

**AN APPROACH TO THE BOOTSTRAP ESTIMATE
OF MEAT PRODUCTION**

ဓာတ်ကြည့်တိုက်
စီးပွားရေးတက္ကသိုလ်
ရန်ကုန်

BY

WIN TUN

JUNE, 1993

**AN APPROACH TO THE BOOTSTRAP ESTIMATE
OF MEAT PRODUCTION**

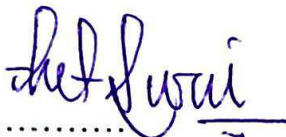
INSTITUTE OF ECONOMICS
YANGON
DEPARTMENT OF STATISTICS

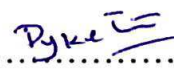
AN APPROACH TO THE BOOTSTRAP ESTIMATE
OF MEAT PRODUCTION

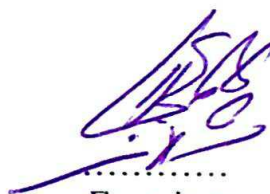
BY

MAUNG WIN TUN

This thesis is submitted to the Board of Examination in Statistics in partial fulfilment of the requirements for the degree of Master of Economics.


.....
Examiner


.....
External examiner


.....
Examiner

Thein Lwin, M.Sc (Rgn), Ph.D. (Prague)
Professor, Department of Statistics
Institute of Economics


.....
Chairman
Board of Examiners
1993

ပါမောက္ခချုပ်
စီးပွားရေးတက္ကသိုလ်၊ ရန်ကုန်





ACKNOWLEDGEMENT

I would like to express my gratitude to Professor U Ko Ko, Head of the Department of Statistics, Institute of Economics, for his permission to write this thesis.

I am deeply indebted to my supervisor, Daw May Thu Naing, Assistant Lecturer, Department of Statistics, for her guidance and support. When she left the department, Dr. Daw Than Toe, Lecturer, Department of Statistics, was asked to take the responsibility of supervising the thesis. I wish to express my indebtedness to Dr. Daw Than Toe for her valuable suggestions and comments.

Finally I would like to express my thanks to Daw Mya Win, Assistant Lecturer, Department of English, Institute of Economics, for editing the manuscript.

CONTENTS

	page
LIST OF TABLES	i
LIST OF FIGURES	ii
SUMMARY	iii
CHAPTER	
I. INTRODUCTION	1
II. JACKKNIFE METHOD OF BIAS REDUCTION	10
2.1 Introduction	10
2.2 Quenouille's Jackknife	11
2.2.1 Bias Reduction in Binomial Distribution	14
2.2.2 Bias Reduction in Uniform Distribution with One Parameter	16
2.2.3 Bias Reduction in Uniform Distribution with Two Parameters	22
2.3 The Generalized Jackknife	27
2.3.1 Bias Reduction in Uniform Distribution with One Parameter	31
2.3.2 Bias Reduction in Uniform Distribution with Two Parameters	33
2.4 Estimation of a Truncation Point	35
III. JACKKNIFE ESTIMATION OF BIAS AND STANDARD ERROR	44
3.1 Introduction	44
3.2 Estimated Bias of an Estimator	45
3.2.1 Estimated Bias of a Variance Estimator	45
3.3 Estimated Standard Error	48
3.3.1 Estimated Standard Error of the Trimmed Mean	48
3.3.2 Estimated Standard Error of the Sample Median	50
3.3.3 Estimated Standard Error of the Sample Variance	51

IV.	BOOTSTRAP METHOD OF ESTIMATION	53
4.1	Introduction	53
4.2	Bootstrap Method	54
4.3	Estimated Bias of an Estimator	56
4.4	Estimated Standard Error	58
4.4.1	Estimated Standard Error of the Sample Mean	59
4.4.2	Estimated Standard Error of the Sample Median	61
4.4.3	Estimated Standard Error for Censored Data	62
4.4.4	Estimated Standard Error in Half-sampling	64
4.5	Confidence Interval for the Median	69
4.5.1	Use of the Binomial Distribution	69
4.5.2	Use of the Percentile Method	70
V.	BOOTSTRAP ESTIMATION FOR SOME MODELS OF MEAT PRODUCTION IN MYANMAR	73
5.1	Introduction	73
5.2	Linear Regression Model	74
5.2.1	Jackknife Estimates	75
5.2.2	Bootstrap Estimates	77
5.3	Application of Bootstrapping to Some Models for Meat Production	79
5.3.1	Selected Models for Meat Production	80
5.3.2	Bootstrap Estimates and Their Distributions	81
	CONCLUSION	96
	APPENDIX	
	REFERENCES	

LIST OF TABLES

	page
<u>Pork Production Model</u>	
5.1 The bootstrap distribution of $\hat{\beta}_0$	82
5.2 The bootstrap distribution of $\hat{\beta}_1$	83
5.3 OLS estimates and bootstrap estimates	84
5.4 Relative errors of bootstrap estimates for different bootstrap sample sizes (N)	85
<u>Fowl Meat Production Model</u>	
5.5 The bootstrap distribution of $\hat{\beta}_0$	86
5.6 The bootstrap distribution of $\hat{\beta}_1$	87
5.7 OLS estimates and bootstrap estimates	88
5.8 Relative errors of bootstrap estimates for different bootstrap sample sizes (N)	89
<u>Beef Production Model</u>	
5.9 The bootstrap distribution of $\hat{\beta}_0$	90
5.10 The bootstrap distribution of $\hat{\beta}_1$	91
5.11 The bootstrap distribution of $\hat{\beta}_2$	92
5.12 OLS estimates and bootstrap estimates	93
5.13 Relative errors of bootstrap estimates for different bootstrap sample sizes (N)	94

LIST OF FIGURES

	page
2.1 The likelihood function of μ, σ	22
 <u>Pork Production Model</u>	
5.1 Histogram for the bootstrap distribution of $\hat{\beta}_0$	82
5.2 Histogram for the bootstrap distribution of $\hat{\beta}_1$	83
 <u>Fowl Meat Production Model</u>	
5.3 Histogram for the bootstrap distribution of $\hat{\beta}_0$	86
5.4 Histogram for the bootstrap distribution of $\hat{\beta}_1$	87
 <u>Beef Production Model</u>	
5.5 Histogram for the bootstrap distribution of $\hat{\beta}_0$	90
5.6 Histogram for the bootstrap distribution of $\hat{\beta}_1$	91
5.7 Histogram for the bootstrap distribution of $\hat{\beta}_2$	92

S U M M A R Y

Interest in the more technical aspects of statistical estimation introduces jackknifing and bootstrapping. Jackknifing is a widely applicable bias reduction tool which can also be used for obtaining a measure of bias and standard error of an estimator. Bootstrapping is another such method which can be seen as a simple and straightforward method for calculating not only the point estimate but also the interval estimate, especially in nonparametric estimation.

In this thesis, Quenouille's jackknife and Gray and Schucany's generalized jackknife are investigated as possible bias reduction methods. Quenouille's jackknife and bootstrapping are looked into as nonparametric estimation methods and applied in the linear regression model. The validity of bootstrapping in linear regression model is then analyzed using the real data. Computer-intensive simulations are carried out and bootstrap distributions, together with the bootstrap estimates are found for meat (pork, fowl meat and beef) production in Myanmar based on yearly data from 1967-68 to 1990-91 . The results show that large number of bootstrap samples furnishes estimates, which are very close to the ordinary least squares estimates. Thus it can be concluded that bootstrapping can be used as an estimation method which can supply the sampling distribution of an ordinary least squares estimator in the case of linear regression models, without the normality assumption.

C H A P T E R I

I N T R O D U C T I O N

In statistical estimation, the unknown characteristics of a population are usually estimated by taking a random sample from the population. The estimation is based on the assumptions that: (1) a random variable X can represent some characteristics of the elements in the population, (2) the probability distribution F is known except for an unknown parameter θ , (3) the value $(x_1, x_2, \dots, x_n)$ of a random sample $(X_1, X_2, \dots, X_n)$ from the probability distribution F can be observed. The value of some statistic, $\hat{\theta}(X_1, X_2, \dots, X_n)$, which is a function of the random sample $(X_1, X_2, \dots, X_n)$ is used to estimate the value of the unknown parameter θ or the value of some function of the unknown parameter. This is called point estimation.

Since a sample statistic is used to estimate θ , it may differ considerably from the true value of θ . It can not be expected to find any method of estimation which can be guaranteed to give an exact value of the parameter on every occasion and for every sample. Hence instead of giving the true value of the parameter by a point estimate, the true value of the parameter is said to be contained in some interval. This is the interval estimation. In this estimation,

two statistics, $\hat{\theta}_1(X_1, X_2, \dots, X_n)$ and $\hat{\theta}_2(X_1, X_2, \dots, X_n)$, where $\hat{\theta}_1(X_1, X_2, \dots, X_n) < \hat{\theta}_2(X_1, X_2, \dots, X_n)$ are calculated from the sample, so that $((\hat{\theta}_1(X_1, X_2, \dots, X_n), \hat{\theta}_2(X_1, X_2, \dots, X_n)))$ constitute an interval and the probability of this interval to contain the value of the parameter can be determined.

In using any estimator, it is naturally desired to know how successful it will be in the task of estimation. If it is possible, one wants to assess how close the estimated value is to the parameter being estimated. However, it is impossible to assess the quality of an estimate in that sense, because one cannot tell how close, his estimate from a particular sample, is to the true value of the parameter θ , which is unknown. If one is willing to settle for an evaluation of the performance of an estimator in terms of probability or equivalently, on how well it performs over a long sequence of trials in which it is used, then something can be said about the degree of success.

In the point estimation of a parameter, some properties of the estimator usually play a central role. One of the properties of a good estimator is unbiasedness. A sample statistic $\hat{\theta}(X_1, X_2, \dots, X_n)$ is said to be unbiased if its expectation is the same as the parameter θ . The expectation is to be taken over all possible samples of the same size n . If the above criterion is not met, the estimator is said to be biased. Biased estimators are usually avoided on the

estimation procedure or if possible, the bias has to be reduced if one wants to use that estimator.

Bias reduction was first introduced by Quenouille(1949). He proposed a procedure for reducing the bias of a serial correlation estimator. His procedure is based on splitting the sample into two half-samples. In his 1956 paper, he generalized this idea into splitting the sample into g groups of size h each and explored its general applicability. The bias reduction procedure suggested by Quenouille is as follows.

Let $(X_1, X_2, \dots, X_n)$ be a sample of independent and identically distributed (i.i.d.) random variables. Let $\hat{\theta}$ be an estimator of the parameter θ based on the sample of size n . Let $\hat{\theta}_{(i)}$ be the corresponding estimator based on the sample of size $(g-1)h$, where the i^{th} group of size h has been deleted. It is defined as,

$$J_i(\hat{\theta}) = g\hat{\theta} - (g-1)\hat{\theta}_{(i)}, \quad i=1,2,\dots,g. \quad (1.1)$$

The estimator with reduced bias, $J(\hat{\theta})$, is formed from $J_i(\hat{\theta})$ as,

$$\begin{aligned} J(\hat{\theta}) &= \frac{1}{g} \sum_{i=1}^g J_i(\hat{\theta}) \\ &= g\hat{\theta} - (g-1) \frac{1}{g} \sum_{i=1}^g \hat{\theta}_{(i)} \\ &= g\hat{\theta} - (g-1)\hat{\theta}_{(.)} \end{aligned} \quad (1.2)$$

$$\text{where } \hat{\theta}_{(.)} = \frac{1}{g} \sum_{i=1}^g \hat{\theta}_{(i)}.$$

In an unpublished work, Tukey subsequently called the g values (1.1) pseudo-values and created the name jackknifed

estimator for (1.2). (Miller(1974a)) Since a boy scout's jackknife is symbolic of a rough-and-ready instrument capable of being utilized in all contingencies and emergencies, Tukey has hoped that the jackknife method would be a rough-and-ready statistical tool. (Miller(1974a))

Quenouille's original justification for the procedure was based upon the assumption of the existence of a certain series expansion for the bias of an estimator. The series expansion for the bias of an estimator can be based on the following expression,

$$E(\hat{\theta}) = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots \quad (1.3)$$

This expansion does exist for a large class of estimators. The fact gives the method the general applicability which Tukey's appellation implies.

The estimator $J(\hat{\theta})$ in (1.2) has the property that it eliminates the order n^{-1} term from a basis of the form (1.3). If (1.3) holds, $E[J(\hat{\theta})] = \theta + O(n^{-2})$, that is, $J(\hat{\theta})$ is biased but the bias is reduced.

Although the general bias reduction procedure needs to be explained in terms of g and h , the bias reduction procedure presented in this thesis centers on $g=n$ and $h=1$. Much of the research on the jackknife has been devoted to this special case. It is the most appealing case because it eliminates any arbitrariness in the formation of the groups, and it is probably the best form of the jackknife to be used in any

problem. For large data bases, however, this may be computationally not feasible. (Miller(1974a))

Quenouille(1956) also gave a way to eliminate the order n^2 term from a basis of the form (1.3) by jackknifing, with weights n^2 , the jackknifed estimator. The second-order jackknife estimator is

$$J(\hat{\theta})^{(2)} = \frac{n^2 J(\hat{\theta}) - (n-1)^2 \sum \hat{\theta}_j/n}{n^2 - (n-1)^2} \quad (1.4)$$

where $\hat{\theta}_j$ is (1.2) applied to the sample of size $(n-1)$ with the j^{th} observation removed.

If (1.3) holds,

$$E[J(\hat{\theta})^{(2)}] = \theta + O(n^{-3}).$$

Schucany, Gray and Owen(1971) generalized the jackknife technique to handle more general forms of bias. Their technique is as follows.

Suppose there are two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ based on all or parts of the data for which the biases factorize in the following manner:

$$E(\hat{\theta}_i) = \theta + f_i(n) b(\theta), \quad i=1,2 \quad (1.5)$$

Then the estimator

$$G(\hat{\theta}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}} \quad (1.6)$$

is exactly unbiased. The usual jackknife (1.2) for $g=n$ fits into the form (1.6) with $\hat{\theta}_1 = \hat{\theta}$, $\hat{\theta}_2 = \hat{\theta}_{(.)}$, $f_1(n) = 1/n$ and $f_2(n) = 1/(n-1)$.

To eliminate k separate terms in the bias, each of which factorizes into distinct functions of n and θ , $k+1$ estimators are required whose expectations are of the form

$$E(\hat{\theta}_i) = \theta + \sum_{j=1}^k f_{ij}(n) b_j(\theta), \quad i=1,2,\dots,k+1. \quad (1.7)$$

The functions $f_{ij}(n)$ are assumed to be known. The generalization of (1.6) is

$$G^{(k)}(\hat{\theta}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \cdot & \cdot & \cdot & \cdot & \hat{\theta}_{k+1} \\ f_{11}(n) & & & & & f_{k+1,1}(n) \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ f_{1k}(n) & & \cdot & \cdot & \cdot & f_{k+1,k}(n) \end{vmatrix}}{\begin{vmatrix} 1 & \cdot & \cdot & \cdot & \cdot & 1 \\ f_{11}(n) & & \cdot & \cdot & \cdot & f_{k+1,1}(n) \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ f_{1k}(n) & & \cdot & \cdot & \cdot & f_{k+1,k}(n) \end{vmatrix}} \quad (1.8)$$

The added generality of the estimator allows the incorporation of additional information not used in the jackknife and hence increases the worth of the estimator as a bias-removal tool in specific cases. The authors suggested that the estimator (1.8) may be more trust worthy than the ordinary jackknife.

It was later called as generalized jackknife estimator by Gray and Schucany(1972) and was applied to different bias reduction situations.

As Tukey has hoped, the jackknife is a general method not only for reducing the bias in an estimator but also for obtaining a measure of the precision of the estimator by nonparametric estimation methods. Tukey(1958) suggested how the jackknife estimate of variance of an estimator could be obtained.

A method, which is an elaboration of the jackknife, was first introduced by Efron(1979). It is the so called bootstrap method. It is a simple and straightforward method for calculating approximated biases, variances, confidence intervals, and so forth, in almost any nonparametric estimation problem. In that sense, this method is more general than the jackknife method of Quenouille-Tukey. It has been shown that, (a) the bootstrap is closely related to the jackknife; (b) under reasonable conditions it gives asymptotically correct results; and (c) for some simple problems which can be analyzed completely, for example, ordinary linear regression, the bootstrap automatically produces standard solutions. Efron's bootstrap is as follows.

Let $(X_1, X_2, \dots, X_n)$ be a random sample of size n from a population with some unknown probability distribution F and let $\theta(X_1, X_2, \dots, X_n; F)$ be the specified random variable of interest, possibly depending upon the unknown distribution F .

Let $\hat{F}$ denote the empirical distribution function of $(X_1, X_2, \dots, X_n)$, this is, the distribution that puts mass $1/n$ at each of the points $X_1, X_2, \dots, X_n$. The bootstrap method is to approximate the distribution of $\theta(X_1, X_2, \dots, X_n; F)$ under F by that of $\theta(X_1^*, X_2^*, \dots, X_n^*; \hat{F})$ under $\hat{F}$ where $(X_1^*, X_2^*, \dots, X_n^*)$ denotes a random sample of size n from $\hat{F}$ and is called a bootstrap sample.

The bootstrap is said to be more widely applicable and perhaps has more sound theoretical basis than the popular Quenouille-Tukey jackknife. (Singh(1981)) It has been justified in two cases. The first justification was shown by Freedman. He has shown that, with i.i.d. errors, the bootstrap approximation to the distribution of the ordinary least square (OLS) estimator is valid. That is, as both the sample size and the number of bootstrap samples increase, the bootstrap distribution converges to the true distribution of the OLS estimator. The second justification was given by Freedman and Peters. They showed by monte-carlo simulations that the bootstrap method provides much more realistic standard errors of the generalized least squares estimators than the conventional asymptotic theory.

The bootstrap method has also been applied to censored sample. Efron(1981) found standard errors and confidence intervals for the parameters of an unknown distribution when the data is subject to right censoring. It can also be used

to obtain distribution-free confidence intervals for OLS estimates of the linear regression model. In the linear regression model, the errors are assumed to be i.i.d. and under the assumption that they are normally distributed, the exact sampling distribution of the OLS estimators of the parameters of the model can be determined. If the errors are not normally distributed, the central limit theorem has to be used to say that the distribution of the OLS estimator is asymptotically normal. For finite samples the normal distribution is only an approximation. This difficulty can be avoided by using the bootstrap method. The bootstrap method can supply confidence intervals for the OLS estimators when the errors are i.i.d. but not normal.

In this thesis, jackknifing and bootstrapping procedures for estimation of parameters will be investigated for various cases including some regression models. Using real data the validity of bootstrapping in some regression models without the assumption of normality of the errors will also be investigated. The bootstrapping is applied to find the sampling distribution of the OLS estimator. The bootstrap estimators so obtained will be compared with the OLS estimators.

C H A P T E R I I

JACKKNIFE METHOD OF BIAS REDUCTION

2.1 Introduction

In the point estimation of a parameter, one of the properties of a good estimator is unbiasedness. Even if an unbiased estimator is not available, one is willing to use the biased estimator after reducing its bias.

The problem of reducing the bias which may be present in an estimator of an unknown parameter has received considerable attention since 1949. The procedure was first introduced by Quenouille(1949) and later was termed the jackknife by Tukey(1958).

A more general procedure was introduced by Schucany, Gray and Owen(1971). Their method is related to the jackknife and in fact includes it as a special case. It was called the generalized jackknife by Gray and Schucany(1972).

In this chapter the bias reduction properties of the jackknife and the generalized jackknife are investigated. Their bias reduction performances in estimating parameters of a set of distributions are assessed.

2.2 The Quenouille's Jackknife

Suppose that the data set consists of an independent and identically distributed (i.i.d.) sample of size n from an unknown probability distribution F ,

$$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F.$$

Having observed values $X_1=x_1, X_2=x_2, \dots, X_n=x_n$, let $\hat{F}$ be the empirical probability distribution function of the data, putting probability mass $1/n$ on each x_i . That is,

$$\begin{aligned} \hat{F}(x) &= \frac{1}{n} (\text{number of } X_i \text{ less than or equal to } x) \\ &= \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \end{aligned} \quad (2.1)$$

where $I_{(-\infty, x]}(X_i)$ is an indicator function and

$$\begin{aligned} I_{(-\infty, x]}(X_i) &= 0 \quad \text{for } X_i > x \\ &= 1 \quad \text{otherwise.} \end{aligned}$$

Sometimes it is called as sample or empirical cumulative distribution function. It is a statistic since it is a function of the sample, for fixed x .

It also is a function of the order statistics $X_{(i)}$, $i=1, 2, \dots, n$ since $\hat{F}(x)$ can be rewritten as

$$\hat{F}(x) = \frac{1}{n} (\text{number of } X_{(i)} \text{ less than or equal to } x).$$

It has the same distribution function as that of the sample mean of a Bernoulli distribution since, if the parent population F is known, it can be shown that

$$\text{Prob}\left(\hat{F}(x) = \frac{k}{n}\right) = \binom{n}{k} [F(x)]^k [1-F(x)]^{n-k}, k=0, 1, \dots, n.$$

From Equation(2.1), it can be seen that $\hat{F}$ is a step-function with jumps of magnitude $1/n$ at each of the observed values.

If $\hat{F}$ is supposed to approximate $F(x)$, $\sup_x |\hat{F} - F|$ is called Kolmogorov-Smirnov distance between $\hat{F}$ and F . Kolmogorov-Smirnov distance can be used to verify whether the theoretical model is satisfactory and agrees with the observed values.

