1-1-2003

Estimation of the population mean using two extremes in order statistics

Vladimir Emil Minev
University of Nevada, Las Vegas

ESTIMATION OF THE POPULATION MEAN USING TWO EXTREMES IN ORDER STATISTICS

by

Vladimir E. Minev

Bachelor of Science
University of Nevada, Las Vegas
1999

A thesis submitted in partial fulfillment
of the requirements for the

Master of Science Degree in Mathematical Sciences
Department of Mathematical Sciences
College of Sciences

Graduate College
University of Nevada, Las Vegas
May 2003

UMI Number: 1414537

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

UNLV
UNIVERSITY OF NEVADA LAS VEGAS

Thesis Approval
The Graduate College
University of Nevada, Las Vegas

APRIL 16, 2003

The Thesis prepared by

Vladimir E. Minev

Entitled

ESTIMATION OF THE POPULATION MEAN USING TWO EXTREMES IN ORDER STATISTICS

is approved in partial fulfillment of the requirements for the degree of

MS MATHEMATICS

Examination Committee Chair

Dean of the Graduate College

Examination Committee Member

Examination Committee Member

Graduate College Faculty Representative

PR/1017-53/1-00

ABSTRACT

An Estimation of the Popuiation Mean
Using Two Extremes in Order Statistics

by

Vladimir E. Minev

Professor Hokwon Cho,Ph.D. Examination Committee Chair
Assistant Professor of Mathematics
University of Nevada, Las Vegas

For most non-parametric statistical inference, the sampling theory of order statistics has been playing a fundamental role because of that properties of the range and the average of the smallest and largest order statistics are useful for estimate the population parameters of both large and small samples.

The applications of the method using order statistics in a given sample to estimate the parametric values appear quite often in the literature. For a certain data set, such as stock market data, the method we are considering may have an advantage in estimating the mean due to the fact that the stock data have a fairly large amount of observations during a given period, even a day.

In addition to estimating the mean, it is of interest to compute $(1-\alpha)$ 100% confidence limits as well. Using the two extremes, $X_{(1)}$ and $X_{(n)}$, we wish to construct a confidence interval for the population mean μ.

TABLE OF CONTENTS

ABSTRACT iii

LIST OF TABLES v

LIST OF FIGURES vi

ACKNOWLEDGEMENTS vii

CHAPTER 1 INTRODUCTION 1
1.1 Terminology 2
1.2 Formulation of the Problem 3

CHAPTER 2 RANDOM VARIABLES OF TWO EXTREMES IN ORDER STATISTICS 5
2.1 Variable with Largest Mean 5
2.2 Variable with Smallest Mean 8
2.3 Variable with Smallest Variance 10
2.4 Variable with Largest Variance 14
2.5 Variable with Average Variance 18
2.6 Uniform Random Variable 20
2.7 Simulated Normal Variable and Derivation of Confidence Interval 22

CHAPTER 3 SIMULATION STUDIES 27
3.1 Numerical Simulation and Testing of Confidence Interval 27
3.2 Application in the Stock Market 28
3.3 Application to Environmental Data 31

CHAPTER 4 CONCLUSION 34

APPENDICES 35
Appendix I 35
Appendix II 40
Appendix III 41

BIBLIOGRAPHY 47

VITA 48

LIST OF TABLES

Table 1.3 Probability of One Success in *n* Trials on *n* Sided Die 3

Table 3.1.1 Probability for Sample Size *n*=45,90% and 95% C.I 27

Table 3.1.2 Probability for Sample Size *n*=100,90% and 95% C.I 28

Table 3.2 DJIA Transactions April 1991-April 2001 29-30

Table 3.3.1 Log-Normal Data 31

Table 3.3.2 Classical and Log-Normal Estimates of the UB 33

LIST OF FIGURES

Figure 1.1 Two DJIA Samples with Same Extremes 2

Figure 2.1 Ordered Sample of the Variable with Largest Mean 7

Figure 2.2 Ordered Sample of the Variable with Smallest Mean 10

Figure 2.3 Ordered Sample of the Variable with Smallest Variance 14

Figure 2.4 Ordered Sample of the Variable with Largest Variance.................. 17

Figure 2.5 Ordered Sample of the Variable with Average Variance 19

Figure 2.6 Ordered Sample of the Uniform Variable....................................... 22

Figure 2.7 Small Sample Average vs. Uniform Variance Plot.......................... 24

ACKNOWLEDGMENTS

I am deeply indebted to Professor Hokwon Cho for his leadership and tireless pursuit of excellence in completing this thesis. I would also like to say thank you to my wife, Tess, and my children, Michael, Christopher, and Natalie for their support. This thesis is dedicated to my late father, Emil Minev, and my mother, Maria Minev.

CHAPTER 1

INTRODUCTION

Daily trading of a large stock market portfolio requires substantial investment in data base management, computer equipment and development of proprietary software packages. Assuming market of 2500 securities with an average of 100 trades, an estimated 8 megabytes of storage space may be needed per day.

In addition, an astute adviser or investor will probably have a database on weekly and monthly closings, plus a large economic database on all securities under supervision. In many cases, especially during bear markets, a company stock can experience long declines (sometimes measured in years), when the economic fundamentals will improve and informed exchange, corporate, and professional investors will reenter the stock market on the long side. This sets the stage for bottom formation and possible stock reversal. The characteristics of the trading range are exhibited in the charts of two samples of monthly Dow Jones Industrial Average (DJIA) closing prices form April 1999 to June 2000 and from September 1999 to November 2000. Although the samples are spaced 6 months apart, the sample extremes are the same.

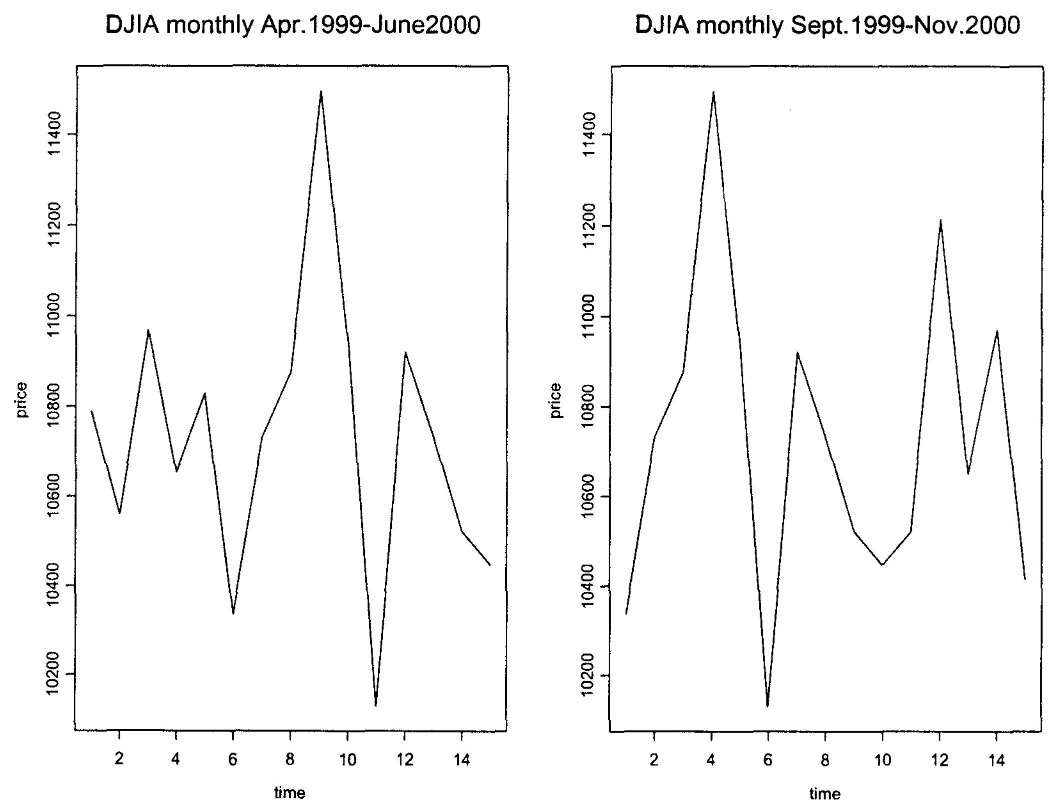

Figure 1.1 Two DJIA samples with the same extremes.

1.1 Terminology

Can we estimate the population mean of a stock by using its yearly high and low?. The goal of this thesis is to find a solution to the question poised above by developing a theorem based on the smallest and largest order statistics and then uses the results to construct confidence interval for the population mean of Normal Distribution.

The following terminology will be used in the thesis:

1. Bear Market: declining market in stocks.
2. Bull Market: advancing market in stocks.

3. Stock reversal: a change in a long-term trend of the stock.
4. Long position: owning a stock.
5. Short position: selling borrowed stock hoping to buy it at a lower price and return it to the lender (usually a brokerage company).
6. Overhead resistance: price range where a stock is likely to experience some selling.
7. Support area: price range were stock may find buying interest.

