NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

SAMPLE SIZE DETERMINATION FOR ESTIMATION OF SENSOR DETECTION PROBABILITIES BASED ON A TEST VARIABLE

by

Okan Oymak

June 2007

Thesis Advisor:	Lyn R. Whitaker
Second Reader:	Samuel E. Buttrey

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2007	Master's Thesis

4. TITLE AND SUBTITLE: Sample Size Determination for Estimation of Sensor Detection Probabilities Based on a Test Variable

5. FUNDING NUMBERS

6. AUTHOR(S) Okan Oymak

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

In this thesis, we study procedures and required sample sizes for estimating the probability of detection as a function of range to target for sensor systems as evaluated by the U.S. Army Yuma Proving Ground. First, we examine the problem within the context of a binomial experiment in order to improve the current estimation method used by the U.S. Army Yuma Proving Ground. Specifically, we evaluate the coverage probabilities and lengths of widely used confidence intervals for a binomial proportion and report the required sample sizes for some specified goals. Although the required sample sizes turn out to be impracticably large, we provide the U.S. Army Yuma Proving Ground with a better understanding of the usual confidence intervals and variability inherent in their current estimation scheme. Second, we show that confidence intervals for a probability of detection as a function of range based on the fit of a simple linear logistic regression model perform much better than the usual confidence intervals for a binomial proportion. Using an empirical approach based on a controlled set of simulations, we then determine the required sample size within the experimental region of interest.

14. SUBJECT TERMS
Sample Size, Binomial Proportion, Confidence Interval, Coverage Probability, Experimental Design, Logistic Regression

15. NUMBER OF PAGES
121

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

SAMPLE SIZE DETERMINATION FOR ESTIMATION OF SENSOR DETECTION PROBABILITIES BASED ON A TEST VARIABLE

Okan Oymak
Major, Turkish Army
B.S., Turkish Military Academy, 1991

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2007

Author: Okan Oymak

Approved by: Lyn R. Whitaker
Thesis Advisor

Samuel E. Buttrey
Second Reader

James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

In this thesis, we study procedures and required sample sizes for estimating the probability of detection as a function of range to target for sensor systems as evaluated by the U.S. Army Yuma Proving Ground. First, we examine the problem within the context of a binomial experiment in order to improve the current estimation method used by the U.S. Army Yuma Proving Ground. Specifically, we evaluate the coverage probabilities and lengths of widely used confidence intervals for a binomial proportion and report the required sample sizes for some specified goals. Although the required sample sizes turn out to be impracticably large, we provide the U.S. Army Yuma Proving Ground with a better understanding of the usual confidence intervals and variability inherent in their current estimation scheme. Second, we show that confidence intervals for a probability of detection as a function of range based on the fit of a simple linear logistic regression model perform much better than the usual confidence intervals for a binomial proportion. Using an empirical approach based on a controlled set of simulations, we then determine the required sample size within the experimental region of interest.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
A. BACKGROUND 1
B. DOSE-RESPONSE PROBLEMS 3
C. OBJECTIVE OF THE STUDY 4
D. ORGANIZATION OF THE STUDY 5

II. LITERATURE REVIEW 7
A. CONFIDENCE INTERVAL METHODS FOR THE BINOMIAL PROPORTION 7
1. The Wald Confidence Interval 7
2. The Wilson Score Confidence Interval 13
3. The Adjusted Wald (Agresti-Coull) Confidence Interval 15
4. The Clopper-Pearson Confidence Interval 20
5. The Jeffreys Prior Interval 23
B. SAMPLE SIZE CALCULATION FOR THE BINOMIAL PROPORTION 25
C. OVERVIEW OF THE LINEAR LOGISTIC REGRESSION MODEL 29
1. Definition 29
2. Interval Estimate for the Binary Response Probability 30
3. Precision of the Estimated Binary Response Probabilities Based on the Fit of a Logistic Regression Model 31

III. SAMPLE PROPORTION-BASED ANALYSIS 33
A. INTRODUCTION 33
B. ASSUMPTIONS 35
C. ANALYSIS THROUGH SIMULATION 35
D. RESULTS OF CALIBRATING THE CONFIDENCE INTERVALS UNDER THE BINNING APPROACH 41
E. CHAPTER SUMMARY 44

IV. LOGISTIC REGRESSION-BASED ANALYSIS 47
A. INTRODUCTION 47
B. LOGISTIC REGRESSION MODEL-BASED ESTIMATORS 48
C. COVERAGE PERFORMANCE OF LOGISTIC REGRESSION MODEL-BASED CONFIDENCE INTERVALS 50
D. MATHEMATICAL MODELS USED IN SIMULATIONS 53
E. ANSWERING THE SAMPLE SIZE QUESTION THROUGH SIMULATION 55
F. COMPARING THE COVERAGE PERFORMANCE OF LARGE-SAMPLE AND NONPARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS 63

G. CHAPTER SUMMARY 65
V. CONCLUSION 67
A. CONCLUDING REMARKS 67
B. FURTHER STUDY SUGGESTIONS 68
APPENDIX A. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE WALD INTERVAL 71
APPENDIX B. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE WILSON INTERVAL 75
APPENDIX C. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE ADJUSTED WALD INTERVAL 79
APPENDIX D. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE CLOPPER-PEARSON INTERVAL 83
APPENDIX E. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE EQUAL-TAILED JEFFREYS PRIOR INTERVAL 87
APPENDIX F. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES OF CONFIDENCE INTERVALS FOR PROBABILITIES BASED ON THE FIT OF A SIMPLE LINEAR LOGISTIC REGRESSION MODEL 91
APPENDIX G. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES OF Bca CONFIDENCE INTERVALS FOR PROBABILITIES BASED ON THE FIT OF A SIMPLE LINEAR LOGISTIC REGRESSION MODEL 93
LIST OF REFERENCES 95
INITIAL DISTRIBUTION LIST 97

LIST OF FIGURES

Figure 1. Coverage Probability for the 95% Wald Confidence Interval; Oscillation Phenomenon for Fixed $p = 0.2$ and Variable $n = 25$ to 100 (From: Brown et al., 2001) 9

Figure 2. Coverage Probabilities for the 95% Wald Confidence Interval (a) $p = 0.25$, (b) $p = 0.05$ (From: Henderson & Meyer, 2001).....10

Figure 3. Standard Interval; Oscillation Phenomenon for Fixed $n = 100$ and variable p (From: Brown et al., 2001).....11

Figure 4. Coverage Probabilities for the Nominal 95% Standard interval (After: Agresti & Coull, 1998).....12

Figure 5. Coverage of the Nominal 99% Wald Interval for fixed $n = 20$ and Variable p (From: Brown et al., 2001).....13

Figure 6. Coverage Probabilities for the 95% Wilson Interval (a) $p = 0.25$, (b) $p = 0.05$ (From: Henderson & Meyer, 2001).....14

Figure 7. Plot of Coverage Probabilities for the Nominal 95% Confidence Intervals for Binomial Proportion p when $n = 25$ (From: Agresti, 2002) .15

Figure 8. Coverage Probabilities for the 95% Wilson Interval when $n = 50$ (From: Brown et al., 2001).....15

Figure 9. Coverage Probabilities for the Binomial Proportion p with Nominal 95% and 99% Wald Confidence Intervals and the Adjusted Interval Based on Adding Four Pseudo Observations, for $n=5$, 10, and 20 (From: Agresti & Caffo, 2000)...17

Figure 10. Mean Coverage Probability as a Function of Sample Size for the Nominal 95% Wald (W) and Adjusted Wald (A) Intervals, When p has (a) a Uniform (0,1) Distribution and (b) a Beta Distribution with $\mu = 0.10$ and $\sigma = 0.05$ (From: Agresti & Coull, 1998).....18

Figure 11. Mean Coverage Probability as a Function of Sample Size for the Nominal 95% Exact (E), Wilson (S), and standard (W) Intervals, When p has (a) a Uniform (0,1) Distribution and (b) a Beta Distribution with $\mu = 0.10$ and $\sigma = 0.05$ (From: Agresti & Coull, 1998).....19

Figure 12. Comparison of the Average Coverage Probabilities (From: Brown et al., 2001).....19
Figure 13. Coverage Probabilities for the Nominal 95% Adjusted Wald and Clopper-Pearson Intervals as a Function of p (After: Agresti & Coull, 1998)..22
Figure 14. Coverage Probabilities for the 95% Clopper-Pearson Interval (a) $p = 0.25$, (b) $p = 0.05$ (From: Henderson & Meyer, 2001).....23
Figure 15. Coverage Probabilities for the 95% Jeffreys Prior Interval, when $n = 50$ (From: Brown et al., 2001).....24
Figure 16. Coverage Probabilities for the 95% Confidence Intervals when $n = 5$.....36
Figure 17. Mean Lengths for the 95% Confidence Intervals when $n = 5$.....38
Figure 18. Function Written in the S-PLUS Language Used to Compute the Equal-tailed Jeffreys Prior Interval Endpoints.....40
Figure 19. 95% Confidence Intervals and Coverage Probabilities for the Wilson Interval Before and After the Calibration.....42
Figure 20. 95% Confidence Intervals and Coverage Probabilities for the Agresti-Coull Interval Before and After the Calibration.....42
Figure 21. 95% Confidence Intervals and Coverage Probabilities for the Clopper-Pearson Interval Before and After the Calibration.....42
Figure 22. 95% Confidence Intervals and Coverage Probabilities for the Equal-Tailed Jeffreys Prior Interval Before and After Calibration.....43
Figure 23. The Effect of Calibration on the Lengths of Confidence Intervals for Each Method.....43
Figure 24. S-PLUS Output for the Logistic Regression Model with Sample Data from Table 5.....49
Figure 25. Coverage Probabilities and Mean Lengths of the 95% Confidence Interval for the Estimated Response as a Function of p for Different Dose Levels with One Observation at each Dose Level..52
Figure 26. The Effect of Doubling the Observations When Dose Level is 51.....53
Figure 27. Coverage Probabilities for the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the First Experimental Design.....57
Figure 28. Mean Length of the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic

Regression Model Under the First Experimental Design.....57
Figure 29. Coverage Probabilities for the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Second Experimental Design.....59
Figure 30. Mean Length of the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Second Experimental Design.....60
Figure 31. Coverage Probabilities for the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Third Experimental Design.....61
Figure 32. Mean Length of the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Third Experimental Design.....61
Figure 33. Coverage Probabilities of the 95% Large Sample and Bca Confidence Intervals Based on the Fit of a Simple Linear Logistic Regression Model When $n = 66$ 64

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1. Standard Interval; Lucky n and Unlucky n for $10 \leq n \leq 50$ and $p = 0.5$ (From: Brown et al., 2001).....11

Table 2. Sample Sizes for Various Values of CI Width, Using Different Approaches when $1 - \alpha = 0.95$ 28

Table 3. Mean Coverage Probabilities of Nominal 95% Confidence Intervals and Root MSEs.....37

Table 4. Mean Coverage Probabilities of the Nominal 95% Confidence Intervals and Root MSEs (Before and After Calibration).....44

Table 5. Sample Data, Where Five Observations are Recorded at each Dose Level.....48

Table 6. Numerical Results Indicating the Effect of Reducing the Number of Observations on the Coverage Performance of the 95% Large Sample Confidence Interval.....51

Table 7. Simulation Results for Model 1 Under the First Experimental Design.....56

Table 8. Simulation Results for Model 1 Under the Second Experimental Design.....59

Table 9. Simulation Results for Model 1 Under the Third Experimental Design.....60

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I am grateful to my beloved country for having given me such an outstanding and invaluable opportunity to study at the Naval Postgraduate School.

I would like to thank Prof. Whitaker and Prof. Buttrey for their continuous guidance and support throughout my research.

I would also like to thank my wife ***Meltem*** and my daughters ***Cansu*** and ***Nehir*** for their love and continual support from thousands of miles away. Without their patience and understanding, I would not have had the perseverance to complete this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Careful planning plays an important role in obtaining practically relevant and statistically valid information from any study. An essential part of this procedure is to determine how large a sample should be relative to the goals of the study, and for studies that are more complex, how observations should be sampled. Too few observations might hamper a study's ability to detect important effects, whereas too many observations increase the cost of the study and can lead to effects that are statistically significant and yet practically inconsequential.

This thesis focuses on experimental design issues with an emphasis on sample size determination for estimating the probability of detection at various ranges for sensor systems whose developmental tests and evaluations are conducted at the U.S. Army Yuma Proving Ground.

We approach the problem of sample size determination for estimation of sensor detection probabilities from two different aspects. First, we examine the problem within the context of a binomial experiment in order to improve the current estimation method used by the U.S. Army Yuma Proving Ground that considers only straight proportions within range intervals (*binning approach*). Using simulation, we evaluate the coverage probabilities and lengths of confidence intervals for binomial proportions and report the required sample sizes for some specified goals utilizing different methods. Second, and again using simulation, we evaluate the coverage probabilities and

lengths of confidence intervals based on logistic regression to get better estimates of the probability of detection with much smaller sample sizes.

The usual confidence interval methods for a binomial proportion that are examined in detail in this thesis are as follows:

- The Wald (Standard Approximate) interval
- The Wilson (Score) interval
- The Adjusted Wald (Agresti-Coull) interval
- The Clopper-Pearson (Exact) interval
- The equal-tailed Jeffreys prior interval

These are just several of the methods that can be used in constructing confidence intervals for the probability of detection p based on observing X number of detections out of n independent trials each with the same probability of detection. These procedures are approximate in the sense that their nominal coverage probability is not the same as their actual coverage probability (the probability that the interval contains the true parameter). Of the confidence intervals reviewed in this thesis, the coverage probabilities of the Wald interval can be significantly less than the nominal confidence level not just for cases when the true (but unknown) probability is near $[0,1]$ boundary but throughout the unit interval. On the other hand, actual coverage of the Clopper-Pearson "exact" intervals is often higher than the intended confidence level. This "exact" procedure is conservative in the sense that it never yields intervals with coverage lower than intended. The remaining three interval methods, namely the Wilson, the Agresti-Coull, and the equal-tailed Jeffreys

prior intervals, turn out to be comparable in terms of their coverage performances and are presented as recommended intervals (e.g., Brown, Cai, and DasGupta, 2001; Henderson and Meyer, 2001; and Agresti and Coull, 1998).

When the design of the experiment to estimate sensor detection probabilities is based on the ***binning approach***, where detections at ranges in a given interval are pooled, our simulation results show that the performance of the Wilson, the Agresti-Coull, and the equal-tailed Jeffreys prior intervals is comparable to the performance based on a binomial experiment. Hence, either of the three can be used depending on preference. However, there are two major drawbacks of the ***binning approach***. The first one is that very large sample sizes are needed to get confidence intervals of reasonable length, and the second one is the lack of ability to estimate the sensor detection probabilities at a specified range.

In our second approach to the problem, our analyses show that by using a parametric model, the U.S. Army Yuma Proving Ground engineers can get much more information out of their samples for the same sample sizes which they currently have. This parametric approach capitalizes on the fact that the probability of detection is a function of range. By analyzing different data sets, we find that an appropriate model for probability of detection as a function of range seems to be a piecewise linear logistic regression model. Furthermore, estimation of the probabilities of detection at various ranges should focus on the middle piece, where the probabilities do not remain

constant. Our simulations based on three different experimental designs[1] show that large-sample confidence intervals for probabilities of detection at various ranges based on the fit of a simple linear logistic regression model perform as well as much more complicated models in terms of their coverage probabilities. Moreover, we find that the use of a logistic regression model reduces the length of the confidence intervals by a considerable amount. The results of our simulations in each of which the sample size varies within the experimental region of interest suggest the following:

- When the model approximates the true probabilities decently, logistic regression model-based estimators are more precise than the sample proportion-based estimators are.
- As the sample size increases within the experimental region of interest, the coverage probabilities of large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model tend to come closer to the nominal confidence level.
- From a practical point of view, experimental design changes that change which ranges are sampled do not have a considerable effect on the coverage probabilities of confidence intervals for a probability based on the fit of a simple linear logistic regression model.
- Large-sample and bootstrap Bca (Bias corrected and accelerated) confidence intervals for a probability based on the fit of a simple linear logistic regression model are competitive in terms of their coverage probabilities.

Based on the findings through our analyses, our recommendations for the U.S. Army Yuma Proving Ground and some important conclusions reached are as follows:

[1] See Section E of Chapter IV for a detailed description of experimental designs.

- First and foremost, when the probability of detection at specified range intervals is estimated using the current ***binning approach***, we recommend that the U.S. Army Yuma Proving Ground engineers consider not only the sample proportions but also the confidence intervals for a binomial proportion. Even though the use of this approach provides estimates for range intervals rather than specific ranges and violates the equal probability of success assumption for each trial in a binomial experiment, our simulations show that the recommended confidence intervals, namely the Agresti-Coull, Wilson, and equal-tailed Jeffreys prior intervals, perform well.

- Second, the U.S. Army Yuma Proving Ground engineers can use a logistic regression model so that they can get much more information out of their samples for the same sample sizes. When this procedure is adopted, estimation of sensor detection probabilities should focus on ranges where the probabilities do not remain constant. Our simulations show that large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model perform much better than the usual confidence intervals for a binomial proportion in terms of their coverage probabilities and lengths.

- Finally, in order to obtain good estimates of sensor detection probabilities at a significance level of 0.05, we recommend that the U.S. Army Yuma Proving Ground engineers use a simple linear logistic regression model and obtain at least 100 observations within the experimental region of interest where the probabilities do not remain constant. In the other two regions where the probabilities remain almost constant, we assess that the current ***binning approach*** that has been taken by the U.S. Army Yuma Proving Ground is appropriate as long as the issues associated with the usual confidence intervals for the binomial proportion are kept in mind.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

Careful planning plays an important role in obtaining practically relevant and statistically valid information from any study. An essential part of this procedure is to determine how large a sample should be relative to the goals of the study, and for studies that are more complex, how observations should be sampled. Too few observations might hamper a study's ability to detect important effects, whereas too many observations increase the cost of the study and can lead to effects that are statistically significant and yet practically inconsequential. This thesis focuses on experimental design issues with an emphasis on sample size determination for estimating the probability of detection at various ranges for sensor systems whose developmental tests and evaluations are conducted by the U.S. Army Yuma Proving Ground.

The U.S. Army Yuma Proving Ground is one of the largest military installations in the world, situated in southwestern Arizona, approximately 24 miles north of the city of Yuma, Arizona. The Proving Ground is used for testing military equipment and encompasses 1,300 square miles (3,367 square kilometers) in the Sonoran Desert ("Yuma Proving Ground," n.d.)

Of the four extreme natural environments recognized as critical in testing military equipment, three are found at the Yuma Proving Ground – desert, cold, and tropic environments. Yuma Test Center capabilities include:

- Ground weapon systems tests
- Helicopter armament and target acquisition systems tests
- Artillery and tank munitions tests
- Cargo and personnel parachutes tests
- Mines and mine-removal systems tests
- Tests of tracked and wheeled vehicles in a desert environment
- Vibration-free, interface-free tests of smart weapon systems (The U.S. Army Yuma Proving Ground, 2006)

For this thesis, we focus on tests designed to estimate sensor detection probabilities at predetermined ranges as an aircraft approaches a target. Because there are always some budgetary constraints that limit the number of test hours available, sample size determination is an important issue. On the other hand, to get good estimates of the probability of detection requires not only a sample of sufficient size but also a method of estimating the probability of detection at different ranges that takes full advantage of all the information available in the sample.

Currently, the experimenters at the U.S. Army Yuma Proving Ground use the small sample proportion of observed detections taken at approximately five different yet similar ranges to the target to estimate the sensor detection probabilities. In essence, they are treating these sensor tests as a sequence of binomial experiments. Experiments that conform either exactly or approximately to the following list of requirements are called binomial experiments (Devore, 2004, p. 120):

- The experiment consists of a sequence of n trials, where n is fixed in advance of the experiment.
- Each trial has exactly two possible outcomes, which we denote by success or failure.
- The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial.
- The probability of each outcome remains the same for each trial.

Because these estimated probabilities are based on such small samples, it becomes important to provide with the experimental results standard errors of these estimates or confidence intervals for the probabilities of detection. There are a number of well-known small sample confidence interval procedures for binomial proportions. These are presented in this thesis, and their properties are studied in the context of the U.S. Army Yuma Proving Ground sensor detection tests.

B. DOSE-RESPONSE PROBLEMS

The problem of estimating the probability of detection as a function of range is equivalent to a large class of problems found in the medical sciences called dose-response problems. There are many situations where clinical experiments tend to yield discrete data. Dose-response experiments are one good example where the responses are binary in most cases (Khuri, Mukherjee, Sinha, & Ghosh, 2006). In dose-response experimental designs, subjects are given varying doses of a drug or medication with the intent of estimating the probability of a specific response to the drug as a function of the dose. Here, the dose level is analogous to the distance to the target, and the probability of response to the drug is analogous to the

probability of detection. There is a large body of literature concerning the analysis of dose-response data. According to Khuri et al. (2006), generalized linear models (GLMs) are appropriate for such data. GLMs are a unified class of regression models for discrete and continuous response variables and have been used routinely in dealing with observational studies. In this regard, logistic regression for binary responses is a special case of GLMs that can be used for estimation of sensor detection probabilities as a function of range and can be a tool to determine the sample size required for getting good interval estimates for the binary response probability. By good estimates we mean that the probability that the interval contains the true parameter (coverage probability) is close to the nominal confidence level at which the interval is constructed.

C. OBJECTIVE OF THE STUDY

The objective of this study is to not only provide insight on how experimental designs can be set up to get good and reliable estimates of sensor detection probabilities, but also to propose a new methodology for getting these estimates. The questions that this thesis seeks to address are as follows:

- Within the context of a binomial experiment, what are the existing confidence interval (CI) methods for the binomial proportion and how do they compare to each other in terms of their coverage probabilities?
- What are the approaches to sample size determination for the binomial proportion?
- How does the precision of an estimated binary response probability based on the fit of a simple linear logistic regression model compare to that of a binomial proportion?

- Based on the findings to the above questions, how many observations are needed at each predetermined range to get good estimates of sensor detection probabilities as an aircraft approaches a target?

D. ORGANIZATION OF THE STUDY

The study includes five chapters. Chapter II presents a literature review of widely used confidence interval methods, approaches to sample size sample determination for the binomial proportion, and the linear logistic regression models. Chapter III uses simulation to analyze the performance of confidence intervals for binomial proportions in terms of their coverage probabilities and lengths within the context of the U.S. Army Yuma Proving Ground experiments. Chapter IV examines the coverage probabilities of confidence intervals based on the fit of a simple linear logistic regression model and presents the results of an empirical approach based on simulation for varying sample sizes and experimental designs. Based on the evidence gathered in Chapter III and IV, Chapter V includes a summary of the study as well as conclusions and recommendations for further study.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

A. CONFIDENCE INTERVAL METHODS FOR THE BINOMIAL PROPORTION

In experiments designed to estimate a binomial proportion p, sample sizes are often computed to ensure that the point estimate $\hat{p}$ will be within a specified distance from the true value with sufficiently high probability (Rahme & Joseph, 1998). Because the sample size needed to estimate a binomial proportion p is closely related to the construction of confidence intervals, this section gives five methods of constructing confidence intervals for the probability of detection p based on observing X number of detections out of n independent trials, each with the same probability of detection. Moreover, to get an idea of how well each of these methods performs, this section compares these methods in terms of their coverage probabilities for varying values of a binomial proportion p and varying sample sizes. The next section continues with an overview of an important problem, namely sample size determination.

1. The Wald Confidence Interval

The Wald confidence interval, also called *the standard approximate confidence interval*, is the one presented in almost all of the introductory statistical textbooks (e.g., Larsen & Marx, 1986; Collett, 1991; Devore, 2004).

