8-2018

Wald Confidence Intervals for a Single Poisson Parameter and Binomial Misclassification Parameter When the Data is Subject to Misclassification

Nishantha Janith Chandrasena Poddiwala Hewage
Stephen F Austin State University, janith.ms@gmail.com

WALD CONFIDENCE INTERVALS FOR A SINGLE POISSON PARAMETER AND BINOMIAL MISCLASSIFICATION PARAMETER WHEN THE DATA IS SUBJECT TO MISCLASSIFICATION

by

NISHANTHA JANITH CHANDRASENA PODDIWALA HEWAGE, 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

AUGUST 2018

WALD CONFIDENCE INTERVALS FOR A SINGLE POISSON PARAMETER AND BINOMIAL MISCLASSIFICATION PARAMETER WHEN THE DATA IS SUBJECT TO MISCLASSIFICATION

by

NISHANTHA JANITH CHANDRASENA PODDIWALA HEWAGE, B.S.

APPROVED:

Kent Riggs, Ph.D., Thesis Director

Robert K. Henderson, Ph.D., Committee Member

Jacob Turner, Ph.D., Committee Member

Garland Simmons, Ph.D., Committee Member

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

ABSTRACT

This thesis is based on a Poisson model that uses both error-free data and error-prone data subject to misclassification in the form of false-negative and false-positive counts. We present maximum likelihood estimators (MLEs), Fisher's Information, and Wald statistics for the Poisson rate parameter and the two misclassification parameters. Next, we invert the Wald statistics to get asymptotic confidence intervals for the Poison rate parameter and false−negative rate parameter. The coverage and width properties for various sample size and parameter configurations are studied via a simulation study. Finally, we apply the MLEs and confidence intervals to one real data set and another realistic data set.

.

ACKNOWLEDGEMENTS

Apart from my efforts, the success of this research project depends mostly on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people, who have been instrumental in the successful completion of this research project.

First and foremost, I owe a debt of gratitude to my supervisor Dr. Kent Riggs. Over the last six months, I have benefited from his inspiring ideas, encouragement and suggestions with this thesis project.

Also, my special thanks goes to thesis committee members Dr. Robert Henderson, Dr. Jacob Turner, and Dr. Garland Simmons. Additionally I want to thank Dr. Lesa Beverly, Dr. William D. Clark, and academic staff members of the Department of Mathematics and Statistics, for their valuable advice, encouragement and guidance during my time at graduate school.

Finally, I am incredibly grateful to my family members, particularly my father, mother and brothers and graduate students for their constant love, moral encouragement and kindness. Furthermore, there is a lengthy list of names of people and institutions who have contributed to my academic success which I am unable to present here due to limited space, but, I am grateful to all of them for their assistance, in bringing this work to completion.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Mathematical statistics is the foundation of the statistics and is the starting place for every statistical model. There are two major components in statistics called descriptive statistics and inferential statistics. Descriptive statistics help summarize, organize, and display a data set. Statistical Inference is used in drawing conclusions about parameters of the population using sample data and some inferential technique. Confidence intervals for parameters is a common inferential technique.

In day to day life, decisions to either do something or not are common. Such binary data are frequently used in a wide range of applications, including survey analysis, criminology, clinical medicine, and information technology. We also see binary or binomial data in epidemiology, which is the study of the distribution of health-related states and events in populations. Bross(1950) [6], Tenenbein(1970) [15], Hochberg(1977) [11], Chen(1979) [8], Viana, Ramakrishnan and Levy (1993) [3], Joseph, Gyorkos and Coupal(1995) [2], and York et al.(1995) [1] published research papers concerning parameter estimation for count data with misclassification for both the binomial and multinomial models.

Count data with a Poisson distribution can also be seen in the areas of epidemiology, market research, and criminal justice. Researchers often propose Poisson models to compare the rates of certain events for different populations. For instance, models to compare mortality rates of different diseases, or to compare crime rates for different neighborhoods. Counts which are used to estimate the Poisson rate of interest could be subject to error on the inferences and rates due to misclassification. Anderson, Bratcher, and Kutron(1994) [7] published an article concerning the Poisson model with misclassification, estimating a Poisson rate allowing for false negatives. Bayesian estimators for the rate of occurrence and probability of false negatives were given by Andersonetal. Also Bayesian methods to estimate Poisson rates in the presence of both false−negative and false−positive misclassification were used by Bratcher and Stamey (2002) [5]. Stamey and Young (2005) [12] published a paper discussing a Poisson model that uses error free data and error-prone data subject to misclassification in the form of false-negative and false-positive counts. In both research papers, a multi-sample procedure was used. In this procedure, a first sample is called training sample or expensive sample that is usually not very large. The second (inexpensive) sample is larger and cheaper than the training sample.

In this thesis, we consider a method to find the confidence interval for a single Poisson parameter and Binomial misclassification parameter when data is subject to misclassification, based on common maximum- likelihood-based asymptotic statistics called Wald statistics. Wald statistics were presented by Abraham Wald in the mid 20^{th} century. We develop a statistical model that uses a double-sampling procedure on binary data that is subject to misclassification. A Monte Carlo simulation is used to investigate the coverage and width properties which related interesting parameters.

The thesis is organized in the following manner. In section 1.1, we discuss some motivating examples and some terminology. Chapter 2 discusses the mathematical theory which is related to the study. Next, in chapter 3, we develop the statistical model for that uses error−prone Poisson data and error−free Poisson, Binomial data to estimate relevant parameter of interest. In chapter 4, we present Wald−based confidence intervals for the Poisson rate and false−negative rate, and study their coverage and width properties via a simulation study. The real wold examples introduced in section 1.1 will be evaluated using the misclassification model in chapter 5. Comments and conclusions are deliberated in chapter 6.

1.1 Motivating Examples

1.1.1 Example One

Mortality data and statements of cause of death on the death certificates are used as major sources of information characterizing the health of population groups. Hence, accuracy on death certificates is very important. Age, sex, date of death, residence, and cause of death are routinely recorded death certificates. Problems with the reliability of the statement of cause of death appear to be related to the current state of medical knowledge, and incomplete information at the time of death. Therefore, there may be major and minor asperities between the medical section of the death record and other sources of clinical and pathological information relating to the death of the patient.

Autopsy is the one of several methods used to supplement or revise the cause of death on death certificates. There are two type of autopsies. One is forensic autopsy and other is clinical autopsy, the latter involves a highly specialized surgical procedure which is an expensive examination of the corpse to identify the cause of death. Data from forensic autopsy can result in misclassification of the cause of death. Therefore, the misclassification model discussed in this thesis can be applied when considering such autopsy results.

This example data taken from Stamey and Young (2005) research paper [12] is based on Kircher, Nelson and Burdo (1985) [16], and includes data based on the death certificates of all people who died in Connecticut in 1980. In that year $28,440$ deaths occurred in Connecticut. Forensic autopsies were performed on 3884 of the decedents. From the sample of 280 autopsy report randomly drawn for the present study, 272 cases were checked by clinical autopsy.

The primary objective of this epidemiology study is to obtain a valid and precise estimate of the expected occurrences of deaths due to digestive disease in the population of the Connecticut. Here incidence refers to the occurrence of new deaths in a Connecticut in 1980. The population of Connecticut in 1980 was 3.108 million. All of these autopsies were performed under the jurisdiction of the state. From the 272 clinical autopsies performed, 32 deaths were attributed to digestive disease. For this same sample, 18 deaths had been attributed to digestive disease by forensic autopsy, of the 18, 16 were correct, 2 were false−positives. 16 forensic autopsies identified as due to some other cause, but death was actually due to digestive disorder (false−negatives). Of all the forensic autopsies conducted in Connecticut in (1980), 6.1% identified the cause of death as digestive disorder.

1.1.2 Example Two

Defects, internal and external corrosion, dents, and gouges are regularly found in oil and gas pipelines. In the majority of cases, defects are minor and have no impact on safety of the pipeline. In some cases, however, defects may can be significant and a repair is necessary. Identifying defects that are critical, and need to be repaired is important. Hence, any method of identifying and assessing critical defects must be accurate and not conservative.

In this example, there are two defect identification approaches, manual inspection and ultrasonic method. Doing inspection manually is more expensive with labour cost and equipment cost versus the ultrasonic method. ABC Technical Service.Inc is interested in the average number of defects per 1000 feet of oil pipelines near Houston, Texas. Using an ultrasonic method, $30,000$ feet were inspected and 145 defects were found.

For $2,000$ feet, ultrasonic and manual inspection methods were used, and 25 defects were identified from the ultrasonic method, but only 23 true defects were identified by the manual method. Of the 25 defects found by ultrasonic method, 16 were classified as false$-$positive and 4 were correct, and there was 14 false$-$negative. The model proposed and evaluated in this study can be applied in each of the examples above.

2 MATHEMATICAL PRELIMINARIES

2.1 Misclassification Models

In day−to−day life, we make classifications that may either be correct or wrong. If the classification is incorrect, that is called a misclassification. For binary data, there are two misclassifications, false−negative and false−positive. As an example, in medical diagnosis, a patient can be classified as not having a specific disease when in fact the patient does have the disease. This is called a false-negative. A false−positive, occurs when the patient is diagnosed to have the disease, when, in really, he does not have the disease. Misclassification can occur because of human error, experimental error, etc. While misclassification may be a common mistake, the consequences can often be severe.

Statistical models may be helpful in accounting for the potentially misclassified data. Such models can allow for appropriate estimates of parameters of interest, as well as misclassification parameters. Binomial, Multinomial, Poisson, negative binomial distributions are just a few of the probability models that may be used. In general, the strategy of developing the model is different from application to application.

In this research, we consider a Poisson model with misclassification, and a double-sampling procedure involving a binomial model that is used to appropriately estimate a Poisson rate parameter and misclassification parameters. From the model we will also develop and study relevant confidence intervals based on the Wald statistics.

2.2 Double Sampling Procedure

Suppose we have two classifying devices, one is perfect (or infallible) while the other device is prone to misclassification (or fallible). Typically, the infallible classifier is expensive in terms of time or money, and the fallible classifier is less expensive. These two devices make up a double sampling procedure. In order to create the first sample, the fallible classifier is applied to a large sample. Note that this ”fallible” sample may included false−positive and false−negative results, but they are not observable. For the second sample, the fallible and infallible classifier are applied to a small sample. In this second sample, false−positive and false−negative counts are observable because of the infallible classifier.

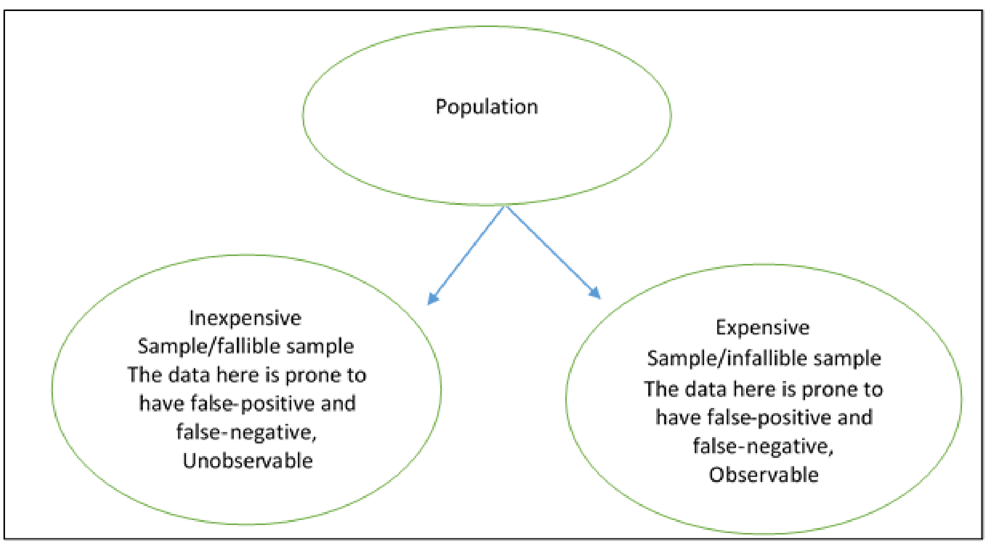


Figure 2.1: Double Sample Procedure

2.3 Binomial Distribution

Swiss mathematician Jakob Bernoulli, is credited with the beginning of the binomial distribution[12]. A Bernoulli experiment is a random experiment when the outcome can be classified as only success or failure (e.g., female or male, life or death, no defective or defective). The binomial distribution models the number of successes from independent Bernoulli trials where the probability of success is constant. The binomial distribution is used to obtain the probability of observing x successes in n trials, with the probability of success on a single trial denoted by p. The formula for the binomial probability mass function is

$$P(x; p, n) = \binom{n}{x} p^x (1-p)^{(n-x)},$$

where

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

and

$$x! = x(x-1)(x-2)\ldots 3\ 2\ 1, \quad 0! = 1, \text{ and } x = 0, 1, 2, \ldots, n$$

2.4 Poisson Distribution

The Poisson distribution is a discrete probability distribution which has many application in statistics, and was discovered by French mathematician Simon Denis Poisson [12]. He derived the distribution using the limit of the binomial distribution. It is used to model the number of occurrences of particular event occurring within a given interval of time, space, or area.

Let X denote the number of occurrences in a in a given interval and assume X follows an approximate Poisson process (described below) with parameter $\lambda > 0$.

The formula for the Poisson probability mass function is

$$p(x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad , x = 0, 1, 2, \ldots$$

.

The assumptions used to support use of the Poisson distribution are,

- Events occur at random in continuous space or time.
- Events occur singly, and the probability of two events occurring simultaneously is zero.
- Events occur uniformly, i.e. the expected number of events in a given interval is proportional to the size of the interval.
- Events occur independently, i.e. the probability of an event occurring in any small interval is independent of the probability of the event occurring in any other small interval.
- The variable is the number of events that occur in an interval of a given magnitude.

However, in practical situations the above assumption might not be valid. For example, blood cells, or yeast cells, or any other organism held in suspension in a liquid, will not remain uniformly spaced through the liquid, unless the liquid is shaken up in its container before each drop is taken to place on a microscope slide. This lack of uniform occurrence can result in a change of mean. Physical arguments like these are always very valuable in deciding whether a Poisson model is appropriate.

In addition, the Poisson distribution has the special property of the mean and variance being equal. Consequently, if we divide the sample variance by the sample mean, we should obtain a number which is very close to one when data really do come from a Poisson distribution, suggesting that true Poisson model is appropriate.

2.5 Likelihood Background

In this section, we discuss maximum likelihood estimation, Fisher's information, and Wald statistics by looking at the appropriate likelihood function. R.A Fisher introduced this most significant development of statistics in the 20^{th} century. He spent ten years, from 1912 to 1922, when he published several papers related to the likelihood function. After these publications, a whole branch of statistical reasoning had been established. Hogg, Mckean, and Craig (2005) [4] provide a nice overview of the likelihood function, which is summarized here.

2.5.1 Likelihood Function

Consider a random vector $X = (X_1, X_2, \ldots, X_n)'$ of dimension n where the X's are jointly distributed with probability density $f(\ X \mid \theta\ )$, where $\theta = (\theta_1, \theta_2, \ldots, \theta_m)'$ is a parameter vector of dimension m contained in the parameter space $\Omega \subset R^m$.

The likelihood function of θ views $f(X \mid \theta)$ as a function of θ given X and is defined as

$$L(\theta) \equiv L(\theta \mid X) = f(X \mid \theta).$$

2.5.2 Maximum Likelihood Estimators

The maximum likelihood estimator(MLE) of θ , if it exists, is defined to be the value, $\hat{\theta} = \hat{\theta}\ (X)$ such that

$$\hat{\theta} = Argmax\{L(\theta)\}.$$

An important measure in large sample likelihood theory is the score function,

$$u(\theta) \equiv \frac{\partial}{\partial \theta} L(\theta),$$

where $\frac{\partial}{\partial\theta}L(\theta) = (\frac{\partial}{\partial\theta_1}L(\theta), \frac{\partial}{\partial\theta_2}L(\theta), \ldots, \frac{\partial}{\partial\theta_m}L(\theta))'$. The MLE is often the solution to the equation $u(\theta) = 0$.

2.5.3 Fisher's Information

The Fisher's information is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ of a distribution that models X.

To obtain the Fisher's information, certain regularity condition are needed. These conditions are given in detail in Hogg, Mckean, Craig " Introduction To Mathematical Statistics" [4].

When the conditions hold, we can compute the $(i,j)^{th}$ element of the $m \times m$ Fisher's information matrix as

$$I_{i,j}(\theta) = \quad -E\left[\frac{\partial \ln(L(\theta))}{\partial\theta_i\ \partial\theta_j{}'}\right] \quad ,i,j = 1,\ldots,m$$

.

2.5.4 Properties of MLE's

The large sample properties on the Likelihood function and Maximum Likelihood Estimators include,

- MLE's become unbiased minimum variance estimators as the sample size increases.
- MLE's have approximately normal distributions and approximate sample variances that can be calculated and used to generate confidence bounds for large sample sizes.

