Variance estimation in repeated samples of size one

Abdul Wajid Rana

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Rana, Abdul Wajid, "Variance estimation in repeated samples of size one " (1994). Retrospective Theses and Dissertations. 11310.
https://lib.dr.iastate.edu/rtd/11310

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Variance estimation in repeated samples of size one

Rana, Abdul Wajid, Ph.D.
Iowa State University, 1994
Variance estimation in repeated samples of size one

by

Abdul Wajid Rana

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
Department: Statistics
Major: Statistics

Approved:
Signature was redacted for privacy.
In Charge of Major Work
Signature was redacted for privacy.
For the Major Department
Signature was redacted for privacy.
For the Graduate College

Iowa State University
Ames, Iowa
1994

Copyright © Abdul Wajid Rana, 1994. All rights reserved.
Dedicated to
my parents,
my wife, Samina and
my lovely sons Asim, Danish and Haseeb.
TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION ........................................... 1
   The Problem ...................................................... 1
   Model 1 : M1 ..................................................... 2
   Model 2 : M2 ..................................................... 3
   Model 3 : M3 ..................................................... 5
   Model 4 : M4 ..................................................... 7
   Model 5 : M5 ..................................................... 7
   Methods of Estimation and Organization of the Dissertation ....... 8

CHAPTER 2. ESTIMATORS BASED ON THE MOVING RANGE
   AND MEAN MOVING RANGE .................................... 10
   Estimation of $\sigma_e$ Based on the Moving Range of 2 Consecutive Observations 10
   Moments of $\hat{\sigma}_e$ when Population Means are not the Same ........ 13
   Estimation of $\hat{\sigma}_e$ Based on the Mean Moving Range ................ 15
   Properties of $\overline{M}_n/d_2$ under Model M1 ...................... 16
   Asymptotic Properties of the Mean Moving Range ..................... 21
   Properties of $\hat{\sigma}_e$ and Bias Correction under Model M2 .......... 26
   Properties of $\hat{\sigma}_e$ and Bias Correction under Model M3 .......... 30
iv

CHAPTER 3. MAXIMUM LIKELIHOOD ESTIMATORS

Maximum Likelihood Estimation of the Variance $\sigma^2$

Estimation under Model M1

Estimation under Model M2

Estimation under Model M3

Estimation under Model M5

Some Properties of ML Estimators under Alternative Models

The Mean of $\sigma^2_{\text{ML}3}$ under Models M4 and M5

The Mean of $\sigma^2_{\text{ML}5}$ under Models M2 and M3

CHAPTER 4. BAYESIAN ESTIMATION

The Gibbs Sampling Algorithm

An Example of Gibbs Sampling for a Simple Two Parameter Normal Model

Bayes Estimation under Model M2

Bayes Estimation under Model M3

Bayes Estimation under Model M4

Bayes Estimation under Model M5

CHAPTER 5. MONTE CARLO STUDIES

Description of the Models Used in the Simulation

Simulation Parameters for Model M1

Simulation Parameters for Model M2

Simulation Parameters for Model M3

Simulation Parameters for Model M4

Simulation Parameters for Model M5
Prior Distributions Used to Produce Bayes Estimators .................. 55
Description of Estimators Used ........................................... 57
Analysis and Conclusions .................................................. 61
  Overall Comparisons .................................................. 61
  Comparisons under $M_{1(1)}$ .......................................... 64
  Comparisons under $M_{1(2)}$ .......................................... 64
  Comparisons under $M_{1(3)}$ .......................................... 64
  Comparisons under $M_{1(4)}, M_{1(5)}$ ................................. 67
  Comparisons under $M_{1(6)}$ .......................................... 67
  Comparisons under $M_{2(1)}, M_{2(2)}, M_{2(3)}$ .................... 67
  Comparisons under $M_{2(4)}, M_{2(5)}, M_{2(6)}$ .................... 67
  Comparisons under $M_{2(7)}, M_{2(8)}, M_{2(9)}$ .................... 74
  Comparisons under $M_{2(10)}, M_{2(11)}, M_{2(12)}$ ............... 74
  Comparisons under $M_{2(13)}, M_{2(14)}, M_{2(15)}$ ............... 74
  Comparisons under $M_{3(1)}, M_{3(2)}, M_{3(3)}$ .................... 74
  Comparisons under $M_{4(1)}, M_{4(2)}, M_{4(3)}$ .................... 83
  Comparisons under $M_5$ .............................................. 83
Analysis of $(\overline{M}_n/d_2)$ as an Estimator of Standard Deviation ...... 83
Comparison of MR and ML2 .................................................. 83

CHAPTER 6. SUMMARY AND RECOMMENDATIONS ......................... 92
  Summary ........................................................................... 92
  Recommendations ......................................................... 94
REFERENCES ................................................................. 95
LIST OF TABLES

Table 2.1: The Mean of $\sigma_e$ when $\sigma_e^2 = 1$ ........................................... 14
Table 2.2: Variance of $W = |X_1 - X_2|$ and $\sigma_e$ when $\sigma_e^2 = 1$ .............. 16
Table 2.3: Numerical Values of $E|X_{j-1} - X_j||X_j - X_{j+1}|$ when $\sigma_e^2 = 1$
(Standard Errors in Parentheses) ............................................................... 19
Table 2.4: Numerical Values of $\text{Corr}(|X_{j-1} - X_j|, |X_j - X_{j+1}|)$ .............. 20
Table 2.5: Values of $\text{Corr}(|X_1 - X_2|, |X_2 - X_3|)$ for Model M2.1 .......... 30
Table 5.1: Models M2 Used in the Simulations ................................................. 55
LIST OF FIGURES

Figure 5.1: Overall Comparison of Estimators $n = 20$ .......................... 62
Figure 5.2: Overall Comparison of Estimators $n = 50$ .......................... 63
Figure 5.3: Comparison of Estimators for Model M1 $n = 20$ ................. 65
Figure 5.4: Comparison of Estimators for Model M1 $n = 50$ ................. 66
Figure 5.5: Comparison of Estimators for Model M1 $n = 20$ ................. 68
Figure 5.6: Comparison of Estimators for Model M1 $n = 50$ ................. 69
Figure 5.7: Comparison of Estimators for Model M2 $n = 20$ ................. 70
Figure 5.8: Comparison of Estimators for Model M2 $n = 50$ ................. 71
Figure 5.9: Comparison of Estimators for Model M2 $n = 20$ ................. 72
Figure 5.10: Comparison of Estimators for Model M2 $n = 50$ ............... 73
Figure 5.11: Comparison of Estimators for Model M2 $n = 20$ ............... 75
Figure 5.12: Comparison of Estimators for Model M2 $n = 50$ ............... 76
Figure 5.13: Comparison of Estimators for Model M2 $n = 20$ ............... 77
Figure 5.14: Comparison of Estimators for Model M2 $n = 50$ ............... 78
Figure 5.15: Comparison of Estimators for Model M2 $n = 20$ ............... 79
Figure 5.16: Comparison of Estimators for Model M2 $n = 50$ ............... 80
Figure 5.17: Comparison of Estimators for Model M3 $n = 20$ ............... 81
Figure 5.18: Comparison of Estimators for Model M3 $n = 50$ ............... 82
Figure 5.19: Comparison of Estimators for Model M4 $n = 20$ ........ 84
Figure 5.20: Comparison of Estimators for Model M4 $n = 50$ ........ 85
Figure 5.21: Comparison of Estimators for Model M5 $n = 20$ ........ 86
Figure 5.22: Comparison of Estimators for Model M5 $n = 50$ ........ 87
Figure 5.23: $\overline{M}/d_2$ as an Estimator of Standard Deviation ........ 88
Figure 5.24: Comparison of MR and ML2 Estimators $n = 20$ ........ 89
Figure 5.25: Comparison of MR and ML2 Estimators $n = 50$ ........ 90
CHAPTER 1. INTRODUCTION

Estimation of a standard deviation (or variance) is an important statistical problem. In context of Statistical Quality Control (SQC), a process standard deviation measures baseline or "common cause" variation. Its estimation is important for purposes of process characterization and subsequent setting of control limits on various process performance measures. This dissertation concerns the estimation of process standard deviation in the absence of replication.

The Problem

We consider the simple model

\[ X_i = \mu_i + \epsilon_i \quad i = 1, 2, \ldots, n \]

where

\[ \epsilon_i \sim \text{IID}(0, \sigma^2_\epsilon) \]
so that given $\mu_1, \mu_2, \ldots, \mu_n$, the variables $X_i$ are independent,

$$X_i \sim N(\mu_i, \sigma^2).$$

In Statistical Quality Control, it is sometimes necessary to periodically take single observations to study a characteristic of a process. This can be modeled as a scenario of taking repeated samples of size $m=1$ from different populations ordered in time. Since the location parameters of these populations may differ significantly, the estimation of variance, $\sigma^2$, or standard deviation poses a problem to be explored.

In practice, estimators of variance (or standard deviation) that strictly speaking are appropriate only supposing that observations are iid, are sometimes used. This is done even though these assumptions are likely to be violated because the population means are changing. The purpose of this study is to estimate the process variance under some simple probability models for $\mu_1, \mu_2, \ldots, \mu_n$ and propose some corrective measures to apply to estimators based on iid assumptions.

Of course, any number of models are possible, but we consider the following models for $X = [X_1, X_2, \ldots, X_n]^T$.

**Model 1:** M1

This is the basic model introduced above. That is, we assume that the $X_i$'s are independently and normally distributed but need not have the same means. i.e.

$$X_i = \mu_i + \epsilon_i \quad i = 1, 2, \ldots, n$$

where
1. $X_i$ is the observed value from $i$th population.

2. $\mu_i$ is the mean of the $i$th population, and

3. $\epsilon_i$'s are $IIN(0,\sigma_\xi^2)$.

In other words, either thinking of the $\mu_i$ as fixed, or thinking conditional on $\mu_1, \mu_2, \ldots, \mu_n$

$$X_i \sim IN(\mu_i, \sigma_\xi^2).$$

Model 2 : M2

To the basic model assumptions M1

$$X_i = \mu_i + \epsilon_i \quad i = 1, 2, \ldots, n$$

we add the assumption that the population means $\mu_i$ follow a first order autoregressive model, AR(1). That is

$$\mu_i = \rho \mu_{i-1} + \nu_i \quad i = 1, 2, \ldots, n$$

where

1. $\mu_i$ is the mean of $i$th population,

2. $\rho = Corr(\mu_{i-1}, \mu_i)$,

3. $\epsilon_i \sim IIN(0, \sigma_\xi^2)$,

4. $\nu_i \sim IIN(0, \sigma_\nu^2)$.
and the $\epsilon_i$ and $\nu_i$ sequences are independent. In this model

$$\mu_i = \rho^i \mu_0 + \sum_{j=1}^{i} \rho^{(i-j)} \nu_j$$

and

$$X_i = \rho^i \mu_0 + \sum_{j=1}^{i} \rho^{(i-j)} \nu_j + \epsilon_i.$$  

In other words,

$$X \sim N(\Psi \mu_0, \Sigma)$$

where $\Psi = [\rho, \rho^2, \ldots, \rho^n]'$, and $\Sigma$, the variance covariance matrix of $X$, has the form

$$\Sigma = \sigma^2_e I_n + \sigma^2_\nu V_n$$

where $I_n$ is the $n \times n$ identity matrix and

$$V_n = \begin{bmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{(n-1)} \\
\rho & (1 + \rho^2) & \rho (1 + \rho^2) & \cdots & \rho^{(n-2)} (1 + \rho^2) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\rho^{(n-1)} & \rho^{(n-2)} (1 + \rho^2) & \rho^{(n-3)} (1 + \rho^2 + \rho^4) & \cdots & \sum_{j=0}^{(n-1)} \rho^{2j}
\end{bmatrix}$$

Note that the variance covariance matrix, $\Sigma$, has linear structure.

**Model 2.1 : M2.1** A useful specialization of model M2 is the case where $\rho = 1$, i.e. where the $\mu_i$'s follow a random walk model,

$$\mu_i = \mu_{i-1} + \nu_i \quad i = 1, 2, \ldots, n.$$  

(1.2)
Under this assumption

\[ \mu_i = \mu_0 + \sum_{j=1}^{i} \nu_j \]  

(1.3)

and

\[ X_i = \mu_0 + \sum_{j=1}^{i} \nu_j + \epsilon_i \]  

(1.4)

Again in other words,

\[ X \sim N(\mu_0, \Sigma) \]  

(1.5)

where \( \mathbf{1} = [1, 1, \ldots, 1]' \) and \( \Sigma = \sigma^2 \mathbf{I}_n + \sigma^2 \mathbf{V}_n \), for

\[ \mathbf{V}_n = \begin{bmatrix}
1 & 1 & 1 & \ldots & 1 \\
1 & 2 & 2 & \ldots & 2 \\
1 & 2 & 3 & \ldots & 3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 2 & 3 & \ldots & n
\end{bmatrix} \]  

(1.6)

The variance covariance matrix, \( \Sigma \), continues to have a linear structure.

Model 3 : M3

To the basic model assumptions M1 we might add the assumption that there is linear drift in \( \mu \)'s i.e.
\begin{equation}
\mu_i = \mu_{i-1} + h
\end{equation}

where

1. \( \mu_i \) is the \( i \)'th population mean, and

2. \( h \) is the drift.

For observed value, \( X_i \), from the \( i \)'th population,

\[ X_i = \mu_i + \epsilon_i \quad i = 1, 2, \ldots, n \]

becomes

\[ X_i = \mu_0 + i h + \epsilon_i \]

with

\[ \epsilon_i \sim N(0, \sigma^2_{\epsilon}). \]

In other words,

\[ X \sim N(\mu_0 + k h, \sigma^2_{\epsilon} I_n) \quad (1.8) \]

for \( k = [1, 2, \ldots, n]' \).
Model 4 : M4

We will also consider a random effects model. To the basic assumptions M1,

\[ X_i = \mu_i + \epsilon_i \quad i = 1, 2, \ldots, n \]

we add the assumptions that the \( \mu_i \) are themselves iid normal random variables independent of the \( \epsilon_i \) sequence. That is for unknown parameters \( \mu_0 \) and \( \sigma^2_{\mu} \) we suppose that the means and errors have the structure

\[ \mu_i \sim IN(\mu_0, \sigma^2_{\mu}). \]

Model 5 : M5

The special case of model M1 when all the \( \mu_i \) are the same, is the \( \rho = 1 \) and \( \sigma^2_{\nu} = 0 \) case of M2, the \( h = 0 \) case of M3, and the \( \sigma^2_{\mu} = 0 \) case of M4. That is, we will also consider the model

\[ X_i = \mu + \epsilon_i \quad i = 1, 2, \ldots, n \]

where

1. \( \mu \) is a constant and

2. \( \epsilon_i \sim IN(0, \sigma^2_{\epsilon}). \)

That is,

\[ X_i \sim IN(\mu, \sigma^2_{\epsilon}). \]
Methods of Estimation and Organization of the Dissertation

Under the different sets of assumptions described above, the following three principles of estimation will be used to develop estimators for $\sigma_E^2$, the parameter of model M1 of primary interest in this dissertation:

1. the Method of Moments,

2. the Maximum Likelihood Method, and

3. the Bayesian Method.

In Chapter 2, we consider an estimation method common in SQC circles, the estimation of standard deviation based on the mean moving range. $\sigma_e = \frac{\overline{M}_n}{d_2}$, where $\overline{M}_n$ is a mean moving range and $d_2$ a standard "control chart constant," is essentially a method of moments estimator and is unbiased for $\sigma_e$ under the $iid$ Model, M5. But if this estimator is to be used under other models, it needs adjustment or corrective measures.

