Abstract

STEIGER, NATALIE MILLER. Improved Batching for Confidence Interval Construction in Steady-State Simulation. (Under the direction of James R. Wilson.)

The primary objectives of this research are formulation and evaluation of an improved batch-means procedure for steady-state simulation output analysis. The new procedure yields a confidence interval for a steady-state expected response that is centered on the sample mean of a portion of the series of simulation-generated responses and satisfies a user-specified absolute or relative precision requirement. We concentrate on the method of nonoverlapping batch means (NOBM), which requires the sample means computed from adjacent batches of observations to be independent and identically distributed normal random variables. For increasing batch sizes and a fixed number of batches computed from a weakly dependent (ϕ -mixing) output process, we establish key asymptotic distributional properties of the vector of batch means and of the numerator and squared denominator of the NOBM t -ratio. These results lead to the Automated Simulation Aalysis Procedure (ASAP), which operates as follows: the batch size is progressively increased until either (a) the batch means pass the von Neumann test for independence, and then ASAP delivers a classical NOBM confidence interval; or (b) the batch means pass the Shapiro-Wilk test for multivariate normality, and then ASAP delivers a corrected confidence interval. The latter correction is based on an inverted Cornish-Fisher expansion for the classical NOBM t -ratio, where the terms of the expansion are estimated via an autoregressive-moving average time series model of the batch means. An extensive experimental performance evaluation demonstrates the advantages of ASAP versus other widely used batch-means procedures.

IMPROVED BATCHING FOR CONFIDENCE INTERVAL CONSTRUCTION IN STEADY-STATE SIMULATION

by

Natalie Miller Steiger

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

INDUSTRIAL ENGINEERING

Raleigh

1999

APPROVED BY:

_____

_____

_____

_____ Chair of Advisory Committee

Biography

Natalie Miller Steiger was born on September 18, 1950 in Vicksburg, MS. She graduated from St. Francis Xavier Academy in 1968. She was a member of the University Honors Program at the University of Southern Mississippi in Hattiesburg, MS where she earned B.S. (1972) and M.S. (1973) degrees in Mathematics.

In 1973, she was employed as a Junior Engineer for South Central Bell in Jackson, MS. In 1975, she began work for Cities Service Oil and Gas Company in Tulsa, OK. She held various positions as analyst and manager while at Cities Service and remained with the company through its acquisition by Occidental Petroleum Company. In 1979, she married David Steiger. She resigned from Occidental Petroleum Company in 1986 to be a full-time mother to Rebecca, who was born in November, 1985. Her son, Benjamin, was born in January, 1988.

In 1993, Natalie and her family moved to Burlington, NC and she entered graduate school at North Carolina State University in Raleigh. After completing her degree in Industrial Engineering, she plans to pursue a career in teaching and research at the university level.

Acknowledgments

I thank Dr. James R. Wilson, Chair of my advisory committee, for the considerable effort he expended and the extreme patience he displayed in guiding me through this research and the composition of this dissertation. His dedication to doing the job well is truly inspiring. I thank Drs. Steve Roberts, Yahya Fathi, and Len Stefanski for serving as advisors on my committee and for their constructive input. I also thank Dr. Jack Silverstein for serving as the graduate representative on my committee. I thank Dr. Jeff Joines for his assistance with programming and computer problems.

I also extend thanks to the many people who supported me in this effort. John and Louise Steiger gave me encouragement when it was most needed. Debbie Edelman kept me grounded in reality. Mandy Frissell, Gayle Canaday, and Ann Frederiksen filled in as mommy when needed and listened to hours of my complaining. My children, Rebecca and Benjamin, endured my bad moods and absences with grace and understanding far beyond their years. But most importantly David, my husband, did all of the above and did it unceasingly.

Contents

List of Tables	vi
List of Figures	viii
1 Introduction	1
1.1 Problem of Confidence Interval Construction in Steady-state Simulation	1
1.2 Analysis of Steady-State Simulation Outputs Using the Method of Nonoverlapping Batch Means (NOBM)	3
1.3 Scope and Objectives of Research	7
1.4 Organization of the Dissertation	9
2 Literature Review	10
2.1 Methods of Confidence Interval Construction	10
2.1.1 Replication/Deletion	10
2.1.2 Spectral Analysis	12
2.1.3 Overlapping Batch Means	14
2.1.4 Regenerative Method	14
2.1.5 Autoregressive Method	17
2.1.6 Standardized Time Series	18
2.2 Fixed Sample Size Approaches to NOBM	21
2.2.1 Classical NOBM	21
2.2.2 Nonclassical NOBM	23
2.3 Sequential Approaches to Batch Means	26
2.3.1 Law and Carson's Procedure	26
2.3.2 Fixed Number of Batches (FNB)	28
2.3.3 Square Root Rule (SQRT)	29
2.3.4 LBatch and ABatch Procedures	29
2.3.5 Two-stage Stopping Procedure for Batch Means	30
3 Convergence Properties of the Batch Means Method	35
3.1 Asymptotic Joint Distribution of Batch Means	35
3.2 Moment Analysis of Components of Batch Means t -Ratio	40
3.3 Behavior of Batch Means Method in Selected Cases	45

3.3.1	Discrete Time Markov Chains	46
3.3.2	M/M/1 Queue Waiting Times	63
3.3.3	Autoregressive Process	68
3.3.4	Exponential Autoregressive Process	72
3.4	Monte Carlo Analysis of Components of Batch Means t -Ratio for Selected Cases	77
3.5	Conclusions	91
4	Formulation of ASAP: Automated Simulation Analysis Procedure	92
4.1	An Alternate Approach to Handling Dependent Batch Means	92
4.2	Overview of ASAP	93
4.3	Testing Batch Means for Independence and Joint Normality	95
4.4	Construction of Time Series Models for Dependent Normal Batch Means	97
4.5	Inverted Cornish-Fisher Correction for Dependence of Normal Batch Means	101
4.6	Fulfilling the Precision Requirement	105
4.7	Advantages of ASAP	106
4.8	Formal Statement of ASAP	107
5	Performance Evaluation of ASAP	111
5.1	Suite of Test Problems	111
5.1.1	Discrete Time Markov Chains	111
5.1.2	Autoregressive and Exponential Autoregressive Processes . . .	112
5.1.3	Queueing Systems	112
5.1.4	Computer Models	113
5.1.5	An Inventory System	114
5.2	Summary of Experimental Results	115
5.3	Discussion of Results	139
5.3.1	Discrete Time Markov Chains	139
5.3.2	Autoregressive and Exponential Autoregressive Processes . . .	140
5.3.3	Queueing Systems	140
5.3.4	Computer Models	142
5.3.5	Inventory Model	143
6	Conclusions and Recommendations	144
6.1	Main Conclusions of the Research	144
6.2	Recommendations for Future Research	147
	Bibliography	149
	A Derivation of Mixed Moments of the Batch Means	153

List of Tables

5.1	Steady-State Expected Waiting Time in Selected Queueing Systems .	113
5.2	Parameters for the Selected Central Server Models	115
5.3	Performance of Batch-Means Procedures for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) Based on 100 Independent Replications of Nominal 90% Confidence Intervals	117
5.4	Performance of Batch-Means Procedures for the 2-State DTMC with High Positive Correlation Structure as Defined by (3.66) and (3.67) Based on 100 Independent Replications of Nominal 90% Confidence Intervals	118
5.5	Performance of Batch-Means Procedures for the 8-State DTMC as Defined by (3.68) and (3.69) Based on 100 Independent Replications of Nominal 90% Confidence Intervals	119
5.6	Performance of Batch-Means Procedures for the AR(1) Process (3.75) with $\varphi = 0.9$ and $\mu_X = 2.0$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	120
5.7	Performance of Batch-Means Procedures for the EAR(1) Process (3.78) with $\varphi = 0.9$ and $\mu_X = 2.0$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	121
5.8	Performance of Batch-Means Procedures for the M/M/1 Queue Waiting Time Process with $\tau = 0.9$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	122
5.9	Performance of Batch-Means Procedures for the M/M/1 Queue Waiting Time Process with $\tau = 0.5$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	123
5.10	Performance of Batch-Means Procedures for the M/M/1 Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	124
5.11	Performance of Batch-Means Procedures for the M/M/1 LIFO Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	125
5.12	Performance of Batch-Means Procedures for the M/M/1 SIRO Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	126

5.13	Performance of Batch-Means Procedures for the $E_4/M/1$ Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	127
5.14	Performance of Batch-Means Procedures for the $M/H_2/1$ Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	128
5.15	Performance of Batch-Means Procedures for the $M/M/2$ Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	129
5.16	Performance of Batch-Means Procedures for the $M/M/1/M/1$ Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	130
5.17	Performance of Batch-Means Procedures for the Time-Shared Model Based on 100 Independent Replications of Nominal 90% Confidence Intervals	131
5.18	Performance of Batch-Means Procedures for the Central Server Model 1 Based on 100 Independent Replications of Nominal 90% Confidence Intervals	132
5.19	Performance of Batch-Means Procedures for the Central Server Model 2 Based on 100 Independent Replications of Nominal 90% Confidence Intervals	133
5.20	Performance of Batch-Means Procedures for the Central Model 3 Based on 100 Independent Replications of Nominal 90% Confidence Intervals	134
5.21	Performance of Batch-Means Procedures for the Central Server Model 4 Based on 100 Independent Replications of Nominal 90% Confidence Intervals	135
5.22	Performance of Batch-Means Procedures for the (s, S) Inventory System Based on 100 Independent Replications of Nominal 90% Confidence Intervals	136
5.23	Performance of LBATCH and ABATCH Algorithms under a Relative Precision Requirement for 2-state DTMC with High Correlation as Defined by (3.66) and (3.67) Based on 100 Independent Replications of Nominal 90% Confidence Intervals	137
5.24	Performance of LBATCH and ABATCH Algorithms under a Relative Precision Requirement for $M/M/1$ Queue $\tau = 0.9$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals	138

List of Figures

2.1	Fishman's Fixed Sample Size Algorithm for Determining Batch Size m and Number of Batches k	24
2.2	Law and Carson's Sequential Algorithm for Determining Batch Size m and Number of Batches k	28
2.3	LBATCH Algorithm for Determining Batch Size m and Number of Batches k	31
2.4	ABATCH Algorithm for Determining Batch Size m and Number of Batches k	32
2.5	Nakayama's Absolute Half-length Two-stage Stopping Procedure . . .	33
2.6	Nakayama's Relative Half-length Two-stage Stopping Procedure . . .	34
3.1	Lag- q Correlation $\text{Corr}(N_i, N_{i+q})$ of Observations of Cost Function $N_i = \eta(X_i)$ on 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66). The SSVC $\sigma^2 = 2.083$ and $\sigma_N^2 \equiv \text{Var}[N_i] = 6.25$	51
3.2	Lag 1 Correlation of the Batch Means with Batches of Size m for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66)	51
3.3	$\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for the Cost Function on 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66). The SSVC $\sigma^2 = 2.083$	52
3.4	Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m	52
3.5	Condition (3.35) on Ratio $\text{E}[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m	53
3.6	Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m	53
3.7	Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]/\text{E}[S_{n,k}^2]}$ for 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m	54

3.8	Lag- q Correlation $\text{Corr}(N_i, N_{i+q})$ of Observations of Cost Function $N_i = \eta(X_i)$ on 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67). The SSVC $\sigma^2 = 618.75$ and $\sigma_N^2 \equiv \text{Var}[N_i] = 6.25$	55
3.9	Lag 1 Correlation of the Batch Means with Batches of Size m for the 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67).	55
3.10	$\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for the Cost Function on 2 State DTMC with High Positive Correlation Structure as Defined by (3.66) and (3.67). The SSVC $\sigma^2 = 618.75$	56
3.11	Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for the 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m	56
3.12	Condition (3.35) on Ratio $\text{E}[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m	57
3.13	Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m	57
3.14	Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]/\text{E}[S_{n,k}^2]}$ for 2-State DTMC with High Positive Correlation Structure as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m	58
3.15	Lag- q Correlation $\text{Corr}(N_i, N_{i+q})$ of Observations of Cost Function $N_i = \eta(X_i)$ on 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69). The SSVC $\sigma^2 = 77.5$ and $\sigma_N^2 \equiv \text{Var}[N_i] = 132.2$	59
3.16	Lag 1 Correlation of the Batch Means with Batches of Size m for the 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69).	59
3.17	$\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for the Cost Function on 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69). The SSVC $\sigma^2 = 77.5$	60
3.18	Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for the 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m	60
3.19	Condition (3.35) on Ratio $\text{E}[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m	61
3.20	Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m	61

3.21	Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/E[S_{n,k}^2]$ for 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m	62
3.22	Lag- q Correlation $\text{Corr}(X_i, X_{i+q})$ of Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$. The SSVC $\sigma^2 = 37440$ and $\sigma_X^2 = 99$	64
3.23	Lag 1 Correlation of the Batch Means with Batches of Size m for the Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$	64
3.24	$\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n of Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$. The SSVC $\sigma^2 = 37440$	65
3.25	Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m	65
3.26	Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m	66
3.27	Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the M/M/1 Queue Wait Time Process where Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m	66
3.28	Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/E[S_{n,k}^2]$ for the M/M/1 Queue Wait Time Process where Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m	67
3.29	Lag- q Correlation $\text{Corr}(X_i, X_{i+q})$ of Observations of an AR(1) Process (3.75) with $\varphi = 0.9$. The SSVC $\sigma^2 = 100$ and $\sigma_X^2 = 5.26$	68
3.30	Lag 1 Correlation of the Batch Means with Batches of Size m for an AR(1) Process (3.75) with $\varphi = 0.9$	69
3.31	$\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for an AR(1) Process (3.75) with $\varphi = 0.9$. The SSVC $\sigma^2 = 100$	69
3.32	Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ and Batches Size m	70
3.33	Condition (3.35) Ratio $E[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size m	70
3.34	Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size m	71
3.35	Condition (3.37) on Coefficient of Variation, $\sqrt{\text{Var}[S_{n,k}^2]}/E[S_{n,k}^2]$, for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size m	71
3.36	Lag- q Correlation $\text{Corr}(X_i, X_{i+q})$ of Observations of an EAR(1) Process (3.78) with $\varphi = 0.9$. The SSVC $\sigma^2 = 76$ and $\sigma_X^2 = 4.0$	73
3.37	Lag 1 Correlation of the Batch Means with Batches of Size m for an EAR(1) Process (3.78) with $\varphi = 0.9$	73
3.38	$\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for an EAR(1) Process (3.78) with $\varphi = 0.9$. The SSVC $\sigma^2 = 76$	74
3.39	Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m	74

3.40	Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\overline{X}(m)]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m	75
3.41	Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\overline{X}(m)]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m	75
3.42	Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/E[S_{n,k}^2]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m . . .	76
3.43	Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the Alternating Correlation 2-State DTMC Defined by (3.65) and (3.66) for $k = 30$ Batches of Size $m = 8, m = 16$ and $m = 32$	79
3.44	Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the Alternating Correlation 2-State DTMC Defined by (3.65) and (3.66) for $k = 30$ Batches of Size $m = 8, m = 16$ and $m = 32$	80
3.45	Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the High Positive Correlation 2-State DTMC Defined by (3.66) and (3.67) for $k = 30$ Batches of Size $m = 16, m = 256$ and $m = 512$	81
3.46	Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the High Positive Correlation 2-State DTMC Defined by (3.66) and (3.67) for $k = 30$ Batches of Size $m = 16, m = 256$ and $m = 512$	82
3.47	Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the Damped Sinusoidal Correlation 8-State DTMC Defined by (3.68) and (3.69) for $k = 30$ Batches of Size $m = 16, m = 32$ and $m = 64$	83
3.48	Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the Damped Sinusoidal Correlation 8-State DTMC Defined by (3.68) and (3.69) for $k = 30$ Batches of Size $m = 16, m = 32$ and $m = 64$ . . .	84
3.49	Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the M/M/1 Queue Waiting Times with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size $m = 16, m = 2048$ and $m = 4096$	85
3.50	Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the M/M/1 Queue Waiting Times with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size $m = 16, m = 2048$ and $m = 4096$	86
3.51	Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16, m = 32$ and $m = 64$	87

3.52	Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$	88
3.53	Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$	89
3.54	Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$	90
4.1	ASAP Algorithm.	108

Chapter 1

Introduction

1.1 Problem of Confidence Interval Construction in Steady-state Simulation

In discrete-event simulation, we are often interested in estimating the steady-state mean μ_X of a stochastic output process $\{X_i : i \geq 1\}$ generated by a single, though long, simulation run. Assuming the target process is stationary and given a time series of length n from this process, we see that a natural estimator of μ_X is the sample mean, given by

$$\overline{X}(n) = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.1)$$

We also require some indication of this estimator's precision; and typically a confidence interval (CI) for μ_X is constructed at a certain confidence level $1 - \alpha$, where $0 < \alpha < 1$.

If the X_i 's are independent and identically distributed (i.i.d.) random variables, then the Central Limit Theorem (CLT) from classical statistics states that

$$\sqrt{n} [\overline{X}(n) - \mu_X] \xrightarrow[n \rightarrow \infty]{D} N(0, \sigma_X^2), \quad (1.2)$$

where: σ_X^2 is the variance of the X_i 's, i.e., $\sigma_X^2 = \text{E}[(X_1 - \mu_X)^2]$, assuming this quantity is positive and finite; the symbol $\xrightarrow[n \rightarrow \infty]{D}$ denotes convergence in distribution; and $N(0, \sigma_X^2)$ denotes a normal distribution with mean 0 and variance σ_X^2 . If σ_X^2 is

known, then an asymptotically valid $100(1 - \alpha)\%$ CI for μ_X is

$$\overline{X}(n) \pm z_{1-\alpha/2} \frac{\sigma_X}{\sqrt{n}}, \quad (1.3)$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of a $N(0, 1)$ distribution. This means that

$$\lim_{n \rightarrow \infty} \Pr \left\{ \mu \in \overline{X}(n) \pm z_{1-\alpha/2} \frac{\sigma_X}{\sqrt{n}} \right\} = 1 - \alpha. \quad (1.4)$$

If we construct a CI for μ_X on each of several prolonged runs of the simulation, then the long-run proportion of these CIs which include (cover) the value μ_X is the *actual coverage probability*. Normally, we would like the CI for μ_X to satisfy two criteria: (a) the CI is narrow enough to be informative, and (b) the actual coverage probability of the CI is close to the nominal coverage probability $1 - \alpha$. The CI (1.3) is based on a sample whose size n is fixed beforehand.

Another way of constructing a CI for μ_X is to continue sampling until some precision criterion is met. A precision requirement involving an *absolute half-length*, δ_a , specifies that the actual half-length of the CI given in (1.3) should be no larger than the threshold value δ_a , i.e.,

$$z_{1-\alpha/2} \frac{\sigma_X}{\sqrt{n}} \leq \delta_a. \quad (1.5)$$

A specification of a *relative half-length*, δ_r , implies that the CI half-length should be no larger than the fraction δ_r of the magnitude of the final sample mean (e.g., $\pm 10\%$) so that

$$z_{1-\alpha/2} \frac{\sigma_X}{\sqrt{n}} \leq \delta_r |\overline{X}(n)|, \quad (1.6)$$

assuming that $\overline{X}(n) \neq 0$.

If the process variance σ_X^2 is unknown, then a consistent estimator of σ_X^2 can be used in its place to construct a confidence interval for μ_X . Typically the sample variance, given by

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}(n))^2, \quad (1.7)$$

is used to estimate σ_X^2 ; and substituting S_n for σ_X in (1.3), (1.5), and (1.6) yields an asymptotically valid $100(1 - \alpha)\%$ confidence interval for μ_X as $\delta_a \rightarrow 0$ (Chow and Robbins 1965) or as $\delta_r \rightarrow 0$ (Nádas 1969)—provided the X_i 's are i.i.d.

Unfortunately, observations of a simulation-generated output process are typically neither independent nor identically distributed. Therefore, the usual method of CI construction from classical statistics described above is not directly applicable. Several methods have been proposed for constructing CIs based on dependent observations. Nonoverlapping batch means (NOBM) is one such method. Brief descriptions of other methods are given in Chapter 2.

1.2 Analysis of Steady-State Simulation Outputs Using the Method of Nonoverlapping Batch Means (NOBM)

In the NOBM method, the sequence of simulation-generated outputs $\{X_i : i = 1, \dots, n\}$ is divided into k adjacent nonoverlapping batches, each of size m_n . (In later chapters we will wish to emphasize m 's dependence upon n ; therefore we use the notation m_n .) For simplicity, we assume that n is a multiple of m_n so that $n = km_n$; thus when k is fixed and $m_n \rightarrow \infty$, we have $n \rightarrow \infty$. The sample mean, $Y_j(m_n)$, for the j th batch is calculated by

$$Y_j(m_n) = \frac{1}{m_n} \sum_{i=m_n(j-1)+1}^{m_n j} X_i \quad \text{for } j = 1, \dots, k. \quad (1.8)$$

Then the grand mean $\bar{Y}(n, k)$ of the individual batch means, given by

$$\bar{Y}(n, k) = \frac{1}{k} \sum_{j=1}^k Y_j(m_n), \quad (1.9)$$

is used as an estimator for μ_X (note that $\bar{Y}(n, k) = \bar{X}(n)$). Naturally, we seek to construct a CI centered on the estimator (1.9).

Irrespective of the method employed in estimating the steady-state mean of a simulation output process, one of the key issues that must be addressed is how to obtain observations of the process that are representative of the steady-state (equilibrium) distribution of the process. Starting the simulation in steady-state operation is typically infeasible: if the experimenter has enough information to sample the simulation's initial conditions from the relevant steady-state distribution, then there is probably no need to perform the simulation in the first place; the mean μ_X of the selected

output process could be readily determined from this information alone. Given that the simulation starts according to an initial distribution other than the steady-state distribution (e.g., the empty-and-idle state for a queueing system), some of the early observations of the selected output process will not be representative of the process in steady state. This is known as the *initial transient* or *initial bias* problem, since a sample mean that includes these observations will likely be biased.

To overcome the initial transient problem, we start collecting statistics after an initial “warm-up” period. The number of observations that should be eliminated prior to the start of data collection is an area of research in itself. See, for example, Law and Kelton (1991) for a discussion. For purposes of this research, we plan to formulate a way to handle this initial bias problem in the context of the NOBM method for simulation output analysis. However, our primary focus will be on overcoming the problem of correlation between batch means and on the rate of convergence of the batch means to joint normality, since these issues affect the actual coverage of the confidence interval delivered by a batch-means procedure.

To continue the introductory discussion of NOBM methods for steady-state simulation output analysis, we must introduce some basic nomenclature. For simplicity, we will assume the selected output process $\{X_j\}$ is *stationary* (or stationary in the strict sense), i.e., the joint distribution of the X_j ’s is insensitive to time shifts. If the process $\{X_j : j = 1, 2, \dots\}$ is stationary, then for all positive integers $r, p, j_1, j_2, \dots, j_r$ and for all real cutoff values $x_1, x_2, \dots, x_r$, we have

$$\Pr[X_{j_1} \leq x_1; X_{j_2} \leq x_2; \dots; X_{j_r} \leq x_r] = \Pr[X_{p+j_1} \leq x_1; X_{p+j_2} \leq x_2; \dots; X_{p+j_r} \leq x_r].$$

A stochastic process $\{X_j\}$ is said to be *weakly stationary* (or stationary in the wide sense) if: (a) $E[X_j] = \mu_X$ for all j ; and (b) the covariance between two observations X_p and X_{p+q} , separated by q successive observations depends only on the lag q and not on p ,

$$\gamma(q) \equiv E[(X_p - \mu_X)(X_{p+q} - \mu_X)], \quad q = 0, \pm 1, \pm 2, \dots \quad (1.10)$$

Stationary processes with a finite marginal variance are, therefore, weakly stationary.