- Likelihood functions can be used to test hypotheses about models and parameters.

2.5.5 Wald Statistics and Confidence Intervals

The Wald statistic is a first order asymptotic statistic composed from a function of the Fisher's information matrix.

Suppose $\theta = (\theta_1, \theta_2, \theta_3)'$, where θ_1,θ_2 and θ_3 are scalars. θ_1 is the parameter of interest and θ_2 and θ_3 are the nuisance parameters. Then the Wald statistic for θ_i is

$$W_i \equiv \frac{\hat{\theta}_i - \theta_i}{I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)},$$

for $i = 1, 2, 3$.

Where, $\hat{\theta}_i$ are MLE's of θ_i and $I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ is the corresponding (i, i) diagonal element of $I[\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3]^{-1}$. W is asymptotically $N(0, 1)$ and W^2 is asymptotically χ^2_1.

In order to develop a confidence interval for parameters θ_i, consider the following probability statements,

$$P\left(-Z_{\alpha/2} \leq W \leq Z_{\alpha/2}\right) \approx 1 - \alpha,$$

$$P\left(-Z_{\alpha/2} \leq \frac{\hat{\theta}_i - \theta_i}{I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)} \leq Z_{\alpha/2}\right) \approx 1 - \alpha,$$

$$P\left(-Z_{\alpha/2}\sqrt{I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)} \leq \hat{\theta}_i - \theta_i \leq Z_{\alpha/2}\sqrt{I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)}\right) \approx 1 - \alpha.$$

Hence, a $(1 - \alpha)100\%$ confidence interval for θ_i can be expressed as

$$\hat{\theta}_i + Z_{\alpha/2}\sqrt{I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)} \geq \theta_i \geq \hat{\theta}_i - Z_{\alpha/2}\sqrt{I^{ii}(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)},$$

where $Z_{\alpha/2}$ is the corresponding standard normal percentile.

3 THE MISCLASSIFICATION MODEL

In this mathematical model, we utilize the Poisson and binomial distributions with a double sampling procedure to account for misclassification. We have two samples, an infallible or training sample (expensive sample) and a fallible sample (inexpensive sample).

Infallible sample: using the infallible and fallible classifier, we take a small sample of size A_0. Note: A_0 may not be a positive integer, but might be any positive real number reflecting an area, distance, or time frame of general interest. We able to observe the number of false negatives (X_0), the number of false positive (Y_0), the number of true occurrence (T_0), and the error−prone (as labeled by the fallible classifier) count of occurrences, Z_0. Note that $Z_0 = T_0 + Y_0 - X_0$. Here, we assume $T_0 \sim Pn(A_0\lambda)$, where λ is the expected number of true occurrence over one unit of interest. Also, the maximum numbers of false negatives is T_0, and we assume $(X_0|T_0 = t_0) \sim Bi(t_0, \theta)$ where θ is the rate of false negative results for the fallible classifier. Given the Poisson nature of occurrences, the false positive random variable has no upper bound. Therefore, we model the true count and the false positive count as independent Poisson random variables. We assume $Y_0 \sim Pn(A_0\phi)$, where ϕ is the the expected number of false−positives over one unit of interest. Hence $Z_0 \approx Pn(A_0\mu)$, where $\mu = \lambda(1-\theta) + \phi$, the expected number of occurrences when using the fallible classifier over one unit of interest. As a summary, λ is the occurrence rate parameter that is the main parameter of interest, ϕ is the occurrence rate of false positive observations, and θ is the probability of a single false negative observation.

Fallible sample: using only the fallible classifier, we obtain a large sample of size A. Note: A also may not be a positive integer, but insted might be positive real number reflecting an area, distance, or time frame of general interest. In this data set, we have error−prone data that possibly contains false−positives (Y), false−negatives (X) counts, and true occurrences (T). However, we cannot observe such counts. Rather, we can only observe Z, where $Z = T + Y - X$. Assuming a similar distribution structure to the infallible sample, we have that $Z \sim Pn(A\mu)$. In this study, we ultimately will investigate the confidence intervals for λ and θ, based on a Wald statistic.

3.1 Joint Probability Mass Function of Misclassification Model

With the assumption and using double sampling as described, we need the probability mass functions of the following random variables to derive the joint probability mass function. In summary:

- $T_0 \sim Pn(A_0\lambda)$
- $Y_0 \sim Pn(A_0\phi)$
- $(X_0|T_0 = t_0) \sim Bi(t_0, \theta)$
- $Z \sim Pn(A\mu)$

where,

$$
\begin{aligned}
A_0 &= \text{number of sample units within the time or space interval in the traning sample,} \\
A &= \text{number of sample units within the time or space interval in the fallible sample,} \\
T_0 &= \text{number of true incidences in the sample } A_0, \\
Y_0 &= \text{number of false−positives in the sample } A_0,
\end{aligned}
$$

$$
\begin{aligned}
X_0 &= \text{number of false}-\text{negatives in the sample } A_0, \\
Z_0 &= \text{number of occurrences identified by the fallible classifier in the sample } A_0, \\
Z &= \text{number of occurrences in observed the sample } A \text{ (using only the fallible classifier)}, \\
\lambda &= \text{expected number of true occurrence over one unit of time or space}, \\
\theta &= \text{probability of be a single false}-\text{negative observation, and} \\
\phi &= \text{expected number of false}-\text{positives over one unit of time or space.}
\end{aligned}
$$

As mentioned in the beginning of the chapter, the number of true occurrences (incidences) (T_0) and the number of false$-$positives (Y_0) are assumed to be independent. However, true occurrences (T_0) and false$-$negatives (X_0) are dependent, and a binomial model is appropriate when the effective training sample (infallible sample) size is small. The Poisson model for T_0 actually assumes a sufficiently large search area. Consequently,$A_0 - t_0$ (where $T_0 = t_0$ scaled to the appropriate unit same as A_0) is also large enough to follow the Poisson, rather than the Binomial distribution model. Again, assuming ϕ is small, T_0 and Y_0 are approximately independent. Hence, we have known probability mass functions (pmf),

$$
\begin{aligned}
p_{T_0}(t_0) &= \begin{cases} \frac{(A_0\lambda)^{t_0} e^{-A_0\lambda}}{t_0!}, & \text{for } t_0 \geq 0, \end{cases} \\
p_{Y_0}(y_0) &= \begin{cases} \frac{(A_0\phi)^{y_0} e^{-A_0\phi}}{y_0!}, & \text{for } y_0 \geq 0, \text{ and} \end{cases} \\
p_{X_0|T_0}(x_0|t_0) &= \begin{cases} \binom{t_0}{x_0}\theta^{x_0}(1-\theta)^{(t_0-x_0)}, & \text{for } 0 \leq x_0 \leq t_0. \end{cases}
\end{aligned}
$$

The joint pmf of the X_0 and T_0 can now be determined from the conditional distribution of X_0 and T_0:

$$p_{X_0|T_0}(x_0|t_0) = \frac{p_{X_0Y_0}(x_0,t_0)}{p_{T_0}(t_0)}$$

$$p_{X_0T_0}(x_0,t_0) = p_{T_0}(t_0) \times p_{X_0|T_0}(x_0|t_0)$$

$$p_{X_0T_0}(x_0,t_0) = \frac{(A_0\lambda)^{t_0}e^{-A_0\lambda}}{t_0!}\binom{t_0}{x_0}\theta^{x_0}(1-\theta)^{(t_0-x_0)}$$

$$p_{X_0T_0}(x_0,t_0) = \frac{A_0^{t_0}}{x_0!(t_0-x_0)!}\left(\lambda^{t_0}e^{-A_0\lambda}\theta^{x_0}(1-\theta)^{(t_0-x_0)}\right)$$

where $0 \leq x_0 \leq t_0$, $t_0 \geq 0$, $\lambda > 0$, $0 \leq \theta \leq 1$.

The joint pmf of X_0, T_0 and Y_0:

$$\begin{aligned}
p_{X_0Y_0T_0}(x_0,y_0,t_0) &= p_{X_0,T_0}(x_0,t_0) \times p_{Y_0}(y_0) \\
&= \frac{A_0^{t_0}}{x_0!(t_0-x_0)!}\left(\lambda^{t_0}e^{-A_0\lambda}\theta^{x_0}(1-\theta)^{(t_0-x_0)}\right)\frac{(A_0\phi)^{y_0}e^{-A_0\phi}}{y_0!} \\
&= \frac{A_0^{t_0+y_0}}{x_0!y_0!(t_0-x_0)!}\left(\lambda^{t_0}e^{-A_0\lambda}\theta^{x_0}(1-\theta)^{(t_0-x_0)}\phi^{y_0}e^{-A_0\phi}\right),
\end{aligned}$$

where $t_0, y_0 \geq 0$, $0 \leq x_0 \leq t_0$ $\lambda, \phi > 0$, $0 \leq \theta \leq 1$.

Now we need to find the distribution of the Z_0(number of occurrence), which can be derived using transforming technique related to $Z_0 = T_0 + Y_0 - X_0$. T_0 follow the Poisson distribution with parameter λ A_0, Y_0 also follow the Poisson distribution with parameter ϕ A_0, and X_0 is depend on T_0 and follows conditional binomial distribution with sample size t_0 and probability of successes θ. In here T_0 and Y_0 are independent random variable and also X_0 and Y_0 are independent random variables. Since only

T_0 and X_0 are dependent, that is only necessary to find the distribution of $T_0 - X_0$, using

$$
\begin{aligned}
p_{X_0T_0}(x_0, t_0) &= p_{T_0}(t_0) \times p_{X_0|T_0}(x_0|t_0) \\
p_{X_0T_0}(x_0, t_0) &= \frac{(A_0\lambda)^{t_0} e^{-A_0\lambda}}{t_0!} \binom{t_0}{x_0} \theta^{x_0}(1-\theta)^{(t_0-x_0)}
\end{aligned}
$$

where $t_0 \geq 0$, $0 \leq x_0 \leq t_0$, $\lambda > 0$, $0 \leq \theta \leq 1$.

Consider the random variables $Z_1 = T_0 - X_0$ and $Z_2 = T_0$, then

$$
\begin{aligned}
X_0 &= Z_2 - Z_1, \\
T_0 &= Z_2, \\
\text{where} \quad Z_2 \geq 0, \quad 0 &\leq Z_1 \leq Z_2.
\end{aligned}
$$

Then,we have:

$$
p_{Z_1Z_2}(z_1, z_2) = \frac{(A_0\lambda)^{z_2} e^{-A_0\lambda}}{z_2!} \binom{z_2}{z_2 - z_1} \theta^{z_2-z_1}(1-\theta)^{z_1}
$$

Thus, the probability mass function of Z_1 is

$$
\begin{aligned}
p_{Z_1}(z_1) &= \sum_{z_2=z_1}^{\infty} P_{Z_1Z_2}(z_1, z_2) \\
&= \sum_{z_2}^{\infty} \frac{(A_0\lambda)^{z_2} e^{-A_0\lambda}}{z_2!} \binom{z_2}{z_2 - z_1} \theta^{z_2-z_1}(1-\theta)^{z_1}
\end{aligned}
$$

$$
\begin{aligned}
&= (1-\theta)^{z_1} \sum_{z_2=z_1}^{\infty} \frac{(A_0\lambda)^{z_2} e^{-A_0\lambda}}{z_2!} \frac{z_2!}{z_1!(z_2-z_1)!} \theta^{z_2-z_1} \\
&= \frac{(\lambda A_0)^{z_1}(1-\theta)^{z_1}}{z_1!} \sum_{z_2=z_1}^{\infty} \frac{1}{(z_2-z_1)!} e^{-A_0\lambda} (\theta A_0\lambda)^{z_2-z_1} \\
&= \frac{[\lambda A_0(1-\theta)]^{z_1}}{z_1!} e^{-A_0\lambda} \sum_{z_2=z_1}^{\infty} \frac{(\theta A_0\lambda)^{z_2-z_1}}{(z_2-z_1)!}.
\end{aligned}
$$

Let consider $y = z_2 - z_1$. Hence

$$
\begin{aligned}
p_{Z_1}(z_1) &= \frac{[\lambda A_0(1-\theta)]^{z_1}}{z_1!} e^{-A_0\lambda} \sum_{y=0}^{\infty} \frac{(\theta A_0\lambda)^{y}}{y!} \\
&= \frac{[\lambda A_0(1-\theta)]^{z_1}}{z_1!} e^{-A_0\lambda}\ e^{\theta A_0\lambda} \\
&= \frac{[\lambda A_0(1-\theta)]^{z_1}}{z_1!} e^{-A_0\lambda(1-\theta)} \qquad z_1 \geq 0
\end{aligned}
$$

Thus, Z_1 follows the Poisson distribution with parameter $A_0\lambda(1-\theta)$.

We need to find the distribution of Z_0 that is equal to summation of Z_1 and Y_0. But, Z_1 and Y_0 are independent Poisson random variables, hence, Z_0 has a Poisson distribution with parameter $A_0\lambda(1-\theta) + A_0\theta$ [4]. So,

$$
p_{Z_0}(z_0) = \frac{[A_0(\lambda(1-\theta)+\phi)]^{z_0}}{z_0!} e^{-A_0(\lambda(1-\theta)+\phi)}, \quad z_0 \geq 0.
$$

According to the assumption the distribution on counts in fallible sample are similar to the corresponding infallible sample counts. Therefore,

$$
p_{Z}(z) = \frac{[A(\lambda(1-\theta)+\phi)]^{z}}{z!} e^{-A(\lambda(1-\theta)+\phi)} \quad z \geq 0.
$$

3.2 Likelihood Function of Misclassification Model

Using the pmf in section 3.1, the likelihood function for the proposed misclassification model is:

$$L(\lambda, \phi, \theta) \;=\; p_{T_0}(t_0)p_{Y_0}(y_0)p_{X_0T_0}(x_0, t_0)p_Z(z)$$

$$L(\lambda, \phi, \theta) \;=\; K\left(\lambda^{t_0}e^{-\lambda A_0}\phi^{y_0}e^{-\phi A_0}\theta^{x_0}(1-\theta)^{(t_0-x_0)}(\lambda(1-\theta)+\phi)^z e^{-A\lambda(1-\theta)-A\phi}\right),$$

where, K is a constant involving combinatorial terms that are not a function of the parameters.

The log−likelihood, $l(\lambda, \phi, \theta)$=ln $(L(\lambda, \phi, \theta))$, is a one−to−one and order preserving transformation of $L(\lambda, \phi, \theta)$, and hence, MLE's can be formed using this transformation. The log−likelihood is

$$\begin{aligned}
l(\lambda, \phi, \theta) &= \ln\left(K(\lambda^{t_0}\exp^{-\lambda A_0}\phi^{y_0}\exp^{-\phi A_0}\theta^{x_0}(1-\theta)^{(t_0-x_0)}(\lambda(1-\theta)+\phi)^z\exp^{-A\lambda(1-\theta)-A\phi})\right)\\
&= \ln K + t_0\ln\lambda - \lambda A_0\ln e + y_0\ln\phi - \phi A_0\ln e + x_0\ln\theta\\
&\quad +(t_0-x_0)\ln(1-\theta) + z\ln(\lambda(1-\theta)+\phi) - (A\lambda(1-\theta)+A\theta)\ln e\\
&= \ln K + t_0\ln\lambda - \lambda A_0 + y_0\ln\phi - \phi A_0 + x_0\ln\theta + (t_0-x_0)\ln(1-\theta)\\
&\quad +z\ln(\lambda(1-\theta)+\phi) - (A\lambda(1-\theta)+A\theta).
\end{aligned}$$

3.3 Maximum Likelihood Estimator

To find the estimators, we first derive the partial derivatives of $l(\lambda, \phi, \theta)$ and set them equal to zero, generating three equations whose solution yields MLE's for λ, ϕ, and θ. The partial derivatives with respective to λ, ϕ, θ are:

$$\frac{\partial l}{\partial \lambda} = \frac{t_0}{\lambda} - A_0 + \frac{z(1-\theta)}{(\lambda(1-\theta)+\phi)} - A(1-\theta),$$

$$\frac{\partial l}{\partial \phi} = \frac{y_0}{\phi} - A_0 + \frac{z}{(\lambda(1-\theta)+\phi)} - A, \text{ and}$$

$$\frac{\partial l}{\partial \theta} = \frac{x_0}{\theta} - \frac{(t_0 - x_0)}{(1-\theta)} - z\frac{\lambda}{(\lambda(1-\theta)+\phi)} + A\lambda.$$

Setting $\frac{\partial l}{\partial \lambda} = 0$, $\frac{\partial l}{\partial \phi} = 0$, $\frac{\partial l}{\partial \theta} = 0$, we get the following estimating equations:

$$t_0(\lambda(1-\theta)+\phi) - \lambda(\lambda(1-\theta)+\phi)A_0 + z(1-\theta)\lambda - A(1-\theta)\lambda(\lambda(1-\theta)+\phi) = 0 \quad (3.1)$$

$$y_0(\lambda(1-\theta)+\phi) - \phi(\lambda(1-\theta)+\phi)A_0 + z\phi - A\phi(\lambda(1-\theta)+\phi) = 0 \quad (3.2)$$

$$x_0(1-\theta)(\lambda(1-\theta)+\phi) - \theta(\lambda(1-\theta)+\phi)(t_0-x_0) - z\lambda\theta(1-\theta) + A\lambda\theta(1-\theta)(\lambda(1-\theta)+\phi) = 0 \quad (3.3)$$

After solving above equation 3.1, 3.2 and 3.3, the maximum likelihood estimators for λ,ϕ, and θ are:

$$\hat{\lambda} = \alpha_1 \frac{(t_0 + z(1 - \frac{y_0}{z_0}))}{A_0} + \alpha_2 \left(\frac{x_0}{A_0} \right), \quad (3.4)$$

$$\hat{\phi} = \frac{y_0(z + z_0)}{z_0(A + A_0)}, \quad (3.5)$$

$$\hat{\theta} = \frac{x_0(A_0 + A)}{A_0(z + t_0 + \frac{x_0}{A_0}A - \frac{y_0}{z_0}z)}, \quad (3.6)$$

where $z_0 = t_0 + y_0 - x_0$, $\alpha_1 = \frac{A_0}{(A+A_0)}$ and $\alpha_1 + \alpha_2 = 1$.