In Chapter 3, we consider the maximum likelihood estimation of $\sigma_E^2$ under different models. The consequences of using the maximum likelihood estimator of $\sigma_E^2$ for the $iid$ model M5 under other models are discussed. Appropriate corrective measures for this estimator if it is to be used under other models are considered.

In Chapter 4, the Bayesian estimation of $\sigma_E^2$ is considered under various models. Since, for the more complicated models, closed forms of posterior distributions do not seem to exist, an approach based on Gibbs sampler has been used to calculate the Bayes estimates.

Chapter 5 describes a Monte-Carlo study of the performance of all the estimators considered in Chapters 2 through 4 under various different models. Comparisons are
based on the MSE criteria.

Chapter 6 consists of summary of the study and some recommendations to practitioners.
CHAPTER 2. ESTIMATORS BASED ON THE MOVING RANGE AND MEAN MOVING RANGE

The range of a sample of size larger than 1 is widely used as a measure of dispersion in Statistical Quality Control. It is a quantity that is often plotted on control charts, and the mean of several ranges is often used as the basis of a simple estimate of standard deviation. If two or more values are observed from the same normal population, an unbiased estimator of standard deviation can be found by dividing the range by the "control chart constant," \( d_2 \). (And several such estimators are often averaged.)

We now describe how, in the unreplicated situation considered here, an estimator of \( \sigma_e \) can be produced based on the mean of moving ranges of consecutive observations. A discussion of the consequences of using the typical estimator based on the mean moving range (which is appropriate under \( iid \) assumptions) when the observations actually follow other models is given. Corrections to the usual estimator are proposed if it is to be used under other models.

**Estimation of \( \sigma_e \) Based on the Moving Range of 2 Consecutive Observations**

Suppose, in two consecutive samples of size \( m = 1 \), we have
\[ X_1 \sim N(\mu_1, \sigma_1^2) \]
\[ X_2 \sim N(\mu_2, \sigma_2^2) \]

where, \( X_1 \) and \( X_2 \) are statistically independent. The observations have same variance but different means, \( \mu_1 \) and \( \mu_2 \). For this \( n = 2 \) situation, the moving range is defined as the absolute difference between \( X_1 \) and \( X_2 \) i.e.,

\[ W = |X_1 - X_2|. \]

The expected value of the moving range, \( W \), is

\[
EW = E |X_1 - X_2| = E \left| \frac{(X_1 - X_2) - (\mu_1 - \mu_2)}{\sqrt{2\sigma_e}} + \frac{\mu_1 - \mu_2}{\sqrt{2\sigma_e}} \right| \sqrt{2\sigma_e} = E |Z + \delta_{12}|(\sqrt{2\sigma_e})
\]

where \( Z \sim N(0,1) \) and \( \delta_{12} = \frac{(\mu_1 - \mu_2)}{\sqrt{2\sigma_e}} \). Now,

\[
E |Z + \delta_{12}| = \int_{-\infty}^{+\infty} \frac{|z + \delta_{12}|}{\sqrt{2\pi}} e^{-z^2/2} dz
\]

\[
= \frac{-1}{\sqrt{2\pi}} \int_{-\delta_{12}}^{\delta_{12}} (z + \delta_{12})e^{-z^2/2} dz + \frac{1}{\sqrt{2\pi}} \int_{-\delta_{12}}^{+\infty} (z + \delta_{12})e^{-z^2/2} dz
\]

\[
= \frac{-1}{\sqrt{2\pi}} \int_{-\delta_{12}}^{\delta_{12}} ze^{-z^2/2} dz - \delta_{12} \int_{-\delta_{12}}^{\delta_{12}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz + \frac{1}{\sqrt{2\pi}} \int_{-\delta_{12}}^{+\infty} ze^{-z^2/2} dz + \delta_{12} \int_{-\delta_{12}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz
\]

\[
= \sqrt{\frac{2}{\pi}} e^{-\delta_{12}^2/2} + \delta_{12} [P(Z > -\delta_{12}) - P(Z < -\delta_{12})]
\]
\[\begin{align*}
\hat{\sigma}_e &= \begin{cases} 
\sqrt{\frac{\pi}{2}} e^{-\frac{\delta_{12}^2}{2}} \delta_{12} P(|Z| < \delta_{12}) & \text{if } \delta_{12} > 0 \\
\sqrt{\frac{\pi}{2}} e^{-\frac{\delta_{12}^2}{2}} + |\delta_{12}| P(|Z| < |\delta_{12}|) & \text{if } \delta_{12} < 0
\end{cases} \\
&= 2\phi(\delta_{12}) + |\delta_{12}| P(|Z| < |\delta_{12}|).
\end{align*}\]

Hence,
\[EW = \sqrt{2}\sigma_e \left[2\phi(\delta_{12}) + |\delta_{12}| P(|Z| < |\delta_{12}|)\right],\]
and
\[E\left[\frac{W}{\sigma_e}\right] = \sqrt{2} [2\phi(\delta_{12}) + |\delta_{12}| P(|Z| < |\delta_{12}|)].\]

Under Model M5, (when \(\mu_1 = \mu_2\), \(\delta_{12} = 0\) and \(X_1\) and \(X_2\) are iid, so
\[\phi(\delta_{12}) = \frac{1}{\sqrt{2\pi}}\]
and
\[E\left[\frac{W}{\sigma_e}\right] = \frac{2}{\sqrt{\pi}} = 1.1281521496\]

This is the value of \(d_2\) for samples of size 2.

From the above, when \(X_1\) and \(X_2\) are iid (and normally distributed), an unbiased estimator of \(\sigma_e\) is
\[\hat{\sigma}_e = \frac{W}{d_2}. \tag{2.1}\]
In those cases (rare in practice) that \( \delta_{12} \) may not be assumed to be 0, but is known, an unbiased estimator of the standard deviation, \( \sigma_e \), is

\[
\hat{\sigma}_e = \frac{W}{\sqrt{2 \{2\phi(\delta_{12}) + \delta_{12}P(|Z| < |\delta_{12}|)\}}} \tag{2.2}
\]

If \( |\delta_{12}| \neq 0 \), use of \( \hat{\sigma}_e \) tends to overestimate \( \sigma_e \) by a (multiplicative) factor of

\[
\sqrt{\frac{\pi}{2} \{2\phi(\delta_{12}) + \delta_{12}P(|Z| < |\delta_{12}|)\}}. \tag{2.3}
\]

Another way of saying this is that when \( \delta_{12} \) is not zero, \( \hat{\sigma}_e \) needs to be corrected by multiplying by

\[
\kappa = \frac{1}{\sqrt{\frac{\pi}{2} \{2\phi(\delta_{12}) + \delta_{12}P(|Z| < |\delta_{12}|)\}}}, \tag{2.4}
\]

so that \( \kappa \hat{\sigma}_e \) is an unbiased estimator of \( \sigma_e \).

**Moments of \( \hat{\sigma}_e \) when Population Means are not the Same**

The Mean of \( \hat{\sigma}_e \). Display (2.3) shows that the extent of systematic overestimation of \( \sigma_e \) by \( \hat{\sigma}_e \) depends on the magnitude of difference between population means. Table 2.1 gives values of the quantity (2.3). In Table 2.1, the first column shows the absolute difference in population means. Column 2 gives the expected value of the moving range, \(|X_1 - X_2|\), while the third column shows the multiplicative factor by which \( \hat{\sigma}_e \) overestimates \( \sigma_e \). We see that the magnitude of overestimation increases with the absolute difference between population means. The systematic overestimation is about 6% if the absolute difference, \( |\mu_1 - \mu_2| \), is 0.5, i.e. half the value of \( \sigma_e \). When \( |\mu_1 - \mu_2| = \sigma_e \), \( \hat{\sigma}_e \) overestimates \( \sigma_e \) by 24%. And a systematic overestimation of 86% results when the absolute difference in population means is twice the standard deviation.
Table 2.1: The Mean of $\hat{\sigma}_e$ when $\sigma_e^2 = 1$

| $|\mu_1 - \mu_2|$ | $EW = E|X_1 - X_2|$ | $E\hat{\sigma}_e = E(W/d_2)$ |
|-----------------|-----------------|-----------------|
| 0.00000         | 1.128152        | 1.00000         |
| 0.10000         | 1.13972         | 1.00249         |
| 0.50000         | 1.19764         | 1.06188         |
| 1.00000         | 1.399015        | 1.24017         |
| 1.50000         | 1.709535        | 1.51534         |
| 2.00000         | 2.100425        | 1.86182         |
| 2.50000         | 2.543723        | 2.25477         |
| 3.00000         | 3.017221        | 2.67448         |
| 3.50000         | 3.506115        | 3.10783         |
| 4.00000         | 4.001951        | 3.54351         |
| 4.50000         | 4.500558        | 3.98931         |
| 5.00000         | 5.000143        | 4.43215         |

For large $|\mu_1 - \mu_2|$, the systematic overestimation is (in fractional terms) approximately $\frac{|\mu_1 - \mu_2|}{\sigma_e d_2}$, in the sense that

$$E\frac{\hat{\sigma}_e}{|\mu_1 - \mu_2|} = \frac{[2\Phi(\delta_{12}) + |\delta_{12}|P(\delta_{12} < \delta_{12})]}{d_2|\delta_{12}|}$$

which clearly has the limit $\frac{1}{d_2}$ as $|\delta_{12}| \to \infty$.

The Variance of $\hat{\sigma}_e$. The variance of the moving range $W = |X_1 - X_2|$ is

$$Var(W) = E|X_1 - X_2|^2 - (E|X_1 - X_2|)^2.$$  

Here

$$E|X_1 - X_2|^2 = Var(X_1 - X_2) + (\mu_1 - \mu_2)^2$$

$$= 2\sigma_e^2 + (\mu_1 - \mu_2)^2$$

$$= 2\sigma_e^2(1 + \delta_{12}^2).$$
So

\[ Var(W) = 2\sigma_e^2(1 + \delta_{12}^2) - 2\sigma_e^2[2\phi(\delta_{12}) + |\delta_{12}|P(|Z| < |\delta_{12}|)]^2 \]

\[ = 2\sigma_e^2 \left[ 1 + \delta_{12}^2 - [2\phi(\delta_{12}) + |\delta_{12}|P(|Z| < |\delta_{12}|)]^2 \right]. \]

When \( \mu_1 = \mu_2, \) \( Var(W) = (1 - \frac{2}{\pi})2\sigma_e^2 \) and \( Var(W) \rightarrow 2\sigma_e^2 \) as \( |\mu_1 - \mu_2| \rightarrow \infty. \)

Further, one can show numerically that

\[ \left(1 - \frac{2}{\pi}\right)2\sigma_e^2 \leq Var(W) < 2\sigma_e^2. \quad (2.5) \]

Since

\[ \hat{\sigma}_e = \frac{W}{d_2} \]

\[ Var(\hat{\sigma}_e) = 2\sigma_e^2 \left[ 1 + \delta_{12}^2 - [2\phi(\delta_{12}) + |\delta_{12}|P(|Z| < |\delta_{12}|)]^2 \right]/d_2^2 \]

Table 2.2 gives the values of \( Var(W) \) and \( Var(\hat{\sigma}_e) \) for different values of \( |\mu_1 - \mu_2| \) when \( \sigma_e^2 = 1. \) When \( \sigma_e^2 = 1, \) we see that as \( \delta_{12} \) increases, \( Var(W) \) approaches 2.

**Estimation of \( \hat{\sigma}_e \) Based on the Mean Moving Range**

Given \( n \) samples of size one, the mean moving range is defined as the arithmetic mean of the absolute difference of consecutive observations. In symbols

\[ \overline{M}_n = \frac{\sum_{j=2}^{n} |X_j - X_{j-1}|}{(n - 1)}. \]

We proceed to consider the properties of \( \overline{M}_n/d_2 \) as an estimator of \( \sigma_e \) under the various models introduced in the previous chapter.
Table 2.2: Variance of $W = |X_1 - X_2|$ and $\sigma_e$ when $\sigma_e^2 = 1$

| $|\mu_1 - \mu_2|$ | $Var(W)$ | $Var(W/d_2)$ |
|-----------------|-----------|--------------|
| 0.000000        | 0.727272  | 0.571428     |
| 0.100000        | 0.730901  | 0.574295     |
| 0.500000        | 0.814882  | 0.640264     |
| 1.000000        | 1.045033  | 0.819109     |
| 1.500000        | 1.327489  | 1.043027     |
| 2.000000        | 1.588212  | 1.247881     |
| 2.500000        | 1.779468  | 1.398154     |
| 3.000000        | 1.896372  | 1.490007     |
| 3.500000        | 1.957156  | 1.537765     |
| 4.000000        | 1.984381  | 1.559156     |
| 4.500000        | 1.994971  | 1.567477     |
| 5.000000        | 1.998569  | 1.570304     |

Properties of $\overline{M}_n/d_2$ under Model M1

The expected value of the mean moving range under model M1 is

$$EM_n = \frac{\sum_{j=2}^{n} E|X_{j-1} - X_j|}{(n-1)}$$

$$= \sqrt{2\sigma_e} \left[ \frac{\sum_{j=2}^{n} [2\phi(\delta_{j-1,j}) + |\delta_{j-1,j}|P(|Z| < |\delta_{j-1,j}|)]}{(n-1)} \right]$$

where $\delta_{j-1,j} = \frac{\mu_{j-1} - \mu_j}{\sqrt{2}\sigma_e}$. Clearly, if each $|\mu_{j-1} - \mu_j| = 0$ (model M5 holds) then

$$E \left( \frac{\overline{M}_n}{\sigma_e} \right) = \frac{2}{\sqrt{\pi}}$$

$$= d_2$$
\[ \hat{\sigma}_e = \frac{\overline{M}_n}{d_2} \] is an unbiased estimator of \( \sigma_e \). If the \( \mu_i \) are not all the same, \( \overline{M}_n/d_2 \) systematically overestimates \( \sigma_e \), but the overestimation factor does not exceed

\[ D_{\text{max}} = \frac{\sqrt{2}}{d_2} \left[ 2\phi(\delta_{\text{max}}) + |\delta_{\text{max}}|P(|Z| < |\delta_{\text{max}}|) \right] \] (2.6)

where

\[ \delta_{\text{max}} = \max\{|\delta_{j-1,j}|; j = 1, 2, \ldots, n\} \]

since we have seen numerically that \( E|X_{j-1} - X_j| \) is monotone in \( |\mu_{j-1} - \mu_j| \). In those cases (again, rare in practice) where \( \delta_{\text{max}} \neq 0 \) but the \( \delta_{j-1,j} \) are known, an unbiased estimator of \( \sigma_e \) based on the mean moving range is

\[ \frac{(n-1)\overline{M}_n}{\sqrt{2} \sum_{j=2}^{n} \left[ 2\phi(\delta_{j-1,j}) + |\delta_{j-1,j}|P(|Z| < |\delta_{j-1,j}|) \right]} \]

The variance of \( \overline{M}_n \) is

\[ \text{Var}(\overline{M}_n) = (n-1)^{-2}V_1. \]

where

\[ V = \begin{bmatrix} \text{Var}_{12} & \text{Cov}_{12,23} & 0 & 0 & \ldots \\ \text{Cov}_{12,23} & \text{Var}_{23} & \text{Cov}_{23,34} & 0 & \ldots \\ 0 & \text{Cov}_{23,34} & \text{Var}_{34} & \text{Cov}_{34,45} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \]

for

\[ \text{Var}_{ij} = \text{Var}(|X_i - X_j|) \]

\[ \text{Cov}_{ij,ik} = \text{Cov}(|X_i - X_j|, |X_j - X_k|). \]
\( V \) is an \((n - 1) \times (n - 1)\) symmetric matrix. In terms of the entries of the matrix \( V \), the variance of the mean moving range is

\[
Var(\bar{M}_n) = \frac{\sum Var(j-1)_j + 2\sum Cov(j-1)_j J(j+1)}{(n - 1)^2}.
\] (2.7)

A simple expression exists for the \( Var(|X_{j-1} - X_j|) \), but no such expression seems possible for \( Cov(|X_{j-1} - X_j|, |X_j - X_{j+1}|) \). Since no exact expression exists for \( E|X_{j-1} - X_j||X_j - X_{j+1}| \), Monte Carlo was used to evaluate this expression. Table 2.3 shows numerical values of \( E|X_{j-1} - X_j||X_j - X_{j+1}| \) for different sets of \((\delta(j-1)_j, \delta(j+1))\) when \( \sigma^2 = 1 \) (and standard errors for the numerical values in parentheses). These were obtained as follows. Since

\[
\begin{pmatrix}
(X_{j-1} - X_j) \\
(X_j - X_{j+1})
\end{pmatrix}
\sim N
\begin{pmatrix}
(\mu_{j-1} - \mu_j) \\
(\mu_j - \mu_{j+1})
\end{pmatrix},
\begin{pmatrix}
2 & -1 \\
-1 & 2
\end{pmatrix},
\]

random samples of size 2000 were generated from this bivariate normal distribution. The arithmetic mean of the values of \( |X_{j-1} - X_j||X_j - X_{j+1}| \) was used to approximate the expected value.