Processes that have a steady-state distribution are often *weakly dependent*, in the sense that X_j ’s widely separated from each other in the sequence are almost independent (in the sense of *ϕ -mixing*, which will be completely defined in Chapter

2) so that $\gamma(q) \rightarrow 0$ as q increases. These weakly dependent processes typically obey a CLT for dependent processes of the form

$$\sqrt{n} [\bar{X}(n) - \mu_X] \xrightarrow[n \rightarrow \infty]{D} N(0, \sigma^2), \quad (1.11)$$

where

$$\sigma^2 \equiv \lim_{n \rightarrow \infty} n \text{Var}[\bar{X}(n)] = \sum_{i=-\infty}^{\infty} \gamma(i) = \gamma(0) + 2 \sum_{i=1}^{\infty} \gamma(i) \quad (1.12)$$

is the steady-state variance constant (SSVC) (as distinguished from the process variance σ_X^2). A sufficient condition for the SSVC to exist is that $\sum_{i=-\infty}^{\infty} |\gamma(i)| < \infty$; see, for example, p. 459 Anderson (1971). Note that $\gamma(0) = \text{Var}[X_i]$.

Although some output analysis methods attempt to estimate this steady-state variance constant σ^2 for the construction of the CI (see Chapter 2), NOBM in its classical setting, i.e., when the number of batches is fixed, does not. NOBM seeks to make each batch a “repetition” of the experiment on the process. In order to achieve this, we assume that the batch size is sufficiently large so that the batch means $\{Y_j(m_n) : 1 \leq j \leq k\}$ are i.i.d. normal,

$$\{Y_j(m_n) : 1 \leq j \leq k\} \stackrel{\text{i.i.d.}}{\sim} N[\mu_X, \sigma^2(m_n)/m_n], \quad (1.13)$$

where the symbol $\sim$ is read “is distributed as,”

$$\sigma^2(m_n) = \gamma(0) + 2 \sum_{q=1}^{m_n-1} \left(1 - \frac{q}{m_n}\right) \gamma(q), \quad (1.14)$$

and $\text{Var}[Y_j(m_n)] = \sigma^2(m_n)/m_n$. It follows that $\lim_{m_n \rightarrow \infty} \sigma^2(m_n) = \sigma^2$ and

$$\text{Var}[Y_j(m_n)] \approx \sigma^2/m_n, \quad (1.15)$$

provided that m_n is sufficiently large.

We can now apply a classical result from statistics to compute a confidence interval for μ_X from the batch means $\{Y_j(m_n) : 1 \leq j \leq k\}$. If $\{Z_j : 1 \leq j \leq k\} \stackrel{\text{i.i.d.}}{\sim} N(\mu_Z, \sigma_Z^2)$ so that the $\{Z_i\}$ constitute a random sample of size k from a normal distribution with mean μ_Z and variance σ_Z^2 , then the sample mean $\bar{Z}(k)$ and the sample variance S_k^2 of the $\{Z_j\}$ are independent with

$$\bar{Z}(k) \sim N\left(\mu_Z, \frac{\sigma_Z^2}{k}\right); \quad (1.16)$$

$$\frac{(k-1)S_k^2}{\sigma_Z^2} \sim \chi_{k-1}^2; \quad (1.17)$$

and

$$\frac{\overline{Z}(k) - \mu}{\sqrt{S_k^2/k}} \sim t_{k-1}, \quad (1.18)$$

where t_{k-1} denotes the Student t -distribution with $k-1$ degrees of freedom and χ_{k-1}^2 denotes the chi-square distribution with $k-1$ degrees of freedom. We can then construct an exact $100(1-\alpha)\%$ CI for μ_Z of the form

$$\overline{Z}(k) \pm t_{1-\alpha/2, k-1} \frac{S_k}{\sqrt{k}}. \quad (1.19)$$

Replacing $\overline{Z}(k)$ by $\overline{X}(n)$ and S_k^2 by the sample variance of the batch means

$$S_{n,k}^2 = \frac{1}{k-1} \sum_{j=1}^k [Y_j(m_n) - \overline{X}(n)]^2 \quad (1.20)$$

in (1.16)–(1.18), we have that (1.16)–(1.18) are approximately satisfied as the batch size m_n becomes sufficiently large while the batch count k is fixed. This is because the batch means $Y_1(m_n), \dots, Y_k(m_n)$ become almost independent (since the process is weakly dependent) and almost normally distributed (from an appropriate CLT for dependent processes). Brillinger (1973) showed the validity of the batch means method. Thus as $m_n \rightarrow \infty$ with k fixed so that $n \rightarrow \infty$, an asymptotically valid $100(1-\alpha)\%$ confidence interval for μ_X is

$$\overline{X}(n) \pm t_{1-\alpha/2, k-1} \frac{S_{n,k}}{\sqrt{k}}. \quad (1.21)$$

The asymptotic validity of NOBM depends on both the assumption of approximate independence of the batch means and the assumption of the batch means being approximately normally distributed. The assumption that the $Y_i(m_n)$'s are i.i.d. normal random variables implies that $\overline{X}(n)$ and $S_{n,k}^2$ should be independent random variables.

If we are using a sampling technique that allows for varying the sample size, then our criterion for CI half-length can be specified using the t -statistic and the variance estimator $S_{n,k}^2$. The requirement for an absolute half-length δ_a becomes

$$t_{1-\alpha/2, k-1} \frac{S_{n,k}}{\sqrt{k}} \leq \delta_a; \quad (1.22)$$

and the requirement for a relative half-length expressed in terms of δ_r has the form

$$t_{1-\alpha/2, k-1} \frac{S_{n,k}}{\sqrt{k}} \leq \delta_r |\overline{X}(n)|. \quad (1.23)$$

Chapter 2 describes several procedures that have been proposed for implementing the NOBM method for CI estimation. These procedures address the problem of determining the batch size, m_n , and the number of batches, k , that are required to satisfy the assumptions of independence and normality. Theoretically, if these assumptions are satisfied, then we will get CIs whose actual coverage is close to the nominal coverage. However, because of the difficulty of the problem, research is still ongoing.

1.3 Scope and Objectives of Research

The primary objective of this research is to develop an easy-to-use, reliable batching method for constructing a CI for the expected response of a steady-state simulation model. This method, which we call Automated Simulation Analysis Procedure (ASAP), has a theoretical basis, delivers specified accuracy and performs well on the basis of CI coverage in a variety of problems.

To provide a foundation for ASAP, first we investigate the asymptotic distributional properties of the vector of batch means along with the numerator and denominator of the NOBM t -ratio. Our research includes processes of the type $\{f(X_n) : n \geq 1\}$, where $\{X_n : n \geq 1\}$ is a finite-state-space, Discrete-Time Markov Chain (DTMC) and $f(\cdot)$ is a measurable function (e.g., a cost function associated with the occupied state). This is a nontrivial stochastic process for which a large body of literature exists. We also study other processes for which analytic expressions are available for the mean, variance, and correlations at all lags.

To date, most NOBM procedures have focused on the question of dependence between batch means and have employed various techniques for selecting a batch size at which the batch means appear to be independent or uncorrelated. However, highly dependent systems may require very large batch sizes before this requirement is met. Additionally, the requirement for normality has been largely glossed over by these techniques under the assumption that if the batch size is large enough for the batch means to be approximately independent, then it is large enough for them to be approximately normally distributed. No procedure in the current literature specifically addresses this issue of normality and how it can possibly affect the properties of the resulting CIs. Our method will have a theoretical basis and will take into account the issue of normality as well as independence of the batch means.

Some of the existing NOBM methods require intervention and interpretation by the user. Our method will remove the need for analysis and judgment calls by the user by specifying a beginning sample size, directing the user as to the amount of additional data necessary to achieve the requested precision, and ultimately delivering a CI.

The objectives of this research may be summarized as follows.

1. Investigate the convergence of batch means to joint normality and the convergence of the numerator and squared denominator of the NOBM t -ratio to their respective limiting distributions as the batch size increases with a fixed number of batches.
2. Develop theory and procedures, based on these convergence properties, for determining the batch size and the number of batches required for an asymptotically valid CI centered on the sample mean as the user-specified upper bound on the CI half-length tends to zero.
3. Formulate ASAP based on the foregoing analysis. ASAP operates as follows: the batch size is progressively increased with a fixed batch count until either
  - a) the batch means pass the von Neumann test for independence (Fishman 1978a), and then ASAP delivers the classical NOBM confidence interval (1.21); or
  - b) the batch means pass the Shapiro-Wilk test for multivariate normality (Tew and Wilson 1992), and then ASAP delivers a corrected confidence interval.

In case b), the correction is based on an inverted Cornish-Fisher expansion for the classical NOBM t -ratio (Hall 1983), where the terms of the expansion are estimated by fitting an autoregressive-moving average time series model to the batch means process.

4. Conduct an extensive experimental performance evaluation of ASAP, comparing it to other popular batch-means procedures in a broad range of stochastic systems.
5. Develop easy-to-use software that will automatically determine a batching policy for a given set of observations or will indicate how much more sampling is necessary for a specified precision and confidence level.

1.4 Organization of the Dissertation

In Chapter 2 we briefly review alternative ways of constructing CIs for dependent processes. We include a survey of previously proposed and studied NOBM procedures. Chapter 3 details the theoretical and experimental basis for ASAP. For increasing batch sizes and a fixed number of batches computed from a weakly dependent (ϕ -mixing) process, we establish key asymptotic distributional properties of the vector of batch means and of the numerator and squared denominator of the NOBM t -ratio. Chapter 3 also includes a Monte Carlo analysis of the abovementioned components of the NOBM t -ratio for six selected stochastic processes. A complete description of ASAP is presented in Chapter 4. Chapter 5 summarizes the results of extensive testing of ASAP versus other NOBM methods. Chapter 6 contains the conclusions from our research and recommendations for future study.

Chapter 2

Literature Review

In this chapter we will first discuss alternatives to NOBM in CI construction around the steady-state mean of a stationary weakly dependent process. A summary of the existing literature on NOBM will complete the literature review.

2.1 Methods of Confidence Interval Construction

2.1.1 Replication/Deletion

When employing the method of replication we simulate k independent replicates, each of length n , of the process and obtain kn observations. Denote the i th observation within the j th replication (run) of the simulation model as $X_{j,i}$. Then

$$\overline{X}_j(n) = \sum_{i=1}^n X_{j,i} \quad (2.1)$$

is the sample mean of the j th replicate,

$$\overline{\overline{X}}(n) = \sum_{j=1}^k \overline{X}_j(n) \quad (2.2)$$

is an estimator of μ_X , and an asymptotically valid $(1 - \alpha)100\%$ CI is given by

$$\overline{\overline{X}}(n) \pm t_{1-\alpha/2, k-1} \frac{S_k(n)}{\sqrt{k}}, \quad (2.3)$$

where $S_k^2(n) = \frac{1}{k-1} \sum_{j=1}^k \left(\overline{X}_j(n) - \overline{\overline{X}}(n) \right)^2$. If n is sufficiently large, (1.11) holds and $\overline{X}_j(n) \sim N\left(\mu_x, \frac{\sigma^2}{n}\right)$. Also, the k sample means are independent, therefore the CI in (2.3) is approximately asymptotically valid.

The method of replication causes the initial bias problem to be repeated k times, as opposed to the single long-run methods which encounter the problem only once. If the k independent replicates are not started in steady-state and n is not large enough to dissipate the initial bias, then $\overline{\overline{X}}(n)$ will be a biased estimator of μ_X . In this case if we take k too large, then we run the risk of a narrow CI around a biased estimator of μ_X , which, therefore, may close in around the wrong value. Also n must be large enough for the normality approximation of the replicate means to be in effect.

In an effort to ameliorate the initial bias, Law and Kelton (1991), for example, make a case for deleting an initial portion of each replication before the sample average is calculated and its CI is constructed; hence the name replication/deletion. Here again, we make k independent replicates of length n , but we eliminate a predetermined number, ℓ , of observations from each replicate. The time up to ℓ is called the warm-up period. We include only the $n - \ell$ observations in the calculation of the j th replicate sample mean, i.e., $\overline{X}_j(n, \ell) = \frac{1}{n-\ell} \sum_{i=\ell+1}^n X_{j,i}$. Now $\overline{\overline{X}}(n, \ell) = \frac{1}{k} \sum_{j=1}^k \overline{X}_j(n, \ell)$ is a supposedly less biased estimator of μ_X and we have an asymptotically valid $100(1 - \alpha)\%$ CI generated by taking long enough replicates to achieve our desired coverage, i.e., as $n \rightarrow \infty$, $\Pr \left[\mu_X \in \overline{\overline{X}}(n, \ell) \pm t_{1-\alpha/2, k-1} \frac{S_k(n, \ell)}{\sqrt{k}} \right] = 1 - \alpha$.

Welch (1983) developed a graphical method for determining ℓ . Law and Kelton (1991) propose using an n “much larger than the warm-up period ℓ determined by Welch’s graphical method.” How large n should be is an open problem; Law and Kelton (1991), for example, make no mention of how large n should be to meet the requirement of approximate normality of the replicate means.

Welch’s graphical method involves taking averages “across replicates” of the k occurrences of the i th observation, i.e., $\overline{X}_i = \frac{1}{k} \sum_{j=1}^k X_{j,i}$. Moving averages, $\overline{X}_i(w)$, at various windows w , of these $\overline{X}_i$ ’s are plotted and ℓ is chosen to be greater than the i where they seem to converge (Welch 1983).

The appeal of the replication/deletion method is its simplicity; it is easy to understand and implement. It also has strong theoretical underpinning. From a computational viewpoint, however, this method seems inefficient. We are “throwing away” ℓ observations from each replicate, a total of ℓk observations. In addition, to achieve k truly independent observations of $\overline{X}_j(n)$ we should make the determination of ℓ from prior simulations of k' independent replicates each of length n' . A better use of the computational effort would appear to be in increasing the size of a single-run method. The initial bias problem would thus be encountered only once, instead of k

times.

2.1.2 Spectral Analysis

As mentioned in Chapter 1, some simulation output analysis methods, known as *consistent estimation* methods, seek to estimate σ^2 (SSVC) and then use this estimate in constructing a CI for μ_X around the sample mean $\bar{X}(n)$. It is theoretically based on the CLT for weakly dependent processes given in (1.11). Spectral analysis is one such method. We are assuming the discrete time stochastic process is stationary and is therefore weakly stationary.

We will assume that the series of covariances $\{\gamma(r) : -\infty < r < \infty\}$ exists for the stationary discrete-time stochastic process we are simulating. The Fourier transform of this sequence is given by

$$f(\lambda) = \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma(r) \cos(\lambda r); \quad (2.4)$$

where $\lambda \in [-\pi, \pi]$ is called the frequency. The function $f(\lambda)$ is the spectral density (see Anderson 1971). Obviously, from (1.12) it follows that $\sigma^2 = 2\pi f(0)$, so we would like an estimate $\hat{f}(0)$ of $f(0)$, based on the n observations $X_1, X_2, \dots, X_n$.

The natural estimator of the covariance $\gamma(r)$ is

$$\hat{\gamma}_n(r) = \frac{1}{n} \sum_{t=1}^{n-r} (X_t - \bar{X}(n)) (X_{t+r} - \bar{X}(n)). \quad (2.5)$$

Given n observations from the simulation, however, we can estimate at most $(n-1)$ of the $\gamma(r)$ terms in (2.4) with the corresponding $\hat{\gamma}_n(r)$, so our estimator of σ^2 would be

$$\hat{\sigma}_n^2 = 2\pi \hat{f}(0) = \sum_{r=-(n-1)}^{n-1} \hat{\gamma}_n(r). \quad (2.6)$$

Because of the large number of terms involved in the expression of the estimator $\hat{\sigma}_n^2$, this estimator's variance does not go down to 0 as $n \rightarrow \infty$. If we use a truncated estimator

$$\sum_{r=-m}^m \hat{\gamma}_n(r), \quad (2.7)$$

where $m \ll n$ ($\ll$ denotes much less than), then we can reduce the variance of our estimator. Although bias will be worse, the net effect will be a better estimator.

To achieve a consistent estimator in the mean-square sense, i.e., $\text{MSE}[\hat{f}(0)] \rightarrow 0$ as $n \rightarrow \infty$, we require $m \rightarrow \infty$ but $m/n \rightarrow 0$. The MSE of the estimator $\hat{\theta}$ of a parameter θ is given by:

$$\text{MSE}(\hat{\theta}) = \text{E} \left[(\hat{\theta} - \theta)^2 \right] = \text{Bias}^2(\hat{\theta}) + \text{Var}(\hat{\theta}). \quad (2.8)$$

See section 6.2.3 of Priestley (1981) for conditions under which this $\text{MSE}[\hat{f}(0)] \rightarrow 0$.

Since we are assuming the process is weakly dependent, the sample covariances at the shorter lags tell us more about the dependency structure of the sequence than those at the longer lags. We can improve the estimator even further, if we could somehow weight these more informative $\hat{\gamma}_n(r)$'s, by using an estimator such as

$$2\pi\hat{f}(0) = \sum_{r=-m}^m w_n(r) \hat{\gamma}_n(r), \quad (2.9)$$

where $w_n(r)$ is an even function, called the lag window function, that is greatest where r is close to 0 and decreases as r is farther from 0.

There are several weight functions that are frequently used in spectral density estimation. Two popular ones are Parzen and Tukey-Hanning (see, for example, Anderson 1971).

One weight function of particular interest is the modified Bartlett window function, $w_n(r) = 1 - |r|/m$, which yields the following as the truncated sample spectral density:

$$2\pi\hat{f}(0) = \sum_{r=-(m-1)}^{(m-1)} \left(1 - \frac{|r|}{m}\right) \hat{\gamma}_n(r), \quad (2.10)$$

for m such that $m \rightarrow \infty$ and $n/m \rightarrow \infty$ as $n \rightarrow \infty$. It turns out that the overlapping batch means (OBM) variance estimator, defined below, is almost equal algebraically to the estimator of σ^2 obtained through the spectral estimation method, using the modified Bartlett window as the weight function (see Section 2.1.3).

Since we now have a consistent estimator $\hat{\sigma}_n^2$ for the SSVC σ^2 , we may now apply the CLT in (1.11) to obtain the CI

$$\overline{X}(n) \pm z_{1-\alpha/2} \frac{\hat{\sigma}_n}{\sqrt{n}}. \quad (2.11)$$

If the sample covariances are used to estimate σ^2 , it is computationally laborious. However, Fast Fourier Transforms (FFT) are used in practice and are computationally

more efficient. Difficulties with the method lie in decisions that must be made on a weight function and on where to truncate (what value of m to use). The choice for m is the critical and complex issue (see Priestley 1981).

2.1.3 Overlapping Batch Means

The method of overlapping batch means (OBM), introduced by Meketon and Schmeiser (1984), divides the n observations into batches of size m , but each observation starts a new batch. The resulting number of batches is $(n - m + 1)$. The first batch contains observations $X_1, X_2, \dots, X_m$; the second batch, $X_2, X_3, \dots, X_{m+1}$, etc. Let

$$\bar{X}_j(m) = \frac{1}{m} \sum_{i=0}^{m-1} X_{i+j} \quad (2.12)$$

denote the sample mean of the j th batch and $\bar{X}(n)$ the usual estimator of μ_X . Obviously the $\bar{X}_j(m)$'s are not independent, but are correlated. Meketon and Schmeiser (1984) find that

$$\widehat{\text{Var}}[\bar{X}(n)] = \frac{m}{n} \sum_{j=1}^{n-m+1} \left(\bar{X}_j(m) - \bar{X}(n) \right)^2 / (n - 2m + 1), \quad (2.13)$$

the OBM estimator of $\text{Var}[\bar{X}(n)]$, is approximately equal to

$$\frac{1}{n - 2m + 1} \left[\hat{\gamma}_n(0) + 2 \sum_{j=1}^{m-1} \left(1 - \frac{j}{m} \right) \hat{\gamma}_n(j) \right] \quad (2.14)$$

which is approximately the estimator of $2\pi\hat{f}(0)$ derived via the spectral estimation method, using the modified Bartlett window. The approximate equality is the result of some end-effect terms in OBM (Damerdji 1991). So the OBM method of variance estimation is closely related to the spectral variance analysis (at frequency 0) method.

2.1.4 Regenerative Method

In order to apply the regenerative method of output analysis the process must be *regenerative*, i.e., there exists an increasing sequence of random times $T_1, T_2, \dots$, called *regeneration times*, at which the process “starts over,” from a probability standpoint, independently of the past. The following hold: $\Pr[T_1 < \infty] = 1$; $\Pr[0 < \tau_k < \infty] = 1$, for $\tau_k \equiv T_{k+1} - T_k$ and $k \geq 1$; the sequence $\{X_s : 0 \leq s < T_1\}$ is independent of

$\{X_s : s \geq T_1\}$; and the sequence $\{X_s : T_k \leq s < T_{k+1}\}$ is an independent probabilistic replica of the sequence $\{X_s : T_1 \leq s < T_2\}$.

The observations $\{X_i : T_k \leq i < T_{k+1}\}$ form the k th *regenerative cycle*. If $T_1 \geq 2$, then the process does not start in the regenerative state and is said to be *delayed*. An illustration of a regenerative process from queueing theory is the G/G/1 queue, where the regeneration times could be defined as the arrival times of customers who find an empty system.

Letting $Y_j = \sum_{i=T_j}^{T_{j+1}-1} X_i$, $\{(Y_j, \tau_j) : j \geq 1\}$ is an i.i.d. sequence. If we continue with our assumption that the process is stationary and also assume that it is regenerative, then we must have τ_j is aperiodic. If $E[\sum_{i=T_1}^{T_2-1} |X_i|] < \infty$, then the steady-state mean of $\{X_i\}$ is

$$\mu_X = \frac{E[Y_1]}{E[\tau_1]}. \quad (2.15)$$

To apply this theory to the estimation of μ_X from simulation, we simulate, in one long run, k regenerative cycles and obtain k observations of Y_i and τ_i . The total length of the run is denoted $n(k)$ because n will depend on k . The sample estimator for μ_X computed from this method is

$$\hat{\mu}_X(k) = \frac{\bar{Y}(k)}{\bar{\tau}(k)}, \quad (2.16)$$

where

$$\bar{Y}(k) = \frac{1}{k} \sum_{j=1}^k Y_j = \frac{1}{n(k)} \sum_{i=1}^{n(k)} X_i \quad (2.17)$$

and

$$\bar{\tau}(k) = \frac{1}{k} \sum_{i=1}^k \tau_i. \quad (2.18)$$

Although $\bar{Y}(k)$ and $\bar{\tau}(k)$ are unbiased estimators of $E(Y_i)$ and $E(\tau_i)$, the quotient $\bar{Y}(k)/\bar{\tau}(k)$ is not an unbiased estimator of μ_X .

