0,1) \tag{4.3.2}$$

The interval is obtained by solving the inequality for *p*:

$$z\alpha_{/2} \leq \frac{x - np}{\sqrt{\frac{n\hat{p}(1-\hat{p})(N-n)}{N-1}}} \leq z_{1-}\alpha_{/2}$$

Since $z\alpha_{/2} = \ -z_{1-}\alpha_{/2}$ we can rewrite the last inequality using the absolute value

$$\left|\frac{x - np}{\sqrt{\frac{n\hat{p}(1-\hat{p})(N-n)}{N-1}}}\right| \leq z\alpha_{/2} \tag{4.3.3}$$

Using the estimate, $\hat{p} = \frac{x}{n}$ in the radical, we can rewrite the expression as

$$\left|\frac{x - np}{\sqrt{\frac{n\left(\frac{x}{n}\right)\left(1-\frac{x}{n}\right)(N-n)}{N-1}}}\right| \leq z\alpha_{/2} \tag{4.3.4}$$

The last expression can then be reduced and modified to the following

$$\left|\frac{x-np}{\sqrt{\frac{x(n-x)(N-n)}{n(N-1)}}}\right| \leq z\alpha/_2 \quad (4.3.5)$$

The confidence interval using standard normal approximation is

$$\frac{x-\ z\alpha/_2\sqrt{\frac{x(n-x)(N-n)}{n(N-1)}}}{n} \leq p\ \leq \frac{x+\ z\alpha/_2\sqrt{\frac{x(n-x)(N-n)}{n(N-1)}}}{n} \quad (4.3.6)$$

4.4 – The Confidence Interval based on Analog of Wilson Interval

The Wilson interval is an alternate interval which uses the exact standard error $\sigma = (pq)^{1/2}\, n^{-1/2}$ instead of the estimated standard error $\hat{\sigma} = (\hat{p}\hat{q})^{1/2}\, n^{-1/2}$. For the hypergeometric distribution, the inequality to solve is of the following form

$$\left|\frac{x-np}{\sqrt{\frac{np(1-p)(N-n)}{N-1}}}\right| \leq\ z_{1-\alpha/_2}. \quad (4.4.1)$$

Clearing the denominator and squaring both sides to clear the radical gives the following

$$x^2-\ 2nxp+n^2p^2 \leq\ z^2{}_{1-\alpha/_2}\left(\frac{np(1-p)(N-n)}{N-1}\right) \quad (4.4.2)$$

Distributing the terms on the right hand side of the inequality, we obtain

$$x^2-\ 2nxp+n^2p^2 \leq \frac{z^2{}_{1-\alpha/_2}}{N-1}(nNp-\ n^2p-nNp^2+\ n^2p^2) \quad (4.4.3)$$

Rewriting the inequality (4.4.3) by grouping the terms

$$p^2\left[\left(\frac{z^2{}_{1-\alpha/2}}{N-1}\right)(n^2 - nN) - n^2\right] + p\left[\left(\frac{z^2{}_{1-\alpha/2}}{N-1}\right)(nN - n^2) + 2nx\right] - x^2 \leq 0 \quad (4.4.4)$$

To solve for *p*, we can use the quadratic formula. The terms to be used in the quadratic formula would then be

A = $\frac{z^2{}_{1-\alpha/2}}{N-1}(n^2 - nN) - n^2$, B = $\frac{z^2{}_{1-\alpha/2}}{N-1}(nN - n^2) + 2nx$, and C = $-x^2$

$$p = \frac{-B \pm \sqrt{B^2 - 4A(-x^2)}}{2A}$$

$$= \frac{-B \pm \sqrt{B^2 + 4Ax^2}}{2A}$$

$$Let \quad p_1 = \frac{-B - \sqrt{B^2 + 4Ax^2}}{2A} \quad \text{and} \quad p_2 = \frac{-B + \sqrt{B^2 + 4Ax^2}}{2A} \quad (4.4.5)$$

Then $p_1 \leq p \leq p_2$ is the $(1 - \alpha)$ confidence interval for *p* based on the analog of the Wilson interval:

$$\frac{-B - \sqrt{B^2 + 4Ax^2}}{2A} \leq p \leq \frac{-B + \sqrt{B^2 + 4Ax^2}}{2A} \quad (4.4.6)$$

4.5 – Bayesian Estimation for Hypergeometric Distribution

Dyer and Pierce examined prior distributions for hypergeometric sampling (1993). They looked at four different families of prior distributions developed by a Bayesian approach. We shall look at their recommended prior distribution, Polya (beta-binomial) when no prior information is available.