With the knowledge of \( |\mu_{j-1} - \mu_j| \) and \( |\mu_j - \mu_{j+1}| \), Tables 2.1 and 2.3 can be used to find the value of \( Cov(|X_{j-1} - X_j|, |X_j - X_{j+1}|) \) and the corresponding correlation. From Table 2.4, we observe that when \( (\mu_{j-1} - \mu_j) \) and \( (\mu_j - \mu_{j+1}) \) have the same sign, either both negative or both positive, there is negative correlation between \( |X_{j-1} - X_j| \) and \( |X_j - X_{j+1}| \), otherwise the correlation is positive. When \( (\mu_{j-1} - \mu_j) = 0 \), the correlation is maximum for \( (\mu_j - \mu_{j+1}) = 0 \), but diminishes as \( (\mu_j - \mu_{j+1}) \) moves away from zero.
Table 2.3: Numerical Values of $E|X_{j-1} - X_{j}||X_{j} - X_{j+1}|$ when $\sigma_{\varepsilon}^2 = 1$ (Standard Errors in Parentheses)

<table>
<thead>
<tr>
<th>$(\mu_{j-1} - \mu_{j})$</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.1006)</td>
<td>(0.0825)</td>
<td>(0.0677)</td>
<td>(0.0780)</td>
<td>(0.1109)</td>
<td>(0.1435)</td>
<td>(0.1716)</td>
</tr>
<tr>
<td>-2</td>
<td>5.5383</td>
<td>3.6921</td>
<td>2.5403</td>
<td>2.4199</td>
<td>3.3508</td>
<td>5.0070</td>
<td>6.9757</td>
</tr>
<tr>
<td></td>
<td>(0.0825)</td>
<td>(0.0637)</td>
<td>(0.0522)</td>
<td>(0.0631)</td>
<td>(0.0900)</td>
<td>(0.1177)</td>
<td>(0.1435)</td>
</tr>
<tr>
<td>-1</td>
<td>3.7655</td>
<td>2.5403</td>
<td>1.7687</td>
<td>1.6881</td>
<td>2.3598</td>
<td>3.3508</td>
<td>4.6404</td>
</tr>
<tr>
<td></td>
<td>(0.0677)</td>
<td>(0.0522)</td>
<td>(0.0424)</td>
<td>(0.0495)</td>
<td>(0.0663)</td>
<td>(0.0900)</td>
<td>(0.1110)</td>
</tr>
<tr>
<td>0</td>
<td>3.4314</td>
<td>2.4199</td>
<td>1.6881</td>
<td>1.4301</td>
<td>1.6881</td>
<td>2.4199</td>
<td>3.4314</td>
</tr>
<tr>
<td></td>
<td>(0.0780)</td>
<td>(0.0631)</td>
<td>(0.0495)</td>
<td>(0.0429)</td>
<td>(0.0495)</td>
<td>(0.0631)</td>
<td>(0.0780)</td>
</tr>
<tr>
<td>1</td>
<td>4.6404</td>
<td>3.3508</td>
<td>2.3598</td>
<td>1.6881</td>
<td>1.7685</td>
<td>2.5403</td>
<td>3.7655</td>
</tr>
<tr>
<td></td>
<td>(0.1109)</td>
<td>(0.0900)</td>
<td>(0.0663)</td>
<td>(0.0424)</td>
<td>(0.0608)</td>
<td>(0.0744)</td>
<td>(0.0963)</td>
</tr>
<tr>
<td>2</td>
<td>6.9757</td>
<td>5.0070</td>
<td>3.5088</td>
<td>2.4199</td>
<td>2.5403</td>
<td>3.6921</td>
<td>5.3833</td>
</tr>
<tr>
<td></td>
<td>(0.1435)</td>
<td>(0.1177)</td>
<td>(0.0900)</td>
<td>(0.0631)</td>
<td>(0.0522)</td>
<td>(0.0637)</td>
<td>(0.0825)</td>
</tr>
<tr>
<td></td>
<td>(0.1716)</td>
<td>(0.1435)</td>
<td>(0.1109)</td>
<td>(0.0780)</td>
<td>(0.0677)</td>
<td>(0.0825)</td>
<td>(0.1006)</td>
</tr>
</tbody>
</table>
Table 2.4: Numerical Values of $\text{Corr}(|X_{(j-1)} - X_j|, |X_j - X_{(j+1)}|)$

<table>
<thead>
<tr>
<th>$(\mu_{j-1} - \mu_j)$</th>
<th>-10</th>
<th>-5</th>
<th>-3</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-0.40</td>
<td>-0.32</td>
<td>-0.30</td>
<td>-0.27</td>
<td>-0.20</td>
<td>0.00</td>
<td>0.08</td>
<td>0.17</td>
<td>0.19</td>
<td>0.17</td>
<td>0.08</td>
</tr>
<tr>
<td>-5</td>
<td>-0.32</td>
<td>-0.46</td>
<td>-0.43</td>
<td>-0.40</td>
<td>-0.30</td>
<td>0.00</td>
<td>0.22</td>
<td>0.31</td>
<td>0.33</td>
<td>0.30</td>
<td>0.17</td>
</tr>
<tr>
<td>-3</td>
<td>-0.30</td>
<td>-0.43</td>
<td>-0.49</td>
<td>-0.45</td>
<td>-0.34</td>
<td>0.00</td>
<td>0.27</td>
<td>0.36</td>
<td>0.37</td>
<td>0.33</td>
<td>0.19</td>
</tr>
<tr>
<td>-2</td>
<td>-0.27</td>
<td>-0.40</td>
<td>-0.45</td>
<td>-0.44</td>
<td>-0.32</td>
<td>0.05</td>
<td>0.30</td>
<td>0.38</td>
<td>0.36</td>
<td>0.31</td>
<td>0.17</td>
</tr>
<tr>
<td>-1</td>
<td>-0.20</td>
<td>-0.30</td>
<td>-0.34</td>
<td>-0.32</td>
<td>-0.18</td>
<td>0.14</td>
<td>0.32</td>
<td>0.30</td>
<td>0.27</td>
<td>0.22</td>
<td>0.08</td>
</tr>
<tr>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.14</td>
<td>0.22</td>
<td>0.14</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1</td>
<td>0.08</td>
<td>0.22</td>
<td>0.27</td>
<td>0.30</td>
<td>0.32</td>
<td>0.14</td>
<td>-0.18</td>
<td>-0.32</td>
<td>-0.34</td>
<td>-0.30</td>
<td>-0.20</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.31</td>
<td>0.36</td>
<td>0.38</td>
<td>0.30</td>
<td>0.05</td>
<td>-0.32</td>
<td>-0.44</td>
<td>-0.45</td>
<td>-0.40</td>
<td>-0.27</td>
</tr>
<tr>
<td>3</td>
<td>0.19</td>
<td>0.33</td>
<td>0.37</td>
<td>0.36</td>
<td>0.27</td>
<td>0.00</td>
<td>-0.34</td>
<td>-0.45</td>
<td>-0.49</td>
<td>-0.43</td>
<td>-0.30</td>
</tr>
<tr>
<td>5</td>
<td>0.17</td>
<td>0.30</td>
<td>0.33</td>
<td>0.31</td>
<td>0.22</td>
<td>0.00</td>
<td>-0.30</td>
<td>-0.40</td>
<td>-0.43</td>
<td>-0.46</td>
<td>-0.32</td>
</tr>
<tr>
<td>10</td>
<td>0.08</td>
<td>0.17</td>
<td>0.19</td>
<td>0.17</td>
<td>0.08</td>
<td>0.00</td>
<td>-0.20</td>
<td>-0.27</td>
<td>-0.30</td>
<td>-0.32</td>
<td>-0.40</td>
</tr>
</tbody>
</table>
Asymptotic Properties of the Mean Moving Range

Now we study some large sample properties for the mean moving range, $\overline{M}_n$. Proposition 2.1 states the a.e. convergence of $\overline{M}_n - E\overline{M}_n$ to zero. This is a kind of almost sure consistency of $\sigma_e = \overline{M}_n/d_2$ for its expectation in model M1.

**Proposition 2.1** Let $\{X_j\}_{j=1}^\infty$ be a sequence of independent random variables such that $X_j \sim N(\mu_j, \sigma_e^2)$. Then, for any fixed sequence of means $\{\mu_j\}_{j=1}^\infty$

$$\overline{M}_n - E\overline{M}_n \xrightarrow{a.e.} 0.$$ 

**Proof:** Let $Y_j = |X_j - X_{j+1}|$. Define another sequence of random variables $\{W_j\}$ by $W_j = Y_j - EY_j$. We will show that with $\hat{n} = (n - 1)$, and $S_{\hat{n}} = \sum_{j=1}^{\hat{n}} W_j$

$$\frac{S_{\hat{n}}}{\hat{n}} \xrightarrow{a.e.} 0 \text{ as } \hat{n} \to \infty.$$ 

Using the obvious notations for the sums of the odd and even indexed terms making up $S_{\hat{n}}$, we may write

$$S_{\hat{n}} = S_{\hat{n}}^{\text{odd}} + S_{\hat{n}}^{\text{even}} + R_{\hat{n}}$$

or

$$\overline{W}_{\hat{n}} = \overline{W}_{\hat{n}}^{\text{odd}} + \overline{W}_{\hat{n}}^{\text{even}} + \overline{R}_{\hat{n}}$$

where

$$\overline{W}_{\hat{n}}^{\text{odd}} = \text{the arithmetic mean of the } W_j \text{'s with } j \text{ odd and } j \leq (\hat{n} - 1),$$
$W_{\text{even}}$ = the arithmetic mean of the $W_j$'s with $j$ even and $j \leq \tilde{n}$,
$c_{\text{odd}}^\tilde{n} = \begin{cases} \frac{1}{2} & \text{if } \tilde{n} \text{ is even} \\ \frac{(\tilde{n}-1)}{2\tilde{n}} & \text{if } \tilde{n} \text{ is odd,} \end{cases}
$c_{\text{even}}^\tilde{n} = \begin{cases} \frac{1}{2} & \text{if } \tilde{n} \text{ is even} \\ \frac{(\tilde{n}-1)}{2\tilde{n}} & \text{if } \tilde{n} \text{ is odd,} \end{cases}$

and

$R_{\tilde{n}} = \begin{cases} 0 & \text{if } \tilde{n} \text{ is even} \\ W_{\tilde{n}} & \text{if } \tilde{n} \text{ is odd.} \end{cases}$

Lamperti (1966, page 31) proves the following form of the strong law of large numbers:

Let $X_1, X_2, \ldots$ be independent random variables with means $\mu_i$ and variances $\sigma_i^2$.

If $\sum_{j=1}^{\infty} (\sigma_j^2 / j^2) < \infty$, then

$$P \left( \lim_{n \to \infty} \left[ \frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{\mu_1 + \mu_2 + \ldots + \mu_n}{n} \right] = 0 \right) = 1$$

Since by (2.5) $2\sigma_\tilde{c}^2 \left(1 - \frac{2}{\tilde{n}}\right) \leq \text{Var}(Y_j) = \text{Var}(W_j) \leq 2\sigma_\tilde{c}^2$ the condition of the Lamperti's strong law is satisfied for $W_{\tilde{n}}^{\text{odd}}$ and $W_{\tilde{n}}^{\text{even}}$, so

$$W_{\tilde{n}}^{\text{odd}} \xrightarrow{\text{a.e.}} 0$$

and

$$W_{\tilde{n}}^{\text{even}} \xrightarrow{\text{a.e.}} 0.$$
Also,

\[ c_{\tilde{n}}^{\text{odd}} \to 1/2 \]

and

\[ c_{\tilde{n}}^{\text{even}} \to 1/2 \quad \text{as } \tilde{n} \to \infty. \]

We thus need only to show that \( R_{\tilde{n}} / \tilde{n} \) a.e. \( \to 0 \).

By Chebyshev's inequality, for \( \epsilon > 0 \)

\[ P \left( \left| R_{\tilde{n}} / \tilde{n} \right| > \epsilon \right) \leq \frac{\text{Var}(R_{\tilde{n}})}{\tilde{n}^2 \epsilon^2} \leq 2 \sigma_{\epsilon}^2 / \tilde{n}^2 \epsilon^2. \]

So

\[ \sum P \left( \left| R_{\tilde{n}} / \tilde{n} \right| > \epsilon \right) < \infty. \]

Thus, by the Borel Cantelli lemma,

\[ P \left( \left| R_{\tilde{n}} / \tilde{n} \right| > \epsilon \text{ i.o.} \right) = 0, \]

which implies \( \exists \) a set \( \Omega_\epsilon \) of probability 1 such that for each \( \omega \in \Omega_\epsilon \) \( \exists n(\omega, \epsilon) \) such that \( \tilde{n} > n(\omega, \epsilon) \) implies that

\[ |R_{\tilde{n}} / \tilde{n}| < \epsilon. \quad (2.8) \]

Let \( \epsilon \) run through a sequence of values decreasing to zero, for example \( \{1/i\} \). Then, \( \Omega_0 = \bigcap_{i=1}^\infty \Omega_{1/i} \) will have probability 1 since \( P(\Omega_0) = \lim_i P(\Omega_{1/i}) \). On the set \( \Omega_0 \)

\[ R_{\tilde{n}} / \tilde{n} \to 0. \]
We note that Proposition 2.1 guarantees an almost sure convergence to 0 of \( \bar{M}_n - E\bar{M}_n \) for each fixed sequence of means. Models M2 and M4 add to the assumptions of model M1 distributional assumptions on the sequence \( \{\mu_j\}_1^\infty \), while models M3 and M5 specify particular restrictions on fixed sequence of means. As such, Proposition 2.1 immediately provides almost sure convergence results for \( \bar{M}_n \) under M3 and M5, and provided one can show the almost sure convergence of the function of \( \{\mu_j\}_1^\infty \) symbolized as \( E\bar{M}_n \) in Proposition 2.1, under models M2 and M4 as well.

Proposition 2.2 is a central limit theorem for \( \bar{M}_n \) under assumptions on the sequence \( \{\mu_j\}_1^\infty \) like those of model M1, M3, M4 and M5 that make the sequence \( \{\|X_j - X_{j+1}\|\} \) stationary.

**Proposition 2.2** Suppose that assumptions on the sequence \( \{\mu_j\}_1^\infty \) are added to the basic model assumptions M1 in such a way that

1. the sequence \( \{\|X_j - X_{j+1}\|\} \) is stationary and \( t \)-dependent, and

2. \( \text{Corr}(\|X_{j-1} - X_j\|, \|X_j - X_{j+1}\|) > -1/2 \).