1.2 Formulation of the Problem

Suppose that we have a n-sided fair die such that one side is marked S, and n-1 sides are marked F where n is positive finite integer greater or equal to 4. Consider an experiment of throwing the die n times and let X be the number of "S" observed. Since this forms a binomial experiment, the probability of such an event is

$$P(X=1)=(1-\frac{1}{n})^{n-1}.$$

N	20	1000	1,000,000
P(X=1)	.3774	.3680	.3678
1/P(X=k)	2.6500	2.7169	2.7182

Table 1.3. Probability of getting X=1

Let X_1, X_2, X_3... X_n be an independent and identically distributed (iid) sample from a population having an unknown distribution F. Consider the smallest and largest order statistics of the sample, $X_{(1)}$ and $X_{(n)}$ respectively. We substitute $X_{(1)}$ with S and $X_{(n)}$ with F in the above experiment. This leads to the following theorem:

Theorem 1: Let $X_1, X_2, \ldots\ldots X_n$ *be a random sample of size* $n \geq 4$ *from continuous but unknown distribution* $F(X_i)$ *with* $E[X_i] = \mu$ *and* $\operatorname{var}(X_i) = \sigma^2 < \infty$. *Let* $X_{(1)} \leq X_{(2)} \leq X_{(3)} \leq, \ldots\ldots X_{(n)}$ *be the order statistics and assume that the extremes of* $X_{(1)}$ *and* $X_{(n)}$ *have been observed. Then, each future sample of size* $n \geq 4$ *with the smallest and largest order statistics* $X_{(1)}$ *and* $X_{(n)}$ *will have sample mean*

$$(i)\ X_{(1)} + \frac{R}{n} < \bar{X} < X_{(n)} - \frac{R}{n},$$

and sample variance

$$(ii)\ \frac{R^2}{2n} < s^2 < \frac{R^2}{4},$$

where R *is the range,* $X_{(n)} - X_{(1)}$.

CHAPTER 2

RANDOM VARIABLES OF TWO EXTREMES IN ORDER STATISTICS

2.1 Variable with Largest Mean

Definition 2.1: A random variable X_{LM} is said to have the largest mean if

$$X_{LM} = \begin{cases} X_{(1)}, & \text{with probability } \dfrac{1}{n} \quad (2.11) \\ X_{(n)}, & \text{with probability } \dfrac{n-1}{n} \quad (2.12) \end{cases}$$

where $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics of previously observed sample of size n from continuous not necessarily symmetrical distribution.

The conditional expected value of X_{LM} is

$$E(X_{LM} \mid X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{x_{(1)} + (n-1)x_{(n)}}{n} = x_{(n)} - \frac{R}{n}\,. \qquad (2.13)$$

Since the second moment is given by

$$E(X^2_{LM} \mid X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{x^2_{(1)} + (n-1)x^2_{(n)}}{n} \qquad (2.14)$$

The variance of X_{LM} is

$$\operatorname{var}(X_{LM} \mid X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{n[x^2_{(1)} + (n-1)x^2_{(n)}] - [(n-1)x_{(n)} + x_{(1)}]^2}{n^2}$$

$$= \frac{x^2_{(1)}(n-1) + x^2_{(n)}[n^2 - n - n^2 + 2n - 1] - 2(n-1)x_{(1)}x_{(n)}}{n^2}$$

$$= \frac{R^2(n-1)}{n^2}. \qquad (2.15)$$

In particular, for sufficiently large *n*,

$$\operatorname{var}(X_{LM} \mid X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{R^2}{n} \qquad (2.16)$$

where R is the range of the sample, $x_{(n)} - x_{(1)}$. Now we proceed with proof that the mean of any future sample from continuous distribution of size n with order statistics $X_{(1)}$ and $X_{(n)}$ is smaller than the mean of X_{LM}.

Proof of Theorem 1: (RHS of (i))

Let $X_{(i)} = X_{(1)} + \varepsilon_i$ *where* $2 \le i \le n$, $\varepsilon_2 < \varepsilon_3 < \varepsilon_4 \ldots < \ldots \varepsilon_n$. *Then,*

$$X_{(n)} - \frac{R}{n} = \frac{n(X_{(1)} + \varepsilon_n) - (X_{(1)} + \varepsilon_n - X_{(1)})}{n} > \frac{X_{(1)} + \sum_{i=2}^{n-1}(X_{(1)} + \varepsilon_i) + X_{(1)} + \varepsilon_n}{n},$$

or

$$\frac{nX_{(1)} + n\varepsilon_n - \varepsilon_n}{n} > \frac{nX_{(1)} + \sum_{i=2}^{n}\varepsilon_i}{n}.$$

Thus $nX_{(1)} + n\varepsilon_n - \varepsilon_n > nX_{(1)} + (\sum_{i=2}^{n-1}\varepsilon_i) + \varepsilon_n$ *and by simplifying we get,*

$$n\varepsilon_n - \varepsilon_n > (\sum_{i=2}^{n-1}\varepsilon_i) + \varepsilon_n .$$

By assumption $\varepsilon_2 < \varepsilon_3 < \varepsilon_4 \ldots < \ldots\varepsilon_n$ *and certainly* $(n-2)\varepsilon_n > \sum_{i=2}^{n-1}\varepsilon_i$. *This implies that* $(n-2)\varepsilon_n + \varepsilon_n > \sum_{i=2}^{n-1}\varepsilon_i + \varepsilon_n$ *or* $n\varepsilon_n - \varepsilon_n > (\sum_{i=2}^{n-1}\varepsilon_i) + \varepsilon_n$. *This completes the proof.*

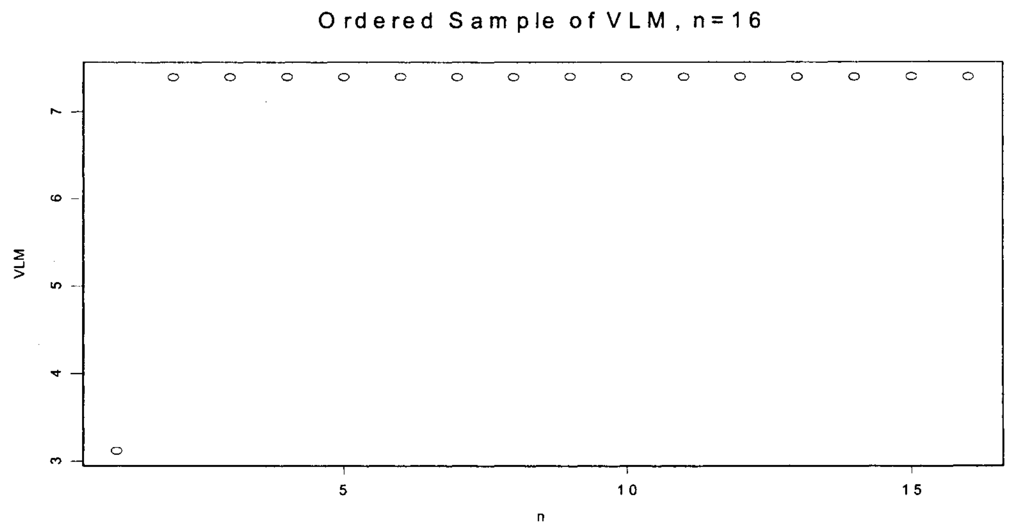

Figure 2.1. Ordered sample generated from distribution of the variable with the largest mean.

The chart shows that $X_{(1)}$=3.116, the rest of observations are equal to $X_{(n)}$=7.395. $X_{(1)}$ and $X_{(n)}$ are generated from N(5,1), sample size n=16.

2.2 Variable with Smallest Mean

Definition 2.2 : The random variable X_{SM} is said to have the smallest mean if

$$X_{SM} = \begin{cases} X_{(1)}, & \text{with probability } \dfrac{n-1}{n} \quad (2.21) \\ X_{(n)}, & \text{with probability } \dfrac{1}{n} \quad (2.22) \end{cases}$$

where $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics of previously observed sample of size n from continuous not necessarily symmetrical distribution.

The conditional expected value of X_{SM} is

$$E(X_{SM} | X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{x_{(1)}(n-1) + x_{(n)}}{n} = x_{(1)} + \frac{R}{n}. \quad (2.23)$$

The second moment is given by

$$E(X^2_{SM} | X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{x^2_{(1)}(n-1) + x^2_{(n)}}{n}. \quad (2.24)$$

The variance of X_{SM} is

$$\begin{aligned} \text{var}(X_{SM} | X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) &= \frac{n[x^2_{(1)}(n-1) + x^2_{(n)}] - [(n-1)x_{(1)} + x_{(n)}]^2}{n^2} \\ &= \frac{x^2_{(1)}[n^2 - n - n^2 + 2n - 1] + x^2_{(n)}(n-1) - 2(n-1)x_{(1)}x_{(n)}}{n^2} \\ &= \frac{R^2(n-1)}{n^2}. \end{aligned} \quad (2.25)$$

For sufficiently large n,

$$\text{var}(X_{LM} | X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{R^2}{n}. \quad (2.26)$$

where R is the range of the sample, $x_{(n)} - x_{(1)}$.

Now we proceed with proof that the mean of any future sample from continuous distribution of size n with order statistics $X_{(1)}$ and $X_{(n)}$ is greater than the mean of X_{SM} .