The strong law of large numbers (SLLN) states that if $V_1, V_2, \dots$ is a sequence of i.i.d. random variables with $E[V_i] = \mu_V$, $E|V_i| < \infty$ and if $\bar{V}(n)$ is the usual size n sample estimator of μ_V ,

$$\lim_{n \rightarrow \infty} \bar{V}(n) = \mu_V, \quad (2.19)$$

with probability (w.p.) 1. As a consequence of the SLLN, if $E[\sum_{i=T_1}^{T_2-1} |X_i|] < \infty$ and $E[\tau_1] < \infty$, then $\bar{Y}(k)/\bar{\tau}(k)$ is a strongly consistent estimator, i.e., as $k \rightarrow \infty$, $\bar{Y}(k)/\bar{\tau}(k) \rightarrow \mu_X$ (w.p.) 1.

Define a new random variable $Z_j = Y_j - \mu_X \tau_j$. Then $E[Z_j] = 0$ and

$$\sigma_Z^2 = \text{Var}[Z_j] = E[(Y_j - \mu_X \tau_j)^2] = E[Y_j^2] - 2\mu_X E[Y_j \tau_j] + \mu_X^2 E[\tau_j^2]. \quad (2.20)$$

Because the Z_j 's are i.i.d. random variables, under the condition that $0 < \sigma_Z^2 < \infty$, the CLT for the i.i.d. case holds, i.e.,

$$\frac{\overline{Z}(k)}{\sqrt{\sigma_Z^2/k}} \xrightarrow[k \rightarrow \infty]{D} N(0, 1). \quad (2.21)$$

Based on the k observations of Y_j and τ_j , we can use the following to estimate σ_Z^2 :

$$\hat{\sigma}_Z^2(k) \equiv \frac{1}{k} \sum_{j=1}^k Y_j^2 - 2\hat{\mu}_X(k) \frac{1}{k} \sum_{j=1}^k Y_j \tau_j + \hat{\mu}_X^2(k) \frac{1}{k} \sum_{j=1}^k \tau_j^2. \quad (2.22)$$

By replacing σ_Z^2 with $\hat{\sigma}_Z^2(k)$ and dividing numerator and denominator by $\bar{\tau}(k)$ in (2.21) we obtain

$$\frac{\hat{\mu}_X(k) - \mu_x}{\sqrt{\frac{\hat{\sigma}_Z^2(k)}{k[\bar{\tau}(k)]^2}}} \xrightarrow[k \rightarrow \infty]{D} N(0, 1) \quad (2.23)$$

Therefore, for k sufficiently large, an approximate asymptotically valid CI is:

$$\hat{\mu}_X(k) \pm z_{1-\alpha/2} \frac{\sqrt{\hat{\sigma}_Z^2(k)/k}}{\bar{\tau}(k)}. \quad (2.24)$$

The regenerative procedure described above is based on fixing the number of regenerative cycles k and simulating until that number of cycles has been observed. Shedler (1993) outlines a two-stage approach to implementing this procedure. A short preliminary pilot run is used to make estimates of μ_X , σ_Z^2 and $E[\tau_1]$ and then the number k of required regenerative cycles for a relative length of $\pm 100\delta_r\%$ is determined by $(z_{1-\alpha/2} \hat{\sigma}_Z)^2 / (\delta \hat{\mu}_X \bar{\tau}(k))^2$, rounded up.

We can also use regenerative theory to obtain strongly consistent estimators and asymptotic CIs from a simulation of fixed length n . Now the observed number of regenerative cycles k depends on n ; we will denote this dependency by $k(n)$. Shedler (1993) points out (2.23) will hold when k is replaced by $k(n)$ and the resulting estimator $\mu_x(k(n))$ is strongly consistent with bias of order $1/n$. Shedler (1993) also notes that we can reduce bias by continuing the simulation past n to the next regeneration point $k(n) + 1$. The new estimator $\mu_X(k(n) + 1)$ has bias of order $1/n^2$.

The regenerative method is appealing for several reasons: it is easy to understand, eliminates the initial bias problem, and has truly i.i.d. observations. However, there are several disadvantages to this method. We must deal with identifying regeneration points, planning the simulation length, and the possibility of a large amount of computation time. The cycle lengths are unknown prior to the actual running of the simulation and may be very long. Therefore, an extremely long simulation may be required to observe only a few cycles. Because the estimators are ratios, Bratley, Fox, and Schrage (1987) are suspicious of CIs, based on (asymptotic) normality, applied to finite sample sizes.

2.1.5 Autoregressive Method

The Autoregressive Method introduced in the simulation context by Fishman (1973, 1977, and 1978b) assumes a covariance stationary process, which is implied by our assumption of strict stationarity. The other assumption required for this method is that the process is well represented by the autoregressive model:

$$\sum_{j=0}^p b_j (X_{i-j} - \mu_X) = \epsilon_i, \text{ for } i = p+1, p+2, \dots \quad (2.25)$$

where p is the order of the model, $b_0 = 1$, $\mu_X = E[X_i]$, and $\{\epsilon_i\}$ is assumed to be an i.i.d. sequence of random variables with $E[\epsilon_i] = 0$ and $\text{Var}[\epsilon_i] = \sigma_\epsilon^2$ (Law and Kelton 1991).

It can be shown that if the process $\{X_i : i = p, p+1, \dots\}$ satisfies (2.25), then $n\text{Var}[\bar{X}(n)] \rightarrow \sigma_\epsilon^2 / \left(\sum_{j=0}^p b_j\right)^2$ as $n \rightarrow \infty$. Fishman's procedure (Fishman 1973) allows for making sample estimators of the order $\hat{p}$; the constants $\hat{b}_j$ for $j = 1, \dots, \hat{p}$; and the variance $\hat{\sigma}_\epsilon$ form the sample estimators of the covariances $\hat{\gamma}(q)$. If $\hat{b} = 1 + \sum_{j=1}^{\hat{p}} \hat{b}_j$, then

$$\widehat{\text{Var}}[\bar{X}(n)] = \frac{\hat{\sigma}_\epsilon^2}{n^2(\hat{b})} \quad (2.26)$$

is the sample estimator of $\text{Var}[\bar{X}(n)]$. The $100(1 - \alpha)\%$ CI for μ_X is :

$$\bar{X}(n) \pm t_{\hat{f}, 1-\alpha/2} \sqrt{\widehat{\text{Var}}[\bar{X}(n)]}, \quad (2.27)$$

where $\hat{f}$ is the estimated degrees of freedom given by:

$$\hat{f} = \frac{n\hat{b}}{2 \sum_{j=0}^{\hat{p}} (\hat{p} - 2j) \hat{b}_j}. \quad (2.28)$$

Although Wold's Decomposition Theorem (Anderson 1971) tells us that any stationary sequence can be written as an *infinite* order autoregressive model, in the application to simulation output analysis we must of course restrict our model to finite order and to parameters which are sample estimates. In the end, we run the risk of not adequately describing the process.

2.1.6 Standardized Time Series

The method of Standardized Time Series (STS), introduced by Schruben (1983), is a so-called "cancellation" method because in a certain limit theorem the SSVC, σ^2 , cancels out and there is, then, no need to estimate it. The STS method assumes that a functional central limit theorem (FCLT) for the (standardized) partial sums of the process is in force. Such a result holds for processes that are weakly dependent in the sense we discussed earlier, i.e., observations that are far from each other in time are almost independent. See the end of this section for a discussion of processes which obey a FCLT.

If $X_1, \dots, X_n$ are observations obtained from the simulation of the process and $E[X_i] = \mu_X$, then we know from (1.11) that $\sqrt{n}(\bar{X}(n) - \mu_X) \xrightarrow[n \rightarrow \infty]{D} \sigma N(0, 1)$. Let the partial sum process be given by

$$S_n = \sum_{i=1}^n X_i, \quad n = 1, 2, \dots \quad (2.29)$$

Then from the SLLN, under regularity conditions, $S_n/n \rightarrow \mu_X$ w.p. 1 as $n \rightarrow \infty$, and from the CLT, $\sqrt{n}(S_n/n - \mu_X) \xrightarrow[n \rightarrow \infty]{D} \sigma N(0, 1)$. Now let

$$Y_n(t) = \frac{S_{[nt]}}{n}, \quad 0 \leq t \leq 1, \quad (2.30)$$

where $[nt]$ is the integer portion of nt . Schruben (1983) explains that the range of t is $[0, 1]$ as a part of "standardizing" the series, by transforming its length to unit length. Again, from the SLLN, we can show that $Y_n(t) \rightarrow \mu_X t$ with probability 1 as $n \rightarrow \infty$.

Now consider the function

$$Z_n(t) = \sqrt{n} \left(\frac{S_{[nt]}}{n} - \mu_X t \right), \quad 0 \leq t \leq 1. \quad (2.31)$$

Note that $Z_n(t)$ is a function in t , but for given t , $Z_n(t) \xrightarrow[n \rightarrow \infty]{D} N(0, \sigma^2 t)$ by the CLT.

We are now ready to state the FCLT for weakly dependent processes:

$$Z_n(\cdot) \xrightarrow[n \rightarrow \infty]{D} \sigma B(\cdot), \quad (2.32)$$

where $\{B(t) : t \geq 0\}$ is the standard Brownian motion process. Brownian motion is characterized by the following properties:

- $B(0) = 0$;
- $\{B(t) : t \geq 0\}$ has continuous sample paths;
- $\{B(t) : t \geq 0\}$ has stationary and independent increments;
- for every $t > 0$, $B(t)$ is normally distributed with mean 0 and variance t .

For $t = 1$, (2.32) may be written

$$Z_n(1) = \sqrt{n} \left(\frac{S_n}{n} - \mu_X \right) \xrightarrow[n \rightarrow \infty]{D} \sigma B(1) \stackrel{D}{=} \sigma N(0, 1), \quad (2.33)$$

where $\stackrel{D}{=}$ means equality in distribution.

To employ (2.33) in constructing a CI, we would have to know σ . However, a function g of $\{S_n : n \geq 0\}$ can be chosen to cancel the scaling constant σ . The continuous function $g : C[0, 1] \rightarrow \Re$, should have the following properties (Glynn and Iglehart 1990):

- $g(\alpha x) = \alpha g(x)$, for $\alpha > 0$ and for all $x \in C[0, 1]$,
- $g(x - \beta k) = g(x)$ for all $\beta \in \Re$ and $x \in C[0, 1]$, where $k(t) = t$,
- $\Pr[g(B) > 0] = 1$,
- $\Pr[B \in D(g)] = 0$.

The set D is the set of discontinuities of the function g in the space $C[0, 1]$, the space of real-valued continuous functions from the $[0, 1]$ interval into $\Re$. (See Billingsley 1968, for details.)

By application of the continuous mapping theorem and the converging-together theorem, (see Glynn and Iglehart 1990, for details) we have that if g is a function with the above properties, then

$$\frac{Z_n(1)}{g(Z_n)} \xrightarrow[n \rightarrow \infty]{D} \frac{\sigma B(1)}{g(\sigma B)} = \frac{\sigma B(1)}{\sigma g(B)} = \frac{B(1)}{g(B)} \quad . \quad (2.34)$$

The SSVC σ^2 does not appear in the right-hand side limit, and, therefore, need not be estimated. To apply the STS method, we must find a function g to use, as well as the distribution of $B(1)/g(B)$.

Glynn and Iglehart (1990) show that NOBM, where the number of batches is fixed, is a STS method. They define the function g by:

$$g_k(x) = \left[\left(\frac{k}{k-1} \right) \sum_{i=1}^k (\Delta_k x(i/k) - x(1)/k)^2 \right]^{1/2}, \quad (2.35)$$

where $\Delta_h x(t) = x(t) - x(t - 1/h)$. It can be shown that g_k has the properties listed above (see Glynn and Iglehart 1990 for details). Now $\Delta_k B(i/k)$ for $i = 1, \dots, k$ are increments of standard Brownian motion, i.e. are i.i.d. normal random variables with zero mean and variance $1/k$. The quantity $B(1)/k$ is the sample mean of these increments and therefore, $B(1)/g_k(B)$ has a Student's t -distribution with $k-1$ degrees of freedom. The function $g(Z_n)$ is equal to $\sqrt{n} (S_{n,k}/\sqrt{k})$, where $S_{n,k}$ is given in (1.20). Glynn and Iglehart (1990) give the functions that correspond to the STS area and maximum methods, introduced by Schruben (1983).

We now discuss some processes which obey a FCLT and therefore are candidates for application of the STS method. Glynn and Iglehart (1990) list the following as processes for which the FCLT exists: stationary (and some non-stationary) ϕ -mixing processes, strictly stationary strongly mixing processes, associated strictly stationary processes, and regenerative processes. Let $\mathcal{F}_u^v$ be the σ -field of events generated by the process $X_u, X_{u+1}, \dots, X_{u+v}$. Mathematically, the process $\{X_i\}$ is strongly mixing if, for all $k \geq 1$,

$$\sup_{Q \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+j}^\infty} |\Pr(Q \cap F) - \Pr(Q) \Pr(F)| \leq \alpha_j, \quad (2.36)$$

where $\lim_{j \rightarrow \infty} \alpha_j = 0$, and ϕ -mixing if

$$\sup_{Q \in \mathcal{F}_1^k, F \in \mathcal{F}_{k+j}^\infty} |\Pr(F|Q) - \Pr(F)| \leq \phi_j, \quad (2.37)$$

where $\lim_{j \rightarrow \infty} \phi_j = 0$.

Although ϕ -mixing is stronger than strongly mixing, both of these definitions imply that events far from each other in time are almost independent. We have previously discussed, in Section (2.1.4), what it means for a process to be regenerative. Schruben (1983) states that m -dependent processes also qualify as ϕ -mixing if m is bounded. Using the notation above, a process is m -dependent if the σ -fields $\mathcal{F}_1^k$ and $\mathcal{F}_{k+m+1}^\infty$ are independent (for all $k \geq 1$).

2.2 Fixed Sample Size Approaches to NOBM

2.2.1 Classical NOBM

We use the term *classical* NOBM to refer to the nonoverlapping batch means method when the number of batches is held fixed. If we fix our sample size n and the number of batches k prior to running the simulation, then we have to live with whatever CI results (it may be too wide to be informative). We would like to choose k and m that give us the “best” results. Let

$$H_{\alpha,k}(n) = t_{\alpha/2,k-1} \frac{S_{n,k}}{\sqrt{k}} \quad (2.38)$$

denote the half-length of the CI given in (1.21). Although here we are considering n to be fixed, other procedures using NOBM allow n to change. Schmeiser (1982) gives the assumptions of the method of NOBM as:

1. Initial transient effects have been removed;
2. For a run length of n , a number of batches k^* and associated batch size $m^* = n/k^*$ exist such that the dependency and nonnormality of the batch means are negligible for $m \geq m^*$ and $k \leq k^*$; and
3. The problem of n/k not being integer is insignificant.

Note that assumption 2 above implies that n is large enough so that dependency and nonnormality is negligible for some k .

What do we choose for k and m ? And on what criteria do we base this choice? As a starting point, we would like a value of k that would generate a CI that would cover the mean with probability at least equal to our desired confidence level, i.e., we

would like $\Pr [|\overline{X}(n) - \mu_X| \leq H_{\alpha,k}(n)] \simeq 1 - \alpha$. Based on this criterion alone, $k = 2$ is the best choice because it yields a large $H_{\alpha,k}(n)$, which in turn means that our CI will have a high probability of covering the mean. However, $H_{\alpha,k}(n)$ may be so large that we do not achieve an informative CI. Also, $\Pr [|\overline{X}(n) - \mu_X| \leq H_{\alpha,k}(n)]$ will most likely be larger than $1 - \alpha$ in actuality.

We might wish to consider the second criterion, $E[H_{\alpha,k}(n)]$. These two criteria compete with each other so there is a trade-off between expected half-length and coverage. Other measures Schmeiser (1982) considers are: the standard deviation of the half-length, $\sqrt{\text{Var}[H_{\alpha,k}(n)]}$, the coefficient of variation of the half-length, $\text{CV}[H_{\alpha,k}(n)] \equiv \sqrt{\text{Var}[H_{\alpha,k}(n)]}/E[H_{\alpha,k}(n)]$, and for all $\mu_1 \neq \mu_X$, the function

$$\beta_{\alpha,k}(\mu_1) \equiv \Pr [|\overline{X}(n) - \mu_1| \leq H_{\alpha,k}(n)]. \quad (2.39)$$

This function may be thought of as the probability of a type II error in hypothesis testing. Therefore, smaller $\beta_{\alpha,k}(\mu_1)$ for $\mu_1 \neq \mu_X$ means better performance.

Based on properties of the half-length, Schmeiser's (1982) results, derived analytically indicate that more batches are needed for smaller α (i.e., higher confidence levels). However, a number of batches larger than 30 had negligible effect on the properties of the half-length.

Schmeiser (1982) favors the function $\beta_{\alpha,k}(\mu_1)$ as a performance measure, since it takes into account both $\overline{X}(n)$ and $S_{n,k}$ together, as opposed to the moments of the half-length $H_{\alpha,k}(n)$ which are functions of $S_{n,k}$ only. Also, $\beta_{\alpha,k}$ is a coverage function, directly related to coverage of the mean, which was our first criterion for an acceptable k .

Plots of the coverage function $\beta_{\alpha,k}(\mu_1)$ (Schmeiser 1982) in terms of covering points removed from the actual mean μ_X by factors δ_k of the $\text{Var}[\overline{X}(n)]$ show that the decrease in $\beta_{\alpha,k}(\mu_1)$ due to increasing k by 1: (a) decreases as k increases, (b) decreases as α increases, and (c) is small when $\delta_k < 1$.

Schmeiser (1982) leans toward choosing a small k to avoid degrading coverage and reaches the conclusion that $k > 30$ should rarely be required to achieve desired results in most simulations. Consequently, he recommends a number of batches between 10 and 30. A caveat is included that tests for independence and normality have low power for small sample sizes.

2.2.2 Nonclassical NOBM

One approach to fixed sample size NOBM starts with a fixed total sample size of n and allows for varying the number of batches k and consequently the batch size m . Fishman (1978a) proposed a procedure for determining batch size for a sample of size n which relies on a statistical test for independence between batch means. Let $\rho_1(m)$ denote the lag-one correlation:

$$\rho_1(m) \equiv \text{Corr}[Y_j(m), Y_{j+1}(m)]. \quad (2.40)$$

The von Neumann test for independence tests the null hypothesis $H_0 : \rho_1(m) = 0$ using the statistic:

$$C_k = 1 - \frac{\sum_{j=1}^{k-1} (Y_j(m) - Y_{j+1}(m))^2}{2 \sum_{j=1}^k (Y_j(m) - \bar{X}(n))^2}. \quad (2.41)$$

If the $Y_j(m)$'s are normally distributed, then under H_0 ,

$$C_k \sim N\left(0, \frac{k-2}{k^2-1}\right), \quad (2.42)$$

for k as small as 8. Fishman (1978a) points out that for small m it is doubtful that the $Y_j(m)$'s are normally distributed. However, if n is large then k is large (when m is small) and the large sample properties suggest that we can approximate the distribution of $C_k/\sqrt{(k-2)/(k^2-1)}$ with the $N(0, 1)$ distribution, provided that the $Y_i(m)$'s are independent and identically distributed. Also as m increases (and k decreases) we expect that the distribution of the $Y_i(m)$'s approaches the normal distribution and we can reasonably assume approximate normality of the C_k as long as $k \geq 8$. If the covariance is assumed to be a monotone decreasing function, then a one-tail test is appropriate. So if we desire a level β test (level β means the $\text{Pr}[\text{rejecting } H_0 | H_0 \text{ is true}] \leq \beta$), then we will accept the null hypothesis that $\rho_1(m) = 0$ if $C_k \leq z_{1-\beta} \sqrt{(k-2)/(k^2-1)}$.

Fishman's (1978a) algorithm (given in Figure 2.1 begins with a total sample size of n . On the first iteration, the batch size, m , is set to one; the number of batches, k , is set to n ; and the batch means $Y_1(m), \dots, Y_k(m)$ and the test statistic C_k are calculated. If the null hypothesis is rejected, then the batch size is doubled and new batch means and C_k are calculated. If the null hypothesis is accepted, then a CI of

1. $m \leftarrow 1$.
2. $k \leftarrow \lfloor n/m \rfloor$.
3. If $k < 8$, then return indicating failure.
4. Compute the batch means $Y_1(m), \dots, Y_k(m)$ using (1.8).
5. Compute C_k using (2.41).
6. If $C_k > z_{1-\beta} \sqrt{(k-2)/(k^2-1)}$, then take $m \leftarrow 2m$ and go to 2.
7. Otherwise, $m^* \leftarrow m$ and $k^* \leftarrow k$.
8. Compute S_{n,k^*}^2 using (1.20).
9. Return with $(1 - \alpha)\%$ CI constructed using (1.21).

Figure 2.1: Fishman’s Fixed Sample Size Algorithm for Determining Batch Size m and Number of Batches k .

the form (1.21) is constructed. If the batch size becomes so large that the number of batches $k < 8$, then the algorithm fails to determine an optimal batch size k^* .

Fishman(1978a) tested the algorithm with the $M/M/1$ queue, varying both the utilization factor and the run length. Sixty replications at each level of utilization and each value of n were made. If the algorithm failed to determine m^* , the algorithm was terminated. In general, larger initial n gave better coverage, although most fell short of the nominal 95%. The poorest performance was on the cases where utilization was high, but the sample size was small. Relatively few replications failed to find an m^* and those which did fail were at the higher levels of utilization and lower values of starting n .

Fishman (1978a) shows that underestimation of the variance becomes worse for fixed n and increasing utilization. He suggests that a potential source of error in the algorithm is underestimation of the m^* required for independence of the batch means, thereby resulting in an estimated variance which is too small. This prompted him to increase β and thereby impose a stronger requirement for independence. The higher β is, the more likely we are to reject H_0 whether it is true or false. So by increasing β we will increase the $\Pr[\text{rejecting } H_0 | H_0 \text{ is false}]$ which would seem to be our primary concern. Although increasing β in general increased the value of m^* reached and also increased the number of runs which failed to find an m^* , the coverages did not improve. Fishman discounts the nonnormality of the batch means: “Although the histograms (of the batch means) revealed a clear departure from normality, the error

in using a t distribution is generally regarded as being insensitive to this departure.”

Fishman theorizes that a short sample where the observations are highly correlated (as in the $M/M/1$ queue with utilization 0.9) may be insufficient to disclose the variation in the underlying stochastic process and we can only improve performance by increasing n .