Dyer and Pierce (1993) described the hypergeometric sampling as taking a sample of size n drawn without replacement from a population of size *N*. Let *k* be the number of failures and *N* – *k* be the number of successes. The sampling distribution of *X*, the number of failures in the sample is the hypergeometric distribution with the pdf

$$f(x|k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad \max\{0, n-(N-k)\} \leq x \leq \min\{k, n\} \qquad (4.5.1)$$

The prior distribution is of the form *g(k;ω)* where ω is a hyperparameter. The prior provides a priori information which can be limited or not available about the true value of *k* which is combined with the sample data to produce the posterior distribution for *k*

$$\rho(k|x;\omega) = \frac{\binom{k}{x}\binom{N-k}{n-x} g(k;\omega)}{\sum_{k=x}^{N-n+x}\binom{k}{x}\binom{N-k}{n-x} g(k;\omega)}, \quad k = x, x+1, \dots, N-n+x \qquad (4.5.2)$$

where ω is the scalar vector parameter of the prior. The posterior is used to make inferences about the unknown quantity, *k*. A concern of statisticians about the selection of a prior distribution is the extra information that may be contained in the prior distribution (Dyer, 1993). As a method of avoiding this issue, many authors have developed non-informative priors. The confidence interval for k is of the form $k_1 \leq k \leq k_2$, where k_1 and k_2 are such that

$$\sum_{k=k_1}^{k_2} \rho(k|x;\omega) \geq 1 - \alpha$$

4.5.1 – Bayesian Estimation with Binomial Prior

One of the possibilities is to impose the binomial prior distribution on k. If a population of size *N* is drawn at random, where each item has an unknown probability *θ* of being defective. For given *θ*, the sampling distribution for *k*, the number of defectives in the sample of size *N*, is binomial

$$f(k|\theta) = \binom{N}{k}\theta^k(1-\theta)^{N-k}, \qquad k = 0, 1, \ldots, N \;\; and \;\; \omega = \theta \tag{4.5.1.1}$$

The unconditional marginal distribution of *k* is then given as follows

$$f(k) = \int_0^1 f(k|\theta)\;\; g(k;\theta)\, d\theta$$

$$= \int_0^1 \binom{N}{k}\theta^k(1-\theta)^{N-k}\, g(k;\theta)\, d\theta \tag{4.5.1.2}$$

The prior distribution for θ can be used to obtain a marginal distribution which can be a prior distribution for *k*.

In this case, the marginal distribution of *x* (for a given value of *θ*) takes the form

$$p(x|\theta) = \binom{N}{n}^{-1} \sum_{k=x}^{N-n+x} \binom{k}{x}\binom{N-k}{n-x}\binom{N}{k}\theta^k(1-\theta)^{N-k} \tag{4.5.1.3}$$

Observe that the binomial coefficients can be expanded and simplified by multiplying the numerator and denominator by $n!(N-n)!$

$$\binom{k}{x}\binom{N-k}{n-x}\binom{N}{k} = \frac{k!\,(N-k)!\,N!}{x!\,(k-x)!\,(n-x)!\,(N-k-n+x)!\;(N-k)!\,k!}$$

$$= \frac{N!\ n!\ (N-n)!}{n!\ (N-n)!\quad x!\,(n-x)!\,(N-n-(k-x))!\,(k-x)!}$$

$$= \binom{N}{n}\binom{n}{x}\binom{N-n}{k-x} \qquad (4.5.1.4)$$

Therefore, using k – x = j, j = 0, 1, … , N – n, we derive the following marginal distribution

$$p(x|\theta) = \binom{N}{n}^{-1} \sum_{k=x}^{N-n+x} \binom{N}{n}\binom{n}{x}\binom{N-n}{k-x}\theta^k(1-\theta)^{N-k}$$

$$= \binom{n}{x} \sum_{k=x}^{N-n+x} \binom{N-n}{k-x}\theta^k(1-\theta)^{N-k}$$

$$= \binom{n}{x} \sum_{j=0}^{N-n} \binom{N-n}{j}\theta^{j+x}(1-\theta)^{N-j-x}$$

$$= \binom{n}{x}\theta^x(1-\theta)^{n-x} \sum_{j=0}^{N-n} \binom{N-n}{j}\theta^{j}(1-\theta)^{N-n-j}$$

$$= \binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad (4.5.1.5)$$

since $\sum_{j=0}^{N-n} \binom{N-n}{j}\theta^{j}(1-\theta)^{N-n-j} = 1$

Therefore, in this case, the marginal distribution of *x* given θ is Binomial (n, θ). Then the posterior distribution of k given x is of the form

$$\rho(k|x;\theta) = \frac{\binom{k}{x}\binom{N-k}{n-x}\binom{N}{k}\theta^k(1-\theta)^{N-k}}{\binom{N}{n}\binom{n}{x}\theta^x(1-\theta)^{n-x}}$$

