THE USE OF RESAMPLING TECHNIQUES FOR LIFETIME DATA ANALYSIS IN INDUSTRIAL ENGINEERING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ÇANKAYA UNIVERSITY

BY

MİRAY HANIM ASLAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF
MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

AUGUST 2007

Title of the Thesis : **The Use of Resampling Techniques for Lifetime Data Analysis in Industrial Engineering**

Submitted by **Miray Hanım ASLAN**

Approval of the Graduate School of Natural and Applied Sciences, Çankaya University

Prof. Dr. Ç. Özhan ULUATAM
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Levent KANDİLLER
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Fetih YILDIRIM
Supervisor

Examination Date :10.08.2007

Examining Committee Members

Prof. Dr. Levent KANDİLLER	(Çankaya Univ.)	
Prof. Dr. Fetih YILDIRIM	(Çankaya Univ.)	
Dr. Özlem İLK	(METU)	Özlem

STATEMENT OF NON-PLAGIARISM

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name	: Miray Hanım ASLAN
Signature	:
Date	: 10.08.2007

ABSTRACT

THE USE OF RESAMPLING TECHNIQUES FOR LIFETIME DATA ANALYSIS IN INDUSTRIAL ENGINEERING

ASLAN, Miray Hanım
M.S.c., Department of Industrial Engineering
Supervisor : Prof. Dr. Fetih YILDIRIM

August 2007, 92 pages

This study concerns with estimating the parameters in lifetime of fragile population and the ratio of fragile population to the fragile and durable (mixed) population by using trunsored models (unification of truncated and censored models) approach. The purpose of this study is to illustrate the bootstrap resampling method used for the parameter estimation in trunsored models. The bootstrap method is especially convenient to make statistical inference when distributional assumptions are not valid. Therefore, trunsored models with bootstrapping, which follow a consistent strategy in statistical inference and data analysis, lead to more accuracy for evaluation.

Like many real world cases, the thermal endurance data in material failure analysis do not follow any distribution perfectly. Furthermore, time and cost limitations prevent to observe a great number of data to analyze accurately. Thus, the trunsored model approach with bootstrapping is thought as potential to reduce the cost of destructive testing due to reduced frequency of testing, to prevent failures and to improve product reliability. The approach presented in this study may also be applied to many other real life problems.

Keywords: Bootstrap, Censored Data, Lifetime Analysis, Resampling Methods, Trunsored Models.

ÖZ

ENDÜSTRİ MÜHENDİSLİĞİNDE YAŞAM SÜRESİ ANALİZLERİ İÇİN YENİDEN ÖRNEKLEME YÖNTEMLERİNİN KULLANILMASI

ASLAN, Miray Hanım
Yükseklisans, Endüstri Mühendisliği Anabilim Dalı
Tez Yöneticisi : Prof. Dr. Fetih YILDIRIM

Ağustos 2007, 92 sayfa

Bu çalışma, kırılgan populasyonun ömür süresi parametrelerini ve kırılgan populasyonun durağan ve kırılgan (karışık) populasyona oranını trunsored (budanmış ve sansürlenmiş modellerin birleşimi) model yaklaşımı kullanarak tahmin etmekle ilgilidir. Bu çalışmanın amacı, trunsored modellerde parametre tahmini yapmak için bootstrap yeniden örnekleme yönteminin kullanılmasıdır. Bootstrap metodu, özellikle dağılım varsayımının geçerli olmadığı durumlarda istatistiksel çıkarsama yapmak için elverişlidir. Bu nedenle, istatistiksel çıkarsama ve veri analizlerinde tutarlı bir strateji takip eden bootstrapli trunsored modelleri, daha çok kesin sonuçlar elde etmeye öncülük etmektedir.

Birçok gerçek dünya vakaları gibi, malzeme arıza analizlerinde, ısı dayanımı verileri hiçbir dağılıma tam olarak uymamaktadır. Ayrıca, zaman ve maliyet kısıtları daha doğru analiz için çok büyük sayıda gözlem yapmayı engellemektedirler. Bu nedenle, bootstrapli trunsored model yaklaşımı test sıklıklarının azalmasından dolayı tahrip edici testlerin maliyetini düşürmek, arızaları azaltmak ve ürün güvenilirliğini geliştirmek için bir potansiyel olarak düşünülmüştür. Bu çalışmada gösterilen yaklaşımlar diğer birçok gerçek hayat problemine de uygulanabilir.

Anahtar Kelimeler: Bootstrap, Sansürlü Veri, Yaşam Süresi Analizi, Yeniden Örnekleme Yöntemleri, Trunsored Modeller

ACKNOWLEDGMENTS

I would like to thank to Prof. Dr. Fetih YILDIRIM for his guidance, advice, criticism, patience, encouragements, and understanding during the development of this thesis.

I would also like to thank Prof. Dr. Levent KANDİLLER and Dr. Özlem İLK for reading and evaluating this thesis.

I would like to express my deepest gratitude to my family and thank my dearest dad, Turgut ASLAN (I know he is always with me), my mom, Rahime ASLAN, and my sister, Sonay ASLAN for their love and support throughout the study.

I would like to thank all my friends in Çankaya University. I feel myself very lucky to have the chance to work with such great people: Engin TOPAN, Erkan BİBER, İlter ÖNDER, İpek SEYRAN TOPAN, Serdar SOYSAL and Derya GELAL.

Finally, I would like to express my special thanks to Ender YILDIRIM. In my hardest times he was always near. I would like to thank him for his support, patience, care, understanding, existence, and for everything that I could not list.

TABLE OF CONTENTS

STATEMENT OF NON-PLAGIARISM iii
ABSTRACT iv
ÖZ v
ACKNOWLEDGMENTS vi
TABLE OF CONTENTS vii
LIST OF FIGURES ix
LIST OF TABLES xi
CHAPTER 1 1
INTRODUCTION 1
CHAPTER 2 3
RESAMPLING METHODS 3
2.1 An Overview of Resampling 3
2.2 Nonparametric Approach 7
2.2.1 Permutation and Randomization 9
2.2.2 Cross – Validation 12
2.2.3 Jackknife 13
2.3 Parametric Approaches 16
CHAPTER 3 18
BOOTSTRAP RESAMPLING 18
3.1 An Overview of Bootstrap Method 18
3.2 Arguments on Bootstrapping: A Literature Review 20
3.3 Bootstrapping 25
3.4 Plug-In Principle 27
3.5 Algorithm and Principle of the Bootstrap 28

3.6 The Bootstrap Estimate of Standard Error 29
3.7 Bootstrap Estimate of Bias and the Bias Correction 31
3.8 Bootstrap Confidence Interval 33
3.8.1 Standard Bootstrap Confidence Interval 34
3.8.2 Bootstrap-t Confidence Interval 35
3.8.3 Bootstrap Percentile Confidence Interval 36
3.8.4 Bias-Corrected and Accelerated Bootstrap Confidence Interval 37
3.8.5 Approximate Bootstrap Confidence Interval 39
CHAPTER 4 40
AN INDUSTRIAL ENGINEERING APPLICATION: LIFETIME DATA ANALYSIS 40
4.1 Definition of Reliability 42
4.2 Achieving Reliability 43
4.3 Reliability Data 45
4.3.1 Types of Reliability Data 46
4.4 Trunsored Model 50
4.4.1 Possible Applications of the Trunsored Model 51
4.4.2 Construction of Trunsored Models 52
4.5 Problem Definition and Lifetime Experiment 58
4.6 Bootstrapping with Censoring Mechanism and its Application 60
4.7 Lifetime Estimation for the Trunsored Data Model 62
CHAPTER 5 67
CONCLUSION 67
REFERENCES R1
APPENDIX A A1
SAMPLE FORM OF EXPERIMENT RESULT A1
APPENDIX B A2
DESCRIPTIVE STATISTICS AND EXPERIMENT RESULTS A2
APPENDIX C A5
BOOTSTRAP AND TRUNSORED MODEL RESULTS A5

LIST OF FIGURES

Figure 3.1: Schematic representation of the bootstrap process for estimating standard error of a statistic s(x) 26
Figure 4.1: Representation of Complete Data over Time 47
Figure 4.2: Representation of Right Censored Data over Time 49
Figure 4.3: Representation of Interval Censored Data over Time 49
Figure 4.4: Representation of Left Censored Data over Time 50
Figure 4.5: Kaplan Meier Estimated Survival Curve of the Lamp Lifetimes 63
Figure C.1: Histogram of Bootstrap Samples for B=1000 .... A5
Figure C.2: Quantile-Quantile Plot of Bootstrap Samples for B=1000 .... A5
Figure C.3: Histogram of Bootstrap Samples for B=2000 .... A6
Figure C.4: Quantile-Quantile Plot of Bootstrap Samples for B=2000 .... A6
Figure C.5: Histogram of Bootstrap Samples for B=3000 .... A7
Figure C.6: Quantile-Quantile Plot of Bootstrap Samples for B=3000 .... A7
Figure C.7: Histogram of Bootstrap Samples for B=4000 .... A8
Figure C.8: Quantile-Quantile Plot of Bootstrap Samples for B=4000 .... A8
Figure C.9: Histogram of Bootstrap Samples for B=5000 .... A9
Figure C.10: Quantile-Quantile Plot of Bootstrap Samples for B=5000 .... A9
Figure C.11: Histogram of Bootstrap Samples for B=6000 .... A10
Figure C.12: Quantile-Quantile Plot of Bootstrap Samples for B=6000 .... A10
Figure C.13: Histogram of Bootstrap Samples for B=7000 .... A11
Figure C.14: Quantile-Quantile Plot of Bootstrap Samples for B=7000 .... A11
Figure C.15: Histogram of Bootstrap Samples for B=8000 .... A12
Figure C.16: Quantile-Quantile Plot of Bootstrap Samples for B=8000 .... A12
Figure C.17: Histogram of Bootstrap Samples for B=9000 .... A13

Figure C.18: Quantile-Quantile Plot of Bootstrap Samples for B=9000 .... A13
Figure C.19: Histogram of Bootstrap Samples for B=10000.................................. A14
Figure C.20: Quantile-Quantile Plot of Bootstrap Samples for B=10000 .... A14
Figure C.21: Estimated Bias vs. Number of Replications A15
Figure C.22: Estimated Ratio of Fragile Population to Mixed Population vs. Number of Replications A15
Figure C.23: Percentiles vs. Number of Replications... A16
Figure C.24: Standard Errors of Mean Lifetime vs. Number of Replications A16
Figure C.25: Estimated Mean Lifetime vs. Number of Replications....................... A17
Figure C.26: Estimated Standard Error vs. Number of Replications....................... A17

LIST OF TABLES

Table 4.1: Experiment Results of a Lot 60

Table 4.2: Bootstrap Resampling Results for Each Replication 66

Table B.1: Descriptive Statistics of the Lifetime Data A3

Table B.2: Experiment Results A3

CHAPTER 1

INTRODUCTION

The perspective of statistical computing has been revolutionized by recent advances in the power of computers. This modern perspective has been improved and oriented the computer intensive techniques. The important aspect to this orientation is the difficulty and complexity of the tasks by using traditional methods. In particular, drawing of statistical inferences from a set of data always required high computer power which made the analysis practical. Thus, this study demonstrates one of the computer intensive techniques, resampling method, as an alternative for traditional methods. Resampling methods are powerful, modern tools that can be used to make statistical inference and to investigate the behavior of any estimation. The main trigger mechanism to use a resampling method is thought as either elimination or restriction of any unverifiable assumption about the data to a minimum. In this study, a specific resampling method, bootstrap resampling, which uses the information contained in a single sample from the population of interest, is mostly focused on. In particular, the nonparametric bootstrap method is used to draw statistics based conclusions since it requires fewer assumptions relative to both parametric and traditional methods.

The primary aim of this study is to develop an efficient industrial engineering application of bootstrap resampling method. Thus, the application is chosen from reliability analysis to develop original solutions that arise in modern engineering designs. The problem taken into account must seriously consider reliability because lifetime data with incomplete observations frequently arises in reliability. A new incomplete data model, i.e. trunsored model, will be used to analyze the data. This method provides not only an estimate for the ratio of the fragile population to the mixed fragile and durable populations, but also tests the hypothesis that the ratio is equal to a prescribed value with estimated confidence intervals. Furthermore, trunsored model approach with bootstrapping can cope with time and cost limitations in lifetime data analysis. Therefore, for modern engineering design, trunsored model approach with bootstrapping is thought as potential to reduce the cost of destructive testing due to reduced frequency of testing, to prevent failures and to improve product reliability.

This study presents the fundamentals of the bootstrap method with an original application. Developing the mathematical and logical background of resampling methods is required for understanding the bootstrap procedure. Thus, resampling methods are defined in Chapter 2 of this study. Since the main philosophy is drawing the pictures of resampling with a framework that comes from wider to narrower, Chapter 3 presents the details of bootstrap method. The basics of lifetime data analysis and the industrial engineering case study are proposed in Chapter 4. Finally, conclusive remarks are provided in Chapter 5.

CHAPTER 2

RESAMPLING METHODS

The roles of computing in statistics became important with the increases in computer power, decreases in computing costs and recent advances in the information technology. Several terms are used when referring to computer intensive statistical methods, including ‘resampling’, ‘Monte Carlo Simulation’, ‘permutation’, ‘randomization’, ‘jackknife’ and ‘bootstrapping’. The generic terms ‘**resampling**’ and ‘**computer-intensive methods**’ refer to all methods in which the observed data are used to generate a reference distribution by means of randomization (Fortin et al. (2002)). Hence, this chapter includes general background of resampling and computer intensive methods. Section 2.1 explains principles and overview of the resampling. Section 2.2 describes nonparametric resampling methods. Finally, Section 2.3 briefly reviews parametric resampling methods.

2.1 An Overview of Resampling

Statistical inference relies on some statistics (or estimates) that are functions of the data. Their sampling distributions depend upon the underlying population and therefore are unknown. Traditional methods used in statistical inference are generally based on postulated probability models. Even if the probability model has held, the conclusions are often made by asymptotical and approximate results. Thus, instead of traditional methods, resampling methods are proposed to provide strategies for estimating or approximating the sampling distribution of a statistic or its characteristics as well as making statistical inference. Since inferential statistics include estimators (i.e. the functions of the data and the statistics computed from a random sample and used to estimate parameters of population distributions) as well as statistics used in hypothesis tests, resampling plays such an important role as the architecture.

One important aspect of resampling is the growing importance of data analyses based on recycling the scores constituting a data set and collection of computer intensive techniques (Fortin et al. (2002)). These techniques generate distributions of statistics by repeating the data analysis many times on replicate data sets (resamples) that are based on an observed set of data (Lunneborg (2000), p. 78). The generated distribution is then used to assess the significance of a statistic calculated from the observed data. Significance is evaluated under the assumption that the statistic computed by using the observed data is sampled from the distribution generated with a randomization mechanism.

In most applied statistical analysis, statistical procedure models a physical process via random samples. Thus, random samples are at the heart of statistical inference and are in the concept of sampling distribution and resampling methods. A resampling method simulates the model with easy-to-manipulate symbols via data-generating mechanism to produce new hypothetical samples. Generated samples try to act as a population and are the introductory part of resampling methods. Therefore, the intellectual advantage of the resampling methods relative to traditional methods is the data generation mechanism for achieving an enormous approximation to the population.

When we compare traditional sampling methods with resampling methods, one can have the following assertions: resampling methods have fewer assumptions than traditional methods (Crowley (1992)). For example, resampling method does not require that the data has analytically known distribution. Hence, an important feature of resampling is that statistical significance is evaluated based on empirical distributions generated from the observed sample. This ‘distribution-free’ alternative to parametric statistics is quite appealing to reliability analysis, which statistical inferences have to be made with small data sets that do not meet the assumed parametric distribution, will be mentioned in the application part of this study in later chapters. While resampling methods may involve fewer assumptions as stated before, this does not mean no assumption exists. Hence, caution must be exercised because the random sampling procedures often assume that data of independent observation and this assumption is invalid when the data are spatially or temporally autocorrelated (Cressie (1993), Edgington (1995), Efron and Tibshirani (1993), Good (2000, pp. 25-29). The reliability of the statistical analysis applied by resampling methods depends on the validity of this assumption. Also, in resampling, the observed data are assumed to be a representative picture of the entire population. The essential idea then is to make statistical inference based on an artificial resample, which is drawn from the data.

Resampling procedure has some benefits with respect to traditional methods and approaches and there exists some motivation to use this procedure. First of all, traditional approaches rely strongly on postulated probability models as stated before. Conclusions of traditional methods are frequently based on asymptotical or approximate properties that increase the effect of biases. These approaches, however, determine the various properties of a particular estimate with various assumptions about the underlying population distribution. On the other hand, there are many situations where the determination of estimators’ properties is not so straightforward. However, resampling techniques can provide a solution in such situations that include, but are not limited to, the following properties:

Distributional assumption violation/inadequacy: Classical procedures rely on the distributional assumptions regarding to the population of interest. When the population is not well-defined or the sample size is small, the analyst should be sceptical whether the usefulness of the theoretical distributions available or not and hence may wish to use “nonparametric approaches”.

Non-random samples: An important classical assumption is that the sample is random and certain processes of inferring population quantities from a sample require this assumption for validity. However, there are situations where the sample might not be random: for example, “self-selected” samples obtained via certain types of questionnaire in which people elect to be a part of the sample rather than being chosen by the experimenter.

Small sample sizes: Many traditional methods for estimating various properties of a certain characteristic of a population rely on the assumption of a “large” sample size. Thus, for smaller samples, these methods may result in invalid estimates of the various properties of a population.

Intractable calculations: In some cases, either the distributional assumptions made for the random variable of interest or the particular nature of the estimator may prevent finding explicit mathematical statements for the various properties of the estimator because the mathematical calculations required to do so are intractable.

Different resampling plans result in different resampling techniques. The main difference relies on the distribution assumption. In classical statistical theory, it is usually assumed that there is a particular mathematical model, with adjustable constants or parameters that fully determine the function. In general, such a model is called as parametric and the data generated from the underlying model is in the family of distributions. However, when no such mathematical model is used and any explicit assumptions does not require about the population's distribution, then the statistical inferences are made by nonparametric methods. Besides, independent and identically distributed random variables should be used. Even if there is a plausible parametric model, a nonparametric analysis can still be useful to assess the robustness of conclusions drawn from a parametric analysis (Davison and Hinkley (1997), p. 11). Some basics of parametric and nonparametric approaches are given in the following sections.

2.2 Nonparametric Approach

In many practical situations that appear in resampling, it is useful to have available statistical methods which do not depend upon specific parametric models. Therefore, nonparametric resampling approaches, which do not rely on any specific assumptions about the form of the probability distribution and have extremely different methodology from parametric approaches, play a central role in statistical inference. The observed data (the sample) used in this approach should come from empirical distribution which puts equal probabilities at each sample value. When the sample values are thought of as the outcomes of independent and identically distributed random variables $X_1,...,X_n$, the equal probabilities will be n^{-1} at each sample value x_i. The corresponding estimate of F is the empirical distribution function (EDF) $\hat{F}$, which is defined as the sample proportion

$$\hat{F}(x)=\frac{\#\left\{x_{(i)} \leq x\right\}}{n} \qquad \text{for } i=1,...,n \tag{2.1}$$

where $\#\left\{A\right\}$ means the number of times the event A occurs; or

$$\hat{F}(x)=\begin{cases} 0, & x<x_{(1)} \\ i/n, & x_{(i)} \leq x<x_{(i+1)} \\ 1, & x \geq x_{(n)} \end{cases}, \tag{2.2}$$

where $x_{(i)}$ is the i^{th} ordered value of x. More formally,

$$\hat{F}(x)=\frac{1}{n}\sum_{i=1}^{n} H(x-x_{(i)}),$$

where $H(u)$ is the unit step function which jumps from 0 to 1 at u=0. It should be noticed that the values of the EDF are fixed $\left(0, \frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n}\right)$, so the EDF is equivalent to its points of increase, the ordered values $x_{(1)} \leq ... \leq x_{(n)}$ of the data. When there are repeated values in the sample, as would often occur with the discrete data, the EDF assigns probabilities proportional to the sample frequencies at each distinct observed value y. The EDF plays the role of fitted model when no mathematical form is assumed for F (i.e. distribution-free), analogous to a parametric cumulative distribution function (CDF) with parameters replaced by their estimates (Davison and Hinkley (1997), pp. 11-12).

Nonparametric resampling procedure has some benefits with respect to parametric methods and approaches. There exists some motivation to use this procedure. These include, but are not limited to:

- If the sample size is very small, there may be no alternative to using a nonparametric resampling method unless the nature of the population distribution is known exactly.
- Nonparametric resampling makes fewer assumptions about the data.
- Nonparametric resampling are available to analyze data which are inherently in ranks or categorical.
- Nonparametric resampling is typically much easier to learn.

In contrast to these motivations, there may be rarer cases in which the use of a resampling method can fail. These include, but are not limited to:

- A nonparametric resampling method is less powerful than a parametric resampling when all the assumptions of the parametric one are met.
- Certain assumptions, which are associated with nonparametric resampling methods, e.g. the observations are independent, may not be held.

Different nonparametric resampling methods are improved in the statistical inference. Friedl and Stampfer (2002) denote that these plans are generally referred as setting of all possible resamples to be taken and their weighting. Rather, the inference is based upon repeated sampling within the same sample without a distribution assumption. The resampling procedure is applied within the light of one of four major methods, i.e. Bootstrap, Jackknife, Cross-Validation, Permutation-Randomization, which the literature has mostly addressed for testing and estimation. Although today they are unified or improved under a common theme, it is important to note that these four techniques were developed by different people at different periods of time for different purposes (Friedl and Stampfer (2002)). In the following subsections, Permutation-Randomization, Cross-Validation and Jackknife will be explained. The main topic of this study, the bootstrap method, is going to be given with details in the next chapter.

2.2.1 Permutation and Randomization

Permutation test is a nonparametric procedure that calculates an attained significance level of a test statistic by comparing it with its resampled values. Hence, it is designed to condition out the unknown sampling distribution. This test utilizes resamples that are drawn without replacement from the observations. The distribution of the resampled values is called permutation distribution, and plays the role of the null distribution in parametric testing problems. The applicability of permutation tests relies on the property that some observations are exchangeable under the null hypothesis, whereas under the alternative hypothesis they are not (Friedl and Stampfer (2002)). Thus, this test is the nonparametric version of hypothesis tests. Deeper insight into this topic is provided by Lehmann (1997) and Good (2006).