Song and Schmeiser (1995) investigate the effects of batch size m on the mean square error (MSE) of the variance estimator. They seek the optimal batch size which yields the minimum MSE. However, we do not know how the coverage of the CI is affected by choosing a batch size based on this criterion. The MSE of the variance estimator is given by:

$$\begin{aligned} \text{MSE}(\widehat{\text{Var}}[\overline{X}(n)]) &= \text{E} \left[\left(\widehat{\text{Var}}[\overline{X}(n)] - \text{Var}[\overline{X}(n)] \right)^2 \right] \\ &= \text{Bias}^2(\widehat{\text{Var}}[\overline{X}(n)]) + \text{Var}[\widehat{\text{Var}}[\overline{X}(n)]], \end{aligned} \quad (2.43)$$

where $\text{Bias} \equiv \text{E}[\widehat{\text{Var}}[\overline{X}(n)] - \text{Var}[\overline{X}(n)]]$. Song and Schmeiser (1995) derive an approximation of the MSE of the variance estimator as a function of m :

$$\text{MSE}(\widehat{\text{Var}}(m)) \approx \left(\frac{c_b^2 \gamma_1^2}{n^2 m^2} + \frac{m c_v \gamma_0^2}{n^3} \right) \gamma^2(0). \quad (2.44)$$

The optimal batch size is given by:

$$m^* \approx \left(2n \left(\frac{c_b^2}{c_v} \right) \left(\frac{\gamma_1}{\gamma_0} \right)^2 \right)^{1/3}, \quad (2.45)$$

where c_b and c_v depend on the method of variance estimation being used. The ratio γ_1/γ_2 is called the balance point, where $\gamma_1 \equiv \sum_{h=-\infty}^{\infty} \rho_h = 1 + 2 \sum_{h=1}^{\infty} \rho_h$ and $\gamma_0 \equiv \sum_{h=-\infty}^{\infty} |h| \rho_h = 2 \sum_{h=1}^{\infty} h \rho_h$ and ρ_h is the correlation at lag h , for the underlying process $X_i : i = 1, 2, \dots$

The sensitivity of the MSE at points close to m^* was also considered by calculating the second derivative of (2.44). Song and Schmeiser (1995) considered four variance estimation methods. Using our notation, the estimator they study for NOBM is $(1/k)S_{n,k}^2$. For NOBM they give $c_b = 1$ and $c_v = 2$. While NOBM compares favorably on the basis of asymptotic bias, it has relatively large asymptotic variance compared to the other three methods. The estimator which is a linear combination of the NOBM estimator and the STS area estimator has smallest variance. The MSE of the NOBM estimator is more sensitive at m^* , but m^* is smaller for NOBM than for the other three methods.

Based on their asymptotic results, Song and Schmeiser (1995) give the following as the approximate optimal MSE batch size for a finite sample of size n :

$$\vec{m}^* \approx 1 + \left(2n \left(\frac{c_b^2}{c_v} \right) \left(\frac{\gamma_1}{\gamma_0} \right)^2 \right)^{1/3}, \quad (2.46)$$

where γ_0 and γ_1 are estimated from the data. Song and Schmeiser (1995) give empirical results for the OBM method.

2.3 Sequential Approaches to Batch Means

In using the fixed sample size procedure, we have assumed that the fixed total sample size is large enough that dependence and nonnormality of the batch means is negligible. We are then forced to live with whatever length of CI results. Often the resulting CI is too wide to be informative, i.e., it does not possess the desired length. Sequential procedures, however, allow for increasing the sample size at each iteration.

In this work, we are concentrating on the NOBM method of CI construction and will therefore discuss in some detail sequential procedures which have been proposed for the NOBM method. However, it should be mentioned that sequential procedures exist for other methods, for example Lavenberg and Sauer (1977) and Fishman (1977) discuss sequential procedures for regenerative simulations and Heidelberger and Welch (1981a, 1981b, and 1983) use sequential procedures in a spectral estimation context.

2.3.1 Law and Carson's Procedure

Law and Carson (1979) developed a sequential procedure for NOBM for covariance stationary processes. They point out three potential sources of error when the method of NOBM is used: (i) bias in the estimator $\hat{\sigma}^2$ due to a batch size m that is too small for the $Y_j(m)$'s to be uncorrelated, (ii) nonnormality of the $Y_j(m)$'s, and (iii) the process $\{X_i\}$ may not in fact be covariance stationary. Law and Carson discount the effect of nonnormality for a number of batches $k \geq 20$ and focus on the problem of bias caused by correlation of the batch means.

The fixed number of batches (FNB) rule calls for fixing the number of batches and sequentially increasing the batch size until the sample lag-one correlation estimator of $\rho_1(m)$ between batch means is less than some specified level, e.g., $\hat{\rho}_1(m) \leq 0.05$. See Section 2.3.2 for further discussion of the FNB rule.

However, empirical studies have shown we may quite often achieve an acceptable $\hat{\rho}_1(m)$ when in fact it can be shown analytically that $\rho_1(m)$ is greater than the specified level. The problem is that estimators of correlation are biased and have large variance when the number of samples is small. Law and Carson (1979) found that some systems required as many as 400 batches for a precise estimate of $\rho_1(m)$.

This prompted them to investigate the possibility of using ℓk batches of size m to make inferences about the correlation between k batches of size ℓm . Law and Carson (1979) chose 34 stochastic processes where $\rho_1(m)$ could be computed analytically and plotted $\rho_1(m)$ vs. m . Three types of behavior were observed: (i) lag one correlation strictly decreases to zero, (ii) lag one correlation changes directions (one or more times) then strictly decreases to zero, and (iii) lag one correlation is less than zero.

The algorithm (given in Figure 2.2 from Law and Carson (1979)) begins by fixing a level of correlation, c , as a stopping value; a relative half-length, $\gamma > 0$; and positive integers ℓ, k, n_0, n_1 (ℓk even, $\ell k/2$ even, $n_0 < n_1 < 2n_0$ each divisible by ℓk). Then n_1 observations are collected. The observations are divided into ℓk batches of size m . The lag one correlation coefficient is calculated from these observations using the jackknifed estimator:

$$\bar{\rho}_1(\ell k, m) = 2\hat{\rho}_1(\ell k, m) - [\hat{\rho}_1^1(\ell k/2, m) + \hat{\rho}_1^2(\ell k/2, m)]/2, \quad (2.47)$$

where $\hat{\rho}_1^1(\ell k/2, m)$ and $\hat{\rho}_1^2(\ell k/2, m)$ are the usual lag one correlation estimators based on the first and last $\ell k/2$ batches, respectively. Law and Carson use this jackknifed estimator because in general it is less biased.

If the acceptable level of correlation is not achieved, i.e., if $\bar{\rho}_1(\ell k, m) \not\leq c$, then the total number of observations is increased and the additional observations are collected. The total number of observations collected doubles every other iteration. When the desired level of correlation is reached, i.e., $\bar{\rho}_1(\ell k, m) \leq c$, the total observations are divided into k batches of size ℓm and a confidence interval is constructed, using (1.21). If the desired relative half-length is not achieved, then more observations are collected and the algorithm is continued. Law and Carson (1979) tested the algorithm on 14 stochastic processes and got good results for a 90% confidence level, using $k = 40$ and $\ell = 10$ chosen because empirical studies showed that 400 batches were adequate to obtain precise correlation estimates. They report that for a CI with high probability of being valid, the choices of $\gamma = 0.075$ and $c = 0.4$ are appropriate. Note that though the choice of c seems high, it is used as the stopping criterion for the correlation of

0.	Fix positive integers ℓ, k, n_0, n_1 (ℓk even, $\ell k/2$ even, $n_0 < n_1 < 2n_0$, n_1 and $2n_0$ each divisible by ℓk), the stopping value $c > 0$, the relative half-length $\gamma > 0$; let $i = 1$; collect n_1 observations.
1.	a. Divide the n_i observations into ℓk batches of size $m = n_i/(\ell k)$. Compute $\bar{\rho}_1(\ell k, m)$ from $Y_j(m)$ ($j = 1, \dots, \ell k$). If $\bar{\rho}_1(\ell k, m) \geq c$, GO TO 2. If $\bar{\rho}_1(\ell k, m) \leq 0$, GO TO 1c. Otherwise, GO TO 1b. b. Divide n_i into $\ell k/2$ batches of size $2m$. Compute $\bar{\rho}_1(\ell k/2, 2m)$ from $Y_1(2m), \dots, Y_{\ell k/2}(2m)$. If $\bar{\rho}_1(\ell k/2, 2m) < \bar{\rho}_1(\ell k, m)$, go to 1c. Otherwise, GO TO 2. c. Divide n_i into k batches of size ℓm. Compute $\bar{X}(n_i)$ and $S_{n_i, k}^2$. Compute the half-length $H_{\alpha, k}(n_i)$ using (2.38). If $H_{\alpha, k}(n_i)/\bar{X}(n) < \gamma$, then construct a confidence interval using (1.21) and STOP. Otherwise, GO TO 2.
2.	$i \leftarrow i + 1$. $n_i \leftarrow 2n_{i-2}$. Collect the additional $n_i - n_{i-1}$ observations and GO TO 1a.

Figure 2.2: Law and Carson's Sequential Algorithm for Determining Batch Size m and Number of Batches k .

ℓk batches of size m ; when, in the end, k batches of size ℓm are used for the variance estimator. (See Law and Carson 1979 for details on the relationship of $\rho(m)$ to $\rho(\ell m)$.)

2.3.2 Fixed Number of Batches (FNB)

If the process $\{X_i : i = 1, 2, \dots\}$ is weakly stationary for batch size m , then the batch means process $\{Y_j(m) : j = 1, 2, \dots\}$ is also weakly stationary. If $m_n \equiv n/k$, it can be shown that (Alexopoulos and Seila 1996, for example):

$$\text{Var}[\bar{X}(n)] = \frac{\text{Var}[\bar{X}(m_n)]}{k} \left(1 + \frac{n \text{Var}[\bar{X}(n)] - m_n \text{Var}[\bar{X}(m_n)]}{m_n \text{Var}[\bar{X}(m_n)]} \right). \quad (2.48)$$

If we fix k and let $n \rightarrow \infty$ then $\text{Var}[\bar{X}(n)] / (\text{Var}[\bar{X}(m_n)]/k) \rightarrow 1$. This leads to the FNB rule: fix the number of batches and let the batch size m_n increase as the total sample size n increases. The FNB rule has some limitations. (See Alexopoulos and Seila 1996 and Fishman 1996.) The sample estimator $m_n S_{n, k}^2$, when k is fixed, is not a consistent estimator of the SSVC σ^2 , which results in CIs which tend to be wider

than ones constructed using a consistent estimator. Also, statistical fluctuations in the half-length of the CI do not diminish relative to statistical fluctuations in the sample mean. Fishman (1996) characterizes this behavior as “unappealing,” particularly compared to the i.i.d case, but does not comment on its relative significance.

2.3.3 Square Root Rule (SQRT)

The problems with the FNB rule lead us to consider a scheme where both the number of batches k_n and the batch size m_n grow as the total sample size n does. Here the number of batches is indexed by n to indicate that it is a function of the sample size. The SQRT rule is such a procedure where, k_n and m_n are chosen such that $\lim_{n \rightarrow \infty} n^{1/2} m_n = 1$ and $\lim_{n \rightarrow \infty} n^{1/2} k_n = 1$. This method, however, tends to underestimate $\text{Var}[\bar{X}(n)]$ for fixed n , which results in poorer coverage. See Fishman (1996).

2.3.4 LBATCH and ABATCH Procedures

We see that both the FNB and SQRT rules have advantages and disadvantages. Although for a given n the FNB rule yields CIs with better coverage than the SQRT rule (because of SQRT rule’s tendency to underestimate σ^2), the mean CI length tends to exceed that of the CI generated from the SQRT rule. For both rules, however, the coverages of the generated CIs approach $1 - \alpha$ as $n \rightarrow \infty$. So the SQRT rule exhibits greater statistical efficiency as $n \rightarrow \infty$. See Fishman (1996) for graphical comparisons of the behavior of the variance estimators and coverage properties of the FNB and SQRT rules.

Fishman (1996) suggests that dynamically shifting between the FNB and SQRT rules would retain the advantages of both methods and reduce the effects of their disadvantages. Such a procedure will result in a random batch size and number of batches. The LBATCH procedure (Fishman 1996) given in Figure 2.3 below, begins with a fixed number of batches and tests for correlation between batch means. If the test detects correlation, then the FNB rule is employed, in order to reduce the correlation as fast as possible. The number of batches remains fixed; the total number of observations and hence the batch size is doubled; and additional observations are taken. This process continues until the test no longer detects correlation between batch means. Then the SQRT rule is used, i.e., additional observations are taken and

the number of batches and the batch size are taken to be $O(\sqrt{n})$. This rule aims at making the distribution of the statistic $(\bar{X}(n_j) - \mu_X) / \sqrt{S_{n_j, k_j}^2 / k_j}$, where n_j and k_j are the sample size and the number of batches on the j th iteration, converge to $N(0, 1)$ as fast as possible. In the LBATCH and ABATCH procedures detailed below, both the batch size and the number of batches grow by a factor of approximately $\sqrt{2}$ on each application of the SQRT rule, so that the total sample size grows by a factor of 2 on each iteration. The SQRT rule for determining batch size and number of batches is employed on all future iterations (i.e., increases to the total sample size n and collecting of additional observations) until a predetermined number of observations is reached. For additional information on and possible modifications to the LBATCH procedure, see Fishman (1996).

The ABATCH procedure is similar to the LBATCH, except the test for independence is repeated at all iterations. If correlation between batch means is detected at some point after the SQRT rule has been invoked, the procedure returns to the FNB rule for the next iteration. Future iterations continue according to the FNB rule, until the hypothesis of independence is again accepted. This can cause the procedure to switch back and forth between the two rules.

2.3.5 Two-stage Stopping Procedure for Batch Means

As an alternative to fully sequential procedures, Nakayama (1994) has developed a two-stage procedure for use in CI construction from simulation output. The method begins with a pilot run consisting of a fixed number of batches of equal size. An STS (NOBM) method is applied to the output of this preliminary run and then the number of additional batches necessary to achieve a CI of specified half-length is estimated. Additional observations are then collected and the CI is constructed. We have already discussed the conditions under which STS yields an asymptotically valid CI (see Section 2.1.6). See Figures 2.5 and 2.6 for Nakayama's procedures based on either an absolute or relative CI half-length requirement.

Nakayama (1994) sees some problems with fully sequential procedures which terminate only when the specification for half-length is met: 1) the run length is randomly determined and may result in a very long run using more computer resources than desired, 2) the run may terminate prematurely due to statistical variability, and 3) consistent estimation methods such as spectral analysis and the autoregressive

Figure 2.3: LBATCH Algorithm for Determining Batch Size m and Number of Batches k .

0. Fix integers K =minimum number of batches, MS =minimum batch size, η and N =desired sample size, where $N = 2^\eta MSK$.
 Fix confidence level $1 - \alpha$.
 $k \leftarrow K$, and $m \leftarrow MS$.
 If $M = 1$, $\tilde{M} \leftarrow 3/2$; otherwise $\tilde{M}S \leftarrow \lfloor \sqrt{2}MS + .50 \rfloor$.
 $\tilde{K} \leftarrow \lfloor \sqrt{2}K + .50 \rfloor$.
 Calculate multipliers used to determine batch size and number of batches for the second application of the SQRT rule: $g_m \leftarrow \tilde{M}S/MS$ and $f_k \leftarrow \tilde{K}/K$.
 Set alternate stopping value: $\tilde{N} \leftarrow 2^{\eta-1} \tilde{M}S\tilde{K}$.
 $i \leftarrow 0$.
  1. If $n > N$ or if $n > \tilde{N}$, then STOP.
 $n \leftarrow mk$.
 Collect observations $X_{i+1}, \dots, X_n$.
    2. Compute: batch means $Y_1(m), \dots, Y_k(m)$; $\bar{X}(n)$ as an estimator of μ_X ; and $S_{n,k}$.
    3. Construct CI: $\bar{X}(n) \pm t_{k-1, 1-\alpha/2} S_{n,k} / \sqrt{k}$.
    4. $i \leftarrow n$.
    5. Test H_0 : $Y_1(m), \dots, Y_k(m)$ are independent.
 If H_0 rejected: $m \leftarrow 2m$ (FNB rule) and GO TO 1.
    6. If $m = 1$, then $m \leftarrow 2$ (FNB rule) and GO TO 1.
    7. Calculate new batch size and number of batches according to the SQRT rule:
 $m \leftarrow mg_m$ and $k \leftarrow kf_k$
    8. Calculate new multipliers for batch size and number of batches for next application of the SQRT rule:
 If $g_m = \tilde{M}S/MS$: $g_m \leftarrow 2MS/\tilde{M}S$ and $f_k \leftarrow 2K/\tilde{K}$. GO TO 1.
 Otherwise: $g_m \leftarrow \tilde{M}S/MS$ and $f_k \leftarrow \tilde{K}/K$. GO TO 1.

Figure 2.4: ABATCH Algorithm for Determining Batch Size m and Number of Batches k .

0. $k \leftarrow$ number of batches in first stage ($k \geq 2$) $\epsilon \leftarrow$ desired absolute half-length of CI $m \leftarrow$ batch size   1. Collect mk observations. Calculate batch means $Y_1(m), \dots, Y_k(m)$, grand sample mean $\bar{X}(mk)$, and sample variance $S_{mk,k}$ Calculate $N_a \leftarrow \max\{k, \lceil (S_{mk,k} t_{k-1, 1-\alpha/2} / \epsilon)^2 \rceil\}$ If $N_a = k$, then GO TO 3.   2. Collect $(N_a - k)m$ additional observations. Calculate batch means $Y_{k+1}(m), \dots, Y_{N_a}(m)$. (Note that although Nakayama includes calculation of the means of the new batches as part of his algorithm, all we really need is the grand mean $\bar{X}(N_a m) = (1/N_a m) \sum_{i=1}^{N_a m} X_i$.)   3. Calculate grand mean $\bar{X}(N_a m) = \frac{1}{N_a} \sum_{j=1}^{N_a} Y_j(m)$ (total sample size is now $N_a m$) The resulting $(1 - \alpha)$ CI is: $\bar{X}(N_a m) \pm \epsilon$

Figure 2.5: Nakayama's Absolute Half-length Two-stage Stopping Procedure

method require some sophistication on the part of the user, making them harder to use. The second problem surfaces when the sample mean and sample variance are positively correlated. A small sample variance, resulting in a narrow CI, may occur when the sample mean is small. (See Sargent, Kang and Goldsman 1992.) Therefore, the CI may not have the desired coverage. Nakayama views the third problem as the most limiting and, therefore, concentrates on STS approaches, which do not require a consistent estimator of the variance.

Included in Nakayama 1994 are some “rules of thumb” for choosing the initial batch size m and number of batches k . Lavenberg and Sauer (1977) select $\epsilon \leq 0.025$ for constructing relative length CIs. Nakayama (1994) suggests choosing ϵ in this range and using a first-stage run length of $1/\epsilon^2$, i.e., an initial batch size of $1/(k\epsilon^2)$. For certain queueing systems, when we desire an absolute length CI, we may use Whitt's (1989a and 1989b) approximation of asymptotic variance and his method for estimating a required total run length t^* and choose a run length of $t^*/2$, i.e., a batch size of $t^*/(2k)$. “By doing this, we (hopefully) ensure that the size of the initial batches is sufficiently large so that they are approximately independent and normally distributed.” (Nakayama 1994)

0. $k \leftarrow$ number of batches in first stage ($k \geq 2$). $\epsilon \leftarrow$ desired relative half-length of CI. $m \leftarrow$ batch size.   1. Collect mk observations. Calculate batch means $Y_1(m), \dots, Y_k(m)$, grand mean $\bar{X}(mk) = \frac{1}{k} \sum_{j=1}^k Y_j(m)$, and sample variance $S_{mk,k}$. Calculate $N_a \leftarrow \max\{k, \lceil (S_{mk,k} t_{k-1, 1-\alpha/2} / (\epsilon \bar{X}(mk)))^2 \rceil\}$. If $N_a = k$, then GO TO 3.   2. Collect $(N_a - k)m$ additional observations. Calculate batch means $Y_{k+1}(m), \dots, Y_{N_a}(m)$. (Note that although Nakayama includes calculation of the means of the new batches as part of his algorithm, all we really need is the grand mean $\bar{X}(N_a m) = (1/N_a m) \sum_{i=1}^{N_a m} X_i$.)   3. Calculate grand mean $\bar{X}(N_a m) = \frac{1}{N_a} \sum_{j=1}^{N_a} Y_j(m)$ (total sample size is now $N_a m$) The resulting $(1 - \alpha)$ CI is: $\bar{X}(N_a m) [1 \pm \epsilon]$

Figure 2.6: Nakayama's Relative Half-length Two-stage Stopping Procedure

For choosing the initial number of batches k , Nakayama (1994) favors Schmeiser's (1982) procedure for nonsequential batch means, where the number of batches is kept small (10 to 30) and recommends an initial number of batches of 5 to 15.

Nakayama conducted experiments using the two-stage procedure for both NOBM and Standardized Sum (SS) on the simulation of the waiting time for the M/M/1 queue at two utilization values, 0.5 and 0.8. He calculated the first-stage run size of 6720 based on Whitt's approximations, for the 0.5 utilization case. On 2000 replications, using 6 (for SS) and 7 (NOBM) for the initial number of batches, he got good coverage for both relative and absolute length requirements. However, when the initial batch size was reduced to 1/4 of the original size, the coverage was not as good, demonstrating the need for an adequate batch size. A run length of 759,100 for utilization of 0.8 was indicated by Whitt's approximation, however Nakayama chose to stay with a run length of 6720 in order to show that the batch size required for good results depends on the stochastic process being simulated. Of course, the actual coverage, using this relatively small batch size for the higher utilization, was not good. (See Nakayama 1994 for details.)

Chapter 3

Convergence Properties of the Batch Means Method

As detailed in Section 1.2, the standard batch means method for constructing asymptotically valid confidence intervals centered on the sample mean of a stationary process is based on the assumption that the batch means are i.i.d. normal random variables. In this chapter we investigate both the theoretical and empirical convergence properties of the batch means process obtained from a stationary process. We consider the classical batch means method where the number of batches k is fixed.

3.1 Asymptotic Joint Distribution of Batch Means

We begin by stating and proving a theorem that characterizes the asymptotic behavior of the k -dimensional vector of batch means under fairly general conditions as the batch size $m \rightarrow \infty$.