Using the previous observation about the binomial coefficients in equation 4.5.1.4 the numerator can be rewritten and we arrive at

$$= \frac{\binom{N}{n}\binom{n}{x}\binom{N-n}{k-x}\theta^{k}(1-\theta)^{N-k}}{\binom{N}{n}\binom{n}{x}\theta^{x}(1-\theta)^{n-x}}$$

$$= \binom{N-n}{k-x}\theta^{k-x}(1-\theta)^{N-n-(k+x)} \qquad (4.5.1.6)$$

So the posterior distribution of k has a binomial distribution given *x* and θ.

We can note that since *x* given θ has a binomial distribution, one can use all previously considered methodology for derivation of the confidence interval for $\theta, (\theta_1, \theta_2)$. Then, the confidence interval for k can be drawn as $(N\theta_1, N\theta_2)$ if *N* is large or by using posterior distribution $\rho(k|x;\theta)$ for moderate values of *N*.

4.5.2 – Bayesian Estimation with Polya (Beta-Binomial) Prior

Suppose that *θ* varies across the process according to a prior distribution for *θ* represented by $g(\theta;\omega)$ where *ω* is a hyperparameter. The unconditional distribution of K is given by

$$f(k) = \int_0^1 f(k|\theta)\, g(\theta;\omega)\, d\theta = \int_0^1 \binom{N}{k}\theta^{k}(1-\theta)^{N-k}\, g(\theta;\omega)\, d\theta \qquad (4.5.2.1)$$

We suggest the use of the Polya (beta-binomial) distribution as a prior. This distribution results from hierarchical model where *k* given *θ* is assumed to have a binomial distribution.

$$g(k|\theta, N) = \binom{N}{k}\theta^{k}(1-\theta)^{N-k}, \qquad k = 0, 1, \ldots, N \qquad (4.5.2.2)$$

and θ has the beta distribution with parameters $\omega = (\alpha, \beta)$. The Polya distribution uses the beta distribution which is commonly used with binomial sampling to create a closed form marginal distribution for *k*. The pdf for the beta distribution is

$$\pi(\theta; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, \qquad 0 < \theta < 1 \tag{4.5.2.3}$$

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ with $\alpha > 0$ and $\beta > 0$. Therefore, the marginal distribution of *k* is obtained by integrating out θ from the joint distribution of *k* and θ:

$$g(k|N, \alpha, \beta) = \int_0^1 g(k|\theta, N)\ \pi(\theta; \alpha, \beta)\, d\theta$$

$$= \int_0^1 \binom{N}{k} \theta^k (1-\theta)^{N-k} \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta$$

$$= \frac{\binom{N}{k}}{B(\alpha, \beta)} \int_0^1 \theta^{k+\alpha-1}(1-\theta)^{N+\beta-k-1}\ d\theta \tag{4.5.2.4}$$

Simplifying this expression by rewriting the integral as the beta function $B(k+\alpha, N+\beta-k)$ we obtain

$$g(k|N, \alpha, \beta) = \frac{\binom{N}{k} B(k+\alpha, N+\beta-k)}{B(\alpha, \beta)}$$

Rewriting the beta functions via gamma functions we can simplify the fraction:

$$g(k|N, \alpha, \beta) = \binom{N}{k} \frac{\frac{\Gamma(k+\alpha)\Gamma(N+\beta-k)}{\Gamma(N+\alpha+\beta)}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$$

$$= \binom{N}{k} \frac{\Gamma(\alpha+\beta)\Gamma(k+\alpha)\Gamma(N+\beta-k)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(N+\alpha+\beta)}, \qquad k = 0, 1, \ldots, N \tag{4.5.2.5}$$

This is the Polya distribution with parameters $(N,\ \alpha,\ \beta)$, which can be used as a prior distribution for hypergeometric sampling (Dyer 1993).