The $100(1-\alpha)\%$ Wald confidence interval for a population proportion p is based on a central limit theorem result, which states that

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

is asymptotically standard normal. Therefore,

$$\Pr\left(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

where z_a is the $1\text{-}a$ quantile of the standard normal density, or the value for which the right tail area is a. From this, plugging in $\hat{p}$ for p in the denominator and solving the inequalities for p, the standard approximate confidence interval takes the form:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ (Henderson & Meyer, 2001, p. 338)

According to Brown, Cai, and DasGupta (2001),

> Most students and users no doubt believe that the larger the number n, the better the normal approximation, and thus the closer the actual coverage would be to the nominal level 1-α. Further, they believe that the coverage probabilities of this method are close to the nominal value, except possibly when n is "small" or p is "near" [zero] or [one]. (p. 103)

Brown et al. (2001) point out an interesting phenomenon for the Wald interval. That is, the actual coverage probability of the confidence interval contains non-negligible oscillation as both p and n vary. They present some "lucky" pairs (p, n) such that the actual coverage probability $C(p, n)$ is very close to or larger than the nominal level. On the other hand, they also show the

existence of some "unlucky" pairs (p, n) such that the corresponding $C(p, n)$ is much smaller than the nominal level.

The following examples reveal the drastic changes in coverage that occur in nearby p for fixed n, and in nearby n for fixed p.

It is clear from Figure 1 that the oscillation is significant and the coverage probability does not steadily get closer to the nominal confidence level of 95% as n increases. For instance, $C(0.2, 30) = 0.946$ and $C(0.2, 98) = 0.928$. As can easily be seen, the coverage probability is significantly closer to 0.95 when $n = 30$ than when $n = 98$. From this example, it is obvious that the true coverage probability behaves contrary to conventional wisdom in a very significant way (Brown et al., 2001).

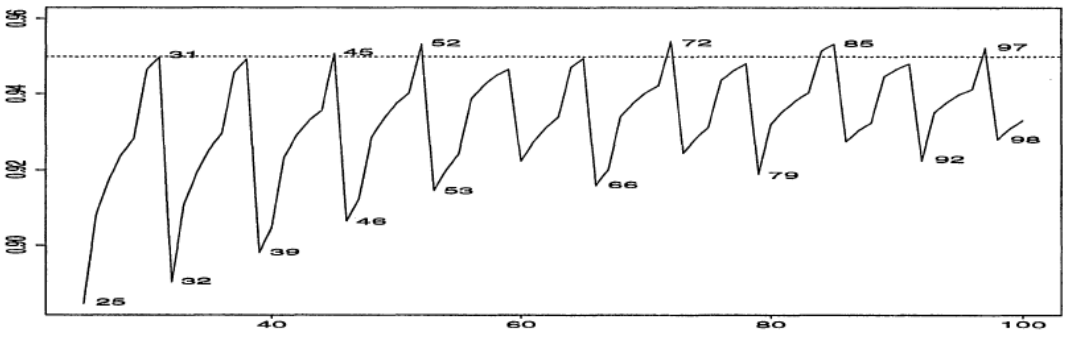

Figure 1. Coverage Probability for the 95% Wald Confidence Interval; Oscillation Phenomenon for Fixed $p = 0.2$ and Variable $n = 25$ to 100 (From: Brown et al., 2001)

In order to see how the 95% Wald or "standard" confidence interval performs under a variety of conditions, Henderson and Meyer (2001) obtained the coverage probabilities as a function of sample size (see Figure 2).

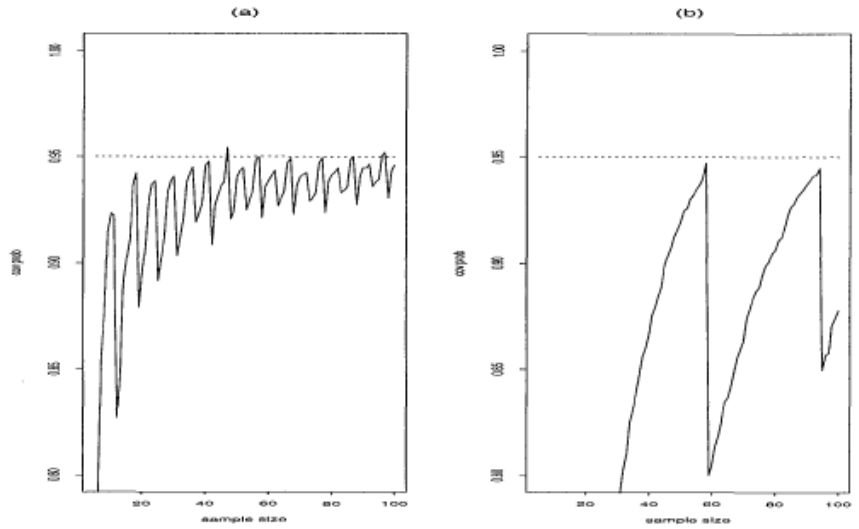

Figure 2. Coverage Probabilities for the 95% Wald Confidence Interval (a) $p = 0.25$, (b) $p = 0.05$ (From: Henderson & Meyer, 2001)

> In Figure [2(a)], p is fixed at 0.25, and coverage probabilities are calculated for each sample size $n = 5$ through $n = 100$. The horizontal line at 0.95 shows the target coverage probability. For some n, the coverage probabilities are near 0.95, but for most, the coverage probabilities are smaller. For p fixed at 0.05, the coverage probabilities, shown in Figure [2(b)], are considerably too small for most n. (Henderson & Meyer, 2001, p. 338)

As part of their study to illustrate the inconsistency, unpredictability, and poor performance of the standard interval Brown et al. (2001) considered the case of $p = 0.5$ and evaluated the actual coverage probability of the 95% Wald interval for $10 \leq n \leq 50$. Table 1 lists the values of "lucky" n (defined as $C(p, n) \geq 0.95$)

and the values of "unlucky" n (defined for specificity as $C(p,n) \leq 0.92$). When $n = 17$, the coverage probability is 0.951, but it equals 0.904 when $n = 18$. Although $p = 0.5$, the coverage is still 0.919 at $n = 40$.

Lucky n	17	20	25	30	35	37	42	44	49
$C(0.5, n)$	0.951	0.959	0.957	.957	0.959	0.953	0.956	0.951	0.956
Unlucky n	10	12	13	15	18	23	28	33	40
$C(0.5, n)$	0.891	0.854	0.908	0.882	0.904	0.907	0.913	0.920	0.919

Table 1. Standard Interval; Lucky n and Unlucky n for $10 \leq n \leq 50$ and $p = 0.5$ (From: Brown et al., 2001)

The following are other examples that display further instances of the inadequacy of the standard interval.

Figure 3 plots the coverage probability of the nominal 95% Wald interval as a function of p when $n = 100$. As shown in Figure 3, despite the large sample size, a significant change in coverage probability is observed in nearby p. The magnitude of oscillation increases significantly as p moves toward zero or one. The general trend of this plot is noticeably below the nominal confidence level of 0.95 except for values of p quite near 0.5 (Brown et al., 2001).

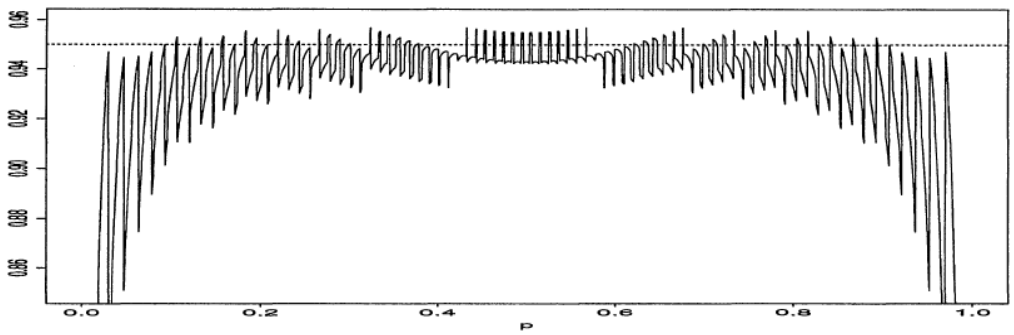

Figure 3. Standard Interval; Oscillation Phenomenon for Fixed $n = 100$ and variable p (From: Brown et al., 2001)

In a study which compares the Wald interval to two other intervals, Agresti and Coull (1998) consider the nominal 95% case and show the erratic and poor behavior of the Wald interval's coverage probability for small n, even when p is not near the boundaries (see Figure 4).

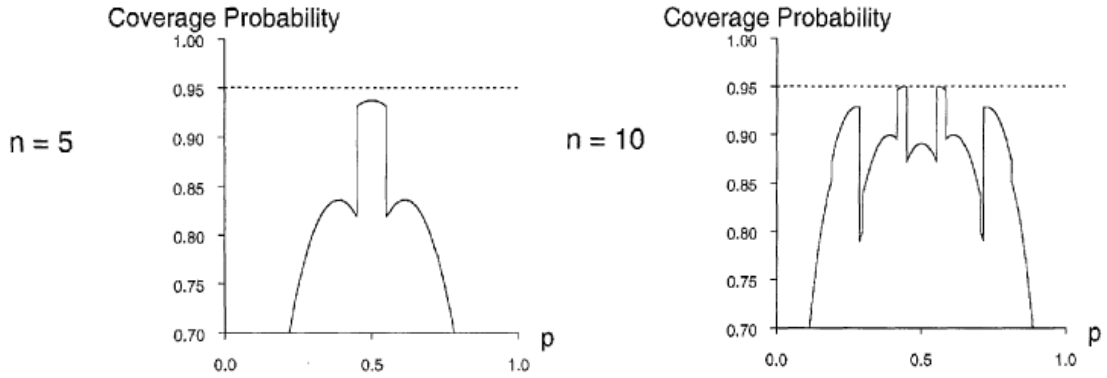

Figure 4. Coverage Probabilities for the Nominal 95% Standard interval (After: Agresti & Coull, 1998)

Another striking fact also shown by Brown et al. (2001) is illustrated in Figure 5, which is a plot of the coverage probability of the nominal 99% Wald interval with $n = 20$ and p from 0 to 1. Besides the oscillation phenomenon similar to the one in Figure 3, it is striking that in this case the coverage probability never reaches the nominal confidence level. As can be seen from Figure 5, the coverage probability is always below 0.99. Brown et al. (2001) report the coverage probability as 0.883 on average. Moreover, their evaluations show that for all $n \leq 45$, the coverage of the 99% Wald interval is strictly smaller than the nominal confidence level for all $0 \leq p \leq 1$.

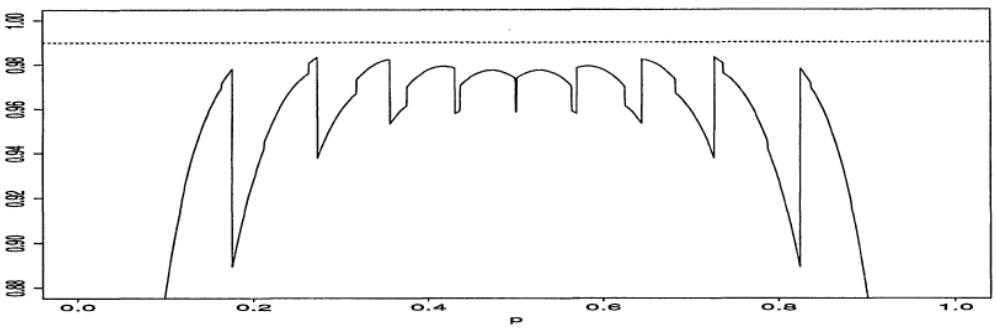

Figure 5. Coverage of the Nominal 99% Wald Interval for fixed n = 20 and Variable p (From: Brown et al., 2001)

From the evaluations reviewed so far, it seems clear that the Wald interval behaves poorly and erratically in terms of its coverage probability, and hence is too risky. Regarding the use of the Wald interval, Newcombe (1998) also strongly recommends that intervals calculated by this method no longer be acceptable for scientific literature (p. 868).

2. The Wilson Score Confidence Interval

This confidence interval, first discussed by Edwin B. Wilson in 1927, is based on inverting the large sample test of the null hypothesis $H_0 : p = p_0$ against the two-sided alternative hypothesis $H_a : p \neq p_0$. Here, the test statistic $(\hat{p} - p_0)/\sqrt{p_0(1 - p_0)/n}$ is approximately normal when H_0 is true. The Wilson interval is the set of p_0 values for which $|\hat{p} - p_0|/\sqrt{p_0(1 - p_0)/n} < z_{\alpha/2}$ (i.e., the set of values for which $H_0 : p = p_0$ is not rejected). This gives an interval of the form

$$\left(\hat{p} + \frac{z^2_{\alpha/2}}{2n} \pm z^2_{\alpha/2}\sqrt{\left[\hat{p}(1 - \hat{p}) + z^2_{\alpha/2}/4n\right]/n}\right)\Big/\left(1 + z^2_{\alpha/2}/n\right) \qquad (1)$$

(Agresti & Coull, 1998, p. 119-120).

Further evaluations by different researchers show how much better the Wilson interval performs in terms of its coverage probability.

The plots in Figure 6 by Henderson and Meyer (2001) illustrate the coverage probabilities of the 95% Wilson interval as a function of sample size. When compared with the plots in Figure 2, it is obvious that the Wilson interval gives coverage probabilities closer to the nominal confidence level.

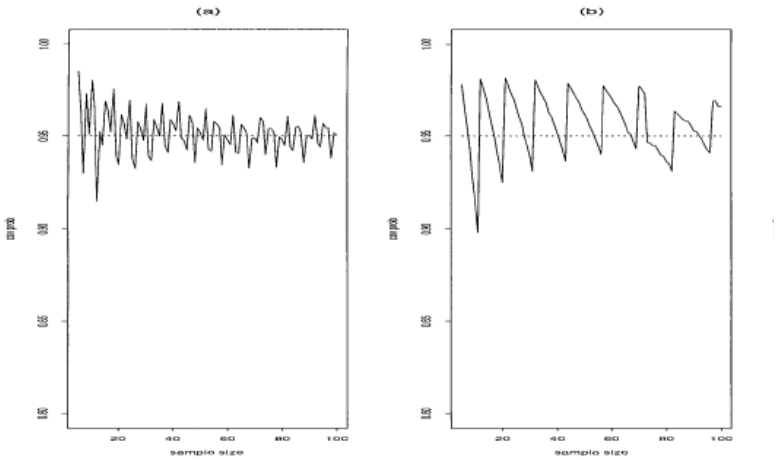

Figure 6. Coverage Probabilities for the 95% Wilson Interval (a) $p = 0.25$, (b) $p = 0.05$ (From: Henderson & Meyer, 2001)

In a similar study in which the coverage probabilities are plotted as a function of a binomial proportion p for the nominal 95% confidence intervals (see Figure 7), Agresti (2002) states the following:

> The score method behaves well, except for some p values close to zero or one. Its coverage probabilities tend to be near the nominal level, not being consistently conservative or liberal. This is a good method unless p is very close to zero or one. (p. 19)

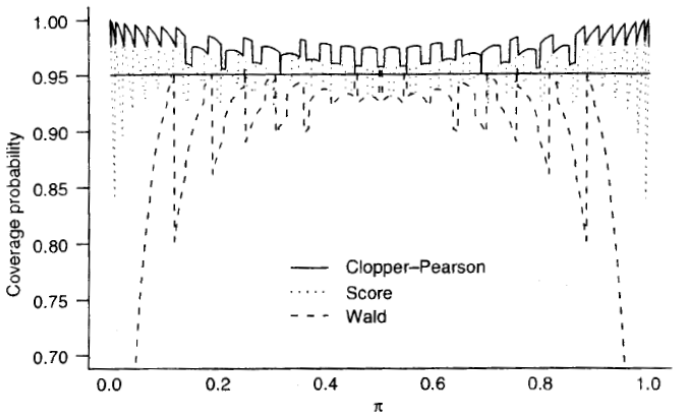

Figure 7. Plot of Coverage Probabilities for the Nominal 95% Confidence Intervals for Binomial Proportion p when $n = 25$ (From: Agresti, 2002)

Having plotted the coverage probabilities as a function of p for fixed $n = 50$, Brown et al. (2001) also reached the same conclusion as Agresti (2002) did (Figure 8). They also found that "coverage of the Wilson interval fluctuates acceptably near $1 - \alpha$, except for p very near zero or one" (p. 110).

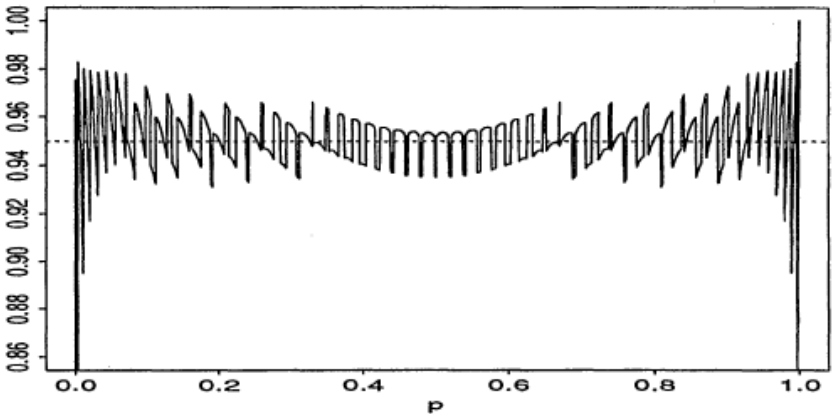

Figure 8. Coverage Probabilities for the 95% Wilson Interval when $n = 50$ (From: Brown et al., 2001)

3. The Adjusted Wald (Agresti-Coull) Confidence Interval

Agresti and Coull (1998) proposed a simple adaptation of the Wald interval that also performs well even for small

samples. As mentioned previously, The Wilson interval is the set of p_0 values for which $|\hat{p} - p_0| / \sqrt{p_0 (1 - p_0)/n} < z_{\alpha/2}$, which is given in Equation 1 and can be rewritten as

$$\hat{p}\left(\frac{n}{n + z^2_{\alpha/2}}\right) + \frac{1}{2}\left(\frac{z^2_{\alpha/2}}{n + z^2_{\alpha/2}}\right)$$

$$\pm z_{\alpha/2}\sqrt{\frac{1}{n + z^2_{\alpha/2}}\left[\hat{p}(1 - \hat{p})\left(\frac{n}{n + z^2_{\alpha/2}}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{z^2_{\alpha/2}}{n + z^2_{\alpha/2}}\right)\right]}$$

With regard to deriving the adjusted Wald interval, the following is given by Agresti and Caffo (2000):

> The midpoint is a weighted average of $\hat{p}$ and 1/2, and it equals the sample proportion after adding $z^2_{\alpha/2}$ pseudo observations, half of each type. The square of the coefficient of $z_{\alpha/2}$ in this formula is a weighted average of the variance of a sample proportion when $p = 1/2$, using $n + z^2_{\alpha/2}$ in place of the usual sample size n. For the 95% case, Agresti and Coull (1998) used this representation to motivate approximating the score interval by the ordinary Wald interval after adding $z^2_{.025} = 1.96^2 \approx 4$ pseudo observations, two of each type. That is, their adjusted "add two successes and two failures" interval has the simple form
>
> $$\tilde{p} \pm z_{.025}\sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n}}}$$
>
> but with $\tilde{n} = (n + 4)$ trials and $\tilde{p} = (X + 2)/(n + 4)$. The midpoint equals that of the 95% [Wilson] confidence interval (rounding $z_{.025}$ to 2.0 for that interval), but the coefficient of $z_{.025}$ uses the variance $\tilde{p}(1 - \tilde{p})/\tilde{n}$ at the weighted average $\tilde{p}$ of $\hat{p}$, and 1/2 rather than the weighted average of the variances; by Jensen's inequality, the adjusted interval is wider than the [Wilson] interval. (p. 280-281)

For confidence levels $(1-\alpha)$ other than 0.95, the adjusted Wald interval adds $t/2$ successes and $t/2$ failures, where $t = z^2_{\alpha/2}$. However, Agresti and Caffo (2000) state that the performance of the adjusted Wald interval with $t = 4$ is much better than the Wald interval for the usual confidence levels.

Figure 9 shows the improvement in performance of the adjusted Wald interval for small samples when compared to the ordinary Wald interval.

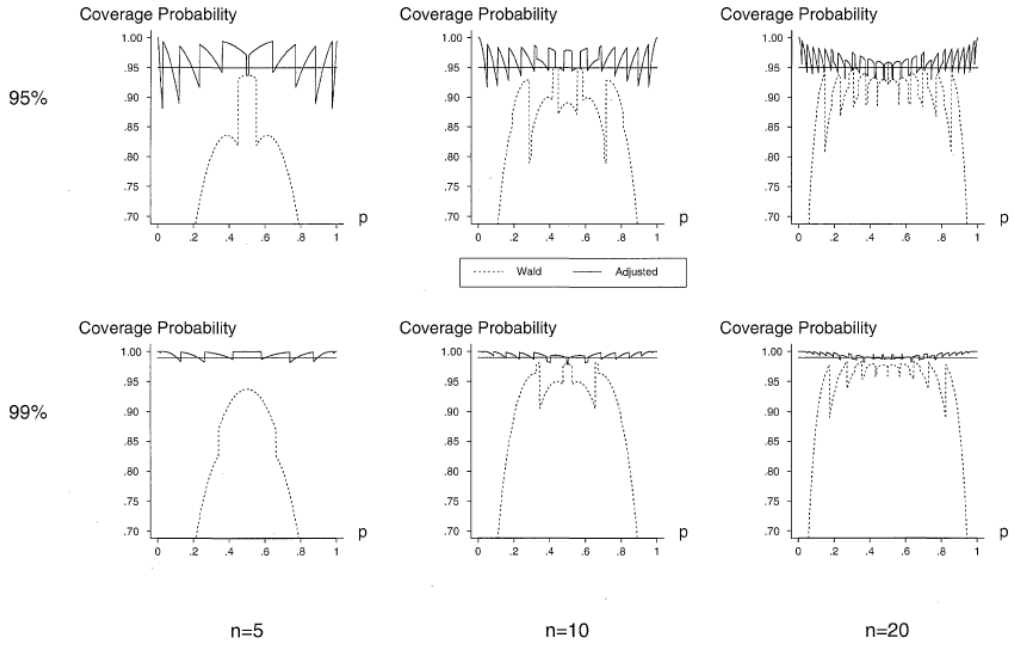

Figure 9. Coverage Probabilities for the Binomial Proportion p with Nominal 95% and 99% Wald Confidence Intervals and the Adjusted Interval Based on Adding Four Pseudo Observations, for *n=5, 10,* and *20* (From: Agresti & Caffo, 2000)

Relative to the Wilson interval, Agresti and Coull (1998) explain the advantage of the adjusted Wald interval by not having spikes with seriously low coverage near $p = 0$ and 1. They also show that, on the average, this simple

adjustment to the Wald interval changes it from highly liberal to slightly conservative (see Figure 10), and to a bit more conservative than the Wilson method (see Figure 11).[2] Their results suggest that the adjusted Wald interval behaves adequately for practical applications for essentially any n regardless of the value of p.