Definition 3.1 *Let $\mathcal{D}$ denote the space of functions on $[0, 1]$ that are right-continuous and have left-hand limits, and let k be a fixed positive integer. For $X, Y \in \mathcal{D}$, let $d(X, Y)$ denote the Skorohod metric on $\mathcal{D}$ (Billingsley 1968): that is, if Λ denotes the class of strictly increasing, continuous mappings of $[0, 1]$ onto itself, then $d(X, Y)$ is defined to be the infimum of those positive δ for which there exists $\lambda \in \Lambda$ such that*

$$\left. \begin{array}{l} \sup_t |\lambda t - t| \leq \delta \\ \sup_t |X(t) - Y(\lambda t)| \leq \delta \end{array} \right\}. \quad (3.1)$$

Note that if $\lambda \in \Lambda$, then so does $\lambda' \equiv \lambda^{-1}$; and as t runs from 0 to 1, so does $t' \equiv \lambda t$. Also $\sup_t |\lambda t - t| = \sup_{t'} |\lambda' t' - t'|$ and $\sup_t |X(t) - Y(\lambda t)| = \sup_{t'} |X(\lambda' t') - Y(t')|$. Thus we see that as a completely equivalent alternative to the definition (3.1), we may take $d(X, Y)$ to be the infimum of all $\delta > 0$ for which there exists $\lambda' \in \Lambda$ such that

$$\left. \begin{array}{l} \sup_{t'} |\lambda' t' - t'| \leq \delta \\ \sup_{t'} |X(\lambda' t') - Y(t')| \leq \delta \end{array} \right\}. \quad (3.2)$$

Definition 3.2 *The function $\psi : \mathcal{D} \rightarrow \mathbb{R}^k$ is defined as follows: for every $X \in \mathcal{D}$,*

$$\psi(X(\cdot, \omega)) \equiv \sqrt{k} \left[X\left(\frac{1}{k}, \omega\right) - X\left(\frac{0}{k}, \omega\right), X\left(\frac{2}{k}, \omega\right) - X\left(\frac{1}{k}, \omega\right), \dots, X\left(\frac{k}{k}, \omega\right) - X\left(\frac{k-1}{k}, \omega\right) \right]. \quad (3.3)$$

Definition 3.3 *Let $\mathcal{K} \subset \mathcal{D}$ denote the subset of functions in $\mathcal{D}$ that are continuous at the multiples of $1/k$ in the unit interval:*

$$\mathcal{K} \equiv \left\{ X \in \mathcal{D} : X\left(\frac{j}{k}\right) = X\left(\frac{j}{k}-\right) \text{ for } j = 1, 2, \dots, k \right\}. \quad (3.4)$$

Definition 3.4 *Let $\{\xi_n : n = 1, 2, \dots\}$ be ϕ -mixing with $\sum_n \sqrt{\phi_n} < \infty$, $E[\xi_0] = \mu$, $\text{Var}[\xi_0] < \infty$ and σ^2 be the SSVV (1.12) of the process $\{\xi_n\}$, where $\sigma^2 > 0$. Let $\sigma = \sqrt{\sigma^2}$ and let k , the number of batches used in the NOBM method, be fixed. Define $U_n \in \mathcal{D}$ to be*

$$U_n(t, \omega) = \frac{1}{\sigma \sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} [\xi_i(\omega) - \mu], \quad 0 \leq t \leq 1. \quad (3.5)$$

Theorem 3.1 *If $\mathcal{D}$, ψ , $\mathcal{K}$, and U_n are specified by Definitions 3.1, 3.2, 3.3 and 3.4 respectively, then:*

1. $\psi(\cdot)$ is continuous at every $X_0 \in \mathcal{K} \subset \mathcal{D}$.
2. If $B(\cdot, \omega)$ is standard Brownian motion as described in Section 2.1.6, then $\psi(B(\cdot, \omega))$ is a k -dimensional multivariate standard normal vector with the $k \times 1$ null mean vector $\mathbf{0}_k$ and the $k \times k$ identity covariance matrix $\mathbf{I}_k$,

$$\psi(B(\cdot, \omega)) \sim N_k(\mathbf{0}_k, \mathbf{I}_k). \quad (3.6)$$

3. The function $\psi(U_n(\cdot, \omega))$ represents the vector of standardized batch means corresponding to the process $\{\xi_n\}$; and as $n \rightarrow \infty$, $\psi(U_n(\cdot, \omega))$ converges in distribution to a standard k -variate normal distribution,

$$\psi(U_n(\cdot, \omega)) \xrightarrow[n \rightarrow \infty]{D} N_k(\mathbf{0}_k, \mathbf{I}_k). \quad (3.7)$$

Proof of 1. Let $\varepsilon > 0$ be an arbitrarily small positive number and choose $X_0 \in \mathcal{K}$ arbitrarily. To demonstrate the continuity of ψ at X_0 , we will show the existence of an $\eta = \eta(X_0, \varepsilon, k) > 0$ such that if $X \in \mathcal{D}$ and $d(X_0, X) < \eta$, then $|\psi(X_0) - \psi(X)| < f(X_0, \varepsilon, k)$ where $f(X_0, \varepsilon, k) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Since $X_0 \in \mathcal{K}$, we can find an $\eta_1 = \eta_1(X_0, \varepsilon, k) > 0$ such that for any $j \in \{0, 1, \dots, k\}$,

$$|t - j/k| < \eta_1 \text{ and } t \in [0, 1] \quad (3.8)$$

implies

$$|X_0(t) - X_0(j/k)| < \varepsilon. \quad (3.9)$$

Choose $\delta = \min\{\eta_1, \varepsilon\}$ and any $X \in \mathcal{D}$ with $d(X_0, X) < \delta$. From the alternate definition (3.2) of $d(X_0, X)$, we know there exists $\lambda_\delta \in \Lambda$ such that

$$\sup_t |\lambda_\delta t - t| < \delta \quad (3.10)$$

and

$$\sup_t |X_0(\lambda_\delta t) - X(t)| < \delta. \quad (3.11)$$

If we let $\psi_j(X)$ denote the j th component of $\psi(X)$ so that

$$\psi_j(X) \equiv \sqrt{k} \left\{ X\left(\frac{j}{k}\right) - X\left(\frac{j-1}{k}\right) \right\} \text{ for } j = 1, \dots, k, \quad (3.12)$$

then we have

$$\|\psi(X_0) - \psi(X)\|_2 = \sqrt{\sum_{i=1}^k (\psi_i(X_0) - \psi_i(X))^2}. \quad (3.13)$$

For $j = 1, \dots, k$, the j th term of under the radical in (3.13) is given by

$$\begin{aligned} & \left[\sqrt{k} \left\{ X_0\left(\frac{j}{k}\right) - X_0\left(\frac{j-1}{k}\right) \right\} - \sqrt{k} \left\{ X\left(\frac{j}{k}\right) - X\left(\frac{j-1}{k}\right) \right\} \right]^2 \\ &= k \left[\left\{ X_0\left(\frac{j}{k}\right) - X\left(\frac{j}{k}\right) \right\} - \left\{ X_0\left(\frac{j-1}{k}\right) - X\left(\frac{j-1}{k}\right) \right\} \right]^2 \\ &= k \left[d_j^2 + d_{j-1}^2 - 2d_j d_{j-1} \right], \end{aligned} \quad (3.14)$$

where

$$d_j \equiv X_0\left(\frac{j}{k}\right) - X\left(\frac{j}{k}\right) \quad \text{for } j = 0, \dots, k. \quad (3.15)$$

Now

$$\left|X_0\left(\frac{j}{k}\right) - X\left(\frac{j}{k}\right)\right| \leq \left|X_0\left(\frac{j}{k}\right) - X_0\left(\lambda_\delta \frac{j}{k}\right)\right| + \left|X_0\left(\lambda_\delta \frac{j}{k}\right) - X\left(\frac{j}{k}\right)\right|. \quad (3.16)$$

By (3.10)

$$\left|\frac{j}{k} - \lambda_\delta \frac{j}{k}\right| \leq \sup_t |t - \lambda_\delta t| < \delta \leq \eta_1 \quad \text{for } j = 0, 1, \dots, k; \quad (3.17)$$

and since $X_0 \in \mathcal{K}$, display (3.17) implies that

$$\left|X_0\left(\frac{j}{k}\right) - X_0\left(\lambda_\delta \frac{j}{k}\right)\right| < \varepsilon \quad \text{for } j = 0, 1, \dots, k \quad (3.18)$$

from (3.8) and (3.9). Also, by (3.11) we have

$$\left|X_0\left(\lambda_\delta \frac{j}{k}\right) - X\left(\frac{j}{k}\right)\right| < \delta \leq \varepsilon, \quad \text{for } j = 0, 1, \dots, k. \quad (3.19)$$

From (3.16), (3.18), and (3.19) it follows that

$$|d_j| \leq 2\varepsilon \quad \text{for } j = 0, 1, \dots, k, \quad (3.20)$$

and

$$\left|k \left[d_j^2 + d_{j-1}^2 - 2d_j d_{j-1}\right]\right| \leq k \left[4\varepsilon^2 + 4\varepsilon^2 + 8\varepsilon^2\right] = 16k\varepsilon^2, \quad (3.21)$$

which, from (3.13) and (3.14), implies that

$$\|\psi(X_0) - \psi(X)\|_2 \leq f(X_0, k, \varepsilon) \equiv \sqrt{16k^2\varepsilon^2} = 4k\varepsilon \quad (3.22)$$

We have shown for each $\varepsilon > 0$ and each $X_0 \in \mathcal{D}_k$, there exists δ such that $X \in \mathcal{D}$ and $d(X_0, X) \leq \delta$ implies that $|\psi(X_0) - \psi(X)| < 4k\varepsilon$. Therefore $\psi(\cdot)$ is continuous at every $X_0 \in \mathcal{K}$.

Proof of 2. From the definition of ψ ,

$$\begin{aligned} \psi(B(\cdot, \omega)) = \sqrt{k} \left[B\left(\frac{1}{k}, \omega\right) - B(0, \omega), B\left(\frac{2}{k}, \omega\right) - B\left(\frac{1}{k}, \omega\right), \dots, \right. \\ \left. B(1, \omega) - B\left(\frac{k-1}{k}, \omega\right) \right]. \end{aligned} \quad (3.23)$$

The k components within the brackets on the right-hand side of (3.23) are the increments of Brownian motion. From the characteristics of Brownian motion, we know that the increment $B(t_2, \omega) - B(t_1, \omega)$ corresponding to the time interval $[t_1, t_2]$ is normally distributed with mean zero and variance $t_2 - t_1$; moreover increments corresponding to nonoverlapping time intervals are independent. (See, for example, p. 274 of Karlin 1966.) Thus, we have

$$\psi(B(\cdot, \omega)) = \sqrt{k}[b_1(\omega), b_2(\omega), \dots, b_k(\omega)], \quad (3.24)$$

where each $b_i(\omega) \sim N(0, 1/k)$ and the $\{b_i(\omega)\}$ are independent. Therefore $\psi(B(\cdot, \omega))$ is a k -dimensional vector whose components are jointly multivariate normal, i.e. $\psi(B(\cdot, \omega)) \sim N_k(\mathbf{0}_k, \mathbf{I}_k)$.

Proof of 3. By Theorem 20.1 of Billingsley (1968),

$$U_n(\cdot, \omega) \xrightarrow[n \rightarrow \infty]{D} B(\cdot, \omega). \quad (3.25)$$

Since $\psi(\cdot)$ is continuous at every point in $\mathcal{K}$ and the Brownian motion process $\{B(t, \omega) : 0 \leq t \leq 1\}$ has a continuous sample path for *each* sample point ω in the underlying probability space (see, for example, p. 64 of Billingsley 1968), it follows that

$$\Pr\{B(\cdot, \omega) \in \mathcal{D} - \mathcal{K}\} = 0. \quad (3.26)$$

Equations (3.25), (3.26) and (3.6) together with the continuous mapping theorem (Corollary 1 of Theorem 5.1, Billingsley 1968) yield the desired result (3.7).

Finally, we show that $\psi(U_n(\cdot, \omega))$ is the vector of standardized and centered batch means corresponding to the process $\{\xi_n\}$ when $n = km$ and the batch size m is a positive integer. Expressing $\psi(U_n(\cdot, \omega))$ in terms of centered partial sums of the form

$$S_\ell \equiv \begin{cases} \sum_{i=1}^{\ell} (\xi_i - \mu), & \text{if } \ell \geq 1, \\ 0 & \text{if } \ell = 0, \end{cases} \quad (3.27)$$

and taking $\ell = \lfloor nj/k \rfloor$ we have

$$\psi(U_n(t, \omega)) = \sqrt{k} \left[\frac{1}{\sigma\sqrt{n}} S_{\lfloor n\frac{1}{k} \rfloor} - \frac{1}{\sigma\sqrt{n}} S_0, \frac{1}{\sigma\sqrt{n}} S_{\lfloor n\frac{2}{k} \rfloor} - \frac{1}{\sigma\sqrt{n}} S_{\lfloor n\frac{1}{k} \rfloor}, \dots, \right. \\ \left. \frac{1}{\sigma\sqrt{n}} S_{\lfloor n\frac{k}{k} \rfloor} - \frac{1}{\sigma\sqrt{n}} S_{\lfloor n\frac{k-1}{k} \rfloor} \right]$$

$$\begin{aligned}
&= \sqrt{k} \left[\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^m (\xi_i - \mu), \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{2m} (\xi_i - \mu) - \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^m (\xi_i - \mu), \dots, \right. \\
&\quad \left. \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{km} (\xi_i - \mu) - \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{(k-1)m} (\xi_i - \mu) \right] \\
&= \left[\frac{1}{\sigma\sqrt{m}} \left(\left(\sum_{i=1}^m \xi_i \right) - m\mu \right), \frac{1}{\sigma\sqrt{m}} \left(\left(\sum_{i=m+1}^{2m} \xi_i \right) - 2m\mu + m\mu \right), \dots, \right. \\
&\quad \left. \frac{1}{\sigma\sqrt{m}} \left(\left(\sum_{i=(k-1)m+1}^{km} \xi_i \right) - km\mu + (k-1)m\mu \right) \right] \\
&= \frac{1}{\sigma\sqrt{m}} [mY_1(m) - m\mu, mY_2(m) - m\mu, \dots, mY_k(m) - m\mu] \\
&= \left[\frac{\sqrt{m}(Y_1(m) - \mu)}{\sigma}, \frac{\sqrt{m}(Y_2(m) - \mu)}{\sigma}, \dots, \frac{\sqrt{m}(Y_k(m) - \mu)}{\sigma} \right]. \quad (3.28)
\end{aligned}$$

Thus it follows that

$$\sqrt{m} [Y_1(m) - \mu, Y_2(m) - \mu, \dots, Y_k(m) - \mu] \xrightarrow[m \rightarrow \infty]{D} N_k(\mathbf{0}_k, \sigma^2 \mathbf{I}_k), \quad (3.29)$$

which is the desired result (3.7). $\square$

3.2 Moment Analysis of Components of Batch Means t -Ratio

In the case of i.i.d. normal random variables, their sample mean is normally distributed and their sample variance is distributed as a scaled chi-square random variable; moreover, these statistics are independent. It follows from Theorem 3.1 that in a large class of steady-state simulation output processes, the numerator and denominator of the t -ratio used in the NOBM method achieve the required properties asymptotically.

We now consider, separately, the convergence properties of the numerator and denominator of the t -ratio used in the NOBM method. This statistic can be written

as

$$t = \frac{\bar{Y} - \mu}{\sqrt{S_{n,k}^2/k}} = \frac{\frac{\bar{Y} - \mu}{\sqrt{\text{Var}[\bar{X}(n)]}} \sqrt{\frac{k \text{Var}[\bar{X}(n)]}{\text{Var}[\bar{X}(m)]}}}{\sqrt{\frac{S_{n,k}^2}{\text{Var}[\bar{X}(m)]}}}, \quad (3.30)$$

where, using simpler notation, $\bar{X}(m) = Y(m_n)$ (the mean of a batch of size $m = m_n$), $\bar{X}(n) = \bar{Y}(n, k)$ ($= \bar{Y}$, the grand mean of $n = km$ observations organized into k nonoverlapping batches each of size m), and $S_{n,k}^2$ (the sample variance of the k batch means) are respectively defined by (1.8), (1.9), and (1.20). Throughout the remainder of this dissertation, we reserve the symbol t to denote the NOBM ratio (3.30).

The NOBM method requires that the batch size m be large enough so that the numerator of (3.30) should be approximately distributed as a $N(0, 1)$ random variable, i.e.,

$$\frac{\bar{Y} - \mu}{\sqrt{\text{Var}[\bar{X}(n)]}} \sqrt{\frac{k \text{Var}[\bar{X}(n)]}{\text{Var}[\bar{X}(m)]}} \sim N(0, 1); \quad (3.31)$$

and the denominator of (3.30) should be approximately distributed as a scaled chi random variable (that is, the square root of chi-square random variable divided by its degrees of freedom),

$$\sqrt{\frac{S_{n,k}^2}{\text{Var}[\bar{X}(m)]}} \sim \sqrt{\frac{\chi_{k-1}^2}{(k-1)}}. \quad (3.32)$$

As surrogates for these distributional requirements, we will impose conditions on the first two moments of the numerator and squared denominator of the NOBM t -ratio. We require m large enough so that

$$\text{Var} \left[\frac{\bar{Y} - \mu}{\sqrt{\text{Var}[\bar{X}(n)]}} \sqrt{\frac{k \text{Var}[\bar{X}(n)]}{\text{Var}[\bar{X}(m)]}} \right] \cong 1, \quad (3.33)$$

or, equivalently,

$$\frac{k \text{Var}[\bar{X}(n)]}{\text{Var}[\bar{X}(m)]} \cong 1. \quad (3.34)$$

The moment conditions for the square of the denominator are that m be large enough to ensure

$$\mathbb{E} \left[\frac{S_{n,k}^2}{\text{Var}[\bar{X}(m)]} \right] \cong 1 \quad (3.35)$$

and

$$\frac{(k-1) \text{Var}[S_{n,k}^2]}{2\text{Var}^2[\bar{X}(m)]} \cong 1. \quad (3.36)$$

From (3.35) and (3.36), it follows that m should be large enough to ensure that the coefficient of variation of $S_{n,k}^2$ should satisfy

$$\frac{\sqrt{\text{Var}[S_{n,k}^2]}}{\mathbb{E}[S_{n,k}^2]} \cong \sqrt{2/(k-1)}. \quad (3.37)$$

There are processes for which the lag q covariances, denoted $\gamma(q) \equiv \text{Cov}(X_i, X_{i+q})$ for $q = 0, \pm 1, \pm 2, \dots$, may be computed analytically. For these processes, using

$$\text{Var}[\bar{X}(m)] = \frac{1}{m} \left[\gamma(0) + 2 \sum_{q=1}^{m-1} \left(1 - \frac{q}{m} \right) \gamma(q) \right], \quad (3.38)$$

we may therefore calculate $\text{Var}[\bar{X}(m)]$, $\text{Var}[\bar{X}(n)]$ and the ratio on the left-hand side of condition (3.34).

We also derive an expression for $\mathbb{E}[S_{n,k}^2]$ in terms of the covariances at various lags for the original (unbatched) process.

$$\begin{aligned} \mathbb{E}[S_{n,k}^2] &= \mathbb{E} \left[\frac{1}{k-1} \sum_{i=1}^k (Y_i(m) - \bar{X}(n))^2 \right] = \frac{k}{k-1} [\text{Var}[Y_i(m)] - \text{Var}[\bar{X}(n)]] \\ &= \frac{k}{k-1} \left\{ \left[\frac{1}{m} - \frac{1}{n} \right] \gamma(0) + \frac{2}{m^2} \sum_{i=1}^{m-1} (m-i) \gamma(i) - \frac{2}{n^2} \sum_{i=1}^{n-1} (n-i) \gamma(i) \right\}. \end{aligned} \quad (3.39)$$

In order to obtain an expression for the $\text{Var}[S_{n,k}^2]$, we make use of the identity

$$\text{Var}[S_{n,k}^2] = \mathbb{E}[(S_{n,k}^2)^2] - [\mathbb{E}(S_{n,k}^2)]^2. \quad (3.40)$$

We can evaluate the second term of the right-hand side of (3.40) from (3.39) derived above. Now we give an expression for the first term of the right-hand side of (3.40),

$$\mathbb{E}[(S_{n,k}^2)^2] = \mathbb{E} \left\{ \left[\frac{1}{k-1} \sum_{i=1}^k (Y_i - \bar{Y})^2 \right]^2 \right\}$$

$$\begin{aligned}
&= \left(\frac{1}{k-1}\right)^2 \left\{ \mathbb{E} \left[\sum_{i=1}^k Y_i^2 \right]^2 - 2k \mathbb{E} \left[\bar{Y}^2 \sum_{i=1}^k Y_i^2 \right] + k^2 \mathbb{E} [\bar{Y}^4] \right\} \\
&= \left(\frac{1}{k-1}\right)^2 \left\{ \left[\sum_{i=1}^k \mathbb{E} Y_i^4 + 2 \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k \mathbb{E} Y_j^2 Y_\ell^2 \right] \right. \\
&\quad - \frac{2}{k} \left[\sum_{i=1}^k \mathbb{E} Y_i^4 + 2 \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k \mathbb{E} Y_j^2 Y_\ell^2 + 2 \sum_{i=1}^k \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k \mathbb{E} Y_i^2 Y_j Y_\ell \right] \\
&\quad + \frac{1}{k^2} \left[\sum_{i=1}^k \mathbb{E} Y_i^4 + 4 \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k \mathbb{E} Y_i^3 Y_j + 6 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \mathbb{E} Y_i^2 Y_j^2 \right. \\
&\quad \left. \left. + 12 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sum_{\substack{\ell=1 \\ \ell \neq i,j}}^k \mathbb{E} Y_i Y_j Y_\ell^2 + 4 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sum_{\substack{\ell=1 \\ \ell \neq i,j}}^{k-1} \sum_{\substack{u=\ell+1 \\ u \neq i,j}}^k \mathbb{E} Y_i Y_j Y_\ell Y_u \right] \right\}. \tag{3.41}
\end{aligned}$$

In order to evaluate the mixed moments in (3.41), we must make some assumptions about the joint distribution of the batch means $\{Y_i : i = 1, \dots, k\}$. From Theorem 3.1, we know that the batch means are asymptotically distributed as i.i.d. normal random variables, provided that the underlying process $\{X_i\}$ is stationary and ϕ -mixing with a finite SSVC. For purposes of evaluating (3.41), we will assume that the batch size m is large enough that the batch means are distributed as multivariate normal random variables, but are not necessarily independent. For simplicity we also assume each batch mean has expected value $\mu_X = 0$. If not, we can always center the batch means.