Permutation test can be employed for continuous as well as for ordinal and nominal data. When the two random samples are taken as $X_1,...,X_m$ from an unknown distribution of $X \sim F_X(.)$ and a random sample $Y_1,...,Y_n$ from an unknown distribution of $Y \sim F_Y(.)$, the null hypothesis is that the mean of two distributions are the same whereas the alternative hypothesis is that the mean of $F_X(.)$ is different from the mean of $F_Y(.)$. One feature of permutation tests is that any test statistic is as easy to use as any other, at least in principle (Davison and Hinkley (1997), pp. 156-158). For this reason, the test statistic, $T(D_n)$, where D_n is a data set with n data points, can be computed for the observed data. If all possible permutations are defined as R, then $R=\binom{M+N}{M}$ where N and M are two subsets of n. If the null hypothesis is true, such permutations are equally likely and there are $\binom{M+N}{M}$ of them. For each i^{th} permutation ($i=1,...,R$), $T^{(i)}$ is the statistic that should be computed and the value of $T(D_n)$ should be compared with the set of values $T^{(i)}$. If the value of $T(D_n)$ falls in the upper and lower $\alpha/2$ tail areas of the $T^{(i)}$ distribution, the null hypothesis is rejected with type 1 error α. The permutation procedure will involve substantial computation unless M and N are small and a random sample of a large number R of the permutations can be taken when the number of permutations is too large.

Permutation method can be used for comparison of two means, estimated survivor functions, testing correlations, and etc. The corresponding examples and the theory can be seen in Davison and Hinkley (1997) and Lehmann (1997) and the detailed algorithm of test can be seen in Efron and Tibshirani (1993, pp. 202-218).

The randomization test is introduced by Fisher (1949, pp. 17-21) as a device for explaining and justifying significance tests, both in sample cases and for complicated experimental designs (Davison and Hinkley (1997), p. 183). The terms ‘permutation test’ and ‘randomization test’ are often used interchangeably. Formally, Fisher used the former term to refer to a method that performed for inference from population, while randomization test was applied to methods for sample-based inference. The term permutation test or randomization test is also used when the test provides the exact significance levels by exhaustive computation of all possible rearrangements (permutations) of the data. In practice, even with current powerful computers, permutation tests can only be performed when the number of observations, n, is small, because the number of permutations increases as the factorial of the sample size (n!). When the number of observations precludes an exact test, an 'approximate randomization' test is used instead. This randomization test generates a subset of the possible permutations because only a subsample of all possible permutations is calculated. Many authors recommend that 10000 or more randomizations should be used while constructing the reference distribution (Crowley (1992), Manly (1997)). The reference distribution of any statistic is obtained using a six-step randomization procedure that repeatedly reallocates the value of the observations over the sample, and then recalculates the statistic to generate the null reference distribution. The algorithm can be summarized as: (i) hypothesis definition; (ii) statistic determination (choose a statistic that already existing or design a new one); (iii) statistic computation for the observed data; (iv) null reference distribution generation by rearranging the order of the observed data over the entire sample by shuffling them randomly (i.e. the values of the response variable are shuffled over all the samples, where each sample keeps its spatial identity); (v) computing the statistic for the randomized data and repeating this step a large number of times; and (vi) comparison of the observed statistic with respect to the reference distribution.

Randomization tests have several advantages and limitations:

- "The randomization tests include flexibility and relative ease of implementation. They support significance testing without distributional assumptions (e.g. normality) and complex designs for which parametric tests do not exist" (Fortin et al. (2002)).
- "For comparable statistics, randomization tests are as powerful as parametric tests when the number of randomizations is large" (Fortin et al. (2002)).

- The test can be used with non-random data and allows user to create a test statistic (Crowley (1992)).
- Results are not necessarily generalizeable to the population. Generalizeable assumption is often made but not verified with standard statistics (Crowley (1992)).
- Randomization and traditional methods give similar significance levels if assumptions of traditional tests hold (Crowley (1992)).
- The test can only be used to test hypotheses comparing two or more groups (Crowley (1992)).

2.2.2 Cross – Validation

Cross validation is another resampling method that was proposed by Kurtz (1948). This method is especially designed for selecting and assessing models with estimation of aggregate error. The paradigm is based on splitting the data set into a training set and a separate assessment set. In general, a sample is randomly divided into two or more subsets, say K (roughly equal-sized pieces). Then, the test results are validated by comparing across subsamples. One piece, i.e. the training set, can be used to test the model that was trained on the remaining K-1 pieces. Therefore for the k^{th} part, the model is fitted to the other K-1 parts of the data, and prediction error of the fitted model is to be calculated while predicting the k^{th} part of the data. To remove the effect of a particular division, this is repeated for all K pieces of data. The results of the K testing procedures are then combined suitably. This cross validation procedure is called as K-fold cross-validation (Efron and Tibshirani (1993), pp. 239-241). In the literature, there also exist more specific cases of K-fold cross-validation such as simple cross-validation, double cross-validation, and multicross-validation. Further details related with cross-validation are given by Stone (1974). Efron and Gong (1983) compares cross validation with other resampling methods. Efron (1983) shows that leave-one-out cross validation reduces the bias of the estimate from $O(1/n)$ to $O(1/n^2)$. Ang (1998) states that cross-validation is problematic because splitting an already small sample increases the risk of artifacts of the subsample, and thus, Ang (1998) recommends to use of jackknife (to be explained in the next section).

The power of cross-validation comes from the reduced number of assumptions and its applicability to complex situations. On the other hand, cross-validation suffers from the same weakness as spilt-half reliability when the sample size is small. By dividing the sample into two halves, each analysis is limited by a smaller number of observations. However, it can have high variability particularly for small sample size. Thus, the sample size must be large enough to fit the model reliably, assess the prediction error reliably and be reasonably independent of the actual split into assessment and trainings sets. The trick to achieve these things even for modest sample sizes is to repeat this procedure for multiple trainings/assessment splits, and to average out the prediction errors.

2.2.3 Jackknife

The resamples can be produced by repeatedly leaving out one observation from the data by the method known as the jackknife. Although the method was coined by Tukey to imply that the method is an all-purpose statistical tool in resampling, jackknife is first proposed by Quenouille (1949). Quenouille (1956) finds out an estimator of bias by using jackknife resamples then it is developed by Tukey (1958) to quantify standard error of an estimate without making distributional assumptions. In the later improvement, jackknife resampling plans generalized for estimating of any statistic of interest.

Jackknife enumerates the reference distribution repeatedly by leaving out one observation at a time and then recalculating the test statistic. When the sample values $\underset{\sim}{y}=(y_1,...,y_n)$ are thought of as the outcomes of independent and identically distributed random variables, and $\hat{\theta}=S(y)$ is the estimator of the statistic of interest θ, the method aims to find bias and standard error estimate of $\hat{\theta}$ by the samples that leave out one observation at a time. The sample

$\underset{\sim}{y}_{(i)}=(y_1,y_2,...,y_{i-1},y_{i+1},...,y_n)$

shows the remaining data set after removing i^{th} observation for $i=1,...,n$, is called as i^{th} jackknife sample (Efron and Tibshirani (1993), pp. 141-143). Let $\hat{\theta}_{(i)}=S(y_{(i)})$ be the i^{th} jackknife replication of $\hat{\theta}$. Then, the jackknife estimate of the statistic θ and estimate of the bias $\widehat{\text{bias}}_{\text{jack}}$ is given by

$$\hat{\theta}_{\text{jack}}=n\hat{\theta}-(n-1)\hat{\theta}_{(.)} \tag{2.3}$$

$$\widehat{\text{bias}}_{\text{jack}}=(n-1)(\hat{\theta}_{(.)}-\hat{\theta}), \tag{2.4}$$

where

$$\hat{\theta}_{(.)}=\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}. \tag{2.5}$$

The jackknife estimate of the standard error defined by

$$\widehat{\text{se}}_{\text{jack}}=\left[\frac{n-1}{n}\sum(\hat{\theta}_{(i)}-\hat{\theta}_{(.)})^2\right]^{1/2}. \tag{2.6}$$

Another way defined in Efron and Tibshirani (1993, p. 145) is to think about the jackknife in terms of the "pseudo values" as

$$\tilde{\theta}_i=n\hat{\theta}-(n-1)\hat{\theta}_{(i)}. \tag{2.7}$$

It is stated that in the special case $\hat{\theta}=\bar{x}$, $\tilde{\theta}_i=x_i$ is observed where x_i is the i^{th} data value. Furthermore, for any $\hat{\theta}$, the formula for $\widehat{\text{se}}_{\text{jack}}$ can be expressed as

$$\widehat{\text{se}}_{\text{jack}}=\left\{\sum_{1}^{n}(\tilde{\theta}_i-\tilde{\theta})^2/\{(n-1)n\}\right\}^{1/2}, \tag{2.8}$$

where $\tilde{\theta}=\sum\tilde{\theta}_i/n$. Here, $\widehat{se}_{jack}$ is like an estimate of the standard error of the mean for the data $\tilde{\theta}_i$, for $i=1,...,n$ owing to fact that the pseudo values are supposed to act as if they were n independent data values. Likewise, an approximate (1-α)% confidence interval can be formed as:

$$\tilde{\theta}\pm t_{n-1}^{(1-\alpha)}\widehat{se}_{jack}, \tag{2.9}$$

where $t_{n-1}^{(1-\alpha)}$ is the $(1-\alpha)^{th}$ percentile of the t distribution on n-1 degrees of freedom. Efron and Tibshirani (1993, p. 145) denote that this interval does not work very well; in particular, it is not significantly better than cruder intervals based on normal theory. Although pseudo values are intriguing, it is not clear whether they are a useful way of thinking about the jackknife.

The methodology of the jackknife changes according to the number of deleted observations. In the simplest case, the jackknife resamples are generated by deleting single cases from the original sample (i.e. delete-one jackknife). A more generalized technique uses resample that relies on multiple deletions (namely, delete-d jackknifes where d is the number of deleted observations). The jackknife often provides a simple and good approximation especially for estimation of standard errors and bias. Like as the other nonparametric resampling methods, the jackknife can be applied to any statistic that is a function of n independent and identically distributed variables. With respect to the resampling plans the idea of cross validation is very similar to the jackknife idea; however, cross validation should not be mixed up with jackknife since both of these resampling procedures are quite different. The major difference is based on their applications in which cross validation is used for model selection and assessment, whereas jackknife provides estimate of bias and variance.

The major motivation for jackknife estimates is that they reduce bias. It is also a nonparametric and easy to be implemented method and it solves a number of problems like other nonparametric methods explained before. For instance, jackknife can be used for any estimator that is a sample analogue of a parameter as in the following: the sample mean as an estimator of the population mean, the sample variance as an estimator of the population variance, the sample minimum as an estimator of the population minimum and so on. However, this method can fail miserably if the statistic $\hat{\theta}$ is not "smooth". Intuitively, the idea of smoothness is that small changes in the data set cause only small changes in the statistic. Efron and Tibshirani (1993, p. 148) state a way that fixes up the inconsistency of the jackknife for non-smooth statistic. The detailed review about the Jackknife has been made by Miller (1974).

2.3 Parametric Approaches

One mathematical route to specifying a sampling distribution is to require the population distribution that has a particular mathematical form. Therefore, it is considered that the observed data (the sample) comes from a specific distribution that can be called as a representative picture of the entire population. The sample values are thought of as the outcomes of independent and identically distributed (iid) random variables $Y_1,...Y_n$ whose probability density function (PDF) is denoted as f and cumulative distribution function (CDF) is denoted as F. The sample is to be used to make inferences about a population characteristic, generally denoted by θ, using a statistic T whose value in the sample is t. The attention is to be focused on the probability distribution of T. When the generating samples come from a known distribution or one may be assumed the probability distributions fitted to samples' data, the statistical analysis relies on the parametric approach.

Hypothesis tests and pivotal variables are used in parametric resampling. The major disadvantages of these methods include model selection error, parameter estimation error, and loss of important serial and cross dependencies in the data, and the difficulty in convincing for the model's validity.

Hypothesis test is a method of sampling distribution that uses a statistic calculated from the sample to test an assertion about the value of a population parameter. In this method, population distribution parameter, θ, is assumed as taking a specific value, θ_0. The value of θ is specified from a statistical hypothesis known as a null hypothesis. So, the null hypothesized value is referred as θ_0. The main aim of the test is to determine what the sampling distribution of the estimator would be if the null hypothesis were correct. Therefore, in the first step, the sample statistic is calculated and the hypothesis is formulated. The null hypothesis (H_0) specifies a value for the population parameter. The decision about which sample statistic should be calculated depends upon the scale used to measure the variable (i.e. a proportion, a mean, etc). In contrast, the alternative hypothesis (H_1) specifies a competing value for the population parameter and is formulated to reflect the proposition the researcher wants to verify. Consistency judgments and the decision either reject the null hypothesis in preference to the alternative or not reject the null hypothesis should be made on statistical grounds. This statistical decision process is referred to as hypothesis testing.

Pivot variable is used to construct confidence intervals. "To avoid the difficulties associated with a shifting sampling distribution, mathematical statisticians have developed pivotal forms for several estimators. The sampling distribution of a pivotal form does not change as we move from one population to another, with a consequent change in the value of the parameter being estimated. The convenience of pivotal forms, particularly for confidence interval estimation and hypothesis testing, has encouraged researchers to use estimators that have pivotal forms"(Lunneborg (2000), pp. 60-61). This method is called as "pivotal inference" because of the results obtained by using pivots and the structure of pivotal inference is improved. Since the parametric approaches are out of the scope of this thesis, details of hypothesis testing and pivotal inference are not included.

CHAPTER 3

BOOTSTRAP RESAMPLING

Bootstrap is a data-based simulation method for assigning measure of accuracy to statistical estimates. “The use of the term *bootstrap** derives from the phrase *to pull oneself up by one's bootstrap*” (Efron and Tibshirani (1993), pp. 5, 10). In statistical data analysis, bootstrap means that one available sample gives rise to many others by resampling. It can be employed in either nonparametric or parametric mode. The nonparametric bootstrap, which is in the main scope of this study and is the original form of bootstrapping, will be described in this chapter. Section 3.1 presents the general framework of bootstrapping. Section 3.2 reviews arguments in the literature. The detailed principle and concept of the bootstrapping are explained in Section 3.3. The principle and algorithm of the bootstrap is described in Sections 3.4 and 3.5. The estimations that can be made via bootstrapping are explained in Sections 3.6, 3.7, and 3.8.

3.1 An Overview of Bootstrap Method

Statistical inference is used in the vast part of the applied statistics to make strategic level decisions. Two of most important problems in applied statistics are the determination of an estimator for a particular parameter of interest and the evaluation

* It is widely thought to be based on one of the eighteenth century Adventures of Baron Munchausen’s, by R.E. Raspe. The Baron had fallen to the bottom of a deep lake. Just when it looked like all was lost, he thought to pick himself up by his own bootstraps.

of the accuracy of that estimator through estimates of the standard error of the estimator. When the estimator was complex and standard approximations were neither appropriate nor accurate, estimation of the standard error of the parameter estimator is the most encountered cases during statistical inference procedure. (Chernick (1999), p. 6).Therefore, Efron (1979) proposed bootstrap resampling methods. This technique, which is commonly used for estimating bias, standard error of an estimator and confidence intervals, is further developed by Efron and Tibshirani (1993) with inferential purposes.

The main assumption of bootstrapping is based on observations' independence whose validity is necessary for other resampling methods as well. It is also assumed that sampling is performed from an infinite population where each observation has the same probability of being chosen each time. In addition, if any parametric assumptions can be made, bootstrapping also provides a way to make statistical inferences.

The basic idea behind the bootstrap method is resampling the data with replacement. Bootstrapping procedure consists of randomly choosing the sample data n times, from an original sample of size n and repeating this large number of times, say B times, with putting the chosen data back into the original set each time. Then by using the resampled (generated) data, the parameters of interests can be estimated.

Bootstrap resampling method has an extensive usage and wide application area because of its special features. Thus, bootstrapping is used in the solution procedure of a wide variety of problems appears in various disciplines including psychology, physics, geology, ecology, ornithology, econometrics, biology, meteorology, genetics, signal and image processing, medicine, engineering, reliability, chemistry, accounting, and etc. "The applications of bootstrapping in these disciplines include not only estimation of biases, standard errors or confidence intervals but also error rate estimation in discriminated analysis, subset selection in regression, density estimation, quartile estimation, p-value adjustment in multiple testing problems, estimating process capability indices, handling missing data problems, and to cope with logistic regression and classification problems, cluster analysis, kriging (i.e., a form of spatial modeling), nonlinear regression, time series analysis, complex surveys and other finite population problems, survival and reliability analysis problems" (Chernick (1999), pp. 6,7).

In this study, possible industrial area applications will be mentioned and a reliability analysis application will be given in the next chapter. Before the application part of the thesis, the bootstrapping mechanism and its all dimensions will be given in the later sections.

3.2 Arguments on Bootstrapping: A Literature Review

Bootstrap resampling strategy was introduced by Efron (1979) to assess the estimators. Therefore, statistical inference and data analysis applications made by the bootstrap resampling method has begun after this append. Although bootstrap method is nearly thirty years old, extensive literatures, researches, and projects exist in a variety of disciplines listed before. Thus, relatively a narrow review will be given. The framework of this review is drawn by Efron and Tibshirani(1993), Davison and Hinkley (1997), Good (2006), and Chernick (1999), which are used as a guide through the vast of this study. Since the basic topic in this study is bootstrapping, in general, the literature is based mostly on bootstrap resampling, its theory and application.

It should be pointed out that bootstrap research began in the late 1970s by Efron as stated before and most important theoretical development had been made after 1980s. Davison and Hinkley (1997, p. ix) denote this publication as major event in statistic since it is synthesizing some of the earlier resampling ideas and establishing a new framework for simulation based statistical analysis. This fact has been proved by the publications after Efron (1979) both in bootstrap resampling and other methods. Efron and Gong (1983) and Diaconis and Efron (1983) are the other introductory level papers about the bootstrap resampling. They argue that the resampling method frees researchers from two limitations of conventional statistics: "the assumption that the data conform to a bell-shaped curve and the need to focus on statistical measures whose theoretical properties can be analyzed mathematically". They also apply the bootstrap method to various types of problems and then compare the results taken from the bootstrap with conventional statistical tests, including the correlation coefficient and principal components. Most of the time, the bootstrap method yielded the same answers that the more conventional methods did. In some cases, bootstrap methods may not give a true picture of every sample, just as conventional tests sometimes find deceptive answers to problems. Efron (1983) compared several variations to the bootstrap estimate. He has also demonstrated the value of the bootstrap in a number of applied and theoretical contexts. Several nonparametric resampling methods are discussed in Efron (1981b) for attaching a standard error to a point estimate such as the jackknife, the bootstrap, half-sampling, subsampling, balanced repeated replications, and etc. Beran (1982) compares the bootstrap with various competitive methods in estimating sampling distributions. Parr (1983) is an early reference comparing the bootstrap, jackknife, and delta method in the context of bias and variance estimation. Efron (1987) shows that the standard approximate intervals based on maximum likelihood theory can be misleading, hence, the accuracy of confidence intervals can be improved based on transformations, bias corrections, and so forth. The proposed intervals incorporate an improvement over previously suggested methods. Three examples of the value of computer intensive inference are provided by Efron (1988). Efron and Tibshirani (1986) show the basic ideas and applications of bootstrap with some examples rather than theoretical considerations. The computational methods for the bootstrap, which are given in Efron (1990), are more efficient than the straightforward Monte Carlo methods usually used. The simplest bootstrap form (one sample nonparametric problem) is taken and bias, variance and approximate confidence interval of some statistics is computed and number of bootstrap replication is reduced. According to Simon and Bruce (1991), the method prevents researchers from simply grabbing the formula for some test without understanding why they chose that test.

In Efron (1992), relevant theoretical statistics of how the bootstrap has impacted is explained by raising six basic theoretical questions. Also, as it is mentioned before, Efron's bootstrap idea is based on iid observations and guaranteed to work with large samples. However, when small sample sizes are involved, it has been discovered through the extensive research that the bootstrap sometimes works better than conventional approaches even with small samples bootstrap resampling.

There also exist some critics for the bootstrap methods. The general question marks are based on the accuracy of the estimates that resampling mechanism yields and making enough experimental trials. In some cases resampling, and so bootstrapping, may be less accurate than conventional methods. Peterson (1991) states that using the numbers over and over again yields nothing, instead of, assumptions have to be made because the analyzer may live to regret that hidden assumptions. Noreen (1989) states several striking aspects of this approach especially for random samples drawing from different populations.

Guidelines for nonparametric bootstrap hypothesis testing are described in Efron (2000). DiCiccio and Romano (1988) present a major review article on the bootstrap and its applications. It shows that the violation of the guidelines can reduce the power of the test. DiCiccio and Efron (1996) give some heuristic overview of bootstrap confidence intervals and some methods to obtain good approximate confidence intervals are given.

Helmers et al. (1992) have mentioned that bootstrap can also be used as an estimation technique to estimate quartile. Mak (2003) is proposed a method for the simultaneous estimation of the variances of a sample statistic for all sample sizes using the bootstrap. He provides sample size determination based on a pilot sample when an explicit expression for the asymptotic variance is either too complex or unavailable at all. Since the bootstrap is employed, the method does not depend on any specific properties of the sample statistic and can therefore be universally implemented in a general computational algorithm. DiCiccio et al. (1992) have presented a new method for the construction of approximate iterated bootstrap confidence intervals which can be calculated by high coverage accuracy in small to moderately sized samples. Bickel and Krieger (1989) use the bootstrap to obtain confidence bands for a distribution function. Hall (1986) describes the number of bootstrap simulations required to construct a percentile–t confidence intervals based on a sample from continuous distribution. He has showed that smaller number of bootstrap simulations cause longer confidence intervals. Hahn and Meeker (1991) briefly discuss bootstrap confidence intervals.

Young and Daniels (1990) use the bias that is introduced in Efron's nonparametric bootstrap as a substitute for the true unknown distribution. Cheng (2001) describes some general procedures for analyzing the results of a simulation experiment using bootstrap resampling. This paper explains the rationale and simple steps needed to implement bootstrapping in estimation as well as distributional properties of the output and its dependence of factors of interest, such as; model fitting; model selection; model validation; sensitivity. Davidson and MacKinnon (2000) present a growing body of evidence from simulation experiments. They indicate that bootstrap tests do indeed yield more reliable inferences than asymptotic tests in a great many cases (for more details see Davidson and MacKinnon (2000)). Bootstrap tests will generally perform better in finite samples than asymptotic tests, and thus, errors are reduced (Hall (1992)).

Bickel and Freedman (1981) demonstrate consistency of the bootstrap under certain mathematical conditions. They also provide a counterexample for consistency of the nonparametric bootstrap.

Hall (1992) presents some sort of smoothness conditions for consistency of bootstrap estimates. Tibshirani (1992) provides some examples of the usefulness of the bootstrap in complex problems. Gine and Zinn (1989) show necessary conditions for the consistency of the bootstrap. However, examples where the bootstrap failed to be consistent due to its inability to meet certain necessary mathematical conditions are shown by Athreya (1987) and Angus (1993). Inconsistency of estimators of the bootstrap distribution is shown by Hall et al. (1993). Martin (2007) describes the construction of bootstrap hypothesis tests which can differ from bootstrap confidence intervals because of the requirements to generate the bootstrap distribution of test statistics under a specific null hypothesis. The performance of bootstrap tests, examining size and power properties of the tests numerically using both real and simulated data is critically assessed. Bootstrap power calculations for some scenarios are also described.