In most cases the parent population distribution F is unknown and $\hat{F}$ is utilized instead of F . $\hat{F}$ is also used in the jackknife method. Let

$$\hat{\theta} = \theta(\hat{F}) = \hat{\theta}(x_1, x_2, \dots, x_n)$$

be an estimate of parameter $\theta = \theta(F)$.

Quenouille's method is based on sequentially deleting point x_i , and recomputing $\hat{\theta}$. Removing point x_i from the data set gives a different empirical probability distribution,

$$\hat{F}_{(i)} : \text{mass } \frac{1}{n-1} \text{ at } x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n,$$

and a corresponding recomputed value of the statistic,

$$\hat{\theta}_{(i)} = \theta(\hat{F}_{(i)}) = \hat{\theta}(x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

Then the i^{th} pseudo-value of the jackknife, denoted by $J_i(\hat{\theta})$, is

$$J_i(\hat{\theta}) = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}, \quad i=1, 2, \dots, n,$$

and the bias-corrected jackknife estimate of θ , denoted by $J(\hat{\theta})$, is the average of $J_i(\hat{\theta})$.

$$\begin{aligned}
J(\hat{\theta}) &= \frac{1}{n} \sum_{i=1}^n J_i(\hat{\theta}) \\
&= n\hat{\theta} - (n-1) \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \\
&= n\hat{\theta} - (n-1) \hat{\theta}_{(.)}
\end{aligned} \tag{2.2}$$

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

The motivation which initially led to $J(\hat{\theta})$ is as follows.

Suppose

$$E(\hat{\theta}) = \theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots, \tag{2.3}$$

where the a_i may be functions of θ but not of n . Then

$$E(\hat{\theta}_{(i)}) = \theta + \frac{a_1}{(n-1)} + \frac{a_2}{(n-1)^2} + \frac{a_3}{(n-1)^3} + \dots,$$

and hence

$$\begin{aligned}
E[J(\hat{\theta})] &= E[n\hat{\theta} - (n-1)\hat{\theta}_{(.)}] \\
&= n E(\hat{\theta}) - (n-1) E(\hat{\theta}_{(i)}) \\
&= n \left(\theta + \frac{a_1}{n} + \frac{a_2}{n^2} + \frac{a_3}{n^3} + \dots \right) \\
&\quad - (n-1) \left(\theta + \frac{a_1}{(n-1)} + \frac{a_2}{(n-1)^2} + \frac{a_3}{(n-1)^3} + \dots \right) \\
&= \theta - \frac{a_2}{n(n-1)} - \frac{(2n-1)a_3}{n^2(n-1)^2} - \dots
\end{aligned} \tag{2.4}$$

Thus when $a_1 \neq 0$, the bias in $J(\hat{\theta})$ is of order n^{-2} while that of $\hat{\theta}$ is of order n^{-1} . Moreover if $a_2 = a_3 = \dots = 0$, $J(\hat{\theta})$ is unbiased.

This property will be illustrated in the following subsections.

2.2.1 Bias Reduction in Binomial Distribution

Let $(X_1, X_2, \dots, X_n)$ be an i.i.d. random sample from the point binomial distribution with parameter p . The random variable X is defined by the number of successes. Then X has the density function defined by

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} ; x=0, 1, \dots, n.$$

Consider

$$\hat{\theta} = \left(\frac{X}{n} \right)^2$$

as an estimator of p^2 because $\frac{X}{n}$ is unbiased for p . Then the expected value of $\hat{\theta}$ is

$$\begin{aligned} E[\hat{\theta}] &= V\left(\frac{X}{n} \right) + \left(E\left(\frac{X}{n} \right) \right)^2 \\ &= \frac{np(1-p)}{n^2} + \frac{(np)^2}{n^2} \\ &= \frac{p(1-p)}{n} + p^2 . \end{aligned}$$

Therefore the bias of $\hat{\theta}$ is

$$\text{bias}(\hat{\theta}) = \frac{p(1-p)}{n}.$$

The value of the estimator $\hat{\theta}_{(i)}$ will be one of two quantities, namely, $\left(\frac{x-1}{n-1} \right)^2$ or $\left(\frac{x}{n-1} \right)^2$, depending upon whether the deleted trial is a success or a failure, respectively. Hence $\hat{\theta}_{(.)}$, which is the average of the n possible values of $\hat{\theta}_{(i)}$ is

$$\begin{aligned} \hat{\theta}_{(.)} &= \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \\ &= \frac{1}{n} \left(x \left(\frac{x-1}{n-1} \right)^2 + (n-x) \left(\frac{x}{n-1} \right)^2 \right) , \end{aligned}$$

since there are X successes to be deleted and $(n-X)$ failures.

This yields

$$\begin{aligned}\hat{\theta}_{(.)} &= \frac{1}{n(n-1)^2} (X^3 - 2X^2 + X + nX^2 - X^3) \\ &= \frac{(n-2)X^2 + X}{n(n-1)^2},\end{aligned}$$

and thus

$$\begin{aligned}J(\hat{\theta}) &= n\hat{\theta} - (n-1)\hat{\theta}_{(.)} \\ &= n\left(\frac{X}{n}\right)^2 - (n-1)\frac{(n-2)X^2 + X}{n(n-1)^2} \\ &= \frac{X^2}{n} - \frac{(n-2)X^2 + X}{n(n-1)} \\ &= \frac{X^2 - X}{n(n-1)} \\ &= \frac{X(X-1)}{n(n-1)}.\end{aligned}$$

The expected value of the jackknife estimator is

$$\begin{aligned}E[J(\hat{\theta})] &= \frac{1}{n(n-1)} E[X(X-1)] \\ &= \frac{1}{n(n-1)} n(n-1)p^2 \\ &= p^2.\end{aligned}$$

Hence this estimator is unbiased for p^2 .

2.2.2 Bias Reduction in Uniform Distribution with One Parameter

In this section, bias reduction for the maximum-likelihood estimator of the parameter θ of a uniform distribution is considered.

Let X be uniformly distributed on the interval $(0, \theta)$. That is,

$$f(x) = \begin{cases} \frac{1}{\theta} & , 0 < x < \theta \\ 0 & , \text{otherwise.} \end{cases} \quad (2.5)$$

The likelihood function is

$$\begin{aligned} L(\theta; x_1, \dots, x_n) &= \prod_{i=1}^n f(x_i; \theta) \\ &= \frac{1}{\theta^n} . \end{aligned}$$

Let $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ be the order statistics. Then it can be seen that the parameter θ lies in the interval $(x_{(n)}, \infty)$. Since the value of θ which maximizes the likelihood function is $x_{(n)}$, the maximum-likelihood estimator of θ is $X_{(n)}$.

In order to compute the expectation of $\hat{\theta}$, the density of $X_{(n)}$, and $E[X_{(n)}]$ are used. The density of $X_{(n)}$ is

$$f_{X_{(n)}}(x) = \frac{n!}{(r-1)! (n-r)!} [F(x)]^{r-1} [1-F(x)]^{n-r} f(x) .$$

Let

$$c = \frac{n!}{(r-1)! (n-r)!} .$$

For the uniform distribution shown in Equation (2.5),

$$\begin{aligned} F(x) &= \frac{x}{\theta} . \\ f_{X_{(n)}}(x) &= c \left(\frac{x}{\theta} \right)^{r-1} \left(1 - \frac{x}{\theta} \right)^{n-r} \frac{1}{\theta} . \end{aligned}$$

The expectation of $X_{(r)}$ is

$$\begin{aligned}
 E[X_{(r)}] &= \int_0^\theta x \cdot c \left(\frac{x}{\theta}\right)^{r-1} \left(1 - \frac{x}{\theta}\right)^{n-r} \frac{1}{\theta} dx \\
 &= c \int_0^\theta \left(\frac{x}{\theta}\right)^r \left(1 - \frac{x}{\theta}\right)^{n-r} dx \\
 &= c\theta \int_0^1 \left(\frac{x}{\theta}\right)^r \left(1 - \frac{x}{\theta}\right)^{n-r} d\frac{x}{\theta} \\
 &= c\theta B(r+1, n-r+1) \\
 &= \frac{n!}{(r-1)!(n-r)!} \theta \frac{r! (n-r)!}{(n+1)!} \\
 &= \frac{r}{n+1} \theta .
 \end{aligned} \tag{2.6}$$

Putting $r=n$ in Equation (2.6), the expectation of $\hat{\theta} = X_{(n)}$ is

$$E[\hat{\theta}] = \frac{n}{n+1} \theta = \theta - \frac{1}{n+1} \theta$$

and consequently the bias of $\hat{\theta}$ is

$$\text{bias}[\hat{\theta}] = -\frac{1}{n+1} \theta . \tag{2.7}$$

In this case it is obvious that $\frac{n+1}{n} \hat{\theta}$ is an unbiased estimator for θ .

The value of $\hat{\theta}_{(i)}$ varies depending upon the observation which is deleted. If the deleted observation X_i is $X_{(n)}$, $\hat{\theta}_{(i)}$ is $X_{(n-1)}$. Otherwise $\hat{\theta}_{(i)}$ is $X_{(n)}$. Now,

$$\begin{aligned}
 \hat{\theta}_{(i)} &= \max [X_{(1)}, X_{(2)}, \dots, X_{(i-1)}, X_{(i+1)}, \dots, X_{(n)}] \\
 &= \begin{cases} X_{(n-1)} & \text{if } X_i = X_{(n)} \\ X_{(n)} & \text{if } X_i \neq X_{(n)} \end{cases}
 \end{aligned}$$

Hence

$$\begin{aligned}\hat{\theta}_{(.)} &= \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \\ &= \frac{1}{n} [X_{(n-1)} + (n-1)X_{(n)}] ,\end{aligned}\quad (2.8)$$

since there is one $X_{(n)}$ to be deleted. The jackknife method yields

$$\begin{aligned}J(\hat{\theta}) &= n\hat{\theta} - (n-1)\hat{\theta}_{(.)} \\ &= nX_{(n)} - (n-1) \frac{X_{(n-1)} + (n-1)X_{(n)}}{n} \\ &= nX_{(n)} - \frac{(n-1)}{n}X_{(n-1)} - \frac{(n-1)^2}{n}X_{(n)} \\ &= \frac{2n-1}{n}X_{(n)} - \frac{(n-1)}{n}X_{(n-1)} .\end{aligned}\quad (2.9)$$

Then, using Equation(2.6)

$$\begin{aligned}E[J(\hat{\theta})] &= \frac{2n-1}{n} E[X_{(n)}] - \frac{(n-1)}{n} E[X_{(n-1)}] \\ &= \frac{2n-1}{n} \frac{n}{n+1} \theta - \frac{(n-1)}{n} \frac{(n-1)}{n+1} \theta \\ &= \frac{\theta}{n(n+1)} (n^2+n-1) \\ &= \theta - \frac{\theta}{n(n+1)} .\end{aligned}$$

Therefore

$$\text{bias}[J(\hat{\theta})] = - \frac{\theta}{n(n+1)} .\quad (2.10)$$

Using Equations (2.7) and (2.10)

$$\text{bias}[J(\hat{\theta})] = \frac{\text{bias}(\hat{\theta})}{n}\quad (2.11)$$

and thus the bias in $\hat{\theta}$ is much reduced by the jackknife method. Since both $\hat{\theta}$ and $J(\hat{\theta})$ are biased estimators, their mean square errors have to be compared in order to see which estimator is a better one.

The mean square error of the original estimator $\hat{\theta}$ is

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= E (\hat{\theta} - \theta)^2 \\ &= E (X_{(n)} - \theta)^2 \\ &= E [X_{(n)}^2 - 2X_{(n)}\theta + \theta^2] . \end{aligned}$$

Now, the expectation of $X_{(n)}^2$ is

$$\begin{aligned} E[X_{(n)}^2] &= \int_0^\theta x^2 c \left(\frac{x}{\theta}\right)^{r-1} \left(1 - \frac{x}{\theta}\right)^{n-r} \frac{1}{\theta} dx \\ &= c\theta \int_0^\theta \left(\frac{x}{\theta}\right)^{r+1} \left(1 - \frac{x}{\theta}\right)^{n-r} dx \\ &= c\theta^2 \int_0^1 \left(\frac{x}{\theta}\right)^{r+1} \left(1 - \frac{x}{\theta}\right)^{n-r} d\frac{x}{\theta} \\ &= c\theta^2 B(r+2, n-r+1) \\ &= \frac{n!}{(r-1)!(n-r)!} \theta^2 \frac{(r+1)!(n-r)!}{(n+2)!} \\ &= \frac{r(r+1)}{(n+1)(n+2)} \theta^2 . \end{aligned} \tag{2.12}$$

Using Equations (2.6) and (2.12)

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \frac{n}{n+2} \theta^2 - \frac{2n}{n+1} \theta^2 + \theta^2 \\ &= \theta^2 \frac{n^2+n-2n^2-4n+n^2+3n+2}{(n+2)(n+1)} \end{aligned}$$

That is,

$$\text{MSE}(\hat{\theta}) = \frac{2\theta^2}{(n+2)(n+1)} . \quad (2.13)$$

The mean square error of the jackknife estimator is

$$\begin{aligned} \text{MSE}[J(\hat{\theta})] &= E[J(\hat{\theta}) - \theta]^2 \\ &= E \left(\frac{2n-1}{n} X_{(n)} - \frac{(n-1)}{n} X_{(n-1)} - \theta \right)^2 \\ &= E \left(\left(\frac{2n-1}{n} \right)^2 X_{(n)}^2 - \frac{2(2n-1)(n-1)}{n^2} X_{(n-1)} X_{(n)} \right. \\ &\quad \left. + \frac{(n-1)^2}{n^2} X_{(n-1)}^2 - 2 \left(\frac{2n-1}{n} X_{(n)} - \frac{(n-1)}{n} X_{(n-1)} \right) \theta + \theta^2 \right) . \end{aligned}$$

Now, the joint density of $X_{(r)}$ and $X_{(s)}$ for $1 \leq r \leq s \leq n$ is

$$f_{X_{(r)}, X_{(s)}}(x, y) = \frac{n! [F(x)]^{r-1} [F(y)-F(x)]^{s-r-1} [1-F(y)]^{n-s} f(x) f(y)}{(r-1)! (s-r-1)! (n-s)!} \quad \text{for } x < y.$$

$$\begin{aligned} f_{X_{(n-1)}, X_{(n)}}(x, y) &= \frac{n!}{(n-2)!} \left(\frac{x}{\theta} \right)^{n-2} \left(\frac{y}{\theta} - \frac{x}{\theta} \right)^0 \left(1 - \frac{y}{\theta} \right)^0 \frac{1}{\theta} \frac{1}{\theta} . \\ &= \frac{n(n-1)}{\theta^2} \left(\frac{x}{\theta} \right)^{n-2} \end{aligned}$$

Thus,

$$\begin{aligned} E[X_{(n-1)} X_{(n)}] &= \int_0^\theta \int_0^y xy \frac{n(n-1)}{\theta^2} \left(\frac{x}{\theta} \right)^{n-2} dx dy \\ &= \int_0^\theta \frac{n(n-1)}{\theta^n} y \left(\int_0^y x^{n-1} dx \right) dy \\ &= \int_0^\theta \frac{n(n-1)}{\theta^n} \frac{y^{n+1}}{n} dy \\ &= \frac{n-1}{\theta^n} \frac{\theta^{n+2}}{n+2} \\ &= \frac{n-1}{n+2} \theta^2 . \end{aligned} \quad (2.14)$$

Using Equations (2.6), (2.12) and (2.14)

$$\begin{aligned}
 \text{MSE}[J(\hat{\theta})] &= \left(\frac{2n-1}{n} \right)^2 \frac{n}{n+2} \theta^2 - \frac{2(2n-1)(n-1)}{n^2} \frac{n-1}{n+2} \theta^2 \\
 &\quad + \frac{(n-1)^2}{n^2} \frac{(n-1)n}{(n+1)(n+2)} \theta^2 \\
 &\quad - 2 \left(\frac{2n-1}{n} \frac{n}{n+2} - \frac{(n-1)}{n} \frac{n-1}{n+2} \right) \theta^2 + \theta^2 \\
 &= \frac{2(n^2-n+1)}{n^2(n+1)(n+2)} \theta^2 .
 \end{aligned} \tag{2.15}$$

Using Equations (2.13) and (2.15)

$$\text{MSE}[J(\hat{\theta})] = \text{MSE}(\hat{\theta}) \left(1 - \frac{(n-1)}{n^2} \right) . \tag{2.16}$$

Hence, for all $n \geq 2$, from Equation (2.11)

$$\text{bias}[J(\hat{\theta})] < \text{bias}(\hat{\theta})$$

and from Equation (2.16)

$$\text{MSE}[J(\hat{\theta})] < \text{MSE}(\hat{\theta}) .$$

Therefore the jackknife estimator is to be preferred to the maximum-likelihood estimator.

2.2.3 Bias Reduction in Uniform Distribution with Two Parameters

Let the random variable X have a uniform distribution with density given by

$$f(x) = \begin{cases} \frac{1}{2\sqrt{3}\sigma} & , \mu - \sqrt{3}\sigma \leq x \leq \mu + \sqrt{3}\sigma \\ 0 & , \text{otherwise} \end{cases} \quad (2.17)$$

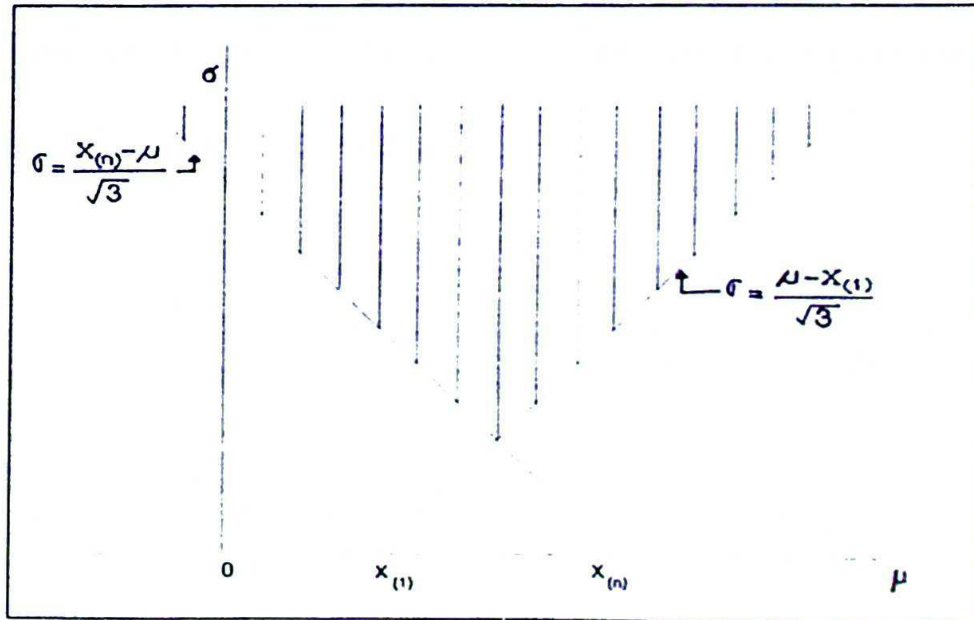
where $-\infty < \mu < \infty$ and $\sigma > 0$.

The likelihood function is

$$\begin{aligned} L(\mu, \sigma; x_1, \dots, x_n) &= \left(\frac{1}{2\sqrt{3}\sigma} \right)^n \prod_{i=1}^n I_{[\mu - \sqrt{3}\sigma, \mu + \sqrt{3}\sigma]}(x_{(i)}) \\ &= \left(\frac{1}{2\sqrt{3}\sigma} \right)^n I_{[\mu - \sqrt{3}\sigma, x_{(n)}]}(x_{(1)}) I_{[x_{(1)}, \mu + \sqrt{3}\sigma]}(x_{(n)}) \\ &= \left(\frac{1}{2\sqrt{3}\sigma} \right)^n I_{[(\mu - x_{(1)})/\sqrt{3}, \infty)}(\sigma) \\ &\quad I_{[(x_{(n)} - \mu)/\sqrt{3}, \infty)}(\sigma) I_{[x_{(1)}, \infty)}(x_{(n)}) , \end{aligned}$$

where $x_{(1)} = \min[x_1, x_2, \dots, x_n]$ and $x_{(n)} = \max[x_1, x_2, \dots, x_n]$.

Figure(2.1): The likelihood function of μ, σ



The likelihood function attains the maximum value at the smallest value of σ . This smallest value of σ is at the intersection of the lines $\mu - \sqrt{3}\sigma = x_{(1)}$ and $\mu + \sqrt{3}\sigma = x_{(n)}$ as shown in Figure(2.1) which shows the likelihood function. The shaded area represents the likelihood of μ, σ . Hence the maximum-likelihood estimators of μ and σ are

$$\hat{\mu} = \frac{x_{(n)} + x_{(1)}}{2}$$

and

$$\hat{\sigma} = \frac{x_{(n)} - x_{(1)}}{2\sqrt{3}}. \quad (2.18)$$

$\hat{\mu}$ is an unbiased estimator of μ and $\hat{\sigma}$ is a biased estimator of σ . It can be shown as follows.