Then
\[ \sqrt{n}(\overline{M}_n - EM_n) \xrightarrow{d} N(0, \sigma^2_{(1)}) \]

where \( \sigma^2_{(1)} = \text{Var}(|X_1 - X_2|) + 2\text{Cov}(|X_1 - X_2|, |X_2 - X_3|) \).

Proof: This result is an immediate consequence of the following form of the central limit theorem (Anderson, 1971):

Let \( \{Y_n\} \) be a sequence of stationary m-dependent random variables with \( EY_1^2 < \infty \).

Assume that \( \sigma_m^2 = \text{Var}(Y_1) + 2 \sum_{k=1}^{m} \text{Cov}(Y_1, Y_{1+k}) > 0 \), then

\[ \sqrt{n}(\overline{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2_m). \]

Consider the immediate implications of Proposition 2.2. Under both models M4 and M5, the sequence \( \{X_j\} \) is an iid sequence. This implies that the sequence \( \{|X_j - X_{j+1}|\} \) is 1-dependent and stationary. Then use of the \( \mu_{j-1} - \mu_j = \mu_j - \mu_{j+1} = 0 \) entry of Table 2.4 shows that the correlation assumption of Proposition 2.2 holds, and that \( \overline{M}_n \) (and so \( \hat{\sigma}_e \)) is asymptotically normal.

Similarly, under model M3, the sequence \( \{|X_j - X_{j+1}|\} \) is 1-dependent and stationary. Then using \( \mu_{j-1} - \mu_j = \mu_j - \mu_{j+1} \) entries of Table 2.4 shows that at least for \( |h| \leq 10\sigma_e \) the correlation assumption seems to hold, and thus \( \overline{M}_n \) (and so \( \hat{\sigma}_e \)) is asymptotically normal.

m-dependent central limit theorems are available in literature for the case where observations are not identically distributed. For example, Orey(1958) gives the following result:

Let \( \{X_i\} \) be a sequence of m-dependent random variables with 0 means and finite variances. Let \( B_n^2 \) be the variance of \( S_n = X_1 + X_2 + \ldots + X_n \). Then

\[ \frac{(S_n)}{B_n} \xrightarrow{d} N(0, 1) \text{ if} \]
(a) \( \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{0}^{\epsilon B_n} X_k^2 \to 0 \) for every \( \epsilon > 0 \), and 
(b) \( \frac{1}{B_n^2} \sum_{k=1}^{n} \int X_k^2 = O(1) \).

It seems likely that additional analysis and Orey's result would show that
\[
\frac{M_n - E M_n}{\sqrt{Var M_n}} \xrightarrow{d} N(0, 1)
\]
under the model assumptions M1 for any sequence of means \( \mu_j \) starting from 1.

**Properties of \( \sigma_e \) and Bias Correction under Model M2**

As described earlier, under model M2, the \( \mu_i \)'s are generated by an autoregressive model i.e. for \( \nu_i \sim N(0, \sigma_{\nu}^2) \) and \( \mu_0 \) an unknown parameter
\[
\mu_i = \rho \mu_{i-1} + \nu_i
\]
or
\[
\mu_i = \rho^i \mu_0 + \rho^{i-1} \nu_1 + \rho^{i-2} \nu_2 + \ldots + \rho \nu_{(i-1)} + \nu_i.
\]
That is,
\[
X_i = \rho^i \mu_0 + \rho^{i-1} \nu_1 + \rho^{i-2} \nu_2 + \ldots + \rho \nu_{(i-1)} + \nu_i + \epsilon_i.
\]
So
\[
(X_{j-1} - X_j) \sim N \left( \rho^{j-1} (1 - \rho) \mu_0, 2 \sigma_e^2 + \sigma_{\nu}^2 \left[ 1 + (1 - \rho) \frac{1 - \rho^{2(j-1)}}{(1 + \rho)} \right] \right).
\]
By arguments similar to those on page 12,
\[
E|X_{j-1} - X_j| = \sigma_e \sqrt{2 + \frac{\sigma_{\nu}^2}{\sigma_e^2} \left[ 1 + (1 - \rho) \frac{1 - \rho^{2(j-1)}}{(1 + \rho)} \right]} \left[ 2\phi(\delta_j) + |\delta_j| P(|Z| < |\delta_j|) \right] \tag{2.9}
\]
where

$$\delta(j) = \frac{\rho j - (1 - \rho) \mu_0}{\sigma_e \sqrt{2 + \frac{\sigma_e^2}{\sigma_0^2} \left[ 1 + (1 - \rho) \frac{(1 - \rho^2(j - 1))}{(1 + \rho)} \right]}}$$

So, in those rare occasions where both $\rho$ and $\frac{\sigma_e^2}{\sigma_0^2}$ are known, an unbiased estimator of $\sigma_e$ based on $|X_{j-1} - X_j|$ is

$$\frac{|X_{j-1} - X_j|}{\sqrt{2 + \left( 1 + (1 - \rho) \frac{(1 - \rho^2(j - 1))}{(1 + \rho)} \right) \frac{\sigma_e^2}{\sigma_0^2} \left[ 2\phi(\delta(j)) + |\delta(j)|P(|Z| < |\delta(j)|) \right]}}$$

That is, if $\rho$ and the ratio $\frac{\sigma_e^2}{\sigma_0^2}$ are known, an adjustment to $\sigma_e$ based on $|X_{j-1} - X_j|$ may be made by multiplying by the factor

$$\kappa_2(j) = \frac{d_2}{\sqrt{2 + \left( 1 + (1 - \rho) \frac{(1 - \rho^2(j - 1))}{(1 + \rho)} \right) \frac{\sigma_e^2}{\sigma_0^2} \left[ 2\phi(\delta(2)) + |\delta(2)|P(|Z| < |\delta(2)|) \right]}}$$

(2.10)

In a similar fashion, under model model M2

$$E\bar{M}_n = \frac{\sigma_e}{(n - 1)} \sum_j \sqrt{2 + \frac{\sigma_e^2}{\sigma_0^2} \left[ 1 + (1 - \rho) \frac{(1 - \rho^2(j - 1))}{(1 + \rho)} \right] \left[ 2\phi(\delta(j)) + |\delta(j)|P(|Z| < |\delta(j)|) \right]}$$

(2.11)

and if $\rho$ and $\frac{\sigma_e^2}{\sigma_0^2}$ are known, an unbiased estimator of $\sigma_e$ can be obtained from $\hat{\sigma}_e = \bar{M}_n / d_2$ as

$$\left( \frac{1}{n - 1} \sum_{j=2}^{n} \frac{1}{\kappa_2(j)} \right)^{-1} \hat{\sigma}_e$$

(2.12)
Model M2.1  Since the random walk model M2.1 is a special case of model M2 where \( \rho = 1 \), it is evident from (2.10) and (2.12) how to correct \( \hat{\sigma}_e \) to produce an unbiased estimator of \( \sigma_e \). Under the random walk model each

\[
\kappa_2(j) = \frac{d_2}{\sqrt{2 + \frac{\sigma_\mu^2}{\sigma_\epsilon^2} \frac{2}{\sqrt{2\pi}}}} = \frac{1}{\sqrt{1 + \frac{\sigma_\mu^2}{2\sigma_\epsilon^2}}}.
\]

This means that \( \hat{\sigma}_e \) tends to overestimate \( \sigma_e \) by a factor of \( \sqrt{1 + \frac{\sigma_\mu^2}{2\sigma_\epsilon^2}} \) and that an unbiased estimator of \( \sigma_e \) is

\[
\frac{\hat{\sigma}_e}{\sqrt{1 + \frac{\sigma_\mu^2}{2\sigma_\epsilon^2}}} = \frac{\overline{M}_n}{\sqrt{\frac{\sigma_\mu^2}{\pi} \left( 2 + \frac{\sigma_\mu^2}{\sigma_\epsilon^2} \right)}}.
\]

Further, arguments similar to those used in the proof of Proposition 2.1 show that under the random walk model M2.1, the function of the sequence \( \{\mu_j\} \) symbolized as \( EM_n \) in the statement of Proposition 2.1 converges almost surely to

\[
\sigma_e \sqrt{\frac{2}{\pi}} \sqrt{2 + \frac{\sigma_\mu^2}{\sigma_\epsilon^2}}.
\]

Thus, under the random walk model M2.1

\[
\hat{\sigma}_e \xrightarrow{a.s.} \sigma_e \sqrt{1 + \frac{\sigma_\mu^2}{2\sigma_\epsilon^2}}.
\]

It is a simple matter to verify that under model M2.1 the sequence \( \{|X_j - X_{j-1}|\} \) is stationary and 1-dependent. Thus, in order to infer the asymptotic normality of \( \overline{M}_n \) (and thus \( \hat{\sigma}_e \)) only the correlation condition of Proposition 2.2 needs to be checked. Note that
\[
\begin{bmatrix}
(X_{j-1} - X_j) \\
(X_j - X_{j+1})
\end{bmatrix}
\sim N\left[
\begin{bmatrix}
0 \\
0
\end{bmatrix},
\begin{bmatrix}
(2\sigma_x^2 + \sigma_p^2) & -\sigma_x^2 \\
-\sigma_x^2 & (2\sigma_x^2 + \sigma_p^2)
\end{bmatrix}
\right],
\]

and
\[
\rho = \text{Corr}[(X_1 - X_2), (X_2 - X_3)] = -\frac{(\sigma_x^2 + \sigma_p^2)}{(2\sigma_x^2 + \sigma_p^2)}
\]

Now, Johnson and Kotz (1972) prove the following:

If $Y$ and $Z$ have a standard bivariate normal distribution with correlation coefficient $\rho$, then

\[
E(\{|YZ|\}) = \frac{2}{\pi} \sqrt{1 - \rho^2 + \rho \sin^{-1} \rho}.
\]

So we have

\[
E[|(X_1 - X_2)(X_2 - X_3)|] = (2\sigma_x^2 + \sigma_p^2)E\left[\frac{(X_1 - X_2)}{\sqrt{2\sigma_x^2 + \sigma_p^2}} \cdot \frac{(X_2 - X_3)}{\sqrt{2\sigma_x^2 + \sigma_p^2}}\right]
\]

\[
= \frac{2(2\sigma_x^2 + \sigma_p^2)}{\pi} \sqrt{1 - \rho^2 + \rho \sin^{-1} \rho}.
\]

Also, $E|X_{j-1} - X_j| = \sqrt{\frac{\pi}{2}} \sqrt{2\sigma_x^2 + \sigma_p^2}$ gives

\[
\text{Cov}(|X_1 - X_2|, |X_2 - X_3|) = \frac{2(2\sigma_x^2 + \sigma_p^2)}{\pi} \sqrt{1 - \rho^2 + \rho \sin^{-1} \rho - 1}
\]

and

\[
\text{Var}(|X_{j-1} - X_j|) = \left(1 - \frac{2}{\pi}\right)(2\sigma_x^2 + \sigma_p^2).
\]

Thus

\[
\text{Corr}(|X_1 - X_2|, |X_2 - X_3|) = \left(\frac{2}{\pi - 2}\right) \left[\sqrt{1 - \rho^2} + \rho \sin^{-1} \rho - 1\right]
\]

\[
= 1.75 \left[\sqrt{1 - \rho^2} + \rho \sin^{-1} \rho - 1\right] \quad (2.15)
\]
Table 2.5: Values of $\text{Corr}(|X_1 - X_2|, |X_2 - X_3|)$ for Model M2.1

<table>
<thead>
<tr>
<th>$\sigma_v^2$</th>
<th>0</th>
<th>0.5</th>
<th>1.0</th>
<th>2.0</th>
<th>5.0</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>-1/2</td>
<td>-0.6</td>
<td>-2/3</td>
<td>-3/4</td>
<td>-6/7</td>
<td>-1</td>
</tr>
<tr>
<td>$\text{Corr}(</td>
<td>X_2 - X_1</td>
<td>,</td>
<td>X_3 - X_2</td>
<td>)$</td>
<td>0.2238</td>
<td>0.3259</td>
</tr>
</tbody>
</table>

Table 2.5 shows that for $\sigma_v^2 = 1$, $\text{Corr}(|X_1 - X_2|, |X_2 - X_3|)$ is minimum when $\sigma_v^2 = 0$ and increases steadily as $\sigma_v^2$ increases, approaching one as $\sigma_v^2 \to \infty$. Thus, the second condition of Proposition 2.2, $\text{Corr}(|X_1 - X_2|, |X_2 - X_3|) > -1/2$, is satisfied and $\bar{M}_n$ (and thus $\hat{\sigma}_e$) is asymptotically normal under the random walk model.

Properties of $\hat{\sigma}_e$ and Bias Correction under Model M3

Under the linear drift model M3,

$$X_i \sim N(\mu_i + i \epsilon, \sigma_v^2) \quad i = 1, 2, \ldots, n$$

and thus

$$\delta_{j-1,j} = \frac{\mu_i - \mu_j}{\sqrt{2 \sigma_v}} = \left( \frac{-h}{\sqrt{2 \sigma_v}} \right).$$

The expected values of the moving range and mean moving range are the same since the $\delta_{j-1,j}$'s are constant. That is

$$E \bar{M}_n = \sqrt{2 \sigma_v} \left[ 2 \phi \left( \frac{-h}{\sqrt{2 \sigma_v}} \right) + \frac{|h|}{\sqrt{2 \sigma_v}} \phi \left( |Z| < \frac{|h|}{\sqrt{2 \sigma_v}} \right) \right] \quad (2.16)$$

where $\phi$ is the standard normal distribution function.
Using $\hat{\sigma}_e$ under model M3 results in overestimation. In those rare cases where one knows the value of $\frac{|h|}{\hat{\sigma}_e}$, $\hat{\sigma}_e$ should be corrected by multiplying by the factor

$$\kappa_3 = \frac{d_2}{\sqrt{2} \left[ 2\phi \left( \frac{-h}{\sqrt{2}\hat{\sigma}_e} \right) + \frac{|h|}{\sqrt{2}\hat{\sigma}_e} P \left( |Z| < \frac{|h|}{\sqrt{2}\hat{\sigma}_e} \right) \right]}.$$ (2.17)

Note that under model M3, $E\bar{M}_n$ given on page 16 is constant. Thus Proposition 2.1 provides the almost sure convergence of $\bar{M}_n$ to its expectation and $\hat{\sigma}_e$ to $\frac{\sigma_e}{\kappa_3}$. Further, Table 2.4 suggests that at least for $|h| \leq 10$ the correlation between $|X_1 - X_2|$ and $|X_2 - X_3|$ exceeds 1/2, so that Proposition 2.2 provides the asymptotic normality of $\bar{M}_n$ (and thus $\hat{\sigma}_e$).