3.4 Interpretation of MLE's

There is an intuitive interpretation for $\hat{\lambda}$, $\hat{\phi}$, and $\hat{\theta}$. For that:

- Recall the quantity z_0 is the number of occurrences observed by the fallible classifier in the training (infallible) sample.
- $\frac{y_0}{z_0}$ is the proportion of false positives in the observed training sample data.
- The estimated rate of false negatives is $\frac{x_0}{A_0}$.

Then we can re−express the estimated parameter:

$$\hat{\lambda} = \frac{z + t_0 + \frac{x_0}{A_0}A - \frac{y_0}{z_0}z}{A_0 + A}.$$

Since λ is the expected number of true occurrences over one unit, and following the double sampling procedure, the numerator of above equation should be the number of true occurrences from the two samples. t_0 is the true occurrences from training sample but the only observable count in fallible sample is z, hence z should be corrected by adding the expected number of false−negatives in fallible sample ($\frac{x_0}{A_0}A$) and subtracting the expected number of false−positives in the fallible sample ($\frac{y_0}{z_0}z$). The denominator is the total sample size $A_0 + A$.

Note that $\hat{\phi}$ can be re−expressed as:

$$\hat{\phi} = \frac{\frac{y_0}{z_0}z + y_0}{(A + A_0)}.$$

For the false−positive parameter, the expected number of false−positives($\frac{y_0}{z_0}Z$) from the error−prone sample should be added with y_0 from training sample to get the numerator. This numerator is then averaged over the total sample size $A_0 + A$.

Note that $\hat{\theta}$ can be re−expressed as:

$$\hat{\theta} = \frac{x_0 + \frac{x_0}{A_0}A}{\text{estimated number of true occurrences from two sample}}.$$

Calculate the "total" false−negatives by adding false−negative from training sample and expected number of false−negatives from fallible sample together. Then it is averaged over the number of true occurrences from the two samples to get an estimate for the probability of a single false−negative observation.

Hence, the estimated number of true occurrences from the two samples is $(z+t_0+\frac{x_0}{A_0}A - \frac{y_0}{z_0}z)$.

3.5 Fishers Information Matrix

To derive Fisher's information matrix for our misclassification model, first we need the second partial derivatives of $l(\lambda, \phi, \theta)$ with respect to λ, ϕ, and θ:

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{t_0}{\lambda^2} - \frac{z(1-\theta)^2}{(\lambda(1-\theta)+\phi)^2}, \tag{3.7}$$

$$\frac{\partial^2 l}{\partial \lambda \partial \phi} = -\frac{z(1-\theta)^2}{(\lambda(1-\theta)+\phi)^2}, \tag{3.8}$$

$$\frac{\partial^2 l}{\partial \lambda \partial \theta} = A + \frac{(z(1-\theta)\lambda)}{((1-\theta)\lambda+\phi)^2} - \frac{z}{((1-\theta)\lambda+\phi)}, \tag{3.9}$$

$$\frac{\partial^2 l}{\partial \phi^2} = -\frac{z}{(1-\theta\lambda+\phi)^2} - \frac{y_0}{\phi^2}, \tag{3.10}$$

$$\frac{\partial^2 l}{\partial \phi \partial \theta} = \frac{z\lambda}{((1-\theta)\lambda+\phi)^2}, \tag{3.11}$$

and

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{z\lambda^2}{(1-\theta)\lambda+\phi)^2} - \frac{t_0 - x_0}{(1-\theta)^2} - \frac{x_0}{\theta^2}. \tag{3.12}$$

Next, we compute the expected values of (3.7) − (3.12) to get $I(\lambda, \phi, \theta)$:

$$-E\begin{bmatrix} \frac{\partial^2 l}{\partial \lambda^2} & \frac{\partial^2 l}{\partial \lambda \partial \phi} & \frac{\partial^2 l}{\partial \lambda \partial \theta} \\ \frac{\partial^2 l}{\partial \phi \partial \lambda} & \frac{\partial^2 l}{\partial \phi^2} & \frac{\partial^2 l}{\partial \phi \partial \theta} \\ \frac{\partial^2 l}{\partial \theta \partial \lambda} & \frac{\partial^2 l}{\partial \theta \phi} & \frac{\partial^2 l}{\partial \theta^2} \end{bmatrix} = \begin{bmatrix} -E(\frac{\partial^2 l}{\partial \lambda^2}) & -E(\frac{\partial^2 l}{\partial \lambda \partial \phi}) & -E(\frac{\partial^2 l}{\partial \lambda \partial \theta}) \\ -E(\frac{\partial^2 l}{\partial \phi \partial \lambda}) & -E(\frac{\partial^2 l}{\partial \phi^2}) & -E(\frac{\partial^2 l}{\partial \phi \partial \theta}) \\ -E(\frac{\partial^2 l}{\partial \theta \partial \lambda}) & -E(\frac{\partial^2 l}{\partial \theta \phi}) & -E(\frac{\partial^2 l}{\partial \theta^2}) \end{bmatrix}$$

$$= \begin{bmatrix} I_{\lambda\lambda} & I_{\lambda\phi} & I_{\lambda\theta} \\ I_{\phi\lambda} & I_{\phi\phi} & I_{\phi\theta} \\ I_{\theta\lambda} & I_{\theta\phi} & I_{\theta\theta} \end{bmatrix}$$

The element of Fisher's Information Matrix are:

$$I_{\lambda\lambda} = -E\left(-\frac{t_0}{\lambda^2} - \frac{z(1-\theta)^2}{(\lambda(1-\theta)+\phi)^2}\right) = \frac{A_0\mu + A\lambda(1-\theta)^2}{\lambda\mu},$$

$$I_{\lambda\phi} = -E\left(-\frac{z(1-\theta)^2}{(\lambda(1-\theta)+\phi)^2}\right) = \frac{A(1-\theta)}{\mu},$$

$$I_{\lambda\phi} = -E\left(-\frac{z(1-\theta)^2}{(\lambda(1-\theta)+\phi)^2}\right) = \frac{A(1-\theta)\lambda}{\mu},$$

$$I_{\phi\lambda} = -E\left(-\frac{z(1-\theta)^2}{(\lambda(1-\theta)+\phi)^2}\right) = \frac{A(1-\theta)}{\mu},$$

$$I_{\phi\phi} = -E\left(-\frac{z}{(1-\theta\lambda+\phi)^2} - \frac{y_0}{\phi^2}\right) = \frac{A_0\mu + A\phi}{\phi\mu},$$

$$I_{\phi\theta} = -E\left(\frac{z\lambda}{((1-\theta)\lambda+\phi)^2}\right) = -\frac{A\lambda}{\mu},$$

$$I_{\phi\theta} = -E\left(\frac{z\lambda}{((1-\theta)\lambda+\phi)^2}\right) = -\frac{A\lambda}{\mu},$$

$$I_{\theta\lambda} = -E\left(A + \frac{(z(1-\theta)\lambda)}{((1-\theta)\lambda+\phi)^2} - \frac{z}{((1-\theta)\lambda+\phi)}\right) = -\frac{A(1-\theta)\lambda}{\mu},$$

$$I_{\theta\phi} = -E\left(\frac{z\lambda}{((1-\theta)\lambda+\phi)^2}\right) = -\frac{A\lambda}{\mu}, \quad \text{and}$$

$$I_{\theta\theta} = -E\left(-\frac{z\lambda^2}{(1-\theta)\lambda+\phi)^2} - \frac{t_0-x_0}{(1-\theta)^2} - \frac{x_0}{\theta^2}\right) = -\frac{A_0\lambda\mu + A\lambda^2\theta(1-\theta)}{\theta(1-\theta)\mu}.$$

4 WALD CONFIDENCE INTERVALS FOR λ AND θ

In this chapter we consider the Wald−based confidence interval for λ and θ. We study their coverage and width properties in a Monte Carlo simulation.

4.1 Inverse of Fisher's Information

To find the Wald statistic from chapter 2 for λ and θ, we need the diagonal elements of the inverse of Fisher's information matrix, $I^{-1}(\lambda, \phi, \theta)$

The $(1,1)$ entry of $I^{-1}(\lambda, \phi, \theta)$ is

$$\begin{aligned} I^{11}(\lambda, \phi, \theta) &= -\frac{\lambda(A(-\theta\lambda + \theta^2\lambda - \phi) - \mu A_0)}{A_0(A(\lambda - \theta\lambda) + \phi) + \mu A_0} \\ &= -\frac{\lambda(A(\theta\lambda(-1+\theta) - \phi) - \mu A_0)}{A_0(A(\lambda(1-\theta) + \phi) + \mu A_0)}, \end{aligned}$$

Using $\mu = \lambda(1-\theta) + \phi$, we can rewrite $I^{11}(\lambda, \phi, \theta)$:

$$\begin{aligned} I^{11}(\lambda, \phi, \theta) &= \frac{\lambda(A(\theta\lambda(1-\theta) + \phi) + \mu A_0)}{A_0(A\mu) + \mu A_0} \\ &= \frac{\lambda(A(\theta\lambda(1-\theta) + \phi) + \mu A_0)}{A_0(A + A_0)\mu}. \end{aligned}$$

The $(2,2)$ entry of $I^{-1}(\lambda, \phi, \theta)$ is

$$I^{22}(\lambda, \phi, \theta) = -\frac{\phi(A(\theta - 1)\lambda - \mu A_0)}{A_0(A(\lambda - \theta\lambda + \phi) + \mu A_0)},$$

and using $\mu = \lambda(1-\theta) + \phi$,we can rewrite as

$$I^{22}(\lambda, \phi, \theta) = \frac{\phi(A(1-\theta)\lambda + A_0\mu)}{A_0(A + A_0)\mu}.$$

The $(3,3)$ entry of $I^{-1}(\lambda,\phi,\theta)$ is

$$\begin{aligned}
I^{33}(\lambda,\phi,\theta) &= \frac{(-1+\theta)\theta(A((-1+\theta)^2\lambda+\phi)\mu A_0)}{\lambda A_0(A((-1+\theta)\lambda-\phi)-\mu A_0)} \\
&= \frac{\theta(\theta-1)(A(\lambda(-1+\theta)^2+\mu)+\mu A_0)}{\lambda A_0(A(\theta-1)\lambda-\phi)-\mu A_0} \\
&= \frac{\theta(\theta-1)(A(\lambda(-1+\theta)^2+\mu)+\mu A_0)}{-\lambda A_0(A(1-\theta)\lambda+\phi)-\mu A_0},
\end{aligned}$$

and using $\mu=\lambda(1-\theta)+\phi$, we can rewrite $I^{33}(\lambda,\phi,\theta)$ as

$$\begin{aligned}
I^{33}(\lambda,\phi,\theta) &= \frac{\theta(\theta-1)(A(\lambda(-1+\theta)^2+\phi)+A-0\mu)}{-A_0\lambda(A\mu)-\mu A_0} \\
&= -\frac{\theta(\theta-1)(A(\lambda(1-\theta)^2+\phi)A_0\mu)}{A_0A\lambda\mu+\mu A_0} \\
&= \frac{\mu(1-\theta)(A(\lambda(\theta-1)^2+\phi)+A_0\mu)}{A_0\mu\lambda(A_0+A)}.
\end{aligned}$$

4.2 Confidence Intervals for λ and θ

Using the general approach described in section 2.5.5, we consider a Wald−based, large sample confidence interval for λ and θ.

4.2.1 Confidence Interval for λ

For the Wald Statistic for λ,

$$W_\lambda = \frac{\hat{\lambda}-\lambda}{\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})}},$$

we have that

$$P\left(-Z_{\alpha/2} \le W \le Z_{\alpha/2}\right) \;\approxeq\; 1-\alpha$$

for "large" sample sizes. Hence,

$$P\left(-Z_{\alpha/2} \leq \frac{\hat{\lambda}-\lambda}{\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})}} \leq Z_{\alpha/2}\right) \approxeq 1-\alpha$$

$$\Rightarrow P\left(-Z_{\alpha/2}\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})} \leq \hat{\lambda}-\lambda \leq Z_{\alpha/2}\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})}\right) \approxeq 1-\alpha$$

$$\Rightarrow P\left(-\hat{\lambda}-Z_{\alpha/2}\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})} \leq -\lambda \leq -\hat{\lambda}+Z_{\alpha/2}\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})}\right) \approxeq 1-\alpha$$

$$\Rightarrow P\left(\hat{\lambda}+Z_{\alpha/2}\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})} \geq \lambda \geq \hat{\lambda}-Z_{\alpha/2}\sqrt{I^{11}(\hat{\lambda},\hat{\phi},\hat{\theta})}\right) \approxeq 1-\alpha$$

$$\Rightarrow P\Bigg(\hat{\lambda}+Z_{\alpha/2}\sqrt{\frac{\hat{\lambda}(A(\hat{\theta}\hat{\lambda}(1-\hat{\theta})+\hat{\phi})+\hat{\mu}A_0)}{A_0(A+A_0)\hat{\mu}}} \geq \lambda \geq$$

$$\hat{\lambda}-Z_{\alpha/2}\sqrt{\frac{\hat{\lambda}(A(\hat{\theta}\hat{\lambda}(1-\hat{\theta})+\hat{\phi})+\hat{\mu}A_0)}{A_0(A+A_0)\hat{\mu}}}\Bigg) \approxeq 1-\alpha$$

Therefore, a large sample $(1-\alpha)100\%$ confidence interval for λ is

$$\hat{\lambda} \pm Z_{\alpha/2}\sqrt{\frac{\hat{\lambda}(A(\hat{\theta}\hat{\lambda}(1-\hat{\theta})+\hat{\phi})+\hat{\mu}A_0)}{A_0(A+A_0)\hat{\mu}}}$$

,

where $Z_{\alpha/2}$ is corresponding standard normal percentile.

4.2.2 Confidence Interval for θ

For the Wald statistic for θ

$$W_\theta = \frac{\hat{\theta} - \theta}{\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})}},$$

we have that

$$P\left(-Z_{\alpha/2} \leq W \leq Z_{\alpha/2}\right) \approxeq 1 - \alpha$$

for "large" sample sizes. Hence,

$$\begin{aligned}
P\left(-Z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})}}) \leq Z_{\alpha/2})\right) &\approxeq 1 - \alpha \\
\Rightarrow P\left(-Z_{\alpha/2}\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})} \leq \hat{\theta} - \theta \leq Z_{\alpha/2}\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})}\right) &\approxeq 1 - \alpha \\
\Rightarrow P\left(-\hat{\theta} - Z_{\alpha/2}\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})} \leq -\theta \leq -\hat{\theta} + Z_{\alpha/2}\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})}\right) &\approxeq 1 - \alpha \\
\Rightarrow P\left(\hat{\theta} + Z_{\alpha/2}\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta}} \geq \theta \geq \hat{\theta} - Z_{\alpha/2}\sqrt{I^{33}(\hat{\lambda}, \hat{\phi}, \hat{\theta})}\right) &\approxeq 1 - \alpha
\end{aligned}$$

$$\Rightarrow P\Bigg(\hat{\theta} + Z_{\alpha/2}\sqrt{\frac{\hat{\mu}(1-\hat{\theta})(A(\hat{\lambda}(\hat{\theta}-1)^2 + \hat{\phi}) + A_0\hat{\mu})}{A_0\hat{\mu}\hat{\lambda}(A_0 + A)}} \geq \theta \geq$$

$$\hat{\theta} - Z_{\alpha/2}\sqrt{\frac{\hat{\mu}(1-\hat{\theta})(A(\hat{\lambda}(\hat{\theta}-1)^2 + \hat{\phi}) + A_0\hat{\mu})}{A_0\hat{\mu}\hat{\lambda}(A_0 + A)}}\ \Bigg) \approxeq 1 - \alpha$$

Therefore, a large sample $(1-\alpha)100\%$ confidence interval for θ is

$$\hat{\theta} \pm Z_{\alpha/2}\sqrt{\frac{\hat{\mu}(1-\hat{\theta})(A(\hat{\lambda}(\hat{\theta}-1)^2 + \hat{\phi}) + A_0\hat{\mu})}{A_0\hat{\mu}\hat{\lambda}(A_0 + A)}}$$

.

where $Z_{\alpha/2}$ is corresponding standard normal percentile.