The density of $X_{(r)}$ is

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} \left(F(x) \right)^{r-1} \left(1-F(x) \right)^{n-r} f(x).$$

Let

$$c = \frac{n!}{(r-1)!(n-r)!}.$$

For the uniform distribution shown in Equation (2.17),

$$F(x) = \frac{1}{2\sqrt{3}\sigma} [x - (\mu - \sqrt{3}\sigma)].$$

Therefore

$$f_{X_{(r)}}(x) = c \left(\frac{x - (\mu - \sqrt{3}\sigma)}{2\sqrt{3}\sigma} \right)^{r-1} \left(1 - \frac{x - (\mu - \sqrt{3}\sigma)}{2\sqrt{3}\sigma} \right)^{n-r} \frac{1}{2\sqrt{3}\sigma},$$

and

$$E[X_{(r)}] = c \int_{\mu - \sqrt{3}\sigma}^{\mu + \sqrt{3}\sigma} x \left(\frac{x - (\mu - \sqrt{3}\sigma)}{2\sqrt{3}\sigma} \right)^{r-1} \left(1 - \frac{x - (\mu - \sqrt{3}\sigma)}{2\sqrt{3}\sigma} \right)^{n-r} \frac{1}{2\sqrt{3}\sigma} dx.$$

Let $y = \frac{x - (\mu - \sqrt{3}\sigma)}{2\sqrt{3}\sigma}$.

Then, $dy = \frac{1}{2\sqrt{3}\sigma} dx$

and

$$\begin{aligned}
 E[X_{(n)}] &= c \int_0^1 [2\sqrt{3}\sigma y + (\mu - \sqrt{3}\sigma)] y^{r-1} (1-y)^{n-r} dy \\
 &= c \left(\int_0^1 2\sqrt{3}\sigma y^r (1-y)^{n-r} dy + (\mu - \sqrt{3}\sigma) \int_0^1 y^{r-1} (1-y)^{n-r} dy \right) \\
 &= c [2\sqrt{3}\sigma B(r+1, n-r+1) + (\mu - \sqrt{3}\sigma) B(r, n-r+1)] \\
 &= \frac{n!}{(r-1)!(n-r)!} \left(2\sqrt{3}\sigma \frac{r!(n-r)!}{(n+1)!} + (\mu - \sqrt{3}\sigma) \frac{(r-1)!(n-r)!}{n!} \right) \\
 &= \frac{r}{n+1} 2\sqrt{3}\sigma + (\mu - \sqrt{3}\sigma) . \tag{2.19}
 \end{aligned}$$

Equation(2.19) gives,

$$E[X_{(1)}] = \frac{1}{n+1} 2\sqrt{3}\sigma + (\mu - \sqrt{3}\sigma)$$

$$E[X_{(n)}] = \frac{n}{n+1} 2\sqrt{3}\sigma + (\mu - \sqrt{3}\sigma)$$

$$\begin{aligned}
 E[\hat{\mu}] &= E\left[\frac{X_{(n)} + X_{(1)}}{2} \right] \\
 &= \frac{1}{2} \left(\frac{n}{n+1} 2\sqrt{3}\sigma + (\mu - \sqrt{3}\sigma) + \frac{1}{n+1} 2\sqrt{3}\sigma + (\mu - \sqrt{3}\sigma) \right) \\
 &= \mu
 \end{aligned}$$

which shows that $\hat{\mu}$ is an unbiased estimator of μ .

Expected value of $\hat{\sigma}$ is

$$\begin{aligned}
 E[\hat{\sigma}] &= E\left(\frac{X_{(n)} - X_{(1)}}{2\sqrt{3}}\right) \\
 &= \frac{1}{2\sqrt{3}} \left(\frac{n-1}{n+1} 2\sqrt{3}\sigma \right) \\
 &= \frac{n-1}{n+1} \sigma
 \end{aligned} \tag{2.20}$$

which shows that $\hat{\sigma}$ is a biased estimator of σ .

The bias of $\hat{\sigma}$ is

$$\text{bias}(\hat{\sigma}) = -\frac{2}{n+1} \sigma \tag{2.21}$$

The jackknife estimator as shown in Equation (2.2) can be formed as follows-

The estimator $\hat{\theta}_{(1)}$ for estimating $\theta = \sigma$ will take one of three values, namely, $\frac{1}{2\sqrt{3}} [X_{(n)} - X_{(2)}]$ or $\frac{1}{2\sqrt{3}} [X_{(n)} - X_{(1)}]$ or $\frac{1}{2\sqrt{3}} [X_{(n-1)} - X_{(1)}]$ with frequency 1, n-2 and 1, respectively.

Therefore

$$\begin{aligned}
 \hat{\theta}_{(1)} &= \frac{1}{n} \left(\frac{n-1}{2\sqrt{3}} [X_{(n)} - X_{(1)}] + \frac{1}{2\sqrt{3}} [X_{(n-1)} - X_{(2)}] \right) \\
 &= \frac{1}{n} \left((n-1) \hat{\sigma} + \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}} \right)
 \end{aligned} \tag{2.22}$$

and thus

$$\begin{aligned}
 J[\hat{\theta}] &= n\hat{\theta} - (n-1)\hat{\theta}_{(1)} \\
 &= n\hat{\sigma} - \frac{(n-1)}{n} \left((n-1) \hat{\sigma} + \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}} \right) \\
 &= n\hat{\sigma} - \frac{(n^2 - 2n + 1)}{n} \hat{\sigma} - \frac{(n-1)}{n} \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}} \\
 &= \frac{2n-1}{n} \hat{\sigma} - \frac{(n-1)}{n} \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}}
 \end{aligned}$$

From Equations (2.19) and (2.20), the expected value of the jackknife estimator is

$$\begin{aligned}
 E[J(\hat{\theta})] &= \frac{2n-1}{n} E[\hat{\sigma}] - \frac{(n-1)}{n} \frac{1}{2\sqrt{3}} [E[X_{(n-1)}] - E[X_{(2)}]] \\
 &= \frac{2n-1}{n} \frac{n-1}{n+1} \sigma - \frac{(n-1)}{n} \frac{1}{2\sqrt{3}} \left(\frac{n-3}{n+1} 2\sqrt{3}\sigma \right) \\
 &= \frac{(n-1)(2n-1-n+3)\sigma}{n(n+1)} \\
 &= \frac{(n-1)(n+2)\sigma}{n(n+1)} \\
 &= \frac{n(n+2)\sigma}{n(n+1)} - \frac{(n+2)\sigma}{n(n+1)} \\
 &= \sigma + \frac{n\sigma}{n(n+1)} - \frac{(n+2)\sigma}{n(n+1)} \\
 &= \sigma - \frac{2\sigma}{n(n+1)}
 \end{aligned}$$

which shows that $J(\hat{\theta})$ is also a biased estimator. The bias of $J(\hat{\theta})$ is

$$\text{bias}[J(\hat{\theta})] = - \frac{2\sigma}{n(n+1)} \quad (2.23)$$

From Equations (2.21) and (2.23)

$$\text{bias}[J(\hat{\theta})] = \frac{\text{bias}(\hat{\sigma})}{n} \quad (2.24)$$

Thus $\text{bias}[J(\hat{\theta})] < \text{bias}(\hat{\theta})$, and bias in the estimator is much reduced by the jackknife method.

2.3 The Generalized Jackknife

It has been proved that the jackknife method reduces the bias of an estimator. In this section, it will be proved that the bias reduction property of the generalized jackknife method is better than that of the jackknife method.

Let θ be an unknown parameter, and let $X_1, X_2, \dots, X_n$ be n independent and identically distributed observations from the probability distribution F . Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are two different estimators of the parameter θ and each of these estimators is biased such that

$$E(\hat{\theta}_k) - \theta = b_k(n, \theta) \neq 0 \quad k=1, 2.$$

Let

$$R = \frac{b_1(n, \theta)}{b_2(n, \theta)} \neq 1. \quad (2.25)$$

Then the generalized jackknife estimator is defined by

$$G(\hat{\theta}) = \frac{\hat{\theta}_1 - R\hat{\theta}_2}{1-R} \quad (2.26)$$

Hence

$$\begin{aligned} E[G(\hat{\theta})] &= \frac{E(\hat{\theta}_1) - R E(\hat{\theta}_2)}{1 - R} \\ &= \frac{\theta + b_1(n, \theta) - R\theta - Rb_2(n, \theta)}{1 - R} \\ &= \theta. \end{aligned}$$

Since θ is an unknown parameter, the values of $b_1(n, \theta)$, $b_2(n, \theta)$ in Equation(2.25) will not be known in general. Even so, there will be many interesting cases in which R is known.

Suppose the functions $b_k(n, \theta)$ may be written as the product of separate functions of n and θ , that is,

$$b_k(n, \theta) = f_k(n) b_k(\theta), \quad k=1, 2.$$

Further suppose that the nature of the estimator $\hat{\theta}_2$ is similar to that of $\hat{\theta}_1$ to the extent that

$$b_1(\theta) = b_2(\theta) = b(\theta).$$

Then

$$R = \frac{f_1(n)}{f_2(n)}, \quad (2.27)$$

so that when $f_k(n)$ is known, R is known. The generalized jackknife estimator

$$G(\hat{\theta}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}} \quad (2.28)$$

is exactly unbiased. It is clear that $J(\hat{\theta})$ is a special case of $G(\hat{\theta})$ by taking $\hat{\theta}_1 = \hat{\theta}$, $\hat{\theta}_2 = \hat{\theta}_{(.)}$, $f_1(n) = 1/n$ and $f_2(n) = 1/(n-1)$ in Equation(2.28).

To eliminate k separate terms in the bias, each of which factorizes into distinct functions of n and θ , $k+1$ estimators are required whose bias are of the form

$$E(\hat{\theta}_i) - \theta = \sum_{j=1}^{\infty} f_{ij}(n) b_j(\theta), \quad i=1, 2, \dots, k+1. \quad (2.29)$$

The functions $f_{ij}(n)$ are assumed to be real numbers such that

$$\begin{vmatrix} 1 & \dots & 1 \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix} \neq 0.$$

The generalization of Equation(2.28) is

$$G^{(k)}(\hat{\theta}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \dots & \hat{\theta}_{k+1} \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \cdot & \dots & \cdot \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix}}{\begin{vmatrix} 1 & \dots & 1 \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \cdot & \dots & \cdot \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix}} \quad (2.30)$$

Then

$$E[G^{(k)}(\hat{\theta})] = \theta + B_G(n, \theta) ,$$

where

$$B_G(n, \theta) = \frac{\begin{vmatrix} B_1 & \dots & B_{k+1} \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \cdot & \dots & \cdot \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix}}{\begin{vmatrix} 1 & \dots & 1 \\ f_{11}(n) & \dots & f_{k+1,1}(n) \\ \cdot & \dots & \cdot \\ f_{1k}(n) & \dots & f_{k+1,k}(n) \end{vmatrix}} .$$

and

$$B_i = \sum_{j=k+1}^{\infty} f_{ij}(n) b_j(\theta) , \quad i=1,2,\dots,k+1$$

Hence if

$$E(\hat{\theta}_i) - \theta = \sum_{j=1}^k f_{ij}(n) b_j(\theta) , \quad i=1,2,\dots,k+1 ,$$

then $G^{(k)}(\hat{\theta})$ is an unbiased estimator of θ .

One method for selecting the k estimators $\hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_{k+1}$ is to simply extend the method of Quenouille as follows-

Let $k < n$ and let $\hat{\theta}_{j_2}, \hat{\theta}_{j_2, j_3}, \dots, \hat{\theta}_{j_2, j_3, \dots, j_{k+1}}$ be the estimators resulting by restricting $\hat{\theta}_1$ to the samples obtained by

deleting at random $1, 2, \dots, k$ of the observations, respectively.

Then $\hat{\theta}_i$ are defined by

$$\hat{\theta}_i = \bar{\theta}_{j_2, j_3, \dots, j_i}, \quad i=2, 3, \dots, k+1, \quad (2.31)$$

the average over the $\binom{n}{i-1}$ resulting statistics.

When Equation (2.31) holds and the bias in $\hat{\theta}_1$ is a power series in $1/n$, that is,

$$E[\hat{\theta}_1 - \theta] = \frac{a_1}{n} + \frac{a_2}{n^2} + \dots, \quad (2.32)$$

the bias in $\hat{\theta}_i$ becomes

$$E[\hat{\theta}_i - \theta] = \frac{a_1}{n-i+1} + \frac{a_2}{(n-i+1)^2} + \dots$$

Hence when Equations (2.31) and (2.32) hold

$$G^{(k)}(\hat{\theta}) = \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \dots & \hat{\theta}_{k+1} \\ \frac{1}{n} & \frac{1}{n-1} & \dots & \frac{1}{n-k} \\ \dots & \dots & \dots & \dots \\ \frac{1}{n^k} & \frac{1}{(n-1)^k} & \dots & \frac{1}{(n-k)^k} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \dots & 1 \\ \frac{1}{n} & \frac{1}{n-1} & \dots & \frac{1}{n-k} \\ \dots & \dots & \dots & \dots \\ \frac{1}{n^k} & \frac{1}{(n-1)^k} & \dots & \frac{1}{(n-k)^k} \end{vmatrix}}. \quad (2.33)$$

Then $E[G^{(k)}(\hat{\theta}) - \theta] = 0(n^{-k-1})$ if the $a_i \neq 0$, and $G^{(k)}(\hat{\theta})$ is unbiased if $a_{k+1} = a_{k+2} = \dots = 0$. By taking $\hat{\theta} = \hat{\theta}_1$ in Equation (2.33), the

particular form of $G^{(k)}(\hat{\theta})$ equals to $J^{(k)}(\hat{\theta})$. Therefore when $f_{ij}(n)=[1/(n-i+1)^j]$, $G^{(k)}(\hat{\theta})=J^{(k)}(\hat{\theta})$.

For some of the cases mentioned in Section 2.2, the bias reduction property of the generalized jackknife estimator can be illustrated in the following sub-sections.

2.3.1 Bias Reduction in Uniform Distribution with One Parameter

In Section 2.2.2, it has been proved that the jackknife estimator eliminates the order n^{-1} term in the bias of the original estimator, but that estimator is biased. For the same case it can be shown that the generalized jackknife estimator is an unbiased estimator as follows-

Consider

$$\hat{\theta}_1 = \hat{\theta} = X_{(n)} \quad (2.34)$$

and

$$\hat{\theta}_2 = \hat{\theta}_{(.)} = \frac{1}{n} [(n-1)X_{(n)} + X_{(n-1)}] \quad (2.35)$$

It has been shown that

$$E(\hat{\theta}_1) - \theta = -\frac{\theta}{n+1},$$

and Equation(2.6) gives that

$$\begin{aligned} E(\hat{\theta}_2) &= \frac{n-1}{n+1} \theta + \frac{n-1}{n(n+1)} \theta \\ &= \frac{n-1}{n} \theta. \end{aligned}$$

The bias of second estimator is

$$E(\hat{\theta}_2) - \theta = -\frac{\theta}{n}.$$

Therefore

$$\begin{aligned}
 G(\hat{\theta}) &= \frac{\begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 \\ f_1(n) & f_2(n) \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ f_1(n) & f_2(n) \end{vmatrix}} \\
 &= \frac{\begin{vmatrix} X_{(n)} & \frac{1}{n} [(n-1)X_{(n)} + X_{(n-1)}] \\ \frac{1}{n+1} & \frac{1}{n} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ \frac{1}{n+1} & \frac{1}{n} \end{vmatrix}} \\
 &= 2X_{(n)} - X_{(n-1)} . \tag{2.36}
 \end{aligned}$$

Thus

$$\begin{aligned}
 E[G(\hat{\theta})] &= 2 \frac{n}{n+1} \theta - \frac{n-1}{n+1} \theta \\
 &= \theta .
 \end{aligned}$$

Since $G(\hat{\theta})$ is unbiased,

$$\begin{aligned}
 \text{MSE}[G(\hat{\theta})] &= \text{Var}[G(\hat{\theta})] \\
 &= E[G(\hat{\theta})^2] - [E[G(\hat{\theta})]]^2 \\
 &= E[4X_{(n)}^2 - 4X_{(n)}X_{(n-1)} + X_{(n-1)}^2] - \theta^2
 \end{aligned}$$

Using Equations (2.12) and (2.14),

$$\begin{aligned}
 \text{Var}[G(\hat{\theta})] &= 4 \frac{n}{n+2} \theta^2 - 4 \frac{n-1}{n+2} \theta^2 + \frac{n(n-1)}{(n+1)(n+2)} \theta^2 - \theta^2 \\
 &= \frac{(4n+4+n^2-n-n^2-3n-2) \theta^2}{(n+1)(n+2)}
 \end{aligned}$$

That is,

$$\text{Var}[G(\hat{\theta})] = \frac{2\theta^2}{(n+1)(n+2)} \quad (2.37)$$

Comparing this result with the MSE of the original estimator in Equation(2.13), it is obvious that the generalized jackknife estimator completely eliminates the bias without increasing the MSE.

2.3.2 Bias Reduction in Uniform Distribution with Two Parameters

In this sub-section, the generalized jackknife estimator is found for the case in Section 2.2.3 .

Consider

$$\hat{\sigma}_1 = \hat{\sigma} = \frac{X_{(n)} - X_{(1)}}{2\sqrt{3}}$$

and

$$\hat{\sigma}_2 = \hat{\sigma}_{(.)} = \frac{1}{n} \left[(n-1)\hat{\sigma} + \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}} \right] .$$

Then

$$E(\hat{\sigma}_1) - \sigma = -\frac{2}{n+1} \sigma$$

and Equations (2.19) and (2.20) provide,

$$\begin{aligned} E(\hat{\sigma}_2) &= \frac{1}{n} \left[(n-1) \left(1 - \frac{2}{n+1} \right) \sigma + \frac{n-3}{n+1} \sigma \right] \\ &= \frac{1}{n} \left[(n-1)\sigma - \frac{(2n-2-n+3)}{n+1} \sigma \right] \\ &= \frac{n-2}{n} \sigma . \end{aligned}$$

Thus

$$E(\hat{\sigma}_2) - \sigma = -\frac{2}{n} \sigma .$$

Hence

$$\begin{aligned}
 G(\hat{\theta}) &= \frac{\begin{vmatrix} \hat{\sigma} & \frac{1}{n} [(n-1)\hat{\sigma} + \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}}] \\ -\frac{2}{n+1} & -\frac{2}{n} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ -\frac{2}{n+1} & -\frac{2}{n} \end{vmatrix}} \\
 &= (n+1)\hat{\sigma} - (n-1)\hat{\sigma} - \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}} \\
 &= 2\hat{\sigma} - \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}} \\
 &= \frac{X_{(n)} - X_{(1)}}{\sqrt{3}} - \frac{X_{(n-1)} - X_{(2)}}{2\sqrt{3}}
 \end{aligned}$$

Thus

$$\begin{aligned}
 E[G(\hat{\theta})] &= \frac{2(n-1)}{(n+1)} \sigma - \frac{(n-3)}{(n+1)} \sigma \\
 &= \sigma .
 \end{aligned}$$

Therefore it can be concluded that the bias reduction property of the generalized jackknife is better than that of the ordinary jackknife and that for the two cases considered it provides unbiased estimators.

2.4 Estimation of a Truncation Point

In many practical applications of statistical theory the random variables which enter the mathematical model are restricted to a finite range; even through the assumed distribution, which is satisfactory in all other aspects, may ordinarily be defined upon random variables whose range is unbounded. In these instances the estimation of the truncation point may be of definite practical importance.

Suppose an estimation of a truncation point θ is desired and a random sample of size n from the distribution $F(x)$ truncated to $x \leq \theta$ ($0 < F(\theta) < 1$) is obtained. Estimators are based upon the order statistics $X_1 < X_2 < \dots < X_n$, where the unordered X 's are assumed to be independent and identically distributed according to some distribution

$$F_\theta(x) = \frac{F(x)}{F(\theta)} \quad (2.38)$$

belonging to the class of functions which admits a Taylor series expansion about θ . Estimator of θ which are unbiased to order n^{-k} ($1 \leq k \leq n$) for all F in this class are constructed as linear combinations of order statistics, using a general method of construction proposed by Quenouille.

Application of Quenouille's method to the problem at hand simply requires a judiciously chosen initial estimator $\hat{\theta}$ having a bias term in the form of a power series in $1/n$. A natural choice for estimating the upper bound of X is the largest of the n observations on X , that is, $\hat{\theta} = \max[X_1, \dots, X_n] = X_{(n)}$.