The moment generating function of such a multivariate normal distribution is given by

$$M(\boldsymbol{\theta}) = \exp \left(\frac{1}{2} \boldsymbol{\theta} \boldsymbol{\Sigma} \boldsymbol{\theta}^T \right), \tag{3.42}$$

where $\boldsymbol{\theta}$ is a k -dimensional row vector and $\boldsymbol{\Sigma}$ is the $k \times k$ covariance matrix of the k batch means, with element σ_{ij} denoting the covariance between the batch means Y_i and Y_j , i.e., $\sigma_{ij} = \text{Cov}[Y_i, Y_j]$. By taking the appropriate partial derivatives of (3.42) and evaluating them at $\boldsymbol{\theta} = \mathbf{0}_k$, we can derive expressions for the mixed moments of the form $\mathbb{E}[Y_i^a Y_j^b Y_u^c Y_\ell^d]$ in equation (3.41) in terms of the covariances of the batch

means. (See Appendix A for details.) Substituting in (3.41) values for the mixed moments given in Appendix A yields

$$\begin{aligned}
\mathbb{E} \left[\left(S_{n,k}^2 \right)^2 \right] &= \left(\frac{1}{k-1} \right)^2 \\
&\times \left\{ \left[3 \sum_{i=1}^k \sigma_{ii}^2 + 2 \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k (\sigma_{jj} \sigma_{\ell\ell} + 2\sigma_{j\ell}) \right] \right. \\
&- \frac{2}{k} \left[3 \sum_{i=1}^k \sigma_{ii}^2 + 2 \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k (\sigma_{jj} \sigma_{\ell\ell} + 2\sigma_{j\ell}) + 2 \sum_{i=1}^k \sum_{j=1}^{k-1} \sum_{\ell=j+1}^k (\sigma_{ii} \sigma_{j\ell} + 2\sigma_{ij} \sigma_{i\ell}) \right] \\
&+ \frac{1}{k^2} \left[3 \sum_{i=1}^k \sigma_{ii}^2 + 12 \sum_{i=1}^k \sum_{\substack{j=1 \\ j \neq i}}^k \sigma_{ii} \sigma_{ij} + 6 \sum_{i=1}^{k-1} \sum_{j=i+1}^k (\sigma_{ii} \sigma_{jj} + 2\sigma_{ij}^2) \right. \\
&+ 12 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sum_{\substack{\ell=1 \\ \ell \neq i,j}}^k (\sigma_{\ell\ell} \sigma_{ij} + 2\sigma_{ij} \sigma_{i\ell}) \\
&\left. \left. + 4 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sum_{\substack{\ell=1 \\ \ell \neq i,j}}^{k-1} \sum_{\substack{u=\ell+1 \\ u \neq i,j}}^k \sigma_{uj} \sigma_{i\ell} + \sigma_{iu} \sigma_{j\ell} + \sigma_{\ell u} \sigma_{ij} \right] \right\}. \quad (3.43)
\end{aligned}$$

Recall that

$$\sigma_{ij} \equiv \text{Cov}(Y_i(m), Y_j(m)) \quad (3.44)$$

denotes the covariance at lag $|i-j|$ between the batch means for batches of size m . We can write an expression for σ_{ij} in terms of the covariances in the original process (Fishman 1978a) at various lags, i.e., at lag $|i-j|$ the covariance between batch means for batches of size m is

$$\sigma_{ij} = \frac{1}{m} \sum_{q=-m+1}^{m-1} \left(1 - \frac{|q|}{m} \right) \gamma(m|i-j|+q). \quad (3.45)$$

In the next section, using (3.39), (3.40), (3.43) and (3.45), we compute $\mathbb{E}[S_{n,k}^2]$ and $\text{Var}[S_{n,k}^2]$ for varying batch sizes in processes for which analytic expressions are available for the covariance function of the original process. This allows us to investigate numerically how increasing batch size m affects the first two moments of the square of the denominator of the NOBM t -ratio.

If the batch size m is large enough that the batch means $\{Y_i(m) : i = 1, \dots, k\}$ are i.i.d. normal, then the grand mean, $\bar{Y} = \bar{X}(n)$, and the sample variance of the batch means, $S_{n,k}^2$, are independent. Again, using the assumption that the batch means are jointly distributed as multivariate normal random variables each with expected value zero, we can evaluate the covariance of $\bar{Y}$ and $S_{n,k}^2$:

$$\text{Cov}(\bar{Y}, S_{n,k}^2) = \text{E}[\bar{Y} S_{n,k}^2] - \text{E}[\bar{Y}] \text{E}[S_{n,k}^2]. \quad (3.46)$$

Obviously, under our assumptions, the second term on the right-hand side of (3.46) is zero; therefore we have

$$\begin{aligned} \text{Cov}(\bar{Y}, S_{n,k}^2) &= \text{E}[\bar{Y} S_{n,k}^2] \\ &= \frac{1}{k-1} \text{E} \left[\bar{Y} \sum_{i=1}^k (Y_i - \bar{Y})^2 \right] \\ &= \frac{1}{k-1} \text{E} \left[\bar{Y} \sum_{i=1}^k Y_i^2 - k \bar{Y}^3 \right] \\ &= \frac{1}{k-1} \text{E} \left[\frac{1}{k} \left\{ \sum_{i=1}^k Y_i^3 + \sum_{i=1}^k \sum_{j=1; j \neq i}^k Y_i^2 Y_j \right\} \right. \\ &\quad \left. - \frac{1}{k^2} \left\{ \sum_{i=1}^k Y_i^3 + 3 \sum_{i=1}^k \sum_{j=1; j \neq i}^k Y_i^2 Y_j + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^k \sum_{\substack{\ell=1 \\ \ell \neq i, j}}^k Y_i Y_j Y_\ell \right\} \right] \\ &= 0. \end{aligned} \quad (3.47)$$

The mixed moments in (3.47) are all zero (see Appendix A); therefore, if the batch means are distributed as multivariate normal random variables, then $\bar{Y}$ and $S_{n,k}^2$ have zero correlation. While this does not necessarily imply that these statistics are independent, it provides some basis for the standard simulation analysis method of nonoverlapping batch means.

3.3 Behavior of Batch Means Method in Selected Cases

We considered several processes for which analytic expressions are available for the covariances at various lags. Four of these are cases that Song and Schmeiser (1995)

studied in investigating their estimates of optimal mean-squared-error batch sizes.

3.3.1 Discrete Time Markov Chains

The first type of process is a real valued function on the state space of an irreducible aperiodic Discrete Time Markov Chain (DTMC). We can write the covariance expressions for DTMCs following the approach used by Hazen and Pritsker (1980) for continuous time Markov chains.

The following definitions are used in our analysis:

- The response function

$$\eta(\cdot) : \{0, 1, \dots, d\} \rightarrow \Re \quad (3.48)$$

is a real-valued function on the state space $\{0, 1, \dots, d\}$ of a $(d+1)$ -state DTMC $\{X_i : i = 0, 1, \dots\}$;

- The vector of stationary probabilities is

$$\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_d); \quad (3.49)$$

i.e., π_ℓ is the steady state probability that the chain $\{X_i\}$ is in state ℓ for $\ell = 0, 1, \dots, d$;

- The response vector

$$\mathbf{h} = (\eta(0), \eta(1), \dots, \eta(d))^T \quad (3.50)$$

is the $(d+1)$ -dimensional row vector of real values associated with the $d+1$ states;

- The $(d+1) \times (d+1)$ matrix

$$\mathbf{\Pi} = \begin{bmatrix} \boldsymbol{\pi} \\ \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{bmatrix} \quad (3.51)$$

has the vector of stationary probabilities in each row; and

- The $(d+1) \times (d+1)$ matrix

$$\mathbf{\Delta} = \text{diag}(\pi_0, \pi_1, \dots, \pi_d) \quad (3.52)$$

has the vector of stationary probabilities on the diagonal and zeros elsewhere.

If the initial state X_0 of the Markov chain is sampled according to its steady state distribution, then the associated process

$$\{N_i = \eta(X_i) : i = 0, 1, \dots\} \quad (3.53)$$

is covariance stationary, and the covariance at lag q may be written

$$\text{cov}(N_\ell, N_{\ell+q}) = \mathbf{h}^T \mathbf{\Delta} (\mathbf{P}^q - \mathbf{\Pi}) \mathbf{h} \text{ for } q = 0, 1, 2, \dots, \quad (3.54)$$

where $\mathbf{P}^0 \equiv \mathbf{I}_{d+1}$ (Hazen and Pritsker 1980). Now if the one-step transition matrix $\mathbf{P}$ has distinct real eigenvalues, then the q th power of $\mathbf{P}$ may be calculated as

$$\mathbf{P}^q = \mathbf{V} \mathbf{\Lambda}^q \mathbf{V}^{-1}, \quad (3.55)$$

where

$$\mathbf{V} = [\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_d]$$

is the $(d+1) \times (d+1)$ matrix of right eigenvectors of $\mathbf{P}$ so that

$$\mathbf{P} \mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{for } i = 0, 1, \dots, d,$$

and

$$\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_d)$$

is the diagonal matrix of associated eigenvalues (Hadley 1961, p. 251).

We can also write a matrix-analytic expression for the SSVC of the process (3.53). Since the SSVC, σ^2 , is the sum of the covariances at all lags, we have

$$\sigma^2 = \mathbf{h}^T \mathbf{\Delta} \left\{ \mathbf{I}_{d+1} - \mathbf{\Pi} + 2 \sum_{\ell=1}^{\infty} (\mathbf{P}^\ell - \mathbf{\Pi}) \mathbf{h} \right\}. \quad (3.56)$$

The following theorem provides a simplified computational formula for σ^2 .

Theorem 3.2 *If the $(d+1)$ -state DTMC $\{X_i : i = 0, 1, \dots\}$ has one step transition matrix $\mathbf{P}$ and associated $(d+1) \times (d+1)$ matrix $\mathbf{\Pi}$ defined by (3.51), then the SSVC of the associated process (3.53) is given by*

$$\begin{aligned}\sigma^2 &\equiv \lim_{q \rightarrow \infty} \cdot \text{Var} \left[\frac{1}{q} \sum_{i=1}^q N_i \right] \\ &= \mathbf{h}^T \mathbf{\Delta} \left\{ \mathbf{I}_{d+1} - \mathbf{\Pi} + 2(\mathbf{I}_{d+1} - \mathbf{P} + \mathbf{\Pi})^{-1} (\mathbf{P} - \mathbf{\Pi}) \right\} \mathbf{h},\end{aligned}\quad (3.57)$$

where $\mathbf{h}$ and $\mathbf{\Delta}$ are defined in (3.50) and (3.52), respectively.

Proof. It is easy to prove by induction that

$$(\mathbf{P} - \mathbf{\Pi})^\ell = \mathbf{P}^\ell - \mathbf{\Pi} \quad \text{for } \ell = 1, 2, \dots. \quad (3.58)$$

It follows that

$$\lim_{\ell \rightarrow \infty} (\mathbf{P} - \mathbf{\Pi})^\ell = \lim_{\ell \rightarrow \infty} \mathbf{P}^\ell - \mathbf{\Pi} = \mathbf{\Pi} - \mathbf{\Pi} = \mathbf{0}; \quad (3.59)$$

and thus Theorem 1.11.1 of Kemeny and Snell (1960) ensures that $\mathbf{I}_{d+1} - (\mathbf{P} - \mathbf{\Pi}) = \mathbf{I}_{d+1} - \mathbf{P} + \mathbf{\Pi}$ has inverse

$$(\mathbf{I}_{d+1} - \mathbf{P} + \mathbf{\Pi})^{-1} = \sum_{\ell=0}^{\infty} (\mathbf{P} - \mathbf{\Pi})^\ell. \quad (3.60)$$

In view of (3.58), we see that

$$\sum_{\ell=1}^{+\infty} (\mathbf{P}^\ell - \mathbf{\Pi}) = \sum_{\ell=1}^{\infty} (\mathbf{P} - \mathbf{\Pi})^\ell. \quad (3.61)$$

Letting

$$\mathbf{G} = \sum_{\ell=1}^{\infty} (\mathbf{P} - \mathbf{\Pi})^\ell \quad (3.62)$$

yields

$$\begin{aligned}(\mathbf{P} - \mathbf{\Pi}) \mathbf{G} &= \sum_{\ell=2}^{\infty} (\mathbf{P} - \mathbf{\Pi})^\ell \\ &= \mathbf{G} - (\mathbf{P} - \mathbf{\Pi}).\end{aligned}\quad (3.63)$$

Using (3.60) to solve (3.63) for $\mathbf{G}$, we have

$$\mathbf{G} = (\mathbf{I}_{d+1} - \mathbf{P} + \mathbf{\Pi})^{-1} (\mathbf{P} - \mathbf{\Pi}). \quad (3.64)$$

Substituting (3.64) into (3.56) yields the desired result. $\square$

For our analysis, we chose 3 DTMCs, each having a different correlation structure. The first is the cost function on a simple 2-state DTMC with probability transition matrix

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.25 & 0.75 \\ 0.75 & 0.25 \end{pmatrix} \end{matrix}. \quad (3.65)$$

This chain demonstrates an alternating positive and negative correlation structure, as illustrated in Figure 3.1. The cost vector associated with the states is

$$\mathbf{h} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 5 \\ 10 \end{matrix} \end{matrix}. \quad (3.66)$$

The second case is a cost function on the state space of a 2-state DTMC with high positive correlation. This case has the same cost vector as the first, but its probability transition matrix is

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix} \end{matrix}. \quad (3.67)$$

The third case is the cost function on the state space of an 8-state DTMC with a correlation structure that resembles a damped sine wave. The chain was designed to loosely represent some sort of cyclical behavior, where the chain has high probability of moving from one state to another and then starting the cycle over. The probability transition matrix for this chain is

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{pmatrix} 0.028 & 0.800 & 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.032 \\ 0.028 & 0.028 & 0.800 & 0.028 & 0.028 & 0.028 & 0.028 & 0.032 \\ 0.028 & 0.028 & 0.028 & 0.800 & 0.028 & 0.028 & 0.028 & 0.032 \\ 0.028 & 0.028 & 0.028 & 0.028 & 0.800 & 0.028 & 0.028 & 0.032 \\ 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.800 & 0.028 & 0.032 \\ 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.800 & 0.032 \\ 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.032 & 0.800 \\ 0.800 & 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.028 & 0.032 \end{pmatrix} \end{matrix}. \quad (3.68)$$

The revenue vector associated with the 8 states is

$$\mathbf{h} = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \begin{pmatrix} 0 & 5 & 10 & 15 & 20 & 25 & 30 & 35 \end{pmatrix}. \end{matrix} \quad (3.69)$$

The probability transition matrices for these three chains have distinct real eigenvalues, so we were able to use the MATLAB software (Hanselman and Littlefield 1997) to obtain the eigenvalues and their corresponding eigenvectors. We could then compute the covariances of the original processes as well as the covariances for the batch means processes using the matrix-analytic equations (3.54), (3.55), and (3.57).

Figures 3.1, 3.8 and 3.15 show the correlation structure of the original processes, while Figures 3.2, 3.9 and 3.16 show the correlation of the batch means process. Figures 3.4, 3.11 and 3.18 show the convergence to the moment condition (3.34). Figures 3.5, 3.12 and 3.19 display convergence to condition (3.35) for the three DTMCs. Figures 3.6, 3.13 and 3.20 show convergence to condition (3.36). Finally, Figures 3.7, 3.14 and 3.21 exhibit the convergence to condition (3.37). It is obvious from these figures that the desired values are achieved at smaller batch sizes for processes having some sort of alternating correlation structure than for processes with a slowly decaying correlation structure.

Figure 3.1: Lag- q Correlation $\text{Corr}(N_i, N_{i+q})$ of Observations of Cost Function $N_i = \eta(X_i)$ on 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66). The SSVC $\sigma^2 = 2.083$ and $\sigma_N^2 \equiv \text{Var}[N_i] = 6.25$.

Figure 3.2: Lag 1 Correlation of the Batch Means with Batches of Size m for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66)

Figure 3.3: $\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for the Cost Function on 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66). The SSVC $\sigma^2 = 2.083$.

Figure 3.4: Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m .

Figure 3.5: Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\overline{X}(m)]$ for 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m .

Figure 3.6: Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\overline{X}(m)]$ for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m .

Figure 3.7: Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/\mathbb{E}[S_{n,k}^2]$ for 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) for $k = 30$ Batches of Size m .

Figure 3.8: Lag- q Correlation $\text{Corr}(N_i, N_{i+q})$ of Observations of Cost Function $N_i = \eta(X_i)$ on 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67). The SSVC $\sigma^2 = 618.75$ and $\sigma_N^2 \equiv \text{Var}[N_i] = 6.25$.

Figure 3.9: Lag 1 Correlation of the Batch Means with Batches of Size m for the 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67).

Figure 3.10: $\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for the Cost Function on 2 State DTMC with High Positive Correlation Structure as Defined by (3.66) and (3.67). The SSVC $\sigma^2 = 618.75$.

Figure 3.11: Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for the 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m .

Figure 3.12: Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m .

Figure 3.13: Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the 2-State DTMC with High Positive Correlation as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m .

Figure 3.14: Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/\text{E}[S_{n,k}^2]$ for 2-State DTMC with High Positive Correlation Structure as Defined by (3.66) and (3.67) for $k = 30$ Batches of Size m .

Figure 3.15: Lag- q Correlation $\text{Corr}(N_i, N_{i+q})$ of Observations of Cost Function $N_i = \eta(X_i)$ on 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69). The SSVC $\sigma^2 = 77.5$ and $\sigma_N^2 \equiv \text{Var}[N_i] = 132.2$.

Figure 3.16: Lag 1 Correlation of the Batch Means with Batches of Size m for the 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69).

Figure 3.17: $\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for the Cost Function on 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69). The SSVC $\sigma^2 = 77.5$.

Figure 3.18: Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for the 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m .

Figure 3.19: Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m .

Figure 3.20: Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m .

Figure 3.21: Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/\text{E}[S_{n,k}^2]$ for 8-State DTMC with Damped Sinusoidal Correlation as Defined by (3.68) and (3.69) for $k = 30$ Batches of Size m .

3.3.2 M/M/1 Queue Waiting Times

The second type of process we examined was the sequence of customer waiting times $\{X_i : i = 1, 2, \dots\}$ in the M/M/1 queue. Letting $\tau = \lambda/\mu$ denote the server utilization (or traffic intensity), where λ is the arrival rate and μ is the service rate, the mean queue waiting time is

$$\mu_X = \frac{\lambda}{\mu(\mu - \lambda)}; \quad (3.70)$$

and the variance of the waiting time is

$$\sigma_X^2 = \frac{\lambda(2\mu - \lambda)}{\mu^2(\mu - \lambda)^2} \quad (3.71)$$

The marginal distribution of each waiting time in steady-state operation is mixed with c.d.f.

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - \tau, & x = 0 \\ 1 - \tau e^{-(\mu - \lambda)x}, & x > 0. \end{cases} \quad (3.72)$$

Thus each waiting time is zero with probability $1 - \tau$; and with probability τ , it is exponential with mean $(\mu - \lambda)^{-1}$. Again the correlations for this process are available analytically. From Song and Schmeiser (1995) the lag- q correlation, $\gamma(q)$, for the wait time process in the M/M/1 queue is given by

$$\gamma(q) = \frac{1 - \tau^2}{2\pi\lambda^2} \int_0^{4\tau/(1+\tau)^2} t^{q+3/2} \left[\left(4\tau/(1+\tau)^2 \right) - t \right]^{1/2} (1-t)^{-3} dt, \quad (3.73)$$

where $\tau < 1$. The SSVC can be written (Wilson 1979) as

$$\sigma^2 = \frac{\tau^3(\tau^3 - 4\tau^2 + 5\tau + 2)}{\lambda^2(1 - \tau)^4}. \quad (3.74)$$

We used MATLAB software (Hanselman and Littlefield 1997) to evaluate numerically the integral in (3.73) for utilization $\tau = 0.9$. As can be seen from Figures 3.22 through 3.28, the rate of convergence to the moment conditions is extremely slow.

Figure 3.22: Lag- q Correlation $\text{Corr}(X_i, X_{i+q})$ of Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$. The SSVC $\sigma^2 = 37440$ and $\sigma_X^2 = 99$.

Figure 3.23: Lag 1 Correlation of the Batch Means with Batches of Size m for the Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$.

Figure 3.24: $\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n of Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$. The SSVC $\sigma^2 = 37440$.

Figure 3.25: Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.26: Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for Wait Times in the M/M/1 Queue with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.27: Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for the M/M/1 Queue Wait Time Process where Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.28: Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/E[S_{n,k}^2]$ for the M/M/1 Queue Wait Time Process where Utilization $\tau = 0.9$ for $k = 30$ Batches of Size m .

3.3.3 Autoregressive Process

We also investigated the convergence properties of the AR(1) process, which is given by

$$X_i = \mu_X + \varphi(X_{i-1} - \mu_X) + \epsilon_i \quad \text{for } i = 1, 2, \dots, \quad (3.75)$$

where $|\varphi| < 1$ and the ϵ_i 's are i.i.d. normal with mean zero and variance σ_ϵ^2 so that the X_i 's have mean μ_X and variance $\sigma_X^2 = \sigma_\epsilon^2 / (1 - \varphi^2)$. To ensure that (3.75) defines a stationary process, we take $X_0 \sim N(\mu_X, \sigma_X^2)$. The lag- q correlation for the AR(1) process is

$$\rho_q = \varphi^{|q|} \quad \text{for } q = 0, \pm 1, \pm 2, \dots; \quad (3.76)$$

and the SSVC is given by

$$\sigma^2 = \sigma_X^2 \left(\frac{1 + \varphi}{1 - \varphi} \right) = \frac{\sigma_\epsilon^2}{(1 - \varphi)^2} \quad (3.77)$$

(Box and Jenkins 1976, p. 58). Figures 3.29 through 3.35 show the convergence to the moment conditions.

Figure 3.29: Lag- q Correlation $\text{Corr}(X_i, X_{i+q})$ of Observations of an AR(1) Process (3.75) with $\varphi = 0.9$. The SSVC $\sigma^2 = 100$ and $\sigma_X^2 = 5.26$.

Figure 3.30: Lag 1 Correlation of the Batch Means with Batches of Size m for an AR(1) Process (3.75) with $\varphi = 0.9$.

Figure 3.31: $\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for an AR(1) Process (3.75) with $\varphi = 0.9$. The SSVC $\sigma^2 = 100$.

Figure 3.32: Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ and Batches Size m .

Figure 3.33: Condition (3.35) Ratio $\text{E}[S^2_{n,k}]/\text{Var}[\bar{X}(m)]$ for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.34: Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.35: Condition (3.37) on Coefficient of Variation, $\sqrt{\text{Var}[S_{n,k}^2]}/\text{E}[S_{n,k}^2]$, for an AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

3.3.4 Exponential Autoregressive Process

The last process for which we computed the analytic correlations in investigating convergence to the moment conditions was the exponential autoregressive order 1 process (EAR(1)):

$$X_i = \begin{cases} (\mu_X - 1/\lambda) + \varphi [X_{i-1} - (\mu_X - 1/\lambda)] & \text{w.p. } \varphi \\ (\mu_X - 1/\lambda) + \varphi [X_{i-1} - (\mu_X - 1/\lambda)] + \epsilon_i & \text{w.p. } 1 - \varphi, \end{cases} \quad (3.78)$$

for $i = 1, 2, \dots$, $0 < \varphi < 1$ and where the ϵ_i 's are i.i.d. exponential with rate $\lambda = 1/\sigma_X$ so that the X_i 's have mean μ_X and variance

$$\sigma_X^2 = \frac{1}{\lambda^2}. \quad (3.79)$$

As in the AR(1) process, the autocorrelation function has the form

$$\rho_q = \varphi^{|q|} \text{ for } q = 0, \pm 1, \pm 2, \dots; \quad (3.80)$$

and the SSVC is given by

$$\sigma^2 = \sigma_X^2 \left(\frac{1 + \varphi}{1 - \varphi} \right) = \frac{1}{\lambda^2} \left(\frac{1 + \varphi}{1 - \varphi} \right) \quad (3.81)$$

(Song and Schmeiser 1995). In the EAR(1) process the marginal distribution of each X_i is exponential with mean $1/\lambda$. This means larger batch sizes should be required to satisfy the assumption that the batch means are normally distributed as compared to, for instance, the AR(1) process where the marginal distribution of each X_i is normal. In the case selected, we chose $\varphi = 0.9$ and $\lambda = 0.5$.