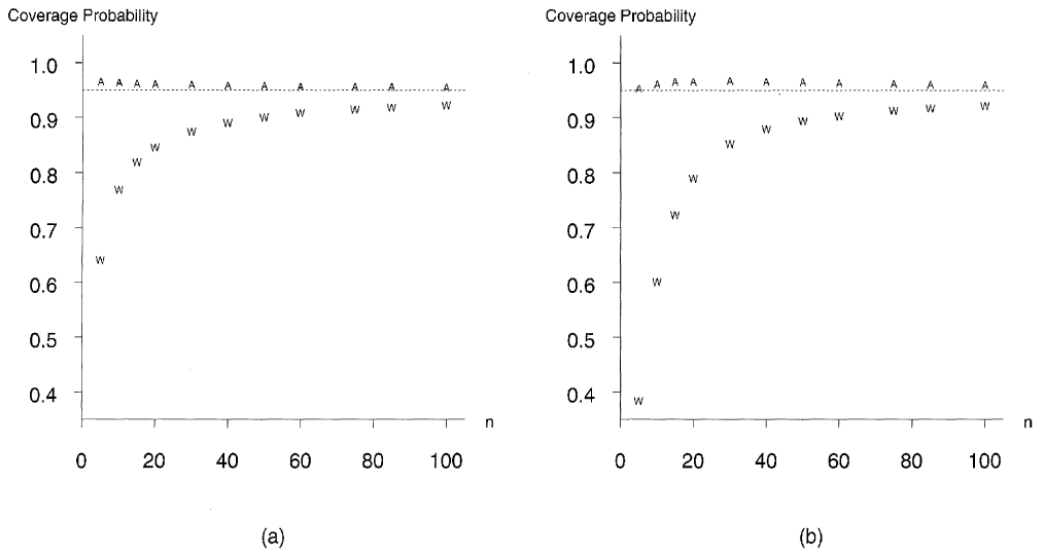

Figure 10. Mean Coverage Probability as a Function of Sample Size for the Nominal 95% Wald (W) and Adjusted Wald (A) Intervals, When p has (a) a Uniform (0,1) Distribution and (b) a Beta Distribution with $\mu = 0.10$ and $\sigma = 0.05$ (From: Agresti & Coull, 1998)

[2] The coverage performance of the Exact (Clopper-Pearson) interval will be addressed later in this chapter.

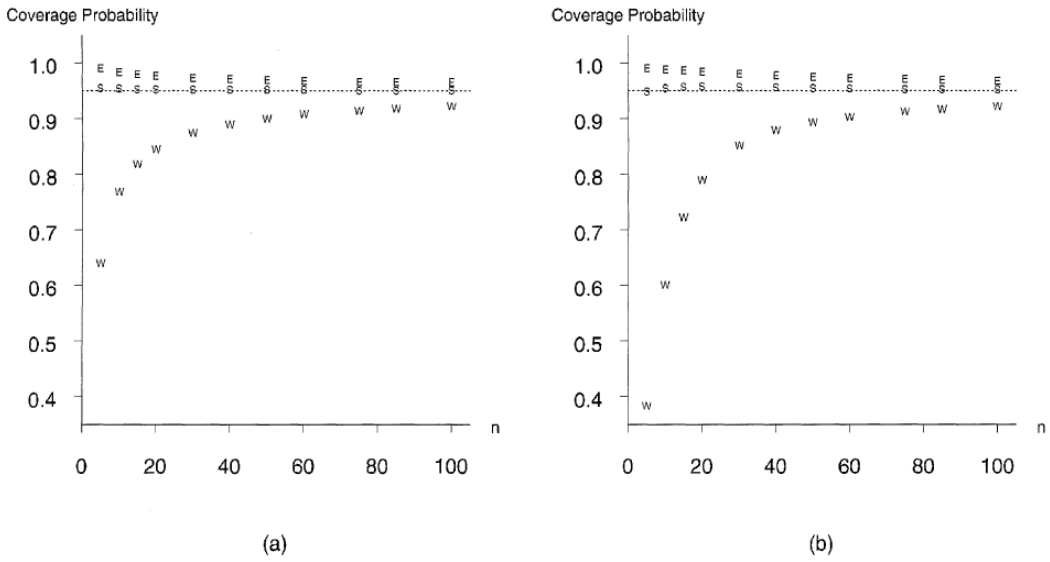

Figure 11. Mean Coverage Probability as a Function of Sample Size for the Nominal 95% Exact (E), Wilson (S), and standard (W) Intervals, When p has (a) a Uniform (0,1) Distribution and (b) a Beta Distribution with $\mu = 0.10$ and $\sigma = 0.05$ (From: Agresti & Coull, 1998)

The results of another study conducted by Brown et al. (2001) generally support those of Agresti and Coull (1998). The adjusted Wald interval turns out to be slightly conservative in terms of average coverage probability, especially for small n (see Figure 12).[3]

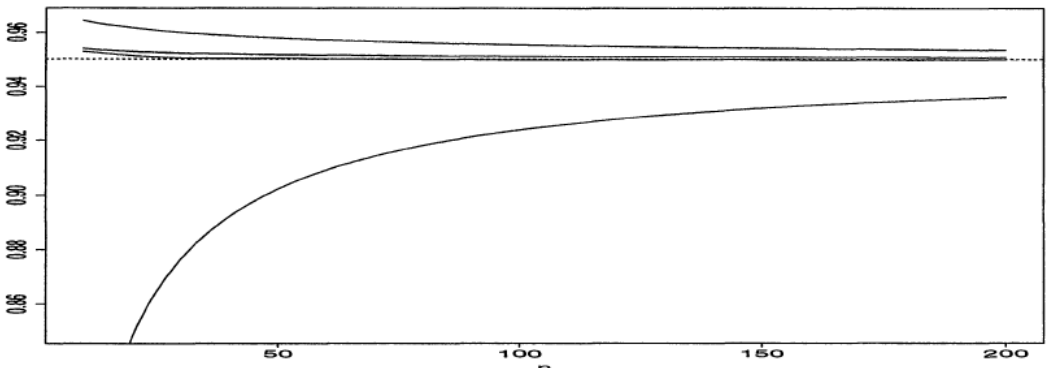

Figure 12. Comparison of the Average Coverage Probabilities (From: Brown et al., 2001)

[3] From top to bottom: the Agresti-Coull interval, the Wilson interval, the Jeffreys Prior interval, and the Wald interval. The nominal confidence level is 0.95.

Based on their analyses, the recommendation of Brown et al. (2001) differs from that of Agresti and Coull. They recommend the adjusted Wald interval for practical use when $n \geq 40$. For $n \leq 40$, their recommendations are the Wilson interval and the Jeffreys prior interval, both of which will be examined later in this chapter.

4. The Clopper-Pearson Confidence Interval

The Clopper-Pearson interval for p is based on inverting the binomial test of $H_0 : p = p_0$ versus $H_a : p \neq p_0$. Some authors refer to this interval as the "exact" procedure because it uses the exact binomial distribution of $n\hat{p}$ rather than a normal approximation. The Clopper-Pearson interval has endpoints that are the solutions in p_0 to the equations

$$\sum_{k=x}^{n} \binom{n}{k} p_0^k \left(1 - p_0\right)^{n-k} = \alpha/2$$

and

$$\sum_{k=0}^{n} \binom{n}{k} p_0^k \left(1 - p_0\right)^{n-k} = \alpha/2$$

except that the lower bound is 0 when $x = 0$ and the upper bound is 1 when $x = n$, where x is the observed number of successes in n trials. This interval estimator is guaranteed to have coverage probability of at least $1 - \alpha$ for every possible value of p. When $x = 1, 2, \ldots, n - 1$, the confidence interval equals

$$\left[1 + \frac{n - x + 1}{x\, F_{2x,\, 2(n-x+1),\, 1-\alpha/2}}\right]^{-1} < p < \left[1 + \frac{n - x}{(x + 1)\, F_{2(x+1),\, 2(n-x),\, 1-\alpha/2}}\right]^{-1}$$

where $F_{a,b,c}$ denotes the 1-c quantile from the F distribution with degrees of freedom a and b. Similarly, the lower endpoint is the $\alpha/2$ quantile of a beta distribution with parameters x and n-x+1, and the upper end point is the $1 - \alpha/2$ quantile of a beta distribution with parameters $x + 1$ and $n - x$ (Agresti & Coull, 1998, p. 119).

In regards to the performance and the general characteristics of the Clopper-Pearson interval, Agresti and Coull (1998) plot the coverage probabilities as a function of p when $n = 5$ and $n = 10$ (see Figure 13). They reach the following conclusions:

> This procedure is necessarily conservative, because of the discreteness of the binomial distribution (Neyman, 1935), just as the corresponding exact test (without supplementary randomization on the boundary of critical region) is conservative. For any fixed parameter value, the actual coverage probability can be much larger than the nominal confidence level unless n is quite large, and we believe it is inappropriate to treat this approach as optimal for statistical practice. (p. 119)

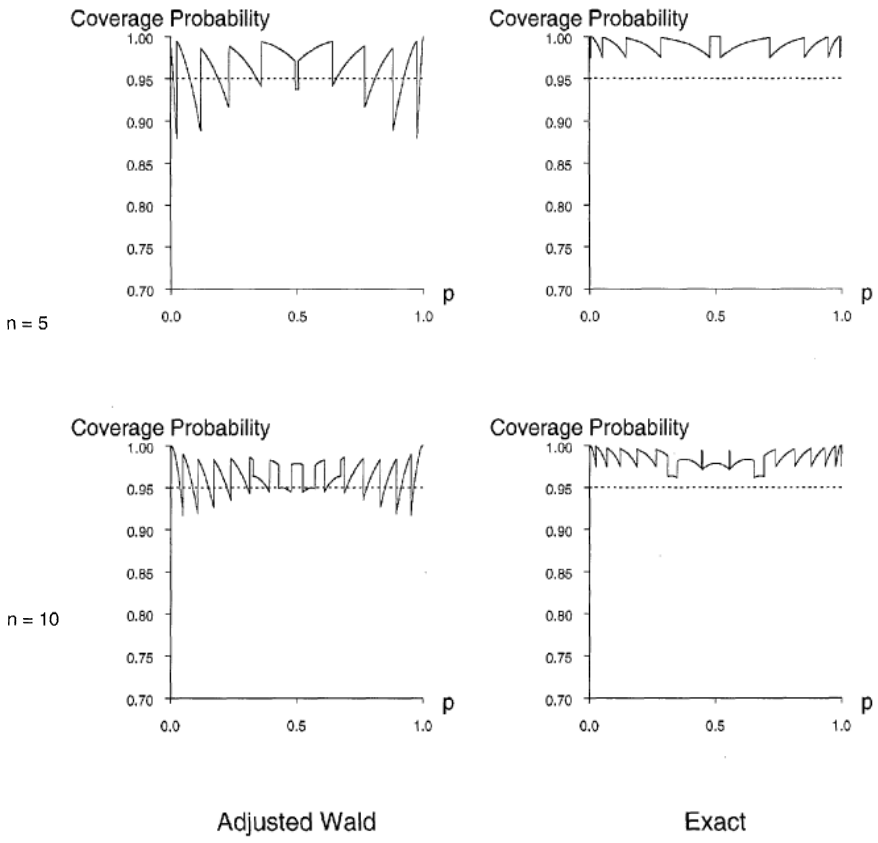

Figure 13. Coverage Probabilities for the Nominal 95% Adjusted Wald and Clopper-Pearson Intervals as a Function of p (After: Agresti & Coull, 1998)

The plots shown in Figure 14 also illustrate the conservative coverage of the Clopper-Pearson interval for different sample sizes when $p = 0.25$ and 0.05.

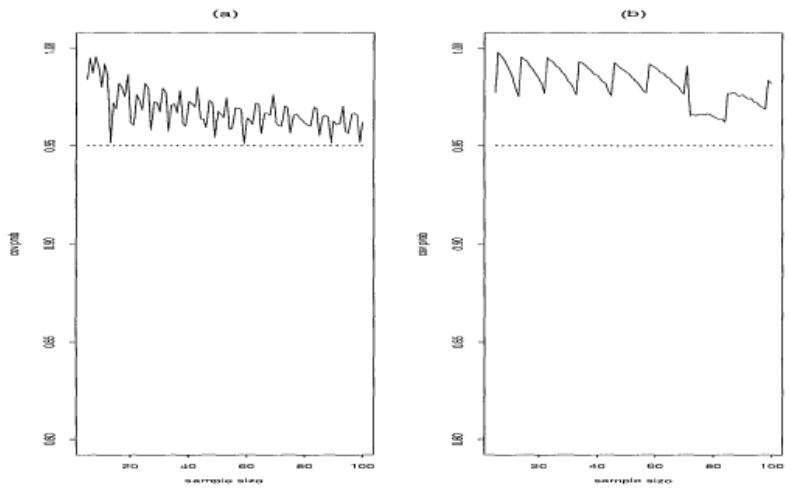

Figure 14. Coverage Probabilities for the 95% Clopper-Pearson Interval (a) $p = 0.25$, (b) $p = 0.05$ (From: Henderson & Meyer, 2001)

Moreover, the following findings of Brown et al. (2001) in regards to the coverage performance of the Clopper-Pearson interval also support those mentioned so far:

> This interval guarantees that the actual coverage probability is always equal to or above the nominal confidence level. However, for any fixed p, the actual coverage probability can be much larger than $1 - \alpha$ unless n is quite large, and thus, the confidence interval is rather inaccurate in this sense... The Clopper-Pearson interval is wastefully conservative and is not a good choice for practical use, unless strict adherence to the prescription $C(p, n) \geq 1 - \alpha$ is demanded. (p. 113)

5. The Jeffreys Prior Interval

The Jeffreys prior interval is the equal-tailed Bayesian interval using Jeffreys prior $\text{Beta}(½, ½)$, which is considered as non-informative. The Bayesian approach combines prior information about the parameter p with the data to get the posterior information. Suppose

$X \sim Binomial(n, p)$ and suppose p has a prior distribution $Beta(\alpha_1, \alpha_2)$; then the posterior distribution of p is $Beta(X + \alpha_1, n - X + \alpha_2)$. Thus, the $100(1-\alpha)\%$ equal-tailed Jeffreys prior interval is

$$\left[B(\alpha/2, X + 1/2, n - X + 1/2), B(1 - \alpha/2, X + 1/2, n - X + 1/2)\right]$$

where $B(\alpha, m_1, m_2)$ denotes the α quantile of a $Beta(m_1, m_2)$ distribution. The lower bound of the confidence interval is zero when $X = 0$ and the upper bound is one when $X = n$ (Brown et al., 2001).

In Figure 15, it is obvious that the coverage of the Jeffreys interval is qualitatively similar to that of the Wilson interval over most of the parameter space $[0, 1]$. Refer to Figure 8 for the comparison.

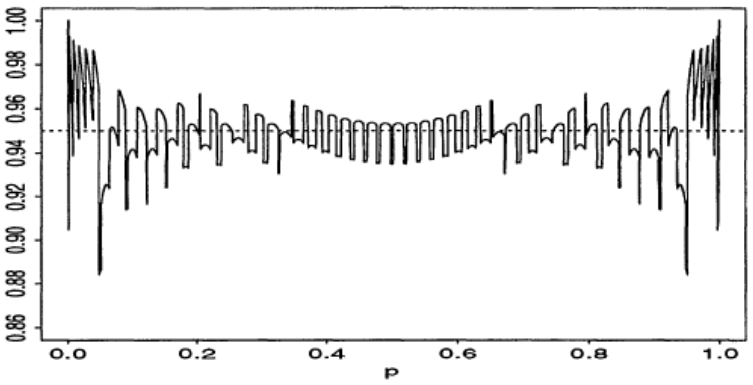

Figure 15. Coverage Probabilities for the 95% Jeffreys Prior Interval, when $n = 50$ (From: Brown et al., 2001)

Agresti and Coull (1998) also point out that the Bayesian confidence intervals with beta priors that are only weakly informative perform well.

When Figure 12 is examined once again, it is seen that the average coverage of the Jeffreys prior interval is very close to the nominal confidence level. As a result of their

analyses, Brown et al. (2001) recommend the Jeffreys prior interval as a serious and credible candidate for practical use when $n \leq 40$.

B. SAMPLE SIZE CALCULATION FOR THE BINOMIAL PROPORTION

Estimating a binomial proportion is the aim of many studies. In these types of studies, sample size is important because of its effect on the precision of the observed proportions (Eng, 2003).

Suppose that the U.S. Army Yuma Proving Ground engineers want to estimate the sensor detection probability p at a certain range in a series of n independent Bernoulli trials, where n is yet to be determined. Regardless of n, it is known that the point estimator for p will be X/n, where X is the number of successes (detections) out of n trials. It is also known that the standard deviation of the estimate will decrease as n increases. Therefore, as the sample size increases, so does the precision of the estimate (Larsen & Marx, 1986).

Unfortunately, the greater the sample size, the more budget the study requires. The budget and resources allocated to an experimental study may not always allow for a large sample size. As stated by Larsen & Marx (1986), the experimenters are thus faced with a trade-off. On one hand, they wish to have as precise an estimator as possible, and on the other hand, they have to keep costs to a minimum. These two conflicting objectives raise the following question: what is the smallest sample size that will guarantee (with a probability of $1-\alpha$) that the point estimate will be some specified distance, d, of p?

In the studies designed to measure a characteristic in terms of a proportion, the well-known sample size formula based on the normal approximation to the binomial distribution is

$$n = \left\lceil \frac{z_{\alpha/2}^2 p\,(1 - p)}{d^2} \right\rceil \qquad (2)$$

where $z_{\alpha/2}$ is the upper $100(1 - \alpha)$ percentile of the normal distribution, d is the half-width of the confidence interval, and $\lceil a \rceil$ denotes the smallest integer larger than a (Rahme & Joseph, 1998).

According to Larsen & Marx (1986), Equation 2 is not acceptable because it involves the unknown parameter p. However, since $0 \leq p \leq 1$, the product $p(1 - p)$ will always be less than or equal to $1/4$. Therefore,

> [one] can insure that Equation [2] is satisfied in even the most "difficult" of situations (when p is actually $1/2$) by choosing as the sample size the smallest n such that
>
> $$n \geq \frac{z_{\alpha/2}^2}{4d^2}.$$ (p. 281) (3)

For instance, suppose that the U.S. Army Yuma Proving Ground engineers want to estimate the probability of sensor detection at a certain range. They want to have a 95% probability that their final estimate of p is correct to within 0.05 (i.e., they want the half-width of the confidence interval to be 0.05 with probability 0.95). According to Equation 3, n should be 385, which seems apparently too large a sample size to be achieved by the Yuma Test Center.

If the value of p is available based on prior information, Larsen and Marx (1986) suggest that it may be possible to reduce substantially the necessary sample size by not making the $p(1-p) = 1/4$ assumption. However, for well-known confidence interval-based sample size formulae where the parameter of interest is a proportion p, Kupper & Hafner (1989) recommend that, when economically feasible, researchers use the maximum sample size computed assuming that $p(1-p) = 1/4$.

Equation 2 is in fact based on the Wald interval. Devore (2004) gives another sample size formula that is based on the Wilson interval. With notation altered to match that of this thesis, the equation for the sample size n necessary to give an interval with a desired precision is given by

$$n = \frac{2z_{\alpha/2}^2 pq - z_{\alpha/2}^2 w^2 \pm \sqrt{4z^4 pq\left(pq - w^2\right) + w^2 z_{\alpha/2}^4}}{w^2} \qquad (4)$$

where w is the specified width of the confidence interval and $q = 1 - p$.

In the above example, where the width of the confidence interval is desired to be 0.10 with probability 0.95, the maximum sample size that Equation 4 yields is 381.

The sample sizes that will be obtained by using Equations 2 and 4 are both approximate. In a study where exact sample size determination for binomial experiments was examined, Rahme and Joseph (1998) provide an algorithm that calculates the exact sample sizes under a modified criterion. In their modified criterion, instead of the

interval length of $2d$ centered at $\hat{p} = X/n$, the highest density interval of length $\leq 2d$ containing $\hat{p}$ is considered. For the example given above, they report the required sample size as 370. See Rahme and Joseph (1998) for more details on an exact sample size calculation using the modified criterion.

Moreover, an exact Bayesian approach to sample size is given by Joseph, Wolfson, and Berger (1995) using the worst outcome criterion (WOC), which is also based on highest-density intervals. Refer to Joseph et al. (1995) for more details on WOC.

Table 2 lists the sample sizes computed by the aforementioned confidence interval-based formulae and some calculation results obtained by Rahme and Joseph (1998) and Joseph et al. (1995).

CI Width (w)	Sample Sizes Based on			
	The Wald Interval	The Wilson Interval	The Modified Criterion by Rahme & Joseph	WOC Criterion by Joseph et al.
0.50	16	12	NA	12
0.40	25	21	NA	21
0.30	43	39	NA	40
0.25	62	58	NA	59
0.20	97	93	97	93
0.10	385	381	370	381

Table 2. Sample Sizes for Various Values of CI Width, Using Different Approaches when $1 - \alpha = 0.95$

As can be seen from the table, within the context of a binomial experiment, different approaches to sample size calculations lead to almost the same sample size, which

could be impracticably large for the experiments designed to estimate the sensor detection probabilities, especially when the precision of the estimate is required to be high.

C. OVERVIEW OF THE LINEAR LOGISTIC REGRESSION MODEL

Logistic regression has been increasingly used in a wide variety of applications as mentioned in Chapter I. In terms of answering the primary thesis question of sample size determination for estimation of sensor detection probabilities as a function of range to the target, this section provides general information about simple logistic regression models and focuses on estimating the binary response probabilities and the precision of the estimates. The main reason in doing so is to introduce the fact that the precision of the estimated detection probabilities based on the fit of a simple linear logistic regression model is quite good when compared to those based on estimating the binomial proportions. Refer to Agresti (2002) and Collett (1991) for further details in regards to fitting a linear logistic model to the binary data and conducting model diagnostics.

1. Definition

Logistic regression models, also called ***logit models***, are generalized linear models (GLMs) with a binomial random component and logit link function (Agresti, 2002, p. 123).

For a binary response variable Y and an explanatory variable X (which in our case is the range to the target), let $p(x) = P(Y = 1 | X = x) = 1 - P(Y = 0 | X = x)$. The logistic regression model is given by

$$p(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$$

Equivalently, the log odds, called the ***logit***, has the linear relationship

$$\text{logit}\left[p\left(x\right)\right] = \log\left[\frac{p\left(x\right)}{1 - p\left(x\right)}\right] = \alpha + \beta x \quad \text{(Agresti, 2002, p. 166)}$$

The function that relates $p\left(x\right)$ to the linear component $\alpha + \beta x$ is generally known as the ***link function*** (Collett, 1991, p. 56).

2. Interval Estimate for the Binary Response Probability

A confidence interval for the corresponding true response probability at x_0 is best obtained by constructing a confidence interval for $\text{logit}\left[p\left(x_0\right)\right]$ and then transforming the resulting limits to give an interval estimate for $p\left(x_0\right)$ itself (Collett, 1991, p. 88).

For fixed $x = x_0$, the estimator of $\text{logit}\left[\hat{p}\left(x_0\right)\right]$ is $\hat{\alpha} + \hat{\beta}x_0$, where $\hat{\alpha}$ and $\hat{\beta}$ are maximum likelihood estimators of α and β. The large-sample standard error (se) for $\text{logit}\left[\hat{p}\left(x_0\right)\right]$ is given by

$$se\left(\hat{\alpha} + \hat{\beta}x_0\right) = \sqrt{se^2\left(\hat{\alpha}\right) + x_0^2\, se^2\left(\hat{\beta}\right) + 2\, x_0\, \text{cov}\left(\hat{\alpha}, \hat{\beta}\right)}$$

where

$$\text{cov}\left(\hat{\alpha}, \hat{\beta}\right) = corr\left(\hat{\alpha}, \hat{\beta}\right) se\left(\hat{\alpha}\right) se\left(\hat{\beta}\right)$$

A 95% confidence interval for $\text{logit}\left[p\left(x_0\right)\right]$ is then $\left(\hat{\alpha} + \hat{\beta}x_0\right) \pm z_{1-0.05/2} se\left(\hat{\alpha} + \hat{\beta}x_0\right)$ where $z_{1-0.05/2} \approx 1.96$ (Agresti, 2002).

3. Precision of the Estimated Binary Response Probabilities Based on the Fit of a Logistic Regression Model

In order to estimate $p(x_0)$, by ignoring the model fit one could simply use the sample proportions (i.e., the saturated model) and construct one of the well-performing confidence intervals mentioned in Section A.