**Properties of $\hat{\sigma}_e$ and Bias Correction under Model M4**

In this model

$$X_j \sim \text{IN}(\mu_j, \sigma^2)$$

where $\sigma^2 = (\sigma_e^2 + \sigma_{\mu_j}^2)$. So clearly,

$$E\hat{\sigma}_e = E\sigma_e = \sqrt{\sigma_e^2 + \sigma_{\mu_j}^2}$$

In those cases where $\frac{\sigma_{\mu_j}}{\sigma_e}$ is known from previous studies, an unbiased estimator of $\sigma_e$ is given by

$$\kappa_4 \hat{\sigma}_e$$

where $\kappa_4 = \frac{1}{\sqrt{\frac{\sigma_{\mu_j}^2}{\sigma_e^2} + 1}}$. 
Under model M4, $E\bar{M}_n$ given on the top of page 16 can be shown (using essentially the same arguments as used to prove Proposition 2.1) to converge almost surely to its (constant) expectation

$$\frac{d_2}{\kappa_4} \sigma_e.$$

Thus, invoking Proposition 2.1 one has the almost sure convergence of $\bar{M}_n$ (and thus $\hat{\sigma}_e$). Further, the asymptotic normality of $\bar{M}_n$ follows from Proposition 2.2, once one realizes that under model M4, the fact that the $X_j$'s are iid implies that the sequence $\{|X_{j-1} - X_j|\}$ is 1-dependent and stationary and that correlations similar to those producing the $\delta_{j-1,j} = \delta_{j,j+1} = 0$ entry of Table 2.4 show the correlation between $|X_1 - X_2|$ and $|X_2 - X_3|$ to be positive.
CHAPTER 3. MAXIMUM LIKELIHOOD ESTIMATORS

Maximum likelihood is a widely used estimation technique. For \( X = [X_1, \ldots, X_n] \) a random vector of observations with joint density \( f_n(x, \Theta) \) over \( n \)-dimensional Euclidean space \( \mathbb{R}^n \) suppose \( \Omega \subset \mathbb{R}^s \) contains the unknown parameter. Given \( x \), the likelihood function is the function of \( \Theta \)

\[
L(\Theta) = f_n(x|\Theta).
\]

Any \( \hat{\Theta} = \hat{\Theta}(x) \in \Omega \) which maximizes \( L(\Theta) \) over \( \Omega \) is known as a maximum likelihood estimate (MLE) of the unknown parameter \( \Theta \). MLE's do not always exist and if an MLE exists, it may not be unique. Under some regularity conditions, in \( iid \) models with increasing \( n \), MLE's can be shown to be consistent estimators. Further, (again under regularity conditions in \( iid \) models), if \( \hat{\Theta}_n \) is a consistent sequence of roots of the likelihood equations

\[
\frac{\partial}{\partial \Theta_j} \log f_n(X|\Theta) = \frac{\partial}{\partial \Theta_j} \sum_{i=1}^{n} \log f(X_i|\Theta) = 0
\]

\( j = 1, 2, \ldots, s \), then as \( n \to \infty \)

\[
\sqrt{n}(\hat{\Theta}_n - \Theta) \xrightarrow{d} N(0, nI(\Theta)^{-1}).
\]

i.e., in large samples the distribution of \( \hat{\Theta}_n \) is approximately \( s \)-variate normal with mean \( \Theta \) and covariance matrix \( I(\Theta)^{-1} \), where \( I(\Theta) = (I_{jk}(\Theta))_{jk=1,2,\ldots,s} \) is the
Fisher information matrix evaluated at the true $\Theta$. This is the reason that MLE's are regarded as efficient estimators.

**Maximum Likelihood Estimation of the Variance $\sigma^2_e$**

**Estimation under Model M1**

Under the basic model M1

$$X_i \sim \text{IN}(\mu_i, \sigma^2_e) \quad i = 1, 2, \ldots, n$$

we have an impossible estimation problem. There are $n$ observations, but $(n + 1)$ parameters to be estimated ($\sigma^2_e, \mu_1, \mu_2, \ldots, \mu_n$). Progress can only be made by adding more assumptions to the basic model. There follows in this chapter some discussion of what is feasible in the way of maximum likelihood estimation under the additional assumptions of models M2, M3 and M5.

**Estimation under Model M2**

In model M2, the population means are assumed to follow an AR(1) model as described in Chapter 1. Here, $\mathbf{X} = [X_1, X_2, \ldots, X_n]' \sim \mathcal{N}(\mu_0, \Sigma)$ where $\Psi = [\rho, \rho^2, \ldots, \rho^n]'$ and $\Sigma = \sigma^2_e I_n + \sigma^2_p V_n$ with $V_n$ as defined in Chapter 1.

If $\mu_0$ and $\rho$ are known, maximum likelihood estimators of $\sigma^2_e$ and $\sigma^2_p$ are solutions to the likelihood equations

$$\text{tr}(\hat{\sigma}_e^2(m^2) I_n + \hat{\sigma}_p^2(m^2) V_n)^{-1} = \text{tr}(\hat{\sigma}_e^2(m^2) I_n + \hat{\sigma}_p^2(m^2) V_n)^{-1}$$

and

$$\text{tr}(\hat{\sigma}_e^2(m^2) I_n + \hat{\sigma}_p^2(m^2) V_n)^{-1} V_n = \text{tr}(\hat{\sigma}_e^2(m^2) I_n + \hat{\sigma}_p^2(m^2) V_n)^{-1}$$
\[ V_n \left( \hat{\sigma}_e^2(ml2) I_n + \hat{\sigma}_\nu^2(ml2) V_n \right)^{-1} C \]

where

\[ C = (X - \mu_0 \Psi) (X - \mu_0 \Psi)' \]  

(3.1)

There is at least one solution \( \hat{\sigma}_e^2(ml2) \) and \( \hat{\sigma}_\nu^2(ml2) \) such that

\[ \hat{G} = (\hat{\sigma}_e^2(ml2) I_n + \hat{\sigma}_\nu^2(ml2) V_n) \]  

(3.2)

is positive definite. If there is more than one solution to the likelihood equations, the absolute maximum of the likelihood function is obtained by the solution minimizing \( |\hat{G}| \). (Anderson 1969, 1970)

If both \( \mu_0 \) and \( \Sigma \) are unknown (but \( \rho \) is known), the maximum likelihood estimator of \( \mu_0 \) is given by

\[ \hat{\mu}_0 = \left( \frac{\Psi' \hat{\Sigma}^{-1} \Psi}{\Psi' \hat{\Sigma}^{-1} \Psi} \right) \]  

(3.3)

Anderson (1973) proposed the following iterative procedure which gives estimators asymptotically equivalent to maximum likelihood estimators. From (3.2) the likelihood equations can be written as

\[ tr(\hat{\Sigma}^{-1} \hat{\Sigma}^{-1} \hat{\sigma}_e^2(ml2)) + tr(\hat{\Sigma}^{-1} \hat{\Sigma}^{-1} V_n \hat{\sigma}_\nu^2(ml2)) = tr(\hat{\Sigma}^{-1} \hat{\Sigma}^{-1} C) \]  

(3.4)

\[ tr(\hat{\Sigma}^{-1} \hat{\Sigma}^{-1} V_n \hat{\sigma}_e^2(ml2)) + tr(\hat{\Sigma}^{-1} V_n \hat{\Sigma}^{-1} V_n \hat{\sigma}_\nu^2(ml2)) = tr(\hat{\Sigma}^{-1} V_n \hat{\Sigma}^{-1} C) \]  

(3.5)

Let \( \hat{\sigma}_e^2(ml2)(0) \) and \( \hat{\sigma}_\nu^2(ml2)(0) \) be an initial set of values and \( \hat{\sigma}_e^2(ml2)(i) \) and \( \hat{\sigma}_\nu^2(ml2)(i) \) be solutions to the set of equations

\[ tr(\hat{\Sigma}^{-1} \hat{\Sigma}^{-1} \hat{\sigma}_e^2(ml2)) + tr(\hat{\Sigma}^{-1} V_n \hat{\sigma}_\nu^2(ml2)) = tr(\hat{\Sigma}^{-1} V_n \hat{\Sigma}^{-1} C) \]  

(3.6)
\[
tr(\hat{\Sigma}_{(i-1)}^{-1} V_n \hat{\Sigma}_{(i-1)}^{-1}) \hat{\sigma}_e^2 + tr(\hat{\Sigma}_{(i-1)}^{-1} V_n \hat{\Sigma}_{(i-1)}^{-1} V_n) \hat{\sigma}_p^2
= tr(\hat{\Sigma}_{(i-1)}^{-1} V_n \hat{\Sigma}_{(i-1)}) \hat{C}
\]

where \( \hat{\Sigma}_{(i-1)} = \hat{\sigma}_e^2(\hat{\Sigma}_{(i-1)}^{-1}) I_n + \hat{\sigma}_p^2(\hat{\Sigma}_{(i-1)}^{-1}) V_n \) \( i = 1, 2, \ldots \) If \( \hat{\Sigma}_{(i-1)} \) is nonsingular, the matrix of coefficients in (3.6) and (3.7) is positive definite (Anderson 1970).

Anderson suggests using initial estimates from the procedure proposed by C.R. Rao (1972). This involves solving the equations

\[
tr(V_n) \hat{\sigma}_e^2 + tr(V_n) \hat{\sigma}_p^2 = tr(C)
\]

and

\[
tr(V_n) \hat{\sigma}_e^2 + tr(V_n) \hat{\sigma}_p^2 = tr(V_n C)
\]

simultaneously. Whittle (1953, 1954), A.M. Walker (1964) and Hannan (1970) have shown that \( \sqrt{n}(\hat{\sigma}_e^2 - \sigma_e^2) \) has a limiting normal distribution under model M2.

**Model M2.1** Since the random walk model M2.1 is a special case of model M2 the above procedure can be used to estimate \( \sigma_e^2 \) and \( \sigma_p^2 \) in the random walk model (upon using the \( \rho=1 \) version of Anderson’s method).

**Estimation under Model M3**

As described earlier, in model M3

\[
X_i = \mu_0 + ih + \epsilon_i.
\]

This is the simple linear regression model with values for independent variate given in \([1, 2, \ldots, n]^t\) and \( h \) the regression coefficient. The maximum likelihood estimators
of \( \mu_0, h \) and \( \sigma_\varepsilon^2 \) are well known to be

\[
\hat{\mu}_0 = \left[ \bar{X} - \frac{(n+1)h}{2} \right] \tag{3.10}
\]

\[
\hat{h} = \frac{\sum_{i=1}^{n} iX_i - \frac{(\sum_{i=1}^{n} i)(\sum_{i=1}^{n} X_i)}{n}}{\sum_{i=1}^{n} i^2 - \frac{(\sum_{i=1}^{n} i)^2}{n}} = \frac{\sum_{i=1}^{n} iX_i - \frac{(n+1)}{2} \sum_{i=1}^{n} X_i}{\frac{n(n^2-1)}{12}} \tag{3.11}
\]

and

\[
\hat{\sigma}_\varepsilon^2(m1) = \frac{\sum_{i=1}^{n} X_i^2 - 12 \left[ \sum_{i=1}^{n} iX_i - \frac{(n+1)}{2} \sum_{i=1}^{n} X_i \right]^2}{n(n^2-1)} \tag{3.12}
\]

respectively.

**Estimation under Model M5**

Under the iid model M5

\[
X_i \sim \text{iid } N(\mu, \sigma_\varepsilon^2) \quad i = 1, 2, \ldots, n.
\]

Maximum likelihood estimation of \( \sigma_\varepsilon^2 \) is described in any book on the theory of statistical inference. The maximum likelihood estimator (MLE) of \( \sigma_\varepsilon^2 \) is

\[
\hat{\sigma}_\varepsilon^2(m5) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \tag{3.13}
\]

Since

\[
E\hat{\sigma}_\varepsilon^2(m5) = \frac{(n-1)}{n} \sigma_\varepsilon^2 \tag{3.14}
\]
\( \hat{\sigma}_e^2(ml5) \) is a biased estimator of \( \sigma_e^2 \) with bias

\[
B = \frac{\sigma_e^2}{n}.
\]  

(3.15)

Clearly as \( n \) increases the bias tends to zero.

The MLE \( \hat{\sigma}_e^2(ml5) \) has the \( \frac{\sigma_e^2}{n} \chi^2(n - 1) \) distribution under model M5. The variance of \( \hat{\sigma}_e^2(ml5) \) under model M5 is

\[
\text{Var}(\hat{\sigma}_e^2(ml5)) = \frac{2(n - 1)}{n^2} \sigma_e^4.
\]  

(3.16)

so that the mean squared error of \( \hat{\sigma}_e^2(ml5) \) is

\[
\text{MSE}(\hat{\sigma}_e^2(ml5)) = \frac{2(n - 1)}{n^2} \sigma_e^4 + \frac{\sigma_e^4}{n^2}
\]

(3.17)

\[
= \frac{(2n - 1)}{n^2} \sigma_e^4
\]

(3.18)

Standard arguments show that under model M5, \( \hat{\sigma}_e^2(ml5) \) is both almost surely consistent and asymptotically normal.

Some Properties of ML Estimators under Alternative Models

The Mean of \( \hat{\sigma}_e^2(ml3) \) under Models M4 and M5

\( \hat{\sigma}_e^2(ml3) \) can be written as

\[
\frac{1}{n}X'(I - P)X
\]

for \( P = D(D'D)^{-1}D' \), where \( D' = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
1 & 2 & \cdots & n
\end{bmatrix} \).

Searle (1971, p.55) shows that when \( X \) is \( N(\mu, V) \)

\[
E(X'AX) = tr(AV) + \mu'A\mu.
\]
So under model M5,

\[ E\sigma_e^2 = \frac{(n-2)}{n} \sigma_e^2 + \frac{(n-1^t P_1)}{n} \mu^2. \] \hspace{1cm} (3.19)

And, under the model M4

\[ E\sigma_e^2 = \frac{(n-2)}{n} (\sigma_e^2 + \sigma_\mu^2) + \frac{(n-1^t P_1)}{n} \mu^2. \]

The Mean of \( \sigma_e^2 \) under Models M2 and M3

The maximum likelihood estimator of \( \sigma_e^2 \) under the iid model M5 is

\[ \sigma_e^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}. \] \hspace{1cm} (3.20)

Under model M2,

\[ X_i = \rho^i \mu_0 + \sum_{j=1}^{i} \rho^{(i-j)} \nu_j + \epsilon_i \]

and

\[ \bar{X} = \frac{\sum_{j=1}^{n} \rho^j}{n} \mu_0 + \frac{\sum_{j=1}^{n} (1 - \rho^{n-(j-1)}) \nu_j}{n(1 - \rho)} + \tau, \]

so that

\[ (X_i - \bar{X}) = \left( \rho^i - \frac{\sum_{j=1}^{n} \rho^j}{n} \right) \mu_0 + \sum_{j=1}^{i} \rho^{(i-j)} \nu_j - \frac{\sum_{j=1}^{n} (1 - \rho^{n-(j-1)}) \nu_j}{n(1 - \rho)} + (\epsilon_i - \tau). \]
Since $E\nu_i$, $E\epsilon_i$, $E_{j\neq k}\nu_j\nu_k$, $E_{j\neq k}\epsilon_j\epsilon_k$ and $E\nu_j\nu_k$ are all zero,

\[
E \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} \left( \rho^i - \frac{\sum_{j=1}^{n} \rho^j}{n} \right)^2 \mu_0^2 + (n-1)\sigma_e^2 + \sigma^2 \left[ \sum_{i=1}^{n} \sum_{j=1}^{n} \rho^2(i-j) + \frac{\sum_{j=1}^{n} (1 - \rho^{n-(j-1)} \rho^2)}{n(1-\rho)^2} \right]
\]

\[
= \sum_{i=1}^{n} \left( \rho^i - \frac{\sum_{j=1}^{n} \rho^j}{n} \right)^2 \mu_0^2 + (n-1)\sigma_e^2 + \sigma^2 \left[ \sum_{j=0}^{(n-1)} (n-j)\rho^2j + \frac{1}{n} \sum_{j=1}^{n} \left( 1 + \rho + \rho^2 + \ldots + \rho^j \right)^2 \right]
\]

\[
= \sum_{i=1}^{n} \left( \rho^i - \frac{\sum_{j=1}^{n} \rho^j}{n} \right)^2 \mu_0^2 + (n-1)\sigma_e^2 + \sigma^2 \left[ \sum_{j=0}^{(n-1)} (n-j)\rho^2j + \frac{1}{n} \sum_{j=1}^{n} \left( \sum_{i=1}^{j} \rho^{i-1} \right)^2 \right],
\]

and

\[
E\sigma^2_{\epsilon}^{(m=5)} = \frac{1}{n} \sum_{i=1}^{n} \left( \rho^i - \frac{\sum_{j=1}^{n} \rho^j}{n} \right)^2 \mu_0^2 + \frac{(n-1)}{n} \sigma_e^2
\]
If \( \rho = 1 \), we are in the random walk model M2.1. For this model

\[
E \hat{\sigma}_e^2(ml5) = \frac{(n - 1)}{n} \sigma_e^2 + \frac{2}{3} \left[ \sum_{i=1}^{n} i - \frac{1}{n} \sum_{i=1}^{n} i^2 \right].
\]  

(3.22)

Since

\[
\sum_{i=1}^{n} i = \frac{n(n + 1)}{2}
\]

and

\[
\sum_{i=1}^{n} i^2 = \frac{n(n + 1)(2n + 1)}{6}.
\]

(3.23)

(3.24)

(3.22) can be shown to yield

\[
E \hat{\sigma}_e^2(ml5) = \frac{(n - 1)}{n} \sigma_e^2 \left[ 1 + \frac{(n + 1)}{6} \frac{\sigma_v^2}{\sigma_e^2} \right].
\]

(3.25)

For large \( n \), \( \hat{\sigma}_e^2(ml5) \) is badly biased (upward) as an estimator of \( \sigma_e^2 \) under model M2.1.