Proof of Theorem 1: (LHS of (i))

Let $X_{(i)} = X_{(1)} + \varepsilon_i$ *where* $2 \le i \le n$, $\varepsilon_2 < \varepsilon_3 < \varepsilon_4 \ldots < ..\varepsilon_n$. *Then,*

$$X_{(1)} + \frac{R}{n} = \frac{nX_{(1)} + X_{(1)} + \varepsilon_n - X_{(1)}}{n} < \frac{X_{(1)} + \sum_{i=2}^{n-1}(X_{(1)} + \varepsilon_i) + X_{(1)} + \varepsilon_n}{n} .$$

Hence,

$$\frac{nX_{(1)} + \varepsilon_n}{n} < \frac{nX_{(1)} + \sum_{i=2}^{n-1}\varepsilon_i + \varepsilon_n}{n} .$$

Thus $nX_{(1)} + \varepsilon_n < nX_{(1)} + (\sum_{i=2}^{n-1}\varepsilon_i) + \varepsilon_n$ *and by simplifying, we get that* $0 < (\sum_{i=2}^{n-1}\varepsilon_i)$. *By assumption,* $0 < \varepsilon_2 < \varepsilon_3 < \varepsilon_4 \ldots < ..\varepsilon_n$ *and* $\sum_{i=2}^{n-1}\varepsilon_i > 0$. *And this completes the proof.*

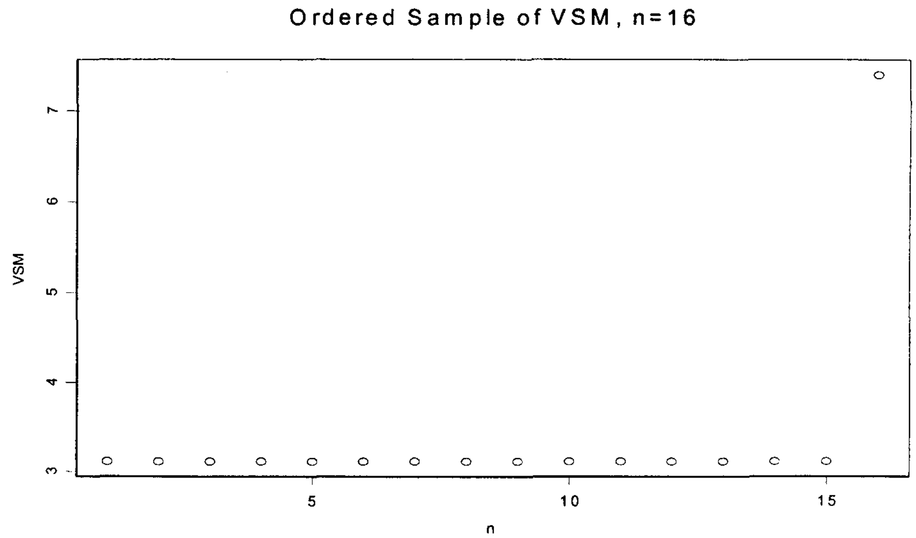

Figure 2.2. Ordered sample generated from distribution of the variable with the smallest mean.

The chart shows 15 observations of $X_{(1)}$=3.116, $X_{(n)}$=7.395. $X_{(1)}$ and $X_{(n)}$ are generated from N(5,1), sample size n=16.

2.3 Variable with Smallest Variance

Definition 2.3 : The random variable X_{SV} is said to have the smallest variance if

$$X_{SV} = \begin{cases} X_{(1)}, & \text{with probability } \frac{1}{n} \quad (2.31) \\ \frac{X_{(1)} + X_{(n)}}{2}, & \text{with probability } \frac{n-2}{n} \quad (2.32) \\ X_{(n)}, & \text{with probability } \frac{1}{n} \quad (2.33) \end{cases}$$

where $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics of previously observed sample of size n from continuous not necessarily symmetrical distribution. Consider the first two moments of X_{SV}.

The conditional expectation of X_{SV} is

$$E(X_{SV}\,|X_{(1)}=x_{(1)},X_{(n)}=x_{(n)}) = x_{(1)}(\frac{1}{n}) + \frac{(n-2)(x_{(1)}+x_{(n)})}{2n} + x_{(n)}(\frac{1}{n})$$
$$= \frac{2x_{(1)}+(n-2)(x_{(1)}+x_{(n)})+2x_{(n)}}{2n}$$
$$= \frac{(x_{(1)}+x_{(n)})}{2}. \quad (2.34)$$

The second moment is given by

$$E(X^2_{SV}\,|X_{(1)}=x_{(1)},X_{(n)}=x_{(n)}) = x^2_{(1)}(\frac{1}{n}) + \frac{(n-2)[x_{(1)}+x_{(n)}]^2}{4n} + x^2_{(n)}(\frac{1}{n})$$
$$= \frac{4x^2_{(1)}+(n-2)[x_{(1)}+x_{(n)}]^2+4x^2_{(n)}}{4n}. \quad (2.35)$$

The variance of X_{SV} is

$$\mathrm{var}(X_{SV}\,|X_{(1)}=x_{(1)},X_{(n)}=x_{(n)}) =$$
$$= \frac{4x^2_{(1)}+n[x_{(1)}+x_{(n)}]^2-2[x_{(1)}+x_{(n)}]^2+4x^2_{(n)}}{4n} - [\frac{x_{(1)}+x_{(n)}}{2}]^2$$
$$= \frac{4x^2_{(1)}+n[x_{(1)}+x_{(n)}]^2-2[x_{(1)}+x_{(n)}]^2+4x^2_{(n)}}{4n} - \frac{n[x_{(1)}+x_{(n)}]^2}{4n}$$
$$= \frac{2x^2_{(1)}-4x_{(1)}x_{(n)}+2x^2_{(n)}}{4n} = \frac{[x_{(1)}-x_{(n)}]^2}{2n}$$
$$= \frac{R^2}{2n} \quad (2.36)$$

where R is the range of the sample, $x_{(n)} - x_{(1)}$. Now we proceed with proof that the variance of any future sample from continuous distribution of size n with order statistics $X_{(1)}$ and $X_{(n)}$ is greater than the variance of X_{SV}.

Lemma 1: For a given set of order statistics $X_{(1)} \le X_{(2)} \le X_{(3)} \le, \ldots\ldots X_{(n)}$ *there exists* $k > 0$ *such that* $\frac{X_{(1)}+X_{(2)}+,\ldots\ldots+X_{(n-1)}+X_{(n)}}{n} = k[\frac{X_{(1)}+X_{(n)}}{2}]$.

Proof:

We have shown in sections 2.1 and 2.2 that $\bar{X}_{SM} < \frac{X_{(1)} + X_{(2)} +, \ldots + X_{(n-1)} + X_{(n)}}{n}$

and $\frac{X_{(1)} + X_{(2)} +, \ldots + X_{(n-1)} + X_{(n)}}{n} < \bar{X}_{LM}$. *Hence,* $X_{(1)} + \frac{R}{n} < k[\frac{X_{(1)} + X_{(n)}}{2}]$ *and*

$k[\frac{X_{(1)} + X_{(n)}}{2}] < X_{(n)} - \frac{R}{n}$. *Solving for* k *we get,* $\frac{2X_{(1)}}{X_{(1)} + X_{(n)}} + \frac{2R}{n(X_{(1)} + X_{(n)})} < k$ *and*

$k < \frac{2X_{(n)}}{X_{(1)} + X_{(n)}} - \frac{2R}{n(X_{(1)} + X_{(n)})}$.

Taking the limit as n *goes to infinity and dividing by* $X_{(n)}$ *it follows that*

$$\frac{2X_{(1)}/X_{(n)}}{1 + X_{(1)}/X_{(n)}} < k \text{ and } k < \frac{2}{1 + X_{(1)}/X_{(n)}}. \tag{2.37}$$

<u>*Case 1*</u>*: If* $\frac{X_{(1)}}{X_{(n)}} = \delta$, δ *is infinitesimal,* $0 < \delta < 1$. *Then,* $Lim_{\delta \to 0}(\frac{2\delta}{1+\delta}) < k$ *leads to*

$0 < k$. *And,* $k < Lim_{\delta \to 0}(\frac{2}{1+\delta})$ *leads to* $k < 2$.

<u>*Case 2*</u>*: If* $\frac{X_{(1)}}{X_{(n)}} = 1 - \delta$, δ *is infinitesimal,* $0 < \delta < 1$. *Then,* $Lim_{\delta \to 0}(\frac{2 - 2\delta}{1 + 1 - \delta}) < k$

leads to $k > 1$. *Letting* $k < Lim_{\delta \to 0}(\frac{2}{1 + 1 - \delta})$ *leads to* $k < 1$.

Therefore, the domain of k *is* $(0,1) \cup (1,2)$. (2.38)

Proof of Theorem 1: (LHS of (ii))

Let X_{sub} *be the random variable generating the future sample. We will represent* X_{sub} *as sum of the Uniform or Smallest Mean variable and an appropriate disturbance term (Hays, 1981).*

Case 1: For symmetrical distributions we can express the subsequent sample as $X_{sub} = X_{UV} + \gamma_i$, *where* γ_i *is a random variable with mean* μ *and variance* σ^2.

Then, $\mathrm{var}(X_{sub}) = \mathrm{var}(X_{UV}) + \mathrm{var}(\gamma_i) + 2\mathrm{cov}(X_{UV}\gamma_i)$. *Since* $\mathrm{var}(X_{sub}) > \mathrm{var}(X_{UV})$, *it is simple to show that* $\mathrm{var}(X_{SV}) < \mathrm{var}(X_{UV})$, *that is,* $\frac{R^2}{2n} < \frac{R^2(n+1)}{12(n-1)}$ *given that* $0 < n^2 - 5n + 6$ *is always true for* $n \geq 4$.

Case 2: For asymmetrical distributions we can express the future sample as $X_{sub} = X_{SM} + \gamma_i$, *where* γ_i *is a random variable with mean* μ *and variance* σ^2.

Since X_{SM} *has variance equal to* $\frac{R^2(n-1)}{n^2}$,

$$\mathrm{var}(X_{sub}) = \frac{R^2(n-1)}{n^2} + \mathrm{var}(\gamma_i) + 2\mathrm{cov}(X_{SM}\gamma_i).$$

But, $\mathrm{var}(X_{sub}) > \frac{R^2(n-1)}{n^2}$, *hence we can show that* $\mathrm{var}(X_{SV}) < \frac{R^2(n-1)}{n^2}$, *that is,*

$\frac{R^2}{2n} < \frac{R^2(n-1)}{n^2}$ *since* $0 < n-2$ *is always true for* $n \geq 4$.