There exists several review papers that compare the resampling methods and argue that which method is best under which conditions. The basic comparisons are to be made among the three resampling methods: bootstrap, jackknife, and cross-validation. The principles of these three methods are similar, but bootstrap is defined as "more thorough" procedure since it can generate many sub-samples than others. Efron (1982) compares these resampling methods. Through resampling procedure vast of the paper found that the bootstrap resampling method provides less biased and more consistent results than the jackknife method does. On the other hand, in rare, such as Mooney and Duval (1993), it is suggested to use jackknife since it has been largely studied than the others. However, unlike the jackknife the bootstrap is applicable more widely, and gives better results in many situations. Therefore jackknifing is recommended mainly if one needs only a variance estimator. Efron and Tibshirani

(1993, pp. 10-15) state that unlike randomization tests, statistical inference for the bootstrap applies to the population characteristics. The common use of bootstrap methods is to provide estimates of standard errors, approximate confidence intervals for unconventional statistics and approximate probability estimates relative to some null hypotheses. Bootstrap hypothesis tests are often inferior to tests based on parametric or permutation methods when such equivalent methods exist. There exists an important difference between randomized and bootstrap sample. Unlike randomization method, a bootstrap sample is generated by sampling with replacement from the original sample. Therefore, randomization method is appropriate when the order or association between parts of the data is assumed to be important.

3.3 Bootstrapping

Bootstrap method is an application of simulation ideas to the problem of statistical inference. The idea particularly comes from where the bootstrap method is assumed as an imitation of real world. Thus, the real (original) world has observable information about the population whereas the bootstrap world try to act as the original world because it is created from original by data generating mechanism, i.e. the bootstrapping.

The bootstrapping aims to draw best representative picture or imitation of the population. The critical point, here, is to obtain the best imitation, i.e. bootstrap samples. The bootstrap sample has the same size with the original sample and consists of members of the original sample. However, the appearance of the data changes from sample to sample such as some datum may appear zero times whereas some may appear more than one times. The bootstrap samples are generated with replacement from original data set and generation is repeated many times as schematized in Efron and Tibshirani (1993, p.13) (see Figure 3.1) to estimate a parameter, the standard error or a confidence interval for the parameter or test a hypothesis about the parameter in situation where there is a random sample from an unknown distribution. A bootstrap sample $X^*=\left(X_1^*,X_2^*,\ldots,X_n^*\right)$ is drawn from the random sample $X=\left(X_1,X_2,\ldots,X_n\right)$. The star (*) notation indicates that X^* is not the real world's data set, but it is a resampled or imitated version of X.

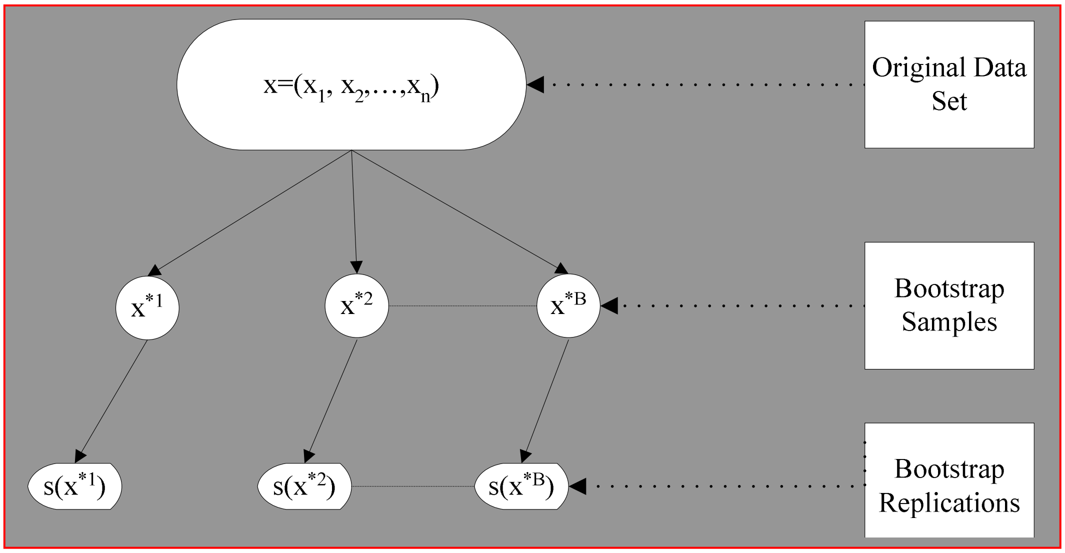

Figure 3.1: Schematic representation of the bootstrap process for estimating standard error of a statistic s(x)

Let F be a probability distribution. A random sample, which is taken from this distribution, is denoted by X. Let **x** be a vector of random data points x_i such as $\mathbf{x}=\left(x_1,x_2,\ldots,x_n\right)$ and $t(.)$ be a numerical evaluation procedure, then

$$\theta=t(F) \tag{3.1}$$

will be some numerical evaluation to the distribution function F. Thus, (3.1) represents the parameter of F (Efron (1993)). The statistic S(.) can be used to calculate an estimate $\hat{\theta}$ of the parameter θ as

$$\hat{\theta}=S(x) \tag{3.2}$$

when the distribution F is unknown. In this case, the empirical distribution $\hat{F}$, which is a discrete distribution, is an estimate of distribution F and is obtained by assigning a probability $\frac{1}{n}$ on each x_i as stated before. Since it is a random sampling with replacement procedure, some values may be observed several times in sample $\underset{\sim}{\mathbf{x}}$. Thus, empirical distribution can be denoted by the proportion of times each value occurs and the probability of the k^{th} item in the empirical distribution would be

$$\hat{f}_k = \frac{\#\{x_i = k\}}{n}. \tag{3.3}$$

The equation (3.3) is also defined as the frequency of the value k in the sample and the EDF would be

$$\hat{F} = \left(\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_K\right). \tag{3.4}$$

In equation (3.4) K represents the number of different values in the sample where $K \leq n$, the sample size. This procedure is used to estimate the parameter(s) of the unknown distribution.

3.4 Plug-In Principle

The plug-in principle is a simple method of estimating parameter from samples. The plug-in estimate of parameter

$$\theta = t(F) \tag{3.5}$$

is

$$\hat{\theta} = t(\hat{F}), \tag{3.6}$$

where the parameter θ is estimated by the function of the empirical distribution $\hat{F}$. In equation (3.5), the estimate $\hat{\theta}$ was defined trough the sample $\underset{\sim}{\mathbf{x}}$, which is the same as equation (3.6), since the empirical distribution is constructed from the sample $\underset{\sim}{\mathbf{x}}$. "The statistics like $\hat{\theta}$ that are used to estimate parameters are sometimes called summary statistics, as well as estimates and estimators" (Efron and Tibshirani (1993), pp. 35-36). The bias and standard error of plug-in estimate is going to be used in the bootstrap application. The bootstrap's advantage here is that it produces biases and standard errors in an automatic way, no matter how complicated the functional mapping of (3.5).

The plug-in principle is less good in situations where there is information about F other than that provided by the sample $\underset{\sim}{\mathbf{x}}$. Thus, it is not recommended to use a plug-in estimate of a statistic in parametric approaches (Efron and Tibshirani (1993), pp. 35-37).

3.5 Algorithm and Principle of the Bootstrap

Bootstrap algorithm has three basic steps defined as bootstrap samples, bootstrap replications and bootstrap estimates. These are given as following:

Step 1 (Bootstrap Samples): B independent bootstrap samples x^{*b}, where $b=1,...,B$, are drawn.

Step 2 (Bootstrap Replications): The bootstrap replications of $\hat{\theta}$ for each of the B independent samples should be evaluated. Efron and Tibshirani (1993, pp. 45-49) define this evaluation as fallows:

Let $S(.)$ be a statistic, and $S(x)$ would be the statistic taken from the original data set stated before. Then, $S(x^*)$ is the bootstrap replication of the statistic, e.g. if $S(.)$ is the sample mean, $S(x^*)$ is the mean of the bootstrap sample. The bootstrap replication of the estimate $\hat{\theta}$ is then

$$\hat{\theta}^*=S(x^*). \tag{3.7}$$

The equation (3.7) represents an estimate for parameter θ based on the bootstrap data set x^*. Thus, for each sample b, $b=1,...,B$

$$\hat{\theta}(b)^*=S(x^{*b}) \tag{3.8}$$

should be evaluated. The quantity of (3.7) is the result of applying same function $S(.)$ to x^* as was applied to x. For example, if $S(x)$ is the sample mean $\bar{x}$ then $S(x^*)$ is the mean of the bootstrap data set, $\bar{x}^* = \sum_{i=1}^{n} x_i^*/n$.

Step 3 (Bootstrap Estimates): The bootstrap estimate of the statistic is evaluated by using B bootstrap estimate.

The number of bootstrap replications should be decided based on the target accuracy of the estimates. However, there is a trade-off between computation time and the accuracy of the approximation to the sampling distribution of $\hat{\theta}$. In this case parallel computational technique (each resample can be done independently of the others) can be used to cope with this trade-off. Unfortunately, most of authors propose to use a general rule of thumb where B is taken as 1000 (Mooney and Duval (1993); Efron and Tibshirani (1993)).

There exist a basic assumption to generate samples by using bootstrap method and its algorithm. Under certain conditions, the variability of $\hat{\theta}$ around the value θ can be assessed via the variability of $\hat{\theta}_i^*$ around the value $\hat{\theta}$. This assumption is called as bootstrap principle. In many situations, it is expected that this condition is satisfied. However, there are some situations in which the assumption does not hold and bootstrap resampling method is inappropriate.

3.6 The Bootstrap Estimate of Standard Error

The bootstrap is firstly introduced as a computer intensive method especially for estimating standard error of some estimator, say $\hat{\theta}$, in 1979. The bootstrap estimate of standard error requires no theoretical calculations, and is available no matter how mathematically complicated the estimator $\hat{\theta}$ may be. The estimation of standard error and the bias depend on the notation of a bootstrap sample. Therefore, $X=\left(X_1,X_2,...,X_n\right)$ be a random sample from an unknown probability distribution F and a parameter of interest $\theta=t(F)$ is to be estimated on the basis of X. Let define again the empirical distribution $\hat{F}$ be an estimate of distribution F and be obtained by assigning a probability $\frac{1}{n}$ on each value x_i for $i=1,...,n$. Thus, an estimate $\hat{\theta}=S(x)$ should be calculated where the $S(x)$ may be the plug-in estimate $t(\hat{F})$ as in (3.6). A bootstrap sample $X^*=\left(X_1^*,X_2^*,...,X_n^*\right)$ is taken from $\hat{F}$ and corresponding to a bootstrap data set x^* is a bootstrap replication of $\hat{\theta}$ as in the equation (3.7).

Let the standard error of a statistic $\hat{\theta}$ be $se_F(\hat{\theta})$. Then, the bootstrap estimate of the standard error of the statistic $\hat{\theta}$, $se_{\hat{F}}(\hat{\theta}^*)$, is a plug-in estimate that uses empirical distribution function $\hat{F}$ instated of unknown distribution F. Consequently, the bootstrap estimate of $se_F(\hat{\theta})$, which is the standard error of $\hat{\theta}$ for data sets of size n randomly sampled from $\hat{F}$, is called the *ideal bootstrap estimate* of standard error of $\hat{\theta}$.

The ideal bootstrap estimate of standard error of $\hat{\theta}$ can be evaluated by the algorithm given in Section 3.5. The algorithm works by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications, and estimating the standard error of $\hat{\theta}$ by the empirical standard deviation of the replications such as

$$\widehat{se}_B=\left\{\frac{\sum_{b=1}^{B}\left[\hat{\theta}^*(b)-\hat{\theta}^*(.)\right]^2}{B-1}\right\}^{1/2}, \tag{3.9}$$

where

$$\hat{\theta}^*(.)=\frac{\sum_{b=1}^{B}\hat{\theta}^*(b)}{B}. \tag{3.10}$$

and B is the number of replications. The limit of $\widehat{se}_B$ as B goes to infinity is the ideal bootstrap estimate of $se_F(\hat{\theta})$, which is given by

$$\lim_{B\to\infty}\widehat{se}_B=se_{\hat{F}}=se_{\hat{F}}(\hat{\theta}^*). \tag{3.11}$$

The limit in (3.11) means that an empirical standard deviation approaches the population standard deviation as the number of replications grows large. The ideal bootstrap estimate defined above is a nonparametric bootstrap estimate since it is based on $\hat{F}$, the nonparametric estimate of the population F.

3.7 Bootstrap Estimate of Bias and the Bias Correction

The standard error is a measure of accuracy for an estimator $\hat{\theta}$. There are also some other measures like bias that provides different frameworks of $\hat{\theta}$'s behavior for statistical accuracy. Bias can be defined as the difference between the expectation of an estimator $\hat{\theta}$ and the quantity θ being estimated. The bootstrap resampling method is also used as an important tool to estimate the bias of an estimator and the bootstrap algorithm can be used for the estimation of bias with some adjustments and adaptations.

Let's assume that same statistical conditions given previous section is still valid. An unknown probability distribution F has given data $X^*=\left(X_1^*,X_2^*,...,X_n^*\right)$ by random sampling. The real valued parameter $\theta=t(F)$ is to be estimated and for this reason an estimator $\hat{\theta}=S(x)$ is taken to find the amount of bias. The bias of $\hat{\theta}=S(x)$ (an estimate of θ) is defined as

$$bias_F=bias_F(\hat{\theta},\theta)=E_F\left[s(x)\right]-t(F) \tag{3.12}$$

Estimator's performance can be evaluated by comparing the associated amount of bias. Thus, a large quantity of the bias is an undesirable case. Unbiased estimates, which are mathematically shown as $E_F\left[s(x^*)\right] - t\,(\hat{F})$, promote the scientific objectivity of the estimation process. Plug-in estimate given in (3.12) does not have to be necessarily unbiased, but they tend to have small biases compared to the magnitude of their standard errors (Efron and Tibshirani (1993), pp. 124-125). This is why the plug-in principle is used in bootstrap resampling.

The bootstrap estimate of bias of an estimator $\hat{\theta}=S(x)$ that is shown in a different form from (3.12) can be defined as

$$\text{bias}_{\hat{F}}=E_{\hat{F}}\left[S(x^*)\right]-t(\hat{F})\,. \qquad (3.13)$$

In equation (3.13), $\hat{F}$ is substituted with F in the formula given in (3.12). Efron and Tibshirani (1993, pp.124-125) describe that $t(\hat{F})$, i.e. the plug-in estimate of θ, may differ from $\hat{\theta}=S(x)$. They also show that the $\text{bias}_{\hat{F}}$ is plug-in estimate of bias_F, whether or not $\hat{\theta}$ is the plug-in estimate of θ.

The ideal bootstrap estimate of bias $\text{bias}_{\hat{F}}$ can be found with the bootstrap algorithm given before. However, some adaptations have to be made in the last step as finding approximate bootstrap expectation $E_{\hat{F}}\left[S(x^*)\right]$ by the average

$$\hat{\theta}^*(.)=\frac{\sum_{b=1}^{B}\hat{\theta}^*(b)}{B}=\frac{\sum_{b=1}^{B}s(x^{*b})}{B}. \qquad (3.14)$$

Then, the bias of bootstrap estimate based on B replications, $\widehat{\text{bias}}_B$, will be

$$\widehat{\text{bias}}_B=\hat{\theta}^*(.)-t(\hat{F}). \qquad (3.15)$$

The mean square error then is

$$\widehat{\text{MSE}}=\frac{1}{B}\sum_{i=1}^{B}\left(\hat{\theta}_i^*-t(\hat{F})\right)^2.$$

In the statistical inference, in general, estimation of the bias is required to correct the estimator $\hat{\theta}$ so that the accuracy of the estimator increases. If the $\widehat{\text{bias}}$ is an estimate of $\text{bias}_F(\hat{\theta},\theta)$, then the bias corrected estimator is

$$\bar{\theta}=\hat{\theta}-\widehat{\text{bias}}. \tag{3.16}$$

If the (3.15) equals to (3.16), bias corrected estimator is given by

$$\bar{\theta}=2\hat{\theta}-\hat{\theta}^*(.). \tag{3.17}$$

Efron and Tibshirani (1993, p. 138) noticed that the bias correction can be dangerous in practice due to high variability in $\widehat{\text{bias}}$. Even if $\bar{\theta}$ is less biased than $\hat{\theta}$, it may have substantially greater standard error. Thus, correcting the bias may cause a larger increase in the standard error and a larger mean square error is observed. Therefore, this should be checked with the bootstrap. If $\widehat{\text{bias}}$ is small compared to the estimated standard error $\widehat{\text{se}}$, then it is safer to use $\hat{\theta}$ than $\bar{\theta}$. If $\widehat{\text{bias}}$ is large compared to the estimated standard error $\widehat{\text{se}}$, then it may indicates that the statistic $\hat{\theta}=S(x)$ is not an appropriate estimate of the parameter θ.

3.8 Bootstrap Confidence Interval

Confidence intervals are statistical inference tools used in applied statistics. The confidence intervals combine the interval estimation and hypothesis testing into a single statistical inference procedure and give the range of plausible values for the statistics. Confidence intervals (or interval estimates) are often more useful than just a point estimate. In the statistical inference procedure, if point and interval estimates are taken together, they give information about what the best guess is for the parameter to be estimated and how far in error that guesses might be. Most confidence intervals are approximate and favorite approximation being the standard interval

$$\hat{\theta}\pm z^{(\alpha)}\hat{\sigma}$$

where $\hat{\theta}$ is a point estimate of the parameter of interest θ, $\hat{\sigma}$ is an estimate of $\hat{\theta}$'s standard deviation, and $z^{(\alpha)}$ is the $100\alpha^{th}$ percentile of a standard normal distribution. The trouble with "standard intervals" is that they are based on an asymptotic approximation that can be quite inaccurate in practice (DiCiccio and Efron (1996)). Over the years statisticians have developed tricks for improving standard interval with bias corrections and parameter transformations. There are several approaches to construct an approximate 100(1-α)% confidence interval for θ using the bootstrap sample. The original approach described by Efron is known as the percentile confidence interval. The other main approaches are discussed in DiCiccio and Efron (1996); Efron and Gong (1983) and Efron and Tibshirani (1986, 1993). They state that the bootstrap confidence interval can be evaluated from automatic algorithms for carrying out these improvements without human intervention. For producing good confidence intervals and improving the accuracy, five kinds of bootstrap confidence intervals have been developed: the standard bootstrap (SB) confidence interval, bootstrap-t confidence interval, the percentile bootstrap (PB) confidence interval, bias corrected and accelerated (BC_α), and approximate bootstrap confidence intervals (ABC).

3.8.1 Standard Bootstrap Confidence Interval

Standard bootstrap confidence intervals are based on the assumption that the estimator $\hat{\theta}$ is normally distributed with mean θ and variance σ^2. An approximate 100(1-α)% confidence interval is given by

$$\hat{\theta} \pm z_{\alpha/2}\sqrt{Var_{bs}[\hat{\theta}]}\,, \qquad (3.18)$$

where

$$Var_{bs}[\hat{\theta}] = \frac{\sum_{b=1}^{B}(\hat{\theta}_{(b)} - \hat{\theta}_{(.)})^2}{B-1}$$

is bootstrap estimate of variance. The standard confidence interval is improved by using the equation (3.17), the bias corrected estimate of θ. Then, the new form of standard confidence interval defined as

$$2\hat{\theta}-\hat{\theta}_{(.)}\pm z_{\alpha/2}\sqrt{\mathrm{Var}_{bs}[\hat{\theta}]}\,. \tag{3.19}$$

It should be noticed that, the estimated variance is the square of standard error presented in the previous sections and $\hat{\theta}$ is the plug-in estimate for statistic θ. The standard confidence interval here gives the range of plausible values for the statistic θ. Thus, it is estimated in which range is the actual value of θ likely to be.

3.8.2 Bootstrap-t Confidence Interval

The standard confidence interval holds for normal distributions and any other distribution when the sample size n is large (i.e. the distribution of $\hat{\theta}$ approaches to normal distribution according to central limit theorem). But for small sample sizes the standard confidence intervals are not accurate. Thus, without making the normality assumptions an accurate bootstrap confidence interval can be obtained. Bootstrap-t confidence intervals, which are improved for small samples, are evaluated by following procedure:

First, B bootstrap samples are generated in order to approximate pivot that is computed for each sample x^{*b} as

$$Z^{*}(b)=\frac{\hat{\theta}^{*}(b)-\hat{\theta}}{\widehat{se}^{*}(b)}\,, \tag{3.20}$$

where $\hat{\theta}^{*}(b)=S(x^{*b})$, stands for the value of estimator $\hat{\theta}$ for the bootstrap sample x^{*b}; $\widehat{se}^{*}(b)$ is the estimated standard error of $\hat{\theta}^{*}$ for the sample x^{*b}, $b=1,2,...,B$. The approximate pivot means that its distribution is approximately the same for each value of θ (i.e. independent of θ).

Second, the α^{th} percentile of $Z^*(b)$ in (3.20) estimated by $\hat{t}^{(\alpha)}$ and defined as

$$\frac{\#\left\{Z^*(b) \le \hat{t}^{(\alpha)}\right\}}{B}=\alpha\,. \tag{3.21}$$

Equation (3.21) is the proportion of the number of observed values lower than and equal to $\hat{t}^{(\alpha)}$ to the number of replication B. If $B\alpha$ is not an integer, then $\left[(B+1)\alpha\right]$ is used.

Finally, the bootstrap-t confidence interval will be

$$\left(\hat{\theta}\text{-}\hat{t}^{1-\alpha}\widehat{se};\hat{\theta}\text{-}\hat{t}^{\alpha}\widehat{se}\right). \tag{3.22}$$

The number of bootstrap replications here should be set large enough (i.e. $B \ge 1000$) to provide an accurate confidence interval.

3.8.3 Bootstrap Percentile Confidence Interval

Another approach to bootstrap confidence interval is based on the percentiles of the bootstrap distribution of a statistic. Since this confidence interval has somewhat different view of the standard normal theory, it results a bootstrap confidence interval with reasonable stability in practice.