On the other hand, the precision of estimated binary response probabilities that would be obtained by using logistic regression is much better. In regards to this issue, Agresti (2002) states:

> [w]hen the logistic regression model truly holds, the model-based estimator of probability is considerably better than the sample proportion. The model has only two parameters to estimate, whereas the saturated model has a separate parameter for every distinct value of x...Reality is a bit more complicated. In practice, the model is not exactly the true relationship between [p(x)] and x. However, if it approximates the true probabilities decently, its estimator still tends to be closer than the sample proportion to the true value. The model smoothes the sample data, somewhat dampening the observed variability. The resulting estimators tend to be better unless each sample proportion is based on extremely large sample. (p. 173-174)

THIS PAGE INTENTIONALLY LEFT BLANK

III. SAMPLE PROPORTION-BASED ANALYSIS

A. INTRODUCTION

In this chapter, the performances of the confidence intervals described in Section B of Chapter II are analyzed through simulation in terms of their coverage probabilities and lengths for the experimental setup used by the U.S. Army Yuma Proving Ground.

In general, the actual coverage probability of a confidence interval for a binomial proportion p could be estimated through simulation as follows (Henderson & Meyer, 2001):

- First, a large number of random samples are drawn from a binary population with population parameter p and sample size n.
- Second, $100(1-\alpha)\%$ confidence intervals are calculated for each sample.
- Third, the proportion of these confidence intervals that contain p is computed. This is the simulated coverage probability.

One can also compute the actual coverage probabilities exactly for any given sample size n and binomial proportion p by computing confidence intervals for $x = 0$ through n, where x is the number of successes and n is the number of trials. For example, suppose $n = 15$ and $p = 0.25$. The 95% Wilson confidence interval for $x = 1$ is $(0.012, 0.298)$, and for $x = 7$ is $(0.248, 0.699)$. These two intervals, as well as those for $1 \le x \le 7$, capture the true parameter $p = 0.25$. If $x = 0$ or $x \ge 8$, the confidence interval does not capture p. The actual coverage probability is then the probability

that the number of observed successes is between one and seven (inclusive) in a binomial trial with $n = 15$ and $p = 0.25$ as shown below.

$$P(1 \le X \le 7) \quad \text{when } X \sim Binomial(n = 15, p = 0.25)$$
$$= \sum_{i=0}^{7} \binom{n}{i} p^i (1-p)^{n-i} - \sum_{i=0}^{0} \binom{n}{i} p^i (1-p)^{n-i}$$
$$= 0.9693$$

The estimated coverage probability through simulation for the example given above is 0.9691.

The simulation is based on the ***binning approach***, which is currently being used by the U.S. Army Yuma Proving Ground. In this approach, the flight path is divided into approximately evenly spaced range intervals, and the number of detections out of n trials for each range interval is recorded. This approach can also be referred to as a ***sample proportion-based approach***. Similar to what the U.S. Army Yuma Proving Ground engineers do, the number of bins used in the simulation is set to 20, and the number of observations obtained for each bin (range interval) is five. At this point, it should be noted that the probability of detection is not the same for all five trials in each of the 20 bins. Therefore, the model for the probability of detection differs from the assumptions for inference about a binomial proportion p in that, here, the probability of detection is increasing as the range to target decreases. One should keep in mind that this phenomenon is likely to affect the coverage probabilities and lengths of the intervals calculated for each bin by introducing bias.

Moreover, this chapter reports the results of an approach one might try in an attempt to calibrate the confidence intervals to obtain narrower ones with coverage performance similar to the ones prior to calibration.

B. ASSUMPTIONS

The detection of an aircraft by a sensor depends on several factors such as range, altitude, radar cross section of target, weather conditions, and how well trained the radar operators are.

Since the data provided by the U.S. Army Yuma Proving Ground consist of a binary response variable (detection, no detection) and a predictor variable (range), this thesis will seek to answer the question of determining sample size for the estimation of sensor detection probabilities assuming that all factors except for range are fixed.

C. ANALYSIS THROUGH SIMULATION

Because of its similarity to the distribution of actual observed responses, for demonstration purposes the model describing the relationship between the observed response and the range is chosen to be

$$p_i = \frac{1}{1 + e^{x_i}}$$

where $Y_i \sim Binomial(n_i = 1, p_i)$. Software written in the S-PLUS language that implements simulations that mimic the approach taken by the U.S. Army Yuma Proving Ground is presented in Appendices A through E.

Figure 16 illustrates the actual coverage probabilities as a function p for the five different confidence interval methods reviewed in Chapter II when the number of observations in each range interval is five.

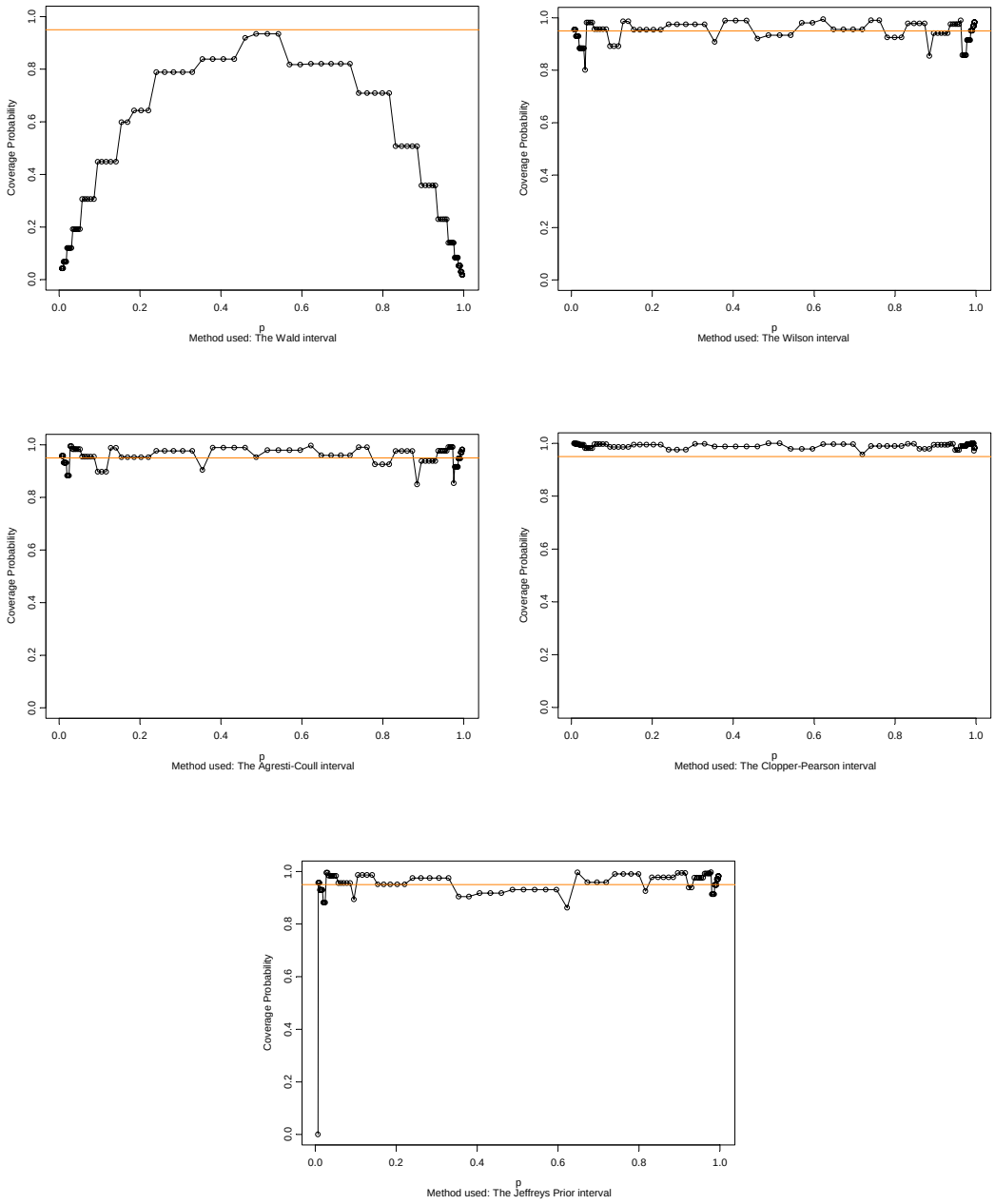

Figure 16. Coverage Probabilities for the 95% Confidence Intervals when $n = 5$

In terms of coverage probabilities, the Wald interval behaves poorly. The coverage probabilities are typically less than the 95% nominal confidence level, which means that in the repeated trials throughout the simulation, fewer than 95% of the computed intervals capture the true

population parameter. The Clopper-Pearson interval has coverage probabilities bounded below by the 95% nominal confidence level. However, the typical coverage is much higher than that level. On the other hand, the Wilson, Agresti-Coull, and equal-tailed Jeffreys prior intervals turn out to be comparable.

Table 3 reports the mean coverage probabilities ($\bar{C}_n(p) = \int C_n(p)\,dp$) as well as the root mean squared error of the coverage probabilities ($Root\ MSE = \sqrt{\int \left(C_n(p) - [1-\alpha]\right)^2 dp}$). Root MSE is provided to describe how far the actual coverage probabilities typically fall from the nominal confidence level (Agresti & Coull, 1998).

Method	Mean Coverage Probability	Root MSE
Wald	0.641	0.388
Wilson	0.945	0.033
Agresti-Coull	0.953	0.031
Exact	0.980	0.040
Jeffreys Prior	0.945	0.037

Table 3. Mean Coverage Probabilities of Nominal 95% Confidence Intervals and Root MSEs

The mean actual coverage probability for the Wald interval is too small. On the other hand, the Clopper-Pearson interval is very conservative. When compared with the Wilson and the equal-tailed Jeffreys prior interval, the Agresti-Coull interval has a better mean coverage probability. Moreover, the root MSE values indicate that

the variability about the nominal 95% confidence level is smaller for the Agresti-Coull and the Wilson intervals than for the others.

Besides coverage, length is also important in evaluating the confidence intervals. Figure 17 plots the mean confidence interval lengths for each bin for each method.

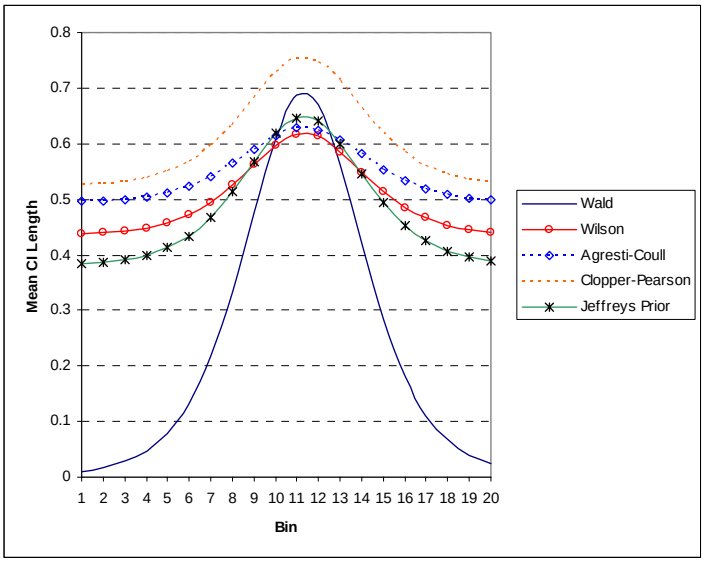

Figure 17. Mean Lengths for the 95% Confidence Intervals when $n = 5$

It is no surprise that the Wald interval is the shortest in bins 1 through 9 and 13 through 20. This is because p is near the boundaries in these range intervals depending on the model used. As stated by Brown et al. (2001), "[The Wald interval] is not really in contention as a credible choice for such values of p because of its poor coverage properties in that region" (p. 111). The Clopper-Pearson interval is the largest over the whole parameter space because of its conservativeness. The Wilson interval is the shortest in bins 10 through 12, where p ranges between 0.35 and 0.72. When compared with the Wilson and

the Agresti-Coull interval, the equal-tailed Jeffreys prior is the shortest in bins 1 through 8 and 14 through 20. As mentioned in Chapter II, the Agresti-Coull interval is always a bit larger than the Wilson interval over the whole parameter space.

Based on the analysis done so far and the review in Chapter II, when the ***binning approach*** is adopted to estimate sensor detection probabilities the use of the Wald interval and the Clopper-Pearson interval is not recommended. While the Wald interval performs poorly for any values of n and p, the Clopper-Pearson interval is highly conservative and yields confidence intervals unnecessarily large. The Wilson, Agresti-Coull, and equal-tailed Jeffreys prior intervals can have coverage probabilities lower than the nominal confidence levels; however, their typical coverage probability is close to that level. In forming a confidence interval, Agresti and Coull (1998) ask and answer the following question:

> In forming a 95% confidence interval, is it better to use an approach that guarantees that the actual coverage probabilities are ***at least*** .95 yet typically achieves coverage probabilities of about .98 or .99, or an approach giving narrower intervals for which the actual coverage probability could be less than .95 but is usually quite close to .95? For most applications, we prefer the latter. (p. 125)

The answer given by Agresti and Coull to the above question also agrees with the recommendations made by Brown et al. (2001).

In choosing one of the three recommended intervals (i.e., the Wilson, Agresti-Coull, or equal-tailed Jeffreys prior intervals), the experimenters are faced with a trade-

off. On one hand, they want to have narrower confidence intervals; on the other hand, they want these intervals to have good coverage probabilities. For the current situation, despite the wider confidence intervals, one may use the Agresti-Coull interval depending on its better coverage performance. One can also use the Wilson interval or the equal-tailed Jeffreys prior interval because the coverage performance of these intervals is comparable. The only challenge in using the equal-tailed Jeffreys prior is the need for a statistical software package to compute the endpoints of the interval. Nevertheless, the following function written in the S-PLUS language and shown in Figure 18 can be used to compute the equal-tailed Jeffreys prior interval endpoints:

```
function(n = 5, k = seq(0, n, 1), alpha = 0.05)
{
       # Arguments
       # n: Number of trials
       # k: Number of successes
       # alpha: Significance level
       #---------------------------
       lo <- rep(0, length(k))
       up <- rep(1, length(k))
       lo[k == n] <- qbeta(alpha/2, k[k == n] + 1/2, n - k[k == n] + 1/2)
       up[k == 0] <- qbeta(1 - alpha/2, k[k == 0] + 1/2, n - k[k == 0] + 1/2)
       index <- (0 < k) & (k < n)
       lo[index] <- qbeta(alpha/2, k[index] + 1/2, n - k[index] + 1/2)
       up[index] <- qbeta(1 - alpha/2, k[index] + 1/2, n - k[index] + 1/2)
       data.frame(Num.Success = k, Lower.CL = lo, Upper.CL = up, Width = up - lo)
}
```

Figure 18. Function Written in the S-PLUS Language Used to Compute the Equal-tailed Jeffreys Prior Interval Endpoints

D. RESULTS OF CALIBRATING THE CONFIDENCE INTERVALS UNDER THE BINNING APPROACH

For the sensor detection problem, the probability of detection decreases with range to target. A simple approach to incorporate this feature is to let the confidence limits in each bin provide information about the adjustability of others in the subsequent as well as previous bins. Such a calibration procedure to get narrower confidence intervals with similar coverage probabilities works as follows:

- Starting from the first bin where the probability of detection is high, the lower confidence limit is compared with the ones in the subsequent bins and is replaced with the maximum lower confidence limit if there is one.
- A different procedure applies for adjustment of the upper confidence limits; therefore, this time, starting from the second bin, the upper confidence limit is compared with the one/ones in the previous bin/bins and is replaced with the minimum upper confidence limit if there is one.
- Notation for both procedures described above can be written as follows:

$$L_k = \max_{k \le i \le n_{bin}} \{L_i\}, \qquad U_k = \min_{1 \le i \le k} \{U_i\}$$

where n_{bin} is the number of bins, $[L_k, U_k]$ is the confidence interval for the k^{th} bin, and $k = \{1, 2, ..., n_{bin}\}$.

Using the procedures described above, Figures 19 through 22 plot the 95% confidence intervals and coverage probabilities for the Wilson, Agresti-Coull, Clopper-Pearson, and equal-tailed Jeffreys prior methods before and after the calibration. Due to the poor coverage performance, results for the Wald interval are not shown. Confidence intervals and coverage probabilities after calibration are in blue to enable comparisons.

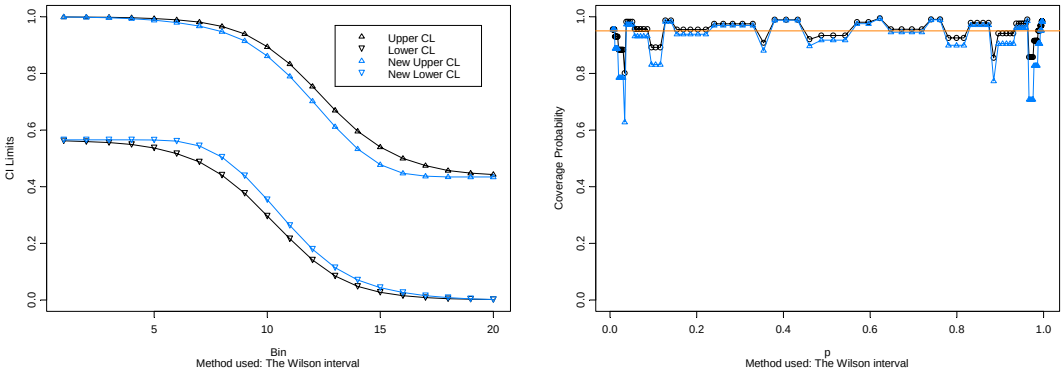

Figure 19. 95% Confidence Intervals and Coverage Probabilities for the Wilson Interval Before and After the Calibration

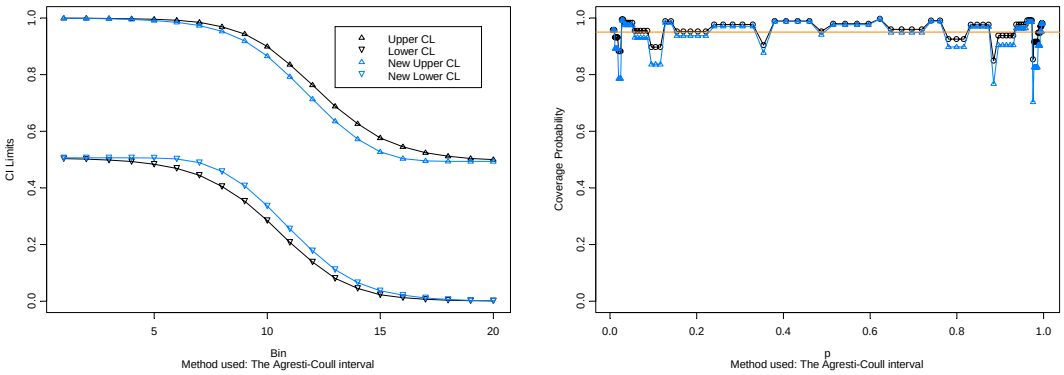

Figure 20. 95% Confidence Intervals and Coverage Probabilities for the Agresti-Coull Interval Before and After the Calibration

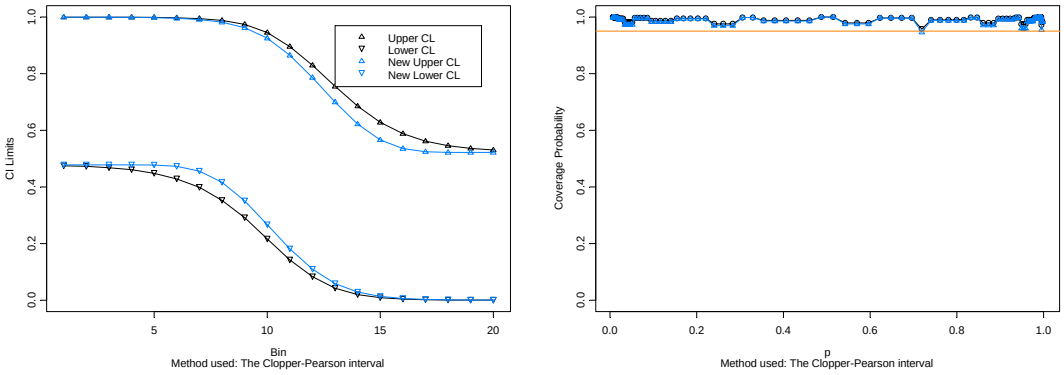

Figure 21. 95% Confidence Intervals and Coverage Probabilities for the Clopper-Pearson Interval Before and After the Calibration

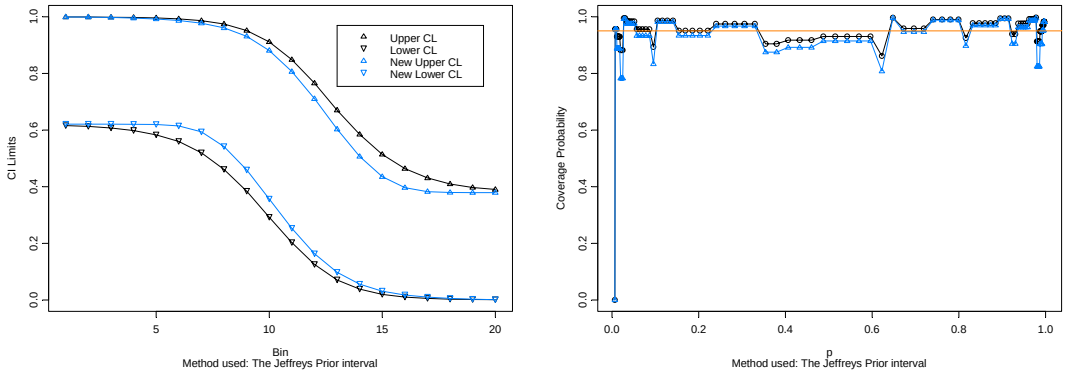

Figure 22. 95% Confidence Intervals and Coverage Probabilities for the Equal-Tailed Jeffreys Prior Interval Before and After Calibration

Figure 23 also illustrates the effect of calibration on the lengths of confidence intervals for each method.

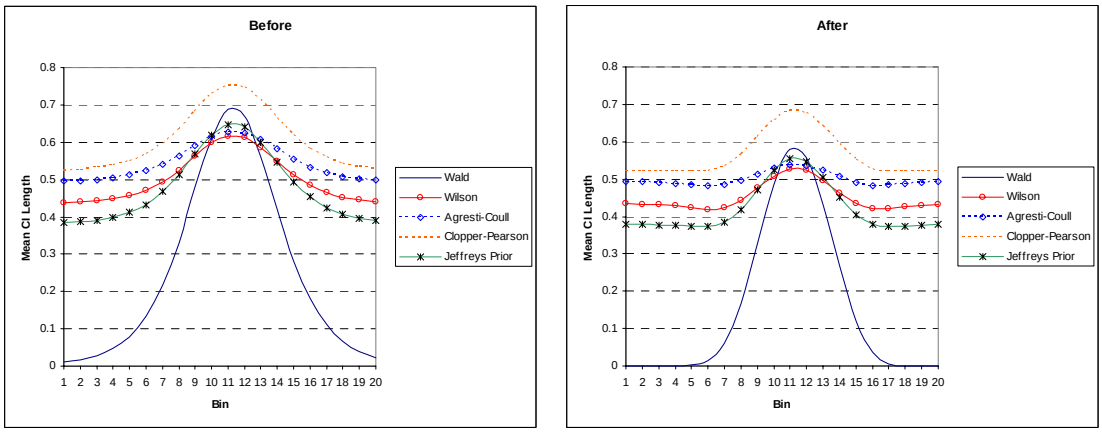

Figure 23. The Effect of Calibration on the Lengths of Confidence Intervals for Each Method

As seen from Figures 19 through 23, calibration causes the coverage probabilities to drop down over the whole parameter space while it provides narrower intervals as intended. Now the question is: do these calibrated intervals still perform well enough in terms of their coverage probabilities? To answer this question, Table 4

reports the mean coverage probabilities and the root MSEs of the actual coverage probabilities for each confidence interval.