Figure 3.36: Lag- q Correlation $\text{Corr}(X_i, X_{i+q})$ of Observations of an EAR(1) Process (3.78) with $\varphi = 0.9$. The SSVC $\sigma^2 = 76$ and $\sigma_X^2 = 4.0$.

Figure 3.37: Lag 1 Correlation of the Batch Means with Batches of Size m for an EAR(1) Process (3.78) with $\varphi = 0.9$.

Figure 3.38: $\sigma^2(n) = n\text{Var}[\bar{X}(n)]$ for Samples of Size n for an EAR(1) Process (3.78) with $\varphi = 0.9$. The SSVC $\sigma^2 = 76$.

Figure 3.39: Condition (3.34) on Ratio $k\text{Var}[\bar{X}(n)]/\text{Var}[\bar{X}(m)]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.40: Condition (3.35) on Ratio $E[S_{n,k}^2]/\text{Var}[\bar{X}(m)]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.41: Condition (3.36) on Ratio $\text{Var}[S_{n,k}^2]/\text{Var}^2[\bar{X}(m)]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

Figure 3.42: Condition (3.37) on Coefficient of Variation $\sqrt{\text{Var}[S_{n,k}^2]}/\text{E}[S_{n,k}^2]$ for an EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size m .

3.4 Monte Carlo Analysis of Components of Batch Means t -Ratio for Selected Cases

After we numerically evaluated the moment conditions for various batch sizes as described in the previous section, we simulated the six selected cases to study the empirical convergence to the distributional requirements. We looked at the distributions of numerator and squared denominator of the NOBM t -ratio (3.30) separately and compared the distributions of these target statistics observed at various batch sizes to the theoretically required distributions.

The number of batches was fixed at $k = 30$, and the initial batch size was usually chosen to be $m = 8$. For each selected process, the batch size was increased until each observed frequency distribution (as depicted by a histogram and by a plot of the empirical c.d.f. based on 2,000 replications of the target statistic) was judged to approximate the corresponding theoretical distribution (as depicted by the corresponding p.d.f. and c.d.f., which are overlaid on their respective plots). It was not possible to display all the various batch sizes at which simulations were performed. In Figures 3.43 through 3.54, three batch sizes are displayed for each case. Results for the batch size of $m = 16$ are always displayed. All simulations were started by sampling the selected process from its steady-state distribution; and thus all simulation-generated data sets were realizations of stationary stochastic processes.

The “worst” case, i.e., the one requiring the largest batch size for the observed distributions to achieve adequate convergence to their theoretical counterparts, was case of the M/M/1 waiting times with utilization $\tau = 0.9$. As can be seen from Figures 3.49 and 3.50, the largest batch size simulated for this case was $m = 4096$. It should be noted that this case has the most slowly decaying correlation structure, as well as a mixed marginal distribution with an exponential tail. Both factors cause serious deviations from the assumptions of independence and normality of the batch means when the batch size is small.

In the other five cases, the largest batch size required for reasonable convergence to the desired limiting distributions was 512. This was the case of the highly correlated DTMC defined by (3.66) and (3.67). Again the slowly decaying correlation (see Figure 3.8) and the discreteness of the underlying distribution work to make convergence to the desired properties slow.

An interesting insight to be gained from these investigations was that often at

relatively small batch sizes, even though the numerator displayed a variance that was not 1, the numerator was still centered, suggesting that the dependency of the batch means was a problem that remained after the batch size was large enough to ensure the distribution has assumed the desired shape.

Figure 3.43: Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the Alternating Correlation 2-State DTMC Defined by (3.65) and (3.66) for $k = 30$ Batches of Size $m = 8$, $m = 16$ and $m = 32$.

Figure 3.44: Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the Alternating Correlation 2-State DTMC Defined by (3.65) and (3.66) for $k = 30$ Batches of Size $m = 8$, $m = 16$ and $m = 32$.

Figure 3.45: Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the High Positive Correlation 2-State DTMC Defined by (3.66) and (3.67) for $k = 30$ Batches of Size $m = 16$, $m = 256$ and $m = 512$.

Figure 3.46: Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the High Positive Correlation 2-State DTMC Defined by (3.66) and (3.67) for $k = 30$ Batches of Size $m = 16$, $m = 256$ and $m = 512$.

Figure 3.47: Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the Damped Sinusoidal Correlation 8-State DTMC Defined by (3.68) and (3.69) for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$.

Figure 3.48: Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the Damped Sinusoidal Correlation 8-State DTMC Defined by (3.68) and (3.69) for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$.

Figure 3.49: Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the M/M/1 Queue Waiting Times with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 2048$ and $m = 4096$.

Figure 3.50: Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the M/M/1 Queue Waiting Times with Utilization $\tau = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 2048$ and $m = 4096$.

Figure 3.51: Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$.

Figure 3.52: Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the AR(1) Process (3.75) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$.

Figure 3.53: Relative Frequency and Cumulative Relative Frequency of the Numerator (3.31) of the t -Statistic (3.30) Based on 2000 Replications of the EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$.

Figure 3.54: Relative Frequency and Cumulative Relative Frequency of the Squared Denominator (3.32) of the t -Statistic (3.30) Based on 2000 Replications of the EAR(1) Process (3.78) with $\varphi = 0.9$ for $k = 30$ Batches of Size $m = 16$, $m = 32$ and $m = 64$.

3.5 Conclusions

Both the theoretical and empirical convergence properties of the batch means process of the DTMCs, the AR(1) process, and the EAR(1) process lead us to suspect that many times the joint distribution of the batch means converges to an approximately multivariate normal distribution at smaller batch sizes than are required to achieve approximate independence of the batch means. The distributions of the numerator of the NOBM t -statistic (3.30) for these three types of processes are symmetric at even small batch sizes. The plots of the observed distributions of the numerator overlaid with the theoretical $N(0, 1)$ density show that at small batch sizes the observed distributions display an incorrect variance (scaling) but are nearly bell-shaped. The distributions of the squared denominator in these cases closely resemble the appropriate $\chi^2_{k-1}/(k-1)$ distribution. Of the cases we studied, only in the M/M/1 queue waiting time process did we find marked asymmetry in the distribution of the numerator of the NOBM t -statistic (3.30) together with marked deviation from the appropriate distribution of the denominator at moderate batch sizes.

In processes having strong positive dependence, a batch size that is too small (resulting in batch means with large positive correlations) causes the classical batch-means method to underestimate the variance of the grand mean $\bar{Y}$, resulting in CIs that possess inadequate coverage probability. Conversely, in processes with alternating correlation structure, the classical batch-means estimator $S^2_{n,k}/k$ may overestimate $\text{Var}[\bar{Y}]$. In either case, the resulting CI may not possess the desired coverage.

In the next chapter, we formulate an algorithm for determining an appropriate batch size m and an overall sample size n that exploits the property of the joint multivariate distribution of the batch means at relatively small batch sizes.

Chapter 4

Formulation of ASAP: Automated Simulation Analysis Procedure

4.1 An Alternate Approach to Handling Dependent Batch Means

From our review in Chapter 2 of previous approaches to implementing the method of batch means for analysis of steady-state simulation output processes, we concluded that dependence between the final batch means computed in these procedures is the main cause of failure in practical applications. Such a failure occurs when a batch-means procedure delivers a confidence interval for a steady-state expected response whose probability of covering that expectation deviates significantly from the user-specified nominal coverage probability. We concluded from the results of our theoretical and empirical analyses detailed in Chapter 3 that if the vector of batch means has a multivariate normal distribution, then the first two moments of the square of the denominator (3.32) of the classical batch means t -statistic (3.30) are close to the first two moments of a chi-square random variable with $k - 1$ degrees of freedom divided by $(k - 1)$. Additionally, although the numerator (3.31) of the t -statistic (3.30) may not display the correct variance, i.e., the variance may not be equal to one, the multivariate normality of the batch means results in a numerator that is normally distributed with expected value zero. We have also shown that the numerator and squared denominator have zero correlation when the batch means are multivariate normal. We believe, therefore, that if we have a batch size large enough so that

the batch means have a joint distribution that is approximately multivariate normal, then we may reasonably assume that the denominator of the t -statistic (3.30) possesses the required distribution and that the numerator and the denominator of the t -statistic (3.30) are independent; and if these assumptions hold, then we must make a correction to the classical batch means confidence interval (1.21) to compensate for the failure of the numerator of (3.30) to possess a variance of one. The proposed correction to (1.21) is based on an inverted Cornish-Fisher expansion (Hall 1983) for the t -statistic (3.30) in which the terms of the expansion are estimated by fitting an autoregressive–moving average time series model (Box and Jenkins 1976) to the series of final batch means. This approach should result in improved CI coverage at smaller batch sizes, even when the batch means do not appear to be independent. These considerations motivated the development of a new batch-means procedure that is described in the rest of this chapter.

4.2 Overview of ASAP

The simulation analysis procedure developed in this chapter is called ASAP (for Automated Simulation Analysis Procedure); and it requires the following user-supplied inputs:

1. a simulation-generated realization $\{x_j : j = 1, 2, \dots, n\}$ of the target output process from which the steady-state expected response μ_X is to be estimated;
2. a confidence coefficient α specifying that the desired confidence-interval coverage probability is $1 - \alpha$; and
3. an absolute or relative precision requirement specifying the final confidence-interval half-length in terms of (a) a maximum absolute half-length H^* , or (b) a maximum relative fraction r^* of the magnitude of the final grand mean $\bar{Y}$.

ASAP delivers the following outputs:

1. a nominal $100(1 - \alpha)\%$ confidence interval for μ_X having the form

$$\bar{Y} \pm H \text{ where } H \leq H^* \text{ or } H \leq r^*|\bar{Y}|, \quad (4.1)$$

provided no additional simulation-generated observations are required;

2. a new total sample size n to be supplied to the algorithm; or
3. the estimated final sample size N^* , final batch size m^* , and final batch count k^* required to deliver a valid confidence interval of the form (4.1) that satisfies the user-specified precision requirement.

If additional observations of the target process must be generated by the user's simulation model before a confidence interval with the required precision can be delivered, then ASAP must be called again with the additional data; and this cycle of simulation followed by analysis may be repeated several times before ASAP finally delivers a confidence interval.

On each iteration of ASAP, the algorithm operates as follows. The simulation outputs are divided into a fixed number of batches (namely, 96 batches); and batch means are computed. The first two batches are discarded, and the remaining 94 batch means are tested for independence. If the test for independence fails, then the batch means are tested for joint multivariate normality. If the normality test fails, then the batch size is increased by a factor of $\sqrt{2}$ and the process is repeated until one of the tests is passed.

Upon acceptance of either the hypothesis of independence or the hypothesis of joint multivariate normality of the batch means, a CI is constructed—either the usual NOBM CI (1.21) (in the case of acceptance of independence) or a corrected CI (in the case of acceptance of multivariate normality). The correction uses an inverted Cornish-Fisher expansion (Hall 1983 and Kendall, Stuart and Ord 1987) of the NOBM t -statistic whose terms are estimated by fitting an autoregressive-moving average model to the batch means process. Subsequent iterations of ASAP that are performed to satisfy the user-specified precision requirement (if there is one) do not repeat testing for independence or multivariate normality of the overall set of batch means. These subsequent iterations require additional sampling, computing the additional batch means, and reconstructing the CI, again discarding the first two batches of the overall data set (consisting of all original observations plus any additional observations required by ASAP). Successive iterations of ASAP continue until the precision requirement is met.

4.3 Testing Batch Means for Independence and Joint Normality

ASAP begins with a user-specified initial batch size $m_1 = 16$ and requires data for $k_1 = 96$ initial batches to be collected. The results of our tests and experiments with the algorithm show that ASAP performs well with this initial batch size, even for processes that are highly dependent and exhibit marked departures from normality. There were several reasons, which are detailed in the following discussion, for choosing an initial batch count of 96. While a total of $k_1 m_1 = 1536$ observations may be more than is actually needed in a few cases, such a sample size is usually easy and inexpensive to generate.

The first two batches of data are excluded from computations of batch-means statistics in an effort to overcome the initial bias problem. Let $k_1^* = k_1 - 2 = 94$ denote the number of batch means retained for confidence-interval construction. The k_1^* retained batches are tested for independence using von Neumann's ratio (2.41) of the sample mean-square successive difference to the sample variance (von Neumann 1941, Fishman 1978a). This test is described in Chapter 2. We note here that the more observations used for the test, the more powerful the test is, i.e. the more capable the test is of detecting type II errors. This is one reason for starting with 96 batches. Batch means procedures to date have used fewer initial batches and consequently run a greater risk of initially accepting the hypothesis of independence when it should be rejected.

Since our studies detailed in Chapter 3 show that the correlation of the batch means process is not always a monotone decreasing function of the batch size, we chose to use a two-sided test for the independence of the batch means with size $\alpha_{\text{ind}} = 0.20$. This is comparable to using a one-sided test of size 0.10 in the LATCH and ABATCH procedures (Fishman and Yarberr 1997). If the $k_1^* = 94$ batch means pass the independence test, then the classical batch means confidence interval (1.21) is constructed with midpoint $\bar{Y}$ and half-length

$$H = t_{1-\alpha/2, k_1^*-1} \frac{S_{n, k_1^*}}{\sqrt{k_1^*}}. \quad (4.2)$$

No correction is made to the confidence interval because presumably none is needed if the batch means are independent (Fishman 1978a, Fishman and Yarberr 1997).

If the test for independence fails, then ASAP constructs 16 vectors each consisting of four adjacent batch means. Two batch means between each set of four are ignored in an effort to obtain approximately independent 4-dimensional vectors of batch means, i.e.,

$$\underbrace{Y_3(m), Y_4(m), Y_5(m), Y_6(m)}_{\text{1st } (4 \times 1) \text{ vector}}, \underbrace{Y_7(m), Y_8(m)}_{\text{ignored}}, \underbrace{Y_9(m), Y_{10}(m), Y_{11}(m), Y_{12}(m)}_{\text{2nd } (4 \times 1) \text{ vector}}, \\ \underbrace{Y_{13}(m), Y_{14}(m)}_{\text{ignored}}, \dots, \underbrace{Y_{93}(m), Y_{94}(m), Y_{95}(m), Y_{96}(m)}_{\text{16th } (4 \times 1) \text{ vector}}. \quad (4.3)$$

We apply the Shapiro-Wilk test for multivariate normality (Malkovich and Afifi 1973, Tew and Wilson 1992) to the resulting 16 vectors, each consisting of four adjacent batch means. This is another reason for the choice of 96 initial batches. To test for r -dimensional multivariate normality, we require a random sample of size at least $\max\{2r, 16\}$ from the corresponding r -dimensional distribution (Tew and Wilson 1992). Although joint normality of these selected sets of $r = 4$ adjacent batch means is not sufficient to ensure joint normality of all $k = 96$ batch means (see exercise 15.20 on p. 504 of Kendall, Stuart and Ord 1987), the results reported in Chapter 5 of an extensive experimental evaluation of ASAP's performance strongly suggest that testing for joint normality in $r \times 1$ vectors each composed of $r = 4$ adjacent batch means yields good performance in many applications of ASAP.

For the multivariate normality test, we used the size $\alpha_{\text{mvn}} = 0.10$ based on results of experimenting with the parameter α_{mvn} . In practice, ASAP appears to perform well even where there is mild departure from multivariate normality of the batch means.

On the i th iteration of ASAP, we let k_i and m_i respectively denote the batch count and the batch size; and we take $k_1 = 96$, $m_1 = 16$ on the first iteration of the procedure. An additional iteration of ASAP will be required if the following conditions occur on iteration i :

- a) the independence test (see Section 2.2.2) yields a significant result (that is, independence is rejected) at the level of significance α_{ind} when this test is applied to the k_i batch means for batches of size m_i ; and then
- b) the multivariate Shapiro-Wilk test (Tew and Wilson 1992) yields a significant result (that is, multivariate normality is rejected) at the level of significance

α_{mvn} when this test is applied to the corresponding sample of size 16 consisting of four-dimensional random vectors formed from adjacent batch means.

Now if conditions a) and b) both occur on iteration i of ASAP, then iteration $i + 1$ will be required in which the batch size and batch count are respectively taken to be

$$m_{i+1} = \lfloor \sqrt{2}m_i \rfloor \text{ and } k_{i+1} = k_i \quad (4.4)$$

so that the total required sample size is $n_{i+1} = m_{i+1}k_{i+1}$; and thus the user must generate the additional simulation responses

$$\{x_j : j = n_i + 1, n_i + 2, \dots, n_{i+1}\} \quad (4.5)$$

before executing iteration $i + 1$ of ASAP. We chose to increase the batch size by the factor $\sqrt{2}$ at each iteration so that the total sample size would double on every other iteration.

4.4 Construction of Time Series Models for Dependent Normal Batch Means

If the batch means have failed the test for independence but have passed the test for joint multivariate normality, then we would like to make a correction to the classical batch means CI to adjust for the correlation between the batch means. A standard way of adjusting CIs for nonnormality is to use some version of an inverted Edgeworth expansion for the corresponding point estimator of the parameter of interest (Hall 1983, Kendall, Stuart and Ord 1987). In the next section, we develop expressions for the higher moments and cumulants of the classical NOBM t -statistic (3.30) involving $\text{Var}[\overline{X}(m)]$ and $\text{Var}[\overline{X}(n)]$. These expressions for the cumulants can then be used in an inverted Cornish-Fisher expansion (Hall 1983) for (3.30) to modify the associated CI, correcting for any error in the variance of the numerator (3.31) of the t -statistic. However, in order to compute sample estimators of these cumulants, we must first have sample estimators of $\text{Var}[\overline{X}(m)]$ and $\text{Var}[\overline{X}(n)]$. We first fit an autoregressive-moving average (ARMA) time series model to the sequence of batch means. We use the estimator $\widehat{\text{Var}}[\overline{X}(m)]$ of the variance of the batch-means process based on the ARMA fit together with the estimated ARMA parameters to calculate an estimate of the variance of the grand mean $\overline{X}(n)$.

If the hypothesis of multivariate normality in Section 4.3 is accepted, then an ARMA process of at most order 2 is fit to the set of $k^* = 94$ batch means. (In this section we suppress the index i of the current iteration of ASAP to simplify the notation; no confusion can result from this since the iteration remains the same throughout the discussion in this section.) Obtaining a good result from the ARMA fitting process generally requires over 50 observations (Box and Jenkins 1976, p. 18), which is the final reason for choosing an initial number of batches close to 100.

Fits of five possible ARMA models are attempted: AR(1), AR(2), MA(1), MA(2), and ARMA(1,1). Adapting the notation in Box and Jenkins (1976) to the notation used here, we let $\{Y_{j-2} : j = 3, \dots, k\}$ represent the $k^* = k - 2$ retained observations of the batch means process and we let $\{\tilde{Y}_{j-2} \equiv Y_{j-2} - \mu_X : j = 3, \dots, k\}$ denote the corresponding deviations from the steady-state mean μ_X . The ℓ th observation of an AR(1) process can be expressed as

$$\tilde{Y}_\ell = \phi_1 \tilde{Y}_{\ell-1} + a_\ell \quad \text{for } \ell = 1, 2, \dots, \quad (4.6)$$

where ϕ_1 is the autoregressive parameter and a_ℓ is an independent normal “shock” with variance σ_a^2 . Similarly, for the second-order autoregressive process, AR(2), we have

$$\tilde{Y}_\ell = \phi_1 \tilde{Y}_{\ell-1} + \phi_2 \tilde{Y}_{\ell-2} + a_\ell \quad \text{for } \ell = 1, 2, \dots. \quad (4.7)$$

Fits of first- and second-order moving average processes are also made, where

$$\tilde{Y}_\ell = a_\ell - \theta_1 a_{\ell-1} \quad \text{for } \ell = 1, 2, \dots \quad (4.8)$$

represents the MA(1) process and

$$\tilde{Y}_\ell = a_\ell - \theta_1 a_{\ell-1} - \theta_2 a_{\ell-2} \quad \text{for } \ell = 1, 2, \dots \quad (4.9)$$

is the MA(2) process. A finite-order moving average process may be represented as an infinite-order autoregressive process. However, if the process is truly MA(1), then an autoregressive representation of it would not be parsimonious. Also, a moving average representation of an autoregressive process would require more parameters than would be necessary with a correctly chosen AR model. That is why we consider both types of models in our algorithm. Sometimes, though, it may be necessary to include both autoregressive and moving average terms in the model to parsimoniously

represent the process (Box and Jenkins 1976). Therefore, we allow for a fifth model, an ARMA(1,1), given by

$$\tilde{Y}_\ell = \phi_1 \tilde{Y}_{\ell-1} + a_\ell - \theta_1 a_{\ell-1} \quad \text{for } \ell = 1, 2, \dots \quad (4.10)$$

IMSL routines (IMSL Problem Solving Software Systems 1987) are used to estimate the autoregressive–moving average parameters, the residual variance, σ_a^2 , and the process variance, $\sigma_Y^2 = \text{Var}[Y_\ell]$, for the five ARMA models. Then the “best” fit of the five is chosen. Preference is given to the AR(1) model. If the MA(1) model has a residual variance less than or equal to 75% of the residual variance from the AR(1) model, then it is chosen as the “best” fit. A second-order model (that is, AR(2), MA(2), or ARMA(1,1)) is chosen only if its residual variance is less than or equal to 50% of the residual variance from the first-order model.

The estimators of $\text{Var}[\overline{X}(m)]$ and the parameters from the ARMA fit are then used to estimate $\text{Var}[\overline{X}(n)]$:

$$\widehat{\text{Var}}[\overline{X}(n)] = \widehat{\text{Var}}[\overline{X}(k^*m)] = \frac{1}{k^*} \sum_{q=-k^*+1}^{k^*-1} \left(1 - \frac{|q|}{k^*}\right) \hat{\gamma}_m(q), \quad (4.11)$$

where $\hat{\gamma}_m(q)$ denotes the estimated lag- q covariance of the batch means Y_{j-2} , $j = 3, \dots, k$ based on the fitted time series model. We now briefly review the equations for obtaining the covariances of the process for the five models. A complete discussion may be found in Box and Jenkins (1976, Chapter 3).

For an AR(1) process (4.6), the covariance at lag q is given by

$$\text{Cov}(\tilde{Y}_\ell, \tilde{Y}_{\ell+q}) = \frac{\phi_1^q}{1 - \phi_1^2} \sigma_a^2, \quad \text{for } q = 0, 1, \dots \quad (4.12)$$

Thus if (4.6) is an adequate model of the batch means process for batches of size m and if $\hat{\phi}_1$ and $\hat{\sigma}_a$ denote the usual maximum likelihood estimates of ϕ_1 and σ_a respectively (Box and Jenkins 1976, Chapter 7), then the estimated covariances in (4.11) are

$$\hat{\gamma}_m(q) = \frac{\hat{\phi}_1^q}{1 - \hat{\phi}_1^2} \hat{\sigma}_a^2 \quad \text{for } q = 0, 1, \dots; \quad (4.13)$$

and a similar approach is taken when other ARMA models are fitted to the batch means process for a given batch size m .