The bias in $\hat{\theta}$ is computed by first making the probability transformation

$$Y_n = F_{\theta}(X_{(n)}) = F(X_{(n)})/F(\theta) ,$$

so that

$$X_{(n)} = H_{F_{\theta}}(Y_n) .$$

The function $H_{F_{\theta}}(Y_n)$ may be expanded in a Taylor series about the point $Y_n=1$, to yield (omitting the subscript for H)

$$X_{(n)} = \theta + (Y_n-1) H'(1) + \frac{(Y_n-1)^2}{2!} H''(1) + \frac{(Y_n-1)^3}{3!} H'''(1) + \dots \quad (2.39)$$

Since $(1-Y_n)$ is distributed as Y_1 , the smallest of n independent observations from the uniform distribution on the unit interval,

$$E(Y_n-1)^k = (-1)^k E(Y_1^k) .$$

It can be seen that the probability distribution functions of Y_n and Y_1 are

$$\begin{aligned} F_{Y_n}(y) &= \sum_{i=0}^n \binom{n}{i} [F(y)]^i [1-F(y)]^{n-i} \\ &= [F(y)]^n , \end{aligned}$$

and

$$\begin{aligned} F_{Y_1}(y) &= \sum_{i=1}^n \binom{n}{i} [F(y)]^i [1-F(y)]^{n-i} \\ &= 1 - [1-F(y)]^n . \end{aligned}$$

The density function of Y_1 is

$$f_{Y_1}(y) = n [1-F(y)]^{n-1} f(y) .$$

If observations are randomly selected from the uniform distribution on the unit interval, that is, $f(y)=1$ and $F(y)=y$ for y in $[0,1]$.

$$f_{Y_1}(y) = n [1-y]^{n-1}.$$

Therefore

$$\begin{aligned} E(Y_1^k) &= \int_0^1 y_1^k n(1-y_1)^{n-1} dy_1 \\ &= n \int_0^1 y_1^k (1-y_1)^{n-1} dy_1 \\ &= n B(k+1, n) \\ &= \frac{k! n!}{(n+k)!}. \end{aligned}$$

Hence

$$\begin{aligned} E(Y_n-1)^k &= (-1)^k E(Y_1^k) \\ &= (-1)^k \frac{k! n!}{(n+k)!}. \end{aligned}$$

Then the expected value of Equation(2.39) gives

$$E(X_{(n)}) = \theta - \frac{H'(1)}{(n+1)} + \frac{H''(1)}{(n+1)(n+2)} - \frac{H'''(1)}{(n+1)(n+2)(n+3)} + \dots \quad (2.40)$$

Consequently, the bias of the original estimator $\hat{\theta}$ is

$$\text{bias}[X_{(n)}] = - \frac{H'(1)}{(n+1)} + \frac{H''(1)}{(n+1)(n+2)} - \frac{H'''(1)}{(n+1)(n+2)(n+3)} + \dots \quad (2.41)$$

Similar expansion of $X_{(n-1)}$ as in Equation(2.39) gives

$$X_{(n-1)} = \theta + (Y_{n-1}-1) H'(1) + \frac{(Y_{n-1}-1)^2}{2!} H''(1) + \frac{(Y_{n-1}-1)^3}{3!} H'''(1) + \dots \quad (2.42)$$

Since $(1-Y_{n-1})$ is distributed as Y_2 , the second smallest of n independent observations from the uniform distribution on the unit interval,

$$E(Y_{n-1}-1)^k = (-1)^k E(Y_2^k) .$$

The probability distribution function of Y_2 is

$$\begin{aligned} F_{Y_2}(y) &= \sum_{i=2}^n \binom{n}{i} [F(y)]^i [1-F(y)]^{n-i} \\ &= 1 - [1-F(y)]^n - n[F(y)][1-F(y)]^{n-1} , \end{aligned}$$

and the density function is

$$\begin{aligned} f_{Y_2}(y) &= n [1-F(y)]^{n-1} f(y) + n(n-1) [F(y)][1-F(y)]^{n-2} f(y) \\ &\quad - n f(y) [1-F(y)]^{n-1} \\ &= n(n-1) [F(y)][1-F(y)]^{n-2} f(y) \\ &= n(n-1) y (1-y)^{n-2} \end{aligned}$$

if $y \in [0, 1]$. Then

$$\begin{aligned} E[Y_2^k] &= \int_0^1 y^k n(n-1) y (1-y)^{n-2} dy \\ &= n(n-1) \int_0^1 y^{k+1} (1-y)^{n-2} dy \\ &= n(n-1) B(k+2, n-1) \\ &= \frac{(k+1)! n!}{(n+k)!} . \end{aligned}$$

Therefore

$$\begin{aligned} E[Y_{n-1}-1]^k &= (-1)^k E[Y_2^k] \\ &= (-1)^k \frac{(k+1)! n!}{(n+k)!} , \end{aligned}$$

and hence the expected value of Equation(2.42) becomes

$$E(X_{n-1}) = \theta - \frac{2H'(1)}{(n+1)} + \frac{3H''(1)}{(n+1)(n+2)} - \frac{4H'''(1)}{(n+1)(n+2)(n+3)} + \dots \quad (2.43)$$

It has been shown that in Equations (2.9) and (2.36), if the original estimator is $\hat{\theta}=X_{(n)}$, then the jackknife estimator is

$$J(\hat{\theta}) = \frac{2n-1}{n} X_{(n)} - \frac{n-1}{n} X_{(n-1)} ,$$

and the generalized jackknife estimator is

$$G(\hat{\theta}) = 2X_{(n)} - X_{(n-1)} .$$

Thus

$$E[J(\hat{\theta})]-\theta = -\frac{H'(1)}{n(n+1)} - \frac{(n-2)H''(1)}{n(n+1)(n+2)} + \frac{(2n-3)H'''(1)}{n(n+1)(n+2)(n+3)} + \dots \quad (2.44)$$

and

$$E[G(\hat{\theta})]-\theta = -\frac{H''(1)}{(n+1)(n+2)} + \frac{2H'''(1)}{(n+1)(n+2)(n+3)} - \dots \quad (2.45)$$

Comparison of Equation (2.44) and Equation (2.45) suggests that $J(\hat{\theta})$ and $G(\hat{\theta})$ eliminate the bias of order n^{-1} term, and $G(\hat{\theta})$ should have slightly better bias reduction property.

To find the second-order generalized jackknife estimator $G^{(2)}(\hat{\theta})$, it has been found from Equations (2.34) and (2.35) that

$$\hat{\theta}_1 = \hat{\theta} = X_{(n)} ,$$

and

$$\hat{\theta}_2 = \hat{\theta}_{(.)} = \frac{1}{n} [(n-1)X_{(n)} + X_{(n-1)}]$$

when the original estimator is $X_{(n)}$.

If two distinct trials are deleted from the sample, the estimator $\hat{\theta}_{(ij)}$ assumes the values

$$X_{(n)} , X_{(n-1)} , \text{ and } X_{(n-2)}$$

and does so with frequencies

$$\binom{n-1}{2} , \binom{n-2}{1} , \text{ and } 1 ,$$

respectively. Hence, the third estimator is

$$\hat{\theta}_3 = \bar{\hat{\theta}}_{(ij)} = \frac{2}{n(n-1)} \left(\frac{(n-1)(n-2)}{2} X_{(n)} + (n-2) X_{(n-1)} + X_{(n-2)} \right) . \quad (2.46)$$

In order to find the bias of $\hat{\theta}_3$, the expected value of $X_{(n-2)}$ has to be derived.

Similar expansion of $X_{(n-2)}$ as in Equation(2.39) gives

$$X_{(n-2)} = \theta + (Y_{n-2}-1) H'(1) + \frac{(Y_{n-2}-1)^2}{2!} H''(1) + \frac{(Y_{n-2}-1)^3}{3!} H'''(1) + \dots \quad (2.47)$$

Since $(1-Y_{n-2})$ is distributed as Y_3 , the third smallest of n independent observations from the uniform distribution on the unit interval, then

$$E(Y_{n-2}-1)^k = (-1)^k E(Y_3^k) .$$

The probability distribution function of Y_3 is

$$\begin{aligned} F_{Y_3}(y) &= \sum_{i=3}^n \binom{n}{i} [F(y)]^i [1-F(y)]^{n-i} \\ &= 1 - [1-F(y)]^n - n[F(y)][1-F(y)]^{n-1} \\ &\quad - \frac{n(n-1)}{2} [F(y)]^2 [1-F(y)]^{n-2} \end{aligned}$$

and the density function is

$$\begin{aligned}
 F_{Y_3}(y) &= n[1-F(y)]^{n-1}f(y) + n(n-1)[F(y)][1-F(y)]^{n-2}f(y) \\
 &\quad - nf(y)[1-F(y)]^{n-1} + \frac{n(n-1)(n-2)}{2}[F(y)]^2[1-F(y)]^{n-3}f(y) \\
 &\quad - n(n-1)F(y)f(y)[1-F(y)]^{n-2} \\
 &= \frac{n(n-1)(n-2)}{2}[F(y)]^2[1-F(y)]^{n-3}f(y) \\
 &= \frac{n(n-1)(n-2)}{2}y^2(1-y)^{n-3}
 \end{aligned}$$

if $y \in [0, 1]$. Then

$$\begin{aligned}
 E[Y_3^k] &= \int_0^1 y^k \frac{n(n-1)(n-2)}{2} y^2 (1-y)^{n-3} dy \\
 &= \frac{n(n-1)(n-2)}{2} B(k+3, n-2) \\
 &= \frac{(k+2)! n!}{2(n+k)!}
 \end{aligned}$$

and thus

$$\begin{aligned}
 E[Y_{n-2-1}]^k &= (-1)^k E[Y_3^k] \\
 &= (-1)^k \frac{(k+2)! n!}{2(n+k)!} .
 \end{aligned}$$

The expected value of Equation(2.47) gives

$$\begin{aligned}
 E(X_{(n-2)}) &= \theta - \frac{3H'(1)}{(n+1)} + \frac{6H''(1)}{(n+1)(n+2)} - \frac{10H'''(1)}{(n+1)(n+2)(n+3)} \\
 &\quad + \frac{15H^{(4)}(1)}{(n+1)(n+2)(n+3)(n+4)} - \dots
 \end{aligned} \tag{2.48}$$

Since

$$f_{11}(n) = \frac{1}{n+1}, \quad f_{12}(n) = \frac{1}{(n+1)(n+2)},$$

$$f_{21}(n) = \frac{1}{n}, \quad f_{22}(n) = \frac{1}{n(n+1)},$$

and

$$f_{31}(n) = \frac{1}{n-1}, \quad f_{32}(n) = \frac{1}{n(n-1)},$$

the second-order generalized jackknife is

$$G^{(2)}(\hat{\theta}) = \begin{vmatrix} \hat{\theta}_1 & \hat{\theta}_2 & \hat{\theta}_3 \\ \frac{1}{n+1} & \frac{1}{n} & \frac{1}{n-1} \\ \frac{1}{(n+1)(n+2)} & \frac{1}{n(n+1)} & \frac{1}{n(n-1)} \end{vmatrix} \left(\sum_{j=1}^3 c_2^j \right)^{-1}$$

where

$$\begin{aligned} c_2^1 &= \left(\frac{1}{n^2(n-1)} - \frac{1}{n(n-1)(n+1)} \right) \\ &= \frac{1}{n^2(n-1)(n+1)}, \end{aligned}$$

$$\begin{aligned} c_2^2 &= \left(\frac{1}{(n-1)(n+1)(n+2)} - \frac{1}{n(n-1)(n+1)} \right) \\ &= \frac{-2}{n(n-1)(n+1)(n+2)}, \end{aligned}$$

and

$$\begin{aligned} c_2^3 &= \left(\frac{1}{n(n+2)^2} - \frac{1}{n(n+1)(n+2)} \right) \\ &= \frac{1}{n(n+1)^2(n+2)}. \end{aligned}$$

Now

$$\sum_{j=1}^3 c_2^j = \frac{2}{n^2(n+1)^2(n-1)(n+2)},$$

and thus

$$\begin{aligned} G^{(2)}(\hat{\theta}) &= \frac{1}{2} [(n+1)(n+2)\hat{\theta}_1 - 2n(n+1)\hat{\theta}_2 + n(n-1)\hat{\theta}_3] \\ &= \frac{1}{2} [(n+1)(n+2) - 2(n+1)(n-1) + (n-1)(n-2)] X_{(n)} \\ &\quad + \frac{1}{2} [2(n-2) - 2(n+1)] X_{(n-1)} + X_{(n-2)} \\ &= 3X_{(n)} - 3X_{(n-1)} + X_{(n-2)}. \end{aligned} \quad (2.49)$$

Hence the bias of $G^{(2)}(\hat{\theta})$ is

$$E[G^{(2)}(\hat{\theta}) - \theta] = - \frac{H'''(1)}{(n+1)(n+2)(n+3)} + \frac{3H^{(4)}(1)}{(n+1)(n+2)(n+3)(n+4)} - \dots \quad (2.50)$$

The truncation point estimator $G^{(2)}(\hat{\theta})$ eliminates the terms of bias up to order n^{-2} and the general truncation point estimator $G^{(k)}(\hat{\theta})$ will eliminate the terms of bias up to order n^{-k} .

C H A P T E R I I I

JACKKNIFE ESTIMATION OF BIAS AND STANDARD ERROR

3.1 Introduction

Most of the classical statistical procedures are based on two assumptions: (1) the sample observations are independent and identically distributed and, (2) the underlying distribution is normal. While the first assumption may not be unrealistic in certain situations, it is the second assumption that is rather unrealistic from a practical point of view. To assume normality instead of the nonnormal distributions encountered in practice might lead to erroneous statistical inferences.

Thus there has been an increasing interest in the development of robust statistical methods that do not depend on the assumption of normality. But these methods need a lot of computation which is impractical without the use of computers. With the advent of computers, these methods can now be used extensively. One possible method is called the jackknife method.

The jackknife is a general method not only for reducing the bias in an estimator but also for obtaining a measure for the precision of an estimator by nonparametric estimation methods. The reduction of bias by jackknife and the

generalized jackknife are explained in the Chapter II. A brief review of the jackknife procedure in estimating bias and the standard error concerned is given in the following sections.

3.2 Estimated Bias of an Estimator

Quenouille's method yields the bias-corrected jackknifed estimate of the unknown parameter θ which is from Equation(2.2) as-

$$\begin{aligned} J(\hat{\theta}) &= n\hat{\theta} - (n-1)\hat{\theta}_{(.)} \\ &= \hat{\theta} - (n-1)(\hat{\theta}_{(.)} - \hat{\theta}) . \end{aligned}$$

Hence, the nonparametric estimate of bias provided by the jackknife method is

$$\begin{aligned} \hat{BIAS}_{JACK} &\equiv \hat{\theta} - J(\hat{\theta}) \\ &= (n-1)(\hat{\theta}_{(.)} - \hat{\theta}) . \end{aligned} \quad (3.1)$$

3.2.1 Estimated Bias of a Variance Estimator

Let

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

be an estimator of $\theta = E[X - E(X)]^2$, which is the variance of an unknown probability distribution F .

Then the expected value of $\hat{\theta}$ is

$$E(\hat{\theta}) = \frac{n-1}{n} \theta .$$

since the sample variance, $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, is an unbiased estimator for θ .

Obviously, the bias of $\hat{\theta}$ is

$$\text{bias}(\hat{\theta}) = -\frac{1}{n} \theta . \quad (3.2)$$

The i^{th} point x_i is deleted from the data set and the corresponding recomputed value of $\hat{\theta}$ is

$$\hat{\theta}_{(i)} = \frac{1}{n-1} \sum_{j \neq i}^n (x_j - \bar{x}_{(i)})^2$$

$$\text{where } \bar{x}_{(i)} = \frac{1}{n-1} \sum_{j \neq i}^n x_j .$$

$$\begin{aligned} \sum_{j \neq i}^n (x_j - \bar{x}_{(i)})^2 &= \left[\sum_{j \neq i}^n x_j^2 - (n-1) \bar{x}_{(i)}^2 \right] \\ &= \sum_{i=1}^n x_i^2 - x_i^2 - (n-1) \left[\left(\sum_{j=1}^n x_j - x_i \right) / (n-1) \right]^2 \\ &= \frac{1}{n-1} \left[(n-1) \sum_{i=1}^n x_i^2 - (n-1) x_i^2 - (n^2 \bar{x}^2 - 2n \bar{x} x_i - x_i^2) \right] \\ &= \frac{1}{n-1} \left[(n-1) \sum_{i=1}^n x_i^2 - n x_i^2 - n^2 \bar{x}^2 + 2n \bar{x} x_i \right]. \quad (3.3) \end{aligned}$$

Hence

$$\hat{\theta}_{(i)} = \frac{1}{(n-1)^2} \left[(n-1) \sum_{i=1}^n x_i^2 - n x_i^2 - n^2 \bar{x}^2 + 2n \bar{x} x_i \right] ,$$

and $\hat{\theta}_{(i)}$ becomes,

$$\begin{aligned} \hat{\theta}_{(i)} &= \frac{1}{(n-1)^2} \left[(n-1) \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i^2 - n^2 \bar{x}^2 + 2n \bar{x}^2 \right] \\ &= \frac{1}{(n-1)^2} \left[(n-2) \sum_{i=1}^n x_i^2 - (n-2) n \bar{x}^2 \right] \\ &= \frac{n-2}{(n-1)^2} \left[\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right] \\ &= \frac{n-2}{(n-1)^2} \sum_{i=1}^n (x_i - \bar{x})^2 . \end{aligned}$$

Thus

$$\begin{aligned}
 J(\hat{\theta}) &= n\hat{\theta} - (n-1)\hat{\theta}_{(.)} \\
 &= \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n-2}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 .
 \end{aligned} \tag{3.4}$$

This result is the usual unbiased estimator of θ . In this case $E(\hat{\theta}) = \theta - \frac{1}{n}\theta$. Comparing with Equation(2.3), it can be seen that $a_1 = -\theta$ and $a_j = 0$ for $j \geq 2$. Equation(2.4) shows that $J(\hat{\theta})$ is exactly unbiased.

The jackknife estimate of bias is

$$\begin{aligned}
 \hat{BIAS}_{JACK} &= (n-1)(\hat{\theta}_{(.)} - \hat{\theta}) \\
 &= \frac{n-2}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n-1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \frac{(n^2 - 2n - n^2 + 2n - 1)}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= -\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 .
 \end{aligned} \tag{3.5}$$

Here, the jackknife estimator of bias is unbiased for estimating the true bias of $\hat{\theta}$ shown in Equation(3.2).

The expected value of the original biased estimator is

$$E(\hat{\theta}) = \theta + \text{bias}$$

and the jackknife estimate of bias is

$$\hat{BIAS}_{JACK} \equiv \hat{\theta} - J(\hat{\theta}).$$

In this case, the jackknife estimator is an unbiased estimator of θ , that is,

$$E[J(\hat{\theta})] = \theta.$$

Thus the expected value of the estimated bias is

$$\begin{aligned} E(\hat{BIAS}_{JACK}) &\equiv E(\hat{\theta}) - E[J(\hat{\theta})] \\ &= \text{bias.} \end{aligned}$$

Therefore, it can be concluded that if $J(\hat{\theta})$ can completely eliminate the bias term, $\hat{BIAS}_{JACK}$ is an unbiased estimator of the true bias of estimator $\hat{\theta}$.

3.3 Estimated Standard Error

Tukey(1958) suggested that a nonparametric estimate of standard error can be calculated by using the recomputed statistic $\hat{\theta}_{(i)}$.

Now, variance of $\hat{\theta}$ is

$$\text{var}(\hat{\theta}) = E[\hat{\theta} - E(\hat{\theta})]^2.$$

Tukey's formula for estimating $\text{var}(\hat{\theta})$ is

$$\hat{VAR}_{JACK} = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2. \quad (3.6)$$

Then the jackknife estimate of the standard error is

$$\hat{SE}_{JACK} = \sqrt{\hat{VAR}_{JACK}} \quad (3.7)$$

The estimated standard errors of some statistics will be found in the following sub-sections.

3.3.1 Estimated Standard Error of the Trimmed Mean

Let $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ be the order statistics of a sample and further let α be the proportion of points which are trimmed from each end of the sample. Suppose $n\alpha$ is an integer.

The α trimmed mean is the average of the observations remaining after the removal of the extremes, $n\alpha$ ($\alpha < \frac{1}{2}$) at each end. If the parameter to be estimated is the mean, the sample trimmed mean is

$$\begin{aligned}\bar{X}_T &= \frac{1}{n-2n\alpha} \sum_{i=n\alpha+1}^{n-n\alpha} X_{(i)} , \quad 0 \leq \alpha < \frac{1}{2} \\ &= \frac{1}{n(1-2\alpha)} \sum_{i=n\alpha+1}^{n-n\alpha} X_{(i)} .\end{aligned}$$

Suppose $g=(n-1)\alpha$ is an integer. After the i^{th} point x_i from the data set has been deleted and $(n-1)\alpha$ points have been trimmed from each end of the sample, the recomputed value of the estimator $\hat{\theta}_{(i)}$ is

$$\hat{\theta}_{(i)} = \begin{cases} \frac{1}{n-2g-1} \sum_{i=g+2}^{n-g} x_{(i)} & \text{if } x_i \leq x_{(g+1)} \\ \frac{1}{n-2g-1} \left[\sum_{i=g+1}^{n-g} x_{(i)} - x_{(W_i)} \right] & \text{if } x_i = x_{(W_i)} \\ \frac{1}{n-2g-1} \sum_{i=g+1}^{n-g-1} x_{(i)} & \text{if } x_i \geq x_{(n-g)} \end{cases}$$

where $x_{(g+1)} < x_{(W_i)} < x_{(n-g)}$

It can generally be written as

$$\hat{\theta}_{(i)} = \frac{1}{(n-1)(1-2\alpha)} \left(\sum_{i=g+1}^{n-g} x_{(i)} - x_{(W_i)} \right)$$

where $W_i = \begin{cases} g+1 , & i \leq g+1 \\ i , & g+1 < i < n-g \\ n-g , & i \geq n-g \end{cases}$

and hence

$$\hat{\theta}_{(i)} = \frac{1}{(n-1)(1-2\alpha)} \left(\sum_{i=g+1}^{n-g} x_{(i)} - \frac{1}{n} \sum_{i=1}^n x_{(W_i)} \right)$$

That is,

$$\hat{\theta}_{(.)} = \frac{1}{(n-1)(1-2\alpha)} \left(\sum_{i=g+1}^{n-g} x_{(i)} - \bar{x}_w \right)$$

where $\bar{x}_w = \frac{1}{n} \sum_{i=1}^n x_{(W_i)}.$

Therefore, the jackknife estimate of standard error for the trimmed mean is

$$\begin{aligned} \hat{SE}_{JACK} &= \left(\frac{n-1}{n} \sum_{i=1}^n [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 \right)^{\frac{1}{2}} \\ &= \left(\frac{n-1}{n} \sum_{i=1}^n \frac{1}{(n-1)^2(1-2\alpha)^2} [x_{(W_i)} - \bar{x}_w]^2 \right)^{\frac{1}{2}} \\ &= \left(\frac{1}{n(n-1)(1-2\alpha)^2} \sum_{i=1}^n [x_{(W_i)} - \bar{x}_w]^2 \right)^{\frac{1}{2}}. \end{aligned}$$

3.3.2 Estimated Standard Error of the Sample Median

The trimmed mean with $\alpha \rightarrow 0.5$ is the sample median. When the sample size is odd, that is, $n=2m-1$, the sample median $\hat{\theta}$ is $x_{(m)}$. The jackknife estimate of variance formula fails in this case. In the even sample size, $n=2m$, the sample median is

$$\hat{\theta} = (x_{(m)} + x_{(m+1)}) / 2.$$

Then

$$\hat{\theta}_{(i)} = x_{(i)}, \text{ where } i = \begin{cases} m+1 & , \quad x_i \leq x_{(m)} \\ m & , \quad x_i > x_{(m)} \end{cases}$$

and thus

$$\begin{aligned} \hat{\theta}_{(.)} &= \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \\ &= \frac{1}{2m} [m x_{(m+1)} + m x_{(m)}] \\ &= \frac{1}{2} [x_{(m+1)} + x_{(m)}]. \end{aligned}$$

Therefore

$$\begin{aligned} [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 &= \left[X_{(i)} - \frac{X_{(m+1)} + X_{(m)}}{2} \right]^2 \\ &= X_{(i)}^2 - X_{(i)}[X_{(m+1)} + X_{(m)}] + \frac{[X_{(m+1)} + X_{(m)}]^2}{4}, \end{aligned}$$

and,

$$\begin{aligned} \sum_{i=1}^{2m} [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 &= m[X_{(m+1)}^2 + X_{(m)}^2] - m[X_{(m+1)} + X_{(m)}]^2 + \frac{m}{2} [X_{(m+1)} + X_{(m)}]^2 \\ &= \frac{m}{2} [X_{(m+1)}^2 + X_{(m)}^2 - 2X_{(m+1)}X_{(m)}] \\ &= \frac{n}{4} [X_{(m+1)} - X_{(m)}]^2 \end{aligned}$$

and hence the jackknife estimate of standard error for the sample median is

$$\begin{aligned} \hat{SE}_{JACK} &= \left(\frac{n-1}{n} \sum_{i=1}^n [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 \right)^{1/2} \\ &= \left(\frac{n-1}{4} [X_{(m+1)} - X_{(m)}]^2 \right)^{1/2}. \end{aligned}$$

3.3.3 Estimated Standard Error of the Sample Variance

Let

$$\hat{\theta} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

be an estimate of $\theta = E[X - E(X)]^2$ which is the variance of an unknown probability distribution F . The k^{th} central moments of F and $\hat{F}$ can be defined as

$$\mu_k = E[X - E(X)]^k$$

and

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

respectively. Then the value of $\hat{\theta}_{(i)}$ is

$$\hat{\theta}_{(i)} = \frac{1}{n-2} \sum_{j \neq i}^n (x_j - \bar{x}_{(i)})^2$$

where $\bar{x}_{(i)} = \frac{1}{n-1} \sum_{j \neq i}^n x_j$.