Method	Before Calibration		After Calibration	
	Mean CP	Root MSE	Mean CP	Root MSE
Wilson	0.945	0.033	0.926	0.058
Agresti-Coull	0.953	0.031	0.937	0.047
Clopper-Pearson	0.980	0.040	0.978	0.038
Jeffreys Prior	0.945	0.037	0.930	0.050

Table 4. Mean Coverage Probabilities of the Nominal 95% Confidence Intervals and Root MSEs (Before and After Calibration)

The root MSE values on the far right of Table 4 indicate that the variability about the nominal 95% level is smaller for the Clopper-Pearson interval than for the other three intervals. The mean CP values get worse by 2.00%, 1.68%, and 1.59% for the Wilson, Agresti-Coull, and equal-tailed Jeffreys prior intervals respectively. The only improvement in terms of coverage turns out to be for the Clopper-Pearson interval. However, it is still conservative, and the other three competitors give better confidence intervals without the need for calibration.

E. CHAPTER SUMMARY

In this chapter, we focused on the analysis of selected confidence intervals in terms of their coverage probabilities and lengths, rather than the determination of sample size. As we pointed out in Chapter II, depending on the method used, the required sample sizes to achieve the

same specified goal in a binomial experiment may differ from each other. However, the resulting sample sizes may still turn out be impracticably large due to budget and time constraints. In this case, either the limited budget, or time, or both determine the sample size. The main issue in estimating a binomial proportion then happens to be selecting a method that will provide confidence intervals with acceptable coverage performance.

When the design of the experiment to estimate sensor detection probabilities is based on the ***binning approach***, where detections at ranges in a given interval are pooled, our simulation results show that the performance of the Wilson, Agresti-Coull, and equal-tailed Jeffreys prior intervals is comparable to performance based on a binomial experiment. Hence, either of the three can be used depending on preference. However, there are two major drawbacks of the ***binning approach***. The first one is that very large sample sizes are needed to obtain confidence intervals of reasonable length, and the second one is the lack of ability to estimate the sensor detection probabilities at a specified range. Therefore, the next chapter focuses on finding a better approach to sample size determination.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. LOGISTIC REGRESSION-BASED ANALYSIS

A. INTRODUCTION

This chapter focuses on estimating the probability of detection and studying the properties of corresponding 95% confidence intervals for different sample sizes based on using a logistic regression approach.

We note that for logistic regression the problem of calculating the required sample size when the goal of the study is to obtain 'confidence intervals for the estimated response' with a desired length is complex. Most literature focuses on sample size determination from different perspectives. For example, Hsieh, Bloch, and Larsen (1998) suggest the use of sample size formulae for comparing means or for comparing proportions in order to calculate the required sample size for a simple logistic regression model. Whittemore (1981), on the other hand, proposes a formula that gives approximate sample sizes needed to test hypotheses about the parameters in the case when the probability of response is small.

Unfortunately, there is no closed-form formula that serves the abovementioned goal in the literature. Therefore, an empirical approach based on simulation is adopted to determine the approximate sample size needed to obtain good estimates of sensor detection probabilities. This is done in the sequential generation of design points, where sampling is continued until an acceptable level of the coverage performance is achieved.

B. LOGISTIC REGRESSION MODEL-BASED ESTIMATORS

Before proceeding with the analysis of coverage performance of logistic regression model-based confidence intervals, we will first show numerically why the model-based estimator of probability is considerably better than the sample proportion. Consider the synthetic data set in Table 5, where five observations are recorded at each predetermined distance. Values in the x column are the predetermined distances and will be referred to as ***dose level***. Values in the y column are the observed responses, where a "1" indicates successful detection and a "0" no detection.

x	y	x	y	x	y	x	y
35	1	41	0	47	0	53	1
35	1	41	1	47	1	53	0
35	1	41	1	47	0	53	0
35	1	41	1	47	0	53	1
35	1	41	1	47	0	53	0
36	1	42	1	48	0	54	1
36	1	42	1	48	1	54	0
36	1	42	1	48	0	54	0
36	1	42	1	48	1	54	0
36	1	42	0	48	1	54	0
37	1	43	0	49	0	55	0
37	1	43	1	49	0	55	1
37	1	43	0	49	1	55	0
37	1	43	1	49	1	55	0
37	0	43	1	49	0	55	0
38	0	44	1	50	1	56	0
38	0	44	1	50	0	56	0
38	1	44	1	50	0	56	1
38	0	44	1	50	0	56	1
38	1	44	1	50	1	56	0
39	1	45	1	51	1		
39	1	45	1	51	1		
39	1	45	0	51	1		
39	1	45	1	51	1		
39	1	45	1	51	1		
40	0	46	1	52	0		
40	1	46	1	52	0		
40	0	46	1	52	0		
40	1	46	1	52	1		
40	1	46	1	52	1		

Table 5. Sample Data, Where Five Observations are Recorded at each Dose Level

As mentioned in Chapter II, one can ignore the model fit and simply use sample proportions to estimate sensor detection probability at a certain dose level. For example, the sample proportion estimate at $x = 42$ is $\hat{p} = X / n = 4 / 5 = 0.80$, and the standard error (se) for the sample proportion of 0.80 with only five observations is $\sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{0.8(1 - 0.8)/5} = 0.179$. On the other hand, by using the fitted logistic regression model in Figure 24, S-PLUS reports $se = 0.051$ for the model-based estimate $\hat{p}(x) = 0.756$.

```
> sample.fit <- glm(y~x, family=binomial, data=sample.data)
> summary(sample.fit)

Call: glm(formula = y ~ x, family = binomial, data = sample.data)
Deviance Residuals:
       Min        1Q    Median        3Q      Max
 -1.978554 -1.029833 0.5873538 0.8892233 1.513756
Coefficients:
                Value Std. Error   t value
(Intercept)  6.8061078 1.72231730  3.951715
          x -0.1351629 0.03645694 -3.707467
(Dispersion Parameter for Binomial family taken to be 1 )
    Null Deviance: 144.206 on 109 degrees of freedom
Residual Deviance: 128.1772 on 108 degrees of freedom
Number of Fisher Scoring Iterations: 3

Correlation of Coefficients:
  (Intercept)
x -0.9922685
> predict(sample.fit, type="response", se=T, newdata=data.frame(x=42))
$fit:
         1
 0.7557034
$se.fit:
          1
 0.05133112
$residual.scale:
[1] 1
$df:
[1] 108
```

Figure 24. S-PLUS Output for the Logistic Regression Model with Sample Data from Table 5

While the 95% Wilson and Agresti-Coull confidence intervals based on these five observations are $(0.376, 0.964)$ and $(0.359, 0.975)$ respectively, the model-based 95% confidence interval is $(0.642, 0.842)$. The first thing that draws attention in this example is that the standard error for the sample proportion (0.179) is considerably greater than the one for the model-based estimate (0.051). Logistic regression estimates are much more precise in cases where the logistic regression model is appropriate because all 110 observations are used to estimate the two model parameters. In contrast, only five observations are used to estimate each binomial proportion.

C. COVERAGE PERFORMANCE OF LOGISTIC REGRESSION MODEL-BASED CONFIDENCE INTERVALS

When constructing a confidence interval, one usually wants the actual coverage probability to be close to the nominal confidence level. In this section, we will analyze the coverage performance of large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model for varying sample sizes. For simplicity, the model used in the simulations is the same as the one that was used in Chapter III. Software written in the S-PLUS language to compute coverage probabilities is presented in Appendix F.

Table 6 reports the average coverage probabilities and corresponding root MSEs for three different situations. In the first situation, similar to the original data, the total number of observations was set to 101. To see the effect of reducing the number of observations on coverage

probabilities, the total number of observations was then set to 51 and 26 for the second and third trials respectively.

Dose Level	Number of Observations at Each Dose Level	Total Number of Observations	Average Coverage Probability	Root MSE of Coverage Probabilities
101	1	101	0.9615	0.0097
51	1	51	0.9708	0.0228
26	1	26	0.9734	0.0319

Table 6. Numerical Results Indicating the Effect of Reducing the Number of Observations on the Coverage Performance of the 95% Large Sample Confidence Interval

As observed from Table 6, reducing the number of observations causes the average coverage probability to go up gradually. Root MSEs of coverage probabilities also indicate that the variability about the nominal confidence level gets larger as the number of observations is reduced. Briefly, the less number of observations the model has, the more conservative intervals it produces.

To illustrate the general characteristics of coverage probabilities at three different dose levels and the effect of these on the length of the confidence intervals, Figure 25 plots both the coverage probabilities and the mean confidence interval lengths as a function of p.[4]

[4] DL: Dose level, Obs: Number of observations.

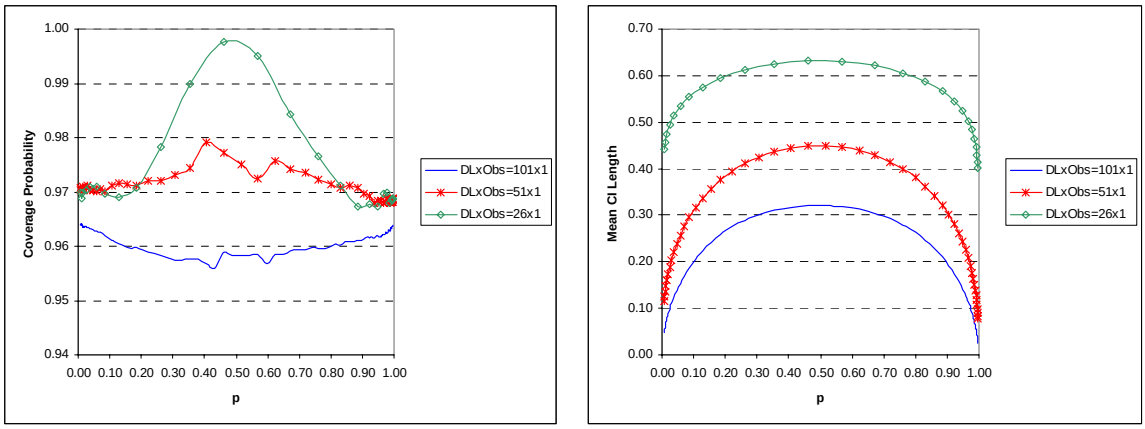

Figure 25. Coverage Probabilities and Mean Lengths of the 95% Confidence Interval for the Estimated Response as a Function of p for Different Dose Levels with One Observation at each Dose Level

The plots in Figure 25 suggest that as the coverage probabilities get farther away from the nominal confidence level, the confidence intervals tend to become wider.

Figure 26, on the other hand, illustrates the effect of changing the experimental design on both the coverage performance and the mean confidence interval lengths. Instead of obtaining one observation at each of the 101 dose levels, we reduced the number of dose levels to 51 and obtained two observations at each of these 51 dose levels. In this design, while the reported average coverage probability is 0.9614, the root MSE is 0.0096 – almost identical to the corresponding values in the case where there is one observation at each of the 101 dose levels. Besides, note that the design change had almost no effect on the mean length of the confidence intervals.

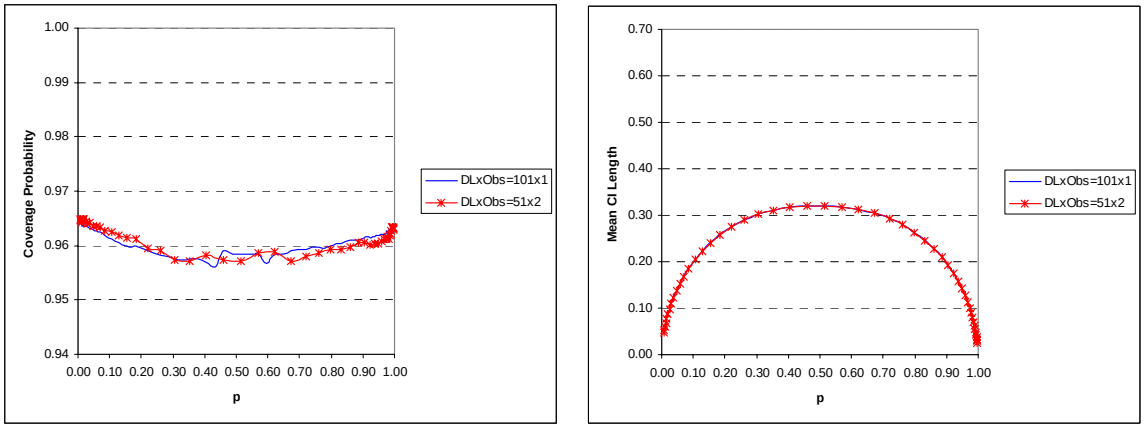

Figure 26. The Effect of Doubling the Observations When Dose Level is 51

The examples and illustrations given so far provide a general idea about the precision of logistic regression model-based estimators and the coverage probabilities of confidence intervals for a probability based on the fit of a simple logistic regression model. Based on these findings, in the next two sections we will continue our analysis in more detail and answer the sample size question using the models obtained from the real data sets.

D. MATHEMATICAL MODELS USED IN SIMULATIONS

Following the analysis of three different data sets provided by the U.S. Army Proving Ground, we obtained three different mathematical models for use in our computer simulations. Each of these models, in fact, revealed similar features in common.

The first similar feature is that all the models are quite close to piecewise linear logistic regression models that in general can be given by

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + \beta_2 (x-a)_+ + \beta_3 (x-b)_+$$

where

$$(x-a)_+ = \begin{cases} x-a & \text{if } x > a \\ 0 & otherwise \end{cases}$$

$$(x-b)_+ = \begin{cases} x-b & \text{if } x > b \\ 0 & otherwise \end{cases}$$

The second similar feature is that in all three models, p is approximately one for $p < a$ and is approximately zero for $p > b$. Only in the middle section $a \leq x \leq b$ does p vary. Besides, in this middle section, the logit of p is approximately linear in x. The primary differences in the models fit to the three data sets are the values of a and b. The second feature is, in fact, worth mentioning. The simulations, in order to check the adequacy of confidence intervals for a probability based on the fit of a simple linear logistic regression model in terms of their coverage probabilities, rely heavily on the model fitted to the synthetic data sets generated by using the mathematical models stated above. The fact that the probabilities in the first and the last pieces (sections or range intervals) are fairly constant causes the simulated responses to be mostly ones in the first section and zeros in the last section. Therefore, a piecewise linear logistic model with four parameters cannot be fitted to most of the synthetic data sets nicely throughout the simulation. When examined closely, it is seen that the parameter estimates and their corresponding standard errors tend to become quite large. In regards to the warning messages about the non-convergence of the iterative process when using a computer package to fit linear logistic models to binary data, Collett (1991) states, "the most likely cause of this phenomenon is that the model is an exact fit to certain binary observations..." (p. 82).

Similar problems also arise when a simple linear logistic regression model with two parameters is fitted separately for the first and the last pieces. Therefore, what we are interested in is to focus on the middle piece, and to analyze the coverage probabilities of confidence intervals in this region for varying samples sizes in different experimental designs.

E. ANSWERING THE SAMPLE SIZE QUESTION THROUGH SIMULATION

As stated in the introduction of this chapter, we look at the problem more empirically. Our approach to sample size determination is to perform a controlled set of simulations for different experimental designs. The first experimental design concerns a design where the dose levels are equally spaced within the experimental region of interest. The second experimental design concerns a design where the dose levels are unequally spaced. In both the first and second design, the number of observations at each dose level is the same. The third experimental design, on the other hand, is a design where the number of observations at unequally spaced dose levels varies. There are in fact two main reasons for setting up three different experimental designs in this study. The first one is the fact that it might not always be possible for the U.S. Army Yuma Proving Ground engineers to obtain observations at equally spaced dose levels, or to obtain the same number of observations at each dose level. The second one is the need to detect whether or not the coverage probabilities are affected considerably by design changes.

For the most part, the simulation results for all of the three models are similar for each of the experimental designs. Therefore, in this chapter, we will present the results pertaining to only one model.

Within the context of the first experimental design, while Table 7 reports summary statistics for eight different set of simulations, Figures 27 and 28 plot the coverage probabilities as a function of p and the mean confidence interval lengths as a function of range respectively.

Number of Observations at each Dose Level	Total Number of Observations	Average CP	Root MSE	Min.CP	Min. CI Length	Max. CI Length
1	33	0.9670	0.1026	0.9544	0.35	0.52
2	66	0.9568	0.0397	0.9534	0.25	0.39
3	99	0.9541	0.0249	0.9518	0.20	0.32
4	132	0.9526	0.0186	0.9485	0.17	0.28
5	165	0.9517	0.0108	0.9495	0.15	0.26
6	198	0.9517	0.0116	0.9498	0.14	0.24
10	330	0.9496	0.0079	0.9473	0.11	0.18
15	495	0.9503	0.0069	0.9488	0.09	0.15

Table 7. Simulation Results for Model 1 Under the First Experimental Design

As can be seen from the table and the figures, when the number of observations at each dose level is one (i.e., sample size is 33), the coverage probabilities tend to be quite above the nominal confidence level of 95%, while

having considerable variability. Besides, the minimum and the maximum mean lengths of the confidence intervals turn out to be too large.

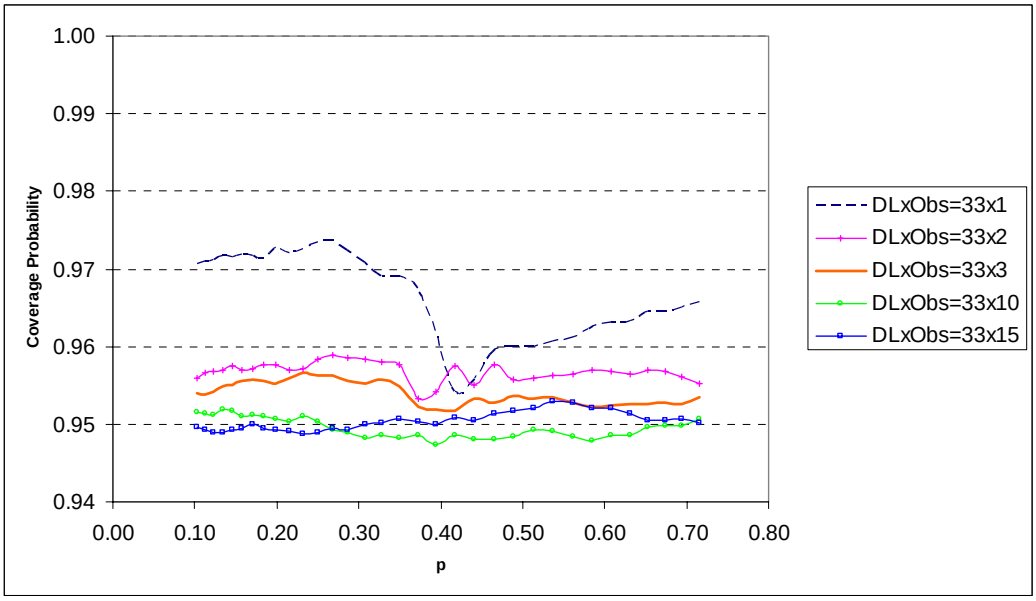

Figure 27. Coverage Probabilities for the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the First Experimental Design

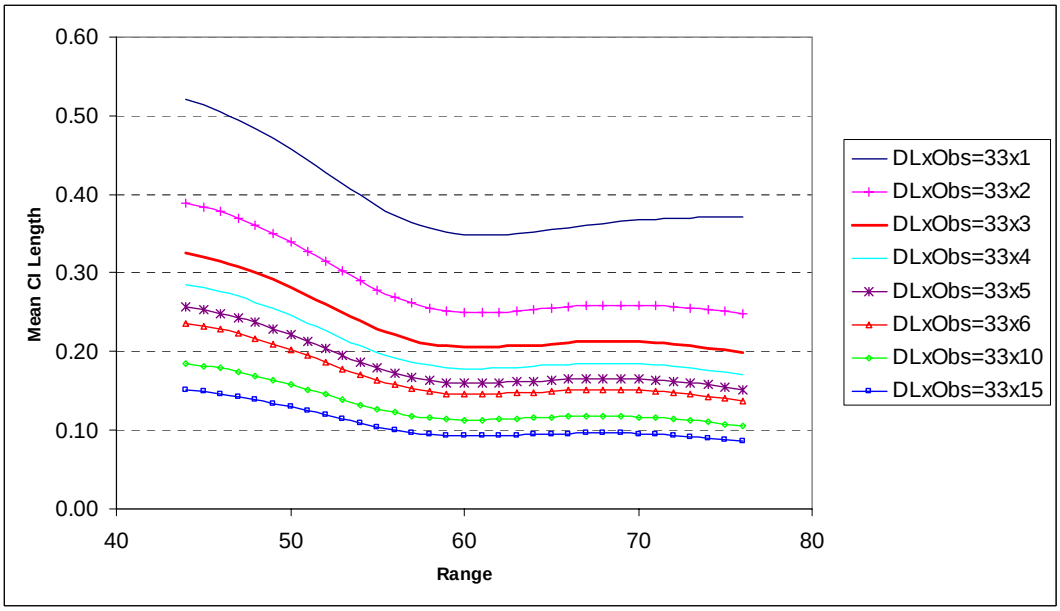

Figure 28. Mean Length of the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the First Experimental Design

As the number of observations within the experimental region of interest increases, the simulation results for the first experimental design suggest the following:

- The coverage probabilities of confidence intervals for a probability based on the fit of a simple linear logistic regression model move closer to the nominal confidence level of 95%.
- The variability of coverage probabilities about the nominal confidence level also gets smaller with the increase in sample size. For instance, when the number of observations at each dose level is one, the root MSE is 0.1026, which is considerably high when compared with those of other sample sizes.
- Although the coverage probabilities may fall below the nominal confidence level for large sample sizes, they are typically very close to that level. For instance, the smallest of the minimum coverage probabilities in Table 7 is 0.9473, when the number of observations at each dose level is set to 10.
- Besides coverage, length is also very important in the evaluation of a confidence interval. As can be seen in Figure 28, the model produces narrower confidence intervals while the increase in sample size improves the coverage probabilities. However, the rate at which the confidence intervals get narrower turns out to be decreasing.

Simulation results for the second and the third experimental designs are also in accordance with those stated above. See Table 8 and Table 9 for summary statistics and Figures 29 through 32 for the coverage probabilities as a function of p and the mean confidence interval lengths as a function of range for these designs.