Under model M3

\[
E \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = h^2 \frac{1}{n} \left[ \sum_{i=1}^{n} i - \frac{i=1}{n} \right]^2 + (n - 1) \sigma_e^2.
\]

(3.26)

Since

\[
\sum_{i=1}^{n} \left[ i - \frac{i=1}{n} \right]^2 = \frac{n(n^2 - 1)}{12}.
\]

\[
E \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 = h^2 \frac{n(n^2 - 1)}{12} + (n - 1) \sigma_e^2.
\]
and

\[ E\hat{\sigma}_e^2 (ml5) = \frac{(n-1)}{n} \left[ \sigma_e^2 + \frac{n(n+1)}{12} h^2 \right]. \]

Once again, for large \( n \), \( \hat{\sigma}_e^2 (ml5) \) is badly biased (upward) as an estimator of \( \sigma_e^2 \) under model M3.
CHAPTER 4. BAYESIAN ESTIMATION

Based on a random quantity $X \sim f(x|\theta)$, using Bayesian ideas to draw inferences about a real parameter $\theta$, we assume that the range of possible values for $\theta$ can be specified and we have some apriori idea about probabilities for $\theta$. That is, in a Bayesian formulation of the estimation problem, the parameter $\theta$ is treated as a random variable with probability density function $\pi(\theta)$, known as the prior density of $\theta$. We can derive the conditional density of $\theta$ given that $X = x$, $\pi(\theta|x)$, called the posterior density of $\theta$ as

$$\pi(\theta|x) = \frac{f(x, \theta)}{f(x)} = \frac{f(x|\theta)}{\int f(x|\theta)\pi(\theta)\,d\theta}$$

That is, $\pi(\theta)$ is updated through the use of $f(x|\theta)$. This posterior density $\pi(\theta|x)$ can be used in different ways. The conditional or posterior mean of $\theta$ given $X = x$,

$$E(\theta|x) = \int \theta \pi(\theta|x)\,d\theta$$

is a function only of $x$ and is the optimal estimator of $\theta$ under squared error loss. Henceforth, the Bayes estimator means the conditional mean of $\theta$ given $x$.

Estimation of the variance $\sigma^2$ using Bayesian approaches has been discussed by Press(1989), Searle(1992) and Berger(1980) for a single sample of size $m$. There,
$X_1, X_2, \ldots, X_m$ are $II\text{N}(\mu, \sigma^2)$ and a natural conjugate prior distribution for $\sigma^2$ is the inverse gamma distribution. Press (1989, page 52) has also discussed the use of "vague" priors for variances of the form

$$g(\sigma^2) \propto \frac{1}{\sigma^2}.$$  

The calculation of posterior densities can be complicated and is sometimes not possible in a closed form. Recently, many advances have been made in numerical and analytic approximation methods for the posterior densities needed in Bayesian inference (e.g. Lindley (1980), Tierney and Kadane (1984), Geweke (1988), Stewart (1984), Naylor and Smith (1982,1984), Shaw (1988), Smith et. al. (1987), Smith, Shaw, Naylor, and Dransfield (1985). But implementation of these techniques requires sophisticated numerical analytic expertise and specialized software. In another development, Gelfand and Smith (1990) have discussed at length the properties of sampling based approaches, stochastic substitution, Gibbs sampling and the sampling-importance-resampling algorithms. We proceed to discuss the Gibbs sampling algorithm in some detail, since we shall use it to approximate marginal posterior means.

### The Gibbs Sampling Algorithm

Geman and Geman (1984) introduced the algorithm that has become known as the Gibbs sampler. This has been widely used in the analysis of complex stochastic models involving large numbers of variables. It is useful when a joint distribution cannot be specified directly, but a set of full conditionals is available. For a vector of observed values $\mathbf{x}$ and an unknown vector of parameters $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$, a
posterior distribution \( \pi(\Theta|x) \) can be calculated, at least up to proportionality, by multiplying the likelihood function by the prior density for \( \Theta \). From this, we can find full conditional distributions \( \pi(\theta_i|x; \theta_j; j \neq i) \).

The Gibbs sampler is a Markovian updating scheme. The Gibbs sampling scheme works as follows: Given some arbitrary starting values \( \Theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}) \), we draw

\[
\begin{align*}
\theta_1^{(1)} &\sim \pi(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)}, x) \\
\theta_2^{(1)} &\sim \pi(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)}, x) \\
&\vdots \\
\theta_k^{(1)} &\sim \pi(\theta_k|\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{k-1}^{(1)}, x)
\end{align*}
\]

Every variable is visited in natural order. In Gibbs sampling, a cycle requires \( k \) random variate generations to move from \( \Theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}) \) to \( \Theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_k^{(1)}) \). In similar fashion, after \( i \) such iterations, we get \( \Theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)}) \).

Under mild regularity conditions, Geman and Geman have proved properties of the Gibbs Sampler that in the present context can be stated in terms of the posterior distribution \( \pi(\Theta|x) \).

(i) Convergence:

\[
\Theta^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)}) \xrightarrow{d} (\theta_1, \theta_2, \ldots, \theta_k) \sim \pi(\theta_1, \theta_2, \ldots, \theta_k|x)
\]

and for any \( s \)

\[
\theta_s^{(i)} \xrightarrow{d} \theta_s \sim \pi(\theta_s|x).
\]
(ii) **Rate of Convergence:** Under the sup norm, the joint density of \((\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)})\) given \(x\) converges to the joint density of \((\theta_1, \theta_2, \ldots, \theta_k)\) given \(x\) at a geometric rate in \(i\).

(iii) **Ergodicity:** For any measurable function \(g(\Theta)\)

\[
\lim_{i \to 0} \frac{1}{i} \sum_{l=1}^{i} g(\theta_1^{(l)}, \theta_2^{(l)}, \ldots, \theta_k^{(l)}) \xrightarrow{a.s.} E[g(\theta_1, \theta_2, \ldots, \theta_k)|x]
\]

provided \(E[g(\theta_1, \theta_2, \ldots, \theta_k)|x]\) exists.

A simulated sample of size \(m\) from the marginal distribution of \(\theta_s\) can be taken by repeating the procedure \(m\) times giving \(m\) iid \(k\)-tuples \((\theta_1^{(i)}, \ldots, \theta_k^{(i)})\), \(j = 1, 2, \ldots, m\). For a large \(i\), \((\theta_1^{(i)}, \ldots, \theta_k^{(i)})\) can be taken as a sample of size \(m\) from the posterior distribution of \(\theta_s|x\). Detailed discussions of properties of the Gibbs sampler are given by Gelfand and Smith (1990), Besag and Green (1993) and Gilks et al. (1993).

**An Example of Gibbs Sampling for a Simple Two Parameter Normal Model**

We now establish the form of the Gibbs sampler for a simple 2 parameter model. Consider the model

\[
X = \lambda \Delta + \zeta
\]

where \(\zeta \sim MVN(\mathbf{0}, \sigma^2 \mathbf{G})\). We suppose \(\Delta\) and \(G\) are given, while \(\lambda\) and \(\sigma^2\) are unknown parameters. (Assuming certain parametric functions to be known, models M2, M4 and M5 can be thought as special cases of this model.) We consider the following independent priors:

1. \(\lambda \sim N(\lambda_0, \sigma^2_\lambda)\), and
2. \( \sigma_\varepsilon^2 \sim IG(\alpha, \beta) \).

for \( \sigma_\lambda^2, \alpha \) and \( \beta \) known. Gelfand and Smith (1990) have discussed how to calculate the marginal posterior densities in this model when \( G = I \) and \( \Delta = 1 \) under various priors.

The joint density of \((X, \lambda, \sigma_\varepsilon^2)\) is

\[
f(x, \lambda, \sigma_\varepsilon^2 | \lambda_0, \sigma_\lambda^2, \alpha, \beta) = f(x | \lambda, \sigma_\varepsilon^2) \pi(\lambda | \lambda_0, \sigma_\lambda^2) \pi(\sigma_\varepsilon^2 | \alpha, \beta)
\]

This is a two variable \((\lambda, \sigma_\varepsilon^2)\) system. To apply Gibbs sampler we need the conditionals \( \pi(\sigma_\varepsilon^2 | x, \lambda) \) and \( \pi(\lambda | x, \sigma_\varepsilon^2) \). The full conditionals are identified as follows.

\[
f(x, \lambda, \sigma_\varepsilon^2) = \frac{e^{-\frac{1}{2\sigma_\varepsilon^2}(x-\lambda \Delta)' G^{-1}(x-\lambda \Delta)}}{(2\pi)^n/2 \sigma_\varepsilon^{n/2} |G|^{1/2}} \times \frac{1}{\sqrt{2\pi \sigma_\lambda^2}} e^{(\lambda-\lambda_0)^2/2\sigma_\lambda^2} \times \frac{\sigma_\varepsilon^{2-(\alpha+1)}}{\Gamma(\alpha \beta) \sigma_\varepsilon^2}.
\]

Then, considered as a function of \( \sigma_\varepsilon^2 \)

\[
f(x, \lambda, \sigma_\varepsilon^2) \propto \frac{\sigma_\varepsilon^{2-(\alpha+\frac{n}{2}+1)}}{(2\pi)^{(n+1)/2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \left[ \frac{1}{\beta} + \frac{(x-\lambda \Delta)' G^{-1}(x-\lambda \Delta)}{2} \right]}.
\]

So \( \pi(\sigma_\varepsilon^2 | x, \lambda) \sim IG \left( \alpha + \frac{n}{2}, \left[ \frac{1}{\beta} + \frac{(x-\lambda \Delta)' G^{-1}(x-\lambda \Delta)}{2} \right]^{-1} \right) \).

Similarly

\[
f(x, \lambda, \sigma_\varepsilon^2) \propto k_{\varepsilon} \left( \frac{\lambda_0}{\sigma_\lambda^2} + \frac{\Delta' G^{-1} \Delta}{\sigma_\varepsilon^2} \right)^{\lambda^2 - 2\lambda} \left( \frac{\lambda_0}{\sigma_\lambda^2} + \frac{\Delta' G^{-1} x}{\sigma_\varepsilon^2} \right) \left( \frac{\lambda_0}{\sigma_\lambda^2} + \frac{\Delta' G^{-1} \Delta}{\sigma_\varepsilon^2} \right)^{\lambda^2 - 2\lambda - 2}.
\]
giving
\[
\pi(\lambda | \underline{\mu}, \sigma_e^2) \sim N \left[ \frac{\frac{\lambda_0}{\sigma_\lambda^2} + \frac{\Delta'G^{-1}x}{\sigma_e^2}}{\frac{1}{\sigma_\lambda^2} + \frac{\Delta'G^{-1}\Delta}{\sigma_e^2}}, \left[ \frac{1}{\sigma_\lambda^2} + \frac{\Delta'G^{-1}\Delta}{\sigma_e^2} \right]^{-1} \right].
\]

In Gibbs sampling for this model one starts with an initial values of \(\lambda, \lambda(0)\).

For example, a suitable value might be the arithmetic mean, \(\underline{x}\). When iteration starts, a value \(\sigma_e^2(1)\) is drawn from \(\pi(\sigma_e^2 | \underline{\mu}, \lambda(0))\). Using \(\sigma_e^2(1), \lambda(1)\) is drawn from \(\pi(\lambda | \underline{\mu}, \sigma_e^2(1))\). So the first iteration is completed. The second iteration starts with \(\lambda(1)\). Although this system permits inference about both \(\lambda\) and \(\sigma_e^2\), our primary interest lies in the estimation of \(\sigma_e^2\). Under squared error loss \(E(\sigma_e^2 | \underline{x})\) is the Bayes estimator of \(\sigma_e^2\). By the virtue of Ergodicity property of the Gibbs sampler, for large \(I\), \(\frac{1}{I} \sum_{l=1}^{I} \sigma_e^2(l)\) is a sensible approximation of the Bayes estimator \(E(\sigma_e^2 | \underline{x})\).

**Bayes Estimation under Model M2**

Recall from page 4 that here \(X \sim MVN(\Psi \mu_0, \Sigma)\) where \(\Psi = [\rho, \rho^2, \ldots, \rho^n]'\) and \(\Sigma = \sigma_e^2 I_n + \sigma_\eta^2 V_n\). Under the assumptions that both \(\rho\) and \(\eta = \frac{\sigma_\eta^2}{\sigma_e^2}\) are known, \(\Sigma\) can be written as

\[
\Sigma = \sigma_e^2 (I_n + \eta V_n) = \sigma_e^2 G
\]

and model M2 is a special case of the 2 parameter model (4.4). Thus under the priors

\[
\mu_0 \sim N(\bar{\mu}, \sigma_0^2) \quad \text{and} \quad \sigma_e^2 \sim IG(\alpha, \beta)
\]

the scheme just discussed produces an approximately Bayes estimator for \(\sigma_e^2, \sigma_e^2(62)\).
Bayes Estimation under Model M3

As mentioned in Chapter 3, M3 is a special case of the simple linear regression model. Press (1989, page 125) has discussed the estimation of $\sigma^2_e$ under vague priors. We consider a procedure based on the Gibbs Sampler to approximate the Bayes estimator of $\sigma^2_e$ under other priors. In model M3 there is simple linear dependency between $X_i$ and $i$, i.e.

$$X_i|i = \mu_0 + hi + \epsilon_i \quad i = 1,2,\ldots,n. \quad (4.5)$$

where $\epsilon_i \sim \text{IN}(0, \sigma^2_e)$. We consider the following independent priors on the parameters of model M3:

1. $\mu_0 \sim N(\mu_0, \sigma^2_1)$,
2. $h \sim N(\hat{h}, \sigma^2_2)$, and
3. $\sigma^2_e \sim IG(\alpha, \beta),$

where $\mu_0, \hat{h}, \sigma^2_1, \sigma^2_2, \alpha$ and $\beta$ are all known.