4.3 A Monte Carlo Simulation

In this Monte Carlo simulation, we used R to study the width and coverage properties of the Wald−based intervals for λ and θ. The R-code can be seen in Appendix A. We perform the simulation by first generating the infallible sample counts from their respective distributions: $T_0 \sim Pois(A_0\lambda)$. If the t_0 is greater than zero, then we create the $X_0 \sim Bi(t_0, \theta)$ otherwise x_0 set to zero, $Y_0 \sim Pois(A_0\phi)$, and Z_0 can be found using the T_0, Y_0, and X_0 ($Z_0 = T_0 + Y_0 - X_0$). Z_0 is forced to be one when its become zero because to prevent the numerical singularities. Next, the fallible sample is generated using the Poisson model with rate $\mu = \lambda(1-\theta) + \phi$: $Z \sim Pois(A\mu)$. This simulation study was carried out for different fallible and training sample sizes. For the confidence interval for λ the infallible sample sizes were chosen $A_0 = 1, 5, 10, 20$, and 50. For the confidence interval for θ, $A_0 = 5, 10, 20$, and 50 . The fallible sample size chosen in both setting were $A = 10, 50, 100, 1000$. The parameter configuration chosen for the study of the confidence interval for λ were: $\lambda = 1, 2, 3, 4, \ldots, 100$, $\theta = 0.05, 0.25$, and $\phi = 1, 5$. The parameter configurations chosen for the study of the Confidence interval for θ were: $\theta = 0.02, 0.03, 0.04, \ldots, 0.98$; $\lambda = 5, 10, 15, \phi = 1, 5$. A 90% nominal confidence level has been considered for each simulation and $10,000$ iterations were performed in each simulation. The simulation results were used to create plots of estimated coverages and estimated average widths of confidence interval using Minitab statistical software. Those graphs appear in Appendix B.

4.3.1 Simulation Interpretation for Confidence Interval for λ

In Appendix B, Figures 6.1 and 6.2 display the estimated actual coverage of the confidence interval for λ. When $\theta = 0.05$, $\phi = 1$, and $A_0 = 1$, we see the confidence interval often undercovers. This is especially true when $A = 1000$, or when $\left(\frac{A_0}{A}\right) = 0.001$. This particular curve converges to the nominal level much more slowly than for smaller A. One reason of this phenomenon might be the large ratio of tainted data to good data.

For second plot of first column in Appendix B, Figure 6.1 when $\phi = 5$, the coverages converge more quickly to the nominal level. This is also true for plots in Figure 6.2 first column with the increase in θ. One reason for improved coverage might be more observed false−negative and false−positive counts in the infallible sample, which helps to better estimate θ and ϕ, and hence more appropriately account for the false−negative and false−positive misclassification. We also see that for increasing A_0 respectively from 1 to $5, 10, 20$, and 50 (Note plots for infallible sample size of $A_0 = 20, A_0 = 50$ that are not included because of similarity with plots of $A_0 = 10$ in Appendix B, Figures 6.1 and 6.2.), the confidence interval for λ covers quite well. Overall, for larger infallible sample size, θ, and ϕ the confidence interval for λ has good coverage properties.

Figures 6.3 and 6.4 in Appendix B display estimated average widths for the confidence interval for λ. First, we note that the estimated average width decreases with an increased A for a fixed A_0. More interesting perhaps is that the estimated average widths increase with larger θ and ϕ. Another interesting aspect is that the estimated average width is getting smaller when the training sample size increases. Overall, the estimated average widths are reduced when more counts in the training sample may be observed. These counts include the false−positive and false−negative counts.

4.3.2 Simulation Interpretation for Confidence Interval for θ

When we look in Appendix B, Figures 6.5 and 6.6 for the estimated coverage of the confidence interval for θ, we can see relatively upside down 'U' shape curves for each ratio $\left(\frac{A_0}{A}\right)$ in each plot. In all configurations, the ratio $\left(\frac{A_0}{A}\right)$ seems to have no effect on the coverage properties. However, as λ increase, the rate of convergence quickens.

We also see improvements in converges with A_0 increases, which is intuitive. We also note that as ϕ goes from 1 to 5, there is not much influence on coverage.

From Appendix B, Figures 6.7 and 6.8, we see reduction in the estimated average width as A_0 increase, and λ grows. The reason for this may be there are more counts of true occurrence with the Poisson distribution with large λ and A_0. Also, changes in ϕ seem to have very little effect on the estimated average widths.

As mentioned above, the interested upside down 'U' shape, appears for the both estimated coverage confidence interval and estimated average width of the confidence interval. One possible reason for this shape is that the variance of the binomial distribution that is quadratic in its probability of success. The upside down 'U' shape is not exactly symmetric, rather, its shape is changing with the infallible sample size and θ. When the sample size is small and $\phi = 1$, the upside down 'U' shape is slightly skewed right, but with larger samples $\phi = 5$ the shape approaches symmetry. Overall in this simulation study of the estimated coverage and estimated average of confidence interval of θ, we have approximately symmetric shapes.

4.3.3 Overall Conclusion

According to the above interpretations, in order to get good coverage and width properties, we need to have a sufficiently large training sample to get the good classification of misclassified data. Most importantly, when we have large counts for true occurrences, false−positives, false−negatives from the training sample by infallible classifier and fallible classifier, we get more useful information from fallible sample, hence, we can more precisely and accurately estimate parameters in this model.

5 APPLICATION ON REAL DATA

In this section, we consider the data sets from chapter one, section 1.1 to illustrate the usefulness of the model, and obtain confidence intervals for the parameter of interest in each example. First, we need to identify the proper sample units and their measured counts for each sample. Subsections of this chapter will discuss single sample confidence intervals and compare these with double sample confidence intervals using the misclassification model in an effort to show the utility of the model with the double−sampling scheme.

5.1 Engage With Example One

Suppose, there is interest in the rate of death due to digestive disease per $10,000$ person−years, death rate per $10,000$ person−years of classified as digestive disease but cause of death is not digestive disease (false−positive misclassification rate) , and probability of death cause by digestive disease but classified otherwise (probability of a false− negative observation). Hence, the search sample sizes need to be in per $10,000$ person−years units used commonly by epidemiologists.

In this example, we have the following classifiers.

Infallible classifier (expensive) : Clinical autopsy

Fallible classifier (less expensive) : Forensic autopsy

In the training sample there were 272 individuals receiving both clinical autopsy and forensic autopsy.

In the fallible sample, there were 3604 individual death certificates with only forensic autopsy.

To convert the sample counts to search a sample size in $10,000$ person−years, we need the following formula for incidence proportion (Risk) [7],

$$\frac{\text{Number of deaths in sample}}{\text{Size of population at start of period}} \times 10,000.$$

Now, Stamey and Young (2005), calculated the two sample sizes in $10,000$ person−years. According to the formula above, it appearances they used approximately 29.4% of population (913752) in Connecticut in 1980 as a size of population at start of period. This might be the part of the population above age 54, which is mean age of training sample [16]. There is not much information in Stamey and Young (2005), and Kircher, Nelson and Burdo (1985) papers, about the person−years. However, there has to have been some assumption which is made by Stamey and Young (2005) (which is unclear in the paper) in order to get the following search sample sizes:

$$\begin{aligned} A_0 &= 3 \text{ in } 10,000 \quad \text{person}-\text{year units} \\ A &= 39.4 \text{ in } 10,000 \;\; \text{person}-\text{year units.} \end{aligned}$$

The observed counts in training sample are:

$$\begin{aligned} t_0 \;\; (\text{true number of deaths due to digestive disease }) &= 32 \\ y_0 \;\; (\text{number of deaths are incorrectly classified as due to digestive disease }) &= 02 \\ x_0 \;\; (\text{number of deaths caused by digestive disease but classified otherwise}) &= 16 \\ z_0 \;\; (\text{number of deaths classified by fallible classifier as due to digestive disease}) &= 18 \end{aligned}$$

The only observed count in fallible sample is:

$$z \;\; (\text{number of deaths classified by fallible classifier as due to digestive disease}) \; = 219$$

z is calculated using 6.1% of deaths classified as due to digestive disease from forensic autopsies of all deaths in Connecticut in 1980 ($3604 \times \frac{6.1}{100} \approx 219$).

The parameters in this example are defined as follows:

$$\begin{aligned}
\lambda &= \text{average number of deaths per } 10,000 \text{ person}-\text{years due to digestive disease} \\
\theta &= \text{probability of death caused by digestive disease but classified otherwise} \\
\phi &= \text{average number of deaths per } 10,000 \text{ person}-\text{years due to digestive disease but} \\
&\quad\ \text{classified as due to non}-\text{digestive disease}
\end{aligned}$$

The double−sample−based MLE values and confidence interval formulas are used to estimate the parameters. The 90% asymptotic confidence intervals for parameters noted above and widths of these confidence intervals are represented in Table 5.1.

Parameter	Estimate	90% asymptotic confidence interval	Width of confidence internal
λ	10.3018	$[7.9374, 12.6663]$	4.7289
θ	0.5177	$[0.4058, 0.6295]$	0.2237
ϕ	0.6210	$[0, 1.3298]$	1.3298

Table 5.1: Parameter Estimates, Confidence Intervals and Width of Confidence Interval Using Double Sample Procedure for Digestive Disease Data

A simulation was conducted using parameters near these parameter estimates in Table 5.1 with $A_0 = 3$ and $A = 39.4$. These simulation results for the confidence interval for λ are in the Appendix B, Figure 6.9, indicating that our confidence interval on λ has appropriate coverage in this example. Figure 6.10 indicates our confidence interval on θ might slightly undercover as compared to the nominal level of 90%.

Note that the single sample (only training sample) confidence intervals for λ, θ, and ϕ are:

Large sample $(1-\alpha)100\%$ confidence interval for λ:

$$\frac{t_0}{A_0} \pm Z_{\alpha/2}\sqrt{\frac{t_0}{A_0}}$$

Large sample $(1-\alpha)100\%$ confidence interval for θ:

$$\frac{x_0}{t_0} \pm Z_{\alpha/2}\sqrt{\frac{\left(1-\frac{x_0}{t_0}\right)\frac{x_0}{t_0}}{t_0}}$$

Large sample $(1-\alpha)100\%$ confidence interval for ϕ:

$$\frac{y_0}{A_0} \pm Z_{\alpha/2}\sqrt{\frac{y_0}{A_0}}.$$

The computed 90% single sample confidence intervals for λ, θ, and ϕ are:

Parameter	Estimate	90% asymptotic confidence interval	Width of confidence internal
λ	10.6667	$[5.2942, 16.0392]$	10.7450
θ	0.5	$[0.3546, 0.6454]$	0.2908
ϕ	0.6667	$[0, 2.0098]$	2.0098

Table 5.2: Parameter Estimates, Confidence Intervals and Width of Confidence Interval for Only Training Sample Data in Example one

According to the Table 5.1 and 5.2, the estimated death rate due to digestive disease is 3.42% smaller for double sample procedure than for the single sample procedure, while the estimated probability of a false−negative is 3.5% bigger and the estimated false−positive rate is 6.85% smaller. One reason for these differences might be the larger fallible sample.

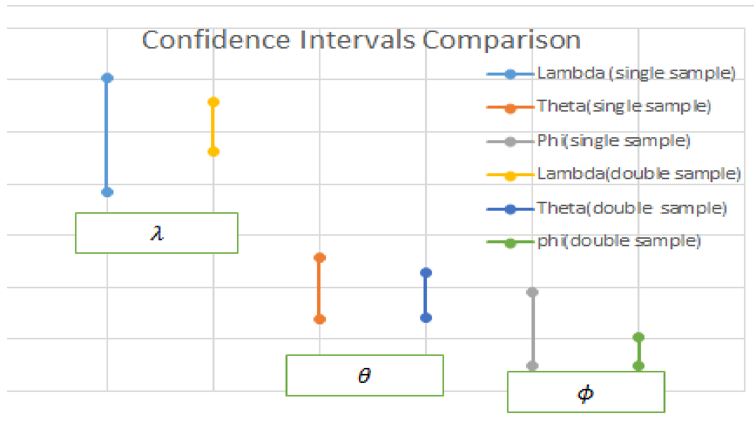


Figure 5.1: Comparison of Single and Double Sample CI for Example One

We can see clearly in the Table 5.2 and Figure 5.1, the width of all the confidence intervals from using only the training sample (single sample) are larger than the Wald confidence intervals from the double sample procedure. Furthermore, the confidence interval width ratios of the single sample confidence intervals versus the double sample confidence intervals for λ, θ, and ϕ, respectively 2.2726, 1.2999, and 1.5113. Hence, using the double sample process, we get the more precise information about the parameters than when only using the single infallible sample data to estimate the parameters.

5.2 Engage With Example Two

From the fictitious, yet realistic example two, from chapter one section 1.1, we have the following classifiers:

Infallible classifier (expensive) : Manual method

Fallible classifier (less expensive) : Ultrasonic method

Training sample (A_0) $= 2,000$ feet oil pipeline

Fallible sample (A) $= 30,000$ feet oil pipeline

To convert to per $1,000$ feet oil pipeline, we need the following formula for incidence rate,

$$\frac{\text{Search feet of oilpipeline}}{\text{1000 feet}}.$$

Hence, our search sample sizes are:

$$\begin{aligned} A_0 &= 2 \;\; (1,000 \text{ feet oil pipeline}) \\ A &= 30 \;\; (1,000 \text{ feet oil pipeline}). \end{aligned}$$

The observed counts in the training sample are:

$$\begin{aligned} t_0 \;\; (\text{true number of defect of oilpipeline }) &= 23 \\ y_0 \;\; (\text{number of defects are classified incorrectly as defect }) &= 16 \\ x_0 \;\; (\text{number of defects are classified as non-defects}) &= 14 \\ z_0 \;\; (\text{number of defects are observed by fallible classifier}) &= 25 \end{aligned}$$

The observed counts in the fallible sample is:

$$z \;\; (\text{number of defects are observed by fallible classifier}) \qquad = 145$$

The parameters in this example have the following definitions:

$$\begin{aligned} \lambda &= \text{Average number of defects per 1,000 feet of oil pipeline} \\ \theta &= \text{Probability of missing a defect by ultrasonic method} \\ \phi &= \text{Average number of falsely declared defects per 1,000 feet oil pipeline} \\ &\quad\ \text{by ultrasonic method} \end{aligned}$$

Their respective MLE's and 90% confidence intervals are given in Table 5.3.

Parameter	Estimate	90% asymptotic confidence interval	Width of confidence internal
λ	8.9125	$[5.5680, 12.2569]$	6.6889
θ	0.7854	$[0.6482, 0.9225]$	0.2743
ϕ	3.4	$[2.0434, 4.7565]$	2.7131

Table 5.3: Parameter Estimates, Confidence Intervals and Width of Confidence Interval for the Parameters for Example Two

A simulation was also conducted using parameters near these parameter estimates in Table 5.3 with $A_0 = 3$ and $A = 30$. These simulation results for the confidence interval for λ are in the Appendix B, Figure 6.11, indicating that our confidence interval on λ has appropriate coverage in this example. Figure 6.12 indicates our confidence interval on θ might slightly undercover as compared to the nominal level of 90%.

The large sample single sample confidence intervals (CI) for example two are given in Table 5.4.

Parameter	Estimate	90% asymptotic confidence interval	Width of confidence internal
λ	11.5	$[5.9216, 17.0784]$	11.1568
θ	0.6086	$[0.4412, 0.7760]$	0.3348
ϕ	8	$[3.3472, 12.6527]$	9.3055

Table 5.4: Parameter Estimates, Confidence Intervals and Width of Confidence Interval for the Parameters Corresponding to Only Infallible Sample Data for Example Two

The point estimate of the average number of defects per $1,000$ feet of oil pipeline in single sample procedure is reduced by 22.5% using the double sample procedure. The estimated average number of false defects per $1,000$ feet oil pipeline in single sample is decreased by 57.5% in the double sample. In the single sample procedure, the estimated probability of missing a defect is increased by 29.0% using the double sample procedure. One reason for the great disparities is likely due to the small training sample and large false negative misclassification.

Figure 5.2 compares the confidence intervals under the double sampling scheme and only training single sampling scheme.