Number of Observations at each Dose Level	Total Number of Observations	Average CP	Root MSE	Min.CP	Min. CI Length	Max. CI Length
1	33	0.9654	0.0926	0.9589	0.34	0.55
2	66	0.9564	0.0407	0.9505	0.25	0.42
3	99	0.9538	0.0229	0.9515	0.20	0.35
4	132	0.9524	0.0169	0.9502	0.18	0.31
5	165	0.9532	0.0190	0.9512	0.16	0.28
6	198	0.9525	0.0154	0.9509	0.14	0.25
10	330	0.9498	0.0086	0.9469	0.11	0.20
15	495	0.9486	0.0094	0.9470	0.09	0.16

Table 8. Simulation Results for Model 1 Under the Second Experimental Design

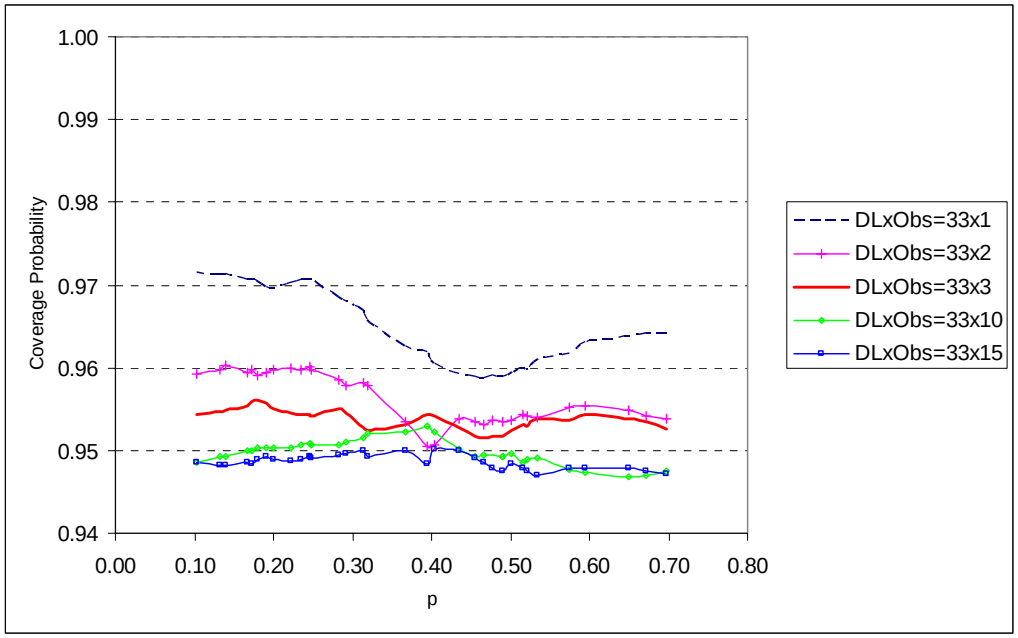

Figure 29. Coverage Probabilities for the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Second Experimental Design

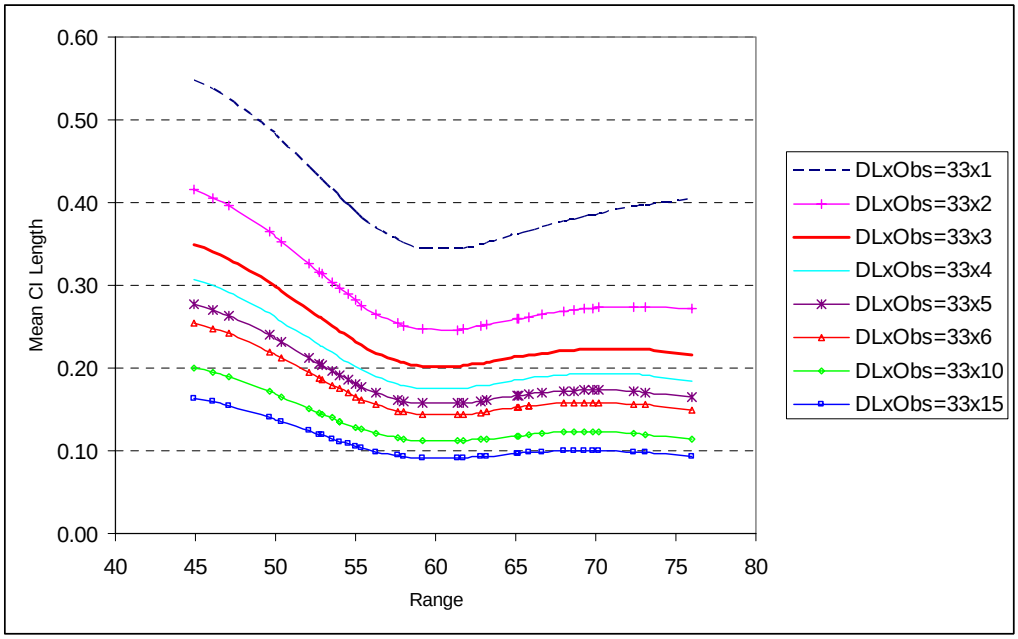

Figure 30. Mean Length of the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Second Experimental Design

Number of Observations at each Dose Level	Total Number of Observations	Average CP	Root MSE	Min.CP	Min. CI Length	Max. CI Length
Varies	33	0.9615	0.0667	0.9543	0.35	0.49
Varies	66	0.9560	0.0530	0.9510	0.25	0.40
Varies	99	0.9521	0.0241	0.9477	0.19	0.34
Varies	132	0.9534	0.0382	0.9507	0.16	0.29
Varies	165	0.9534	0.0416	0.9436	0.16	0.26
Varies	198	0.9521	0.0288	0.9507	0.14	0.25
Varies	330	0.9519	0.0308	0.9499	0.11	0.18
Varies	495	0.9508	0.0166	0.9497	0.09	0.16

Table 9. Simulation Results for Model 1 Under the Third Experimental Design

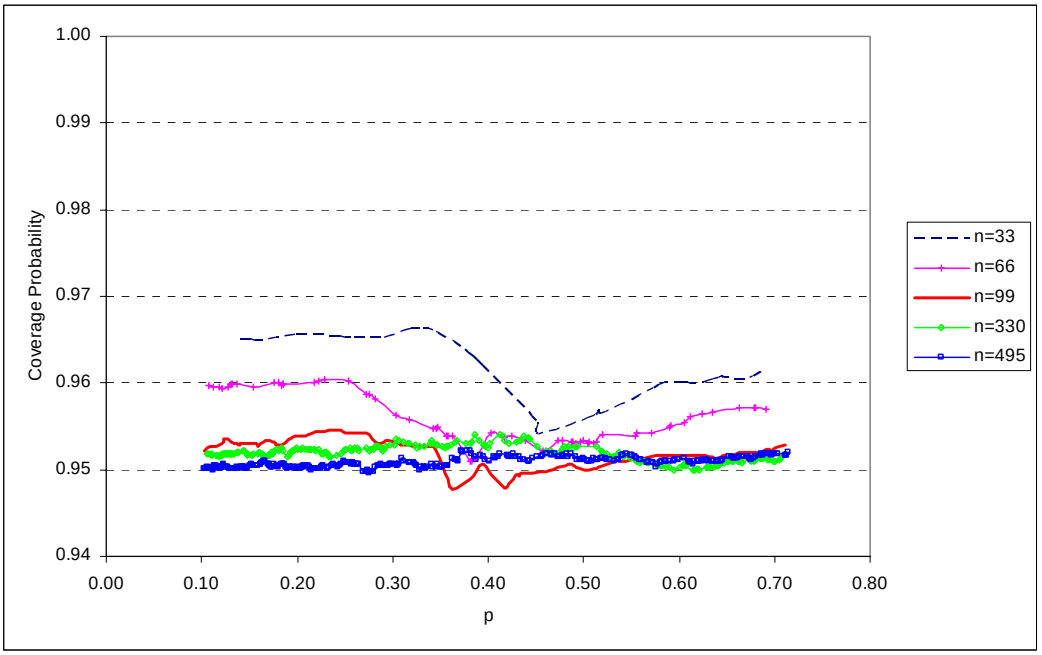

Figure 31. Coverage Probabilities for the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Third Experimental Design

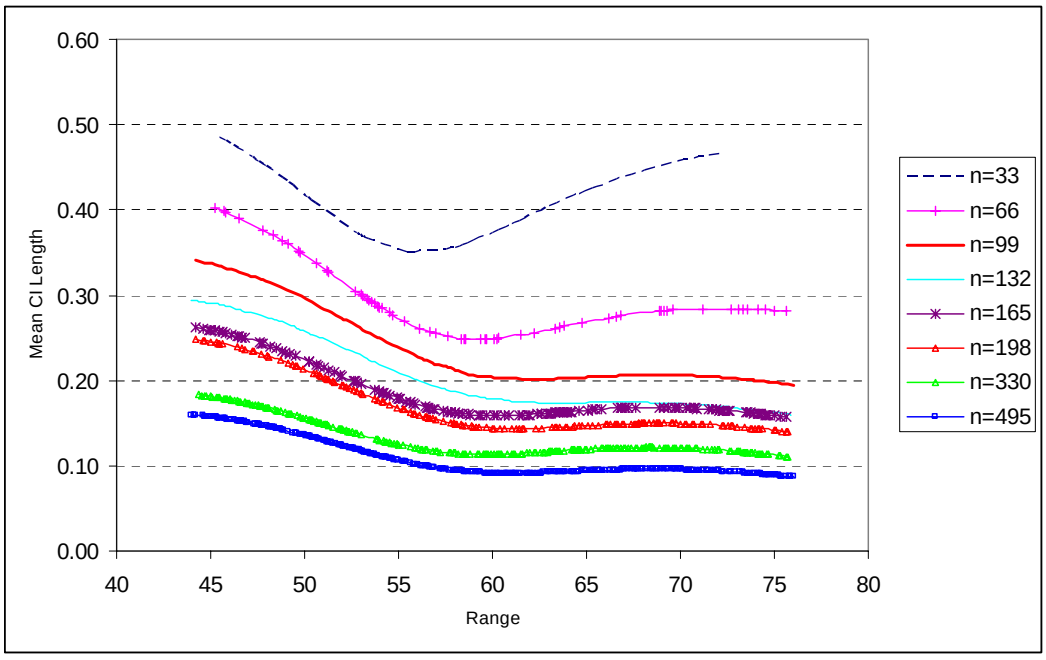

Figure 32. Mean Length of the 95% Confidence Interval Based on the Fit of a Simple Linear Logistic Regression Model Under the Third Experimental Design

In order to evaluate if the true average coverage probabilities are affected by the experimental design change, we carried out an analysis of variance F test at significance level 0.05. Although the evidence allows us to conclude that the true average coverage probability depends on the experimental design, we assess that there is not a practical difference, because an acceptable level of coverage performance is achieved especially when the sample size is increased within the experimental region of interest.

In the light of the evidence gathered so far, we suggest that under any of the three experimental designs, the Yuma Proving Ground engineers obtain at least 100 observations within the experimental region of interest where the probability of detection does not remain constant. If the goal is to produce narrower confidence intervals together with more improved coverage probabilities, then the number of observations can go up to 500 depending on the budget and time allocated to the experiment.

As a continuation of our study, we also compared the coverage probabilities of large-sample confidence intervals for a probability based on the fit of a simple logistic regression model with those of the nonparametric bootstrap confidence intervals. In this regard, the next section provides a comparison when the sample size is 66 within the context of the first experimental design.

F. COMPARING THE COVERAGE PERFORMANCE OF LARGE-SAMPLE AND NONPARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS

According to Efron and Tibshirani (1993), one of the principal goals of the bootstrap theory is to produce good confidence intervals automatically. "Good" means that the bootstrap intervals should closely match exact confidence intervals in those special situations where statistical theory yields an exact answer, and should give dependably accurate coverage probabilities in all situations. Among the several methods for confidence interval construction using the bootstrap, the nonparametric Bca (bias-corrected and accelerated) confidence intervals are presented as a substantial improvement over the percentile method in both theory and practice, and are said to come close to the criteria stated above, though their coverage probabilities can still be erratic for small sample sizes.

Due to their improved performance, we chose to compare the coverage probabilities of nonparametric Bca confidence intervals with those of large-sample confidence intervals. The software written in the S-PLUS language to compute the coverage probabilities of the nominal 95% Bca intervals is in Appendix G. Figure 33 plots the coverage probabilities for the 95% large-sample and the Bca confidence intervals for a probability based on the fit of a simple logistic regression model under the first experimental design when the sample size is 66.

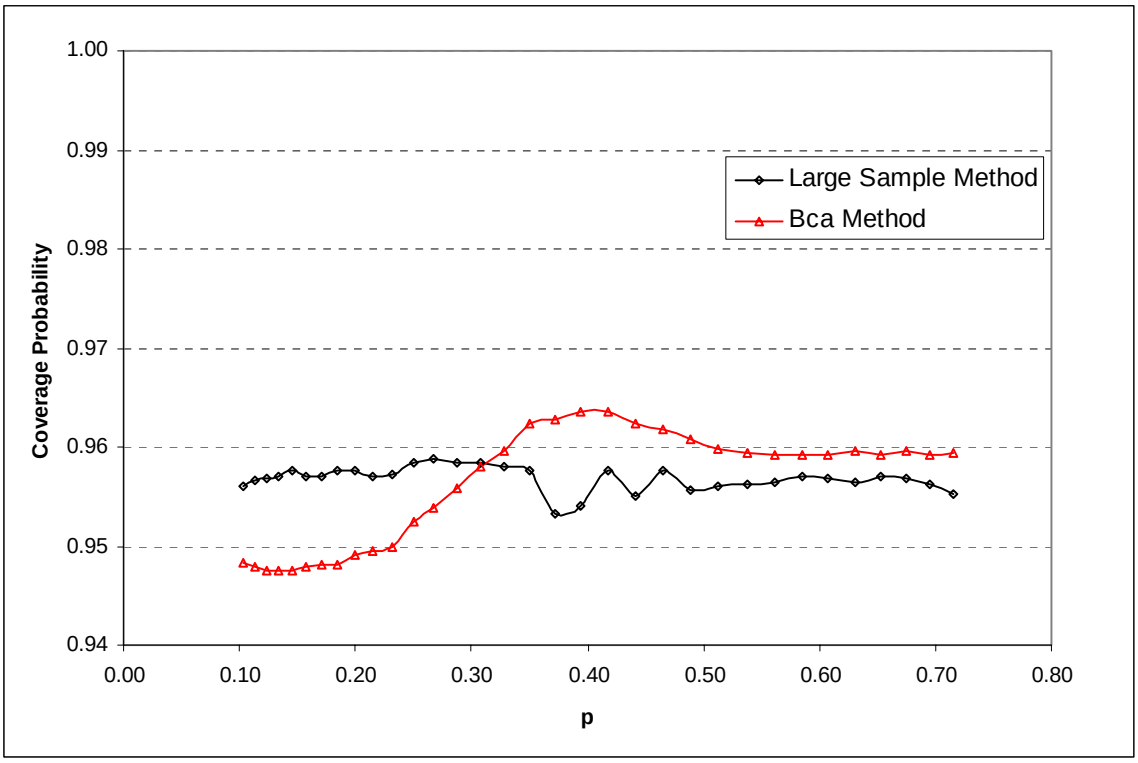

Figure 33. Coverage Probabilities of the 95% Large Sample and Bca Confidence Intervals Based on the Fit of a Simple Linear Logistic Regression Model When $n = 66$

According to the simulation results, the average coverage probability of the Bca confidence interval is 0.9558, and the root MSE of the coverage probabilities is 0.0473. When these values are compared with those of the large-sample confidence interval (0.9568 and 0.0397 respectively), it turns out that both methods are competitive. However, the coverage performance of the large-sample confidence interval seems better than that of the Bca confidence interval. As can be seen from Figure 33, while the Bca interval has coverage probabilities less than the large-sample interval when $0.103 < p < 0.307$, it remains a little bit conservative when $0.328 < p < 0.715$. Our evaluations at this point show that for the recommended

sample sizes within the experimental region of interest, the large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model perform well in terms of their coverage probabilities as long as the logistic regression model is fitted to the data carefully.

G. CHAPTER SUMMARY

In this chapter, we first showed that the logistic regression model-based estimator of probability is considerably better than the sample proportion. With this motivation in mind, we then examined the coverage probabilities of large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model for varying sample sizes within the experimental region of interest under three different experimental designs. The first of the two main reasons for setting up three different experimental designs in this study was the fact that it might not always be possible for the Yuma Proving Ground engineers to obtain observations at equally spaced dose levels, or to obtain the same number of observations at each dose level. The second reason was the need to detect if the coverage probabilities would be affected considerably by design change. Lastly, we compared the coverage probabilities of large-sample confidence intervals with those of nonparametric Bca confidence intervals to cross-validate our results.

Based on our evaluations, some of the important conclusions reached are as follows.

- When the model approximates the true probabilities in a decent manner, logistic regression model-based estimators are more precise than the sample proportion-based estimators are.

- As the sample size increases within the experimental region of interest, the coverage probabilities of large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model tend to come closer to the nominal confidence level.
- From a practical point of view, experimental design changes do not have a considerable effect on the coverage probabilities of confidence intervals for a probability based on the fit of a simple linear logistic regression model.
- Large-sample and nonparametric Bca confidence intervals for a probability based on the fit of a simple linear logistic regression model are competitive in terms of their coverage probabilities.
- At least 100 observations should be obtained within the experimental region of interest in order to obtain good estimates of sensor detection probabilities.

V. CONCLUSION

A. CONCLUDING REMARKS

In this thesis, we approach the problem of sample size determination for estimation of sensor detection probabilities from two different aspects. First, we examine the problem within the context of a binomial experiment in order to improve the current estimation method used by the U.S. Army Yuma Proving Ground that considers only straight proportions within range intervals (***binning approach***). Using simulation, we evaluate the coverage probabilities and lengths of confidence intervals for binomial proportions and report the required sample sizes for some specified goals through the utilization of different methods. Second, again using simulation, we evaluate the coverage probabilities and lengths of confidence intervals based on logistic regression to obtain better estimates of the probability of detection with much smaller sample sizes.

Based on the findings through our analyses, our recommendations for the U.S. Army Yuma Proving Ground and some important conclusions reached are as follows:

- First and foremost, when the probability of detection at specified range intervals is estimated using the current ***binning approach***, we recommend that the U.S. Army Yuma Proving Ground engineers consider not only the sample proportions but also the confidence intervals for a binomial proportion. This is because confidence intervals are a fundamentally more ambitious measure of statistical accuracy than proportions. Even though the use of this approach provides estimates for range intervals rather than specific ranges and violates the fourth assumption of a binomial experiment as stated in

Section A of Chapter I, our simulations show that the recommended confidence intervals, namely the Agresti-Coull, Wilson, and equal-tailed Jeffreys prior intervals, perform well.

- Second, the U.S. Army Yuma Proving Ground engineers can use a parametric model so that they can obtain much more information out of their samples for the same sample sizes. An appropriate model in this case seems to be a piecewise linear logistic regression model dependent upon the analyses conducted on three data sets provided by the U.S. Army Yuma Proving Ground. Due to the reasons stated in Section D of Chapter IV, when this procedure is adopted estimation of sensor detection probabilities should focus on ranges where the probabilities do not remain constant. Our simulations under three different experimental designs show that large-sample confidence intervals for a probability based on the fit of a simple linear logistic regression model perform much better than the confidence intervals for a binomial proportion discussed in Chapter II in terms of their coverage probabilities and lengths. Besides, nonparametric Bca confidence intervals for a probability based on the fit of a simple linear logistic regression model also confirm our results.

- Finally, in order to get good estimates of sensor detection probabilities at a significance level of 0.05, we recommend that the U.S. Army Yuma Proving Ground engineers use a simple linear logistic regression model and obtain at least 100 observations within the experimental region of interest where the probabilities do not remain constant. In the other two regions, where the probabilities remain almost constant, we assess that the current ***binning approach*** that has been taken by the U.S. Army Yuma Proving Ground is appropriate as long as the issues discussed in Chapter II are kept in mind.

B. FURTHER STUDY SUGGESTIONS

- Due to the data provided by the U.S. Army Yuma Proving Ground, we restricted our analyses only to one predictor variable, namely range. A further study may attempt to answer the sample

size question considering other factors such as type and radar cross section of aircraft together with range within the context of a logistic regression.

- In response to the primary thesis question, we adopted an empirical approach based on a controlled set of simulations. Another further study, on the other hand, may focus on the proper choice of designs needed to fit logistic regression models. By design we mean the determination of the settings of the predictor variables that result in adequate predictions of the response of interest throughout the experimental region. That is, a further study may focus on optimally selecting the number of dose levels (ranges at which observations are taken) within the experimental region, and then determining the number of observations at each of these dose levels with respect to a given optimality criterion for a fixed sample size. Refer to Khuri et al. (2006) for a detailed discussion about the approaches to solving such design problems.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE WALD INTERVAL

```
function(n = 5, bin.number = 20, nrep = 100000, alpha = 0.05)
{
	x.t <- seq(-6, 5, 11/(bin.number * n))
	x <- x.t[-1]
	z <- qnorm(1 - alpha/2)
	#1. CREATE A MATRIX WHOSE ROWS CONTAIN nrep BERNOULLI R.V.'s
	y.mat <- matrix(nrow = length(x), ncol = nrep)
	for(i in 1:length(x)) {
		y.mat[i,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(x[i])))
	}
	#2. COMPUTATION OF nrep phats FOR EACH BIN OF LENGTH n,
	#   AND STORING THEM IN A bin.number x nrep MATRIX
	lb <- seq(1, length(x) - n + 1, n)
	ub <- seq(n, length(x), n)
	p.hat.mat <- matrix(nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		p.hat.mat[i,  ] <- apply(y.mat[lb[i]:ub[i],  ], MARGIN = 2, mean)
	}
	#3. COMPUTATION OF (1-alpha)100 WALD CONFIDENCE INTERVALS
	l.mat <- matrix(nrow = bin.number, ncol = nrep)
	u.mat <- matrix(nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		l.mat[i,  ] <- p.hat.mat[i,  ] - z * sqrt((p.hat.mat[i,  ] *
			(1 - p.hat.mat[i,  ]))/n)
		u.mat[i,  ] <- p.hat.mat[i,  ] + z * sqrt((p.hat.mat[i,  ] *
			(1 - p.hat.mat[i,  ]))/n)
	}
	# Replace values that are greater than 1 with 1.0,
	# and values that are less than 0 with 0.0
	lo.mat <- replace(l.mat[], which(l.mat[] < 0), 0)
	up.mat <- replace(u.mat[], which(u.mat[] > 1), 1)
	#4. COMPUTE THE CONFIDENCE INTERVAL WIDTHS FOR PHASE 1
	width.mat <- up.mat - lo.mat
	mean.width.mat <- as.matrix(apply(width.mat, 1, mean))
	#5. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 1
	mean.lo.mat <- as.matrix(apply(lo.mat, 1, mean))
	mean.up.mat <- as.matrix(apply(up.mat, 1, mean))
	#6. COMPUTE THE COVERAGE PROBABILITIES FOR PHASE 1
	p.i.vector <- 1/(1 + exp(x))
	p.i.mat <- matrix(p.i.vector, nrow = n, ncol = bin.number)
	cp.mat <- matrix(nrow = n, ncol = bin.number)
	for(i in 1:bin.number) {
		for(j in 1:n) {
			cp.mat[j, i] <- sum((lo.mat[i,  ] < p.i.mat[j, i]) &
				(p.i.mat[j, i] < up.mat[i,  ]))/nrep
		}
	}
	cp.vector <- as.vector(cp.mat)
	#7. PLOT THE COVERAGE PROBABILITIES AS A FUNCTION OF p
	plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
		"Coverage Probability", ylim = c(0, 1))
	title(sub = "Method used: The Wald interval")
	abline(1 - alpha, 0, col = 5)
	#8. REARRANGE LOWER CONFIDENCE LIMITS FOR PHASE 2
	new.lo.mat <- lo.mat
	max.fn <- function(k, lo.mat)
	{
```

```
        n.row <- dim(lo.mat)[1]
        apply(lo.mat[k:n.row,  ], MARGIN = 2, max)
}
new.lo.mat[1:dim(lo.mat)[1] - 1,  ] <- t(sapply(1:(dim(lo.mat)[1] -
        1), max.fn, lo.mat = lo.mat))
#9. REARRANGE UPPER CI's FOR PHASE 2
new.up.mat <- up.mat
min.fn <- function(k, up.mat)
{
        apply(up.mat[k:1,  ], 2, min)
}
new.up.mat[2:dim(up.mat)[1],  ] <- t(sapply(2:dim(up.mat)[1], min.fn,
        up.mat = up.mat))
#10. COMPUTE THE NEW CONFIDENCE INTERVAL WIDTHS FOR PHASE 2
new.width.mat <- new.up.mat - new.lo.mat
new.mean.width.mat <- as.matrix(apply(new.width.mat, 1, mean))
#11. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 2
new.mean.lo.mat <- as.matrix(apply(new.lo.mat, 1, mean))
new.mean.up.mat <- as.matrix(apply(new.up.mat, 1, mean))
#12. COMPUTE THE NEW COVERAGE PROBABILITIES FOR PHASE 2
new.cp.mat <- matrix(nrow = n, ncol = bin.number)
for(i in 1:bin.number) {
        for(j in 1:n) {
                new.cp.mat[j, i] <- sum((new.lo.mat[i,  ] < p.i.mat[
                        j, i]) & (p.i.mat[j, i] < new.up.mat[i,  ]))/nrep
        }
}
new.cp.vector <- as.vector(new.cp.mat)
#13. PLOT LOWER AND UPPER CONFIDENCE LIMITS
mean.lo.vector <- as.vector(mean.lo.mat)
mean.up.vector <- as.vector(mean.up.mat)
new.mean.lo.vector <- as.vector(new.mean.lo.mat)
new.mean.up.vector <- as.vector(new.mean.up.mat)
plot(1:bin.number, mean.lo.vector, type = "o", pch = 6, xlab = "Bin",
        ylab = "CI Limits")
title(sub = "Method used: The Wald interval")
points(1:bin.number, mean.up.vector, type = "o", pch = 2)
points(1:bin.number, new.mean.lo.vector, type = "o", pch = 6, col = 6)
points(1:bin.number, new.mean.up.vector, type = "o", pch = 2, col = 6)
legend(13, 0.97, c("Upper CL", "Lower CL", "New Upper CL",
        "New Lower CL"), marks = c(2, 6, 2, 6), col = c(1, 1, 6, 6))
#14. PLOT THE OLD & THE NEW COVERAGE PROBABILITIES AS A FUNCTION OF p
plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
        "Coverage Probability", ylim = c(0, 1))
title(sub = "Method used: The Wald interval")
points(p.i.vector, new.cp.vector, type = "o", pch = 2, col = 6)
abline(1 - alpha, 0, col = 5)
#15. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 1
target <- rep(1 - alpha, length(x))
mse <- (rev(cp.vector) - target)^2
a.mse <- rep(0, each = length(mse))
p <- rev(p.i.vector)
for(i in 1:(length(mse) - 1)) {
        a.mse[i + 1] <- 0.5 * (mse[i] + mse[i + 1]) * (p[i + 1] - p[
                i])
}
RMSE <- sqrt(sum(a.mse))
#16. MEAN COVERAGE PROBABILITY for PHASE 1
cp <- rev(cp.vector)
mcp <- rep(0, length(cp))
for(i in 1:(length(cp) - 1)) {
        mcp[i + 1] <- 0.5 * (cp[i] + cp[i + 1]) * (p[i + 1] - p[i])
}
```

```
	MCP <- sum(mcp)
	#17. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 2
	mse.new <- (rev(new.cp.vector) - target)^2
	a.mse.new <- rep(0, each = length(mse.new))
	for(i in 1:(length(mse.new) - 1)) {
		a.mse.new[i + 1] <- 0.5 * (mse.new[i] + mse.new[i + 1]) * (
			p[i + 1] - p[i])
	}
	RMSE.new <- sqrt(sum(a.mse.new))
	#18. MEAN COVERAGE PROBABILITY for PHASE 2
	cp.new <- rev(new.cp.vector)
	mcp.new <- rep(0, length(cp.new))
	for(i in 1:(length(cp.new) - 1)) {
		mcp.new[i + 1] <- 0.5 * (cp.new[i] + cp.new[i + 1]) * (p[i +
			1] - p[i])
	}
	MCP.new <- sum(mcp.new)
	#19. RETURN RESULTS
	Table.1 <- data.frame("Mean Lower Limit" = mean.lo.mat,
		"Mean Upper Limit" = mean.up.mat, "Mean CI Width" =
		mean.width.mat)
	Table.2 <- data.frame("Mean Lower Limit" = new.mean.lo.mat,
		"Mean Upper Limit" = new.mean.up.mat, "Mean CI Width" =
		new.mean.width.mat)
	Table.3 <- data.frame(Root.MSE = RMSE, Mean.CP = MCP, Root.MSE.New =
		RMSE.new, Mean.CP.New = MCP.new)
	return(t(cp.mat), t(new.cp.mat), Table.1, Table.2, Table.3)
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE WILSON INTERVAL