Equation(3.3) provides that

$$\hat{\theta}_{(i)} = \frac{1}{(n-1)(n-2)} \left[(n-1) \sum_{i=1}^n x_i^2 - nx_i^2 - n^2 \bar{x}^2 + 2n\bar{x}x_i \right],$$

and thus

$$\begin{aligned} \hat{\theta}_{(.)} &= \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} \\ &= \frac{1}{(n-1)(n-2)} \left[(n-2) \sum_{i=1}^n x_i^2 - (n-2)n\bar{x}^2 \right]. \end{aligned}$$

Then, the jackknife estimate of standard error can be derived as follows-

$$\begin{aligned} \hat{\theta}_{(i)} - \hat{\theta}_{(.)} &= \frac{1}{(n-1)(n-2)} \left[\sum_{i=1}^n x_i^2 + 2n\bar{x}x_i - nx_i^2 - 2n\bar{x}^2 \right] \\ &= \frac{1}{(n-1)(n-2)} \left[\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) - (nx_i^2 - 2n\bar{x}x_i + n\bar{x}^2) \right] \\ &= \frac{1}{(n-1)(n-2)} [n\hat{\mu}_2 - n(x_i - \bar{x})^2] \\ &= \frac{1}{(n-1)(n-2)} \left[\sum_{i=1}^n (x_i - \bar{x})^2 - n(x_i - \bar{x})^2 \right] \end{aligned}$$

$$[\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 = \frac{1}{(n-1)^2(n-2)^2} [n^2\hat{\mu}_2^2 - 2n^2\hat{\mu}_2(x_i - \bar{x})^2 + n^2(x_i - \bar{x})^4]$$

and

$$\sum_{i=1}^n [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 = \frac{n^3}{(n-1)^2(n-2)^2} (\hat{\mu}_2^2 - 2\hat{\mu}_2^2 + \hat{\mu}_4).$$

Hence

$$\begin{aligned} \hat{SE}_{JACK} &= \left(\frac{n-1}{n} \sum_{i=1}^n [\hat{\theta}_{(i)} - \hat{\theta}_{(.)}]^2 \right)^{\frac{1}{2}} \\ &= \left(\frac{n^2}{(n-1)(n-2)^2} (\hat{\mu}_4 - \hat{\mu}_2^2) \right)^{\frac{1}{2}}, \end{aligned} \quad (3.8)$$

Here $\hat{SE}_{JACK}$ agrees with the true standard error of $\hat{\theta}$, which is

$$se(\hat{\theta}) = \left(\frac{1}{n} \left(\mu_4 - \frac{n-3}{n-1} \mu_2^2 \right) \right)^{\frac{1}{2}}.$$

C H A P T E R I V

BOOTSTRAP METHOD OF ESTIMATION

4.1 Introduction

Parametric methods include classical statistical techniques which are based on specific assumptions about the population sampled. In the distribution-free and nonparametric methods, specific assumptions about the distribution are not made. They are replaced by very general assumptions and the analysis is based on some function of the sample observations whose sampling distribution can be determined without the knowledge of specific distribution function of the random variable concerned. The assumption most frequently required is simply that the population be continuous. More restrictive assumptions are sometimes made, for example, that the population is symmetrical, but not to the extent that the distribution is specifically postulated. Jackknife and bootstrap estimation methods can be included in the distribution-free and nonparametric methods. They can be used in statistical estimation as alternatives to parametric methods.

Besides the classical methods, bias and standard error of an estimator can be found by such nonparametric methods as jackknife and bootstrap. Jackknife method has been discussed

in the previous chapter. In this chapter, the general procedure of bootstrap method, and how bootstrap estimates of bias and standard error are calculated will be discussed. Methods for finding the confidence intervals will also be investigated. The discussion will be extended to the cases of censored data and half-sampling.

4.2 Bootstrap Method

Bootstrap Method is a nonparametric method which is used to estimate the bias and standard error of an estimator.

In many statistical problems, the probability distribution function F though unknown, an i.i.d. random sample $(X_1, X_2, \dots, X_n)$ from F can be observed. Let $\theta(F)$ be some interesting property of F . Then, it may be estimated by $\hat{\theta} = \theta(\hat{F})$, where $\hat{F}$ is the empirical distribution function putting probability mass $1/n$ at each observed value x_i . Some measure of accuracy to $\theta(\hat{F})$ is to be assigned.

Let $\sigma(F)$ be that measure of accuracy of $\theta(\hat{F})$. Then the bootstrap estimate of accuracy becomes

$$\hat{\sigma}_{\text{BOOT}} = \sigma(\hat{F}). \quad (4.1)$$

In other words, $\hat{\sigma}_{\text{BOOT}}$ is the measure of accuracy that would be obtained if the true F equals $\hat{F}$. Computer simulation methods are used to calculate $\hat{\sigma}_{\text{BOOT}}$. A bootstrap sample $(X_1^*, X_2^*, \dots, X_n^*)$ is drawn from $\hat{F}$, in which each X_j^* independently takes value x_j with probability $1/n$, $j=1, 2, \dots, n$.

In other words, $(X_1^*, X_2^*, \dots, X_n^*)$ is an independent sample of size n drawn with replacement from the set of observations $(X_1, X_2, \dots, X_n)$. This gives a bootstrap empirical distribution function $\hat{F}^*$, the empirical distribution function of the n values $X_1^*, X_2^*, \dots, X_n^*$, and a corresponding bootstrap value $\hat{\theta}^* = \theta(\hat{F}^*)$. This procedure is repeated, independently, a large number of times, B , giving bootstrap values $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$. Then $\hat{\sigma}_{BOOT}$, which is the estimated measure of accuracy of $\hat{\theta}$, is approximated by that measure of accuracy of the $\hat{\theta}^*$.

The whole procedure can be stated as the bootstrap algorithm which can be explained as follows-

Let $R = R(X, F)$ be a random variable of interest where $X = (X_1, X_2, \dots, X_n)$ indicates the entire i.i.d. sample $(X_1, X_2, \dots, X_n)$. Having the observed value $X = x$, some aspects of R 's distribution can be estimated as follows-

Step 1 : The empirical distribution function $\hat{F}$ is constructed putting mass $1/n$ at each observation.

$$\hat{F} : \text{mass } \frac{1}{n} \text{ at } x_i, i=1, 2, \dots, n. \quad (4.2)$$

Step 2 : A bootstrap sample is drawn from $\hat{F}$ and R^* is calculated.

$$X_1^*, X_2^*, \dots, X_n^* \stackrel{iid}{\sim} \hat{F}$$

and

$$R^* = R(X^*, \hat{F}) \quad (4.3)$$

Step 3 : Step 2 is repeated, independently, a large number of times, B . The $R^{*1}, R^{*2}, \dots, R^{*B}$ are obtained and the interested aspects of R 's distribution can be calculated.

The above algorithm is usually followed whenever bootstrap distribution and bootstrap estimates of some interested aspects of an estimator are to be calculated. In the following sections, this algorithm is used repeatedly in order to estimate the bias and, standard error of the estimator by the bootstrap method.

4.3 Estimated Bias of an Estimator

Let $(X_1, X_2, \dots, X_n)$ be an i.i.d. random sample of size n from an unknown probability distribution F . Further, let a statistic $\hat{\theta} = \theta(\hat{F}) = \hat{\theta}(X_1, X_2, \dots, X_n)$ be an estimator of $\theta = \theta(F)$ and a random variable $R = R(X, F)$ be $\hat{\theta} - \theta$, where $X = (X_1, X_2, \dots, X_n)$ indicates the entire i.i.d. sample $X_1, X_2, \dots, X_n$. Then the bias of $\hat{\theta}$ is

$$\text{bias}(\hat{\theta}) = E(R) = E(\hat{\theta} - \theta) . \quad (4.4)$$

The calculation procedure of the bootstrap estimate of bias is as follows-

Step 1 : The empirical distribution function is constructed.

$$\hat{F} : \text{mass } \frac{1}{n} \text{ at } x_i, \quad i=1, 2, \dots, n.$$

Step 2 : A bootstrap sample is drawn from $\hat{F}$, and $\hat{\theta}^*$ and R^* are calculated.

$$X_1^*, X_2^*, \dots, X_n^* \stackrel{\text{iid}}{\sim} \hat{F}$$

$$\hat{\theta}^* = \hat{\theta}(X_1^*, X_2^*, \dots, X_n^*) = \theta(\hat{F}^*) \quad (4.5)$$

where $\hat{F}^*$ is the empirical probability distribution of the bootstrap sample:

$$\hat{F}^* : \text{mass } \frac{M_i^*}{n} \text{ on } x_i, \quad i=1, 2, \dots, n \quad (4.6)$$

where M_i^* is the number of times x_i appears in the bootstrap sample. Then

$$R^* = R(X^*, \hat{F}) = \hat{\theta}^* - \hat{\theta} \quad (4.7)$$

Step 3 : Step 2 is independently repeated a large number of times, B . $R^{*1}, R^{*2}, \dots, R^{*B}$ are then obtained and the bootstrap estimate of bias is calculated by

$$\begin{aligned} \hat{\text{BIAS}}_{\text{BOOT}} &= E. R^* \\ &= \frac{1}{B} \sum_{b=1}^B R^{*b} \\ &= \frac{1}{B} \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}) \\ &= \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*b} - \hat{\theta} \\ &= \hat{\theta}^* - \hat{\theta} \end{aligned} \quad (4.8)$$

This algorithm is similarly used repeatedly in order to estimate the standard error by the bootstrap method.

4.4 Estimated Standard Error

Let $(X_1, X_2, \dots, X_n)$ be an i.i.d. random sample of size n from an unknown probability distribution F . Further let a statistic $\hat{\theta} = \theta(\hat{F}) = \hat{\theta}(X_1, X_2, \dots, X_n)$ be an estimator of $\theta(F)$. The standard error of $\hat{\theta}$ is written as

$$se = \sigma(F) .$$

The bootstrap estimate of standard error is $\sigma(\cdot)$ evaluated at $\hat{F}$.

$$\hat{SE}_{BOOT} = \sigma(\hat{F}) . \quad (4.9)$$

If the function $\sigma(F)$ is known the bootstrap estimate of standard error can simply be calculated. Usually the function $\sigma(F)$ cannot be written down explicitly. Thus using a monte-carlo algorithm, the bootstrap estimate of standard error can be calculated as follows-

Step 1 : The empirical distribution function is constructed putting mass $\frac{1}{n}$ at each observation,

$$\hat{F} : \text{mass } \frac{1}{n} \text{ at } x_i, \quad i=1, 2, \dots, n.$$

Step 2 : A bootstrap sample is drawn from $\hat{F}$ and $\hat{\theta}^*$ is calculated.

$$X_1^*, X_2^*, \dots, X_n^* \stackrel{iid}{\sim} \hat{F}$$

and

$$\hat{\theta}^* = \hat{\theta}(X_1^*, X_2^*, \dots, X_n^*) . \quad (4.10)$$

Step 3 : Step 2 is independently repeated B times, obtaining bootstrap replications $\hat{\theta}^1, \hat{\theta}^2, \dots, \hat{\theta}^B$, and $\hat{SE}_{BOOT}$ can be computed as

$$\hat{SE}_{BOOT} = \left(\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^b - \hat{\theta}^{\cdot})^2 \right)^{\frac{1}{2}} \quad (4.11)$$

$$\text{where } \hat{\theta}^{\cdot} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^b .$$

If B approaches ∞ , then $\hat{SE}_{BOOT}$ would exactly equal $\sigma(\hat{F})$. In most cases there is no point in taking B so large for $\hat{SE}_{BOOT}$ to be close to $\sigma(\hat{F})$, since $\sigma(\hat{F})$ itself will be highly variable for estimating the true $se = \sigma(F)$. In practice B is often taken as at most 1,000 times.

4.4.1 Estimated Standard Error of the Sample Mean

Let $(X_1, X_2, \dots, X_n)$ be an i.i.d. random sample of size n from an unknown probability distribution F and further let a statistic $\theta(\hat{F}) = \hat{\theta}(X_1, X_2, \dots, X_n) = \bar{X}$ be an estimator of $\theta(F) = E(X)$.

The variance of $\bar{X}$ is

$$\text{var}(\bar{X}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$

Let $(X_1^*, X_2^*, \dots, X_n^*)$ be an i.i.d. random sample from $\hat{F}$.

The bootstrap sample mean is

$$\bar{X}^* = \frac{1}{n} \sum_{i=1}^n X_i^* .$$

The variance under this sampling scheme is calculated using the sense of step function of $\hat{F}$.

$$\hat{VAR}_B(\bar{X}) = \text{var}(\bar{X}^*)$$

That is,

$$\begin{aligned}\hat{\text{VAR}}_{\text{BOOT}}(\bar{X}) &= E. [\bar{X}^* - E.(\bar{X}^*)]^2 \\ &= E.(\bar{X}^*)^2 - [E.(\bar{X}^*)]^2 .\end{aligned}$$

$$\begin{aligned}E.(\bar{X}^*) &= E. \left(\frac{1}{n} \sum_{i=1}^n X_i^* \right) \\ &= E.(X_i^*) \\ &= \sum_{i=1}^n x_i [\hat{F}(x_i) - \hat{F}(x_{i-1})] \\ &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= \bar{x}, \text{ the sample mean.}\end{aligned}$$

Then

$$\begin{aligned}\text{var.}(\bar{X}^*) &= E. \left(\frac{1}{n} \sum_{i=1}^n X_i^* \right)^2 - \bar{x}^2 \\ &= \frac{1}{n} E.(X_i^*)^2 - \bar{x}^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n x_i^2 - \bar{x}^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{x})^2 .\end{aligned}$$

The standard error is

$$\hat{\text{SE}}_{\text{BOOT}}(\bar{X}) = \left(\frac{1}{n^2} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} . \quad (4.12)$$

It is slightly different from the usual estimated standard error of $\bar{X}$, which is

$$\begin{aligned}\hat{\text{se}}(\bar{X}) &= [\hat{\text{var}}(\bar{X})]^{\frac{1}{2}} \\ &= \left(\frac{1}{n} \hat{\text{var}}(X) \right)^{\frac{1}{2}} \\ &= \left(\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (4.13)\end{aligned}$$

Comparing Equations (4.12) and (4.13),

$$\hat{SE}_{BOOT}(\bar{X}) = \left(\frac{n-1}{n} \right)^{\frac{1}{2}} \hat{SE}(\bar{X})$$

Thus, the bootstrap estimated standard error of $\bar{X}$ is biased down ward.

Generally, if $\hat{\theta}(X_1, X_2, \dots, X_n)$ is an estimator for any parameter θ , the bootstrap estimate of standard error for $\hat{\theta}(X_1, X_2, \dots, X_n)$ is

$$\hat{SE}_{BOOT} = \left(\text{var.}[\hat{\theta}(X_1^*, X_2^*, \dots, X_n^*)] \right)^{\frac{1}{2}}.$$

4.4.2 Estimated Standard Error of the Sample Median

The bootstrap distribution for the sample median can be calculated theoretically, without recourse to monte-carlo methods. Suppose that sample size n is odd, that is, $n=2m-1$. Then the sample median $\hat{\theta}$ equals $x_{(m)}$, the middle order statistic.

The bootstrap sample $(X_1^*, X_2^*, \dots, X_n^*)$ drawn from $\hat{F}$ has bootstrap sample median $\hat{\theta}^* = X_{(m)}^*$, the m^{th} ordered value of the X_i^* . Let

$$M_j^* = \#\{X_i^* = x_{(j)}\}, \quad j=1, 2, \dots, n$$

where $\#$ represents the number.

The event $\{X_{(m)}^* > x_{(k)}\}$ is equivalent to $\left\{ \sum_{j=1}^k M_j^* \leq m-1 \right\}$, so that

$$\begin{aligned} \text{Prob.}\{\hat{\theta}^* > x_{(k)}\} &= \text{Prob.}\left\{ \sum_{j=1}^k M_j^* \leq m-1 \right\} \\ &= \text{Prob.}\left\{ B_i\left(n, \frac{k}{n}\right) \leq m-1 \right\} \end{aligned} \quad (4.14)$$

where

$$\begin{aligned} \text{Prob.}\{B_i(n, \frac{k}{n}) \leq m-1\} &= \sum_{j=0}^{m-1} b_{j,n}(\frac{k}{n}) \\ &= \sum_{j=0}^{m-1} \binom{n}{j} (\frac{k}{n})^j (1 - \frac{k}{n})^{n-j}, \end{aligned}$$

and

$$\begin{aligned} \text{Prob.}\{\hat{\theta}^* \geq x_{(k)}\} &= \text{Prob.}\left\{\sum_{j=1}^{k-1} M_j^* \leq m-1\right\} \\ &= \text{Prob.}\{B_i(n, \frac{k-1}{n}) \leq m-1\} \end{aligned} \quad (4.15)$$

where

$$\begin{aligned} \text{Prob.}\{B_i(n, \frac{k-1}{n}) \leq m-1\} &= \sum_{j=0}^{m-1} b_{j,n}(\frac{k-1}{n}) \\ &= \sum_{j=0}^{m-1} \binom{n}{j} (\frac{k-1}{n})^j (1 - \frac{k-1}{n})^{n-j}. \end{aligned}$$

Therefore the bootstrap distribution of $\hat{\theta}^*$ is concentrated on the values $x_{(1)} < x_{(2)} < \dots < x_{(n)}$, with bootstrap probability $p_{(k)}$ of being equal to $x_{(k)}$,

$$\begin{aligned} p_{(k)} &\equiv \text{Prob.}\{\hat{\theta}^* = x_{(k)}\} \\ &= \text{Prob.}\{\hat{\theta}^* \geq x_{(k)}\} - \text{Prob.}\{\hat{\theta}^* > x_{(k)}\} \\ &= \sum_{j=0}^{m-1} \{b_{j,n}(\frac{k-1}{n}) - b_{j,n}(\frac{k}{n})\}. \end{aligned} \quad (4.16)$$

The bootstrap estimate of standard error is computed by

$$\hat{SE}_{BOOT} = \left(\sum P_{(k)} x_{(k)}^2 - (\sum P_{(k)} x_{(k)})^2 \right)^{\frac{1}{2}}.$$

4.4.3 Estimated Standard Error for Censored Data

Previous results of the bootstrap method are for uncensored data. In this section is concerned with the bootstrap estimate of standard error for censored data.

Censoring is a property of the sample whereas truncation is a property of the distribution. The observed sample $(X_1, X_2, \dots, X_n)$ of size n is arranged in increasing or decreasing order of magnitude and then it can be censored on the right side or on the left side or on both sides. These are called censoring on the right or censoring on the left or double censoring on both sides, respectively. If the lowest r_1 and the highest r_2 of the sample are censored on both sides, only the remaining variables $m = n - (r_1 + r_2)$ are available for estimation purposes.

Sometimes the variate values occurred outside some observable range, the censoring took place at certain fixed points. This is called Type II censoring. Type I censoring is somewhat different from Type II. It is said to occur when a fixed proportion of the sample size n is censored at the lower and/or upper ends of the range of the variable X .