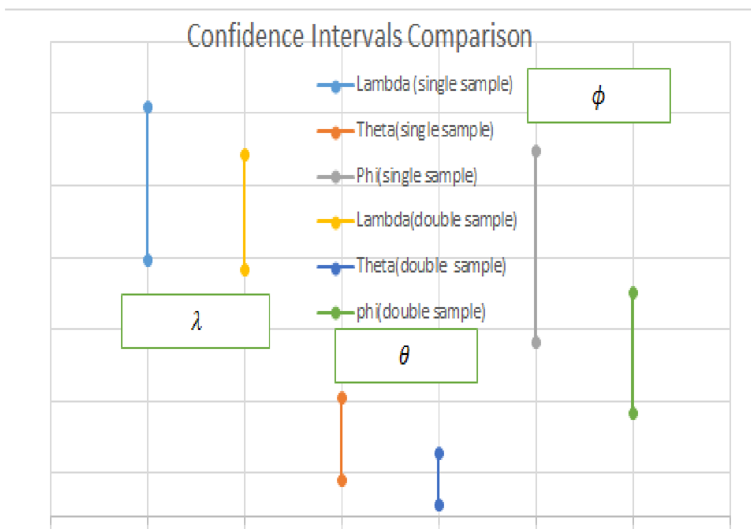


Figure 5.2: Comparison of Single and Double Sample CI for Example Two

In Figure 5.2, the width of confidence intervals for parameters λ, θ, and ϕ in the double sample procedure are much narrowest than the single sample procedure. The width ratios of single sample versus double sample procedure of λ and ϕ are 1.6679 and 3.4298, respectively. The width ratio for θ is greater than one (1.2183), hence, double sample confidence interval is relatively smaller than single sample procedure.

6 COMMENTS

In this paper, we studied confidence intervals for the single Poisson parameter from a model where the data are subject to misclassification, as well as the false−negative probability interval estimators by inverting the appropriate Wald statistics. The Wald confidence intervals coverage and width properties were studied in Monte Carlo simulation for various parameter and sample size configurations. The confidence interval for λ performed well except for small training sample size and large fallible sample size with small θ and ϕ. Also, the confidence interval for θ carried out good coverage and width properties except for the small λ and small training sample. One interesting outcome is that an appropriate fallible sample size depends on the training sample size in order to maintain the good coverage and width properties of the confidence intervals.

Finally, we applied confidence intervals to real data sets involving death certificates and oil pipeline defects. The model and resulting confidence intervals provide better estimates of the average death rate due to digestive failure and average number of defects per $1,000$ feet of oil pipeline, as well as of the misclassification involved in death certificate and defects on oil pipeline for the populations of interest. These two examples, support the appropriate use of a double sample procedure as a better method for finding good confidence intervals for the Poisson parameter of primary interest when the data is subject to misclassification. For future study, we could compute score and profile log−likelihood confidence interval, and study their coverage and width properties. Also, we could study the coverage and width properties for a confidence interval on ϕ. Finally, better understanding the effects of the ratio of training sample and fallible sample could be another study.

BIBLIOGRAPHY

[1] D.Madigan, I. Heuch, J. York, and R.T. Lie, *Birth defects registered by double sampling: A bayesian approach incorporating covariates and model uncertainty.*, Applied Statistics. **44** (1995), 227–242.

[2] T.W.Gyorkos, L. Joseph and L. Coupal, *Bayesian estimation of disease prevalence and parameters of diagnostic tests in the absence of a gold standard.*, Amer. J. Epidemiol. **141** (1995), 263–272.

[3] V. Ramakrishnan, M. Viana and P. Levy, *Bayesian analysis of prevalence from the result of small screening samples.*, Comm. Statist. Theory Methods. **22** (1993), 575–585.

[4] W.J.Mckean, V.R.Hogg, and T.A.Craig, *Introduction to mathematical statistics,sixth edition ed.*, Pearson Prentice, 2005.

[5] T. Bratcher, and J. Stamey, *Estimation of poisson rates with misclassified counts.*, Biometrical J. **44** (2002), 946–956.

[6] I. Bross, *Misclassification in 2× 2 tables.*, Biometrics. **10** (1950), 478–486.

[7] T.L. Bratcher, C. Anderson, and K. Kutran, *Bayesian estimation of population density and visibility .*, Texas J.Sci. **46** (1994), 7–12.

[8] T.T. Chen, *Log-linear models for categorical data with misclassification and double sample.*, J.Amer. Statist. Assoc. **74** (1979), 481–488.

[9] Centers for Disease Control and Prevention (CDC), *An introduction to applied epidemiology and biostatistics*, third edition ed., U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, Atlanta, GA 30333, 2006 (English).

[10] G.M.Clarke, and D.Cooke, *A basic course in statistics*, fourth ed., ARNOLD, 338 E, 1998 (English).

[11] Y. Hochberg, *On the use of double sample schemes in analysing categorical data with misclassification errors.*, J. Amer. Assoc. **72** (1977), 914–921.

[12] D.M.Young, J.D. Stamey, *Maximum likelihood estimation for a poisson parameter with misclassified counts*, Australian and New Zealand Journal of Statistics (2005), 163–172.

[13] R.Choudary Hanumara, John M.Hoenig, and Dennis M.Heisey, *Generalizing double and triple sampling for repeated surveys and partial verification*, Biometrical Journal **44** (2002), 603–618.

[14] J.D. Stamey, K. Riggs, and D.M. Young, *Interval estimation for misclassification rate parameters in a complementary poisson model.*, Communications in Statistics Theory and Method (2009), 159–172.

[15] A. Tenenbein, *A double sampling scheme for estimation from binomial data with misclassification* ., J. Amer. Statist. Assoc. **141** (1970), 263–272.

[16] T.Kircher, J.Nelson, and H. Burdo*The autopsy as a measure of accuracy of the death certificate.*, The New England Journal of Medicine **20** (1985), 1263–1269.

APPENDIX

APPENDIX A :- R-Software Code for Simulation

SIMULATION CODE FOR λ

```
library(xlsx)
N<-10000
L <-c()
U <-c()
total<-c()
est_lambda_hat<-c()
est_phi_hat<-c()
est_mu_hat<-c()
est_theta_hat<-c()
est_lambda_covert<-c()
est_var_lambda_hat<-c()
lamda<-c()
coverage <-c()
width1<-c()
width<-c()
SDAc<-c()
SDCw<-c()
Ao <-1 # sample Size of Infallible
A <-100# sample Size of fallible
e1 <-(Ao/(A+Ao)) # alpha1 e1+e2=1
```

```
e2 <-(1-e1) # alpha2
theta<-0.05
phi<-1
for (m in 1:100)
{
lamda[m]<-(10+m)
}
for(k in 1:100)
 {
 lambda=lamda[k]
 mu<-(lambda*(1-theta)+phi)
for(i in 1:N)
 {
to <- rpois(1,Ao*lambda)
xo <- rbinom(1,size=to,prob=theta)
yo <- rpois(1,Ao*phi)
z  <- rpois(1,A*mu)
zo<-to+yo-xo
if(zo==0)
{zo=1
}
est_lambda_hat[i]<-(e1*(to+z*(1-yo/zo))/(Ao))+
(e2*(xo/Ao)) # Estimate lambda
est_phi_hat[i] <-(yo*(z+zo))/((A+Ao)*zo)
# Estimate Phi
est_theta_hat[i]<-(xo*(Ao+A))/(Ao*(z+to+(xo/Ao)
*A-(yo/zo)*z))  # Estimate Theta
```

```
est_mu_hat[i]<-(est_lambda_hat[i]*(1-est_theta_hat[i])
+est_phi_hat[i])
#print(paste("zo: ",zo))
est_var_lambda_hat[i]<-(est_lambda_hat[i]*
(A*(est_theta_hat[i]*est_lambda_hat[i]*
(1-est_theta_hat[i])+est_phi_hat[i)
+est_mu_hat[i]*Ao))/(Ao*est_mu_hat[i]*(Ao+A))

# Wald 90%confidence for occurence-rate parameter(lambda)

L[i]<-est_lambda_hat[i]-1.645*sqrt(est_var_lambda_hat[i])
U[i]<-est_lambda_hat[i]+1.645*sqrt(est_var_lambda_hat[i])
total[i]<- ifelse(((lambda>L[i])&(lambda<U[i])),1,0)
width1[i]<-(U[i]-L[i])
}
#print(total)
coverage[k] <- mean(total)
width[k]<-mean(width1)
SDAc[k]<-sd(total)/sqrt(10000)
SDCw[k]<-sd(width1)/sqrt(10000)
}
coverage
width
SDAc
SDCw
par(mfrow=c(1,2))
plot(lamda,coverage, pch = 1)
```

```
plot(lamda,width, pch =2 )

write.xlsx(coverage,"C:/Users/Nishantha Janith/Desktop/Reserach out put/Lambda/
NewNewWay/Ao1_theta0.05_phi1/
OutputCov_A100.xlsx")
write.xlsx(width, "C:/Users/Nishantha Janith/Desktop/Reserach out put/Lambda/
NewNewWay/Ao1_theta0.05_phi1/
OutputWid_A100.xlsx")
write.xlsx(SDAc, "C:/Users/Nishantha Janith/Desktop/Reserach out put/Lambda/
NewNewWay/Ao1_theta0.05_phi1/
OutputCov_SD_A100.xlsx")
write.xlsx(SDCw, "C:/Users/Nishantha Janith/Desktop/Reserach out put/Lambda/
NewNewWay/Ao1_theta0.05_phi1
/OutputWid_SD_A100.xlsx")
```

SIMULATION CODE FOR θ

```
library(xlsx)
N<-10000
L <-c()
U <-c()
total<-c()
est_lambda_hat<-c()
est_phi_hat<-c()
est_mu_hat<-c()
est_theta_hat<-c()
est_lambda_covert<-c()
est_var_theta_hat<-c()
```

```
thetaa<-c()

coverage <-c()
width1<-c()
width<-c()
Ac<-c()
Cv<-c()

Ao <-50 # sample Size of Infallible
A <-10# sample Size of fallible
e1 <-(Ao/(A+Ao)) # alpha1 e1+e2=1
e2 <-(1-e1) # alpha2
lambda<-10
phi<-1
for (m in 1:100)
{
thetaa[m]<-(m/100)
}
for (k in 1:100)
 {
 theta=thetaa[k]
 mu<-(lambda*(1-theta)+phi)
for (i in 1:N)
 {
to <- rpois(1,Ao*lambda)
if(to >0)
{
```

```
xo <- rbinom(1,size=to,prob=theta)
}
if(to==0)
{
xo<-0
}
yo <- rpois(1,Ao*phi)
z  <- rpois(1,A*mu)
zo<-to+yo-xo
if(zo==0)
{zo=1
}
est_lambda_hat[i]<-(e1*(to+z*(1-yo/zo))/(Ao))
+(e2*(xo/Ao)) # Estimate lambda
est_phi_hat[i] <-(yo*(z+zo))/((A+Ao)*zo)
# Estimate Phi
est_theta_hat[i]<-(xo*(Ao+A))/(Ao*(z+to+(xo/Ao)
*A-(yo/zo)*z))  # Estimate Theta
est_mu_hat[i]<-(est_lambda_hat[i]*(1-est_theta_hat[i])+est_phi_hat[i])
#print(paste("zo: ",zo))
est_var_theta_hat[i]<-(est_theta_hat[i]*
(1-est_theta_hat[i])*(A*(est_lambda_hat[i]*
(est_theta_hat[i]-1)^2+est_phi_hat[i])
+Ao*est_mu_hat[i]))/(Ao*est_lambda_hat[i]*
est_mu_hat[i]*(Ao+A))
# Wald 90%confidence for occurence-rate parameter(theta)
L[i]<-est_theta_hat[i]-1.645*sqrt(est_var_theta_hat[i)
```

```
U[i]<-est_theta_hat[i]+1.645*sqrt(est_var_theta_hat[i)
total[i]<- ifelse(((theta>L[i])&(theta<U[i])),1,0)
width1[i]<-(U[i]-L[i])
}
#print(total)
coverage[k] <- mean(total)
width[k]<-mean(width1)
SDAc[k]<-sd(total)/sqrt(10000)
SDCw[k]<-sd(width1)/sqrt(10000)
}
coverage
width
SDAc
SDCw
par(mfrow=c(1,2))
plot(thetaa,coverage, pch = 1)
plot(thetaa,width, pch =2 )
write.xlsx(coverage, "C:/Users/Nishantha Janith/Desktop/Reserach output/Theta/
Ao50_lambda15_phi1/OutputCov_A1000.xlsx")
write.xlsx(width, "C:/Users/Nishantha Janith/Desktop/Reserach out put/Theta/
Ao50_lambda15_phi1/OutputWid_A1000.xlsx")
write.xlsx(Ac, "C:/Users/Nishantha Janith/Desktop/Reserach out put/Theta/
Ao50_lambda15_phi1/OutputCov_SD_A1000.xlsx")
write.xlsx(Cv, "C:/Users/Nishantha Janith/Desktop/Reserach out put/Theta/
Ao5_lambda15_phi1/OutputWid_SD_A1000.xlsx")
```

APPENDIX B

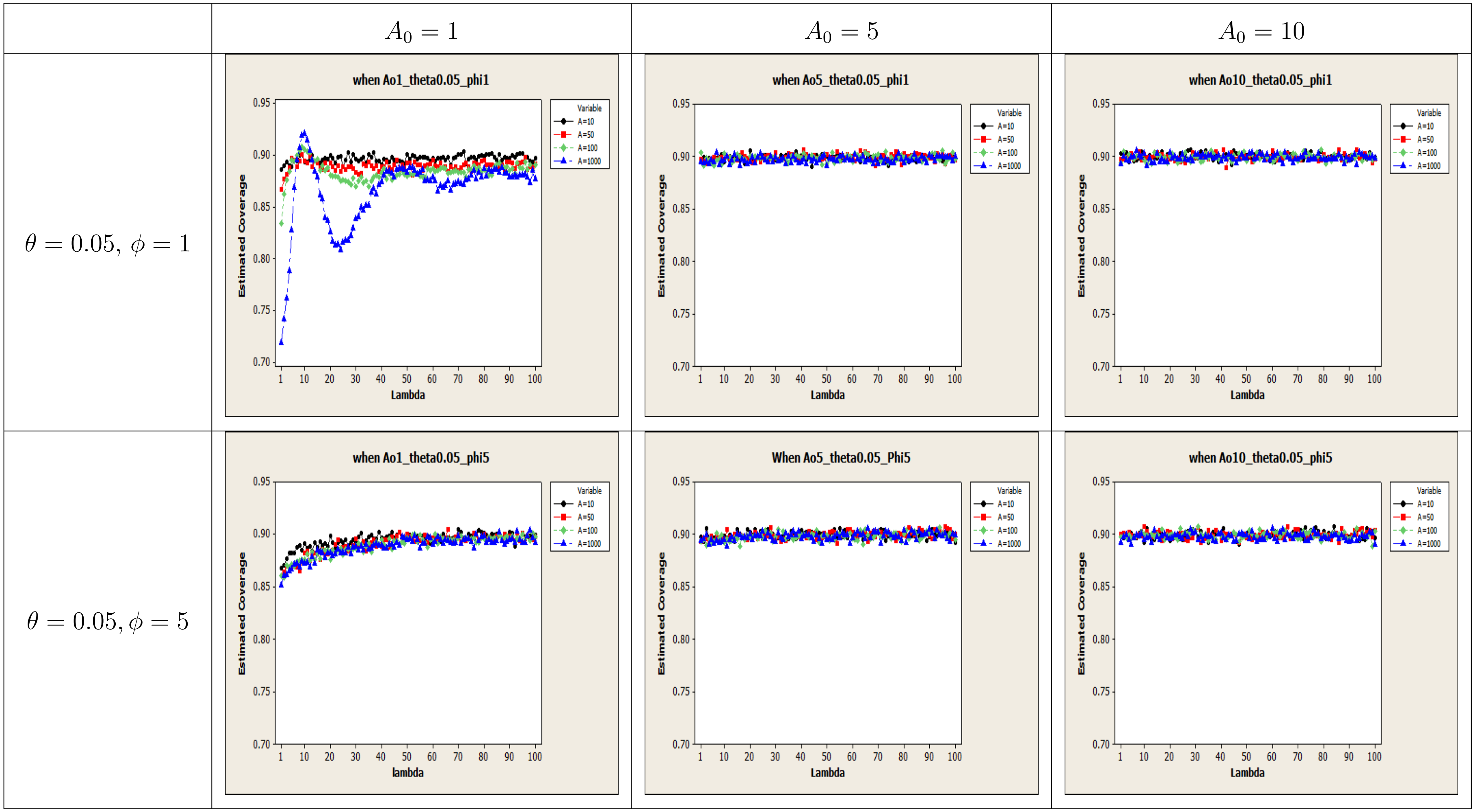


Figure 6.1: Estimated Actual Coverage for Confidence Interval on λ for $\theta = 0.05$ and $\phi = 1, 5$