Let $\hat{\theta}$ be the plug-in estimate of a parameter θ and $\widehat{se}$ be its estimated standard error. If the standard confidence interval is considered again as $\left[\hat{\theta}\text{-}z^{(1-\alpha)}\widehat{se};\hat{\theta}\text{-}z^{(\alpha)}\widehat{se}\right]$, Efron and Tibshirani (1993, pp. 168-169) state that the endpoints of this interval can be described in a way that is particularly convenient for bootstrap calculations. Let $\hat{\theta}^*$ again indicate a random variable drawn from the normal distribution as in the standard confidence interval case $\hat{\theta}^* \sim N(\hat{\theta},\widehat{se}^2)$ then the lower and upper confidence limits will be

$$\hat{\theta}_{\text{lower}}=\hat{\theta}\text{-}z^{(1-\alpha)}\widehat{se} \tag{3.23}$$

$$\hat{\theta}_{upper}=\hat{\theta}-z^{(\alpha)}\widehat{se} \tag{3.24}$$

which are $100\alpha^{th}$ and $100(1-\alpha)^{th}$ percentiles of $\hat{\theta}^*$, respectively. The confidence limits given in (3.23) and (3.24) are more direct approaches for constructing a confidence interval since it uses the upper and lower α values of the bootstrap distribution. If $\hat{G}$ denotes the cumulative distribution function of $\hat{\theta}^*$, the 1-2α percentile interval is defined by the α and 1-α percentiles of $\hat{G}$ given by

$$\left[\hat{\theta}_{\%,lower},\hat{\theta}_{\%,upper}\right]=\left[\hat{G}^{-1}(\alpha),\hat{G}^{-1}(1-\alpha)\right].$$

In other words

$$\left[\hat{\theta}_{\%,lower},\hat{\theta}_{\%,upper}\right]=\left[\hat{\theta}^{*(\alpha)},\hat{\theta}^{*(1-\alpha)}\right]$$

since by definition $\hat{G}^{-1}(\alpha)=\hat{\theta}^{*(\alpha)}$, the $100\alpha^{th}$ percentile of the bootstrap distribution. The confidence interval given above is the ideal form since the number of the replication is assumed to be infinite. However, in practice it is required that some finite number of replications B should be used. Thus, B independent bootstrap data set $x^{*1},x^{*2},...,x^{*B}$ is generated and $\hat{\theta}^*(b)=S(x^{*B})$ is computed for bootstrap replications (b=1,2,...,B). If we denote $\hat{\theta}_B^{*(\alpha)}$ as $100\alpha^{th}$ empirical percentile of the $\hat{\theta}^*(b)$ values, that is, the $B\alpha^{th}$ value in the ordered list of the B replications of $\hat{\theta}^*$. Likewise $\hat{\theta}_B^{*(1-\alpha)}$ be the $100(1-\alpha)^{th}$ empirical percentile. Efron and Tibshirani (1993, pp. 170-176) have shown that percentile confidence interval gives better results than standard bootstrap confidence intervals do.

3.8.4 Bias-Corrected and Accelerated Bootstrap Confidence Interval

Bias-corrected and accelerated bootstrap confidence interval (BC_α) is an improved version of the percentile method providing closely match exact confidence intervals and giving accurate coverage probabilities in all situations. Since neither bootstrap percentile nor bootstrap-t gives such kind of confidence interval, this method is proposed to satisfy more coverage property.

Let $\hat{\theta}^{*(\alpha)}$ indicate the $100\alpha^{th}$ percentile of B bootstrap replication $\hat{\theta}^{*}(1),\hat{\theta}^{*}(2),...,\hat{\theta}^{*}(B)$. The percentile interval $(\hat{\theta}_{lower},\hat{\theta}_{upper})$ of the coverage $1\text{-}2\alpha$, is obtained from

$$\left[\hat{\theta}_{lower},\hat{\theta}_{upper}\right]=\left[\hat{\theta}^{*(\alpha)},\hat{\theta}^{*(1\text{-}\alpha)}\right]. \tag{3.25}$$

In the BC_{α} interval limits are also given by the percentiles as in (3.25) with a $1\text{-}2\alpha$ coverage obtained by

$$\left[\hat{\theta}_{lower},\hat{\theta}_{upper}\right]=\left[\hat{\theta}^{*(\alpha_1)},\hat{\theta}^{*(\alpha_2)}\right], \tag{3.26}$$

where

$$\alpha_1=\Phi\left(\hat{z}_0+\frac{\hat{z}_0+z^{(\alpha)}}{1\text{-}\hat{\alpha}(\hat{z}_0+z^{(\alpha)})}\right) \tag{3.27}$$

and

$$\alpha_2=\Phi\left(\hat{z}_0+\frac{\hat{z}_0+z^{(1\text{-}\alpha)}}{1\text{-}\hat{\alpha}(\hat{z}_0+z^{(1\text{-}\alpha)})}\right). \tag{3.28}$$

In fact, the percentile confidence limits take a different form from (3.27) and (3.28). The basic difference is (3.26) depends on acceleration ($\hat{\alpha}$) and bias-correction ($\hat{z}_0$) where $\Phi(.)$ is the standard normal cumulative distribution function and $z^{(\alpha)}$ is the $100\alpha^{th}$ percentile point of a standard normal distribution. The value of bias correction is evaluated by

$$\hat{z}_0=\Phi^{-1}\left(\frac{\#\left\{\hat{\theta}^{*}(b)<\hat{\theta}\right\}}{B}\right), \tag{3.29}$$

where (3.29) gives the proportion of bootstrap replications less than the original estimate $\hat{\theta}$; hence, we obtain $\hat{z}_0=0$ if exactly half of $\hat{\theta}^{*}(b)$ values are less then or equal to $\hat{\theta}$. Here, $\Phi^{-1}(.)$ gives the inverse function of standard normal cumulative distribution function, e.g., $\Phi^{-1}(.95)=1.645$.

The value of acceleration is evaluated by

$$\hat{\alpha}=\frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(.)}-\hat{\theta}_{(i)}\right)^{3}}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(.)}-\hat{\theta}_{(i)}\right)^{2}\right\}^{3/2}}. \tag{3.30}$$

Equation (3.30) refers to the rate of change of the standard error of $\hat{\theta}$ with respect to true parameter value θ.

It should be noticed that if $\hat{\alpha}$ and $\hat{z}_0$ are zero, then BC_α is the same as the percentile interval. The non-zero values of $\hat{\alpha}$ or $\hat{z}_0$ change the percentiles used for the BC_α interval limits. Efron and Tibshirani (1993, pp. 178-190) have shown how these values correct certain deficiencies of the percentile methods.

3.8.5 Approximate Bootstrap Confidence Interval

Approximate bootstrap confidence interval (ABC) is a method of approximating the BC_α limits without using Monte Carlo replications at all to compensate the requirement of large number of replications. This method works with approximation of bootstrap random sampling results by Taylor series expansions. Since this study basically relies on Monte Carlo replications, this topic is assumed as out of the scope of this study. The detailed information can be seen in Efron and Tibshirani (1993).

CHAPTER 4

AN INDUSTRIAL ENGINEERING APPLICATION: LIFETIME DATA ANALYSIS

The primary aim of this study is to develop an efficient lifetime data analysis that takes into consideration of real life conditions by using nonparametric bootstrap procedure. A new incomplete data model, i.e. trunsored model, will be used to analyze the data. This method provides not only estimate the ratio of the fragile population to the mixed fragile and durable populations, but also tests the hypothesis that the ratio is equal to a prescribed value with estimated confidence intervals. Therefore, a representative application of this method is used in the analysis of thermal endurance of coil used in a special lamp. Reliability of the lamps is determined to observe their performance. In fact, the effectiveness of lamps is determined since the intended performance of the lamps shows their effectiveness and reliability is one of important attribute of effectiveness (Kales (1998), pp. 6-7).

The engineers dealing with reliability eliminates early failures by observing their distribution, eliminating the appropriate debugging method, and the length of the debugging period. Then, they observe the statistical distribution of wearout and determine the preventive replacement periods for the various parts or their design life. Also they pay attention to chance failures and their prevention, reduction or complete elimination in the scope of a reliability improvement program.

The reliability improvement plan should be designed to optimize reliability at the same time reducing costs and increasing output without increasing unit costs and increasing customer satisfaction. The first step in such a program should integrate the reliability and product assurance programs to all available company activities like purchasing, engineering, research, manufacturing, quality control, packaging, shipping and performance feedback. The others should be selection of better raw materials, reduction of the number of components that makes up the product, using reliability check lists in all phases of the product life (design, development, manufacturing, and service life), implementation of an information feedback, analysis and control systems, and implementation of a failure mode and effects analysis. One of the key points here is implementing the reliability improvement program into the manufacturing processes and quality control. For that reason, as a representative industrial engineering application of the bootstrap resampling method, some special reliability data are taken from a firm. The firm controls and analyzes the lifetime of products to quantify the lifetime standards. Since a confidentiality agreement is made with the firm, the name of the firm and the details of the product will not be defined during the study.

In the application procedure, lifetime data analysis is made for an implementation of a reliability program. The main advantage of the study is having a chance to apply the bootstrap resampling method to lifetime data set and analyze the result. In addition to this, different perspectives are obtained via getting better information about the types of failures experienced by parts and systems that aid design, research, and development efforts to minimize these failures, estimation of the failure ratio for both new and old design products, and getting estimations of the required redundancy to achieve the specified reliability. Thus, in this chapter, basics and application of trunsored models will be presented as a new perspective in lifetime data analysis after presentation of some basics of reliability and lifetime data. Section 4.1 defines the reliability. Section 4.2 explains the key points for achieving reliability. Sections 4.3 and 4.4 explain the time and data perspective of reliability and trunsored data model.

Section 4.5 gives the details of the problem and lifetime experiment. Section 4.6 explains how bootstrapping is applied with censored data and its possible application. Finally, Section 4.7 describes the trunsored model constructed for the lifetime estimation.

4.1 Definition of Reliability

The reliability of a component is the probability that the component will perform a specified function under specified operational and environmental conditions, at and throughout a specified time as reported by Kales (1998, p. 7). This probability deals with the laws of random chance of lives or failures as they appear in nature. Thus, reliability refers to the chance, or likelihood, which the device will work properly.

Reliability was developed to provide methods for assuring a product or service functions. These methods consist of techniques for

- determining what can go wrong,
- how it can be prevented from going wrong ,
- if something goes wrong, how it can be quickly recovered and consequences can be minimized.

In order to assure for product or service functions, product and service standards (specifications) must be assessed. Satisfactory reliability specifications are given below, which are pointed out by Doty (1989).

1. state exactly what is wanted;

2. explain the methods and procedures, including sampling and computations and provide the means for test;

3. avoid nonessential quality restrictions that add to cost without adding to utility;

4. conform, as far as possible, to the established commercial standards;

5. explain condition and where the product is to be used;

6. contain a statement of purpose of the product as guide to usage and against misapplication;

7. explain the inspection and testing procedure to be used in determining conformance to the standards, including instrument and personnel;

8. state the applicable standards, including tolerances;

9. contain a statement of time frame;

10. define failures in terms of product use and explain how they are to be measured;

11. state the maintenance procedures and contains.

Adequate performance must be defined in term of a time frame, and hence the standard must include a time limit. In reliability, there are several different types of time frames: total test time, test period, mean test time, mean repair time, allowed repair time, and mission time. Finally, operating conditions and environmental conditions must also be included in the reliability standards.

4.2 Achieving Reliability

Modern programs are implemented for achieving and improving reliability of existing products and for assuring continued high reliability for the next generations. Modern programs involve the followings to achieve and improve reliability:

Emphasis: Increased emphasis is being given to product reliability because of that product is more complicated and emphasis is due to the consumer protection act.

System Reliability: If products become more complex, the chance that they will not function increases. The method of arranging the components affects the reliability of the entire system. Components can be arranged in series, parallel or in combination.

Design: The important aspect of the reliability is the design. It should be as simple as possible. The fewer the number of components have the greater reliability. Another way of achieving reliability is backup or redundant component. When primary component does not function, another component is activated. Parallel arrangement of component is cheaper to have inexpensive redundant components to achieve a particular reliability than to have a single expensive component. Reliability can also be achieved by over design. Using factors of safety can increase the reliability of a product which is determined by Besterfield (2001, pp. 419-443).

Environmental Conditions: Dust, temperature, moisture, and vibration can be cause of an unreliable product. The designer must protect the product from these conditions. Heat shields, rubber vibration mounts, and filters are used to increase the reliability under environmental condition.

Production: The production process is the second most important aspect of reliability. Basic quality control techniques minimize the risk of product unreliability. Production personnel can experiment with process conditions to determine which condition produce the most reliable product.

Transportation: The third aspect of reliability is the transportation of the product to the consumer. The reliability of the product can be greatly affected by the type of handling. Good packaging techniques and shipment evaluation are essential.

Maintenance: While designers try to eliminate the need for customer maintenance, there are many situations where it is not practical or possible. Maintenance should be simple and easy to perform.

4.3 Reliability Data

Bootstrap resampling method is used for the reliability data. Thus, the reasons for collecting reliability data are defined in the following (Meeker and Escobar (1998), p. 2):

- assessing characteristics of materials over warranty period or over the product's design life;
- predicting product reliability;
- predicting product warranty costs;
- providing needed inputs for system-failure risk assessment;
- assessing the effect of a proposed design change;
- assessing whether customer requirements and government regulations have been met;
- tracking the product in the field to provide field to provide information of cause of failure and methods of improving product reliability;
- supporting programs to improve reliability through the use of laboratory experiments, including accelerated life tests;
- comparing components from two or more different manufacturers, materials, production periods, operating environments, and etc;
- checking the velocity of an advertising claim.

It is often necessary to use past experience and observations or other scientific and engineering judgment provide information as input data analysis that requires the use of special statistical methods. Owing to the fact that reliability data have a number of special features, e.g. reliability data are typically *censored* or *truncated* which are defined in the later sections, they are analyzed by using different statistical methods. Hence, as the first of all, types of life time data will be defined in the following sections.

4.3.1 Types of Reliability Data

Statistical models briefly based on data to make predictions. The models are the statistical distributions and the data are the life data or times-to-failure data of the related component. The accuracy of any prediction is directly proportional to the quality and accuracy and completeness of the supplied data. Good data, along with the appropriate model choice, usually results in good predictions. Insufficient data will almost always result in bad predictions.

Some synonyms for reliability data are used in the literature as failure time data, life data, survival data (used in medicine and biological sciences), and event time data (used in social sciences) (Meeker and Escobar (1998), p. 3). In the remaining part of this study, life data or lifetime data is used because of the application part.

In the analysis of life data, the main aim is to use the all available data. In some cases, the data set is incomplete which includes uncertainty as a failure occurred. To accomplish this, the collection types of data are separated into two based categories; complete (all information is available) or incomplete (some of the information is missing) as given in the following:

Complete Data: Complete data means that the results of each sample exist, either observed or known. Through life data analysis, the data would involve the times-to-failure of all units in the whole data set. The whole application is continued up to all data through the sample (see Figure 4.1).

Accelerated Data: It is often the case that systems have very long expected life under normal conditions and a useful test may not be possible for long duration of testing the whole sample to wait and see whether a failure occur or not. During the observation, to obtain data from such kind of application, life testing is conducted at higher than normal stress levels in order to induce failures. The data obtained under high stress level conditions are called accelerated data, which are observed for further implementation of data analysis, is defined by Zacks (1992, p. 195).

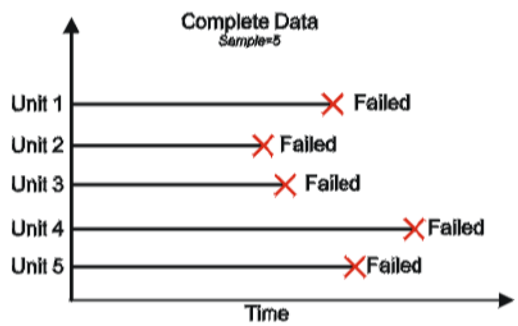

Figure 4.1*: Representation of Complete Data over Time

Incomplete Data: Some units in the sample may not have failed in a specific period of time during the life time analysis of the data. This restriction causes that the failure times of all the units cannot be observed exactly and, hence, this type of life time data is also called as “censored”. There are four types of possible incomplete (or censored) data form, namely, right censored (suspended data), interval censored, left censored, and a special case of censored data as truncated data type. Meeker and Escobar (1998, p. 34) denote the reasons for censoring as in the following:

➢ There exist some restrictions for the life test period. Therefore, analysis of life test data should be completed before all units have failed. Because of the restrictions one can use "time censoring" (Type I censoring) method, in which the unfailed units are removed from the test at a prescribed time or the method "failure censoring" (Type II censoring), in which a life test is terminated after a specified number of failures. Although the failure-censored data can be statistically analyzed easier than time-censored data, failure-censored tests are less common in practice.

➢ In many life tests, since the failures observed only at times of inspection, observations consist of upper and lower bounds on a failure time. Hence, these life time data is known as interval censored data (or inspection data, grouped data, read-out

* http://www.weibull.com/LifeDataWeb/data_classification.htm

data). If a unit has failed at its first inspection, it is the same as a left-censored observation and if a unit has not failed by the time of the last inspection, it is right-censored which are defined in the next section.

- In some situations, products may have more than one cause of failure. If one focus on a particular cause of failure, the failure from other causes is defined as a form of random right censoring.
- In some life tests, units are put on test at different times. This is known as staggered entry. If the data are to be analyzed at a point in time when not all units have failed, the data will be classified as multiply right-censored.

Meeker and Escobar (1998, p. 35) state some assumptions for the use of most reliability models and methods that analyze censored data as in the following:

1. censoring time can be either random or predetermined;

2. censoring time of a unit should depend only on the history of the observed failure-time process in order to the analysis is valid;

3. using future events to stop observing a unit could introduce bias.

Meeker and Escobar (1998, pp. 34-41) defines four types of censoring mechanism as in the following:

1. Right Censored Data: This type of censored life data is the most common case in reliability applications. These data are composed of units that did not fail. The term "right censored" means that the interested event (i.e. the time-to-failure) falls in the right of the data point. In other words, during the operation of the components, the failure occurs at some time after the data point (or to the right on the time scale that can be seen in Figure 4.2).

2. Interval Censored Data: The second type of censoring is called interval-censored data. The failing of components occur during an intervals with uncertainty as shown in the Figure 4.3. This type of data results from tests or conditions where the application is not monitored continually. With a certain sample of components the only information is whether the component failed or did not fail between inspections. This is also called inspection data by some authors.

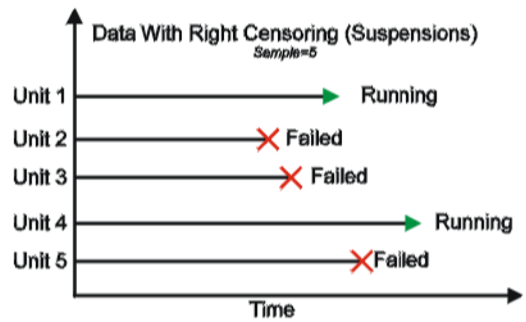

Figure 4.2*: Representation of Right Censored Data over Time

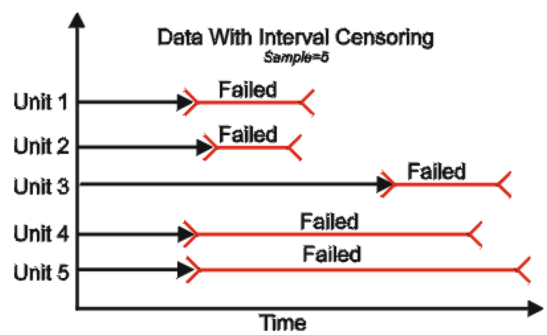

Figure 4.3*: Representation of Interval Censored Data over Time

3. Left Censored Data: Left censored observations occur when a failure is observed before a certain time, i.e. its first inspection time, as it can be seen in the Figure 4.4. If the starting time of interval censored mechanism is zero, then the interval censored data is same as left censored.

4. Truncated Data: In some cases, it may also arise that the lifetime less than some certain threshold may not be observed at all. This type of data is called as truncated. It should be noted that truncation is different from censoring. The general use of this case is observing data after the start of using system or component.

* http://www.weibull.com/LifeDataWeb/data_classification.htm

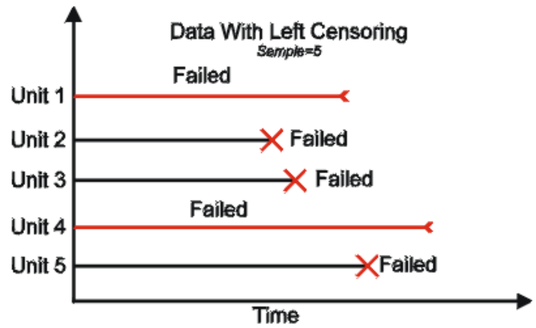

Figure 4.4*: Representation of Left Censored Data over Time

"It is important to distinguish between truncated and censored data. Censoring occurs when there is a bound on an observation (lower bound for observations censored on the right, upper bound for observations censored on the left and both upper and lower bounds for observations that are interval censored). However, truncation arises when even the existence of a potential observation would be unknown if its value were to lie in a certain range" (Meeker and Escobar (1998), p. 266).

4.4 Trunsored Model

Life time data analysis generally express two kinds of incomplete data of importance in estimation: censored data and truncated data which are defined before. Hirose (2005) introduces a third data type, referred to as trunsored data, which is a unification of censored and truncated data. A different view appears in the life data analysis with this unification since it does not only estimate and test the parameter of interests when both truncated and censored data exist but also estimate the ratio of truncated and censored data in new trunsored model. Thus, the trunsored model is proposed as a new incomplete data model to use in reliability and lifetime analysis.

Trunsored models define two types of samples. The first type has some observations that fail and others may not fail by the prescribed time T defined as *fragile samples*

* http://www.weibull.com/LifeDataWeb/data_classification.htm

whereas the second type has the observations that does not fail at all by the prescribed time T defined as *durable samples*. Hirose (2005) discusses that how to estimate the lifetime of a fragile population and corresponding confidence intervals using samples of size n from a mixture of the fragile and durable populations assuming that the ratio p of the fragile population to the durable population is unknown. Let assume that r failures are observed by time T. It is considered that the data as censored when $p=1$, while it is treated that the data is truncated if n is unknown. However, there are actual mixed cases, in which p is unknown and n is known. Thus, the ratio p and the lifetime of the fragile population are to be estimated as in Meeker (1987).

The primary part of the model construction procedure is to find out which population is dominant. If the durable population is dominant (i.e. p is very small), a truncated model approach solve the problem; if the fragile population is dominant (i.e. p is close to 1), a censored model approach would solve the problem as stated in Hirose (2005). There exists a third choice that p may be neither close to 0 nor 1. In this case, the trunsored model approach would propose the solution. If the fragile population is estimated to be dominant, existence of durable population should be searched. If any durable population does not exist, then the estimates of parameter of interest should be obtained by regarding the data as censored.

In following subsections, firstly the details of trunsored models and its relation with bootstrap resampling will be given and then an application will be presented.

4.4.1 Possible Applications of the Trunsored Model

Typical applications of the trunsored model are given as follows:

1. Decision making by manufacturers: Estimating the ratio of the fragile population to the total mixed population provides judgments that the whether manufacturers should recall their products for safety reasons via assessing the ratio at an early stage or not. A small ratio may indicate that the manufacturers can handle failed products on an individual case basis.

2. Assessment of the effectiveness of cancer treatment: When a newly developed cancer treatment is introduced, physicians can assess the effectiveness of the new treatment by comparing the survival rates between the new and old treatments. The survival rate can be estimated at an early stage when the trunsored model is used.