```
function(n = 5, bin.number = 20, nrep = 100000, alpha = 0.05)
{
	x.t <- seq(-6, 5, 11/(bin.number * n))
	x <- x.t[-1]
	z <- qnorm(1 - alpha/2)
	#1. CREATE A MATRIX WHOSE ROWS CONTAIN nrep BERNOULLI R.V.'s
	y.mat <- matrix(nrow = length(x), ncol = nrep)
	for(i in 1:length(x)) {
		y.mat[i,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(x[i])))
	}
	#2. COMPUTATION OF nrep phats FOR EACH BIN OF LENGTH n,
	#   AND STORING THEM IN A bin.number x nrep MATRIX
	lb <- seq(1, length(x) - n + 1, n)
	ub <- seq(n, length(x), n)
	p.hat.mat <- matrix(nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		p.hat.mat[i,  ] <- apply(y.mat[lb[i]:ub[i],  ], MARGIN = 2, mean)
	}
	#3. COMPUTATION OF (1-alpha)100% WILSON CONFIDENCE INTERVALS
	lo.mat <- matrix(nrow = bin.number, ncol = nrep)
	up.mat <- matrix(nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		lo.mat[i,  ] <- (p.hat.mat[i,  ] + z^2/(2 * n) - z * sqrt(
			(p.hat.mat[i,  ] * (1 - p.hat.mat[i,  ]))/n + z^2/
			(4 * n^2)))/(1 + z^2/n)
		up.mat[i,  ] <- (p.hat.mat[i,  ] + z^2/(2 * n) + z * sqrt(
			(p.hat.mat[i,  ] * (1 - p.hat.mat[i,  ]))/n + z^2/
			(4 * n^2)))/(1 + z^2/n)
	}
	#4. COMPUTE THE CONFIDENCE INTERVAL WIDTHS FOR PHASE 1
	width.mat <- up.mat - lo.mat
	mean.width.mat <- as.matrix(apply(width.mat, 1, mean))
	#5. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 1
	mean.lo.mat <- as.matrix(apply(lo.mat, 1, mean))
	mean.up.mat <- as.matrix(apply(up.mat, 1, mean))
	#6. COMPUTE THE COVERAGE PROBABILITIES FOR PHASE 1
	p.i.vector <- 1/(1 + exp(x))
	p.i.mat <- matrix(p.i.vector, nrow = n, ncol = bin.number)
	cp.mat <- matrix(nrow = n, ncol = bin.number)
	for(i in 1:bin.number) {
		for(j in 1:n) {
			cp.mat[j, i] <- sum((lo.mat[i,  ] < p.i.mat[j, i]) &
				(p.i.mat[j, i] < up.mat[i,  ]))/nrep
		}
	}
	cp.vector <- as.vector(cp.mat)
	#7. PLOT THE COVERAGE PROBABILITIES AS A FUNCTION OF p
	plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
		"Coverage Probability", ylim = c(0, 1))
	title(sub = "Method used: The Wilson interval")
	abline(1 - alpha, 0, col = 5)
	#8. REARRANGE LOWER CI's FOR PHASE 2
	new.lo.mat <- lo.mat
	max.fn <- function(k, lo.mat)
	{
		n.row <- dim(lo.mat)[1]
		apply(lo.mat[k:n.row,  ], MARGIN = 2, max)
```

```
}
new.lo.mat[1:dim(lo.mat)[1] - 1,  ] <- t(sapply(1:(dim(lo.mat)[1] -
      1), max.fn, lo.mat = lo.mat))
#9. REARRANGE UPPER CI's FOR PHASE 2
new.up.mat <- up.mat
min.fn <- function(k, up.mat)
{
      apply(up.mat[k:1,  ], 2, min)
}
new.up.mat[2:dim(up.mat)[1],  ] <- t(sapply(2:dim(up.mat)[1], min.fn,
      up.mat = up.mat))
#10. COMPUTE THE NEW CONFIDENCE INTERVAL WIDTHS FOR PHASE 2
new.width.mat <- new.up.mat - new.lo.mat
new.mean.width.mat <- as.matrix(apply(new.width.mat, 1, mean))
#11. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 2
new.mean.lo.mat <- as.matrix(apply(new.lo.mat, 1, mean))
new.mean.up.mat <- as.matrix(apply(new.up.mat, 1, mean))
#12. COMPUTE THE NEW COVERAGE PROBABILITIES FOR PHASE 2
new.cp.mat <- matrix(nrow = n, ncol = bin.number)
for(i in 1:bin.number) {
      for(j in 1:n) {
            new.cp.mat[j, i] <- sum((new.lo.mat[i,  ] < p.i.mat[
                  j, i]) & (p.i.mat[j, i] < new.up.mat[i,  ]))/nrep
      }
}
new.cp.vector <- as.vector(new.cp.mat)
#13. PLOT LOWER AND UPPER CONFIDENCE LIMITS
mean.lo.vector <- as.vector(mean.lo.mat)
mean.up.vector <- as.vector(mean.up.mat)
new.mean.lo.vector <- as.vector(new.mean.lo.mat)
new.mean.up.vector <- as.vector(new.mean.up.mat)
plot(1:bin.number, mean.lo.vector, type = "o", pch = 6, xlab = "Bin",
      ylab = "CI Limits", ylim = c(0, 1))
title(sub = "Method used: The Wilson interval")
points(1:bin.number, mean.up.vector, type = "o", pch = 2)
points(1:bin.number, new.mean.lo.vector, type = "o", pch = 6, col = 6)
points(1:bin.number, new.mean.up.vector, type = "o", pch = 2, col = 6)
legend(13, 0.97, c("Upper CL", "Lower CL", "New Upper CL",
      "New Lower CL"), marks = c(2, 6, 2, 6), col = c(1, 1, 6, 6))
#14. PLOT THE OLD & THE NEW COVERAGE PROBABILITIES AS A FUNCTION OF p
plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
      "Coverage Probability", ylim = c(0, 1))
title(sub = "Method used: The Wilson interval")
points(p.i.vector, new.cp.vector, type = "o", pch = 2, col = 6)
abline(1 - alpha, 0, col = 5)
#15. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 1
target <- rep(1 - alpha, length(x))
mse <- (rev(cp.vector) - target)^2
a.mse <- rep(0, each = length(mse))
p <- rev(p.i.vector)
for(i in 1:(length(mse) - 1)) {
      a.mse[i + 1] <- 0.5 * (mse[i] + mse[i + 1]) * (p[i + 1] - p[
            i])
}
RMSE <- sqrt(sum(a.mse))
#16. MEAN COVERAGE PROBABILITY for PHASE 1
cp <- rev(cp.vector)
mcp <- rep(0, length(cp))
for(i in 1:(length(cp) - 1)) {
      mcp[i + 1] <- 0.5 * (cp[i] + cp[i + 1]) * (p[i + 1] - p[i])
}
MCP <- sum(mcp)
#17. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 2
```

```
	mse.new <- (rev(new.cp.vector) - target)^2
	a.mse.new <- rep(0, each = length(mse.new))
	for(i in 1:(length(mse.new) - 1)) {
		a.mse.new[i + 1] <- 0.5 * (mse.new[i] + mse.new[i + 1]) * (
			p[i + 1] - p[i])
	}
	RMSE.new <- sqrt(sum(a.mse.new))
	#18. MEAN COVERAGE PROBABILITY for PHASE 2
	cp.new <- rev(new.cp.vector)
	mcp.new <- rep(0, length(cp.new))
	for(i in 1:(length(cp.new) - 1)) {
		mcp.new[i + 1] <- 0.5 * (cp.new[i] + cp.new[i + 1]) * (p[i +
			1] - p[i])
	}
	MCP.new <- sum(mcp.new)
	#19. RETURN RESULTS
	Table.1 <- data.frame("Mean Lower Limit" = mean.lo.mat,
		"Mean Upper Limit" = mean.up.mat, "Mean CI Width" =
		mean.width.mat)
	Table.2 <- data.frame("Mean Lower Limit" = new.mean.lo.mat,
		"Mean Upper Limit" = new.mean.up.mat, "Mean CI Width" =
		new.mean.width.mat)
	Table.3 <- data.frame(Root.MSE = RMSE, Mean.CP = MCP, Root.MSE.New =
		RMSE.new, Mean.CP.New = MCP.new)
	return(t(cp.mat), t(new.cp.mat), Table.1, Table.2, Table.3)
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE ADJUSTED WALD INTERVAL

```
function(n = 5, bin.number = 20, nrep = 100000, alpha = 0.05)
{
      x.t <- seq(-6, 5, 11/(bin.number * n))
      x <- x.t[-1]
      z <- qnorm(1 - alpha/2)
      #1. CREATE A MATRIX WHOSE ROWS CONTAIN nrep BERNOULLI R.V.'s
      y.mat <- matrix(nrow = length(x), ncol = nrep)
      for(i in 1:length(x)) {
            y.mat[i,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(x[i])))
      }
      #2.a. OBTAIN THE NUMBER OF SUCCESSES OUT OF n OBERVATIONS FOR EACH BIN
      lb <- seq(1, length(x) - n + 1, n)
      ub <- seq(n, length(x), n)
      num.suc.mat <- matrix(nrow = bin.number, ncol = nrep)
      for(i in 1:bin.number) {
            num.suc.mat[i,  ] <- apply(y.mat[lb[i]:ub[i],  ], MARGIN = 2,
                  sum)
      }
      #2.b. ADD TWO SUCCESSES TO EACH ELEMENT OF num.suc.mat
      adj.suc.mat <- num.suc.mat + 2
      #2.c. COMPUTE THE ADJUSTED p.hat BY DIVIDING EACH ELEMENT OF adj.suc.mat
 BY n+4
      adj.p.hat.mat <- adj.suc.mat/(n + 4)
      #3. COMPUTATION OF (1-alpha)100% ADJUDTED WALD CONFIDENCE INTERVALS
      l.mat <- matrix(nrow = bin.number, ncol = nrep)
      u.mat <- matrix(nrow = bin.number, ncol = nrep)
      for(i in 1:bin.number) {
            l.mat[i,  ] <- adj.p.hat.mat[i,  ] - z * sqrt((adj.p.hat.mat[
                  i,  ] * (1 - adj.p.hat.mat[i,  ]))/(n + 4))
            u.mat[i,  ] <- adj.p.hat.mat[i,  ] + z * sqrt((adj.p.hat.mat[
                  i,  ] * (1 - adj.p.hat.mat[i,  ]))/(n + 4))
      }
      # Replace values > 1 with one, and values < 0 with zero
      lo.mat <- replace(l.mat[], which(l.mat[] < 0), 0)
      up.mat <- replace(u.mat[], which(u.mat[] > 1), 1)
      #4. COMPUTE THE CONFIDENCE INTERVAL WIDTHS FOR PHASE 1
      width.mat <- up.mat - lo.mat
      mean.width.mat <- as.matrix(apply(width.mat, 1, mean))
      #5. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 1
      mean.lo.mat <- as.matrix(apply(lo.mat, 1, mean))
      mean.up.mat <- as.matrix(apply(up.mat, 1, mean))
      #6. COMPUTE THE COVERAGE PROBABILITIES FOR PHASE 1
      p.i.vector <- 1/(1 + exp(x))
      p.i.mat <- matrix(p.i.vector, nrow = n, ncol = bin.number)
      cp.mat <- matrix(nrow = n, ncol = bin.number)
      for(i in 1:bin.number) {
            for(j in 1:n) {
                  cp.mat[j, i] <- sum((lo.mat[i,  ] < p.i.mat[j, i]) &
                        (p.i.mat[j, i] < up.mat[i,  ]))/nrep
            }
      }
      cp.vector <- as.vector(cp.mat)
      #7. PLOT THE COVERAGE PROBABILITIES AS A FUNCTION OF p
      plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
            "Coverage Probability", ylim = c(0, 1))
      title(sub = "Method used: The Agresti-Coull interval")
      abline(1 - alpha, 0, col = 5)
```

```
#8. REARRANGE LOWER CI's FOR PHASE 2
new.lo.mat <- lo.mat
max.fn <- function(k, lo.mat)
{
       n.row <- dim(lo.mat)[1]
       apply(lo.mat[k:n.row,  ], MARGIN = 2, max)
}
new.lo.mat[1:dim(lo.mat)[1] - 1,  ] <- t(sapply(1:(dim(lo.mat)[1] -
       1), max.fn, lo.mat = lo.mat))
#9. REARRANGE UPPER CI's FOR PHASE 2
new.up.mat <- up.mat
min.fn <- function(k, up.mat)
{
       apply(up.mat[k:1,  ], 2, min)
}
new.up.mat[2:dim(up.mat)[1],  ] <- t(sapply(2:dim(up.mat)[1], min.fn,
       up.mat = up.mat))
#10. COMPUTE THE NEW CONFIDENCE INTERVAL WIDTHS FOR PHASE 2
new.width.mat <- new.up.mat - new.lo.mat
new.mean.width.mat <- as.matrix(apply(new.width.mat, 1, mean))
#11. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 2
new.mean.lo.mat <- as.matrix(apply(new.lo.mat, 1, mean))
new.mean.up.mat <- as.matrix(apply(new.up.mat, 1, mean))
#12. COMPUTE THE NEW COVERAGE PROBABILITIES FOR PHASE 2
new.cp.mat <- matrix(nrow = n, ncol = bin.number)
for(i in 1:bin.number) {
       for(j in 1:n) {
              new.cp.mat[j, i] <- sum((new.lo.mat[i,  ] < p.i.mat[
                     j, i]) & (p.i.mat[j, i] < new.up.mat[i,  ]))/nrep
       }
}
new.cp.vector <- as.vector(new.cp.mat)
#13. PLOT LOWER AND UPPER CONFIDENCE LIMITS
mean.lo.vector <- as.vector(mean.lo.mat)
mean.up.vector <- as.vector(mean.up.mat)
new.mean.lo.vector <- as.vector(new.mean.lo.mat)
new.mean.up.vector <- as.vector(new.mean.up.mat)
plot(1:bin.number, mean.lo.vector, type = "o", pch = 6, xlab = "Bin",
       ylab = "CI Limits", ylim = c(0, 1))
title(sub = "Method used: The Agresti-Coull interval")
points(1:bin.number, mean.up.vector, type = "o", pch = 2)
points(1:bin.number, new.mean.lo.vector, type = "o", pch = 6, col = 6)
points(1:bin.number, new.mean.up.vector, type = "o", pch = 2, col = 6)
legend(13, 0.97, c("Upper CL", "Lower CL", "New Upper CL",
       "New Lower CL"), marks = c(2, 6, 2, 6), col = c(1, 1, 6, 6))
#14. PLOT THE OLD & THE NEW COVERAGE PROBABILITIES AS A FUNCTION OF p
plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
       "Coverage Probability", ylim = c(0, 1))
title(sub = "Method used: The Agresti-Coull interval")
points(p.i.vector, new.cp.vector, type = "o", pch = 2, col = 6)
abline(1 - alpha, 0, col = 5)
#15. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 1
target <- rep(1 - alpha, length(x))
mse <- (rev(cp.vector) - target)^2
a.mse <- rep(0, each = length(mse))
p <- rev(p.i.vector)
for(i in 1:(length(mse) - 1)) {
       a.mse[i + 1] <- 0.5 * (mse[i] + mse[i + 1]) * (p[i + 1] - p[i])
}
RMSE <- sqrt(sum(a.mse))
#16. MEAN COVERAGE PROBABILITY for PHASE 1
cp <- rev(cp.vector)
mcp <- rep(0, length(cp))
```

```
	for(i in 1:(length(cp) - 1)) {
		mcp[i + 1] <- 0.5 * (cp[i] + cp[i + 1]) * (p[i + 1] - p[i])
	}
	MCP <- sum(mcp)
	#17. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 2
	mse.new <- (rev(new.cp.vector) - target)^2
	a.mse.new <- rep(0, each = length(mse.new))
	for(i in 1:(length(mse.new) - 1)) {
		a.mse.new[i + 1] <- 0.5 * (mse.new[i] + mse.new[i + 1]) * (
			p[i + 1] - p[i])
	}
	RMSE.new <- sqrt(sum(a.mse.new))
	#18. MEAN COVERAGE PROBABILITY for PHASE 2
	cp.new <- rev(new.cp.vector)
	mcp.new <- rep(0, length(cp.new))
	for(i in 1:(length(cp.new) - 1)) {
		mcp.new[i + 1] <- 0.5 * (cp.new[i] + cp.new[i + 1]) * (p[i +
			1] - p[i])
	}
	MCP.new <- sum(mcp.new)
	#19. RETURN RESULTS
	Table.1 <- data.frame("Mean Lower Limit" = mean.lo.mat,
		"Mean Upper Limit" = mean.up.mat, "Mean CI Width" =
		mean.width.mat)
	Table.2 <- data.frame("Mean Lower Limit" = new.mean.lo.mat,
		"Mean Upper Limit" = new.mean.up.mat, "Mean CI Width" =
		new.mean.width.mat)
	Table.3 <- data.frame(Root.MSE = RMSE, Mean.CP = MCP, Root.MSE.New =
		RMSE.new, Mean.CP.New = MCP.new)
	return(t(cp.mat), t(new.cp.mat), Table.1, Table.2, Table.3)
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE CLOPPER-PEARSON INTERVAL

```
function(n = 5, bin.number = 20, nrep = 100000, alpha = 0.05)
{
	x.t <- seq(-6, 5, 11/(bin.number * n))
	x <- x.t[-1]
	z <- qnorm(1 - alpha/2)
	#1. CREATE A MATRIX WHOSE ROWS CONTAIN nrep BERNOULLI R.V.'s
	y.mat <- matrix(nrow = length(x), ncol = nrep)
	for(i in 1:length(x)) {
		y.mat[i,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(x[i])))
	}
	#2. OBTAIN THE NUMBER OF SUCCESSES OUT OF n OBERVATIONS FOR EACH BIN
	lb <- seq(1, length(x) - n + 1, n)
	ub <- seq(n, length(x), n)
	num.suc.mat <- matrix(nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		num.suc.mat[i,  ] <- apply(y.mat[lb[i]:ub[i],  ], MARGIN = 2,
			sum)
	}
	#3. COMPUTATION OF (1-alpha)100% CLOPPER-PEARSON CONFIDENCE INTERVALS
	lo.mat <- matrix(0, nrow = bin.number, ncol = nrep)
	up.mat <- matrix(1, nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		lo.mat[i,  ][num.suc.mat[i,  ] == n] <- (alpha/2)^(1/n)
		up.mat[i,  ][num.suc.mat[i,  ] == 0] <- 1 - (alpha/2)^(1/n)
		Index <- (0 < num.suc.mat[i,  ]) & (num.suc.mat[i,  ] < n)
		lo.mat[i,  ][Index] <- qbeta(alpha/2, num.suc.mat[i,  ][Index],
			n - num.suc.mat[i,  ][Index] + 1)
		up.mat[i,  ][Index] <- qbeta(1 - alpha/2, num.suc.mat[i,  ][
			Index] + 1, n - num.suc.mat[i,  ][Index])
	}
	#4. COMPUTE THE CONFIDENCE INTERVAL WIDTHS FOR PHASE 1
	width.mat <- up.mat - lo.mat
	mean.width.mat <- as.matrix(apply(width.mat, 1, mean))
	#5. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 1
	mean.lo.mat <- as.matrix(apply(lo.mat, 1, mean))
	mean.up.mat <- as.matrix(apply(up.mat, 1, mean))
	#6. COMPUTE THE COVERAGE PROBABILITIES FOR PHASE 1
	p.i.vector <- 1/(1 + exp(x))
	p.i.mat <- matrix(p.i.vector, nrow = n, ncol = bin.number)
	cp.mat <- matrix(nrow = n, ncol = bin.number)
	for(i in 1:bin.number) {
		for(j in 1:n) {
			cp.mat[j, i] <- sum((lo.mat[i,  ] < p.i.mat[j, i]) &
				(p.i.mat[j, i] < up.mat[i,  ]))/nrep
		}
	}
	cp.vector <- as.vector(cp.mat)
	#7. PLOT THE COVERAGE PROBABILITIES AS A FUNCTION OF p
	plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
		"Coverage Probability", ylim = c(0, 1))
	title(sub = "Method used: The Clopper-Pearson interval")
	abline(1 - alpha, 0, col = 5)
	#8. REARRANGE LOWER CI's FOR PHASE 2
	new.lo.mat <- lo.mat
	max.fn <- function(k, lo.mat)
	{
		n.row <- dim(lo.mat)[1]
```