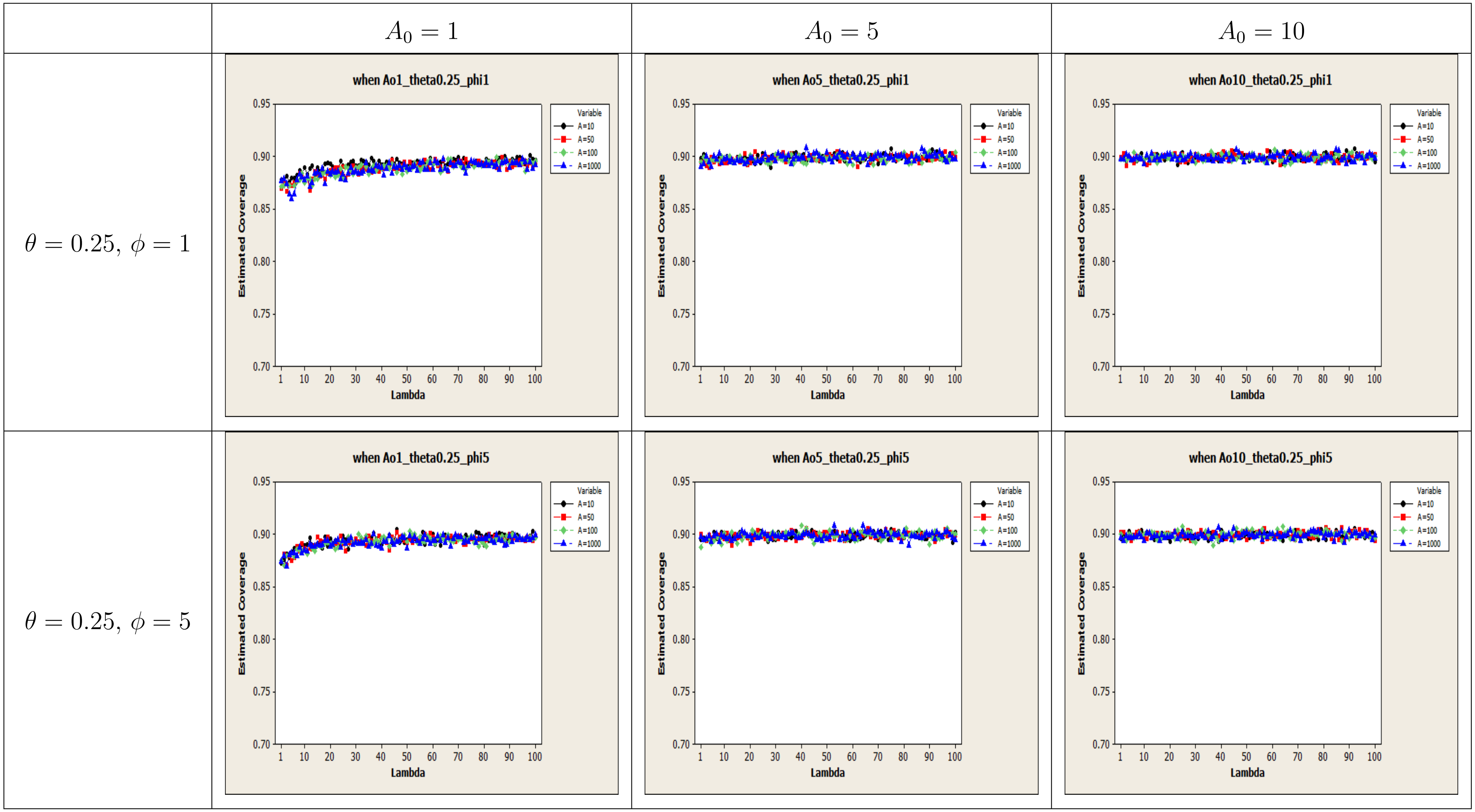


Figure 6.2: Estimated Actual Coverage for Confidence Interval on λ for $\theta = 0.25$ and $\phi = 1, 5$

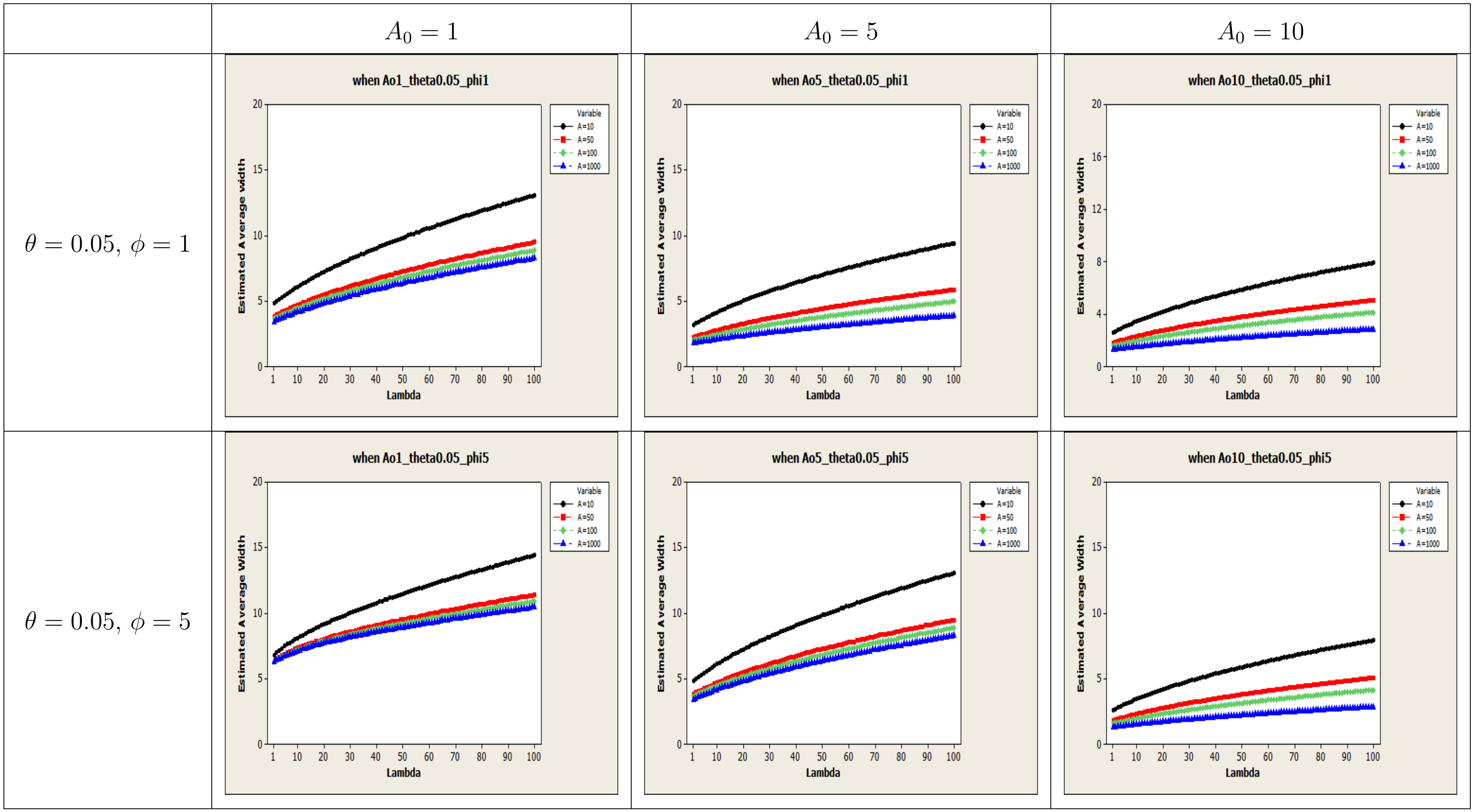


Figure 6.3: Estimated Average Width for Confidence Interval on λ with $\theta = 0.05$ and $\phi = 1, 5$

$\theta = 0.25$, $\phi = 1$	
$\theta = 0.25$, $\phi = 5$	

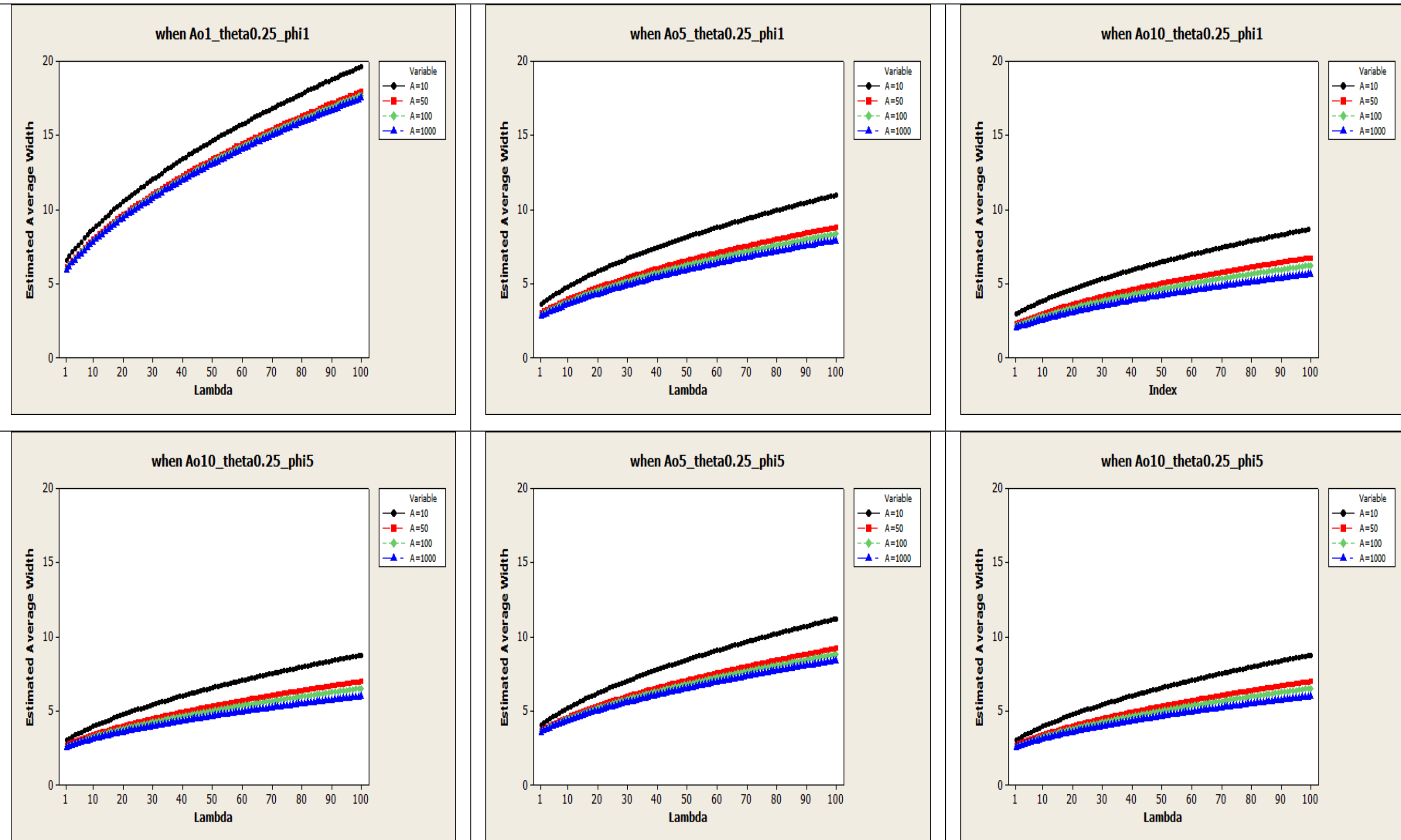


Figure 6.4: Estimated Average Width for Confidence Interval on λ with $\theta = 0.25$ and $\phi = 1, 5$

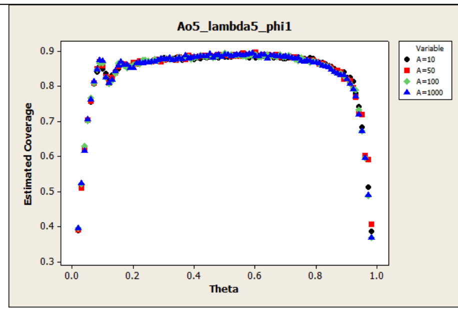
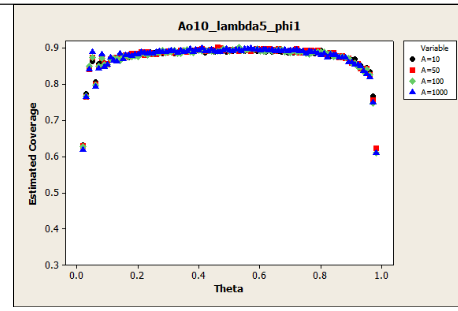
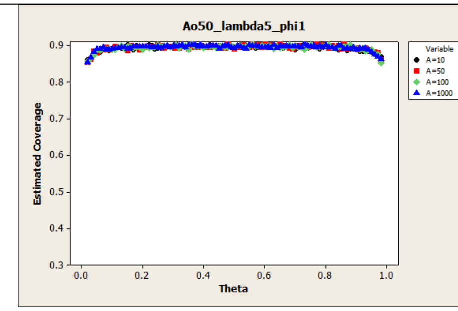
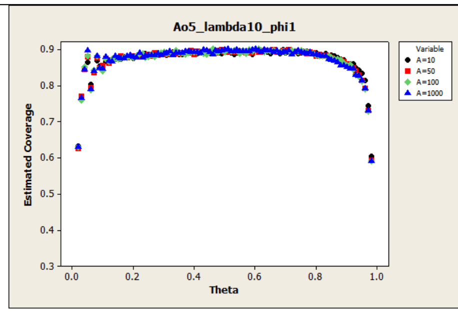
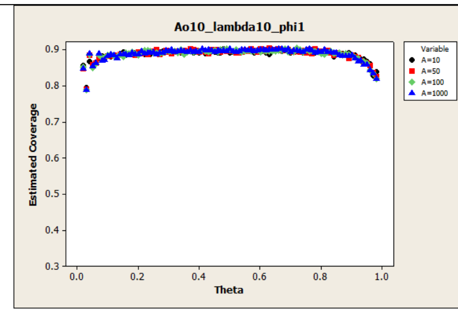
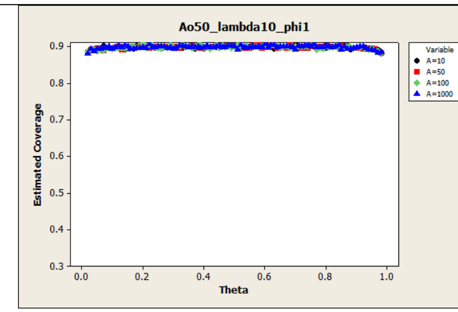
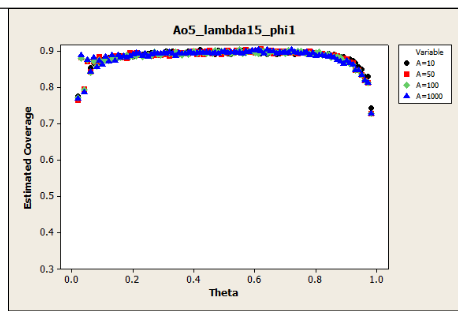
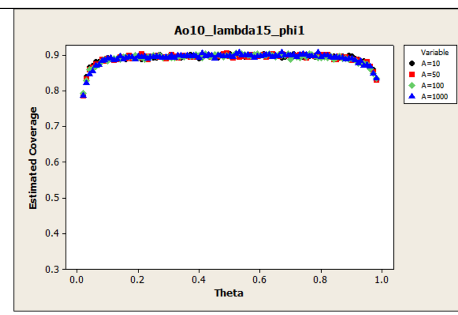
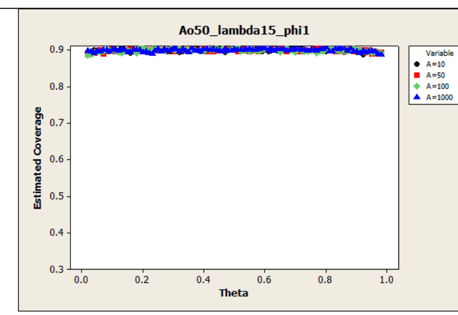
	$A_0 = 5$	$A_0 = 10$	$A_0 = 50$
$\lambda = 05\ \phi = 1$			
$\lambda = 10\ \phi = 1$			
$\lambda = 15\ \phi = 1$			

Figure 6.5: Estimated Actual Coverage for Confidence Interval on θ with Different λ and $\phi = 1$