The joint density of $X, \mu_0, h$ and $\sigma^2_e$ is

$$f(x, \mu_0, h, \sigma^2_e) = f(x|\mu_0, h)\pi(\mu_0|\mu_0, \sigma^2_1)\pi(h|\hat{h}, \sigma^2_2)\pi(\sigma^2_e|\alpha, \beta).$$

This is a three variable $(\mu_0, h, \sigma^2_e)$ system. For the Gibbs sampler, the conditionals

$$\pi(\sigma^2_e|x, \mu_0, h), \pi(\mu_0|x, h, \sigma^2_e) \text{ and } \pi(h|x, \mu_0, \sigma^2_e)$$
are needed. In a manner similar to that used in the analysis for model (4.4)

\[
\pi(\sigma^2_e|x, \mu_0, h) \sim IG\left( \left( \frac{n}{2} + \alpha \right), \left[ \frac{1}{\beta} + \frac{\sum (x_i - \mu_0 - h^2)^2}{2} \right]^{-1} \right),
\]

\[
\pi(\mu_0|x, \sigma^2_e, h) \sim N \left[ \frac{n \sigma^2_e (\bar{x} - \frac{(n+1)h}{2}) + \sigma^2_e \mu_0}{(n \sigma^2_e + \sigma^2_e)}, \frac{\sigma^2_e}{(n \sigma^2_e + \sigma^2_e)} \right],
\]

and

\[
\pi(h|x, \sigma^2_e, \mu_0) \sim N \left[ \frac{\sigma^2_e (\sum x_i - \mu_0 \sum i) + \sigma^2_e h}{\sigma^2_e \sum i^2 + \sigma^2_e}, \frac{\sigma^2_e}{\sigma^2_e \sum i^2 + \sigma^2_e} \right].
\]

We start with initial values of \(\mu_0\) and \(h\), \(\mu_0^{(0)}\), and \(h^{(0)}\). One set of possible candidates for these is the set of maximum likelihood estimators. On the first iteration a value of \(\sigma^2_e^{(1)}\) is taken from \(\pi(\sigma^2_e|x, \mu_0^{(0)}, h^{(0)})\), then a value of \(\mu_0^{(1)}\) is drawn from \(\pi(\mu_0|x, \sigma^2_e^{(1)}, h^{(0)})\). The first iteration is completed by randomly selecting \(h\) from \(\pi(h|x, \sigma^2_e^{(1)}, \mu_0^{(1)})\). The second iteration starts with \(\mu_0^{(1)}\) and \(h^{(1)}\). The approximately Bayes estimator of \(\sigma^2_e\) is then (for large \(I\)) \(\sigma^2_e^{(b3)} = \frac{1}{I} \sum_{l=1}^{I} \sigma^2_e^{(l)}\).

Bayes Estimation under Model M4

Model M4 can be viewed as another special case of the model (4.5) discussed above. For model M4

\[ X \sim N(\mu_0, \sigma^2 I_n) \]

where \(\sigma^2 = \sigma^2_{\mu} + \sigma^2_e = (1 + \bar{y})\sigma^2_e\) with \(\bar{y} = \frac{\sigma^2_{\mu}}{\sigma^2_e}\). We consider the situation where the ratio \(\bar{y}\) is known. Under the priors

\[ \mu_0 \sim N(\mu_0, \sigma^2_{\mu}) \]
and
\[ \sigma^2 \sim IG(\alpha, \beta) \]

the Gibbs sampler yields an approximately Bayes estimator of \( \sigma^2 \), say \( \hat{\sigma}^2(b^4) \); and

the Bayes estimator of \( \sigma^2_{\varepsilon} \) is
\[ \hat{\sigma}^2_{\varepsilon}(b^4) = \frac{\hat{\sigma}^2(b^4)}{1 + \eta}. \]

**Bayes Estimation under Model M5**

Model M5 deals with iid assumptions. Replacing \( \Delta \) and \( G \) by \( 1 \) and \( I \) respectively in the model (4.5) produces model M5, and an approximately Bayes estimator of \( \sigma^2_{\varepsilon} \) is then evident.

If we have prior for \( \sigma^2_{\varepsilon} \) only and \( \mu \) is assumed to be known, a closed form exists for the posterior distribution of \( \sigma^2_{\varepsilon} \). That is, with \( \mu \) known, \( X \sim N(\mu, \sigma^2_{\varepsilon}) \) and the prior \( \sigma^2_{\varepsilon} \sim IG(\alpha, \beta) \), the posterior distribution of \( \sigma^2_{\varepsilon} \) is

\[ g(\sigma^2_{\varepsilon}|\varepsilon) \sim IG\left[(\alpha + \frac{n}{2}), \left(\frac{1}{\beta} + \frac{\sum (x_i - \mu)^2}{2}\right)^{-1}\right] \]

Under the squared error loss function, the Bayes estimator is given by

\[ \hat{\sigma}^2_{\varepsilon}(b^5) = E(\sigma^2_{\varepsilon}|\varepsilon) = \frac{\left(\frac{1}{\beta} + \frac{\sum (x_i - \mu)^2}{2}\right)}{(\alpha + \frac{n}{2} - 1)} \]

Under model M5 this estimator has

\[ E\hat{\sigma}^2_{\varepsilon}(b^5) = \frac{\left(\frac{1}{\beta} + \frac{n\sigma^2_{\varepsilon}}{2}\right)}{(\alpha + \frac{n}{2} - 1)}, \]
\[ \text{Bias} = \frac{\left( \frac{1}{\beta} - (\alpha - 1)\sigma_e^2 \right)}{\left( \alpha + \frac{n}{2} - 1 \right)}, \]

\[ \text{Var}(\hat{\sigma}^2_e) = \frac{\left( \frac{n}{2} \sigma_e^4 \right)}{\left( \alpha + \frac{n}{2} - 1 \right)^2}, \]

and

\[ \text{MSE} = \frac{\left( \frac{1}{\beta} - (\alpha - 1)\sigma_e^2 \right)^2 + \left( \frac{n}{2} \sigma_e^4 \right)}{\left( \alpha + \frac{n}{2} - 1 \right)^2}. \]
CHAPTER 5. MONTE CARLO STUDIES

As is clear from previous chapters, not all of the estimators studied in this thesis have moments that are expressible in closed form. We have thus used simulation techniques to study the performance of the estimators under a number of sets of parameters for the models introduced in Chapter 1.

Description of the Models Used in the Simulation

Throughout our simulation study and without loss of generality, $\sigma^2_{\varepsilon}$ has been taken to be 1.0. Our results could be appropriately scaled to give moments for other values of $\sigma^2_{\varepsilon}$. A description of the parameter values used for the models M1, M2, M3, M4 and M5 in our simulation follows.

Simulation Parameters for Model M1

Under this model, the population means are not the same, but are nonrandom. We investigated the following six deterministic patterns for the means.

1. $M_1^{(1)}$: Oscillating means, in particular 0, -2, -5, -2, 0, 2, 5, 2, 0, ....

2. $M_1^{(2)}$: For the first half of a sequence, the means are zero and for the second half means are 5. That is 0, 0, .... 0, 5, 5, .... 5.
3. $M_1(3)$: Alternating means, in particular -1, +1, -1, +1, ....

4. $M_1(4)$: For the first half of a sequence a linear drift is 3 while for the second half the drift is -3.

5. $M_1(5)$: For the first half of a sequence a linear drift is 3 while for the second half the drift is -1.

6. $M_1(6)$: For the first half of a sequence a linear drift is 1 while for the second half the drift is -1.

Simulation Parameters for Model M2

In this model the means $\mu_i$ follow an AR(1) model. The random walk model, M2.1, is a special case where $\rho = 1.0$. This model involves three parameters $\mu_0$, $\rho$, $\sigma^2_p$ besides $\sigma^2_v$. Table 5.1 gives different sets of values used for the parameters involved. In all cases, $\mu_0 = 0$.

Simulation Parameters for Model M3

For the linear drift model M3, $\mu_0 = 0$ and three values of the drift parameter, $h$, are 0.1, 0.5, 1.0 for models $M_3(1)$, $M_3(2)$ and $M_3(3)$ respectively.

Simulation Parameters for Model M4

The means are assumed to be $\mathcal{I}N(0, \sigma^2_H)$. Choices of parameters $\sigma^2_H = 0.5, 1.0$ and 2.0 give models $M_4(1)$, $M_4(2)$ and $M_4(3)$ respectively.
Table 5.1: Models M2 Used in the Simulations

<table>
<thead>
<tr>
<th>Model</th>
<th>Means follow AR(1) with</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho$</td>
</tr>
<tr>
<td>$M_2(1)$</td>
<td>0.25</td>
</tr>
<tr>
<td>$M_2(2)$</td>
<td>0.25</td>
</tr>
<tr>
<td>$M_2(3)$</td>
<td>0.25</td>
</tr>
<tr>
<td>$M_2(4)$</td>
<td>-0.25</td>
</tr>
<tr>
<td>$M_2(5)$</td>
<td>-0.25</td>
</tr>
<tr>
<td>$M_2(6)$</td>
<td>-0.25</td>
</tr>
<tr>
<td>$M_2(7)$</td>
<td>0.50</td>
</tr>
<tr>
<td>$M_2(8)$</td>
<td>0.50</td>
</tr>
<tr>
<td>$M_2(9)$</td>
<td>0.50</td>
</tr>
<tr>
<td>$M_2(10)$</td>
<td>-0.50</td>
</tr>
<tr>
<td>$M_2(11)$</td>
<td>-0.50</td>
</tr>
<tr>
<td>$M_2(12)$</td>
<td>-0.50</td>
</tr>
<tr>
<td>$M_2(13)$</td>
<td>1.0</td>
</tr>
<tr>
<td>$M_2(14)$</td>
<td>1.0</td>
</tr>
<tr>
<td>$M_2(15)$</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Simulation Parameters for Model M5

For this IID model, $\mu=0$ and $\sigma^2_c=1$.

Prior Distributions Used to Produce Bayes Estimators

The most commonly used prior for a normal variance is the Inverse Gamma distribution. If $Y \sim IG(\alpha, \beta)$, then $E(Y) = \frac{1}{(\alpha-1)\beta}$ and $Var(Y) = \frac{1}{(\alpha-1)^2(\alpha-2)\beta^2}$. Many choices of $\alpha$ and $\beta$ are possible. We have taken $\alpha=3$ and varied $\beta$ to get different means (and variances). Since $\sigma^2_c=1$, the following priors have been used for
\[ \sigma^2_e: \]
\[ \sigma^2_e \sim IG(3, 0.1), \]
\[ \sigma^2_e \sim IG(3, 0.5), \text{ and} \]
\[ \sigma^2_e \sim IG(3, 2.5). \]

These priors have means respectively 5 times \( \sigma^2_e \), equal to \( \sigma^2_e \) and 1/5 of \( \sigma^2_e \).

To specify estimators that are Bayes under M2, we have considered the random walk model M2.1 with \( \eta=0.5, 1.0 \text{ and } 2.0 \). We have then considered estimators that are Bayes in this model, using the prior for \( \mu_0 \)
\[ \mu_0 \sim N(0, 1) \]
and priors for \( \sigma^2_e \)
\[ \sigma^2_e \sim IG(3, 0.1), \]
\[ \sigma^2_e \sim IG(3, 0.5), \text{ and} \]
\[ \sigma^2_e \sim IG(3, 2.5). \]

where apriori all parameters are independent.

To specify a Bayes estimator of \( \sigma^2_e \) under the linear drift model M3, the following independent priors were used for \( \mu_0 \) and \( h \).
\[ \mu_0 \sim N(0, 1) \]
\[ h \sim N(0.5, 1) \]

Bayes estimators for model M4 have been studied using the priors for \( \sigma^2 \). The prior for \( \mu_0 \) is \( N(0,1) \). In model M4, \( \sigma^2 = \sigma^2_\mu + \sigma^2_e = \sigma^2_e (1 + \eta) \) can take three values.
1.5, 2.0 and 3.0 for \( \eta = 0.5, 1.0 \) and 2.0 respectively. We have considered priors for \( \sigma^2 \) as

\[
\begin{align*}
\sigma^2 &\sim IG(3,1/15), \\
\sigma^2 &\sim IG(3,1/3), \text{ and} \\
\sigma^2 &\sim IG(3,5/3).
\end{align*}
\]

having means respectively 5 times \( \sigma^2 \), equal to \( \sigma^2 \) and 1/5 of \( \sigma^2 \) when \( \eta = 0.5 \)

\[
\begin{align*}
\sigma^2 &\sim IG(3,0.05). \\
\sigma^2 &\sim IG(3,0.25), \text{ and} \\
\sigma^2 &\sim IG(3,1.25).
\end{align*}
\]

having means respectively 5 times \( \sigma^2 \), equal to \( \sigma^2 \) and 1/5 of \( \sigma^2 \) when \( \eta = 1.0 \) and

\[
\begin{align*}
\sigma^2 &\sim IG(3,1/30) \\
\sigma^2 &\sim IG(3,1/6), \text{ and} \\
\sigma^2 &\sim IG(3,5/6),
\end{align*}
\]

having means respectively 5 times \( \sigma^2 \), equal to \( \sigma^2 \) and 1/5 of \( \sigma^2 \) when \( \eta = 2.0 \).

**Description of Estimators Used**

The following estimators have been used in the Monte Carlo study.

1. \( \hat{\sigma}_e \) from Chapter 2.

2. The likelihood estimator of \( \sigma_e^2 \) under the random walk model M2.1, from Anderson(1973).
3. The maximum likelihood estimator of $\sigma^2_\epsilon$ under model M3.

4. The maximum likelihood estimator of $\sigma^2_\epsilon$ under model M5.

5. Model M2.1 Bayes estimators supposing $\eta=0.5, 1.0$ and 2.0.


7. Model M4 Bayes estimators of $\sigma^2_\epsilon$ based on priors for $\sigma^2$.

8. Model M5 Bayes estimators for $\sigma^2_\epsilon$.

More precisely, the estimators studied and some abbreviations for them follow.

1. MR: $\hat{\sigma}^2_\epsilon = (\bar{M}_n/d_2)^2$.

2. ML2: Likelihood estimator of $\sigma^2_\epsilon$ under model M2.1. Anderson (1973), $\hat{\sigma}^{(ml2)}_\epsilon$.

   The iterative procedure was stopped when consecutive estimates did not differ by more than 0.001. Negative estimate, if any, was taken as zero.

3. ML3: The ML estimator of $\sigma^2_\epsilon$ under model M3. $\hat{\sigma}^{(ml3)}_\epsilon$.

4. ML5: The ML estimator of $\sigma^2_\epsilon$ under model M5. $\hat{\sigma}^{(ml5)}_\epsilon$.

5. B2a.1: The approximately Bayes estimator of $\sigma^2_\epsilon$ under model M2, with $\rho = 1.0$, $\eta = 0.5$ and priors $\mu_0 \sim N(0,1)$. $\sigma^2_\epsilon \sim IG(3,0.1)$.

6. B2a.2: The approximately Bayes estimator of $\sigma^2_\epsilon$ under model M2, with $\rho = 1.0$, $\eta = 0.5$ and priors $\mu_0 \sim N(0,1)$. $\sigma^2_\epsilon \sim IG(3,0.5)$.

7. B2a.3: The approximately Bayes estimator of $\sigma^2_\epsilon$ under model M2, with $\rho = 1.0$, $\eta = 0.5$ and priors $\mu_0 \sim N(0,1)$. $\sigma^2_\epsilon \sim IG(3,2.5)$. 
8. B2b.1: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M2, with $\rho = 1.0$, 
$\eta = 1.0$ and priors $\mu_0 \sim N(0, 1)$, $\sigma_\varepsilon^2 \sim IG(3, 0.1)$.

9. B2b.2: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M2, with $\rho = 1.0$, 
$\eta = 1.0$ and priors $\mu_0 \sim N(0, 1)$, $\sigma_\varepsilon^2 \sim IG(3, 0.5)$.

10. B2b.3: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M2, with $\rho = 1.0$, 
$\eta = 1.0$ and priors $\mu_0 \sim N(0, 1)$, $\sigma_\varepsilon^2 \sim IG(3, 2.5)$.

11. B2c.1: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M2, with $\rho = 1.0$, 
$\eta = 2.0$ and priors $\mu_0 \sim N(0, 1)$, $\sigma_\varepsilon^2 \sim IG(3, 0.1)$.

12. B2c.2: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M2, with $\rho = 1.0$, 
$\eta = 2.0$ and priors $\mu_0 \sim N(0, 1)$, $\sigma_\varepsilon^2 \sim IG(3, 0.5)$.