3. Severe infectious disease alert: By estimating the case fatality ratio of infectious diseases at an early stage, the people can be alerted to prevent the spread of a disease. The case fatality ratio can be estimated based on the number of infected persons, the number who have died, and the number of survivors. In this case, the (type I) mixed trunsored model is used.

4. Precautions against possible failures: If the items in a system have two phases, one in which the time of failure is observable and the other in which the appearance of a malcondition is observable (but not the time at which the condition changes from good to bad), and if the probability distributions of the time of failure and of the appearance of malconditions have some common relationship, e.g., the distributions have the same shape parameters, then the system manager can estimate the total number of malconditions at an early stage. In this case, the (type II) mixed trunsored model is used.

4.4.2 Construction of Trunsored Models

In many applications, life time data will be collected on a sample of units that are assumed as representation of population. In life time estimation problems, it is assumed that the underlying distribution is a single homogeneous population, and all samples drawn that population will eventually fail (or die). When the sample has incomplete data, the characteristics of the population are determined by regarding the data as censored or truncated. Thus, the type of the sample should be defined by the following inquiry: If the left endpoint, T_0, of the underlying distribution is very large, the failures will not be observed within the prescribed time T, which is extremely smaller then the left endpoint. Then it is assumed that the sample is taken from durable population. On the other hand, it is possible that some failures are observed within the time interval of length T. Then, it is assumed that the sample is drawn from the fragile population. If the sample size is not known, this kind of mixture problem is reduced to a truncated model problem. The lifetime of the fragile population can then be estimated using the conditional likelihood if a parametric model of the underlying probability distribution is assumed. But, since we are dealing with nonparametric models, this case is out of our scope.

After construction of the resampling and bootstrap basics, the problem that will be discussed here is to estimate the lifetime of a fragile population when r failures are observed within T from the mixture of the fragile and durable populations in which the sample size is n, assuming that the ratio, p_m, of the fragile population to the mixed populations is unknown.

In the literature, there are very limited application studies in which fragile and durable populations appear to be mixed. For example, Goldman (1984) discussed the proportion of patients cured by a particular treatment by using Monte Carlo; Meeker (1987) and Hirose (2000, 2002) applied the model to integrated circuit reliability. In these studies, p_m is unknown and n is known, p_m and the lifetime of the fragile population is to be found out.

The motivations for trunsored model construction are summarized as in the following:

1. The fragile and durable populations may be mixed,

2. The ratio of the fragile population to the mixed populations is unknown,

3. The sample size n is known.

If the durable population is dominant, the truncated model approach or if the fragile population is dominant, i.e., p_m is regarded as close to 1, a censored model approach might solve the problem. The critical point here is to determine which model of the censored model or truncated model should be used. In each case, the confidence intervals of the parameters are very different from each other. To decide the model, we need a hypothesis test, $H_0{:}p_m{=}p_0$. This hypothesis test is done via bootstrap confidence intervals.

The following notations are used in the model:

F, f	Cdf, pdf for the fragile population, respectively
G, g	Cdf, pdf for the durable population, respectively
H, h	Cdf, pdf of the linear combination of fragile and durable population, respectively
θ, ϕ	parameter in the [fragile, durable] population
ψ	parameter in the linear combination of fragile and durable populations
s	linear combination parameter for F and G, $-\infty < s < \infty$
p_0	prescribed real number, $0 \le p_0 \le 1$
p_m	the ratio of the fragile population to the mixed populations, $0 \le p_m \le 1$
$\hat{s},\hat{p}_m$	estimated value in [trunsored, mixture] model
r	number of failed observations
n	number of observations
t	Time
t_i	observed failure time, (i=1,...,r)
T	censoring time
T_0	endpoint such that $\inf_{t\ge 0}\{t : G(t) > 0\}$
H_0	null hypothesis that $p_m{=}p_0$
L_{ts},L_m,L_c,L_t	likelihood function for [trunsored, mixture, censored, truncated] model

Hirose (2005) defines following conditions and assumptions of the trunsored models:

1. The probability density functions (pdf) $f(t;\theta)$ and $g(t;\phi)$ are assumed as smooth.

2. The observations finish at the prescribed time T.

3. The failure times, $t_1, t_2, ..., t_r (\le T)$, are observed.

4. The sample size is known, n.

5. $G(T_0)=0, T \ll T_0$.

6. Type I, right censoring model is mainly considered here.

A cdf $H(t;\psi)$, which is a linear combination of $F(t;\theta)$ and $G(t;\phi)$ given by

$$H(t;\psi) = sF(t;\theta) + (1-s)G(t;\phi), \quad (t \ge 0, -\infty < s < \infty) \tag{4.1}$$

where s is a combination parameter. Thus corresponding pdf of H is given in the following form

$$h(t;\psi) = sf(t;\theta) + (1-s)g(t;\phi). \tag{4.2}$$

Then, the likelihood function for the combination model is given in the form as

$$L(\psi) = \{1 - H(T;\psi)\}^{n-r} \prod_{i=1}^{r} h(t_i;\psi). \tag{4.3}$$

If $g(t_i)=0$ and $g(T)=0$ because of the assumption (6), then $L(\psi) \to L_{ts}(\theta, s)$ where

$$L_{ts}(\theta, s) = \{1 - sF(T;\theta)\}^{n-r} \prod_{i=1}^{r} \{sf(t_i;\theta)\}, \quad (-\infty < s < \infty) \tag{4.4}$$

The likelihood function of mixture model is given in the form below by restricting the s as $0 \le s \le 1$,

$$L_m(\theta, p_m) = \{1 - p_m F(T;\theta)\}^{n-r} \prod_{i=1}^{r} \{p_m f(t_i;\theta)\}, \quad (0 \le p_m \le 1) \tag{4.5}$$

where the parameter s is changed to p_m for clarity.

As it is defined before, if $p_m=1$, then we have censored model instead of trunsored model. Thus, the corresponding likelihood function will be

$$L_c(\theta) = \{1 - F(T;\theta)\}^{n-r} \prod_{i=1}^{r} \{ f(t_i;\theta)\} \; . \tag{4.6}$$

Finally, the truncated data model is known as

$$L_t(\theta) = \prod_{i=1}^{r} \left\{ \frac{f(t_i;\theta)}{F(T;\theta)} \right\}. \tag{4.7}$$

The likelihood equations of trunsored model are

$$\frac{\partial \log L_{ts}}{\partial \theta} = (n-r)\frac{\partial \log\{1-sF(T)\}}{\partial \theta} + \sum_{i=1}^{r} \frac{\partial \log\{sf(t_i)\}}{\partial \theta} \doteq 0 \tag{4.8}$$

and

$$\frac{\partial \log L_{ts}}{\partial s} = (n-r)\frac{\partial \log\{1-sF(T)\}}{\partial s} + \sum_{i=1}^{r} \frac{\partial \log\{sf(t_i)\}}{\partial s} \doteq 0 \tag{4.9}$$

From the equation (4.9) we have

$$\hat{s}\hat{F}(T) = \frac{r}{n}. \tag{4.10}$$

If we substitute (4.10) into the (4.8), then we have

$$\frac{\partial \log L_{ts}}{\partial \theta} = -\frac{r}{F(T)}\frac{\partial F(T)}{\partial \theta} + \sum_{i=1}^{r} \frac{\partial \log f(t_i)}{\partial \theta} = 0, \tag{4.11}$$

which is the same as likelihood equation of truncated model

$$\frac{\partial \log L_t}{\partial \theta} = -r\frac{\partial F(T)}{\partial \theta} + \sum_{i=1}^{r} \frac{\partial \log f(t_i)}{\partial \theta} = 0 \; . \tag{4.12}$$

According to the model formulation, Hirose (2005) shows that $\hat{\theta}_m$ of mixture model is the same as $\hat{\theta}_{ts}$ of the trunsored model if $\hat{s} \le 1$, and $\hat{\theta}_m$ is the same as $\hat{\theta}_c$ of the censored model if $\hat{s} \ge 1$. Thus, during the solution procedure for the linear combination model, solution corresponding to the truncated model can always be obtained as long as it exists and the solution corresponding to the censored model can also be obtained by setting p=1 and because of this reason, this linear combination model is referred as the trunsored model (Hirose (2005)). On the other hand, Hirose also shows that if $0.6 \le \hat{p} \le 1$ then the model approaches to the censored model. If

$\hat{p}>1$, the model is defined as imaginal mixture model with a likelihood function given by

$$L_{im}(\theta,p)=\{1-pF(T;\theta)\}^{n-r}\prod_{i=1}^{r}\{pf(t_i;\theta)\}, \qquad (p\geq 0). \tag{4.13}$$

The primary objective in proposing the trunsored model here is to make it easier to test the hypothesis $H_0: p_m=p_0$ where p_m is the ratio of the fragile population to the total mixed population of fragile and durable populations. Here, it is assumed that the fragile samples will eventually fail whereas the durable samples are assumed never to fail. To estimate the parameters in an underlying probability distribution with (right) censored homogeneous observed data, the censored model is used when the total sample size n is known, and the truncated model is often used when n is unknown. When p_0 is close to 1, it is preferable to test the hypothesis, $H_0: p_m=p_0$, before adopting either the censored model or truncated model because the standard errors of the parameters in the truncated model are markedly larger than those in the censored model. The confidence intervals are used to decide whether we reject the null hypothesis or not. Hirose (2005) classifies the patterns of the confidence intervals of the estimates in the mixture model approximately into three categories:

1. Pattern A: The censored model confidence intervals,

2. Pattern B: The truncated model confidence intervals,

3. Pattern C: The combination of these two confidence intervals.

Pattern A should be used only if there are strong indications that the data are censored; even if $\hat{s}$ is close to 1, the confidence intervals in the censored model absolutely differ from those in the mixture model. Pattern B is used after we have a rejected result from the hypothesis test related with the ratio p_m. For both patterns A and B, the confidence intervals of the estimates may be constructed based on the observed Fisher information matrix or the likelihood ratio statistics. For pattern C, however, it is necessary to perform the bootstrap resampling methods (Efron, 1979). Another motivation here is to use bootstrap to have a small $\hat{s}$.

4.5 Problem Definition and Lifetime Experiment

In this study, we demonstrate experimentally how the life-time of new designs can be estimated with the minimum biases as well as how the ratio of the fragile population to the mixed fragile and durable populations can be estimated for both old and new products. Since confidentiality agreement is made with the firm, the name of the firm and the details of product will not be given.

The firm is faced with such problems as:

- The products do not correspond to European standards of quality. Therefore, the firm desire to make reliable lifetime analysis whether they can correspond the standards or not.
- Achieving long lifetime is the main goal in new designs. The capability of monitoring the impact of different design options of the lamps is the key point in these lifetime tests. Besides, it may be guaranteed that the reliability of the chosen optimizations. However, in order to monitor reliable results, lifetime tests should be repeated many times which is costly and time consuming.
- They also desire to decide the best design only with testing the prototypes. However, they experienced that the test results based on classical random sampling do not represent the result of the population of new designs.

They also decide to develop a reliability program and construct a 100% control and test mechanism to monitor the products' reliability which is also costly.

The experimentation of thermal endurance is made in a laboratory. In each experiment, seven lamps (say, one lot) are taken to apply the endurance test because of limited capacity of the testing machine. 10 lots were tested by the machine during the observation. It is assumed that each lot is experimented in same environmental and physical conditions. A simple experiment result form can be seen in Appendix A.

The lifetime of a lamp is to be found out by using accelerated tests from the following formula:

$$\text{LogL} = \text{LogL}_0 + S\left(\frac{1}{K} - \frac{1}{K_w}\right),$$

where

L	Target endurance time[*] (in 30, 60, 90, and 120 days),
L_0	3652 days (nearly 10 years),
K	The temperature[**] (theoretically assumed) of the coil in the experiment (in Kelvin),
K_w	Maximum observed temperature in which the lamp does not fail (in Kelvin),
S	Constant[***] that depends on design of lamp control mechanism and the type of coil insulator used (unitless).

➢ The experiment is stopped at 4th and 24th hours to control the temperature of the coil and observed up to thirty days passed. At the end of the experiment, the failed lamps are detected and the test results are analyzed to estimate the life time of lamps and, hence, the coils.

The censoring time, T, is set as 30 days. We suppose the lifetimes after T are not observed. Thus, we have right censored data set. The total number of observations is 70. Table 5.1 represents the experiment results of a lot. If any lamp fails before 30 days, then it is assumed that the lamp could not pass the test. Hence, all the lifetimes have passed the test and are censored in the table given below.

Descriptive statistics, histogram and the experiment results of the whole lifetime data can be seen in Appendix B.

[*] L=30 in the observed data set.
[**] K=232 C° in the observed data set.
[***] S=4500 in the observed data set.

Table 4.1: Experiment Results of a Lot

Lamp no	1	2	3	4	5	6	7
Starting current (A*)	0.98	0.97	0.96	0.97	0.97	0.98	0.96
Temperature of coil after 4 hour (C°)	189	190	192	198	198	198	197
Temperature of coil after 24 hour (C°)	185	186	186	185	184	184	185
Ending time of the experiment (day)	33	32	31	33	32	33	33

4.6 Bootstrapping with Censoring Mechanism and its Application

The analysis of data from experiments in the development phase and measurements during production plays an important role in manufacturing. Experiments are performed during the development phase to ensure the design fitness for mass production. During production, a large number of measurements in the production control the quality and reliability of the products and processes. As the number of measurements increases, the traditional data analysis approaches its limits, and alternative methods are needed. Thus, bootstrapping is a crucial solution for limited experimental conditions. In many industrial areas, these conditions force the analyzer for using bootstrapping with censoring mechanism.

The bootstrap method for uncensored data is extremely simple and given theoretically in the previous chapter. Bootstrap procedures are developed to support inference for the reliability function because resampling techniques provide a useful methodology for constructing nonparametric confidence intervals.

Let us define the right censored data as of the form $\{(x_1,d_1),(x_2,d_2),...,(x_n,d_n)\}$, where x_i is the ith observation, censored or not, and

$$d_i = \begin{cases} 1, & \text{if } x_i \text{ is uncensored} \\ 0, & \text{if } x_i \text{ is censored} \end{cases} \qquad (i=1,...n).$$

* A: Ampere

For convenience it is assumed $x_1<x_2<...<x_n$ in the calculations below to avoid notational difficulties and some minor technical problems arising from ties.

Bootstrap algorithm has defined for uncensored data set with three basic steps in the previous chapter. The algorithm takes following form to generate censored data (Efron (1981a)):

Step 1: Bootstrap sample, $\left\{(x_1^*,d_1^*),(x_2^*,d_2^*),...,(x_n^*,d_n^*)\right\}$, is drawn with independent sampling n times with replacement from $\hat{F}$, the distribution putting mass 1/n at each point (x_i,d_i).

Step 2: The bootstrap replication of the estimate $\hat{\theta}$, i.e. $\hat{\theta}^*=S(x^*)$, should be evaluated. Step 1 and 2 are repeated independently B times. Thus, we obtain

$\hat{\theta}_1^*,\hat{\theta}_2^*,...,\hat{\theta}_B^*$.

Step 3: The bootstrap estimate, $\hat{\theta}$, of the statistic, θ, is evaluated by using B bootstrap estimate.

This form of the bootstrap requires only that the observed pairs (x_i,d_i) are independently and identically distributed observations from a distribution F. Here the statistic to be considered is of the form

$\hat{\theta}=\theta(\hat{S})$,

where $\hat{S}(t)$ is the Kaplan Meier curve. The Kaplan Meier curve $\hat{S}(t)$ is nearly unbiased estimate of the true survival curve and is given by

$$\hat{S}(t)=\prod_{i=1}^{k_t}\left(\frac{n-i}{n-i+1}\right)^{d_i}. \tag{4.14}$$

Here k_t is the value of k such that $t\in\left[x_k,x_{k+1}\right)$; in other words, the largest observed value, censored or not, equal to or less than t. If there is no censoring, then all $d_i=1$, and $\hat{S}(t)=(n-k_t)/n$, the ordinary right sided cdf. Kaplan and Meier (1958) shows that

$\hat{S}(t)$ is the nonparametric maximum likelihood estimator (MLE) for $S(t)$. Hence, we will use Kaplan-Meier estimator in trunsored model.

The Kaplan-Meier estimating procedure is applied before bootstrapping because of the presence of censored cases. The model is based on estimating conditional probabilities at each time point when an event occurs and taking the product limit of those probabilities to estimate the survival rate at each point in time (Kaplan and Meier (1958)). The Figure 4.5 shows Kaplan Meier estimated survival curve for all 70 data. Of the 70 lifetimes, 23 were exactly observed; i.e. the lamp failed during the experiment. The remaining 47 observations were censored; i.e. the lamps were still working on the 30th day of experiment. Mean and standard deviation of lifetimes are estimated as 34 and 1 days, respectively. The proportions of terminating and surviving events are given as 0.3286 and 0.6714. The proportion of surviving events (i.e. fragile population) will be estimated in the trunsored model analysis.

4.7 Lifetime Estimation for the Trunsored Data Model

A cdf $H(t;\psi)$, which is a linear combination of $F(t;\theta)$ and $G(t;\varphi)$, correspond to the mixture of fragile and durable populations given by

$$H(t;\psi) = pF(t;\theta) + (1-p)G(t;\phi), \quad (t \geq 0,\ 0<p \leq 1), \tag{4.15}$$

where p is a combination parameter. Thus, corresponding pdf of H is given in the following form

$$h(t;\psi) = pf(t;\theta) + (1-p)g(t;\phi). \tag{4.16}$$

If the observation is finished by the prescribed time T, and failure times, $t_1, t_2, ..., t_r (\leq T)$, are observed, then the likelihood function for the mixture model in the form,

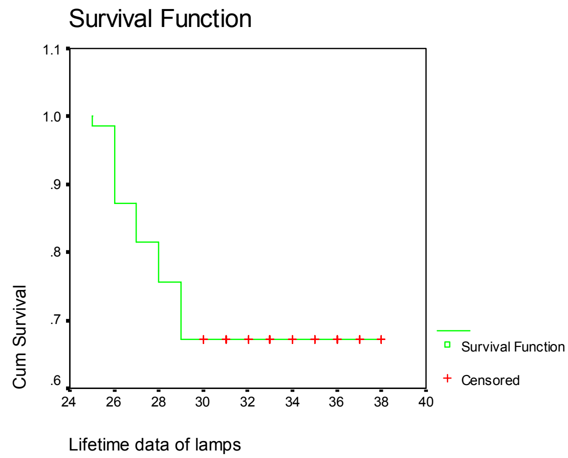

Figure 4.5: Kaplan Meier Estimated Survival Curve of the Lamp Lifetimes

$$L(\psi) = \{1 - H(T;\psi)\}^{n-r} \prod_{i=1}^{r} h(t_i;\psi), \qquad (r/n \le p \le 1) \tag{4.17}$$

where $n=70$ and $r=23$ in the original data set.

Let us assume that $g(t_i) \to 0$ and $G(T) \to 0$ because this population is durable. Then, $L(\psi) \to L_m(\theta, p)$ where

$$L_m(\theta, p) = \{1 - pF(T;\theta)\}^{n-r} \prod_{i=1}^{r} \{pf(t_i;\theta)\}, \qquad (r/n \le p \le 1) \tag{4.18}$$

If p=1, then we have the likelihood for the censored data model,

$$L_c = (\theta) = \{1 - F(T;\theta)\}^{n-r} \sum_{i=1}^{r} f(t_i;\theta). \tag{4.19}$$

On the other hand, the truncated data model can be expressed as

$$L_t(\theta) = \prod_{i=1}^{r} \left\{ \frac{f(t_i;\theta)}{F(T;\theta)} \right\}. \tag{4.20}$$

Hirose (2005) states that the MLE estimate $\hat{\theta}$ corresponding to the mixture model is the same as that corresponding to the truncated model if $p \le 1$, and it is the same as that to the censored model if $p=1$. The number of replication of bootstrapping samples is chanced from 1000 to 10000 and ten types of replication are performed. The histogram and quantile-quantile plots of each replication type can be seen in Appendix from Figure C.1 to Figure C.20. The solution of the likelihood equation in the mixture model can be obtained either by the solution of truncated model or by the solution in the censored model, and thus, this mixture model is called as trunsored model. By maximizing the MLE of the parameters, $(\hat{\theta}_{ts}, \hat{p}_{ts}) \cong (32, 0.68)$. The number of fragile population is estimated by 47.6 in the original data set, where the number of censored data is, in fact, 47. Hence, we have a bias nearly 0.6 for the original data set. However, the results for the generated data sets for all replications with ratio of fragile population to mixed population, mean lifetimes, biases, standard errors, empirical and BC_a percentiles of 2.5%, 5%, 95%, 97.5% are given in Table 4.2. The results can be summarized as in the following:

- The 95% confidence interval of ratio of fragile population to mixed population is (0.69838, 0.73142). Hence, the null hypothesis H_0: p_m=1 is rejected with 0.05 level of significance.
- Thermal endurance of coil approaches to a censored model since $0.69838 \le p \le 0.73142$.
- Bias of the mean lifetime is estimated as nearly zero. This fact again shows the power of bootstrapping with trunsored models (see Appendix, Figure C.21).
- The ratio of fragile population to mixed population is increases if the number of replication increases (see Appendix, Figure C.22).
- Bootstrap percentile confidence intervals give almost same results with the BC_a confidence intervals (see Appendix, Figure C.23).
- Bootstrap estimate of standard error gives better results than jackknife after bootstrap method (JAB) (see Appendix, Figure C.24). In fact, this result illustrates the power of bootstrapping.
- In five of the ten types of replications, estimated mean lifetime is the same as the sample mean (see Appendix, Figure C.25).
- Standard error of the mean lifetime changes from 0.44 to 0.46 (see Appendix, Figure C.26).

- Increasing the number of replication affects only the ratio of fragile population to mixed population because number of replication is chosen as large number as 1000 and greater as proposed in the literature.
- As it can be seen in histograms in Appendix from Figure C.1 to C.20, if the number of replications increases, the form of the model approaches to the censored model.
- Estimated mean lifetimes are changing between 31 and 32 days. Thus, the firm absolutely cannot achieve the European Standards which is 35 days.