Right censored data is of the form $\{(x_1, d_1), (x_2, d_2), \dots, (x_n, d_n)\}$ where x_j is the j^{th} observation, censored or not, and

$$d_j = \begin{cases} 1 & \text{if } x_j \text{ is uncensored} \\ 0 & \text{if } x_j \text{ is censored.} \end{cases}$$

Let the observed pairs (x_i, d_i) be i.i.d. observations from an unknown distribution F and further let some functional θ based on $\{(x_1, d_1), (x_2, d_2), \dots, (x_n, d_n)\}$ be an estimate of θ . Using computer simulation methods, the bootstrap estimate of standard error, $\hat{SE}_{\text{BOOT}}$, can be calculated as follows-

Step 1 The empirical probability distribution is found by putting mass $1/n$ at each point.

$$\hat{F} : \text{mass } \frac{1}{n} \text{ at each } (x_i, d_i), i=1, 2, \dots, n \quad (4.17)$$

Step 2 A bootstrap sample is drawn, by independent sampling, n times with replacement from $\hat{F}$ and $\hat{\theta}^*$ is calculated.

$$(X_1^*, D_1^*), (X_2^*, D_2^*), \dots, (X_n^*, D_n^*) \stackrel{\text{iid}}{\sim} \hat{F},$$

and

$$\hat{\theta}^* = \theta(\text{data}^*), \quad (4.18)$$

where data^* represent the bootstrap data set.

Step 3 Step 2 is independently repeated a large number of times, B , obtaining $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$ and the $\hat{SE}_{\text{BOOT}}$ is estimated as

$$\hat{SE}_{\text{BOOT}} = \left(\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}^*)^2 \right)^{\frac{1}{2}}. \quad (4.19)$$

The procedure for the bootstrap estimate $\hat{SE}_{\text{BOOT}}$ is the same as for the uncensored case, except that the individual data points are now in pairs (x_i, d_i) .

4.4.4 Estimated Standard Error in Half-Sampling

In sampling theory it is natural to consider stratified situations where the sample space χ is a union of disjoint strata χ_h ,

$$\chi = \bigcup_{h=1}^H \chi_h, \quad h=1, 2, \dots, H.$$

The data consist of separate i.i.d. samples from each stratum,

$$X_{hi} \stackrel{iid}{\sim} F_h, \quad i=1,2,\dots,n_h,$$

where F_h is an unknown probability distribution on X_h .

If it is to assign an estimate of standard error to a functional statistic

$$\hat{\theta} = \theta(\hat{F}) = \theta(\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H), \quad (4.20)$$

the bootstrap estimate of $se(\hat{\theta})$ can be obtained by the following algorithm:

Step 1 The empirical probability distribution function for stratum h , $\hat{F}_h$ ($h=1,2,\dots,H$) is found as

$$\hat{F}_h : \text{mass } \frac{1}{n_h} \text{ at } x_{hi}. \quad (4.21)$$

Step 2 The bootstrap samples are drawn independently from the empirical probability distribution function $\hat{F}_h$ and $\hat{\theta}^*$ is calculated.

$$X_{hi}^* \stackrel{iid}{\sim} \hat{F}_h, \quad i=1,2,\dots,n_h, \quad h=1,2,\dots,H.$$

Let $\hat{F}_h^*$ be the distribution putting mass P_{hi}^* on x_{hi} ,

where

$$P_{hi}^* = \frac{\#\{X_{hi}^* = x_{hi}\}}{n_h}.$$

Then

$$\hat{\theta}^* = \hat{\theta}(\hat{F}_1^*, \hat{F}_2^*, \dots, \hat{F}_H^*). \quad (4.22)$$

Step 3 Step 2 is independently repeated B times, bootstrap replications $\hat{\theta}^{*1}, \hat{\theta}^{*2}, \dots, \hat{\theta}^{*B}$ are obtained and bootstrap estimate of $se(\hat{\theta})$ is estimated by

$$\hat{SE}_{BOOT} = \left(\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*b} - \hat{\theta}^*)^2 \right)^{\frac{1}{2}}. \quad (4.23)$$

Suppose $\hat{\theta}$ is a linear functional statistic

$$\hat{\theta} = \sum_{h=1}^H \left\{ \mu_h + \frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_h(X_{hi}) \right\} \quad (4.24)$$

μ_h and $\alpha_h(\cdot)$ being known, though possibly different for different strata. In this case, without using monte-carlo sampling, the bootstrap estimate of standard error can be obtained by standard theory as follows-

Since

$$\begin{aligned} \hat{\theta}^* &= \hat{\theta}(\hat{F}_1^*, \hat{F}_2^*, \dots, \hat{F}_H^*) \\ &= \sum_{h=1}^H \left\{ \mu_h + \frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_h(X_{hi}^*) \right\}, \\ E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H}(\hat{\theta}^*) &= \sum_{h=1}^H \left\{ \mu_h + \frac{1}{n_h} \sum_{i=1}^{n_h} E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H} \alpha_h(X_{hi}^*) \right\} \\ &= \sum_{h=1}^H \left\{ \mu_h + \frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_h(X_{hi}) \right\} \\ &= \sum_{h=1}^H (\mu_h + \alpha_{h.}) , \end{aligned}$$

$$\text{where } \alpha_{hi} = \alpha_h(X_{hi}), \quad \alpha_{h.} = \frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_{hi}.$$

The bootstrap estimate of variance is,

$$\begin{aligned} \hat{VAR}_{BOOT} &= E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H} \left(\hat{\theta}^* - E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H}(\hat{\theta}^*) \right)^2 \\ &= E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H} \left(\sum_{h=1}^H \left(\frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_h(X_{hi}^*) - \alpha_{h.} \right) \right)^2 \\ &= \sum_{h=1}^H E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H} \left(\frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_h(X_{hi}^*) - \alpha_{h.} \right)^2 \\ &= \sum_{h=1}^H \frac{1}{n_h^2} \sum_{i=1}^{n_h} (\alpha_{hi} - \alpha_{h.})^2 \\ &= \sum_{h=1}^H \frac{\hat{\sigma}_h^2}{n_h} , \end{aligned}$$

$\hat{\sigma}_h^2$ being the h^{th} sample variance and

$$\hat{\sigma}_h^2 = \frac{1}{n_h} \sum_{i=1}^{n_h} (\alpha_{hi} - \alpha_h)^2.$$

The bootstrap estimate of standard error is

$$\hat{SE}_{\text{BOOT}} = \left(\sum_{h=1}^H \frac{\hat{\sigma}_h^2}{n_h} \right)^{\frac{1}{2}}, \quad (4.25)$$

The true standard error is

$$\begin{aligned} se(\hat{\theta}) &= \left(\text{var} \left[\sum_{h=1}^H \left\{ \mu_h + \frac{1}{n_h} \sum_{i=1}^{n_h} \alpha_h(X_{hi}) \right\} \right] \right)^{\frac{1}{2}} \\ &= \left(\sum_{h=1}^H \frac{1}{n_h^2} \sum_{i=1}^{n_h} \text{var} \{ \alpha_h(X_{hi}) \} \right)^{\frac{1}{2}} \\ &= \left(\sum_{h=1}^H \frac{\sigma_h^2}{n_h} \right)^{\frac{1}{2}}, \end{aligned} \quad (4.26)$$

where $\sigma_h^2 = \text{var}[\alpha_h(X_{hi})]$.

The expected value of $\hat{VAR}_{\text{BOOT}}$ equals

$$E_{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_H} \left(\sum_{h=1}^H \frac{\hat{\sigma}_h^2}{n_h} \right) = \frac{n_h-1}{n_h} \sum_{h=1}^H \frac{\sigma_h^2}{n_h} \quad (4.27)$$

If $n_h=2$ for all h , which is the case often considered in half-sampling theory, then

$$E(\hat{VAR}_{\text{BOOT}}) = \frac{1}{2} \text{var}(\hat{\theta}).$$

If one wishes to use the half-sampling form of bootstrap, it might be preferred to half-samples from distributions other than $\hat{F}_h$. Let $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ be the sample mean and variance of $\hat{F}_h$.

$$\begin{aligned} \hat{\mu}_h &= E_{\hat{F}_h}(X_h) \\ &= \frac{1}{n_h} \sum_{i=1}^{n_h} X_{hi} \\ &= \frac{X_{h1} + X_{h2}}{2}. \end{aligned}$$

$$\begin{aligned} E_{\hat{F}_h}(X_h^2) &= \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}^2 \\ &= \frac{x_{h1}^2 + x_{h2}^2}{2} . \end{aligned}$$

$$\begin{aligned} \hat{\sigma}_h^2 &= E_{\hat{F}_h}(X_h^2) - [E_{\hat{F}_h}(X_h)]^2 \\ &= \frac{x_{h1}^2 + x_{h2}^2}{2} - \left(\frac{x_{h1} + x_{h2}}{2} \right)^2 . \\ &= \frac{x_{h1}^2 - 2x_{h1}x_{h2} + x_{h2}^2}{4} . \\ &= \left(\frac{x_{h1} - x_{h2}}{2} \right)^2 \\ &= (x_{h1} - \hat{\mu}_h)^2 . \end{aligned}$$

$\tilde{F}$ is defined by

$$\tilde{F}: \text{mass } \frac{1}{n} \text{ at } \tilde{x} = \frac{x_{hi} - \hat{\mu}_h}{\hat{\sigma}_h}, \quad i=1,2; \quad h=1,2,\dots,H. \quad (4.28)$$

And let $\tilde{F}_h$ be the distribution of $\hat{\mu}_h + \hat{\sigma}_h \tilde{X}$, where $\tilde{X} \sim \tilde{F}$. Then, $\tilde{F}_h$ has the same mean and variance as $\hat{F}_h$, but makes use of information from all strata to estimate the distribution in stratum h . Half-sampling from the $\tilde{F}_h$ (i.e. independently selecting $X_{hi}^* \sim \tilde{F}_h$, $h=1,2,\dots,H$ and computing the standard error of $\hat{\theta}(\hat{F}_1^*, \hat{F}_2^*, \dots, \hat{F}_H^*)$ where $\hat{F}_h^*$ puts all its mass on X_{hi}^*) might be better than half-sampling from the $\hat{F}_h$.

4.5 Confidence Interval for the Median

So far, estimation of bias and standard error of a point estimator $\hat{\theta}$ of a real parameter θ have been discussed in somewhat detail. However, a confidence interval for θ is usually preferable. This section concerns with setting approximate confidence interval for the median in small sample nonparametric situations.

4.5.1 Use of the Binomial Distribution

Let F be a distribution on the real line with median θ , defined as

$$\theta = \inf_t [\text{Prob}\{X \leq t\} = 0.5].$$

Having observed an i.i.d. sample $X_i = x_i$, $i=1, 2, \dots, n$ from F , the exact confidence intervals for θ can be constructed using the order statistic $x_{(1)} < x_{(2)} < \dots < x_{(n)}$.

Let the random variable Z be the number of observations, less than the median. That is

$$Z = \#\{X_i < \theta\} . \quad (4.29)$$

Then Z has a binomial distribution with $p = \frac{1}{2}$, $Z \sim \text{Bi}(n, \frac{1}{2})$. The event $\{x_{(k_1)} < \theta \leq x_{(k_2)}\}$ is the same as the event $\{k_1 \leq Z < k_2\}$. Therefore the probability that θ lies in the interval $(x_{(k_1)}, x_{(k_2)}]$ is

$$\begin{aligned} \text{Prob}\{x_{(k_1)} < \theta \leq x_{(k_2)}\} &= \text{Prob}\{k_1 \leq Z < k_2\} \\ &= \sum_{k=k_1}^{k_2-1} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{n-k} \\ &= \sum_{k=k_1}^{k_2-1} \binom{n}{k} \left(\frac{1}{2}\right)^n \end{aligned} \quad (4.30)$$

For each tail, at a given α level of significance,

$$\begin{aligned}
 \alpha &= \text{Prob}\{\theta \leq x_{(k_1)}\} = \text{Prob}\{Z \leq k_1 - 1\} \\
 &= \text{Prob}\{\text{Bi}(n, \frac{1}{2}) \leq k_1 - 1\} \\
 &= \sum_{k=0}^{k_1-1} \binom{n}{k} \left(\frac{1}{2}\right)^n
 \end{aligned} \tag{4.31}$$

and

$$\begin{aligned}
 \alpha &= \text{Prob}\{\theta > x_{(k_2)}\} = \text{Prob}\{Z \geq k_2\} \\
 &= \text{Prob}\{\text{Bi}(n, \frac{1}{2}) \geq k_2\} \\
 &= \sum_{k=k_2}^n \binom{n}{k} \left(\frac{1}{2}\right)^n .
 \end{aligned} \tag{4.32}$$

The $100(1-2\alpha)\%$ confidence interval for θ is

$$(x_{(k_1)}, x_{(k_2)}) , \tag{4.33}$$

where k_1 and k_2 are the values specified in Equations (4.31) and (4.32).

4.5.2. Use of the Percentile Method

This method provides the approximate confidence intervals to any real-valued parameter $\theta(F)$, based on the bootstrap distribution of $\hat{\theta} = \theta(\hat{F})$. Let

$$\hat{CDF}(t) = \text{Prob}\{\hat{\theta}^* \leq t\}$$

be the cumulative distribution function of the bootstrap distribution of $\hat{\theta}^*$. If the bootstrap distribution is obtained by the monte-carlo method then cumulative distribution function of $\hat{\theta}$ is approximated by

$$\hat{CDF}(t) = \frac{\#\{\hat{\theta}^* \leq t\}}{B} . \tag{4.34}$$

For a given 2α level of significance

$$\hat{CDF}(\hat{\theta}_{low}) = \alpha$$

and

$$\hat{CDF}(\hat{\theta}_{up}) = 1-\alpha.$$

Therefore

$$\hat{\theta}_{low}(\alpha) = \hat{CDF}^{-1}(\alpha) \quad (4.35)$$

and

$$\hat{\theta}_{up}(\alpha) = \hat{CDF}^{-1}(1-\alpha). \quad (4.36)$$

Hence an approximate $100(1-2\alpha)\%$ confidence interval for θ is

$$[\hat{\theta}_{low}(\alpha), \hat{\theta}_{up}(\alpha)]. \quad (4.37)$$

The interval consists of the central $1-2\alpha$ proportion of the bootstrap distribution.

Let $\hat{\theta}$ be the sample median. It is an estimate of $\theta(F)$, which is the median of a distribution F . The bootstrap distribution of $\hat{\theta}^*$ depends on the order statistic $x_{(k)}$, as shown in sub-section(4.4.2). Therefore any interval

$$[\hat{\theta}_{low}, \hat{\theta}_{up}] = [x_{(k_1)}, x_{(k_2)}]$$

is the central $100(1-2\alpha)\%$ confidence interval of the bootstrap distribution with

$$\alpha = p(1)+p(2)+\dots+p(k_1-1)+p(k_1)/2. \quad (4.38)$$

where $p(k_1)$ is as given in Equation(4.16).

In this case the percentile method comes very close to giving the classical binomial intervals (4.30). According to Efron(1980), numerical investigation confirms that the agreement is always very good as long as $\alpha \geq 0.01$. A theoretical reason for this agreement is given by Efron as follows-

For the binomial interval, the α level of significance connected with the event $\{\theta \leq x_{(k_1)}\}$ is

$$\begin{aligned}
 \alpha &= \text{Prob}\{\theta \leq x_{(k_1)}\} \\
 &= \text{Prob}\{z \leq k_1 - 1\} \\
 &= \text{Prob}\{\text{Bi}(n, \frac{1}{2}) \leq k_1 - 1\}
 \end{aligned} \tag{4.39}$$

The percentile method, taking into account of splitting the bootstrap probability at the endpoint, amounts to assigning significance level $\hat{\alpha}$ to $\{\theta \leq x_{(k_1)}\}$ where, from Equation(4.38),

$$\begin{aligned}
 \hat{\alpha} &= \frac{1}{2} [\text{Prob}\{\hat{\theta}^* \leq x_{(k_1)}\} + \text{Prob}\{\hat{\theta}^* < x_{(k_1)}\}] \\
 &= \frac{1}{2} [1 - \text{Prob}\{\hat{\theta}^* > x_{(k_1)}\} + 1 - \text{Prob}\{\hat{\theta}^* \geq x_{(k_1)}\}].
 \end{aligned}$$

Using Equations (4.14) and (4.15), $\hat{\alpha}$ becomes

$$\hat{\alpha} = \frac{1}{2} [\text{Prob}\{\text{Bi}(n, \frac{k_1}{n}) \geq m\} + \text{Prob}\{\text{Bi}(n, \frac{k_1-1}{n}) \geq m\}]. \tag{4.40}$$

C H A P T E R V
BOOTSTRAP ESTIMATION FOR SOME MODELS
OF MEAT PRODUCTION IN MYANMAR

5.1 Introduction

How the use of distribution-free and nonparametric methods of jackknifing and bootstrapping in estimating bias and standard errors of the estimators has been discussed in previous chapters. These methods were applied to parameter estimation procedures concerning a single variable. In most cases of application, relationships between two or more than two variables are of primary concern. Regression models are some such models in which the unknown parameters have to be estimated.

In this chapter, jackknifing and bootstrapping used in the estimation of parameters and in finding the standard errors of these estimators for a linear regression model, are discussed in detail. Further, the procedure of bootstrapping, especially in regression models were illustrated for meat production during the 24-year period, from 1967-68 to 1990-91.

First, the ordinary least squares (OLS) estimates of the parameters in a linear regression model for meat (pork, fowl meat and beef) production and their standard errors are calculated. Then bootstrap estimates are found and are compared with the previous estimates.

5.2 Linear Regression Model

The linear regression model is described as

$$Y = C \beta + \epsilon, \quad (5.1)$$

where Y is an $n \times 1$ vector of observations on the dependent variable; C is a $n \times p$ matrix of observations on p explanatory variables, which are assumed to be fixed; β is a $p \times 1$ vector of unknown parameters; and ϵ is an $n \times 1$ vector of random errors. The OLS estimate of β is

$$\hat{\beta} = G^{-1} C' Y \quad (5.2)$$

where $G = C' C$ and it is assumed that G is of full rank p .

The conventional assumption is that the elements of ϵ are i.i.d. normal variates with zero mean and finite variance.

In this case, $\hat{\beta}$, OLS estimator of β , is also normally distributed. If the errors are not assumed to be normally distributed, the central limit theorem has to be used to say that the distribution of the OLS estimators is asymptotically normal. But for finite samples the normal distribution is only an approximation. In this i.i.d. model, jackknifing and bootstrapping, which are distribution-free methods, can be used.