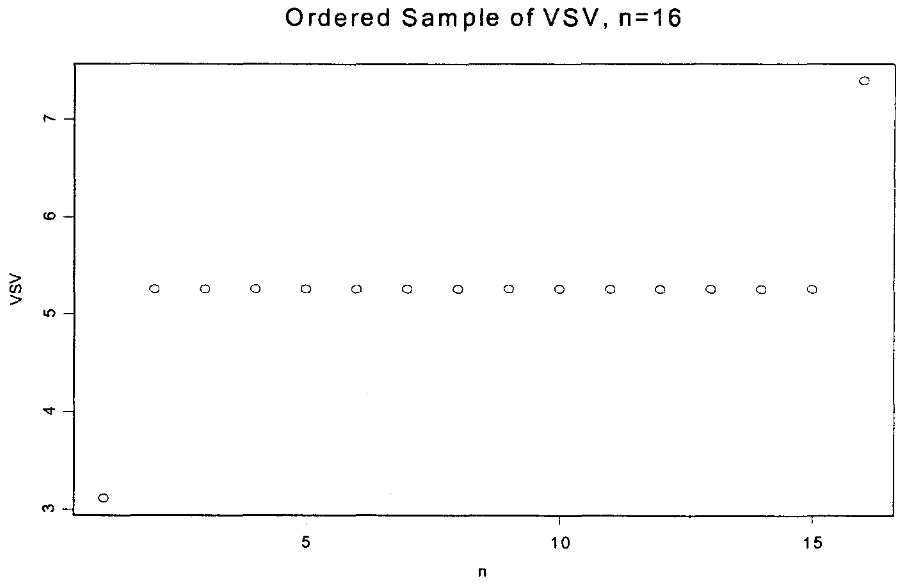

Figure 2.3. Ordered sample generated from distribution of the variable with the smallest variance.

The chart shows 1 observation of $X_{(1)}=3.116$, 14 observations of $(X_{(1)}+X_{(n)})/2=5.255$, $X_{(n)}=7.395$. $X_{(1)}$ and $X_{(n)}$ are generated from N(5,1) ,sample size n=16.

2.4 Variable with Largest Variance

Definition 2.4 : The random variable X_{LV} is said to have the largest variance if

$$X_{LV} = \begin{cases} X_{(1)}, & \text{with probability } \dfrac{m}{n} \quad (2.41) \\ X_{(n)}, & \text{with probability } \dfrac{n-m}{n} \quad (2.42) \end{cases}$$

where $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics from previously observed sample of size n from continuous not necessarily symmetrical distribution and 1<m<n.

The conditional expected value of X_{LV} is

$$E(X_{LV} \mid X_{(1)}=x_{(1)}, X_{(n)}=x_{(n)}) = \frac{mx_{(1)}+(n-m)x_{(n)}}{n}. \tag{2.43}$$

The second moment is given by

$$E(X^2_{LV} \mid X_{(1)}=x_{(1)}, X_{(n)}=x_{(n)}) = \frac{mx^2_{(1)}+(n-m)x^2_{(n)}}{n}. \tag{2.44}$$

The variance is

$$\begin{aligned}
\mathrm{var}(X_{LV} \mid X_{(1)}=x_{(1)}, X_{(n)}=x_{(n)}) &= [\frac{mx^2_{(1)}+(n-m)x^2_{(n)}}{n}] - [\frac{mx_{(1)}+(n-m)x_{(n)}}{n}]^2 = \\
&= \frac{mnx^2_{(1)}+n^2x^2_{(n)}-mnx^2_{(n)}-m^2x^2_{(1)}-2m(n-m)x_{(1)}x_{(n)}-[n-m]^2x^2_{(n)}}{n^2} \\
&= \frac{[mnx^2_{(1)}-m^2x^2_{(1)}]+[n^2x^2_{(n)}-mnx^2_{(n)}]-[2m(n-m)x_{(1)}x_{(n)}]-[(n-m)^2x^2_{(n)}]}{n^2} \\
&= \frac{mx^2_{(1)}(n-m)+nx^2_{(n)}(n-m)-2m(n-m)x_{(1)}x_{(n)}-[n-m]^2x^2_{(n)}}{n^2} \\
&= \frac{(n-m)[mx^2_{(1)}+nx^2_{(n)}-2mx_{(1)}x_{(n)}-nx^2_{(n)}+mx^2_{(n)}]}{n^2} \\
&= \frac{(n-m)[mx^2_{(1)}-2mx_{(1)}x_{(n)}+mx^2_{(n)}]}{n^2} = \frac{m(n-m)R^2}{n^2}
\end{aligned} \tag{2.45}$$

where R is the range of the sample, $x_{(n)} - x_{(1)}$.

The first derivative of the variance with respect to m is given by:

$$\frac{d(\mathrm{var}(x))}{dm} = \frac{nR^2-2mR^2}{n^2}.$$

Setting the first derivative to zero, we get that $\frac{nR^2 - 2mR^2}{n^2} = 0$ or $m = \frac{n}{2}$. Since range is always positive, the second derivative is negative, therefore the variance will be maximized when $m = \frac{n}{2}$.

Based on $m = \frac{n}{2}$, the mean and variance of X_{LV} are given by :

$$E(X_{LV}) = \frac{(x_{(1)} + x_{(n)})}{2} \tag{2.46}$$

$$\text{var}(X_{LV}) = \frac{R^2}{4} \tag{2.47}$$

Now we proceed with proof that the variance of any future sample from continuous distribution of size n with order statistics $X_{(1)}$ and $X_{(n)}$ is smaller than the variance of X_{LV}.

Proof of Theorem 1 : (RHS of (ii))

Let X_{sub} be the random variable generating the future sample. We will represent X_{sub} as sum of the Uniform or Smallest Mean variable and an appropriate disturbance term (Hays, 1981).

Case 1: *For symmetrical distributions we can express the future sample as $X_{sub} = X_{UV} + \gamma_i$, where γ_i is a random variable with mean μ and variance σ^2. Then,* $\text{var}(X_{sub}) = \text{var}(X_{UV}) + \text{var}(\gamma_i) + 2\,\text{cov}(X_{UV}\gamma_i)$. *Since* $\text{var}(X_{sub}) > \text{var}(X_{UV})$, *it is simple to show that* $\text{var}(X_{UV}) < \text{var}(X_{LV})$, *that is,* $\frac{R^2(n+1)}{12(n-1)} < \frac{R^2}{4}$ *since* $0 < n-2$ *is always true for* $n \geq 4$.

Case 2: For asymmetrical distributions we can express the future sample as $X_{sub} = X_{SSM} + \gamma_i$, where γ_i is a random variable with mean μ and variance σ^2.

Since X_{SM} has variance equal to $\frac{R^2(n-1)}{n^2}$,

$\operatorname{var}(X_{sub}) = \frac{R^2(n-1)}{n^2} + \operatorname{var}(\gamma_i) + 2\operatorname{cov}(X_{SM}\gamma_i)$. *But* $\operatorname{var}(X_{sub}) > \frac{R^2(n-1)}{n^2}$, *hence we can show that* $\frac{R^2(n-1)}{n^2} < \operatorname{var}(X_{LV})$, *that is,* $\frac{R^2(n-1)}{n^2} < \frac{R^2}{4}$ *since* $0 < n^2 - 4n + 4$ *is always true for* $n \geq 4$.

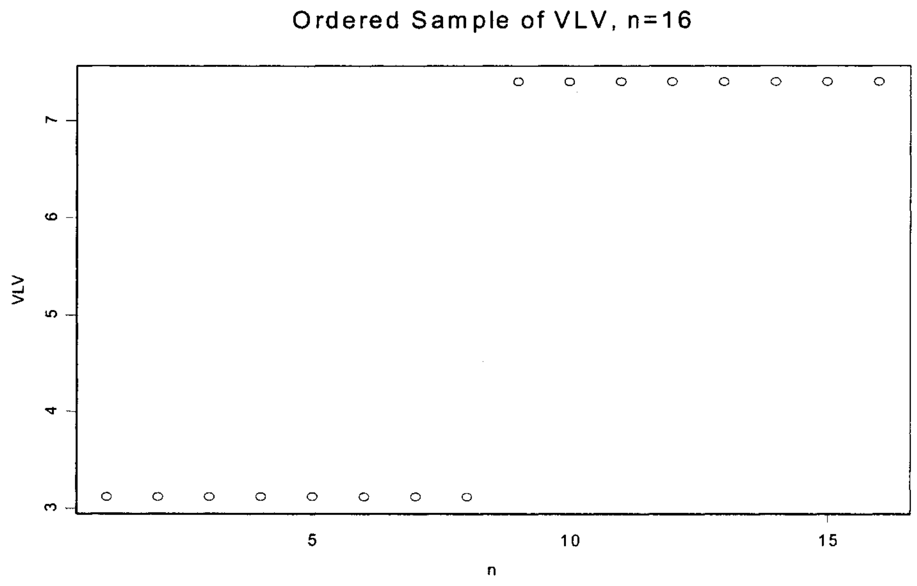

Figure 2.4. Ordered sample generated from distribution of the variable with the largest variance.

The chart shows 8 observations of $X_{(1)}$=3.116, 8 observations of $X_{(n)}$=7.395. $X_{(1)}$ and $X_{(n)}$ are generated from N(5,1) sample size n=16.

2.5 Variable with Average Variance

Definition 2.5 : The random variable X_{AV} is said to have average variance if

$$X_{AV} = \begin{cases} X_{(1)}, & \text{with probability } \frac{1}{n} & (2.51) \\ \frac{3X_{(1)} + X_{(n)}}{4}, & \text{with probability } \frac{n-2}{2n} & (2.52) \\ \frac{X_{(1)} + 3X_{(n)}}{4}, & \text{with probability } \frac{n-2}{2n} & (2.53) \\ X_{(n)}, & \text{with probability } \frac{1}{n} & (2.54) \end{cases}$$

where $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics from previously observed sample of size n from continuous not necessarily symmetrical distribution.