Table 4.2: Bootstrap Resampling Results for Each Replication

B	$\hat{p}_m$	Number of Failures	BIAS	MEAN	SE	SE(JAB)	Empirical Percentiles				BCa Percentiles			
							2.5%	5%	95%	97.5%	2.5%	5%	95%	97.5%
1000	0.673	327	0.003243	31.67	0.4416	0.4579	30.8425	30.97071	32.4	32.57143	30.84286	30.96774	32.4	32.57143
2000	0.689	622	-0.01776	31.65	0.4547	0.4797	30.78536	30.92786	32.40071	32.57143	30.8	30.94448	32.44286	32.61429
3000	0.697	909	-0.003848	31.67	0.4441	0.4386	30.8	30.94286	32.38643	32.52857	30.8	30.92857	32.38571	32.52857
4000	0.711	1156	-0.002289	31.67	0.4444	0.4454	30.8	30.94286	32.4	32.54286	30.78571	30.92857	32.38571	32.52857
5000	0.708	1460	-0.01653	31.65	0.4467	0.4489	30.8	30.91429	32.38571	32.52857	30.82857	30.94286	32.41429	32.57143
6000	0.726	1644	-0.008331	31.66	0.4518	0.4646	30.78571	30.92857	32.4	32.5575	30.81429	30.95714	32.44286	32.61429
7000	0.735	1855	0.006265	31.68	0.4524	0.4575	30.78571	30.92857	32.41429	32.54321	30.72857	30.87143	32.37143	32.48571
8000	0.732	2144	0.002721	31.67	0.4536	0.4651	30.81429	30.94286	32.42857	32.55714	30.8	30.92857	32.41429	32.54286
9000	0.741	2331	-0.0007825	31.67	0.4544	0.4633	30.8	30.91429	32.42857	32.55714	30.8	30.92857	32.44286	32.57143
10000	0.737	2630	0.005281	31.68	0.4485	0.455	30.8	30.94286	32.41429	32.55714	30.75714	30.9	32.37143	32.51429

CHAPTER 5

CONCLUSION

In most applied statistical analysis random samples are at the heart of statistical inference and are in the concept of resampling methods. Experimentations, which are made for obtaining random sample(s), are performed under some restrictions such as time and cost. Although accurate and reliable statistical inferences depend on the sample size, time and cost limitations prevent to obtain a great number of data. The experimenters are generally faced with these restrictions inevitably and try to cope with large amount of biases. However, when the estimator of interest was complex and standard approximations were neither appropriate nor accurate, estimation of the standard error of the parameter estimator is the most encountered cases during statistical inference procedure. On the other hand, statistical inference is used in the vast part of the applied statistics to make strategic level decisions, and thus, the biases should be decreased with optimum time and cost perspectives. Resampling methods are proposed to compensate these problems and restrictions by achieving an enormous approximation to the population. Therefore, the samples that are generated by bootstrapping are used to draw conclusions. There exist some alternative methods for bootstrap resampling procedure. But, literature shows that this method gives the best results when one tries to estimate accurate bias and standard error.

In this study, we statistically analyze lifetime data with incomplete observations by bootstrap method and we use an evolutionary model. Specifically, we try to adapt trunsored data model to a nonparametric approach and construct the model with bootstrap resampling results. Thus, the bootstrapped samples are used as input for the trunsored model. This method was not a conventional reliability approach but a more sophisticated on utilizing incomplete data analysis models. Since the literature proposes limited applications for handling the questions about nonparametric approach in trunsored models, the main difficulty in this adaptation is to provide the best estimates of the parameters of interest that are used in the key points of the model. Therefore, various numbers of replications ($\geq$1000) are performed to generate lifetime data. The estimated parameters are used to find the ratio of truncated and censored models.

The main advantage of this study is having a chance to apply the bootstrap resampling method to a real lifetime data set with an original perspective in trunsored models. In addition to this, during the problem definition and data analysis procedure, different perspectives are obtained via getting better information about the types of failures experienced by parts and systems that aid design, research, and development efforts to minimize these failures, estimation of the failure ratio for both new and old design products, and getting estimations of the required redundancy to achieve the specified reliability. Since the trunsored models are proposed for decision making by manufacturers, the model becomes a potential to reduce cost of experimentations, to prevent failures especially for new design products and to improve product reliability. The framework presented in this study may also be applied for survival analysis such as assessment of the effectiveness of cancer treatment, severe infectious disease alert and precautions against possible failures. On the other hand these models can be used in the industrial engineering applications such as decision making by manufacturers as in our case. Estimating the ratio of the fragile population to the total mixed population provides judgments that the whether manufacturers should recall their products for safety reasons via assessing the ratio at an early stage or not. A small ratio may indicate that the manufacturers can handle failed products on an individual case basis. Because of its growing global market, the product’s manufacturing and performance have become the focus of much research. In today's technological world, consumers demand and expect reliable products. When products fail, most of the time the results are costly. Thus, it is critical to produce correct designed, processed and produced products. The main difference from the traditional methods is that the trunsored models with bootstrapping mechanism provide solutions without making unrealistic assumptions to produce reliable products in limited experimental conditions.

For future research, the proposed adaptation can be compared with the original form of the model by using different resampling and/or bootstrapping procedures. The adaptation can be realized very easily by updating the algorithms of bootstrapping with trunsored data set. In particular, the updated algorithms for trunsored data models with bootstrapping will be practical for measuring the effectiveness of treatment in survival analysis as well as for repairable products in lifetime analysis.

REFERENCES

1. **Ang, R. P.** (1998), Use of Jackknife Statistic to Evaluate Result Reliability, *Journal of General Psychology*, 125, pp. 218-228.

2. **Angus, J. E.** (1993), Asymptotic Theory for Bootstrapping the Extremes, *Commutations Statistics Theory Methods*, 22, pp. 15-30.

3. **Athreya, K. B.** (1987), Bootstrap Estimation of the Mean in the Infinite Variance Case, *Annals of Statistics*, 15, pp. 724-731.

4. **Beran, R. J.** (1982), Estimated Sampling Distributions: The Bootstrap and Competitors, *Annals of Statistics*, 10, pp. 212-225.

5. **Besterfield, H. D.** (2001), *Quality Control*, 6th edition, Prentice Hall.

6. **Bickel, P. J. and Freedman, D. A.** (1981), Some Asymptotic Theory for the Bootstrap, *Annals of Statistics*, 9, pp. 1196-1217.

7. **Bickel, P. J. and Krieger, A. M.** (1989), Confidence Bands for a Distribution Function Using the Bootstrap, *Journal of American Statistical Association*, 84, pp. 95-100.

8. **Cheng, R. C. H.** (2001), Analysis of Simulation Experiments by Bootstrap Resampling, *Proceedings of the 2001 Winter Simulation Conference*, pp. 179-186.

9. **Chernick M. R.** (1999), *Bootstrap Methods a Practitioner's Guide*, John Wiley & Sons, New York.

10. **Cressie, N. A. C.** (1993), *Statistics for Spatial Data*, Wiley, New York.

11. **Crowley, P. H.** (1992), Resampling Methods for Computation-Intensive Data Analysis in Ecology and Evolution, *Annual Review of Ecological Systems*, 23, pp. 405–447.

12. **Doty, L. A.** (1989), *Reliability for the Technology*, Industrial Press Inc, New York.

13. **Davison, A. C. and Hinkley, D. V.** (1997), *Bootstrap Methods and Their Application*, Cambridge University Press, New York.

14. **Davidson, R. and MacKinnon, J. G.** (2000), Improving the Reliability of Bootstrap Tests, *Queen's Institute for Economic Research Discussion Paper*, No. 995.

15. **Diaconis, P. and Efron, B.** (1983), Computer-Intensive Methods in Statistics, Scientific American, 5, pp. 116-130.

16. **DiCiccio, T. J. and Romano, J. P.** (1988), A Review of Bootstrap Confidence Intervals, *Journal of the Royal Statistical Society*, Series B, 50, pp. 338–370.

17. **DiCiccio, T. J., Martin, M. A. and Young, G. A.** (1992), Fast and Accurate Approximate Double Bootstrap Confidence Intervals, *Biometrica*, 79, 2, pp. 285-295.

18. **DiCiccio, T. J. and Efron, B.** (1996), Bootstrap Confidence Intervals, *Statistical Science*, 11, 3, pp. 189- 228.

19. **Edgington, E. S.** (1995), *Randomization Tests*, 3rd Ed., Marcel Dekker, New York.

20. **Efron, B.** (1979), Bootstrap Methods: Another Look at the Jackknife, *Annals of Statistics*, 7, pp. 1-26.

21. **Efron, B.** (1981a), Censored Data and the Bootstrap, *Journal of American Statistical Association*, 76, 374, pp. 312-319.

22. **Efron, B.** (1981b), Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods, *Biometrica*, 68, 3, pp. 589-599.

23. **Efron, B.** (1982), The Jackknife, The Bootstrap and Other Resampling Plans, *Society of Industrial and Applied Mathematics*, CBMS-NSF Monographs, 38.

24. **Efron, B.** (1983), Estimating the Error Rate of a Prediction Rule: Improvements on Cross Validation, *Journal of American Statistical Association*, 78, pp. 316-331.

25. **Efron, B.** (1987), Better Bootstrap Confidence Intervals, *Journal of the American Statistical Association*, 82, 397, pp. 171-185.

26. **Efron, B.** (1988), Three Examples of Computer Intensive Statistical Inference, *Sankhya A*, 50, pp. 338-362.

27. **Efron, B.** (1990), More Efficient Bootstrap Computations, *Journal of the American Statistical Association*, 85, 409, pp. 79-89.

28. **Efron, B.** (1992), Six Questions Raised by Bootstrap, *In Exploring the Limits of Bootstrap*, Wiley, New York, pp. 99-126.

29. **Efron, B.** (1993), Statistics in the 21st Century, *Statistics Computing*, 3, pp. 188-190.

30. **Efron, B.** (2000), The Bootstrap and Modern Statistics, Journal of the American *Statistical Association*, 95, 452, pp. 1293-1296.

31. **Efron, B. and Gong, G.** (1983), A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation, *The American Statistician*, 37, pp. 36-48.

32. **Efron, B and Tibshirani R. J.** (1986), Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy, *Statistical Science*, 1, pp. 54–77.

33. **Efron, B. and Tibshirani, R. J.** (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York.

34. **Fisher, R. A.** (1949), *The Design of Experiments*, Oliver & Boyd, Edinburgh.

35. **Friedl, H. and Stampfer, E.** (2002), Resampling Methods, *Encyclopedia of Environmetrics*, 3, pp. 1768-1770.

36. **Fortin, M. J., Jacquez, G. M. and Shipley, B.** (2002), Computer Intensive Methods, *Encyclopedia of Environmetrics*, 1, pp. 399-402.

37. **Gine, E. and Zinn, J.** (1989), Necessary Conditions of Bootstrap of the Mean, *Annals of Statistics*, 17, pp. 684-691.

38. **Goldman, A. I.** (1984), Survivorship Analysis When Cure is a Possibility: a Monte Carlo Study, *Statistics in Medicine*, 3, pp. 153–167.

39. **Good, P.** (2000), *Permutation Tests: a Practical Guide to Resampling Methods for Hypothesis Testing*, Springer- Verlag, New York.

40. **Good, P.** (2006), *Resampling Methods: A Practical Guide to Data Analysis*, Birkhauser, Boston.

41. **Hall, P.** (1986), On the Number of Bootstrap Simulations Required to Construct a Confidence Interval, *The Annals of Statistics*, 14, 4, pp. 1453-1462.

42. **Hall, P.** (1992), *The Bootstrap and Edgeworth Expansion*, Springer, New York.

43. **Hall, P., Hardle, W. and Simar, L.** (1993), On the Inconsistency of Bootstrap Distribution Estimators, *Computer Statistics Data Analysis*, 16, 11-18.

44. **Hahn G. J. and Meeker W. Q.** (1991), *Statistical Intervals: A Guide for Practitioners*, Wiley, New York.

45. **Helmers, R., Janssen, P. and Veraverbeke, N.** (1992), Bootstrapping U-Quantiles, *In Exploring the Limits of Bootstrap*, pp. 145-155.

46. **Hirose, H.** (2000), *Lifetime Estimation of Defective Products from the Imaginal Mixture of Defective and Non-Defective Products: the Trunsored Data Model*, Kyushu Institute of Technology, Tech. Rep. CSSE-00-10.

47. **Hirose, H.** (2002), *Trunsored Data Analysis with Applications to Field Data*, *In Int. Conf. Statistics and Related Fields* (June 5–9 2002*)*, Honolulu, Hawaii.

48. **Hirose, H.** (2005), The Trunsored Model and Its Application to Lifetime Analysis: Unified Censored and Truncated Models, *IEEE Transactions on Reliability*, 54, 1, pp. 11-21.

49. **Hirose, H.** (2006), The Mixed Trunsored Model with Applications to SARS, *Mathematics and Computers in Simulation*, doi:10.1016/j.matcom.2006.06.031.

50. **Kales, P.** (1998), *Reliability for Technology, Engineering and Management*, Prentice Hall, New Jersey.

51. **Kaplan, E. L. and Meier, P.** (1958), Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, 53, pp. 457-481.

52. **Kurtz, A. K.** (1948). A Research Test of Rorschach Test, *Personnel Psychology*, 1, pp. 41-53.

53. **Lehmann, E. L.** (1997), *Testing Statistical Hypothesis*, 3rd Ed., Springer-Verlag, New York.

54. **Lunneborg, C. E.** (2000), *Data Analysis by Resampling: Concepts and Applications*, Duxbury, New York.

55. **Mak, T. K.** (2003), Estimating Variances for All Sample Sizes by the Bootstrap, *Computational Statistics and Data Analysis*, 46, pp. 459-467.

56. **Manly, B. F. J.** (1997), *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd Edition, Chapman & Hall, London.

57. **Martin, M. A.** (2007), Bootstrap Hypothesis Testing for some Common Statistical Problems: A Critical Evaluation of Size and Power Properties, *Computational Statistics and Data Analysis*, doi: 10.1016/j.csda.2007.01.020.

58. **Meeker, W. Q. and Escobar, L. A.** (1998), *Statistical Methods for Reliability Data*, Wiley, New York.

59. **Meeker, W. Q.** (1987), Limited Failure Population Life Tests: Application to Integrated Circuit Reliability, *Technometrics*, 29, pp. 51-65.

60. **Miller, R. G.** (1974), The Jackknife - a Review, *Biometrika*, 61, pp. 1-15.

61. **Mooney, C. Z. and Duval, R. D.** (1993), *Bootstrapping: a Nonparametric Approach to Statistical Inference*, Sage Publications.

62. **Noreen, E.** (1989), *Computer-Intensive Methods for Testing Hypothesis*, Wiley, New York.

63. **Parr, W. C.** (1983), A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance, *Biometrika*, 70, pp. 719-722.

64. **Peterson, I.** (1991), Pick a Sample, *Science News*, 140, pp. 56-58.

65. **Quenouille, M.** (1949), Approximate Tests of Correlation in Time Series, *Journal of the Royal Statistical Society*, Soc. Series B, 11, pp. 18-84.

66. **Quenouille, M.** (1956), Notes on Bias in Estimation, *Biometrika*, 43, pp. 353-360.

67. **Simon, J. L. and Bruce, P.** (1991), Resampling: A Tool for Everyday Statistical Work, *Chance*, 4, 1, pp. 22-32.

68. **Stone, M.** (1974), Cross-Validatory Choice and Assessment of Statistical Predictions (With Discussion), *Journal of the Royal Statistical Society*, Series B 36, pp. 111-147.

69. **Tibshirani, R.** (1992), Some Applications of the Bootstrap in Complex Problems, *In Exploring the Limits of Bootstrap*, Wiley, New York, pp. 271-277.

70. **Tukey, J. W.** (1958), Bias and Confidence in not Quite Large Samples, *Annals of Mathematical Statistics*, 29, p. 614.

71. **Young, G. A. and Daniels, H. E.** (1990), Bootstrap Bias, *Biometrika*, 77, pp. 179-185.

72. **Zacks, S.** (1992), *Introduction to Reliability Analysis*, Springer-Verlag Inc., New York, USA.

APPENDIX A

SAMPLE FORM OF EXPERIMENT RESULT

sayfa 17 / 22 Rapor No. 0.16.01.02/2006-

IEC 61347-2-9			
Clause	Requirement – Test	Result - Remark	Verdict

	EK 4: sargıların ısıl dayanıklılığı, Madde 13'ün sonuçları		
	Tip referansı :	MH 0070	—
	Anma gerilimi (V)...:	230V	—
	Sargıların malzemeleri ...:	Emeya Bobin Teli	—
	Deney için çıkarılan bileşenler...:	--	—
	Lamba tipi ...:	70 W Metal Halide	—
	tw ...:	130	—
	S ...:	4500	—
	Amaçlanan deney süresi (gün)...:	30	—
	Teorik deney sıcaklığı (°C)...:	232	—

parametre	1	2	3	4	5	6	7
Lambayı başlatma ve çalıştırma	G	G	G	G	G	G	G
İlk başlangıç akımı (A)	0,98A	0,97A	0,96A	0,97A	0,97A	0,98A	0,96A
4 saat sonra sargı sıcaklığı (°C)	189	190	192	198	198	198	197
Etüv sıcaklığı (°C)	162	162	163	162	164	165	164
24 saat sonra sargı sıcaklığı (°C)	185	186	186	185	184	184	185
Deney bitiş süresi (gün)	33	32	31	33	32	33	33
Lambayı başlatma	G	G	G	G	G	G	G
Son başlangıç akımı (A)	G	G	G	G	G	G	G
Yalıtım direnci (MΩ)	G	G	G	G	G	G	G

APPENDIX B

DESCRIPTIVE STATISTICS AND EXPERIMENT RESULTS

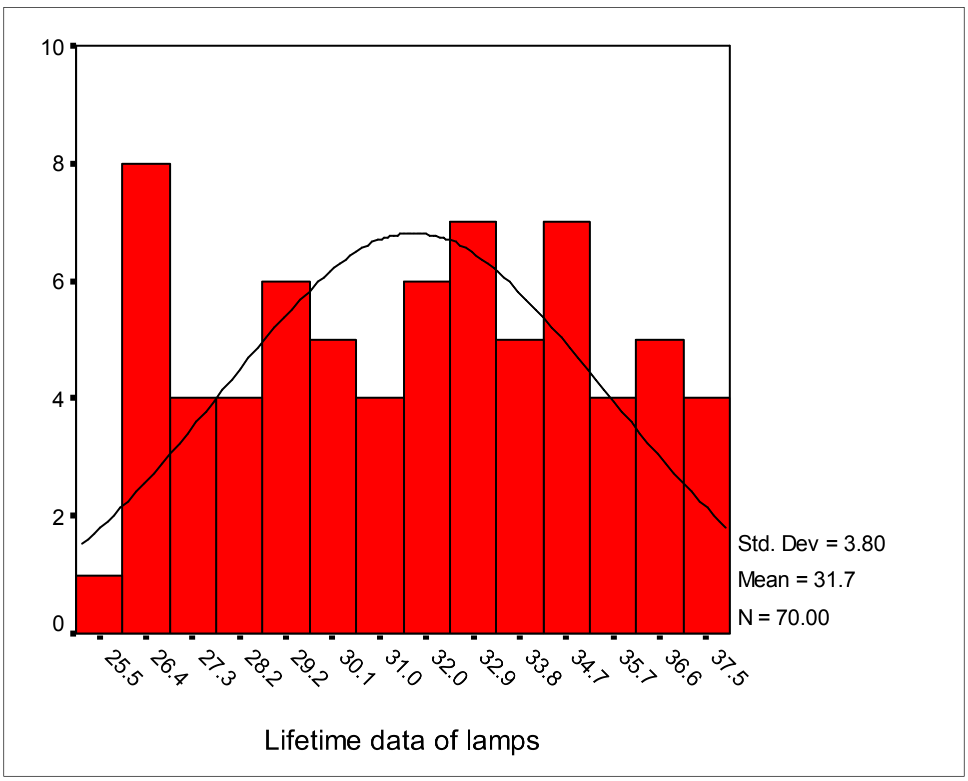

Figure B.1: Histogram of Lifetime Data

Table B.1: Descriptive Statistics of the Lifetime Data

Sample Size	70
Min	25
Max	38
Range	13
Mean	31.67
SE of Mean	0.45
Variance	14.46
Median	32
Skewness	-0.02
Kurtosis	-1.159

Table B.2: Experiment Results

Lamp no	1	2	3	4	5	6	7
Starting current (A)	0.98	0.97	0.96	0.97	0.97	0.98	0.96
Temperature of coil after 4 hour (C°)	189	190	192	198	198	198	197
Temperature of coil after 24 hour (C°)	185	186	186	185	184	184	185
Ending time of the experiment (day)	33	32	31	33	32	33	33
Lamp no	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Starting current (A)	0.96	0.97	0.96	0.97	0.96	0.97	0.96
Temperature of coil after 4 hour (C°)	195	198	195	198	197	197	193
Temperature of coil after 24 hour (C°)	186	185	184	186	185	185	184
Ending time of the experiment (day)	28	26	31	26	26	37	33
Lamp no	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Starting current (A)	0.97	0.96	0.98	0.98	0.96	0.96	0.96
Temperature of coil after 4 hour (C°)	196	190	197	198	191	195	195
Temperature of coil after 24 hour (C°)	186	184	184	186	185	184	186
Ending time of the experiment (day)	32	28	35	25	27	33	31
Lamp no	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Starting current (A)	0.95	0.98	0.98	0.97	0.96	0.98	0.98
Temperature of coil after 4 hour (C°)	189	194	197	197	192	198	197
Temperature of coil after 24 hour (C°)	184	186	185	185	184	186	186
Ending time of the experiment (day)	27	37	33	35	32	37	37
Lamp no	**29**	**30**	**31**	**32**	**33**	**34**	**35**
Starting current (A)	0.97	0.96	0.97	0.98	0.98	0.97	0.96
Temperature of coil after 4 hour (C°)	196	191	192	198	198	196	191
Temperature of coil after 24 hour (C°)	185	184	184	186	184	184	184
Ending time of the experiment (day)	25	26	28	35	38	35	25
Lamp no	**36**	**37**	**38**	**39**	**40**	**41**	**42**
Starting current (A)	0.97	0.97	0.97	0.96	0.97	0.98	0.97
Temperature of coil after 4 hour (C°)	198	197	195	189	195	198	197
Temperature of coil after 24 hour (C°)	186	186	185	184	185	184	185
Ending time of the experiment (day)	26	32	27	36	31	37	37

Table B.2: Experiment Results (Cont'd)

Lamp no	43	44	45	46	47	48	49
Starting current (A)	0.97	0.98	0.97	0.95	0.96	0.98	0.98
Temperature of coil after 4 hour (C°)	193	195	194	190	192	198	195
Temperature of coil after 24 hour (C°)	184	186	186	184	184	186	185
Ending time of the experiment (day)	26	33	26	27	34	31	29
Lamp no	**50**	**51**	**52**	**53**	**54**	**55**	**56**
Starting current (A)	0.98	0.97	0.98	0.98	0.98	0.97	0.96
Temperature of coil after 4 hour (C°)	197	197	194	198	196	196	193
Temperature of coil after 24 hour (C°)	186	186	185	185	184	184	184
Ending time of the experiment (day)	28	35	36	35	29	28	29
Lamp no	**57**	**58**	**59**	**60**	**61**	**62**	**63**
Starting current (A)	0.96	0.96	0.97	0.97	0.96	0.97	0.98
Temperature of coil after 4 hour (C°)	192	190	189	198	195	193	198
Temperature of coil after 24 hour (C°)	184	186	185	186	184	185	186
Ending time of the experiment (day)	26	26	36	27	31	30	28
Lamp no	**64**	**65**	**66**	**67**	**68**	**69**	**70**
Starting current (A)	0.96	0.95	0.98	0.97	0.98	0.97	0.97
Temperature of coil after 4 hour (C°)	190	190	196	196	198	195	195
Temperature of coil after 24 hour (C°)	184	186	185	186	184	186	185
Ending time of the experiment (day)	29	35	33	36	28	37	34