13. B2c.3: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M2, with $\rho = 1.0$, 
$\eta = 2.0$ and priors $\mu_0 \sim N(0, 1)$, $\sigma_\varepsilon^2 \sim IG(3, 2.5)$.

14. B3.1: The approximately Bayes estimator under model M3 with priors $h \sim N(0.5, 1.0)$, 
$\mu_0 \sim N(0, 1)$ and $\sigma_\varepsilon^2 \sim IG(3, 0.1)$.

15. B3.2: The approximately Bayes estimator under model M3 with priors $h \sim N(0.5, 1.0)$, 
$\mu_0 \sim N(0, 1)$ and $\sigma_\varepsilon^2 \sim IG(3, 0.5)$.

16. B3.3: The approximately Bayes estimator under model M3 with priors $h \sim N(0.5, 1.0)$, 
$\mu_0 \sim N(0, 1)$ and $\sigma_\varepsilon^2 \sim IG(3, 2.5)$.

17. B4a.1: The approximately Bayes estimator of $\sigma_\varepsilon^2$ under model M4 with $\eta = 0.5$ 
and priors $\mu_0 \sim N(0, 1)$, $\sigma^2 \sim IG(3, 1/15)$. 
18. B4a.2: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 0.5$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 1/3)$.

19. B4a.3: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 0.5$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 5/3)$.

20. B4b.1: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 1.0$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 0.05)$.

21. B4b.2: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 1.0$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 0.25)$.

22. B4b.3: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 1.0$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 1.25)$.

23. B4c.1: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 2.0$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 1/30)$.

24. B4c.2: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 2.0$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 1/6)$.

25. B4c.3: The approximately Bayes estimator of $\sigma^2_e$ under model M4 with $\eta = 2.0$ and priors $\mu_0 \sim N(0, 1), \sigma^2 \sim IG(3, 5/6)$.

26. B5.1: The Bayes estimator of $\sigma^2_e$ under model M5 with $\mu$ taken to be 0 and prior $\sigma^2_e \sim IG(3.0, 0.1)$.

27. B5.2: The Bayes estimator of $\sigma^2_e$ under model M5 with $\mu$ taken to be 0 and prior $\sigma^2_e \sim IG(3.0, 0.5)$. 
28. B5.3: The Bayes estimator of $\sigma_2^2$ under model M5 with $\mu$ taken to be 0 and prior $\sigma_2^2 \sim IG(3.0, 2.5)$.

For all Bayes estimators above based on the Gibbs sampler, $I = 100$ and 200 estimates were used to estimate $MSE$. Maximum likelihood estimates were used as to initialize the Gibbs sampling algorithm.

Analysis and Conclusions

As described above, we used 28 estimators and 28 sets of parameters for the five different models described earlier. These sets of parameters cover a wide range of circumstances.

Overall Comparisons

Figure 5.1 shows an overall comparison of performances of these estimators for $n = 20$. One can observe that the estimators B2, i.e. the Bayes estimators of $\sigma_2^2$ under the random walk model perform best across the set of models we have employed. The ML3 estimator stands next to MR and the ML2 estimators in terms of their performances. The ML5 estimators have done very poorly.

When $n = 50$, except for the MR estimator, the ranges of MSE's produced by the estimators increase as compared to those for $n = 20$. Figure 5.2 shows that overall, the MR estimator outperforms the ML2 and ML3 estimators for this larger number of observations. If the MR estimator is used, the value of MSE appears to decrease with increasing $n$. Otherwise, the overall picture remains the same. Estimators B4 appear to be doing well but have very large values of MSE for some models.
Figure 5.1: Overall Comparison of Estimators $n = 20$
Figure 5.2: Overall Comparison of Estimators $n = 50$
Now, we look at each model separately and compare the performances of different estimators.

**Comparisons under $M_{1(1)}$**

(i) For $n = 20$ Fig. 5.3 shows that the random walk Bayes estimator (with $\sigma^2_p = 2.0$) $B_{2c.3}$ gives the lowest MSE. Among non-Bayesian estimators, ML2 performs slightly better than the MR estimator. Among other Bayes estimators, $B_{4c.2}$ and $B_{4c.3}$ do equally well. ML3 and ML5 do not perform well.

(ii) For $n = 50$ Fig. 5.4 shows that the estimators $B_{4b.2}$ and $B_{4b.3}$ perform best. The random walk Bayes estimator, $B_{2c.3}$ also does well. Among non-Bayesian estimators, the MR estimator has the lowest MSE and outperforms ML2.

**Comparisons under $M_{1(2)}$**

(i) For $n = 20$ Fig. 5.3 shows that again $B_{2b.3}$ and $B_{2c.2}$ perform best overall. Among non-Bayesian estimators, ML2 and MR give the lowest MSE.

(ii) For $n = 50$ Fig. 5.4 shows that ML2 and MR outperform all other estimators. Among the Bayes estimators, $B_{3.3}$, $B_{4c.2}$ and $B_{4c.3}$ have the lowest value of MSE.

**Comparisons under $M_{1(3)}$**

(i) For $n = 20$ Fig. 5.3 shows that among non-Bayesian estimators, ML3 does best. Overall $B_{2c.3}$ and $B_{4b.2}$ give the lowest values of MSE.

(ii) For $n = 50$ Fig. 5.4 shows that ML3 and ML5 perform best among non-Bayesian estimators, but overall $B_{2c.3}$ and $B_{4b.2}$ give the lowest values of MSE.

In brief, among non-Bayesian estimators ML2 and MR perform best for $n=20,$
Figure 5.3: Comparison of Estimators for Model M1 $n = 20$
Figure 5.4: Comparison of Estimators for Model M1 $n = 50$
but ML2 gives higher values of MSE for $n=50$. In case of alternating population means, $M_1(3)$, estimator ML3 does well. Among Bayes estimators, B2b.2 and B2c.3 do well when $n=20$. Performance of ML5 is poor for these three models.

**Comparisons under $M_1(4)$, $M_1(5)$**

Fig. 5.5 and Fig. 5.6 show that estimator ML2 outperforms all other estimators for $n=20$ as well as $n=50$. Next stands MR. Among Bayes estimators, estimators B2c give the lowest values of MSE. ML3, ML5 and B4 perform poorly.

**Comparisons under $M_1(6)$**

Fig. 5.5 and Fig. 5.6 show that MR gives the smallest value of MSE and next comes ML2 for $n=50$, but that Bayes estimators B2 outperform these for $n=20$. Among Bayes estimators, the estimators B2c are best (for both $n=20$ and $n=50$).

**Comparisons under $M_2(1)$, $M_2(2)$, $M_2(3)$**

Fig. 5.7 and Fig. 5.8 show that among non-Bayesian estimators, ML3 does best and next stands ML5, while MR and ML2 perform poorly (for both $n=20$ and $n=50$). Overall, the Bayes estimators outperform the others. The estimators B2 and B4 work best with larger prior means for $\sigma_e^2$ in cases where $\sigma_p^2$ is large.

**Comparisons under $M_2(4)$, $M_2(5)$, $M_2(6)$**

Fig. 5.9 and Fig. 5.10 show that among non-Bayesian estimators, the conclusions are the same as for models $M_2(1)$, $M_2(2)$, $M_2(3)$. Overall for $n = 20$ Bayes estimators B2 and B4 perform well. For $n=50$, B2 and B4 estimators perform equally well.
Figure 5.5: Comparison of Estimators for Model M1 $n = 20$
Figure 5.6: Comparison of Estimators for Model M1 \( n = 50 \)
Figure 5.7: Comparison of Estimators for Model M2 \( n = 20 \)
Figure 5.8: Comparison of Estimators for Model M2 $n = 50$
Figure 5.9: Comparison of Estimators for Model M2 $n = 20$
Figure 5.10: Comparison of Estimators for Model M2 $n = 50$
Comparisons under $M_2(7)$, $M_2(8)$, $M_2(9)$

Fig. 5.11 and Fig. 5.12 show that among non-Bayesian estimators, ML3 does best for $n=20$ but ML2 has the highest value of MSE. However, for $n=50$, estimator MR outperforms ML3. Overall, Bayes estimators B2 and B4c.1 give the lowest values of MSE.

Comparisons under $M_2(10)$, $M_2(11)$, $M_2(12)$

Fig. 5.13 and Fig. 5.14 show that ML3 and ML5 perform best among non-Bayesian estimators while MR gives the highest value of MSE. Among Bayes estimators, the B2 estimators do well, but overall the B4 estimators outperform the others. The picture does not change much for $n=50$, with B4c.1 performing the best overall.

Comparisons under $M_2(13)$, $M_2(14)$, $M_2(15)$

Fig. 5.15 and Fig. 5.16 show that among non-Bayesian estimators ML2 performs slightly better than MR except for model $M_2(13)$. Overall the estimators B2 have the lowest MSE for both $n=20$ and $n=50$. Estimators ML5, B5 and B4 give large values of MSE.

Comparisons under $M_3(1)$, $M_3(2)$, $M_3(3)$

Fig. 5.17 and Fig. 5.18 show that (as would be expected) ML3 and B3.2 perform the best. However, MR does better than the others. For $M_3(3)$, i.e. $h=1.0$, the Bayes estimators B2c.2 and B2c.3 perform equally well for both values of $n$ but these estimators did poorly for $M_3(1)$ and $M_3(2)$. Both ML5 and the B5 estimators give large values of MSE. Overall the method B2 also works well.
Figure 5.11: Comparison of Estimators for Model M2 \( n = 20 \)
Figure 5.12: Comparison of Estimators for Model M2 $n = 50$
Figure 5.13: Comparison of Estimators for Model M2 $n = 20$
Figure 5.14: Comparison of Estimators for Model $M_2$ $n = 50$
Figure 5.15: Comparison of Estimators for Model M2 $n = 20$
Figure 5.16: Comparison of Estimators for Model M2 $n = 50$
Figure 5.17: Comparison of Estimators for Model M3 $n = 20$
Figure 5.18: Comparison of Estimators for Model M3 $n = 50$
Comparisons under $M_4(1), M_4(2), M_4(3)$

Fig. 5.19 and Fig. 5.20 show that ML3 and ML5 did well among non-Bayesian estimators. The B2 estimators perform well for $n=20$, but poorly for $n=50$. ML2 gives the highest value of MSE. However, the B4 estimators give the lowest values of MSE, as expected.

Comparisons under $M_5$

Fig. 5.21 and Fig. 5.22 show that both ML3 and ML5 perform equally well for this IID model. ML2 does poorly. Overall B4a.1 (for $n=20$), B3.2 and B4c.1 (for $n=50$) outperform the others. For $n=50$, the estimators B2 give high values of MSE.

Analysis of $(\overline{M}_n/d_2)$ as an Estimator of Standard Deviation

Fig. 5.23 shows that the mean moving range estimator, $(\overline{M}_n/d_2)$ performs best in models $M_{4}(1), M_{3}(1), M_{3}(2)$ and $M_5$. In general, it did well for $M_2$ models with small values of $\sigma^2_{\nu}$. However, it is not a good estimator when $\rho$ is negative for $M_2$, or in the cases of $M_{1}(1), M_{1}(3), M_{1}(4)$ or $M_4$ with large $\sigma^2_{\mu}$. As $n$ increases, $MSE$ of $(\overline{M}_n/d_2)$ gets smaller.

Comparison of MR and ML2

We have observed that estimators MR and ML2 seem to be most attractive among the non-Bayesian estimators. Fig. 5.24 and Fig. 5.25 gives a clear comparison of these two estimators.

For $n = 20$, MR and ML2 perform equally well for $M_{1}(1), M_{1}(2)$ and $M_{1}(3)$, but MR does poorly for the $M_{1}(4)$ and $M_{1}(5)$ models. However, MR has the smaller
Figure 5.19: Comparison of Estimators for Model M4 $n = 20$
Figure 5.20: Comparison of Estimators for Model M4 $n = 50$
Figure 5.21: Comparison of Estimators for Model M5 $n = 20$
Figure 5.22: Comparison of Estimators for Model M5 \( n = 50 \)
Figure 5.23: $\overline{M_n}/d_2$ as an Estimator of Standard Deviation
Figure 5.24: Comparison of MR and ML2 Estimators $n = 20$
Figure 5.25: Comparison of MR and ML2 Estimators $n = 50$
value of $MSE$ for linear drift models $M_3(1)$, $M_3(2)$ and $M_3(3)$. MR performs better than ML2 for most of the M2 models. For $n = 50$: The situation does not change much except that ML2 does better than MR for M2 models with $\rho = 1$.

Some of the practical problems which we came across while using ML2 estimator are:

1. Sometimes the estimates may be negative.

2. Since it is an iterative procedure, ML2 estimation takes more computer time as compared to MR estimator.

So due to its parsimonious properties, we might consider MR to be preferable to ML2.
CHAPTER 6. SUMMARY AND RECOMMENDATIONS

Summary

The purpose of this study has been to explore the estimation of variance in repeated sampling of size one from different populations ordered in time, when the population means are potentially changing. We hoped to find an estimator robust to unknown changes in population means. We considered five models, covering a variety of circumstances. Method of moments, maximum likelihood and Bayesian techniques have used to derive estimators.

In Chapter 2, we have discussed properties of a moving range and the mean moving range as estimators of standard deviation. The mean moving range is the basis of a standard approach in SQC for estimating standard deviation. With unequal population means, the mean moving range estimator, $\overline{M}/d_2$, systematically overestimates the standard deviation. The extent of overestimation depends on the magnitude of the difference between population means. However, this overestimation is only 5% if the absolute difference between population means is half the value of standard deviation and becomes 24% when this difference equals the standard deviation. We have studied some large sample properties of the mean moving range (or $\overline{M}/d_2$). The almost sure convergence of $\overline{M} - E\overline{M} \rightarrow 0$ has been shown for any fixed sequence of means $\{\mu_j\}$. This is a kind of almost sure consistency of $\overline{M}/d_2$. 
for its expectation under a model with nonrandom means (and consequently for the
cases of constant means and means with linear drift). A central limit theorem for
$\overline{M}_n$ has been given and its assumptions have been verified for most of the models
considered.

Maximum likelihood is a commonly used estimation technique. In case the pop­
ulation means follow an AR(1) model, an iterative procedure which gives estimators
asymptotically equivalent to maximum likelihood estimators has been used. The

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

(n-divisor) sample variance, $\frac{n}{n-1}$, is a common estimator of population
variance and the maximum likelihood estimator under model M5. Some properties
of ML estimators under alternative models have also been discussed. We have seen
that when population means follow an AR(1) model, the n-divisor sample variance
is badly biased (upward) as an estimator of variance. The same conclusion holds for
the case when there is linear drift in population means.

Bayesian estimation procedures can be employed if one has prior beliefs that can
be quantified in terms of a distribution. The Inverse gamma distribution has been
widely used as prior distribution for variances. In cases where calculation of posterior
densities is not possible in closed form, the Gibbs sampling algorithm has been used
to approximate posterior means. The application of the Gibbs sampler requires full
conditional distributions of the parameters to be estimated. It has nice asymptotic
properties. As discussed in Chapter 5, Bayes estimators for the random walk model
perform best overall.
Recommendations

Since practitioners often encounter data with little or no knowledge about possible variability in the population means in repeated sampling, choice of a suitable estimator is a difficult task. If we a good model for the population means, a suitable estimator can be derived for any estimation method. A real issue is to find some method robust to misspecification of the model for population means. We observed that the Bayes estimator for the random walk with a large value of the variance of the shocks in the random walk performed best across the variety of circumstances. Among non-Bayesian estimators, the estimator MR (based on the mean moving range) and the likelihood estimator under random walk model performed very well. But the mean moving range estimator MR may be preferred over other estimators if one is interested in non-Bayesian methods.
REFERENCES