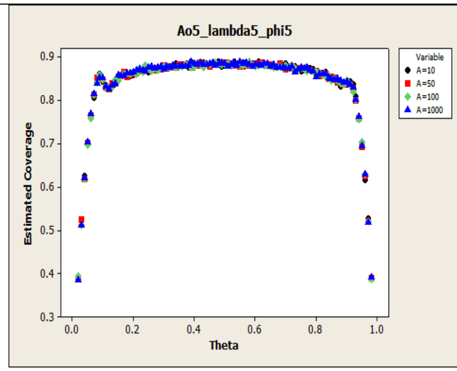
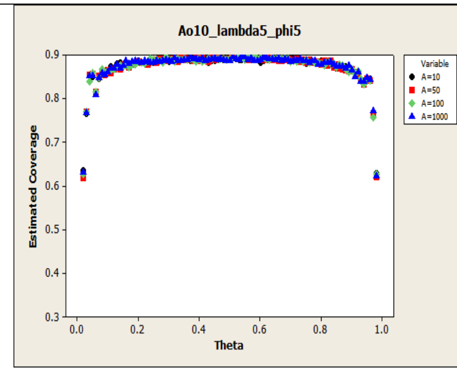
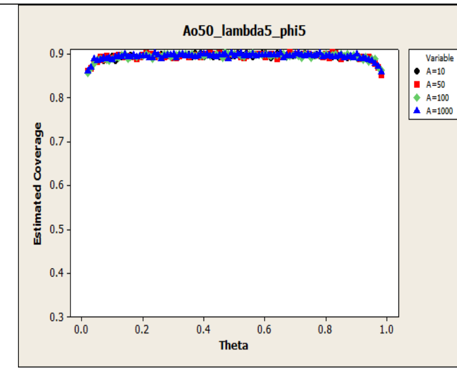
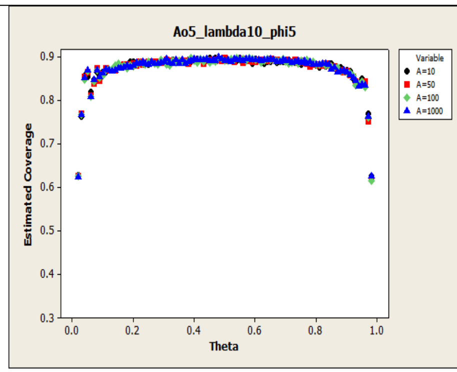
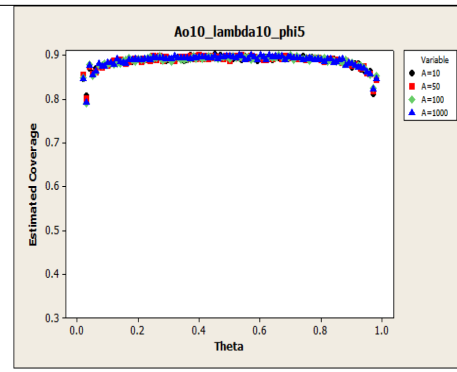
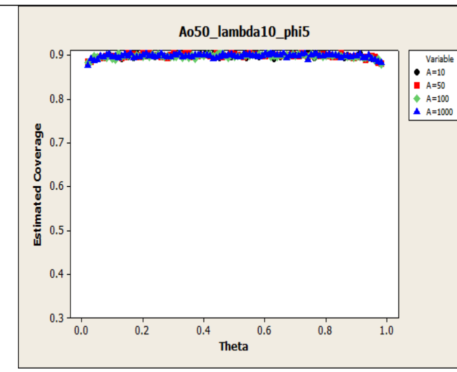
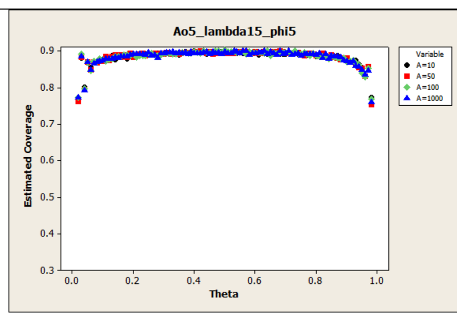
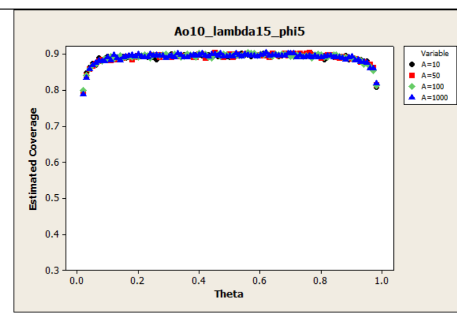
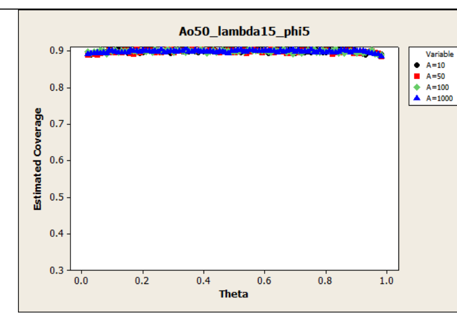
	$A_0 = 5$	$A_0 = 10$	$A_0 = 50$
$\lambda = 05$ $\phi = 5$			
$\lambda = 10$, $\phi = 5$			
$\lambda = 15$, $\phi = 5$			

Figure 6.6: Estimated Actual Coverage for Confidence Interval on θ with Different λ and $\phi = 5$

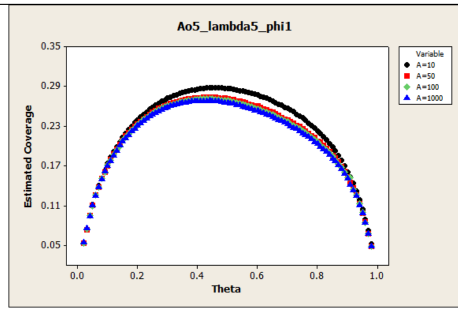
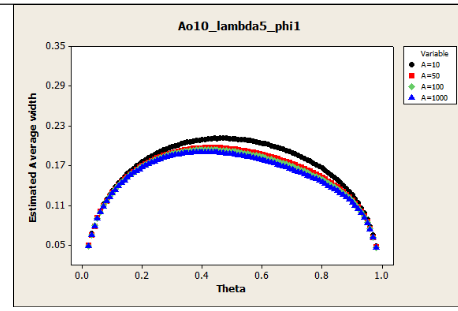
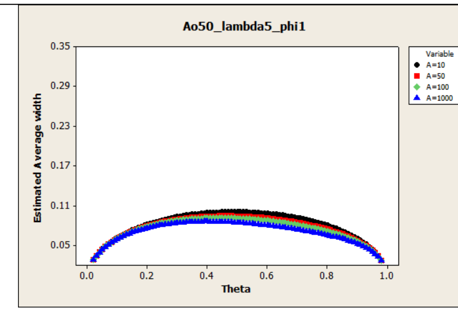
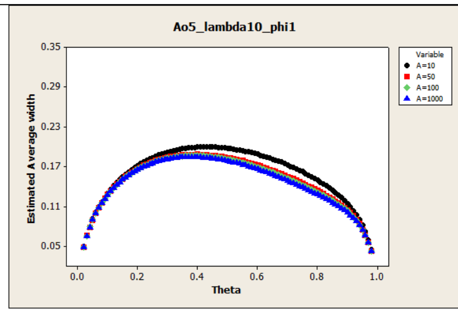
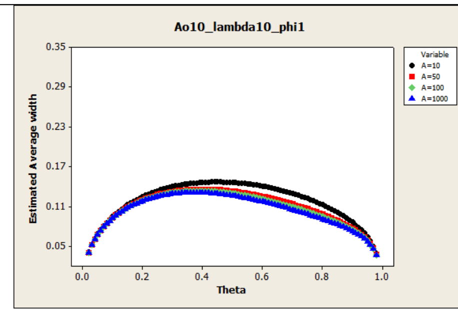
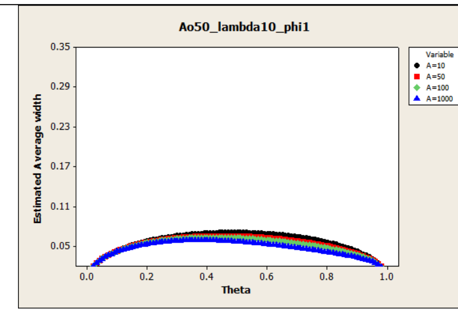
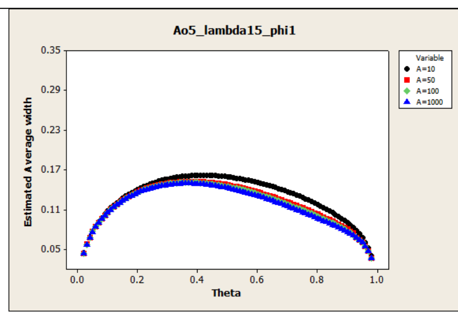
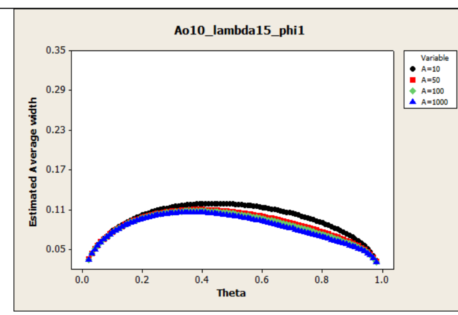
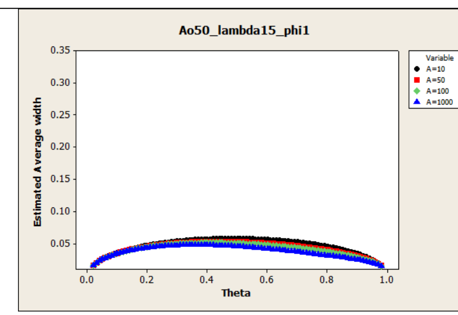
	$A_0 = 5$	$A_0 = 10$	$A_0 = 50$
$\lambda = 05$, $\phi = 1$			
$\lambda = 10$, $\phi = 1$			
$\lambda = 15$, $\phi = 1$			

Figure 6.7: Estimated Average Width for Confidence Interval on θ with Different the λ and $\phi = 1$

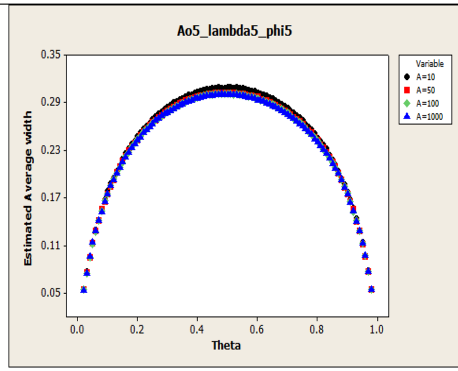
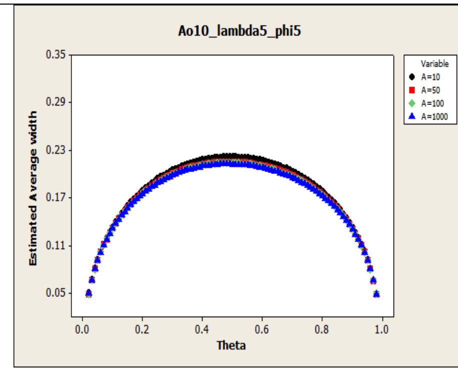
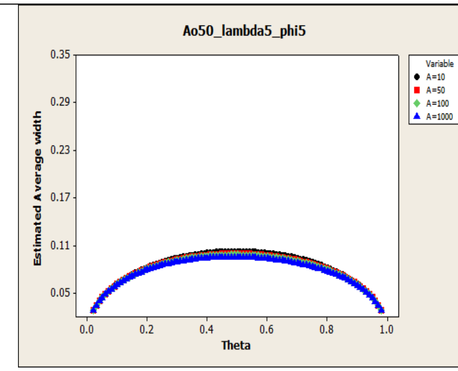
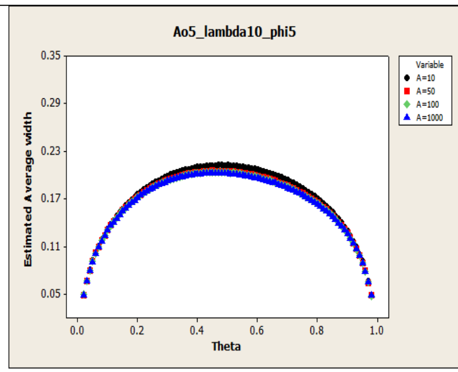
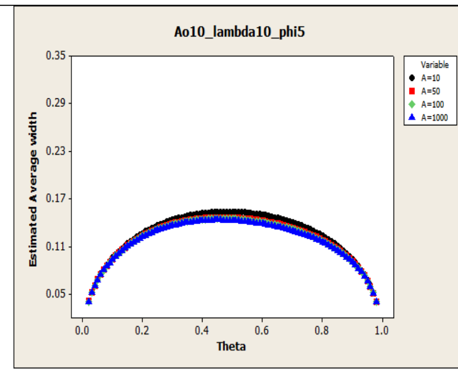
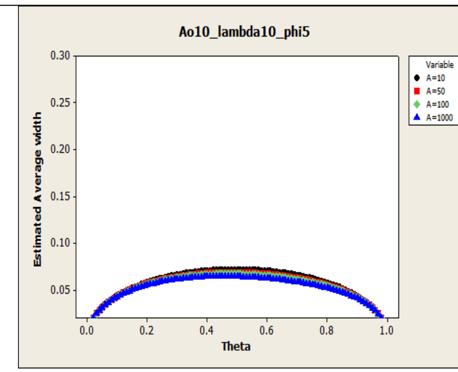
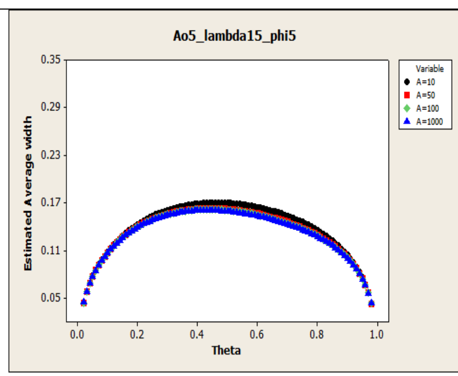
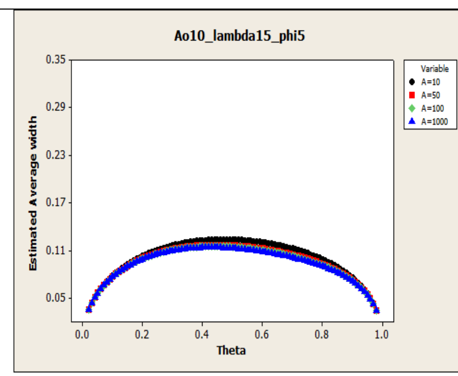
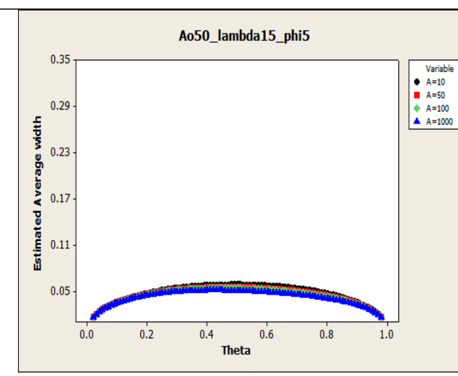
	$A_0 = 5$	$A_0 = 10$	$A_0 = 50$
$\lambda = 05$, $\phi = 5$			
$\lambda = 10$, $\phi = 5$			
$\lambda = 15$, $\phi = 5$			

Figure 6.8: Estimated Average Width for Confidence Interval on θ with Different the λ and $\phi = 5$