APPENDIX C

BOOTSTRAP AND TRUNSORED MODEL RESULTS

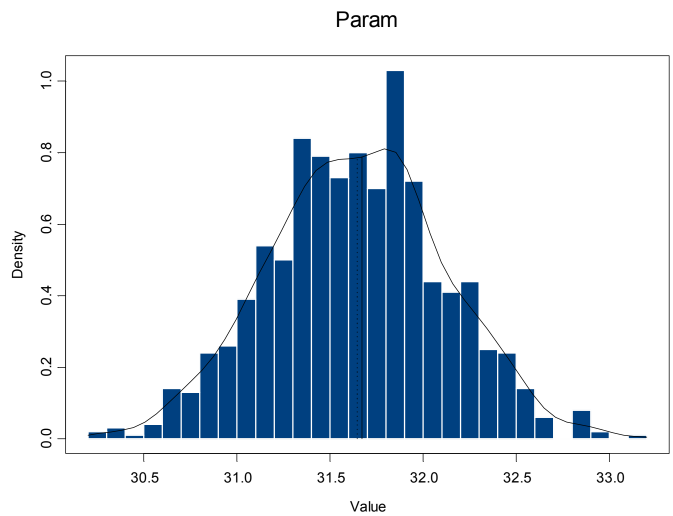

Figure C.1: Histogram of Bootstrap Samples for B=1000

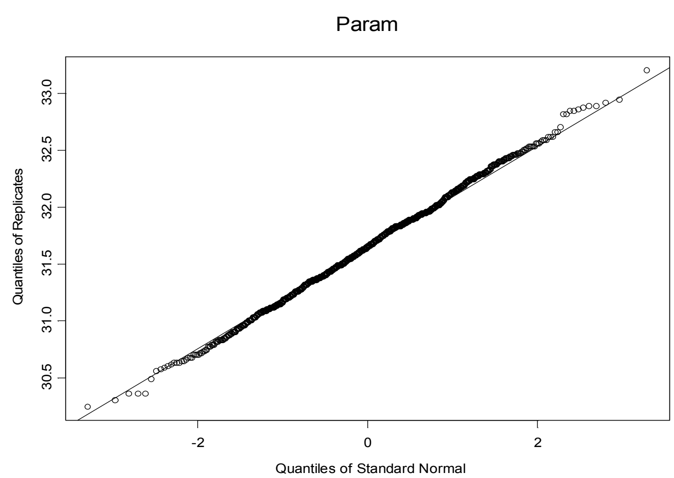

Figure C.2: Quantile-Quantile Plot of Bootstrap Samples for B=1000

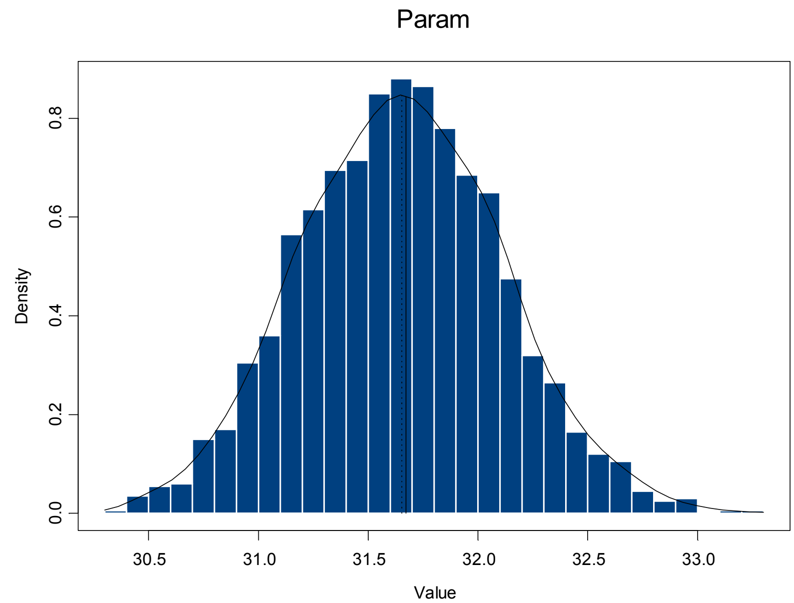

Figure C.3: Histogram of Bootstrap Samples for B=2000

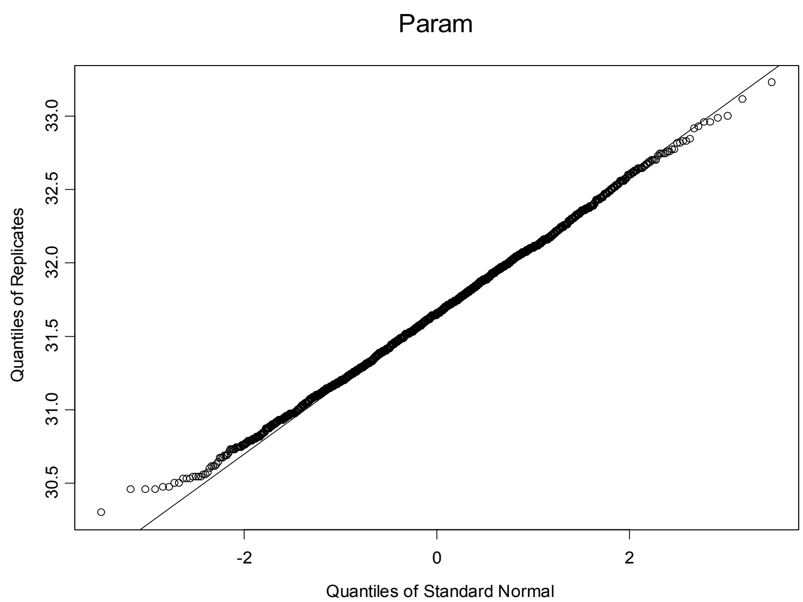

Figure C.4: Quantile-Quantile Plot of Bootstrap Samples for B=2000

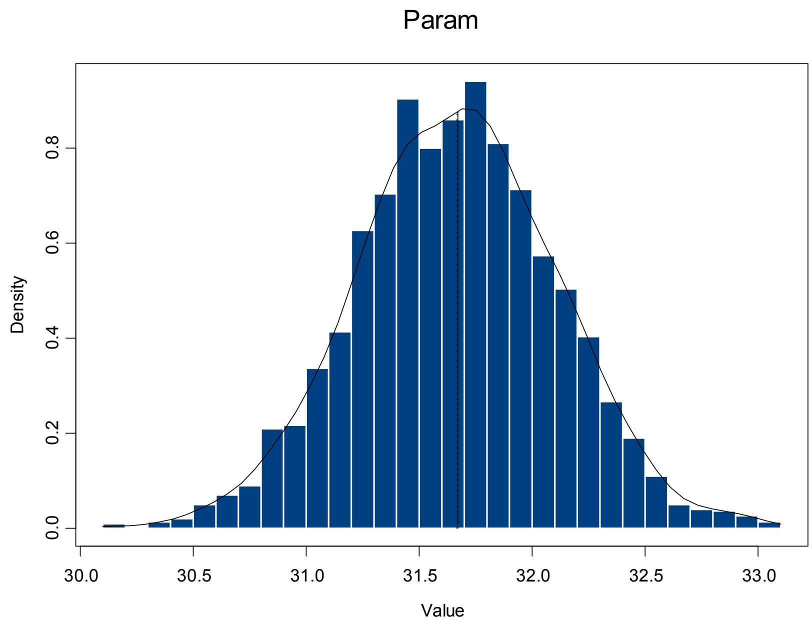

Figure C.5: Histogram of Bootstrap Samples for B=3000

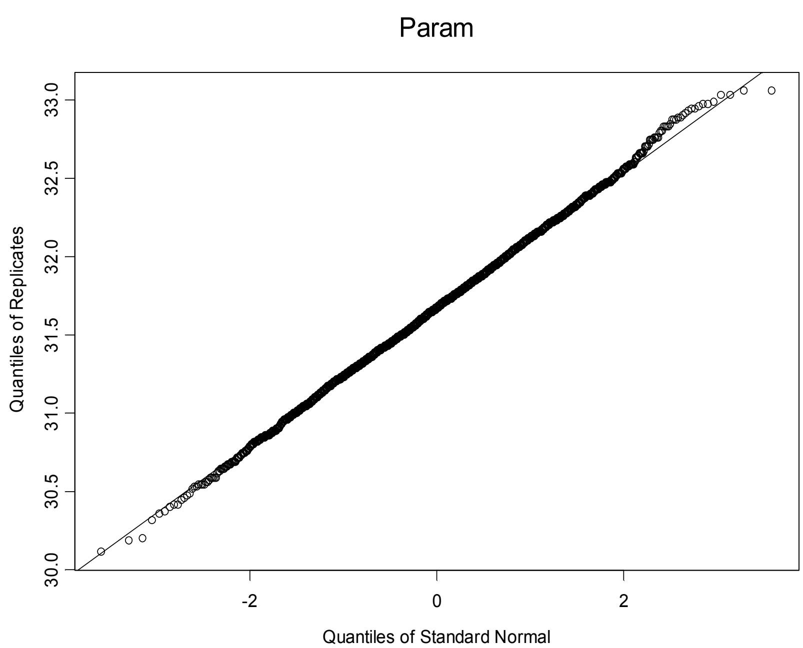

Figure C.6: Quantile-Quantile Plot of Bootstrap Samples for B=3000

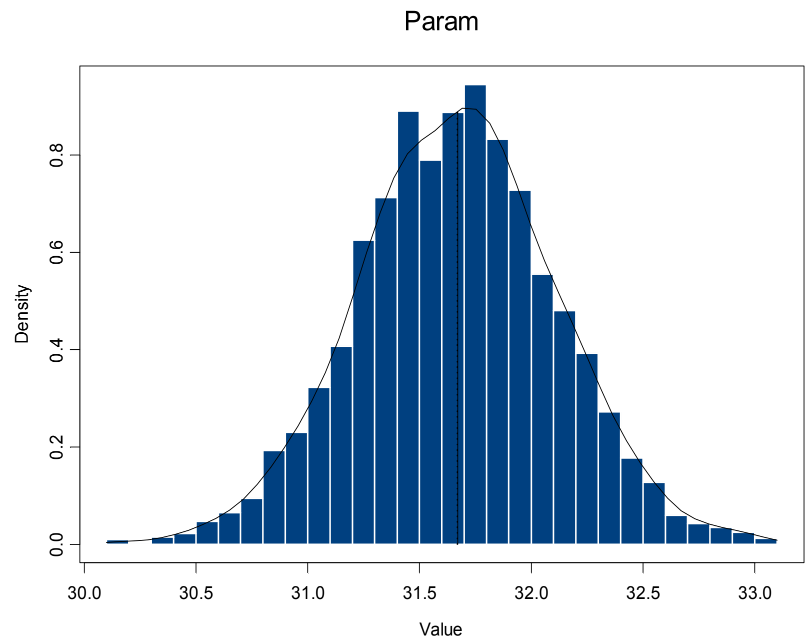

Figure C.7: Histogram of Bootstrap Samples for B=4000

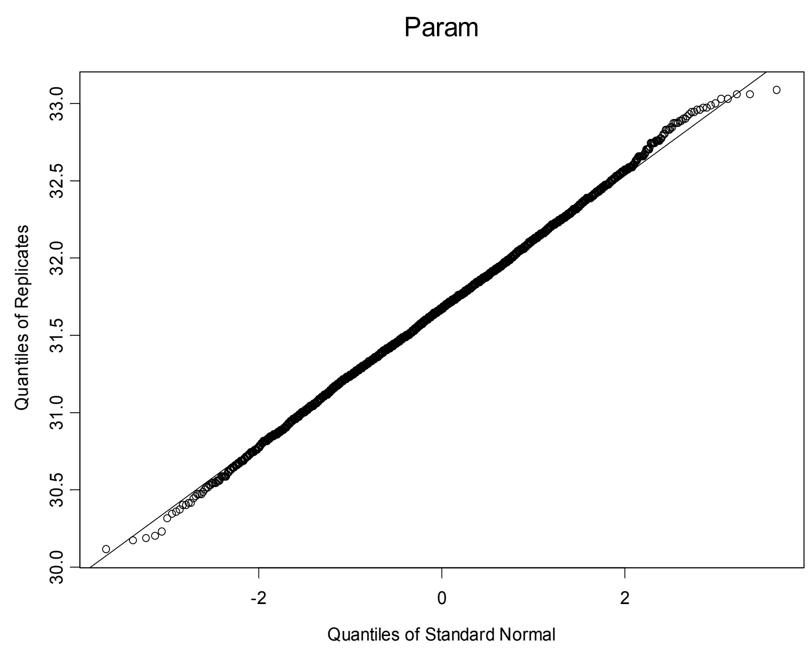

Figure C.8: Quantile-Quantile Plot of Bootstrap Samples for B=4000

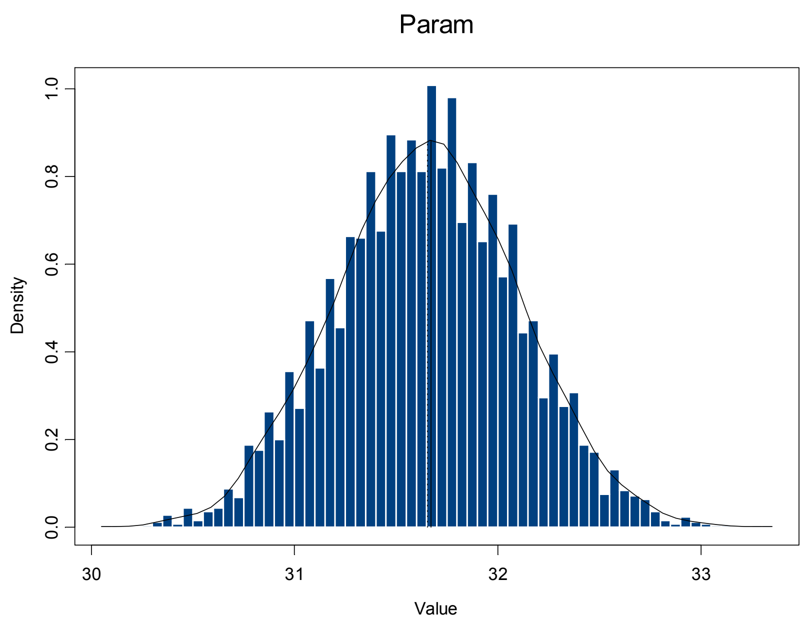

Figure C.9: Histogram of Bootstrap Samples for B=5000

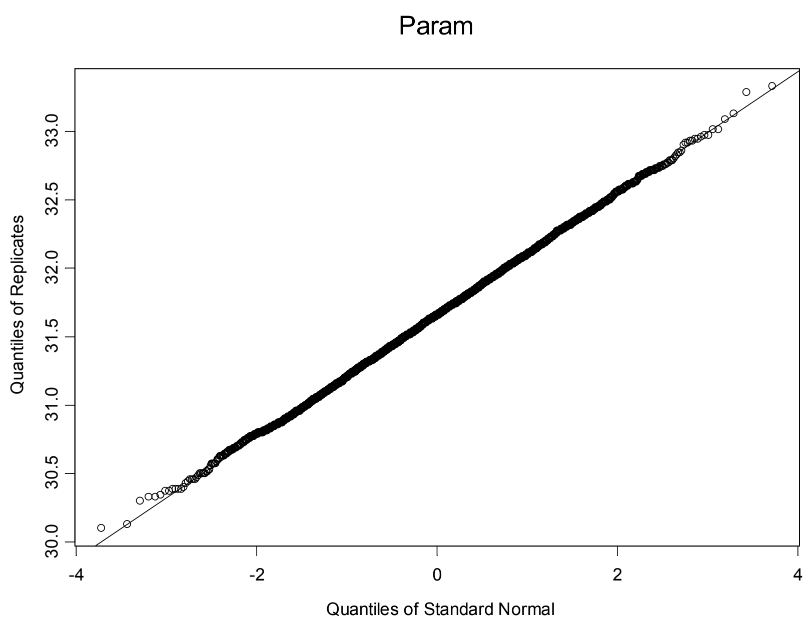

Figure C.10: Quantile-Quantile Plot of Bootstrap Samples for B=5000

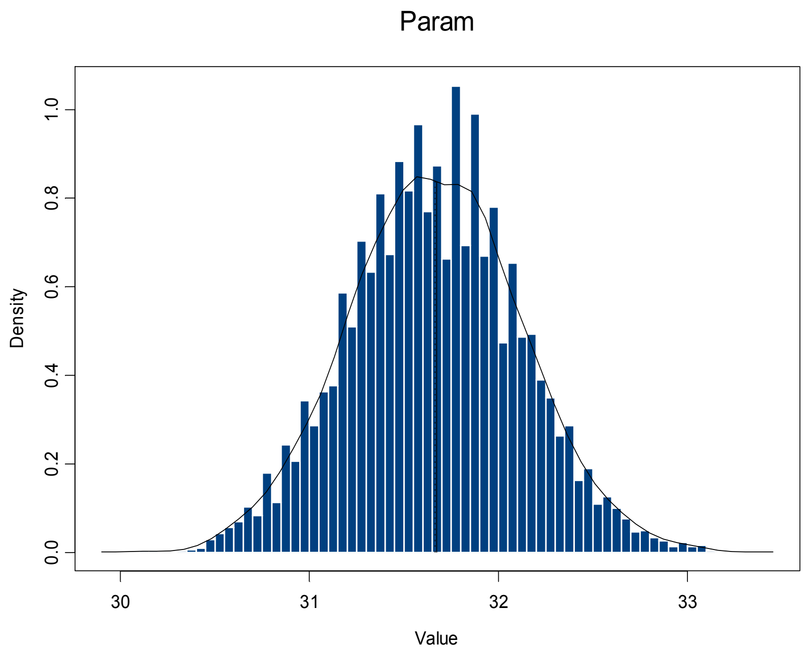

Figure C.11: Histogram of Bootstrap Samples for B=6000

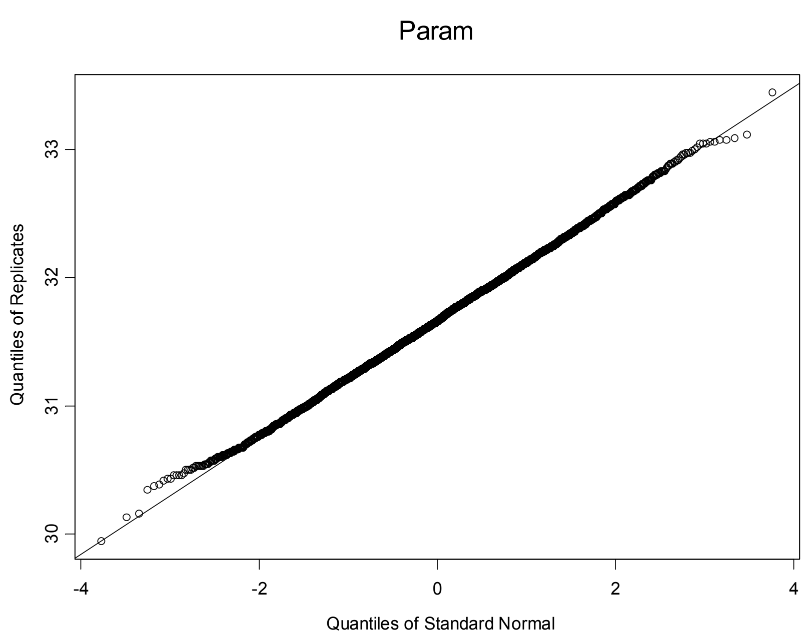

Figure C.12: Quantile-Quantile Plot of Bootstrap Samples for B=6000

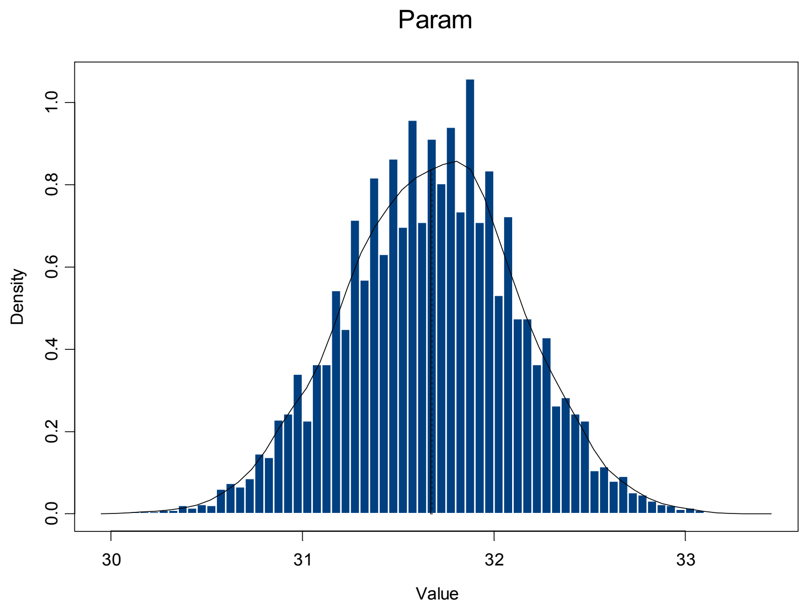

Figure C.13: Histogram of Bootstrap Samples for B=7000