Then, by integration we obtain a marginal distribution for *x*, the number of defectives in the sample of size *n*. The resulting pdf is

$$m(x) = \int_0^1 f(x|\theta)\ \pi(\theta; \alpha, \beta)\ d\theta$$

$$= \int_0^1 \binom{n}{x} \theta^x (1-\theta)^{n-x} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}\, d\theta$$

$$= \frac{\binom{n}{x}}{B(\alpha,\beta)} \int_0^1 \theta^{x+\alpha-1} (1-\theta)^{n+\beta-x-1}\ d\theta$$

$$= \binom{n}{x} \frac{\Gamma(\alpha+\beta)\Gamma(x+\alpha)\Gamma(n+\beta-x)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta)}, \qquad x = 0, 1, \ldots, n \qquad (4.5.2.6)$$

The posterior distribution for *k* is

$$g(k|x; N, n, \alpha, \beta) = \frac{\binom{N}{k} \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)\,\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)\,\Gamma(\alpha+\beta+N)}}{\binom{n}{x} \frac{\Gamma(\alpha+x)\,\Gamma(\beta+n-x)\,\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)\,\Gamma(\alpha+\beta+n)}}$$

After simplifying the fraction and combining the binomial coefficients we arrive at the posterior

$$g(k|x; N, n, \alpha, \beta) = \binom{N-n}{k-x} \frac{\Gamma(\alpha+k)\Gamma(N+\beta-x)\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)\Gamma(N+\alpha+\beta)},$$

$$k = x, x+1, \ldots, N-n+x \qquad (4.5.2.7)$$

Note that the posterior distribution for *k* – *x* is also the Polya distribution with parameters (N – n, α + x, n + β – x). The Polya distribution is then the conjugate prior distribution for hypergeometric sampling.

Then, the posterior on $\gamma = \frac{x}{k}$ is found by replacing k with $\frac{x}{\gamma}$

$$g(\gamma|x;\ N,n,\alpha,\beta) = \binom{N-n}{\frac{x}{\gamma}-x}\frac{\Gamma\left(\alpha+\frac{x}{\gamma}\right)\Gamma(N+\beta-x)\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)\Gamma(N+\alpha+\beta)}$$

$$= \binom{N-n}{x\left(\frac{1}{\gamma}-1\right)}\frac{\Gamma\left(\alpha+\frac{x}{\gamma}\right)\Gamma(N+\beta-x)\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)\Gamma(N+\alpha+\beta)} \qquad (4.5.2.8)$$

CHAPTER FIVE: CONCLUSIONS

In the present thesis, we considered construction of confidence intervals for the binomial and the hypergeometric distributions. Confidence intervals for the binomial proportion, p, has coverage issues when p is near the boundaries 0 or 1 due to the discreteness of the binomial data. Although, to the best of our knowledge, no one studied systematically confidence intervals in the case of the hypergeometric distribution, similar issues will arise. In this situation, if the population size is large, one can either reduce the case to the binomial confidence intervals or credible sets. Otherwise, methods described above can be applied.

LIST OF REFERENCES

Agresti, A. & Coull, B.A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician,* 52, 119 – 126.

Anscombe, F.J. (1948). The transformation of Poisson, binomial and negative binomial data. *Biometrika*, 35, 246 – 254.

Anscombe, F. J. (1956). On estimation binomial response relations. *Biometrika,* 43, 461 – 464.

Berger, J. O. (2003). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York, NY: Springer.

Blyth, C.R. & Still, H.A. (1983). Binomial confidence intervals. *Journal of the American Statistical Association,* 78, 108 – 116.

Box, G.E.P. & Tiao, G.C. (1992). *Bayesian Inference in Statistical Analysis*, New York: Wiley.

Brown, L.D., Cai, T. & DasGupta, A. (2002). Interval Estimation for a Binomial Proportion, *Statistical Science,* 16, 101 – 133.

Brown, L.D., Cai, T. & DasGupta, A. (2001). Confidence Intervals for a Binomial Proportion and Asymptotic Expansions. *The Annals of Statistics*, 30(1), 160 – 201.

Casella, G. & Berger, R.L. (2002). *Statistical Inference* (2nd ed.). Pacific Grove, CA: The Wadsworth Group.

Clopper, C.J. & Pearson E.S. (1934), Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26(4), 404 – 413.

Cox, D.R. & Snell, E.J. (1989). *Analysis of Binary Data* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Dyer, D. & Pierce R. (1993). On the Choice of the Prior Distribution in Hypergeometric Sampling. *Communications in Statistics – Theory and Methods*, 22(8), 2125 – 2146.

Freeman, M. & Tukey, J.W. (1950). Transformations Related to the Angular and the Square Root. *The Annals of Mathematical Statistics*, 21(4), 607 – 611.

Gart, J.J. & Zweifel, J.R. (1967). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika*, 54, 181 – 187.

Hogg, R., McKean, J. & Craig A. (2005). *Introduction to Mathematical Statistics* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Newcombe, R.G. (2001). Logit Confidence Intervals and the Inverse Sinh Transformation. *The American Statistician*, 55(3), 200 – 202.

Wright, T. (1991). *Exact Confidence Bounds when Sampling from Small Finite Universes*. Berlin: Springer-Verlag.