```
        apply(lo.mat[k:n.row,  ], MARGIN = 2, max)
}
new.lo.mat[1:dim(lo.mat)[1] - 1,  ] <- t(sapply(1:(dim(lo.mat)[1] -
        1), max.fn, lo.mat = lo.mat))
#9. REARRANGE UPPER CI's FOR PHASE 2
new.up.mat <- up.mat
min.fn <- function(k, up.mat)
{
        apply(up.mat[k:1,  ], 2, min)
}
new.up.mat[2:dim(up.mat)[1],  ] <- t(sapply(2:dim(up.mat)[1], min.fn,
        up.mat = up.mat))
#10. COMPUTE THE NEW CONFIDENCE INTERVAL WIDTHS FOR PHASE 2
new.width.mat <- new.up.mat - new.lo.mat
new.mean.width.mat <- as.matrix(apply(new.width.mat, 1, mean))
#11. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 2
new.mean.lo.mat <- as.matrix(apply(new.lo.mat, 1, mean))
new.mean.up.mat <- as.matrix(apply(new.up.mat, 1, mean))
#12. COMPUTE THE NEW COVERAGE PROBABILITIES FOR PHASE 2
new.cp.mat <- matrix(nrow = n, ncol = bin.number)
for(i in 1:bin.number) {
        for(j in 1:n) {
                new.cp.mat[j, i] <- sum((new.lo.mat[i,  ] < p.i.mat[
                        j, i]) & (p.i.mat[j, i] < new.up.mat[i,  ]))/nrep
        }
}
new.cp.vector <- as.vector(new.cp.mat)
#13. PLOT LOWER AND UPPER CONFIDENCE LIMITS
mean.lo.vector <- as.vector(mean.lo.mat)
mean.up.vector <- as.vector(mean.up.mat)
new.mean.lo.vector <- as.vector(new.mean.lo.mat)
new.mean.up.vector <- as.vector(new.mean.up.mat)
plot(1:bin.number, mean.lo.vector, type = "o", pch = 6, xlab = "Bin",
        ylab = "CI Limits", ylim = c(0, 1))
title(sub = "Method used: The Clopper-Pearson interval")
points(1:bin.number, mean.up.vector, type = "o", pch = 2)
points(1:bin.number, new.mean.lo.vector, type = "o", pch = 6, col = 6)
points(1:bin.number, new.mean.up.vector, type = "o", pch = 2, col = 6)
legend(13, 0.97, c("Upper CL", "Lower CL", "New Upper CL",
        "New Lower CL"), marks = c(2, 6, 2, 6), col = c(1, 1, 6, 6))
#14. PLOT THE OLD & THE NEW COVERAGE PROBABILITIES AS A FUNCTION OF p
plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
        "Coverage Probability", ylim = c(0, 1))
title(sub = "Method used: The Clopper-Pearson interval")
points(p.i.vector, new.cp.vector, type = "o", pch = 2, col = 6)
abline(1 - alpha, 0, col = 5)
#15. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 1
target <- rep(1 - alpha, length(x))
mse <- (rev(cp.vector) - target)^2
a.mse <- rep(0, each = length(mse))
p <- rev(p.i.vector)
for(i in 1:(length(mse) - 1)) {
        a.mse[i + 1] <- 0.5 * (mse[i] + mse[i + 1]) * (p[i + 1] - p[
                i])
}
RMSE <- sqrt(sum(a.mse))
#16. MEAN COVERAGE PROBABILITY for PHASE 1
cp <- rev(cp.vector)
mcp <- rep(0, length(cp))
for(i in 1:(length(cp) - 1)) {
        mcp[i + 1] <- 0.5 * (cp[i] + cp[i + 1]) * (p[i + 1] - p[i])
}
MCP <- sum(mcp)
```

```
	#17. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 2
	mse.new <- (rev(new.cp.vector) - target)^2
	a.mse.new <- rep(0, each = length(mse.new))
	for(i in 1:(length(mse.new) - 1)) {
		a.mse.new[i + 1] <- 0.5 * (mse.new[i] + mse.new[i + 1]) * (
			p[i + 1] - p[i])
	}
	RMSE.new <- sqrt(sum(a.mse.new))
	#18. MEAN COVERAGE PROBABILITY for PHASE 2
	cp.new <- rev(new.cp.vector)
	mcp.new <- rep(0, length(cp.new))
	for(i in 1:(length(cp.new) - 1)) {
		mcp.new[i + 1] <- 0.5 * (cp.new[i] + cp.new[i + 1]) * (p[i +
			1] - p[i])
	}
	MCP.new <- sum(mcp.new)
	#19. RETURN RESULTS
	Table.1 <- data.frame("Mean Lower Limit" = mean.lo.mat,
		"Mean Upper Limit" = mean.up.mat, "Mean CI Width" =
		mean.width.mat)
	Table.2 <- data.frame("Mean Lower Limit" = new.mean.lo.mat,
		"Mean Upper Limit" = new.mean.up.mat, "Mean CI Width" =
		new.mean.width.mat)
	Table.3 <- data.frame(Root.MSE = RMSE, Mean.CP = MCP, Root.MSE.New =
		RMSE.new, Mean.CP.New = MCP.new)
	return(t(cp.mat), t(new.cp.mat), Table.1, Table.2, Table.3)
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES USING THE EQUAL-TAILED JEFFREYS PRIOR INTERVAL

```
function(n = 5, bin.number = 20, nrep = 100000, alpha = 0.05)
{
	x.t <- seq(-6, 5, 11/(bin.number * n))
	x <- x.t[-1]
	z <- qnorm(1 - alpha/2)
	#1. CREATE A MATRIX WHOSE ROWS CONTAIN nrep BERNOULLI R.V.'s
	y.mat <- matrix(nrow = length(x), ncol = nrep)
	for(i in 1:length(x)) {
		y.mat[i,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(x[i])))
	}
	#2. OBTAIN THE NUMBER OF SUCCESSES OUT OF n OBERVATIONS FOR EACH BIN
	lb <- seq(1, length(x) - n + 1, n)
	ub <- seq(n, length(x), n)
	x.mat <- matrix(nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		x.mat[i,  ] <- apply(y.mat[lb[i]:ub[i],  ], MARGIN = 2, sum)
	}
	#3. COMPUTATION OF (1-alpha)100% JEFFREYS CONFIDENCE INTERVALS
	lo.mat <- matrix(0, nrow = bin.number, ncol = nrep)
	up.mat <- matrix(1, nrow = bin.number, ncol = nrep)
	for(i in 1:bin.number) {
		lo.mat[i,  ][x.mat[i,  ] == n] <- qbeta(alpha/2, x.mat[i,  ][
			x.mat[i,  ] == n] + 1/2, n - x.mat[i,  ][x.mat[i,  ] ==
			n] + 1/2)
		up.mat[i,  ][x.mat[i,  ] == 0] <- qbeta(1 - alpha/2, x.mat[
			i,  ][x.mat[i,  ] == 0] + 1/2, n - x.mat[i,  ][x.mat[
			i,  ] == 0] + 1/2)
		Index <- (0 < x.mat[i,  ]) & (x.mat[i,  ] < n)
		lo.mat[i,  ][Index] <- qbeta(alpha/2, x.mat[i,  ][Index] + 1/
			2, n - x.mat[i,  ][Index] + 1/2)
		up.mat[i,  ][Index] <- qbeta(1 - alpha/2, x.mat[i,  ][Index] +
			1/2, n - x.mat[i,  ][Index] + 1/2)
	}
	#4. COMPUTE THE CONFIDENCE INTERVAL WIDTHS FOR PHASE 1
	width.mat <- up.mat - lo.mat
	mean.width.mat <- as.matrix(apply(width.mat, 1, mean))
	#5. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 1
	mean.lo.mat <- as.matrix(apply(lo.mat, 1, mean))
	mean.up.mat <- as.matrix(apply(up.mat, 1, mean))
	#6. COMPUTE THE COVERAGE PROBABILITIES FOR PHASE 1
	p.i.vector <- 1/(1 + exp(x))
	p.i.mat <- matrix(p.i.vector[-1], nrow = n, ncol = bin.number)
	cp.mat <- matrix(nrow = n, ncol = bin.number)
	for(i in 1:bin.number) {
		for(j in 1:n) {
			cp.mat[j, i] <- sum((lo.mat[i,  ] < p.i.mat[j, i]) &
				(p.i.mat[j, i] < up.mat[i,  ]))/nrep
		}
	}
	cp.vector <- as.vector(cp.mat)
	#7. PLOT THE COVERAGE PROBABILITIES AS A FUNCTION OF p
	plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
		"Coverage Probability", ylim = c(0, 1))
	title(sub = "Method used: The Jeffreys Prior interval")
	abline(1 - alpha, 0, col = 5)
```

```
#8. REARRANGE LOWER CI's FOR PHASE 2
new.lo.mat <- lo.mat
max.fn <- function(k, lo.mat)
{
       n.row <- dim(lo.mat)[1]
       apply(lo.mat[k:n.row,  ], MARGIN = 2, max)
}
new.lo.mat[1:dim(lo.mat)[1] - 1,  ] <- t(sapply(1:(dim(lo.mat)[1] -
       1), max.fn, lo.mat = lo.mat))
#9. REARRANGE UPPER CI's FOR PHASE 2
new.up.mat <- up.mat
min.fn <- function(k, up.mat)
{
       apply(up.mat[k:1,  ], 2, min)
}
new.up.mat[2:dim(up.mat)[1],  ] <- t(sapply(2:dim(up.mat)[1], min.fn,
       up.mat = up.mat))
#10. COMPUTE THE NEW CONFIDENCE INTERVAL WIDTHS FOR PHASE 2
new.width.mat <- new.up.mat - new.lo.mat
new.mean.width.mat <- as.matrix(apply(new.width.mat, 1, mean))
#11. COMPUTE THE MEAN OF LOWER AND UPPER CONFIDENCE LIMITS FOR PHASE 2
new.mean.lo.mat <- as.matrix(apply(new.lo.mat, 1, mean))
new.mean.up.mat <- as.matrix(apply(new.up.mat, 1, mean))
#12. COMPUTATION OF THE NEW COVERAGE PROBABILITIES FOR PHASE 2
new.cp.mat <- matrix(nrow = n, ncol = bin.number)
for(i in 1:bin.number) {
       for(j in 1:n) {
              new.cp.mat[j, i] <- sum((new.lo.mat[i,  ] < p.i.mat[
                     j, i]) & (p.i.mat[j, i] < new.up.mat[i,  ]))/nrep
       }
}
new.cp.vector <- as.vector(new.cp.mat)
#13. PLOT LOWER AND UPPER CONFIDENCE LIMITS
mean.lo.vector <- as.vector(mean.lo.mat)
mean.up.vector <- as.vector(mean.up.mat)
new.mean.lo.vector <- as.vector(new.mean.lo.mat)
new.mean.up.vector <- as.vector(new.mean.up.mat)
plot(1:bin.number, mean.lo.vector, type = "o", pch = 6, xlab = "Bin",
       ylab = "CI Limits", ylim = c(0, 1))
title(sub = "Method used: The Jeffreys Prior interval")
points(1:bin.number, mean.up.vector, type = "o", pch = 2)
points(1:bin.number, new.mean.lo.vector, type = "o", pch = 6, col = 6)
points(1:bin.number, new.mean.up.vector, type = "o", pch = 2, col = 6)
legend(13, 0.97, c("Upper CL", "Lower CL", "New Upper CL",
       "New Lower CL"), marks = c(2, 6, 2, 6), col = c(1, 1, 6, 6))
#14. PLOT THE OLD & THE NEW COVERAGE PROBABILITIES AS A FUNCTION OF p
plot(p.i.vector, cp.vector, type = "o", xlab = "p", ylab =
       "Coverage Probability", ylim = c(0, 1))
title(sub = "Method used: The Jeffreys Prior interval")
points(p.i.vector, new.cp.vector, type = "o", pch = 2, col = 6)
abline(1 - alpha, 0, col = 5)
#15. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 1
target <- rep(1 - alpha, length(x))
mse <- (rev(cp.vector) - target)^2
a.mse <- rep(0, each = length(mse))
p <- rev(p.i.vector)
for(i in 1:(length(mse) - 1)) {
       a.mse[i + 1] <- 0.5 * (mse[i] + mse[i + 1]) * (p[i + 1] - p[
              i])
}
RMSE <- sqrt(sum(a.mse))
#16. MEAN COVERAGE PROBABILITY for PHASE 1
cp <- rev(cp.vector)
```

```
	mcp <- rep(0, length(cp))
	for(i in 1:(length(cp) - 1)) {
		mcp[i + 1] <- 0.5 * (cp[i] + cp[i + 1]) * (p[i + 1] - p[i])
	}
	MCP <- sum(mcp)
	#17. ROOT MEAN SQUARED ERROR of COVERAGE PROBABILITIES for PHASE 2
	mse.new <- (rev(new.cp.vector) - target)^2
	a.mse.new <- rep(0, each = length(mse.new))
	for(i in 1:(length(mse.new) - 1)) {
		a.mse.new[i + 1] <- 0.5 * (mse.new[i] + mse.new[i + 1]) * (
			p[i + 1] - p[i])
	}
	RMSE.new <- sqrt(sum(a.mse.new))
	#18. MEAN COVERAGE PROBABILITY for PHASE 2
	cp.new <- rev(new.cp.vector)
	mcp.new <- rep(0, length(cp.new))
	for(i in 1:(length(cp.new) - 1)) {
		mcp.new[i + 1] <- 0.5 * (cp.new[i] + cp.new[i + 1]) * (p[i +
			1] - p[i])
	}
	MCP.new <- sum(mcp.new)
	#19. RETURN RESULTS
	Table.1 <- data.frame("Mean Lower Limit" = mean.lo.mat,
		"Mean Upper Limit" = mean.up.mat, "Mean CI Width" =
		mean.width.mat)
	Table.2 <- data.frame("Mean Lower Limit" = new.mean.lo.mat,
		"Mean Upper Limit" = new.mean.up.mat, "Mean CI Width" =
		new.mean.width.mat)
	Table.3 <- data.frame(Root.MSE = RMSE, Mean.CP = MCP, Root.MSE.New =
		RMSE.new, Mean.CP.New = MCP.new)
	return(t(cp.mat), t(new.cp.mat), Table.1, Table.2, Table.3)
}
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX F. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES OF CONFIDENCE INTERVALS FOR PROBABILITIES BASED ON THE FIT OF A SIMPLE LINEAR LOGISTIC REGRESSION MODEL

```
function(nrep = 100000, alpha = 0.05)
{
	#-------------------------------------------------------------------------
	# Define the experimental region
	#-------------------------------------------------------------------------
	x <- seq(-6, 5, 11/100)
	#
	y.mat <- matrix(nrow = length(x), ncol = nrep)
	for(i in 1:length(x)) {
		y.mat[i,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(x[i])))
	}
	lo.mat <- matrix(nrow = length(x), ncol = nrep)
	up.mat <- matrix(nrow = length(x), ncol = nrep)
	#-------------------------------------------------------------------------
	# Inner function to fit a logistic regression to a data set, and
	# calculate lower and upper confidence levels for p for each range, x
	#-------------------------------------------------------------------------
	get.fits <- function(y, alpha)
	{
		assign("y", y, frame = 1)
		fit <- glm(y ~ x, family = binomial)
		list.1 <- predict(fit, type = "link", se = T)
		L <- list.1$fit - qnorm(1 - alpha/2) * list.1$se.fit
		U <- list.1$fit + qnorm(1 - alpha/2) * list.1$se.fit
		lo <- 1/(1 + exp( - L))
		up <- 1/(1 + exp( - U))
		c(lo, up)
	}
	#-------------------------------------------------------------------------
	# Fit a glm to each column of y.mat, and collect lower and upper
	# levels in two different matrices
	#-------------------------------------------------------------------------
	assign("x", x, frame = 1)
	new.mat <- apply(y.mat, 2, get.fits, alpha = alpha)
	lo.mat[1:length(x),  ] <- new.mat[1:length(x),  ]
	up.mat[1:length(x),  ] <- new.mat[(length(x) + 1):(2 * length(x)), ]
	width.mat <- up.mat - lo.mat
	mean.ci.width <- apply(width.mat, 1, mean)
	mean.lo <- apply(lo.mat, 1, mean)
	mean.up <- apply(up.mat, 1, mean)
	#-------------------------------------------------------------------------
	# Compute the coverage probabilities
	#-------------------------------------------------------------------------
	cp <- numeric(length(x))
	p.i <- 1/(1 + exp(x))
	for(i in 1:length(x)) {
		cp[i] <- sum((lo.mat[i,  ] < p.i[i]) & (p.i[i] < up.mat[i,
			]))/nrep
	}
	#-------------------------------------------------------------------------
	# Plot the coverage probabilities
	#-------------------------------------------------------------------------
	plot(p.i, cp, type = "o", xlab = "Population Parameter, p", ylab =
		"Coverage Probabilities", ylim = c(0, 1))
```

```
        abline(1 - alpha, 0, col = 6)
        plot(x, mean.lo, type = "l", xlab = "", ylab = "CI")
        points(x, mean.up, type = "l")
        data.frame(Range = x, p.i = p.i, Cov.Prob. = cp, "Lower CL" = mean.lo,
                "Upper CL" = mean.up, "Mean CI Width" = mean.ci.width)
}
```

APPENDIX G. SOFTWARE FOR COMPUTING THE COVERAGE PROBABILITIES OF Bca CONFIDENCE INTERVALS FOR PROBABILITIES BASED ON THE FIT OF A SIMPLE LINEAR LOGISTIC REGRESSION MODEL

```
function(nrep = 20000, B = 1000, alpha = 0.05)
{
	# ------------------------------------------------------------
	# Define the experimental region
	# ------------------------------------------------------------
	x.t <- seq(44, 76, 1)
	x <- rep(x.t, each = 2)
	# ------------------------------------------------------------
	# Generate 'nrep' data sets to be bootstrapped
	# ------------------------------------------------------------
	y.mat <- matrix(nrow = length(x), ncol = nrep)
	for(j in 1:length(x)) {
		y.mat[j,  ] <- rbinom(nrep, size = 1, p = 1/(1 + exp(-
			5.15176333358151 +  0.0962015734743007 * x[j])))
	}
	# ------------------------------------------------------------
	# Create two matrices to store the Bca confidence limits.
	# ------------------------------------------------------------
	lo.mat <- matrix(nrow = length(x), ncol = nrep)
	up.mat <- matrix(nrow = length(x), ncol = nrep)
	# ------------------------------------------------------------
	# Start nonparametric bootstrapping with Bca method
	# ------------------------------------------------------------
	for(i in 1:nrep) {
		#
		# Using the ith column of y.mat, make a data frame
		#
		assign("x", x, frame = 1)
		b.data <- data.frame(x = x, y = y.mat[, i])
		#
		assign("b.data", b.data, frame = 1)
		#
		boot.result <- bootstrap(data = b.data, B = B, statistic =
			predict(glm(y ~ x, data = b.data, family = binomial),
			newdata = data.frame(x = rep(seq(44, 76, 1), each = 2)),
			type = "response"))
		#
		# Assign the Bca confidence limits to a matrix
		Limit <- limits.bca(boot.result)
		#
		# Pass the 1st column of Limit matrix to the ith column of lo.mat
		# The 1st column corresponds the 2.5% percentile
		#
		lo.mat[, i] <- Limit[, 1]
		#
		# Pass the 4th column of Limit matrix to the ith column of up.mat
		# The 4th column corresponds to the 97.5% percentile
		#
		up.mat[, i] <- Limit[, 4]
	}
	width.mat <- up.mat - lo.mat
	mean.ci.width <- apply(width.mat, 1, mean)
	mean.lo <- apply(lo.mat, 1, mean)
	mean.up <- apply(up.mat, 1, mean)
```

```
	#-------------------------------------------------------------------------
	# Compute the coverage probabilities
	#-------------------------------------------------------------------------
	#
	cp <- numeric(length(x))
	p.i <- 1/(1 + exp(-5.15176333358151 + 0.0962015734743007 * x))
	for(i in 1:length(x)) {
		cp[i] <- sum((lo.mat[i,  ] < p.i[i]) & (p.i[i] < up.mat[i,  ]))
			/nrep
	}
	#-------------------------------------------------------------------------
	# Plot the coverage probabilities
	#-------------------------------------------------------------------------
	plot(p.i, cp, type = "o", xlab = "Population Parameter, p", ylab =
		"Coverage Probabilities", ylim = c(0.9, 1))
	abline(1 - alpha, 0, col = 6)
	#
	data.frame(Range = x, p.x = p.i, Cov.Prob. = cp, "Lower CL" = mean.lo,
		"Upper CL" = mean.up, "Mean CI Width" = mean.ci.width)
}
```

LIST OF REFERENCES

Agresti, A. (2002). ***Categorical data analysis*** (2nd ed.). New York: Wiley.

Agresti, A., & Caffo, B. (2000). Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. ***The American Statistician, 54***(4), 280-288.

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. ***The American Statistician, 52***(2), 119-126.

Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. ***Statistical Science***, 16(2), 101-117.

Collett, D. (1991). ***Modeling binary data*** (1st ed.). London: Chapman & Hall.

Devore, J. L. (2004). ***Probability and statistics*** (6th ed.). Canada: Brooks/Cole, Thomson Learning, Inc.

Efron, B., & Tibshirani, R. (1993). ***An introduction to the bootstrap.*** New York: Chapman & Hall.

Eng, J. (2003). Sample size estimation: How many individuals should be studied? ***Radiology, 227***(2), 311. Retrieved March 1, 2007, from http://radiology.rsnajnls.org/cgi/content/abstract/227/2/309.

Henderson, M., & Meyer, M. C. (2001). Exploring the confidence interval for a binomial parameter in a first course in statistical computing. ***The American Statistician, 55***(4), 337-344.

Hsieh, F. Y., Blosch, D. A., & Larsen, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. ***Statistics in medicine, 17***, 03/01/2007-1623-1634. Retrieved March 1, 2007, from http://www3.interscience.wiley.com/cgi-bin/fulltext/3057/PDFSTART.

Joseph, L., Wolfson, D. B., & Berger, R. D. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *The Statistician, 44*(2), 143-154.

Khuri, A. I., Mukherjee, B., Sinha, B. K., & Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science, 21*(3), 376-399.

Kupper, L. L., & Hafner, K. B. (1989). How appropriate are popular sample size formulas? *The American Statistician, 43*(2), 101-105.

Larsen, R. J, & Marx, M. L. (1986). *An introduction to mathematical statistics and its applications* (2nd ed.). Englewood Cliffs, New Jersey: Prentice-Hall.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in medicine, 17*(8), 857-872.

U.S. Army Yuma Proving Ground. (2006). Retrieved December 15, 2006, from http://www.yuma.army.mil/yuma_test_ctr.htm.

Whittemore, A. S. (1981). Sample size for logistic regression with small response probability. *Journal of the American Statistical Association, 76*(373), 27-32.

Yuma Proving Ground. (n.d.). Retrieved December 15, 2006, from http://en.wikipedia.org/wiki/Yuma_Proving-Ground.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Lyn R. Whitaker
   Naval Postgraduate School
   Monterey, California

4. Samuel E. Buttrey
   Naval Postgraduate School
   Monterey, California

5. Robert A. Koyak
   Naval Postgraduate School
   Monterey, California

6. Michael Diehl
   The U.S. Yuma Army Proving Ground
   Yuma, Arizona

7. Kara Harp Okulu Komutanlığı
   Savunma Bilimleri Enstitüsü
   Bakanlıklar, Ankara, Turkey

8. Okan Oymak
   Kara Kuvvetleri Harekat Başkanlığı
   Bakanlıklar, Ankara, Turkey