5.2.1 Jackknife Estimates

Suppose in the model (5.1), i^{th} observation is removed from the equation and let $\hat{\beta}_{(i)}$ be the estimate of $\hat{\beta}$ computed after this removal. Then, $\hat{\beta}_{(i)}$ becomes

$$\hat{\beta}_{(i)} = (C' C - c_i' c_i)^{-1} (C' Y - c_i' Y_i) . \quad (5.3)$$

The inverse of the $p \times p$ matrix $(C' C - c_i' c_i)$ can be derived as follows-

Let

$$A = (C' C - c_i' c_i)$$

Then,

$$(C' C - c_i' c_i) A^{-1} = I$$

$$C' C A^{-1} = I + c_i' c_i A^{-1}$$

$$A^{-1} = G^{-1} + G^{-1} c_i' c_i A^{-1}$$

where

$$G^{-1} = (C' C)^{-1}$$

and,

$$A^{-1} c_i' = G^{-1} c_i' + G^{-1} c_i' c_i A^{-1} c_i' . \quad (5.4)$$

Pre-multiplying Equation(5.4) by the vector c_i gives,

$$c_i A^{-1} c_i' = c_i G^{-1} c_i' + c_i G^{-1} c_i' c_i A^{-1} c_i'$$

$$c_i A^{-1} c_i' - c_i G^{-1} c_i' c_i A^{-1} c_i' = c_i G^{-1} c_i'$$

$$(1 - c_i G^{-1} c_i') c_i A^{-1} c_i' = c_i G^{-1} c_i'$$

$$c_i A^{-1} c_i' = \frac{c_i G^{-1} c_i'}{1 - c_i G^{-1} c_i'} . \quad (5.5)$$

Substituting Equation (5.5) into (5.4) gives

$$A^{-1} c_i' = G^{-1} c_i' + \frac{G^{-1} c_i' c_i G^{-1} c_i'}{1 - c_i G^{-1} c_i'}$$

and,

$$\mathbf{A}^{-1} \mathbf{c}_i' \mathbf{c}_i = \mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i + \frac{\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'}$$

$$\mathbf{A}^{-1} = \mathbf{G}^{-1} + \frac{\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \mathbf{G}^{-1}}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'}$$

Substituting this result in Equation(5.3) gives

$$\begin{aligned} \hat{\beta}_{(i)} &= \left[\mathbf{G}^{-1} + \frac{\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \mathbf{G}^{-1}}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'} \right] (\mathbf{C}' \mathbf{Y} - \mathbf{c}_i' Y_i) \\ &= \mathbf{G}^{-1} \mathbf{C}' \mathbf{Y} + \frac{\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \mathbf{G}^{-1} \mathbf{C}' \mathbf{Y}}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'} - \mathbf{G}^{-1} \mathbf{c}_i' Y_i - \frac{\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' Y_i}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'} \\ &= \hat{\beta} + \frac{1}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'} [\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \hat{\beta} - (1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i') \mathbf{G}^{-1} \mathbf{c}_i' Y_i - \mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' Y_i] \\ &= \hat{\beta} + \frac{1}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'} [\mathbf{G}^{-1} \mathbf{c}_i' \mathbf{c}_i \hat{\beta} - \mathbf{G}^{-1} \mathbf{c}_i' Y_i + \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' \mathbf{G}^{-1} \mathbf{c}_i' Y_i - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' \mathbf{G}^{-1} \mathbf{c}_i' Y_i] \\ &= \hat{\beta} - \frac{\mathbf{G}^{-1} \mathbf{c}_i' (Y_i - \mathbf{c}_i \hat{\beta})}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'} \\ &= \hat{\beta} - \frac{\mathbf{G}^{-1} \mathbf{c}_i' \hat{\epsilon}_i}{1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i'}. \end{aligned}$$

Since

$$\mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' = O(1/n), \quad 1 - \mathbf{c}_i \mathbf{G}^{-1} \mathbf{c}_i' = 1 - O(1/n) \approx 1,$$

and,

$$\hat{\beta}_{(i)} \approx \hat{\beta} - \mathbf{G}^{-1} \mathbf{c}_i' \hat{\epsilon}_i$$

Hence

$$\begin{aligned} \hat{\beta}_{(.)} &= \frac{1}{n} \sum_{i=1}^n \hat{\beta}_{(i)} \\ &\approx \hat{\beta} - \frac{1}{n} \sum_{i=1}^n \mathbf{G}^{-1} \mathbf{c}_i' \hat{\epsilon}_i \\ &= \hat{\beta}, \end{aligned}$$

since $\sum_{i=1}^n \mathbf{G}^{-1} \mathbf{c}_i' \hat{\epsilon}_i = 0.$

Using Equation(3.6), the multivariate version of Tukey's formula for the standard error of $\hat{\beta}$ is

$$\begin{aligned}\hat{\sigma}^2 &= \frac{n-1}{n} \sum_{i=1}^n [\hat{\beta}_{(i)} - \hat{\beta}_{(.)}] [\hat{\beta}_{(i)} - \hat{\beta}_{(.)}]' \\ &\approx \frac{n-1}{n} \sum_{i=1}^n (G^{-1} c_i' \hat{e}_i) (\hat{e}_i c_i G^{-1}) \\ &= \frac{n-1}{n} G^{-1} \left(\sum_{i=1}^n c_i' c_i \hat{e}_i^2 \right) G^{-1} .\end{aligned}\quad (5.6)$$

5.2.2 Bootstrap Estimates

For the model considered in Section(5.2), the OLS estimator of β is

$$\hat{\beta} = G^{-1} C' Y$$

and the $n \times 1$ residual vector, e , is

$$e = Y - C \hat{\beta} . \quad (5.7)$$

In order to calculate bootstrap estimates, first, the n residuals are randomly selected with replacement from the residual vector e and are placed in the $n \times 1$ bootstrap residual vector e^* ,

$$e_1^*, e_2^*, \dots, e_n^* \stackrel{iid}{\sim} \hat{F},$$

where $\hat{F}$ is the empirical distribution function putting mass $1/n$ at each observed residual e_i ,

$$\hat{F}: \text{mass } \frac{1}{n} \text{ at } e_i, \quad i=1, 2, \dots, n,$$

and

$$e^* = (e_1^*, e_2^*, \dots, e_n^*)' . \quad (5.8)$$

Using the OLS estimator $\hat{\beta}$ and the bootstrap residual vector e^* , a new vector of bootstrap dependent variable can be constructed as

$$Y^* = C \hat{\beta} + e^* . \quad (5.9)$$

Using the bootstrap data set Y^* , the parameter vector is re-estimated as

$$\hat{\beta}^* = G^{-1} C' Y^* . \quad (5.10)$$

This procedure is repeated, independently, a large number B of times, obtaining bootstrap estimates $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*B}$. These vectors are used to construct the joint probability distribution of the OLS estimator $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$. The mean value of the bootstrap distribution of the OLS estimator $\hat{\beta}$ is calculated over the full set of $\hat{\beta}^{*b}$ which is

$$\hat{\beta}^{\cdot} = \frac{1}{B} \sum_{b=1}^B \hat{\beta}^{*b} . \quad (5.11)$$

The bootstrap estimate of covariance matrix is

$$\hat{COV} = \frac{1}{B-1} \sum_{b=1}^B (\hat{\beta}^{*b} - \hat{\beta}^{\cdot}) (\hat{\beta}^{*b} - \hat{\beta}^{\cdot})' . \quad (5.12)$$

In this approach, the variance within each of the bootstrap sample selected from an original sample is used to estimate the variance of the population from which the sample has been drawn.

5.3 Application of Bootstrapping to Some Models for Meat Production

In linear regression model, the errors are assumed to be i.i.d. normal variates in order to find the distributions of the OLS estimators. This assumption cannot be justified in some practical cases. In such cases, the bootstrap method can supply the approximate distributions of the OLS estimators. In this section, it will be shown that the bootstrap approximation to the distribution of the OLS estimator is valid, that is, the bootstrap estimates of parameters and their standard errors are very close to the OLS estimates and their standard errors.

This will be shown by considering some models for meat production in Myanmar. It is assumed that meat (pork, fowl meat and beef) production depends on the number of heads (pigs, fowls and cattles) and population in Myanmar. Yearly meat production during the 24 years from 1967-68 to 1990-91 is used in finding the suitable models. The model considered for meat production is

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 P_t + \beta_5 P_{t-1} + \beta_6 P_{t-2} + \epsilon_t \quad (5.13)$$

where Y_t represents the meat (pork, fowl meat and beef) production measured in thousand of viss in year t ; X_t , the number of heads (pigs, fowls and cattles) measured in thousand in year t ; P_t , population measured in thousand in year t ; and ϵ_t , the random error for year t .

The suitable model is selected by using the stepwise, forward and backward procedures supplied in SPSS/PC+ version 4.0 program.

5.3.1 Selected Models for Meat Production

Pork Production Model

The suitable model for pork production is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (5.14)$$

and OLS estimation gives

$$\hat{Y}_i = \begin{matrix} -1555.6 \\ (583.3) \end{matrix} + \begin{matrix} 11.020 \\ (0.264) \end{matrix} X_i \quad (5.15)$$

where the values in the parentheses are standard errors of the estimated parameters.

Fowl Meat Production Model

The suitable model for fowl meat production is

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad (5.16)$$

and OLS estimation gives

$$\hat{Y}_i = \begin{matrix} -6424.0 \\ (1999.7) \end{matrix} + \begin{matrix} 1.878 \\ (0.073) \end{matrix} X_i \quad (5.17)$$

Beef Production Model

The suitable model for beef production is

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 P_i + \epsilon_i \quad (5.18)$$

and OLS estimation gives

$$\hat{Y}_i = \begin{matrix} -1934.1 \\ (408.4) \end{matrix} + \begin{matrix} 1.215 \\ (0.181) \end{matrix} X_i + \begin{matrix} 0.341 \\ (0.049) \end{matrix} P_i \quad (5.19)$$

5.3.2 Bootstrap Estimates and Their Distributions

The bootstrap distributions of the OLS estimators given in this section are based on 1,000 bootstrap samples, that is, $B=1000$. An estimate of the parameter vector β is obtained from each of the bootstrap samples. The bootstrap estimate of β and the corresponding covariance matrix are calculated from the entire set of estimator, $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*B}$.

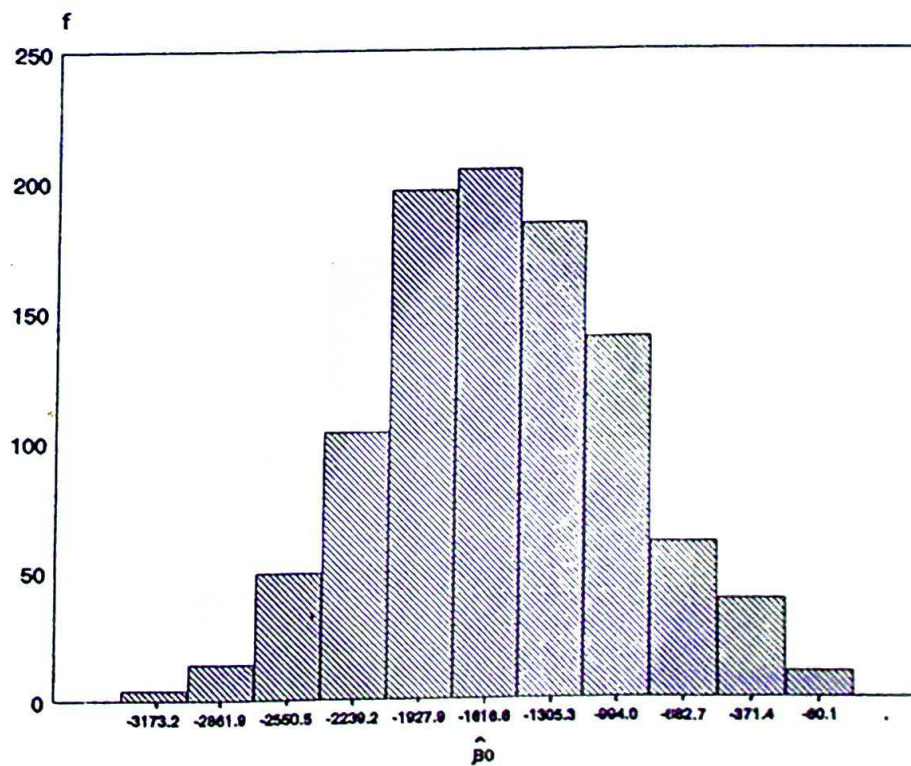
The bootstrap distributions of parameter estimates, $\hat{\beta}$, are found by using the econometric computer program SHAZAM version 5.1. The results are as follows.

Pork Production Model

Table(5.1)
The bootstrap distribution of $\hat{\beta}_0$

Class No.	$\hat{\beta}_0$	f
1	(-3328.8)-(-3017.5)	4
2	(-3017.5)-(-2706.2)	14
3	(-2706.2)-(-2394.9)	49
4	(-2394.9)-(-2083.6)	103
5	(-2083.6)-(-1772.3)	196
6	(-1772.3)-(-1461.0)	204
7	(-1461.0)-(-1149.7)	183
8	(-1149.7)-(-838.4)	139
9	(-838.4)-(-527.1)	60
10	(-527.1)-(-215.8)	38
11	(-215.8)-(-95.5)	10
Total		1000

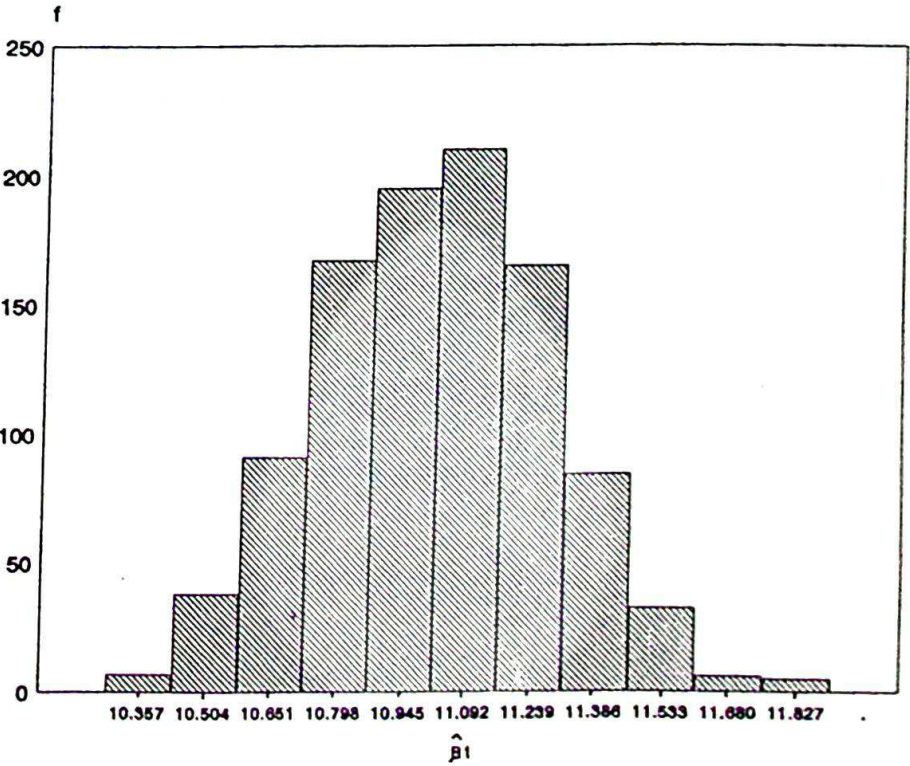
Figure(5.1)
Histogram for the bootstrap distribution of $\hat{\beta}_0$



Table(5.2)
The bootstrap distribution of $\hat{\beta}_1$

Class No.	$\hat{\beta}_1$	f
1	10.283 - 10.430	7
2	10.430 - 10.577	38
3	10.577 - 10.724	91
4	10.724 - 10.871	167
5	10.871 - 11.018	195
6	11.018 - 11.165	210
7	11.165 - 11.312	165
8	11.312 - 11.459	84
9	11.459 - 11.606	32
10	11.606 - 11.753	6
11	11.753 - 11.900	5
Total		1000

Figure(5.2)
Histogram for the bootstrap distribution of $\hat{\beta}_1$



The bootstrap estimation gives

$$\hat{Y}_i = \begin{matrix} -1547.0 \\ (570.2) \end{matrix} + \begin{matrix} 11.016 \\ (0.263) \end{matrix} X_i \quad (5.20)$$

The OLS estimates from Equation(5.15) and bootstrap estimates from Equation(5.20) are compared in Table(5.3).

Table(5.3)
OLS estimates and bootstrap estimates

Parameter	β_0	β_1
<u>OLS Estimation</u>		
Estimate	-1555.6	11.020
Standard error	(583.3)	(0.264)
<u>Bootstrap Estimation</u>		
Estimate	-1547.0	11.016
Standard error	(570.2)	(0.263)

Table(5.4)
Relative errors of bootstrap estimates
for different bootstrap sample sizes(N)

N	$\hat{\beta}_0^1$	Relative ² error(%)	$\hat{\beta}_1^1$	Relative error(%)
100	-1625.6	4.502	11.062	0.377
200	-1568.7	0.840	11.033	0.114
300	-1540.8	0.948	11.019	0.005
400	-1536.6	1.219	11.015	0.043
500	-1530.7	1.598	11.011	0.078
600	-1550.1	0.353	11.022	0.022
700	-1551.1	0.287	11.021	0.012
800	-1549.6	0.386	11.020	0.001
900	-1544.1	0.741	11.017	0.030
1000	-1547.0	0.553	11.016	0.032

If the bootstrap estimates differ by less than 5 per cent from the OLS estimates, it can be said that the bootstrap estimates are very close to the OLS estimates. (Prescott and Stengos(1987)) When $N=1000$, for which the bootstrap distributions were found, the bootstrap estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ differ by 0.55 per cent and 0.03 per cent from the OLS estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively. Even if $N=100$, bootstrap estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$ differ 4.50 per cent and 0.38 per cent respectively. Thus, it can be said that, by the above criterion, the bootstrap estimates are quite close to the OLS estimates for the model considered.

¹ $\hat{\beta}_i$ is the mean calculated over N values of OLS estimate $\hat{\beta}_i$.

² Relative error = $\frac{\text{bootstrap estimate} - \text{OLS estimate}}{\text{OLS estimate}} \times 100$

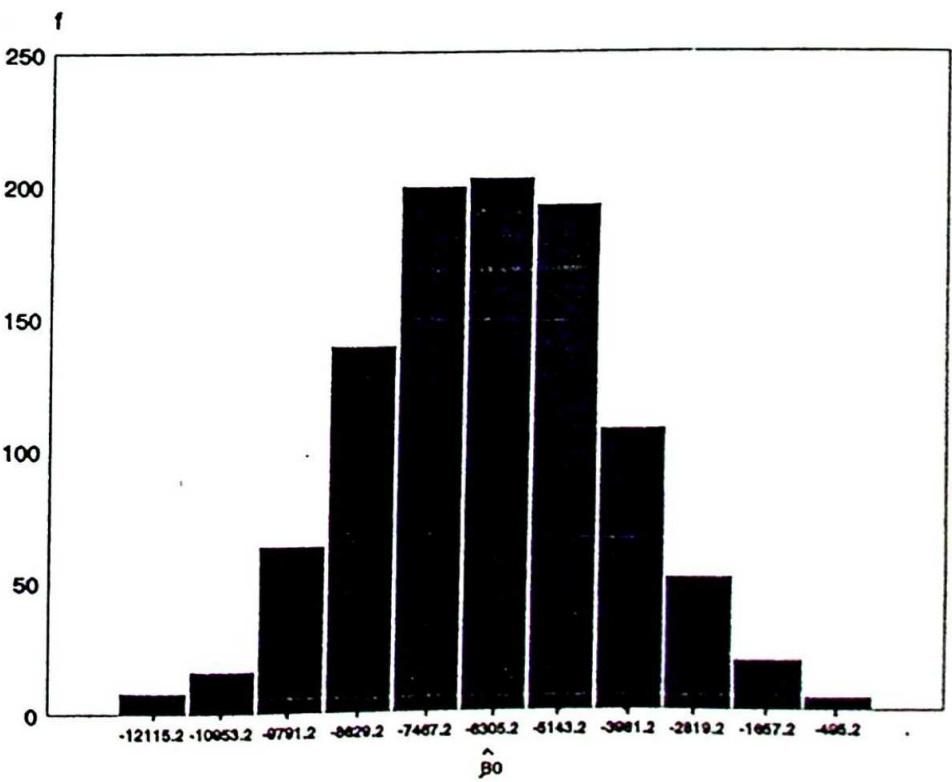
Relative errors are computed from the bootstrap estimates and the OLS estimates with several decimal points. The values shown in the table are the rounded values with 3 decimal points.

Fowl Meat Production Model

Table(5.5)
The bootstrap distribution of $\hat{\beta}_0$

Class No.	$\hat{\beta}_0$	f
1	(-12696.2) - (-11534.2)	8
2	(-11534.2) - (-10372.2)	16
3	(-10372.2) - (-9210.2)	63
4	(-9210.2) - (-8048.2)	139
5	(-8048.2) - (-6886.2)	199
6	(-6886.2) - (-5724.2)	202
7	(-5724.2) - (-4562.2)	192
8	(-4562.2) - (-3400.2)	107
9	(-3400.2) - (-2238.2)	50
10	(-2238.2) - (-1076.2)	19
11	(-1076.2) - (85.8)	5
Total		1000

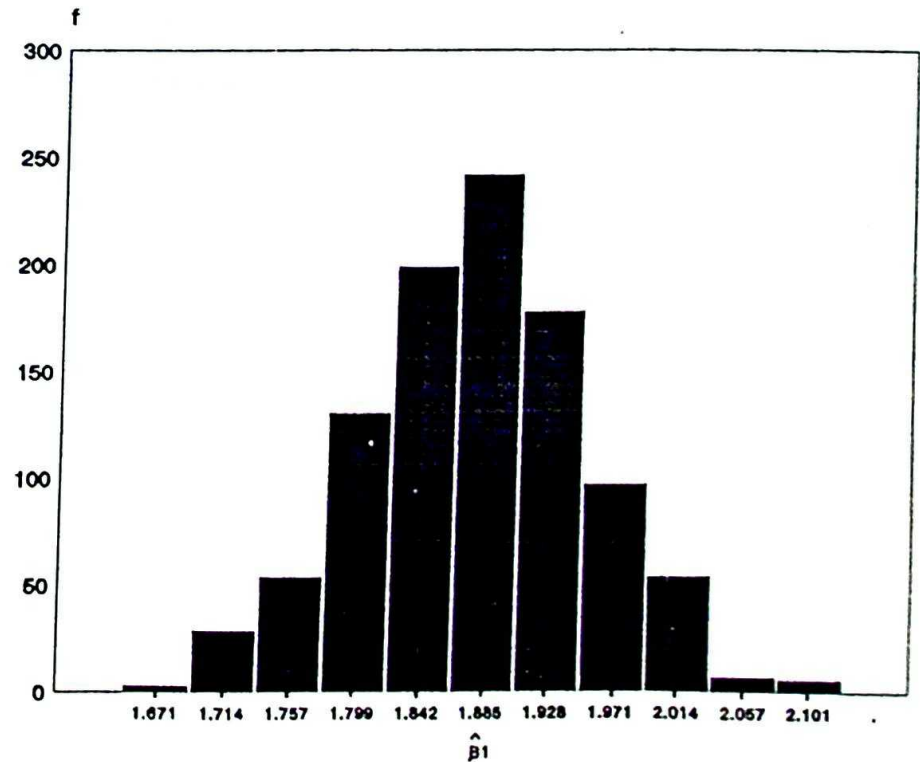
Figure(5.3)
Histogram for the bootstrap distribution of $\hat{\beta}_0$



Table(5.6)
The bootstrap distribution of $\hat{\beta}_1$

Class No.	$\hat{\beta}_1$	f
1	1.649 - 1.692	3
2	1.692 - 1.735	29
3	1.735 - 1.778	54
4	1.778 - 1.821	131
5	1.821 - 1.864	199
6	1.864 - 1.907	242
7	1.907 - 1.950	178
8	1.950 - 1.993	97
9	1.993 - 2.036	54
10	2.036 - 2.079	7
11	2.079 - 2.122	6
Total		1000

Figure(5.4)
Histogram for the bootstrap distribution of $\hat{\beta}_1$



The bootstrap estimation gives

$$\hat{Y}_i = \begin{matrix} -6426.9 \\ (2071.8) \end{matrix} + \begin{matrix} 1.878 \\ (0.075) \end{matrix} X_i \quad (5.21)$$

The OLS estimates from Equation(5.17) and bootstrap estimates from Equation(5.21) are compared in Table(5.7).