The conditional expectation of X_{AV} is

$$E(X_{AV} | X_{(1)}= x_{(1)}, X_{(n)}= x_{(n)}) = x_{(1)}\frac{1}{n} + \frac{(n-2)(3x_{(1)}+ x_{(n)})}{8n} + \frac{(n-2)(x_{(1)}+ 3x_{(n)})}{8n} + x_{(n)}\frac{1}{n}$$

$$= \frac{8x_{(1)}+ (n-2)[3x_{(1)}+ x_{(n)}+ x_{(1)}+ 3x_{(n)}] + 8x_{(n)}}{8n}$$

$$= \frac{8x_{(1)}+ 8x_{(n)}+ [4nx_{(1)}+ 4nx_{(n)}- 8x_{(1)}- 8x_{(n)}]}{8n}$$

$$= \frac{(x_{(1)}+ x_{(n)})}{2}. \tag{2.55}$$

The second moment is

$$E(X^2_{AV} | X_{(1)}= x_{(1)}, X_{(n)}= x_{(n)}) =$$

$$= x^2_{(1)}\frac{1}{n} + \frac{(n-2)[3x_{(1)}+ x_{(n)}]^2}{32n} + \frac{(n-2)[x_{(1)}+ 3x_{(n)}]^2}{32n} + x^2_{(n)}\frac{1}{n}$$

$$= \frac{32x^2_{(1)} + (n-2)([3x_{(1)}+ x_{(n)}]^2 + [x_{(1)}+ 3x_{(n)}]^2) + 32x^2_{(n)}}{32n}$$

$$= \frac{5nx^2_{(1)}+ 6x^2_{(1)} + 5nx^2_{(n)} + 6x^2_{(n)} + 6nx_{(1)}x_{(n)}- 12x_{(1)}x_{(n)}}{16n}. \tag{2.56}$$

The variance of X_{AV} is

$$\operatorname{var}(X_{AV} \mid X_{(1)}=x_{(1)}, X_{(n)}=x_{(n)}) =$$

$$= \frac{5nx^2_{(1)} + 6x^2_{(1)} + 5nx^2_{(n)} + 6x^2_{(n)} + 6nx_{(1)}x_{(n)} - 12x_{(1)}x_{(n)}}{16n} - [\frac{x_{(1)} + x_{(n)}}{2}]^2$$

$$= \frac{5nx^2_{(1)} + 6x^2_{(1)} + 5nx^2_{(n)} + 6x^2_{(n)} + 6nx_{(1)}x_{(n)} - 12x_{(1)}x_{(n)} - 4nx^2_{(1)} - 8nx_{(1)}x_{(n)} - 4nx^2_{(n)}}{16n}$$

$$= \frac{x^2_{(1)}(n+6) - 2(n+6)x_{(1)}x_{(n)} + x^2_{(n)}(n+6)}{16n}$$

$$= \frac{(n+6)(x^2_{(1)} - 2x_{(1)}x_{(n)} + x^2_{(n)})}{16n} = \frac{R^2(n+6)}{16n} = \frac{R^2}{16} + \frac{3R^2}{8n}. \qquad (2.57)$$

For large samples we have that

$$\operatorname{var}(X_{AV}) = Lim_{n\to\infty}(\frac{R^2}{16} + \frac{3R^2}{8n}) = \frac{R^2}{16} \qquad (2.58)$$

where $R = x_{(n)} - x_{(1)}$.

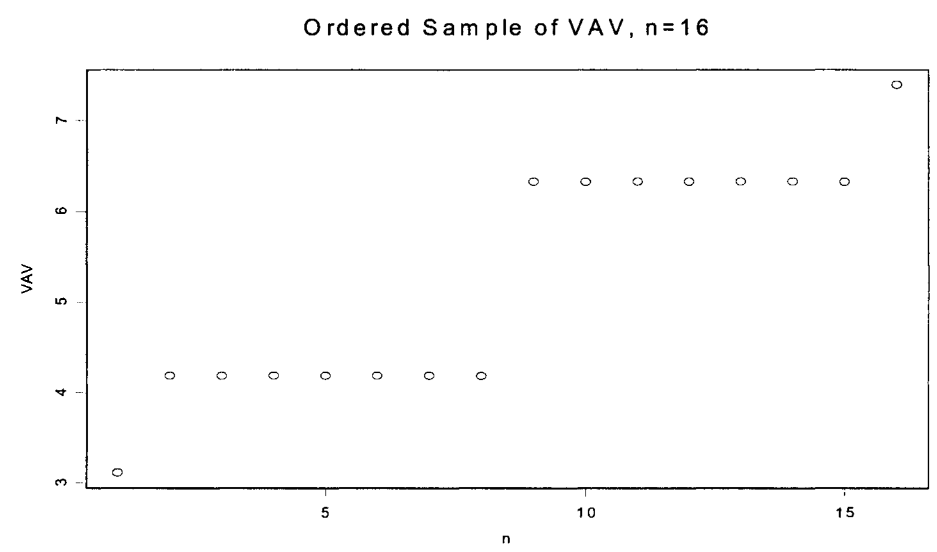

Figure 2.5. Ordered sample generated from distribution of the variable with the average variance.

The chart shows 1 observation of $X_{(1)}$=3.116, 7 observations of $(3X_{(1)}+X_{(n)})/4$ and $(X_{(1)}+3X_{(n)})/4$ each, 1 observation of $X_{(n)}$=7.395. $X_{(1)}$ and $X_{(n)}$

are generated from N(5,1) ,sample size n=16.

2.6 Uniform Random Variable

Definition 2.6 The uniform random variable X_{UV} has distribution

$X_{UV} = X_{(1)} + m(i-1)$ with probability $\frac{1}{n}$, where $i = 1,2,3.....n$,

$$v = \frac{X_{(n)} - X_{(1)}}{n-1} = \frac{R}{n-1}, \tag{2.61}$$

and $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics of previously observed sample of size n from continuous not necessarily symmetrical distribution.

The conditional expected value of X_{UV} is

$$E(X_{UV} \mid X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = \frac{\sum_{i=1}^{n}[x_{(1)} + \frac{R}{n-1}(i-1)]}{n} = \frac{nx_{(1)} + (\frac{x_{(n)} - x_{(1)}}{n-1})[\sum_{i=1}^{n}(i-1)]}{n}$$

$$= \frac{nx_{(1)} + [\frac{x_{(n)} - x_{(1)}}{n-1}][\frac{n(n+1)}{2} - n]}{n}$$

$$= \frac{nx_{(1)} + [\frac{x_{(n)} - x_{(1)}}{n-1}][\frac{n(n-1)}{2}]}{n}$$

$$= \frac{x_{(1)} + x_{(n)}}{2}. \tag{2.62}$$

The second moment is given by

$$E(X^2_{UV}\,|X_{(1)}=x_{(1)},X_{(n)}=x_{(n)}) = \frac{\sum_{i=1}^{n}[x_{(1)}+\frac{R}{n-1}(i-1)]^2}{n}$$

$$=\frac{\sum_{i=1}^{n}[x^2_{(1)}+\frac{2Rx_{(1)}(i-1)}{n-1}+[\frac{R}{n-1}]^2[i-1]^2]}{n}$$

$$=\frac{nx^2_{(1)}+\frac{2Rx_{(1)}}{n-1}[\frac{n(n+1)}{2}-n]+\frac{R^2}{(n-1)^2}[\frac{(n)(n+1)(2n+1)}{6}-\frac{2n(n+1)}{2}+n]}{n}$$

$$=\frac{nx^2_{(1)}+\frac{2(x_{(n)}-x_{(1)})x_{(1)}}{n-1}[\frac{n(n-1)}{2}]+\frac{(x_{(n)}-x_{(1)})^2}{(n-1)^2}[\frac{(n)(2n-1)(n-1)}{6}]}{n}$$

$$=\frac{6n(n-1)x_{(n)}x_{(1)}+n(2n-1)(x^2_{(1)}-2x_{(1)}x_{(n)}+x^2_{(n)})}{6n(n-1)}$$

$$=\frac{x^2_{(1)}(2n-1)+x_{(1)}x_{(n)}(6n-6-4n+2)+x_{(n)}(2n-1)}{6(n-1)}$$

$$=\frac{x^2_{(1)}(2n-1)+2x_{(1)}x_{(n)}(n-2)+x_{(n)}(2n-1)}{6(n-1)}. \tag{2.63}$$

Hence, the variance is

$$\mathrm{var}(X_{UV})=\frac{x^2_{(1)}(2n-1)+2x_{(1)}x_{(n)}(n-2)+x^2_{(n)}(2n-1)}{6(n-1)}-[\frac{x_{(1)}+x_{(n)}}{2}]^2$$

$$=\frac{x^2_{(1)}[4n-2-3(n-1)]+2x_{(1)}x_{(n)}[2n-4-3(n-1)]+x^2_{(n)}[4n-2-3(n-1\,)]}{12(n-1)}$$

$$=\frac{(n+1)[x^2_{(1)}-2x_{(1)}x_{(n)}+x^2_{(n)}]}{12(n-1)}=\frac{(n+1)R^2}{12(n-1)}. \tag{2.64}$$

The large sample variance is given by

$$\mathrm{var}(X_{UV})=\frac{R^2}{12} \tag{2.65}$$

where $R=x_{(n)}-x_{(1)}$.