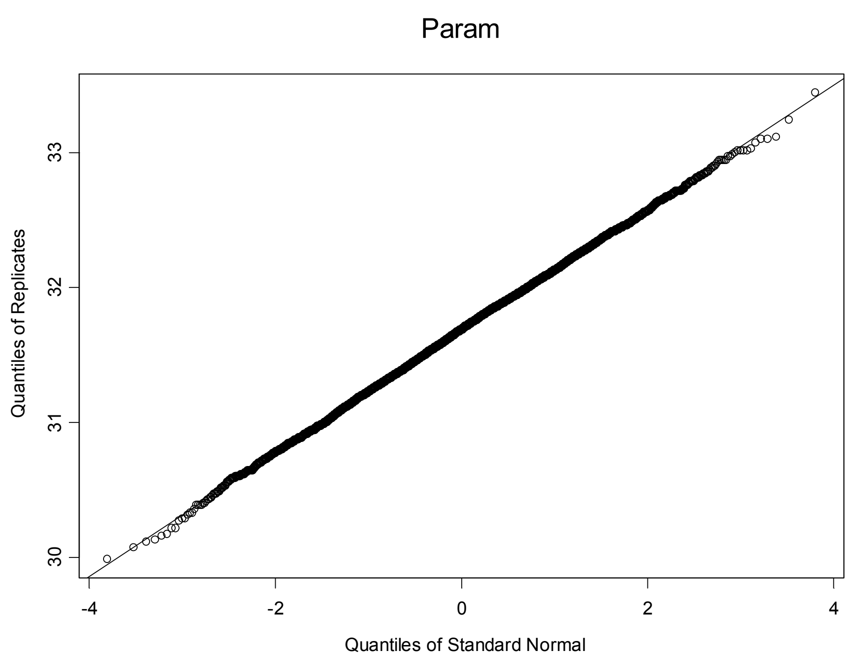

Figure C.14: Quantile-Quantile Plot of Bootstrap Samples for B=7000

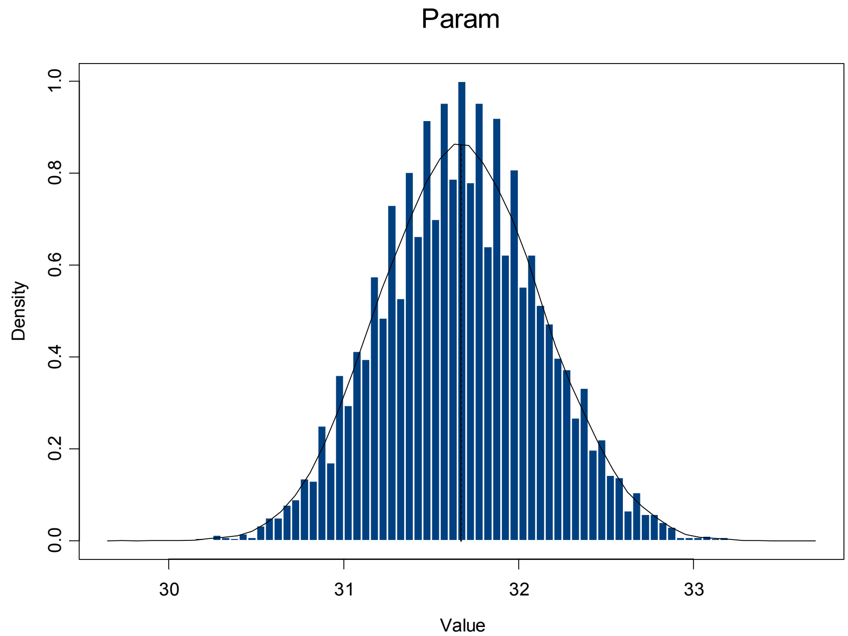

Figure C.15: Histogram of Bootstrap Samples for B=8000

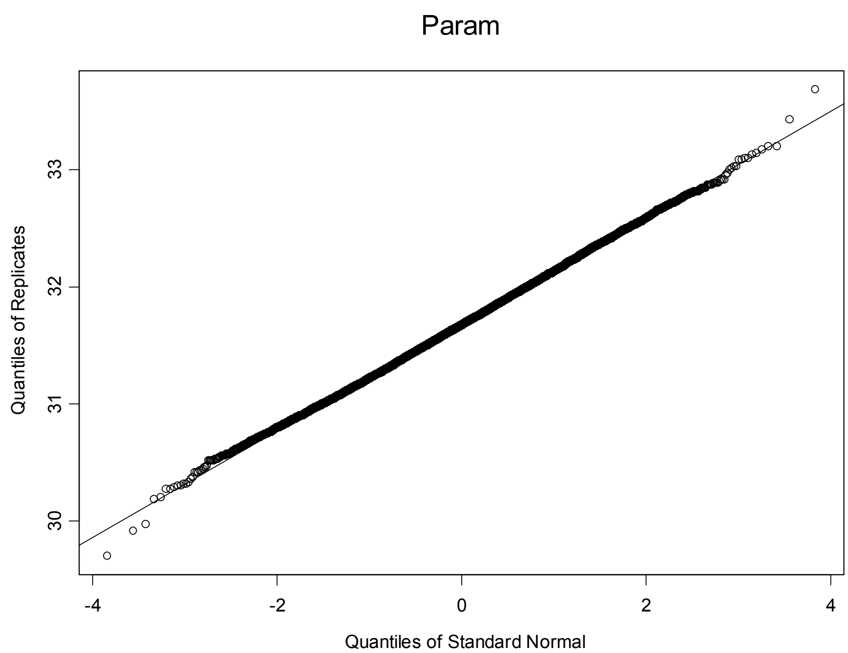

Figure C.16: Quantile-Quantile Plot of Bootstrap Samples for B=8000

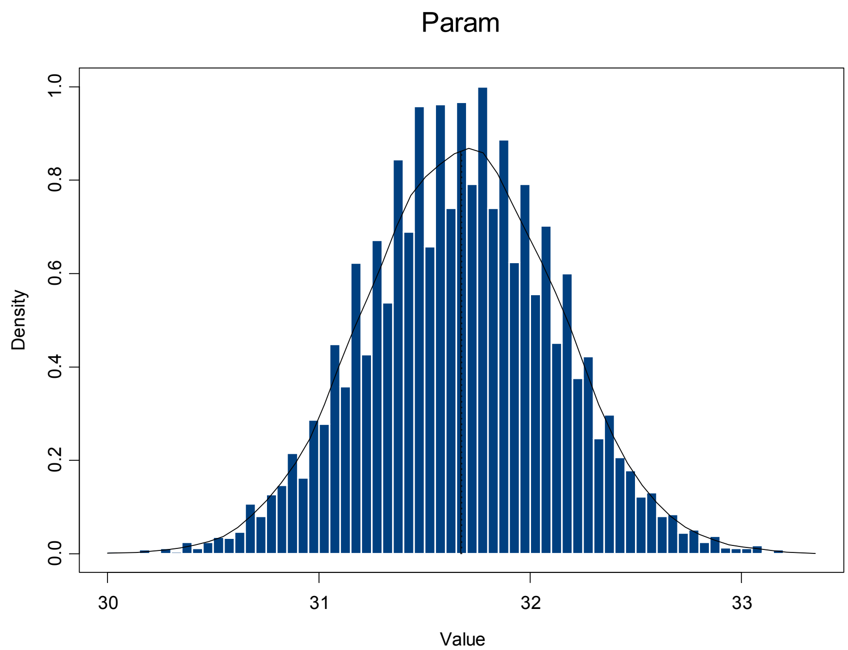

Figure C.17: Histogram of Bootstrap Samples for B=9000

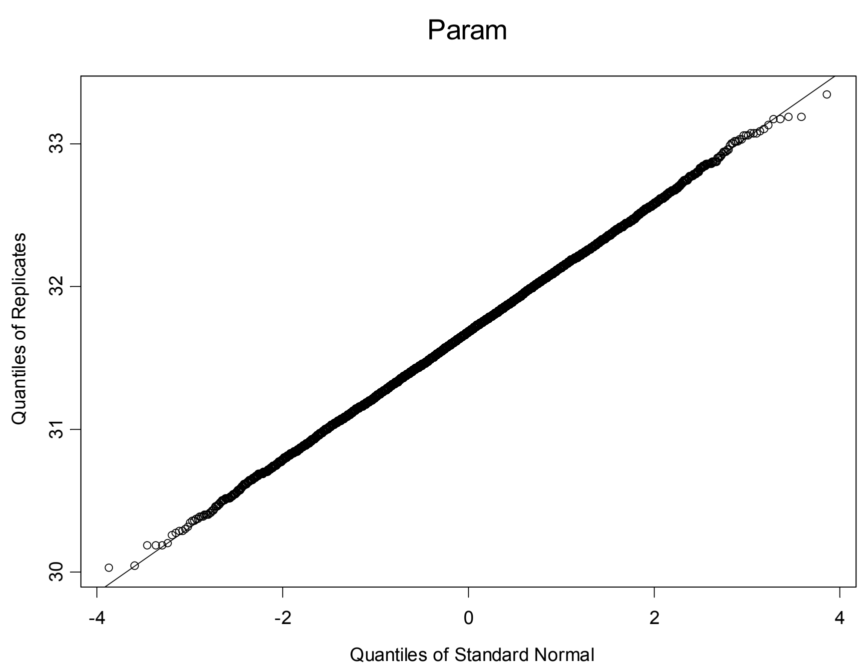

Figure C.18: Quantile-Quantile Plot of Bootstrap Samples for B=9000

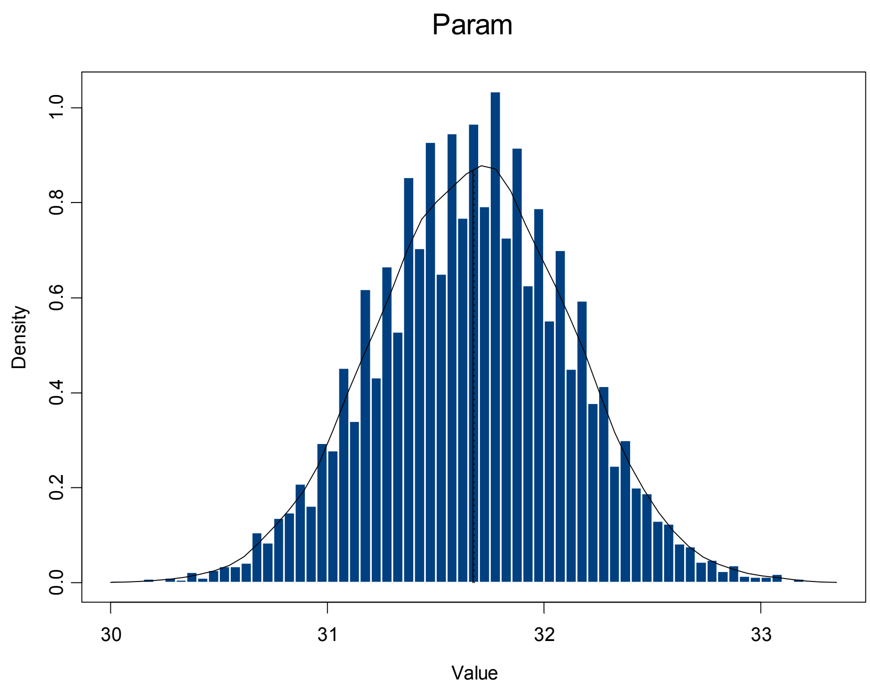

Figure C.19: Histogram of Bootstrap Samples for B=10000

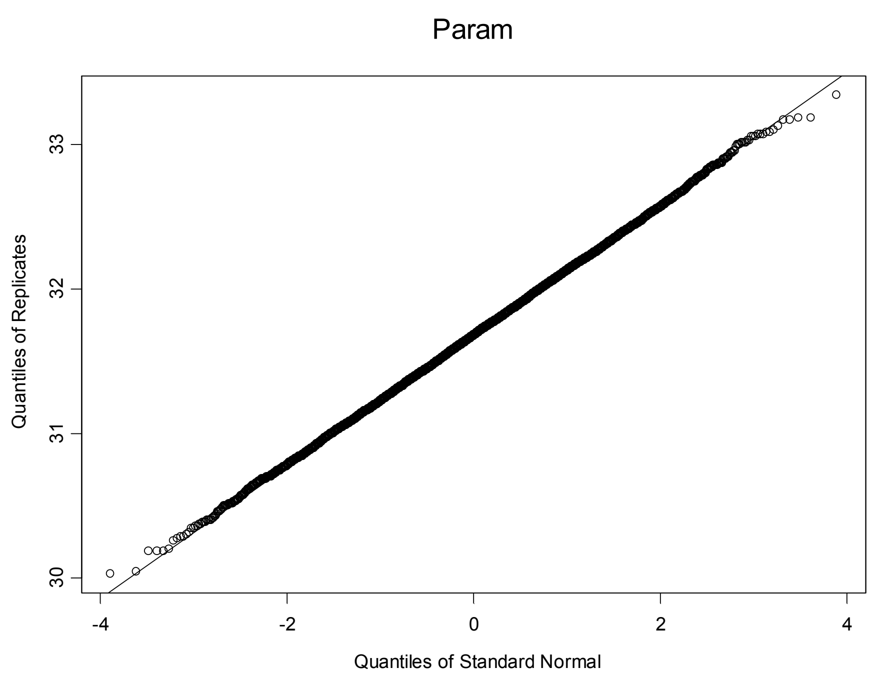

Figure C.20: Quantile-Quantile Plot of Bootstrap Samples for B=10000

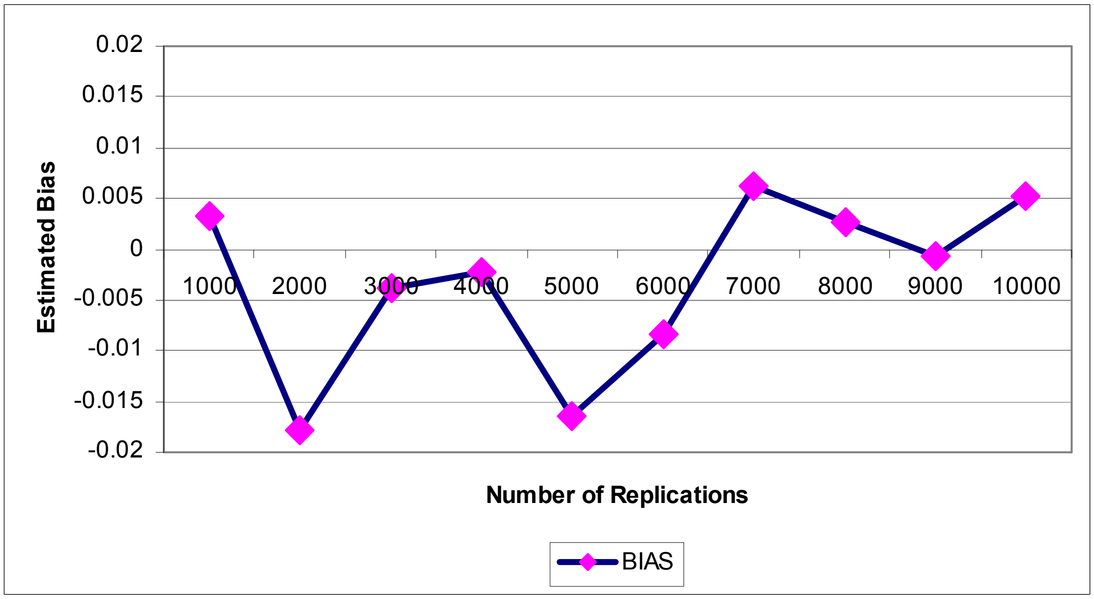

Figure C.21: Estimated Bias vs. Number of Replications

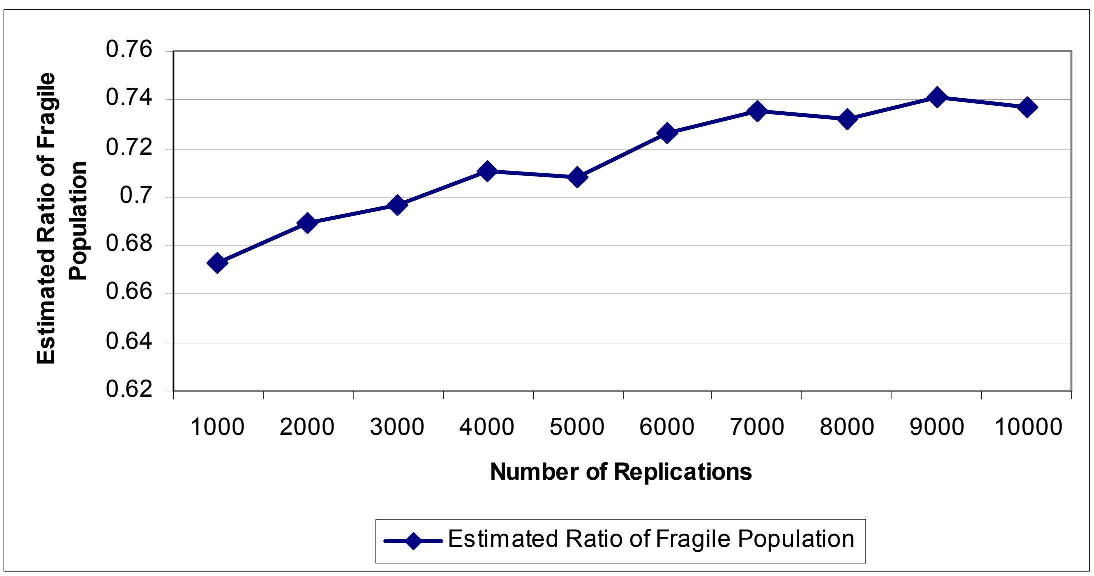

Figure C.22: Estimated Ratio of Fragile Population to Mixed Population vs. Number of Replications

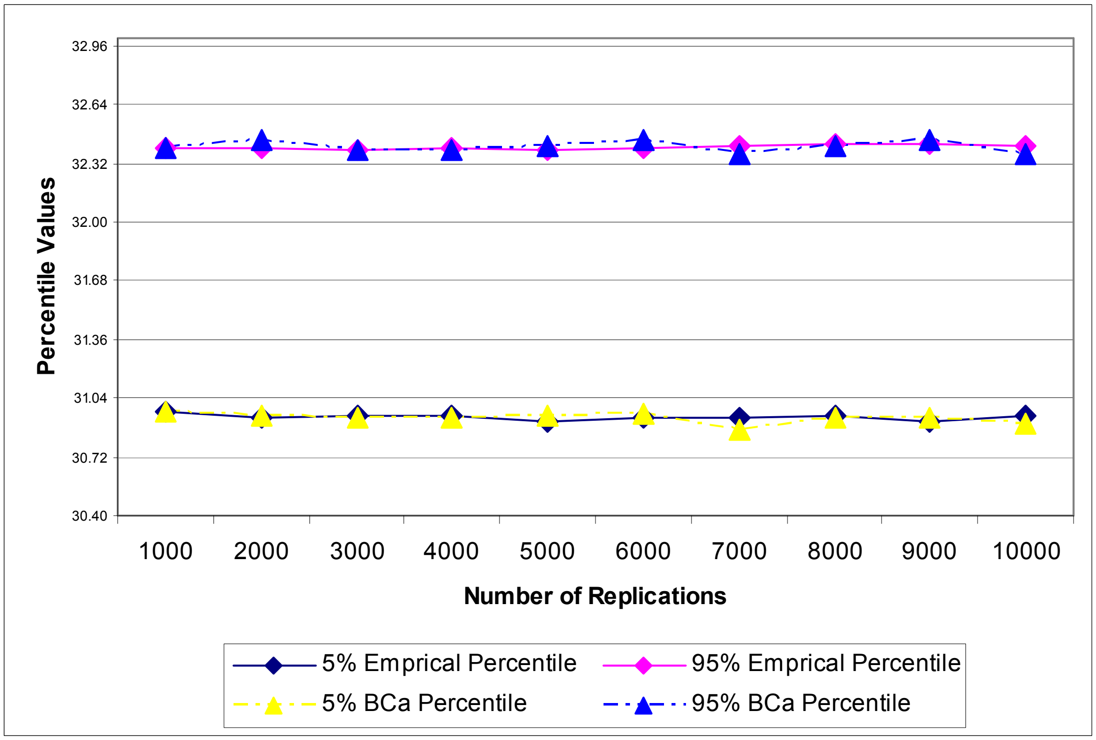

Figure C.23: Percentiles vs. Number of Replications

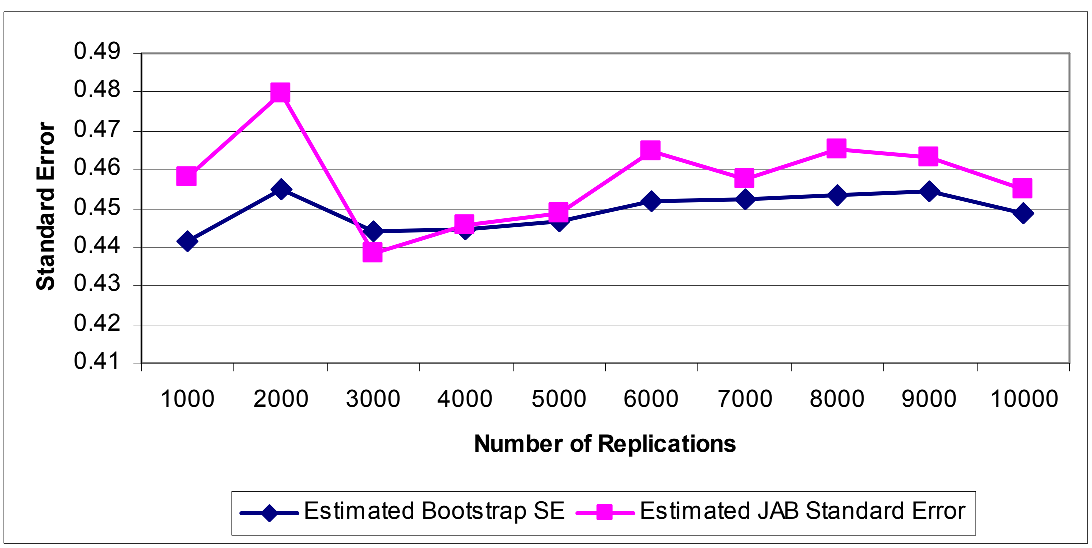

Figure C.24: Standard Errors of Mean Lifetime vs. Number of Replications

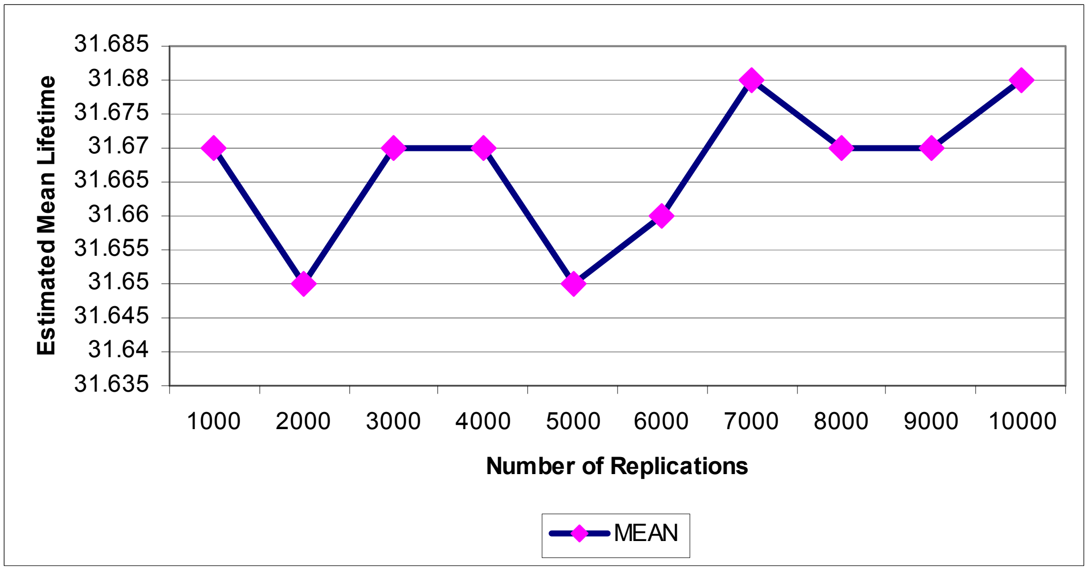

Figure C.25: Estimated Mean Lifetime vs. Number of Replications

Figure C.26: Estimated Standard Error vs. Number of Replications