Table(5.7)
OLS estimates and bootstrap estimates

Parameter	β_0	β_1
<u>OLS Estimation</u>		
Estimate	-6424.0	1.878
Standard error	(1999.7)	(0.073)
<u>Bootstrap Estimation</u>		
Estimate	-6426.9	1.878
Standard error	(2071.8)	(0.075)

Table(5.8)

Relative errors of bootstrap estimates
for different bootstrap sample sizes(N)

N	$\hat{\beta}_0$	Relative error(%)	$\hat{\beta}_1$	Relative error(%)
100	-6278.6	2.264	1.875	0.173
200	-6300.7	1.920	1.875	0.162
300	-6338.6	1.330	1.877	0.064
400	-6323.6	1.564	1.875	0.159
500	-6427.1	0.048	1.878	0.004
600	-6455.0	0.482	1.879	0.069
700	-6414.4	0.150	1.877	0.043
800	-6437.7	0.213	1.878	0.006
900	-6431.9	0.122	1.878	0.005
1000	-6426.9	0.045	1.878	0.016

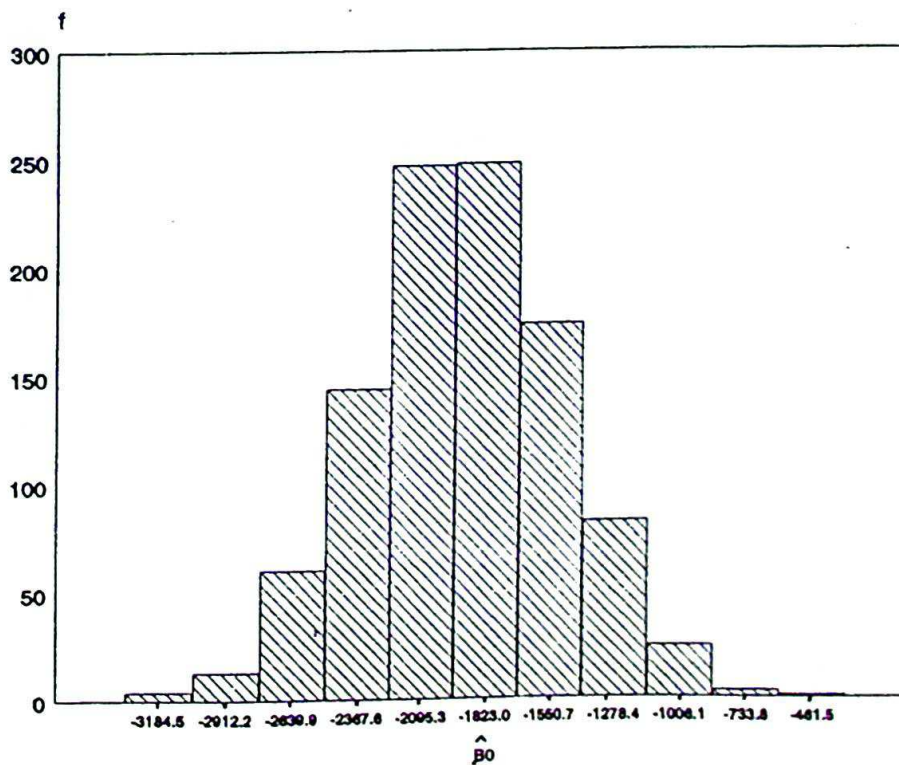
From Table(5.8) it can be seen that the bootstrap estimates are quite close to the OLS estimates and can be used with confidence for the model considered.

Beef Production Model

Table(5.9)
The bootstrap distribution of $\hat{\beta}_0$

Class No.	$\hat{\beta}_0$	f
1	(-3320.6) - (-3048.3)	4
2	(-3048.3) - (-2776.0)	13
3	(-2776.0) - (-2503.7)	60
4	(-2503.7) - (-2231.4)	144
5	(-2231.4) - (-1959.1)	247
6	(-1959.1) - (-1686.8)	248
7	(-1686.8) - (-1414.5)	174
8	(-1414.5) - (-1142.2)	82
9	(-1142.2) - (-869.9)	24
10	(-869.9) - (-597.6)	3
11	(-597.6) - (-325.3)	1
Total		1000

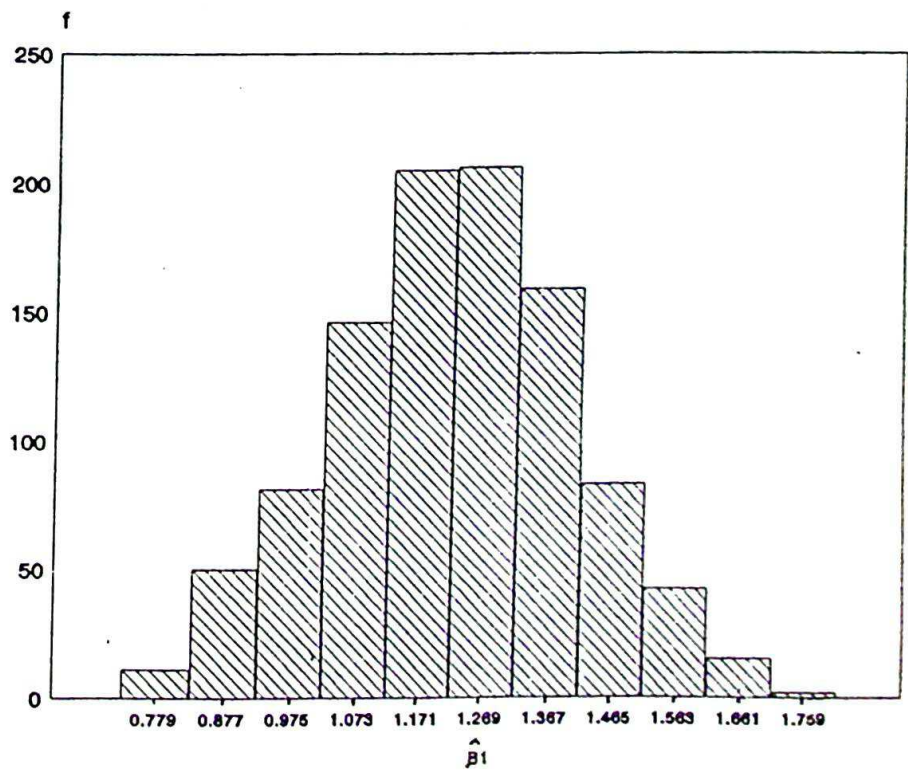
Figure(5.5)
Histogram for the bootstrap distribution of $\hat{\beta}_0$



Table(5.10)
The bootstrap distribution of $\hat{\beta}_1$

Class No.	$\hat{\beta}_1$	f
1	0.730 - 0.828	11
2	0.828 - 0.926	50
3	0.926 - 1.024	81
4	1.024 - 1.122	146
5	1.122 - 1.220	205
6	1.220 - 1.318	206
7	1.318 - 1.416	159
8	1.416 - 1.514	83
9	1.514 - 1.612	42
10	1.612 - 1.710	15
11	1.710 - 1.808	2
Total		1000

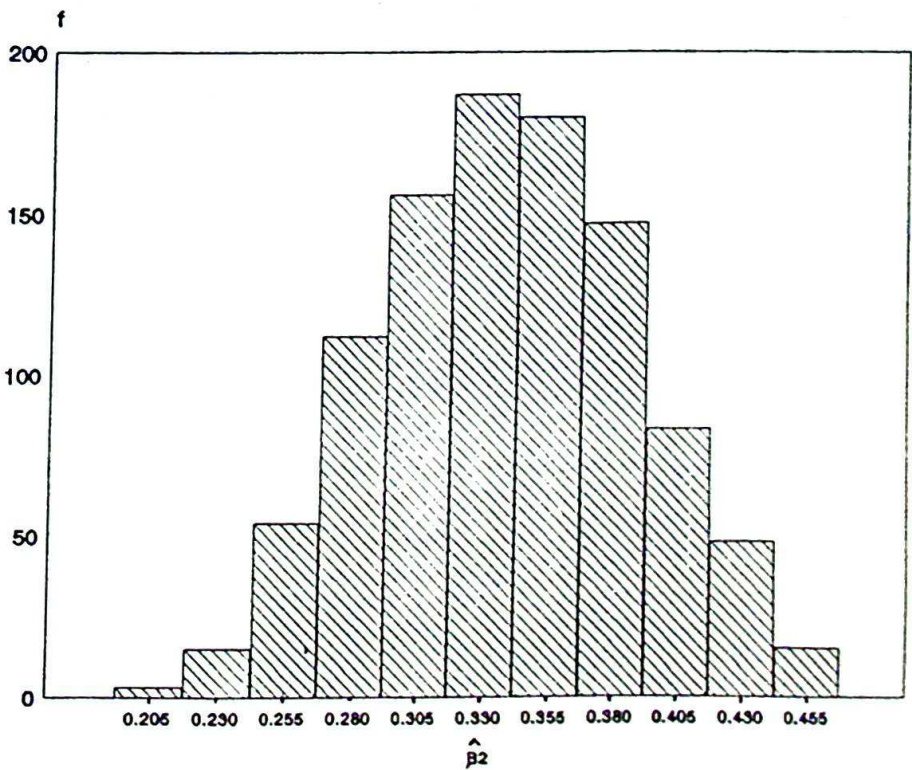
Figure(5.6)
Histogram for the bootstrap distribution of $\hat{\beta}_1$



Table(5.11)
The bootstrap distribution of $\hat{\beta}_2$

Class No.	$\hat{\beta}_2$	f
1	0.192 - 0.217	3
2	0.217 - 0.242	15
3	0.242 - 0.267	54
4	0.267 - 0.292	112
5	0.292 - 0.317	156
6	0.317 - 0.342	187
7	0.342 - 0.367	180
8	0.367 - 0.392	147
9	0.392 - 0.417	83
10	0.417 - 0.442	48
11	0.442 - 0.467	15
Total		1000

Figure(5.7)
Histogram for the bootstrap distribution of $\hat{\beta}_2$



The bootstrap estimation gives

$$\hat{Y}_i = -1924.0 + 1.221 X_i + 0.339 P_i \quad (5.22)$$

(404.4)
(0.183)
(0.049)

The OLS estimates from Equation (5.19) and bootstrap estimates from Equation (5.22) are compared in Table (5.12).

Table(5.12)
OLS estimates and bootstrap estimates

Parameter	β_0	β_1	β_2
<u>OLS Estimation</u>			
Estimate	-1934.1	1.215	0.341
Standard error	(408.4)	(0.181)	(0.049)
<u>Bootstrap Estimation</u>			
Estimate	-1924.0	1.221	0.339
Standard error	(404.4)	(0.183)	(0.049)

Table(5.13)
 Relative errors of bootstrap estimates
 for different bootstrap sample sizes(N)

N	$\hat{\beta}_0$	Relative error(%)	$\hat{\beta}_1$	Relative error(%)	$\hat{\beta}_2$	Relative error(%)
100	-1925.6	0.442	1.221	0.492	0.339	0.588
200	-1941.6	0.389	1.226	0.917	0.338	0.849
300	-1937.1	0.157	1.228	1.046	0.338	1.015
400	-1918.3	0.817	1.226	0.927	0.337	1.072
500	-1909.4	1.277	1.229	1.189	0.336	1.378
600	-1909.8	1.255	1.233	1.447	0.336	1.609
700	-1915.2	0.976	1.226	0.903	0.337	1.073
800	-1920.3	0.715	1.224	0.763	0.338	0.901
900	-1918.9	0.786	1.224	0.700	0.338	0.861
1000	-1924.0	0.524	1.221	0.532	0.339	0.668

It can be seen, from the results shown in Table(5.13), that bootstrap estimates are reliable estimates which can replace the OLS estimates for the model under consideration.

The OLS estimation assumes that the random errors are i.i.d., so does the bootstrap estimation. A consistency check on the bootstrap method is provided by the differences between the OLS estimates and the bootstrap parameter estimates. If the random errors are i.i.d. variates, then these differences should be small.

In the meat production models, the bootstrap estimates are found to be very close to the original OLS estimates as has been expected. More precisely, the bootstrap estimates differ by less than 0.7 per cent from the original OLS estimates. This evidence supports that the bootstrap estimation method is valid to be used when the errors are i.i.d.. Therefore, it is concluded that the bootstrap estimation method can be used to find the approximate distribution of the OLS estimator when the errors can only be assumed to be i.i.d. and normality assumption cannot be justified.

C O N C L U S I O N

In the point estimation of a parameter, unbiasedness is one of the properties of a good estimator. But an unbiased estimator cannot always be obtained. Then a biased estimator has to be used and it will be better if the bias of this estimator can be reduced.

Quenouille's jackknife is a bias reduction method. It is based upon the assumption of the existence of a certain series expansion for the bias of an estimator. Since this type of expansion does exist for a large class of estimators, Quenouille's method is widely applicable. For a variety of cases, it was found that the first-order jackknife estimator eliminates the order n^{-1} term in the bias of an estimator. Thus, to eliminate the order n^{-2} term in the bias, the second-order jackknife can be used.

Another bias reduction method is the generalized jackknife, which is a new family of jackknife based on the assumption that the biases of estimators can be factorized. In so doing, the method can handle more general form of bias. In applying this method to the estimation of a truncation point, it was found that the k^{th} order generalized jackknife eliminates the order n^{-k} term in the bias. It was also found that although the first-order jackknife is biased, the first-

order generalized jackknife is unbiased in the parameter estimation for uniform distributions. Therefore it can be concluded that the generalized jackknife is a method to be used if the bias of different orders is to be reduced simultaneously.

Although Quenouille's jackknife was first introduced for the purpose of reduction bias, it can be used for estimating bias and standard error of an estimator in nonparametric estimation. It was found that Quenouille's jackknife method can supply an unbiased estimator of the estimated bias of a variance estimator. But whenever the jackknife estimator is unbiased, the estimated bias of the original estimator by the jackknife method is also unbiased. It may not be true in general. This is an area to be investigated further by researchers. Next, the method is found to be able to give an estimated standard error of the sample variance which agrees nicely with the true standard error of the estimator.

Another method which can be used for nonparametric estimation of bias and standard error is the bootstrap method. Besides the calculation of approximated biases and standard errors, it can be used to find the distributions of the nonparametric estimators. Thus, bootstrap method is more useful than the jackknife method. For this method, it was found that in estimating the standard error of the sample mean, the estimated standard error is biased downward.

In order to show the use of this method in linear regression model, the method was applied to find the coefficients of meat production models for Myanmar. It was found that the method can successfully be applied to the linear regression model. The bootstrap distributions of the OLS estimators so obtained were found to be concentrated closely around the estimators without the normality assumption. The relative errors so obtained are at most 4.5 per cent and for the bootstrap sample size of 1000, the largest relative error is 0.67 per cent. It can be as small as 0.02 per cent.

The method has the attraction of being a distribution-free method. But it is computer-intensive. If computation is easily assessable, it may be worth applying the method to more complicated models for which the application of the method would require some programming on the part of the researcher.

၆၂၅၃၅
စာကြည့်တိုက်
စီးပွားရေးပစ္စည်း
ရန်ကုန်

A P P E N D I X

Appendix Table (1)
Production of Meat

(viss in thousand)

Sr.	Year	Pork	Fowl Meat	Beef
1	1967-68	11910	23782	15236
2	1968-69	11850	23717	15342
3	1969-70	13980	25742	15703
4	1970-71	15090	29298	15956
5	1971-72	15180	30088	16304
6	1972-73	14070	30002	16289
7	1973-74	15870	30186	17324
8	1974-75	14610	27833	16988
9	1975-76	16200	29709	17244
10	1976-77	18270	31494	17482
11	1977-78	20298	31262	17818
12	1978-79	21063	32430	18452
13	1979-80	20337	36668	19424
14	1980-81	22790	40641	19924
15	1981-82	26632	48505	20711
16	1982-83	29509	54941	21335
17	1983-84	28081	58335	21699
18	1984-85	27828	62070	22034
19	1985-86	33403	65730	22647
20	1986-87	31327	66098	22661
21	1987-88	31861	71542	23063
22	1988-89	33058	76607	23463
23	1989-90	25846	50227	23207
24	1990-91	23442	47923	28684

Source: (1) Report to the Pyithu Hluttaw on the Financial, Economic and Social Conditions of the Socialist Republic of the Union of Burma. (1979-1988)
(2) Review of the Financial, Economic and Social Conditions of the Union of Myanmar. (1989-1991)

Appendix Table (2)
Number of Heads

(in thousand)

Sr.	Year	Pig	Fowl	Cattle
1	1967-68	1175	14947	6558
2	1968-69	1259	15969	6694
3	1969-70	1478	17694	6833
4	1970-71	1605	19116	6993
5	1971-72	1604	19436	7158
6	1972-73	1489	19604	7235
7	1973-74	1461	18868	7267
8	1974-75	1432	18132	7299
9	1975-76	1578	19323	7405
10	1976-77	1780	20151	7526
11	1977-78	1955	20390	7694
12	1978-79	1985	21039	7932
13	1979-80	1961	23915	8307
14	1980-81	2196	26577	8531
15	1981-82	2631	31793	8857
16	1982-83	2884	34405	9147
17	1983-84	2734	36701	9338
18	1984-85	2706	38796	9502
19	1985-86	2953	38791	9718
20	1986-87	2986	38168	9758
21	1987-88	3059	39551	9919
22	1988-89	3199	40169	10091
23	1989-90	2449	27789	9126
24	1990-91	2243	26422	9298

Source: (1) Report to the Pyithu Hluttaw on the Financial, Economic and Social Conditions of the Socialist Republic of the Union of Burma. (1979-1988)
(2) Review of the Financial, Economic and Social Conditions of the Union of Myanmar.(1989-1991)

Appendix Table (3)
Population in Myanmar

(in thousand)		
Sr.	Year	Population
1	1967-68	25867
2	1968-69	26444
3	1969-70	27034
4	1970-71	27637
5	1971-72	28262
6	1972-73	28886
7	1973-74	29521
8	1974-75	29778
9	1975-76	30389
10	1976-77	31009
11	1977-78	31642
12	1978-79	32284
13	1979-80	32939
14	1980-81	33608
15	1981-82	34287
16	1982-83	34976
17	1983-84	35663
18	1984-85	36361
19	1985-86	37073
20	1986-87	37800
21	1987-88	38541
22	1988-89	39297
23	1989-90	40034
24	1990-91	40786

Source: (1) Report to the Pyithu Hluttaw on the Financial, Economic and Social Conditions of the Socialist Republic of the Union of Burma. (1979-1988)
(2) Review of the Financial, Economic and Social Conditions of the Union of Myanmar. (1989-1991)

REFERENCES

1. COCHRAN, W.G.
1977 Sampling Techniques, 3rd edition,
John Wiley, New York.
2. EFRON, B.
1981 Censored Data and the Bootstrap, **J. Amer. Statist. Assoc.**, 76, pp.312-319.
3. EFRON, B.
1982 The Jackknife, the Bootstrap and Other Resampling Plans, SIMA, Philadelphia.
4. FREEDMAN, D.A.
1981 Bootstrapping Regression Models, **Ann. Statist.**, 9, pp.1218-1228.
5. FREEDMAN, D.A., AND PETERS, S.C.
1984 Bootstrapping a Regression Equation: Some Empirical Results, **J. Amer. Statist. Assoc.**, 79, pp.97-106.
6. GIBBONS, J.D.
1971 Nonparametric Statistical Inference,
McGraw-Hill, New York.
7. GRAY, H.L., AND SCHUCANY, W.R.
1972 The Generalized Jackknife Statistic,
Marcel Dekker, New York.
8. HORST, P.
1963 Matrix Algebra for Social Scientists,
Holt, Rinehart and Winston, New York.

9. KENDALL, M.G., AND STUART, A.
1979 The Advanced Theory of Statistics, vol.2,
 4th edition, Griffin, London.
10. MILLER, R.G.
1974a The Jackknife - A Review, *Biometrika*, 61,
 pp.1-17.
11. MILLER, R.G.
1974b An Unbalanced Jackknife, *Ann. Statist.*, 2,
 pp.880-891.
12. MOOD, A.M., GRAYBILL, F.A., AND BOES, D.C.
1974 Introduction to the Theory of Statistics,
 3rd edition, McGraw-Hill International, Tokyo.
13. PRESCOTT, D.M., AND STENGOS, T.
1987 Bootstrapping Confidence Intervals: An
 Application to Forecasting the Supply of Pork,
 Amer. J. Agr. Econ., 69, pp.266-273.
14. QUENOUILLE, M.
1949 Approximate Tests of Correlation in Time
 Series, *J. Roy. Statist. Soc. Ser. B.*, 11,
 pp.18-84
15. QUENOUILLE, M.
1956 Notes on Bias in Estimation, *Biometrika*, 43,
 pp.353-360
16. ROBSON, D.S., AND WHITLOCK, J. H.
1964 Estimation of a Truncation Point, *Biometrika*,
 51, pp.33-40.

17. SCHUCANY, W.R., GRAY, H.L., AND OWEN, D.B.
1971 On Bias Reduction in Estimation, *J. Amer. Statist. Assoc.*, 66, pp.524-533.
18. SILVERMAN, B.W.
1986 Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.
19. SINGH, K.
1981 On the Asymptotic Accuracy of Efron's Bootstrap, *Ann. Statist.*, 9, pp.1187-1195.
20. TIKU, M.L., TAN, W.Y., AND BALAKRISHNAN, N.
1986 Robust Inference, Marcel Dekker, New York.
21. TUKEY, J.
1958 Bias and Confidence in Not Quite Large Samples, Abstract, *Ann. Math. Statist.*, 29, p.614.