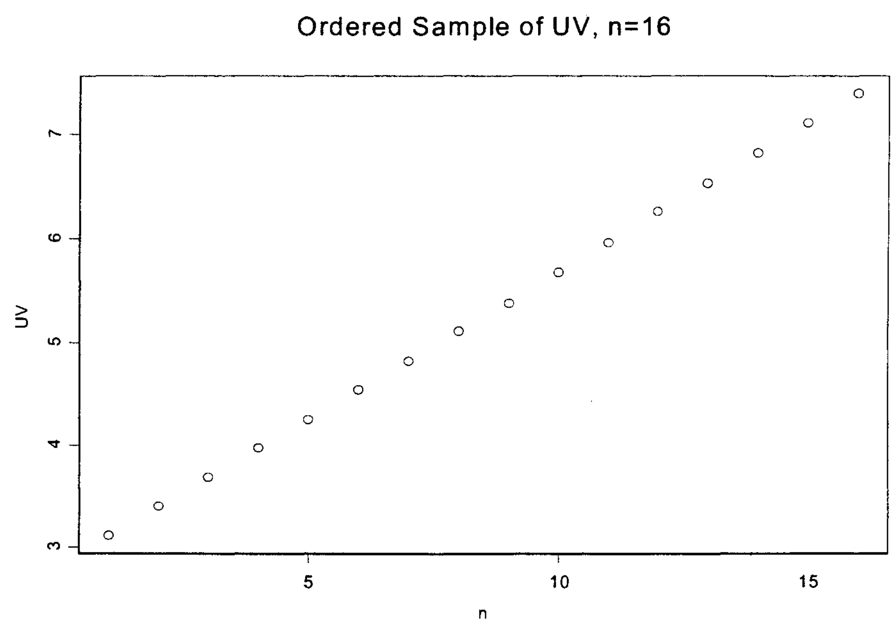

Figure 2.6. Ordered sample generated from the distribution of the Uniform variable.

The chart shows 1 observation of $X_{(1)}$=3.116, 14 observations along the line connecting $X_{(1)}$ and $X_{(n)}$ and 1 observation of $X_{(n)}$=7.395. $X_{(1)}$ and $X_{(n)}$ are generated by N(5,1),sample size n=16.

2.7 Simulated Normal Variable and Derivation of Confidence Interval

Definition 2.7 : The simulated Normal random variable X_{snv} (Hays,1981) has distribution

$$X_{SNV} = X_{UV} + \psi_i ,$$
$$\psi_i \sim N(\xi, \sigma^2). \tag{2.71}$$

where $X_{(1)}$ and $X_{(n)}$ are the smallest and the largest order statistics of previously observed sample of size n from continuous not necessarily symmetrical distribution.

As consequence of (2.38) we assume $\bar{\psi_i} \neq 0$. Then,

$$E(X_{SNV} \big| X_{(1)} = x_{(1)}, X_{(n)} = x_{(n)}) = E(X_{UV}) + E(\psi_i) = \frac{x_{(1)} + x_{(n)}}{2} \pm \bar{\psi_i}, \quad (2.72)$$

$$\text{var}(X_{UV} + \psi_i) = \text{var}(X_{UV}) + \text{var}(\psi_i) + 2\,\text{cov}(X_{UV}, \psi_i). \quad (2.73)$$

Ignoring the covariance term for a moment and subtracting from both sides of (2.73) the $\text{var}(X_{AV})$ for large samples, we get that

$$\text{var}(X_{SNV}) - \frac{R^2}{16} = \frac{R^2}{12} - \frac{R^2}{16} + \text{var}(\psi_i).$$

We can write $\frac{R^2}{48} + \text{var}(\psi) > 0$ since both quantities are positive. This implies that

$0 < \text{var}(\psi) \leq \frac{R^2}{48}$ or $\frac{R^2}{48} < \text{var}(\psi)$.

The variance is bounded above by $x_{(1)}$ because, by definition, the first order statistics of the unknown sample is $X_{SNV} = x_{(1)} + \psi_i$ and, if $\psi_1 = -x_{(1)}$ then $X_{SNV(1)} = 0$, which contradicts the fact that the smallest and largest order statistics are known. Thus we can express the variance of the disturbance term as

$$0 < \text{var}(\psi_i) \leq \frac{R^2}{48} \text{ or } \frac{R^2}{48} < \text{var}(\psi_i) < x_{(1)}. \quad (2.74)$$

In order to correlate the disturbance term with the uniform variable, we can select $\frac{R^2}{48}$ as an estimate of the variance of the disturbance term. It is shown in

Appendix II that the covariance between X_{LV} and X_{SV} is $\text{cov}(X_{LV}, X_{SV}) = \frac{R^2}{2n}$.

Therefore, it is reasonable to expect that the covariance between the uniform variable and the disturbance term will be of the form $\text{cov}(X_{UV}, \psi_i) = \frac{\eta R^2}{2n}$, where $\eta > 0$; and that the covariance will get close to zero for large values of n.

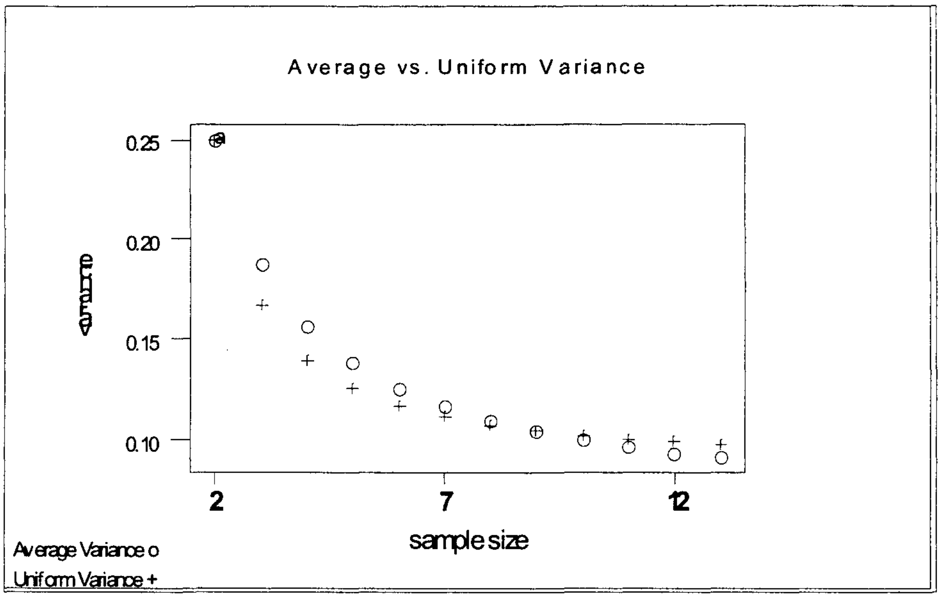

Figure 2.7. Plot of small sample Average Variance vs. small sample Uniform Variance at unit Range, R=1.

The total variance is given by:

$$\text{var}(X_{SNV}) = \frac{R^2}{12} + \frac{R^2}{48} = 0.10417R^2. \tag{2.75}$$

If we take $\text{var}(X_{SNV}) = .10417R^2$, then

$$sdev(X_{SNV}) = 0.3227R. \tag{2.76}$$

The variance of the disturbance term is given by $R^2(1/48)=0.02083R^2$ and the standard deviation is $\sqrt{0.02083R^2} = .1443R$. Since the expectation of the disturbance is assumed to be close to zero, and the standard deviation rather large $0.1443R$, the coefficient of variation $CV > 3$ was found to work well. This leads to the large sample estimate of the disturbance term mean of $0.0403R$.

The mean of our simulated sample is

$$\overline{X}_U = \frac{x_{(1)} + x_{(n)}}{2} + 0.0403R \quad (2.77)$$

or

$$\overline{X}_D = \frac{x_{(1)} + x_{(n)}}{2} - 0.0403R \ . \quad (2.78)$$

We therefore have the following two overlapping confidence intervals:

$$(\frac{x_{(1)} + x_{(n)}}{2} + 0.0403R) \pm \frac{0.3227Rt_{(n-1,1-\frac{\alpha}{2})}}{\sqrt{(n-1)}}, \quad (2.79)$$

$$(\frac{x_{(1)} + x_{(n)}}{2} - 0.0403R) \pm \frac{0.3227Rt_{(n-1,1-\frac{\alpha}{2})}}{\sqrt{(n-1)}}, \quad (2.80)$$

where $R = x_n - x_1$ and t is critical value of a T-distribution with n-1 degrees of freedom and level of significance of $1-\frac{\alpha}{2}$.

Let

$$UL = (\frac{x_{(1)} + x_{(n)}}{2} + 0.0403R) + \frac{0.3227Rt_{(n-1,1-\frac{\alpha}{2})}}{\sqrt{(n-1)}}, \quad (2.81)$$

$$LL = \left(\frac{x_{(1)} + x_{(n)}}{2} - 0.0403R\right) - \frac{0.3227Rt_{(n-1,1-\frac{\alpha}{2})}}{\sqrt{(n-1)}}. \qquad (2.82)$$

Then UL and LL are the limits of the $1-\alpha$ C.I.

CHAPTER 3

SIMULATION STUDIES

3.1 Numerical Simulation and Testing of Confidence Interval

Using S+ computer program MeanTest11 shown in Appendix I, simulation study was performed in order to investigate and illustrate the performance of the procedure we have suggested. The results of the simulation, based on inequalities (2.81) and (2.82), are summarized in the following tables, which contains the normal probabilities and coverage probabilities for various normal distributions. Every row in Table 3.11 and 3.12 corresponds to 200 independent experiments.

Probability	N(5,1)	N(5,4)	N(5,9)
0.90	.915	.88	.935
0.95	.935	.965	.975

Table 3.1.1 Probability for sample size n=45, 90% and 95% confidence intervals.

Probability	N(5,1)	N(5,4)	N(5,9)
0.90	.895	.895	.865
0.95	.88	.90	.915

Table 3.1.2 Probability for sample size n=100, 90% and 95% confidence intervals.