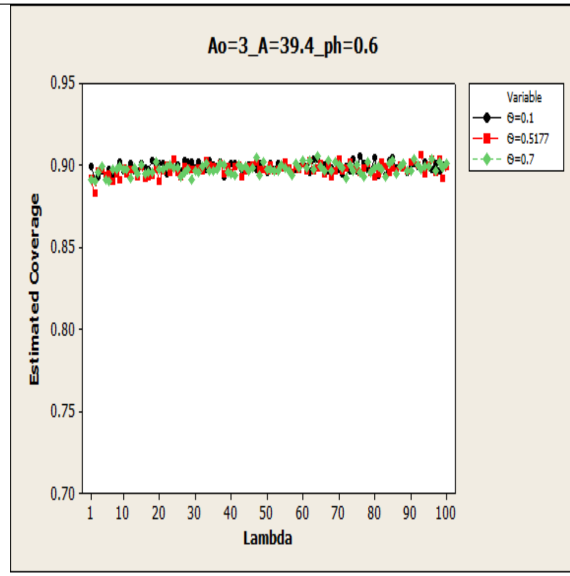
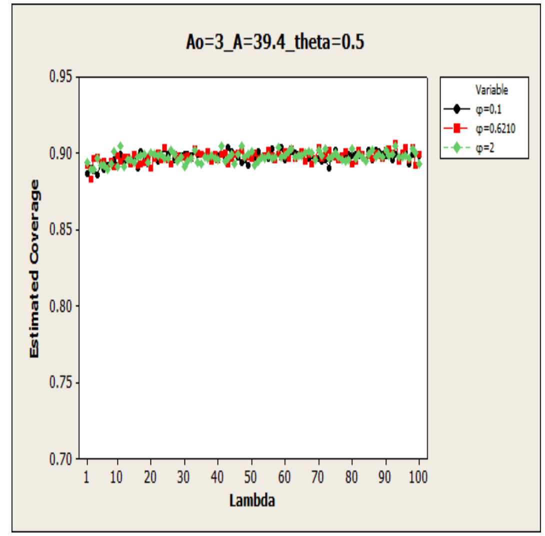
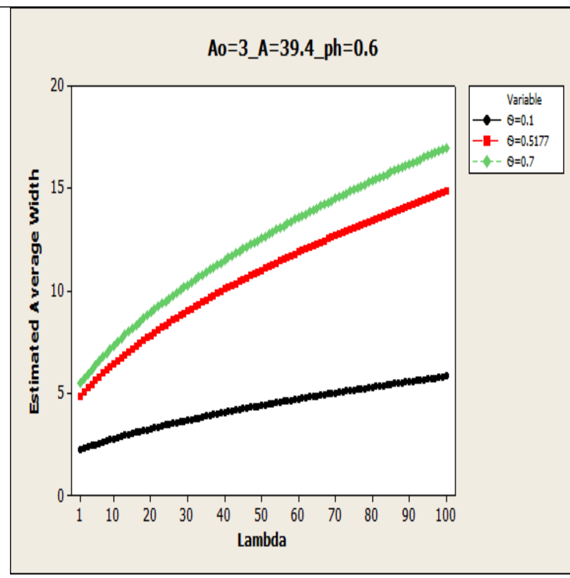
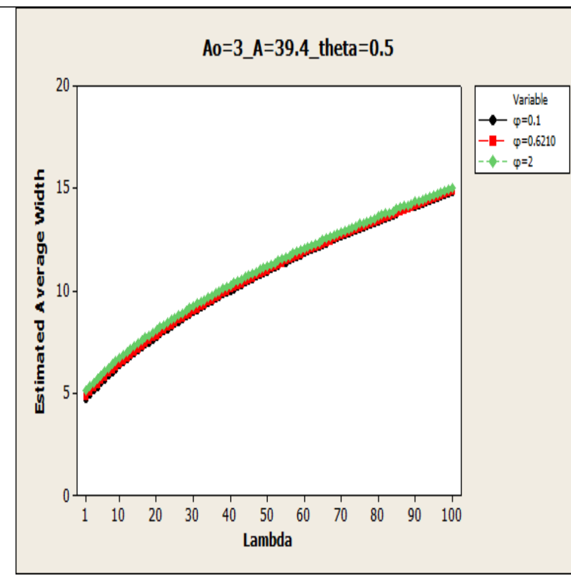
<table>
<tr><td></td><td>$A_0 = 3, \phi = 0.6$</td><td>$A_0 = 3, \theta = 0.5$</td></tr>
<tr><td>$A = 39.4$</td><td>
</td><td>
</td></tr>
<tr><td>$A = 39.4$</td><td>
</td><td>
</td></tr>
</table>

Figure 6.9: Estimated coverage and Estimated Average Width for Confidence Interval on λ for Example One

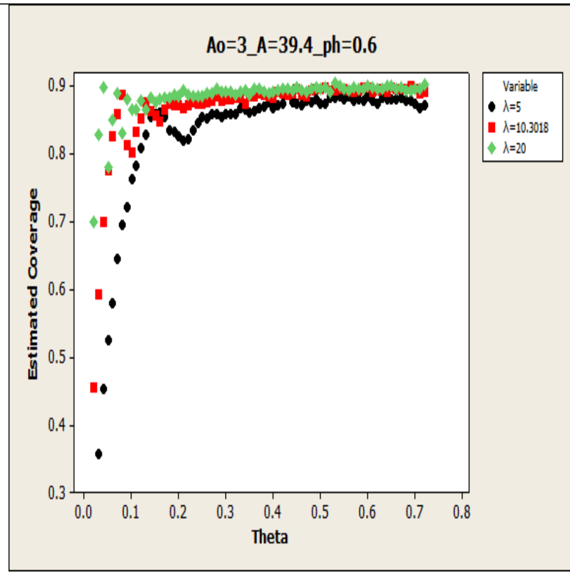
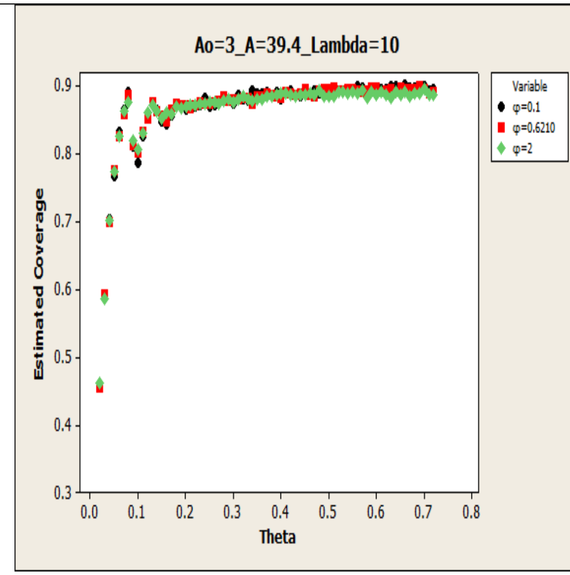
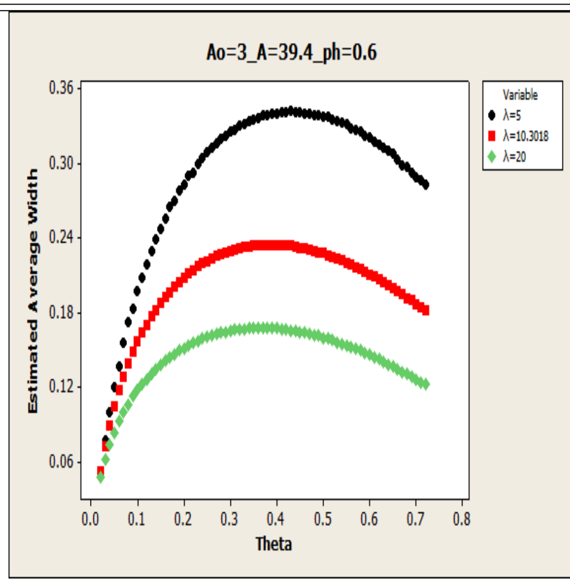
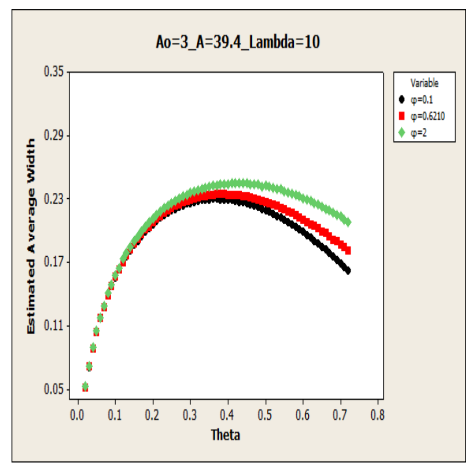
	$A_0 = 3, \phi = 0.6$	$A_0 = 3, \lambda = 10$
$A = 39.4$		
$A = 39.4$		

Figure 6.10: Estimated Coverage and Estimated Average Width for Confidence Interval on θ for Example One

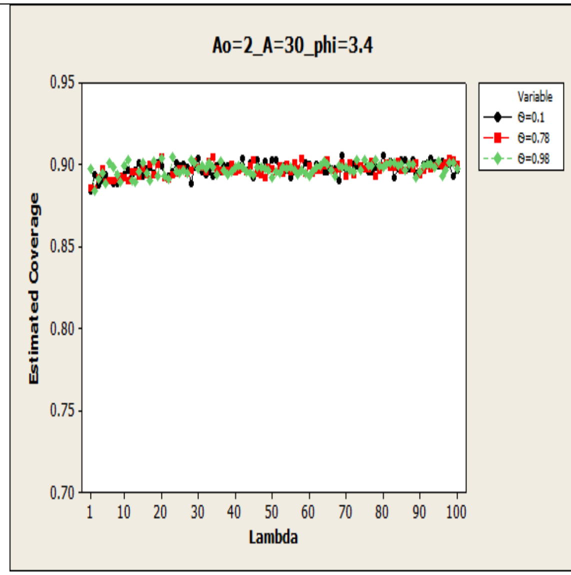
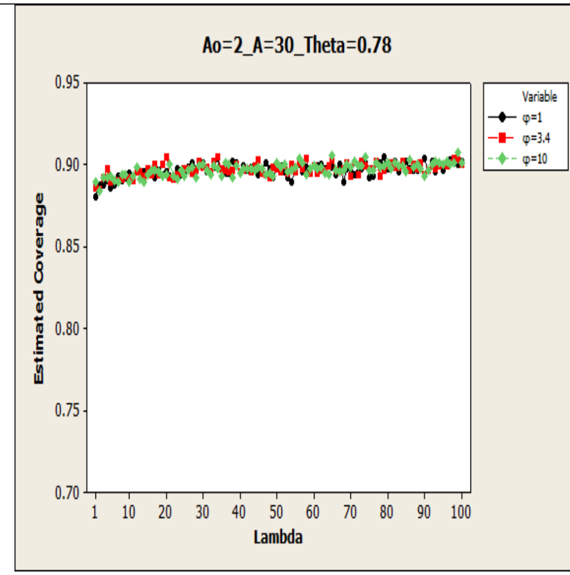
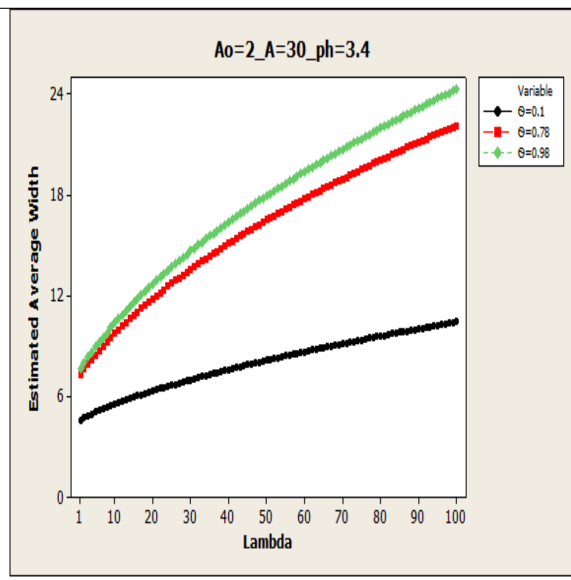
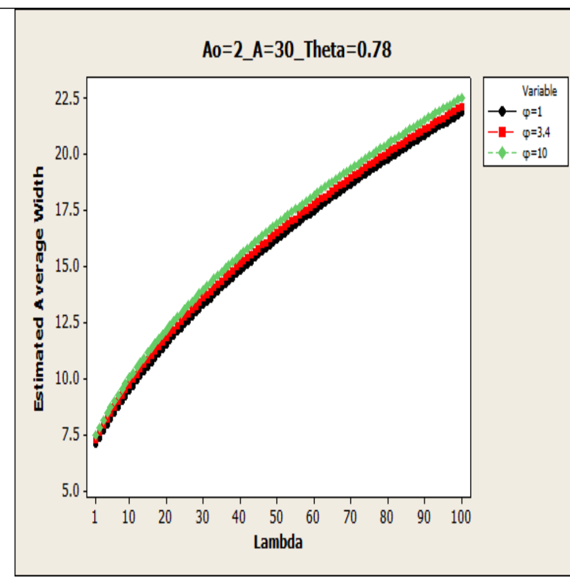
	$A_0 = 2, \phi = 3.4$	$A_0 = 2, \theta = 0.78$
$A = 30$		
$A = 30$		

Figure 6.11: Estimated Coverage and Estimated Average Width for Confidence Interval on λ for Example Two

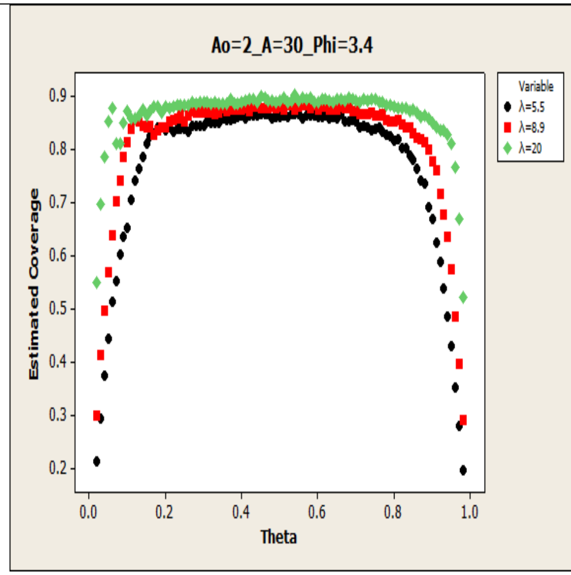
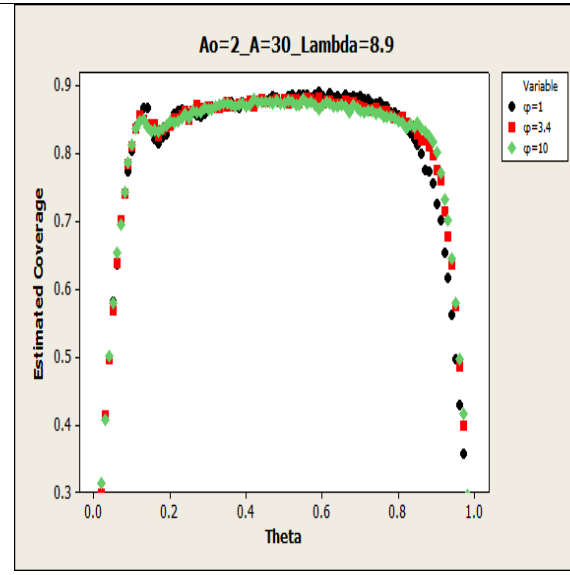
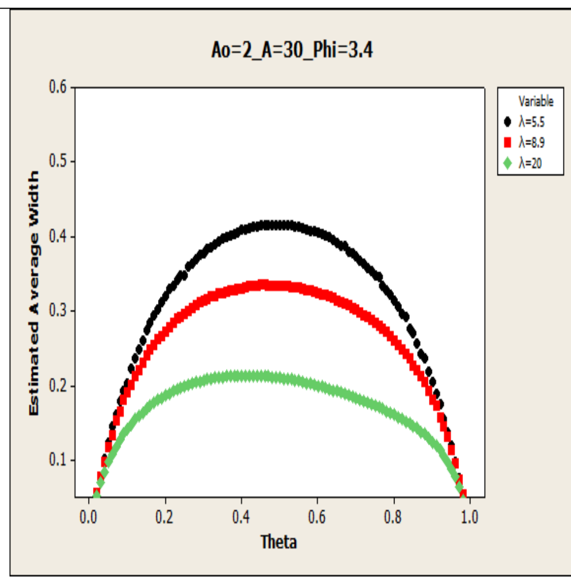
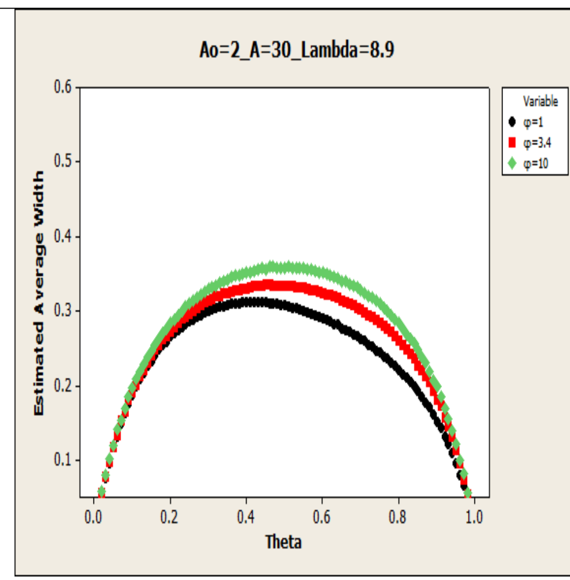
<table>
<tr><td></td><td>$A_0 = 2, \phi = 3.4$</td><td>$A_0 = 2, \lambda = 8.9$</td></tr>
<tr><td>$A = 30$</td><td>
</td><td>
</td></tr>
<tr><td>$A = 30$</td><td>
</td><td>
</td></tr>
</table>

Figure 6.12: Estimated Coverage and Estimated Average Width for Confidence Interval on θ for Example Two

VITA

Nishantha Poddiwla Hewage was born in Horana, Sri Lanka. He attended the Rajarata University of Sri Lanka at Sri Lanka from 2008-2013 and received a four year Bachelor degree in Industrial Mathematics in 2013. He worked as a Teaching assistant in the Department of Physical science and later as a full-time tutor in BSc online degree program at Rajarata University, Sri Lanka between 2013 and 2016. He began work towards a Master degree in Mathematical Sciences at Stephen F.Austin State University, Texas, the USA in the Fall of 2016 and is expected to graduate in August 2018.

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 Nishantha Janith Chandrasena Poddiwla Hewage using LaTeX.