From the data shown in the tables above it appears that the probability coverage is affected by the sample size and the coefficient of dispersion of the distribution being used.

3.2 Application in the Sock Market

In this section, we would like apply results we obtained to a major market average by using the data in Appendix 1 where we have monthly High, Low and Closing for the Dow Jones Industrial Average(DJIA) for the period from April 1991 to April 2001. The summary of the data is shown in Table 3.2 (The data source is www.MSN.com). Using High, Low, sample size of 30, and confidence limit of 0.95, we will use inequality (2.81 and 2.82) to compute an confidence interval for each month, denoting the upper limit as UL and lower limit as LL.

April 2001 High 10043/ Low 9385.4 Close 9485.7

April 1991 High 3030.5/ Low 2848.5 Close 2887.9

Our decision rule is simple:

If the Upper Limit of the next confidence interval is higher, we buy, assuming long position to start with at the closing price.

If the Upper Limit is 2% or lower than the previous limit, we sell at closing price.

If we sustain 2 consecutive losses on the long side (after two flags have been raised), we revert to short position. The rules for short position are opposite to that of a long position.

The next table shows the DJIA transactions for the period April1991-April2001.

Month	Closing	UL	LL	Trade	Price	Gain	Total	Flag
Apr. 01	9485	9821	9607					
Dec. 00	10788	10745	10457	sell	10788			
Nov. 00	10414	10832	10524	buy	10414	108	5651	
Oct. 00	10971	10590	10089					
May. 00	10522	10775	10474	sell	10522			
Apr. 00	10733	11104	10624	buy	10733	188	5543	
Mar. 00	10921	10733	10184	sell	10921			
Feb. 00	10128	10733	10255	sell	10128	-749	5355	F2
Jan. 00	10940	11470	11048					
Nov. 99	10877	10943	10700	buy	10877			
Oct. 99	10729	10546	10221					
Sep. 99	10337	10826	10447	sell	10337	1775	6104	
Aug. 99	10829	11111	10804					

Oct. 98	8592.1	8273.7	7844.4	buy	8592			
Sep. 98	7842.6	7959	7674.4					
Aug. 98	7539.1	8465.8	8000.1	sell	7539	-1344	4329	F1
July. 98	8883.3	9201.5	8997.6	buy	8883			
June.98	8952	8942.4	8737.2	sell	8952	1044	5673	
May. 98	8900	9126.2	8946.8					
Dec. 97	7908.3	7991.6	7781.2	buy	7908			
Nov. 97	7823.1	7732.3	7537					
Sep. 97	7945.3	7902.3	7732.3	sell	7945	2329	4629	
Aug. 97	7622.4	8084.1	7836.9					
Aug. 96	5616.2	5676.3	5593.5	buy	5616			
July. 96	5528.9	5567.6	5372.4	sell	5528	1770	2300	
June.96	5654.6	5699.5	5630					
May. 94	3758.4	3728.4	3670.1	buy	3758			
Apr. 94	3681.7	3661.5	3592.2					
Mar. 94	3636	3787.8	3668.1	sell	3636	331	530	
Feb. 94	3832	3935.3	3874.6					
Nov. 92	3305.2	3276	3227.3	buy	3305			
Oct. 92	3226.3	3222.6	3156.2	sell	3226	199	199	
Sep. 92	3271.7	3335.8	3282.2					
May. 91	3027.5	2973.7	2905.3	buy	3027			
Apr. 91	2887.9	2969.1	2909.9					

Table 3.2 DJIA transactions April1991-April2001

The percentage gain for is given by (points gained)/(high close-low close) or 5651/(9485-2887)=85.6% or 8.56% per year.

3.3 Application to Environmental Data

As a second example of the application of the procedure, we consider an example that is given in Singh, Singh, Engelhard (1997). They studied the lognormal distribution in Environmental data and compared with several methods for estimation of the Upper Bound of the distribution.

440.8517	1013.4986	1857.7698	500.9632	397.9905
110.7144	196.2847	128.2843	1529.9753	5.7978
940.8903	597.5925	1519.5159	181.6512	52.8952

Table 3.3.1 Simulated data, sample of size n = 15 from LN (5,1.5).

It is known that the population has mean of 457.14, variance of 1331.83 and coefficient of variation of 2.91. Sample mean, Variance and Coefficient of Variations are 631.65, 603.13 and 0.96 respectively.

In order to estimate the population mean of the distribution by applying results developed by Gumbel (1954) for population parameters, we obtained that

$E(R)=\sqrt{\dfrac{n}{4-\dfrac{2}{n}}}$ and $\bar{X}_n \le \mu + \dfrac{(n-1)\sigma}{\sqrt{2n-1}}$.

Assuming that X_n is local record, let $\bar{X}_n = E(X_{(n)} - \frac{R}{n}) = X_{(n)} - \frac{1}{n}\sqrt{\frac{n}{4-\frac{2}{n}}}$ and let variance of X_{UV} be an estimator of population variance $s^2 = \frac{R^2}{12}$, where $R = X_{(n)} - X_{(1)}$.

Then, $X_{(n)} = 1857.76, X_{(1)} = 5.79, R = 1851.96, n = 15, \quad \bar{X}_n = 1857.62, \ \hat{\sigma} = 534.61.$

Therefore, the Gumble mean is $1857.62 - \frac{(14)(534.61)}{\sqrt{29}} = 467.77 \geq \mu$.

If $Y = \ln(X)$ *is* $N(\mu, \sigma^2)$ then, X is $LN(\mu_1, \sigma_1^2)$. We note that the parameters of Y and X are related by $\mu_1 = \exp(\mu + .5\sigma^2)$, $\sigma_1^2 = [\exp(2\mu + \sigma^2)][\exp(\sigma^2) - 1]$ and $CV = \frac{\sigma_1}{\mu_1} = \sqrt{\exp(\sigma^2) - 1}$.

Setting $\hat{\mu}_1 = 467.77, \hat{\sigma}_1 = 534.61$ and solving for μ and σ, we get that $\hat{\mu} = 5.73$ and $\hat{\sigma} = 0.91$.

We can now compute the lowest 95% percentile of the distribution as $\exp(5.73 + 1.65 * 0.91)$ or $95\% \geq 1380.22$. The UB estimates above lowest 95% percentile are given by Log – Normal Jackknife at 1534.94 and Log – Normal Chebychev at 2798.63. The H – UCL value of 4590.27 is above the true 95% percentile distribution.

The next table shows the classical and Log - Normal Estimates of the Upper Bound.

Method	Prediction	Method	Prediction
Std.Jackknife	705.88	LN.Jackknife	1534.94
Std.Bootstrap	882.82	LN.Bootstrap	1363.26
Pivotal Bootstrap	977.82	LN.Chebychev	2798.63
CLT	887.81	H-UCL	4570.27
Adjusted CLT	919.81		
Std.Chebychev	1327.75		

Table 3.3.2 Classical and Log-Normal Estimates of the Upper Bound.

CHAPTER 4

CONCLUSION

The methods described in this thesis consider method of estimating the population mean of normal distribution, given the smallest, the largest statistic and sample size. In a way, the six samples developed are similar to lower and upper records, with important difference that the CDF is unknown and seven artificial random variables are generated for the purpose of estimating the population mean.

APPENDIX I

Below is a S+ computer program for generating confidence intervals based on inequality 12).

```
cat (" This is SLM confidence interval test program")

n1 <- as.integer(readline(prompt="Input sample size"))

c1 <- as.integer(readline(prompt="How many intervals ?"))

i <- as.integer

m <- as.double(readline(prompt="Input sample mean"))

sd <- as.double(readline(prompt="Input standart deviation"))

k <- as.double(readline(prompt="Input Chebychev constant"))

test1 <- array(dim=c(c1,7)) #storage array for int.number,low limit, high limit

plot1 <- array(dim=c(2*c1,2)) # array for plotting int.number,low limit, high

p <- as.double(readline(prompt="Input T-sta confidence level"))#

a1 <- as.double # Confidence level

a1 <- 1-p

meancount <- as.integer

meancount <- 0

tv <- qt(p,n1-1)
```

```
xbaru <- as.double# Gumbel up sample mean i.e k(Yn+Y1)/2
xbard <- as.double # down sample mean
sdev <- as.double # Gumbel standard deviation
scount <- as.double # Success counter how many interval bracket population
mean
#############Counters
count1 <- as.integer# Counts how many lower limit are > mean
count2 <- as.integer# Counts how many lower limit are > mean
count1 <- 0
count2 <- 0
m1 <- as.double # slope of idealized sample
sq<- as.double # sum of squares of idealized sample
#################################################################
for(i in 1:c1){
   v1 <- rnorm(n1,m,sd)
   v1 <- sort(v1)
   Y1 <- min(v1)
   YN <- max(v1)
   r <- YN-Y1
   m1 <- r/(n1-1)
   #if(mean(v1)>(Y1+YN)/2){
   #   meancount <- meancount+1}
   #sdev   <- .5*r*((.25+3/(2*n1))^.5)
```

```
# Old MeanTest9 sum of squares and x bar
#sq <- (n1*Y1^2+.16666*m1^2*(n1-1)*n1*(2*n1-1)+m1*Y1*(n1-1)*n1)/n1
  #xbar <- (sq-sdev^2)^.5#  New Estimate of sample  mean
##########################################################################
New Mean
sdev   <- .322*r#.14435*r old sdeviationnew sdev r^2/12+r^2/48
xbaru <- (Y1+YN)/2+.125*sdev
xbard <- (Y1+YN)/2-.125*sdev
   LL  <- xbard-(tv*sdev)/(n1-1)^.5# Gumbel  intervals
   HL  <- xbaru+(tv*sdev)/(n1-1)^.5
   #LL  <- YN-((YN-Y1)/n1)-(n1-1)/(2*n1-1)^.5
   #HL  <- Y1+((YN-Y1)/n1)+(n1-1)/(2*n1-1)^.5
test1[i,1] <- i
test1[i,2] <- LL
test1[i,3] <- HL
test1[i,4] <- (m-LL)
test1[i,5] <- (HL-m)
test1[i,6] <- (Y1)
test1[i,7] <- (YN)
}
for(i in 1:c1){
   if(test1[i,4] < 0){
   test1[i,4] <- -1
```

```
count1 <- count1 +1}
    else if(test1[i,4]>0 )
    {
    test1[i,4] <- 0
}
}
for(i in 1:c1){
  if(test1[i,5] < 0){
  test1[i,5] <- -1
  count2 <- count2+1
  }
  else if(test1[i,5]>0 )
    {
    test1[i,5] <- 0
  }
}
print(test1)
print(count1)
print(count2)
print(a1)
print(xbaru)
print(xbard)
print(sdev)
```

```
scount <-(1-(count1+count2)/c1)
print(scount)
print (meancount/c1)
    for(i in 1:c1){
        plot1[2*i-1,2]<- test1[i,2]
        plot1[2*i,2] <- test1[i,3]
    }
    iv <- rep(1:c1,each=2)
    plot(iv,plot1[,2],type="p",main="500 SLM C.I,m=0,sd=1,k=2,cl=0.95 ")
# This is the latest program using  Average deviation 07/31/02
```

APPENDIX II

Variable with Average Variance (X_{AV}), $X_{AV} = \frac{(X_{LV} + X_{SV})}{2}$.

The expected value is $E(X_{AV}) = E[\frac{X_{LV} + X_{SV}}{2}] = \frac{X_{(n)} + X_{(1)}}{2}$.

The variance is given by

$$\mathrm{var}[\frac{X_{LV} + X_{SV}}{2}] = \frac{1}{4}\mathrm{var}(X_{LV}) + \frac{1}{4}\mathrm{var}(X_{SV}) + 2(\frac{1}{4})co\mathrm{var}(X_{LV}X_{SV}).$$

The covariance $\mathrm{cov}(X_{LV}, X_{SV})$ is

$$\mathrm{cov}(X_{LV}, X_{SV}) = \frac{4X^2_{(1)} + [X_{(1)} + X_{(n)}]^2(n-2) + 4X^2_{(n)}}{4n} - [\frac{X_{(1)} + X_{(n)}}{2}]^2,$$

$$\mathrm{cov}(X_{LV}, X_{SV}) = \frac{R^2}{2n}.$$

The variance of X_{AV} is then given by

$$\mathrm{var}(X_{AV}) = \frac{1}{4}(\frac{R^2}{4}) + \frac{1}{4}(\frac{R^2}{2n}) + \frac{1}{2}(\frac{R^2}{2n}),$$

$$\mathrm{var}(X_{AV}) = \frac{R^2}{16} + \frac{3R^2}{8n}.$$

If we take the limit as n goes to infinity, we get the expression of $\mathrm{var}(X_{AV})$ for

large samples, $Lim_{n\to\infty}(\frac{R^2}{16} + \frac{3R^2}{8n}) = \frac{R^2}{16}$.

APPENDIX III

DJIA monthly H/L April 1991-April 2001

April 2001 High 10043/ Low 9385.4 Close 9485.7

April 1991 High 3030.5/ Low 2848.5 Close 2887.9

10043.0	9385.4	9485.7
10940.5	9047.6	9878.8
11140.1	10225.1	10495.3
11224.4	10325.7	10887.4
11044.7	10158.2	10788.0
11152.0	10204.8	10414.5
11108.8	9571.4	10971.1
11518.8	10439.3	10650.9
11416.0	10428.6	11215.1
10980.3	10303.3	10522.0
11013.0	10161.5	10447.9
11086.7	10163.2	10522.3
11600.4	10128.6	10733.9
11311.3	9611.8	10921.9
11228.4	9760.4	10128.3
11908.5	10610.4	10940.5

11658.7	10798.1	11497.1
11195.3	10449.4	10877.8
10883.1	9884.2	10729.9
11218.4	10055.2	10337.0
11428.9	10487.3	10829.3
11321.6	10595.0	10655.2
11120.2	10334.4	10970.8
11244.4	10373.0	10559.7
11072.3	9707.9	10789.0
10158.6	9163.4	9786.2
9662.8	9025.4	9306.6
9759.4	8994.3	9358.8
9390.8	8610.6	9181.4
9458.0	8573.6	9116.5
8718.3	7399.8	8592.1
8253.8	7379.7	7842.6
8948.2	7517.7	7539.1
9412.6	8786.5	8883.3
9155.0	8524.5	8952.0
9312.0	8761.0	8900.0
9287.3	8715.6	9063.4
8997.1	8377.3	8799.8
8616.7	7987.5	8545.7

8072.9	7391.6	7906.5
8209.6	7563.2	7908.3
7934.5	7334.8	7823.1
8218.3	6933.0	7442.1
8078.4	7556.2	7945.3
8340.1	7580.9	7622.4
8329.0	7613.5	8222.6
7868.4	7214.3	7672.8
7430.2	6891.4	7331.0
7081.2	6315.8	7009.0
7158.3	6532.5	6583.5
7112.9	6683.4	6877.7
6953.5	6319.0	6813.1
6624.0	6206.8	6448.3
6606.3	5975.3	6521.7
6162.8	5833.7	6029.4
5952.1	5550.4	5882.2
5762.0	5507.8	5616.2
5769.9	5170.1	5528.9
5770.6	5559.7	5654.6
5833.0	5327.7	5643.2
5737.1	5382.7	5569.1
5755.9	5395.3	5587.1

5693.4	5319.4	5485.6
5433.2	5000.1	5395.3
5266.7	5016.7	5117.1
5143.1	4719.7	5074.5
4845.1	4638.4	4755.5
4839.5	4594.7	4789.1
4772.6	4552.8	4610.6
4768.0	4530.3	4708.5
4614.2	4394.6	4556.1
4480.7	4278.7	4465.1
4348.9	4129.7	4321.3
4213.7	3935.3	4157.7
4034.6	3809.2	4011.1
3955.6	3794.4	3843.9
3882.2	3639.0	3834.4
3919.9	3612.1	3739.2
3958.3	3736.2	3908.1
3972.7	3804.5	3843.2
3954.5	3722.4	3913.4
3782.6	3611.0	3764.5
3839.9	3603.9	3625.0
3788.8	3609.7	3758.4
3733.2	3520.5	3681.7

3911.8	3544.1	3636.0
3998.1	3811.8	3832.0
4002.8	3715.2	3978.4
3818.9	3673.3	3754.1
3749.9	3585.9	3684.0
3713.6	3541.7	3680.6
3665.5	3501.5	3555.1
3681.7	3523.5	3651.3
3604.9	3443.3	3539.5
3577.3	3445.8	3516.1
3582.2	3402.4	3527.4
3499.4	3338.4	3427.6
3497.3	3334.1	3435.1
3472.9	3262.5	3370.8
3338.1	3219.3	3310.0
3364.9	3229.8	3301.1
3326.5	3176.8	3305.2
3291.4	3087.4	3226.3
3391.4	3226.6	3271.7
3413.2	3200.9	3257.4
3414.9	3255.5	3393.8
3435.2	3242.3	3318.5
3434.0	3316.6	3396.9

3388.0	3141.8	3359.1
3318.4	3176.2	3235.5
3307.5	3193.4	3267.7
3313.5	3119.9	3223.4
3204.6	2832.3	3168.8
3091.9	2861.1	2894.7
3091.0	2925.5	3069.1
3066.6	2963.1	3016.8
3068.7	2835.4	3043.6
3039.6	2897.4	3024.8
3057.5	2879.3	2906.8
3044.5	2834.5	3027.5
3030.5	2848.5	2887.9

BIBLIOGRAPHY

Arnold, Barry. Records. New York City, New York. John Wiley & Sons, Inc. 1998.

Gumbel, E.J. The Maxima of the Mean Largest and of the Range. Annals of Mathematical Statistics, Volume 25, p.76-84, Issue 1. 1954.

Hays, William L. Statistics. 3rd edition. New York City, New York. Holt, Reinhart, and Wilson, Inc. 1981.

Hogg, Robert V. Introduction to Mathematical Statistics. 5th edition. Englewood Hills, New Jersey. 1995.

Johnson, J. Econometric Methods. 2nd edition. New York City, New York. McGraw-Hill Book Company. 1963.

Selby, Samuel M. Standard Mathematical Tables. 18th edition. Cleveland, Ohio. The Chemical Rubber Company. 1964.

Singh, A.K. The Log-normal Distribution In environmental applications. EPA Technology Support Center Issue, 1997.

Venables, W.N. Modern Applied Statistics with S-PLUS. 3rd edition. New York City, New York. Springer-Verlag, Inc. 1999.

VITA

Graduate College
University of Nevada, Las Vegas

Vladimir Emil Minev

Home Address:
9625 Sterling Ridge Avenue
Las Vegas, Nevada 89129

Degree:
Bachelor of Science, 1999
Department of Mathematical Sciences
University of Nevada, Las Vegas

Thesis Title: Estimation of the Population Mean Using Two Extremes in Order Statistics.

Thesis Examination Committee:
Chairperson, Professor Hokwon Cho, Ph. D.
Committee Member, Professor Ashok Singh, Ph. D.
Committee Member, Professor Rohan Dalpatadu, Ph. D.
Graduate Faculty Representative, Professor William Wells, Ph. D.