For an AR(2) process, the autoregressive parameters ϕ_1 and ϕ_2 must satisfy the Yule-Walker equations (Box and Jenkins 1976, p. 60)

$$\left. \begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \end{aligned} \right\}, \quad (4.14)$$

where ρ_1 and ρ_2 are the correlations at lags 1 and 2, respectively. Starting with $\rho_0 = 1$ and solving (4.14) for ρ_1 as a function of ϕ_1 and ϕ_2 , we obtain

$$\rho_1 = \phi_1 / (1 - \phi_2). \quad (4.15)$$

We may then determine the lag- q correlation from the relationship

$$\rho_q = \phi_1 \rho_{q-1} + \phi_2 \rho_{q-2} \quad \text{for } q > 1. \quad (4.16)$$

The variance of the AR(2) process is

$$\sigma_Y^2 = \frac{1}{1 - \rho_1 \phi_1 - \rho_2 \phi_2} \sigma_a^2. \quad (4.17)$$

Thus if (4.7) is an adequate model of the batch-means process for batches of size m and if $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\sigma}_a$ denote the usual maximum likelihood estimates of ϕ_1 , ϕ_2 and σ_a respectively, then these parameter estimates are inserted into (4.15), (4.16) and (4.17) to yield the estimated batch means covariance $\hat{\gamma}_m(q)$ at lag q for $q = 0, 1, \dots$

For the remaining autoregressive-moving average models that we used to represent a batch-means process $\{\tilde{Y}_\ell\}$, the covariance functions are given below. The covariances for the MA(1) process (4.8) are given by

$$\text{Cov}(\tilde{Y}_\ell, \tilde{Y}_{\ell+q}) = \begin{cases} (1 + \theta_1^2) \sigma_a^2 & \text{for } q = 0, \\ -\theta_1 \sigma_a^2 & \text{for } q = 1, \\ 0 & \text{for } q > 1. \end{cases} \quad (4.18)$$

The covariances for the MA(2) process (4.9) are given by

$$\text{Cov}(\tilde{Y}_\ell, \tilde{Y}_{\ell+q}) = \begin{cases} (1 + \theta_1^2 + \theta_2^2) \sigma_a^2 & \text{for } q = 0, \\ -\theta_1(1 - \theta_2) \sigma_a^2 & \text{for } q = 1, \\ -\theta_2 \sigma_a^2 & \text{for } q = 2, \\ 0 & \text{for } q > 2. \end{cases} \quad (4.19)$$

Finally, the covariances for the mixed autoregressive-moving average model MA(1,1), process (4.10) are given by

$$\text{Cov}(\tilde{Y}_\ell, \tilde{Y}_{\ell+q}) = \begin{cases} \frac{1+\theta_1^2-2\phi_1\theta_1}{1-\phi_1^2}\sigma_a^2 & \text{for } q = 0, \\ \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1-\phi_1^2}\sigma_a^2 & \text{for } q = 1, \\ \phi_1 \text{Cov}(\tilde{Y}_\ell, \tilde{Y}_{\ell+q-1}) & \text{for } q > 1. \end{cases} \quad (4.20)$$

4.5 Inverted Cornish-Fisher Correction for Dependence of Normal Batch Means

In this section, we develop an adjustment to the usual CI (1.21) to correct for dependency between the batch means. This correction is based on the first four moments of the t -ratio (3.30). In order to pursue this development, we must make the following three assumptions:

A₁: The batch means have a joint multivariate normal distribution.

A₂: The numerator and denominator of the t -ratio (3.30) are independent.

A₃: The square of the denominator (3.32) of the t -ratio (3.30) is distributed as $\chi_{k-1}^2/(k-1)$.

Assumption A₁ is based on using a batch size large enough to yield a nonsignificant result for the multivariate Shapiro-Wilk test as described in the next-to-last paragraph of Section 4.3. Assumption A₂ is supported by the result (3.47): if the batch means have a multivariate normal distribution, then the numerator and squared denominator of the t -statistic (3.30) are uncorrelated. Finally, results of our studies of the convergence properties of the numerator and squared denominator of the NOBM t -ratio (3.30)—in particular Figures 3.43 through 3.54—suggest that the square of the denominator (3.32) has approximately achieved the distribution of a $\chi_{k-1}^2/(k-1)$ variate if the batch size m is large enough to result in approximate multivariate normality of the batch means. With the above assumptions A₁–A₃, we derive expressions for the first four noncentral moments of the t -statistic (3.30).

Let N and D denote the numerator and denominator of the t -statistic, as given in (3.30). First we observe that

$$\mathbb{E}[N] = \mathbb{E}[N^3] = 0 \quad (4.21)$$

since N is normal with mean zero when its component batch means are jointly normal. Squaring the numerator (3.31) and taking the expected value gives

$$\mathbb{E}[N^2] = \frac{k \text{Var}[\overline{X}(n)]}{\text{Var}[\overline{X}(m)] \text{Var}[\overline{X}(n)]} \mathbb{E}\left[(\overline{Y} - \mu_X)^2\right] = \frac{k \text{Var}[\overline{X}(n)]}{\text{Var}[\overline{X}(m)]} \quad (4.22)$$

since $\mathbb{E}[(\overline{Y} - \mu_X)^2] = \text{Var}[\overline{X}(n)]$. Similarly, the fourth moment of the numerator N is

$$\begin{aligned} \mathbb{E}[N^4] &= \mathbb{E}\left[\left(\frac{\overline{Y} - \mu_X}{\sqrt{\text{Var}[\overline{X}(n)]}} \sqrt{\frac{k \text{Var}[\overline{X}(n)]}{\text{Var}[\overline{X}(n)]}}\right)^4\right] \\ &= \frac{k^2 \text{Var}^2[\overline{X}(n)]}{\text{Var}^2[\overline{X}(m)]} \mathbb{E}\left\{\left[\frac{\overline{Y} - \mu_X}{\sqrt{\text{Var}[\overline{X}(n)]}}\right]^4\right\} \\ &= 3 \frac{k^2 \text{Var}^2[\overline{X}(n)]}{\text{Var}^2[\overline{X}(m)]}, \end{aligned} \quad (4.23)$$

since

$$\frac{\overline{Y} - \mu_X}{\sqrt{\text{Var}[\overline{X}(n)]}} \sim N(0, 1) \quad (4.24)$$

when the batch means are multivariate normal, and the fourth moment of a $N(0, 1)$ random variable is 3.

Now under the assumption A_2 that the numerator N and denominator D are independent, we see that when p is a positive integer,

$$\begin{aligned} \mathbb{E}[t^p] &= \mathbb{E}\left[\frac{N^p}{D^p}\right] \\ &= \mathbb{E}[N^p] \cdot \mathbb{E}[D^{-p}] \\ &= 0 \quad \text{for } p = 1, 3, \end{aligned} \quad (4.25)$$

provided $k \geq 4$ to ensure the existence of $\mathbb{E}[D^{-p}]$ for $p = 1, 3$. From (4.22) it follows that the second moment of the t -statistic, which we will denote μ_2 , can be written

$$\mu_2 = \mathbb{E}[t^2] = \mathbb{E}\left[\frac{N^2}{D^2}\right] = \mathbb{E}[N^2] \mathbb{E}[D^{-2}] = \frac{k \text{Var}[\overline{X}(n)]}{\text{Var}[\overline{X}(m)]} \mathbb{E}\left[\frac{1}{D^2}\right]. \quad (4.26)$$

Based on assumption A₃, we may derive an expression for the expected value on the right-hand side of (4.26):

$$\begin{aligned}
E\left[\frac{1}{D^2}\right] &= E\left[\frac{k-1}{\chi_{k-1}^2}\right] \\
&= (k-1) \int_0^\infty \frac{1}{y} \left(\frac{y^{(k-3)/2} e^{-y/2}}{\Gamma\left(\frac{k-1}{2}\right) 2^{(k-1)/2}} \right) dy \\
&= (k-1) \left(\frac{\Gamma\left(\frac{k-3}{2}\right) 2^{(k-3)/2}}{\Gamma\left(\frac{k-1}{2}\right) 2^{(k-1)/2}} \right) \int_0^\infty \frac{y^{(\frac{k-3}{2}-1)} e^{-y/2}}{\Gamma\left(\frac{k-3}{2}\right) 2^{(k-3)/2}} dy. \quad (4.27)
\end{aligned}$$

Provided $k \geq 4$, the integrand in the final right-hand side of (4.27) is the density function of a χ_{k-3}^2 random variable; and therefore the integral is equal to one. If we simplify the coefficient in (4.27), then we can write

$$E[t^2] = \frac{k \text{Var}[\bar{X}(n)]}{\text{Var}[\bar{X}(m)]} \frac{(k-1)}{(k-3)} \text{ if } k \geq 4. \quad (4.28)$$

Similarly, we see that under the assumption A₃, we have

$$\mu_4 = E[t^4] = E\left[\frac{N^4}{D^4}\right] = E[N^4] E[D^{-4}]. \quad (4.29)$$

Now

$$\begin{aligned}
E\left[\frac{1}{D^4}\right] &= E\left[\frac{(k-1)^2}{(\chi_{k-1}^2)^2}\right] \\
&= (k-1)^2 \int_0^\infty \frac{1}{y^2} \left(\frac{y^{(k-3)/2} e^{-y/2}}{\Gamma\left(\frac{k-1}{2}\right) 2^{(k-1)/2}} \right) dy \\
&= (k-1)^2 \left(\frac{\Gamma\left(\frac{k-5}{2}\right) 2^{(k-5)/2}}{\Gamma\left(\frac{k-1}{2}\right) 2^{(k-1)/2}} \right) \int_0^\infty \frac{y^{(\frac{k-5}{2}-1)} e^{-y/2}}{\Gamma\left(\frac{k-5}{2}\right) 2^{(k-5)/2}} dy. \quad (4.30)
\end{aligned}$$

Recognizing the integrand on the final right-hand side of (4.30) as the density function of a χ_{k-5}^2 random variable, evaluating the integral as 1 and simplifying the coefficient, we obtain the final expression for μ_4 :

$$\mu_4 = E[t^4] = \frac{3k^2(k-1)^2 \text{Var}^2[\bar{X}(n)]}{(k-3)(k-5) \text{Var}^2[\bar{X}(m)]}, \text{ provided } k \geq 6. \quad (4.31)$$

The inverted Cornish-Fisher expansion we will be using involves the cumulants κ_1 , κ_2 , κ_3 and κ_4 of the t -ratio (3.30) computed under assumptions A₁–A₃. From (4.25),

(4.28) and (4.31) together with Section 3.14 of Kendall, Stuart, and Ord (1987), we have

$$\kappa_1 = 0, \quad (4.32)$$

$$\kappa_2 = \mu_2 = \frac{k \text{Var}[\overline{X}(n)](k-1)}{\text{Var}[\overline{X}(m)](k-3)}, \quad (4.33)$$

$$\kappa_3 = \mu_3 = 0, \quad (4.34)$$

and

$$\begin{aligned} \kappa_4 = \mu_4 - \mu_2^2 &= \frac{k^2 \text{Var}^2[\overline{X}(n)]}{\text{Var}^2[\overline{X}(m)]} \left\{ \frac{(k-1)^2}{(k-3)(k-5)} - \frac{(k-1)^2}{(k-3)^2} \right\} \\ &= \frac{2k^2(k-1)^2 \text{Var}^2[\overline{X}(n)]}{(k-3)^2(k-5) \text{Var}^2[\overline{X}(m)]}. \end{aligned} \quad (4.35)$$

Based on an inverted Cornish-Fisher expansion for the classical NOBM t -ratio, an adjusted $100(1-\alpha)\%$ confidence interval for μ_X is

$$\left[\overline{X}(n) - h'(z_{1-\alpha/2}) \frac{S_{n,k}}{\sqrt{k}}, \overline{X}(n) - h'(-z_{1-\alpha/2}) \frac{S_{n,k}}{\sqrt{k}} \right], \quad (4.36)$$

where $n = km$ and

$$h'(z_{1-\alpha/2}) = z_{1-\alpha/2} + (\kappa_1 - \kappa_3/6) + [(\kappa_2 - 1)/2] z_{1-\alpha/2} + (\kappa_3/6) z_{1-\alpha/2}^2, \quad (4.37)$$

for the first-order pivot, or

$$\begin{aligned} h'(z_{1-\alpha/2}) &= z_{1-\alpha/2} + (\kappa_1 - \kappa_3/6) + [(\kappa_2 - 1)/2 - \kappa_4/8] z_{1-\alpha/2} \\ &\quad + (\kappa_3/6) z_{1-\alpha/2}^2 + (\kappa_4/24) z_{1-\alpha/2}^3, \end{aligned} \quad (4.38)$$

for the second-order pivot, and κ_i denotes the i th cumulant of the t -statistic, for $i = 1, 2, 3, 4$ (Hall 1983, Chien 1989).

Under the assumptions A₁–A₃, the first four cumulants of the t -ratio are given by (4.32)–(4.35). By substituting the variance estimator $\widehat{\text{Var}}[\overline{X}(m)]$ from the ARMA fit for $\text{Var}[\overline{X}(m)]$ and by substituting the variance estimator $\widehat{\text{Var}}[\overline{X}(n)]$ of display (4.11) for $\text{Var}[\overline{X}(n)]$ in the expressions (4.33) and (4.35) for κ_2 and κ_4 , we obtain the following approximate $100(1-\alpha)\%$ confidence intervals for μ_X :

$$\overline{X}(n) \pm z_{1-\alpha/2} \left(1 + \frac{\hat{\kappa}_2 - 1}{2} \right) \sqrt{\frac{\widehat{\text{Var}}[\overline{X}(m)]}{k}}, \quad (4.39)$$

or

$$\overline{X}(n) \pm \left[z_{1-\alpha/2} \left(1 + \frac{\hat{\kappa}_2 - 1}{2} - \frac{\hat{\kappa}_4}{8} \right) + \frac{\hat{\kappa}_4}{24} z_{1-\alpha/2}^3 \right] \sqrt{\frac{\widehat{\text{Var}}[\overline{X}(m)]}{k}}, \quad (4.40)$$

depending on which pivot is used.

4.6 Fulfilling the Precision Requirement

The final step in ASAP is to determine if the confidence interval that was constructed meets the user's requirement for precision. The confidence interval may be the one based on a nonsignificant result from the independence test (that is, the batch means pass the test for independence) or the adjusted CI based on a nonsignificant result from the test for multivariate normality of the batch means (that is, the batch means pass the test for multivariate normality). If the relevant requirement

$$H \leq H^* \quad \text{or} \quad H \leq r^* |\overline{Y}| \quad (4.41)$$

for the precision of the confidence interval is satisfied, then ASAP terminates, returning the sample mean $\overline{Y}$ and the CI half-length H . If the precision requirement (4.41) is not satisfied on iteration i of ASAP, then the procedure estimates the number of additional batches k_i^+ required to satisfy (4.41) using batch size m_i ; thus on iteration $i + 1$ of ASAP the batch count and batch size are tentatively taken to be

$$k_{i+1} \leftarrow k_i + k_i^+ \quad \text{and} \quad m_{i+1} \leftarrow m_i \quad (4.42)$$

so that the total required sample size on iteration $i + 1$ is tentatively taken to be $n_{i+1}^* \leftarrow k_{i+1} m_{i+1}$.

For ease in coding ASAP, we wished to specify an upper limit on the number of batches the algorithm could require. Preliminary experiments with ASAP revealed that, except in the most difficult cases, 1500 batches were sufficient to meet relative precision requirements as small as $\pm 5\%$. If the total number of batches k_{i+1} exceeds 1502, then a new tentative batch size,

$$m_{i+1} \leftarrow \lfloor \sqrt{2} m_{i+1} \rfloor, \quad (4.43)$$

is calculated and a new tentative number of batches

$$k_{i+1} \leftarrow \lfloor n_{i+1}^* / m_{i+1} \rfloor \quad (4.44)$$

is computed. The tentative batch size m_{i+1} is repeatedly increased by a factor of $\sqrt{2}$ and the resulting k_{i+1} is recomputed until $k_{i+1} \leq 1502$. Thus if the user-specified precision requirement (4.41) is not satisfied on iteration i of ASAP, then iteration $i+1$ will be required in which the number of batches $k_{i+1} \leq 1502$ and the total sample size is finally taken to be

$$n_{i+1} \leftarrow m_{i+1}k_{i+1}; \quad (4.45)$$

and thus the user must generate the additional simulation responses

$$\{x_j : j = n_i + 1, n_i + 2, \dots, n_{i+1}\} \quad (4.46)$$

before executing iteration $i+1$ of ASAP.

The user then performs iteration $i+1$ of ASAP with the values of m_{i+1} , k_{i+1} and n_{i+1} for the batch size, batch count and total sample size, respectively. The first two batches are again omitted from the calculation of the sample mean. The batch means of $(k_{i+1} - 2)$ batches of size m_{i+1} are computed. If an ARMA model was used in constructing an adjusted CI on the previous iteration, then an updated ARMA fit is made using $(k_{i+1} - 2)$ batches of size m_{i+1} ; moreover, in this situation new estimates of $\text{Var}[\bar{X}(m)]$, $\text{Var}[\bar{X}(n)]$, κ_2 and κ_4 are computed, and the CI (4.39) or (4.40) is constructed. If the CI for the previous iteration was based on batch means that passed the independence test, then the classical NOBM confidence interval (1.21) is constructed with the batch means of $(k_{i+1} - 2)$ batches of size m_{i+1} . If the precision requirement (4.41) is satisfied on iteration $i+1$ of ASAP, then the algorithm terminates, returning $\bar{Y}$ and H . If the required precision is not achieved on iteration $i+1$ of ASAP, then the rest of iteration $i+1$ of ASAP proceeds along the same lines as described by displays (4.42) through (4.45) for iteration i .

4.7 Advantages of ASAP

ASAP has a feature that addresses the initial bias problem, albeit somewhat crudely, by eliminating the first two batches from the calculations of the mean and variance estimators. However, if we have reached a batch size m large enough to pass either the independence test or the multivariate normality test, then by eliminating $2m$ observations we have probably overcome the initial bias.

We feel that ASAP goes beyond conventional batching procedures in that it not only determines a feasible batch size, given a fixed total sample size, but it also ascertains the total sample size needed for steady-state CI construction. While ASAP is clearly a heuristic method and is not guaranteed to deliver valid CIs, it does have a strong basis in theory. It addresses both the assumptions of independence and normality upon which the NOBM method is based. ASAP does not rely entirely on achieving independence of the batch means, which can be a disadvantage in highly dependent systems.

ASAP requires a moderate amount of initial sampling, since $n_1 = 1536$. While this is about twice the initial sample size of 800 that Law and Carson (1979) used for testing their procedure, we found that in most cases in which a precision requirement was given, the first-iteration sample size was never excessive.

By correcting for remaining dependence between batch means, we are rewarded with CIs which approach nominal coverage. The downside to this, of course, is that in processes with high dependency the CI half-length delivered by ASAP can be substantially larger than that delivered by procedures based on approximately independent batch means. However, the precision requirement can compensate for this with more sampling. In theory, where the NOBM variance estimator overestimates the variance (e.g., in processes with alternating correlation structure), we may get smaller CIs than we would without the correction.

4.8 Formal Statement of ASAP

Figure 4.1 displays a formal algorithmic statement of ASAP. Appendix B contains a brief user's manual for the ASAP software.

0. Fix integers $m_1 \leftarrow$ user specified minimum batch size [default: $m_1 \leftarrow 16$],
 $i \leftarrow 1$, $k_1 \leftarrow 96$, $k_1^* \leftarrow 94$, $k_1^+ \leftarrow 0$, $n_0 \leftarrow 1$, and $n_1 \leftarrow 96m_1$.
 Fix confidence level $1 - \alpha$.
 Set **IndTestPassed** $\leftarrow$ ‘no’ and **MVTestPassed** $\leftarrow$ ‘no’.
 If *relative* precision required, then set **RelPrec** $\leftarrow$ ‘yes’ and
      $r^* \leftarrow$ fraction of magnitude of final sample mean that gives
     upper bound H^* on acceptable CI half-length.
 If *absolute* precision required, then set **RelPrec** $\leftarrow$ ‘no’ and
      $H^* \leftarrow$ absolute upper bound on acceptable CI half-length.

1. Collect observations $X_{n_0}, \dots, X_{n_i}$.
 Compute: batch means $Y_1(m_i), \dots, Y_{k_i}(m_i)$ and the grand mean

$$\bar{X}(k_i^* m_i) \leftarrow \frac{1}{k_i^* m_i} \sum_{\ell=2m_i+1}^n X_\ell = \frac{1}{k_i^*} \sum_{j=3}^{k_i} Y_j(m_i).$$

if **MVTestPassed**=‘yes’, **then goto** 6;
if **IndTestPassed**=‘yes’, **then goto** 3;
 Test H_{ind} : $\{Y_3(m_i), \dots, Y_{96}(m_i)\}$ are i.i.d. using (2.41).
if H_{ind} is rejected, **then goto** 5;

2. **IndTestPassed** $\leftarrow$ ‘yes’;
 3. Calculate

$$S_{k_i^* m_i, k_i^*}^2 \leftarrow \frac{1}{k_i^* - 1} \sum_{j=3}^{k_i} \left(Y_j(m_i) - \bar{X}(k_i^* m_i) \right)^2 \quad \text{and} \quad H \leftarrow z_{1-\alpha/2} \frac{S_{k_i^* m_i, k_i^*}}{\sqrt{k_i^*}}$$

Construct CI

$$\bar{X}(k_i^* m_i) \pm H$$

if **RelPrec**=‘yes’ **then** $H^* \leftarrow r^* |\bar{X}(k_i^* m_i)|$.
if $(H \leq H^*)$ **or** $(r^* = 0 \text{ and } H^* = 0)$, **then** deliver CI $\bar{X}(k_i^* m_i) \pm H$
and stop

Figure 4.1: ASAP Algorithm.

4. Calculate additional number of batches needed,

$$k_i^+ \leftarrow \left\lceil \left(\frac{H}{H^*} \right)^2 k_i^* \right\rceil - k_i^*.$$

Set new number of batches $k_{i+1} \leftarrow k_i + k_i^+$, new batch size $m_{i+1} \leftarrow m_i$,

and new sample size $n_{i+1}^* \leftarrow k_{i+1} m_{i+1}$,

if $k_{i+1} \leq 1502$ **then**

$k_{i+1}^* \leftarrow k_i^* + k_i^+$, $n_0 \leftarrow n_i + 1$, $n_{i+1} \leftarrow k_{i+1} m_{i+1}$, and $i \leftarrow i + 1$;

goto 1;

else

while $k_{i+1} > 1502$

begin

$m_{i+1} \leftarrow \lfloor \sqrt{2} m_{i+1} \rfloor$ and $k_{i+1} \leftarrow \lceil n_{i+1}^* / m_{i+1} \rceil$;

end

$n_0 \leftarrow n_i + 1$, $n_{i+1} \leftarrow k_{i+1} m_{i+1}$, and $i \leftarrow i + 1$;

goto 1;

5. Test H_{mvn} :

$\{[Y_{3+(j-1)6}(m_i), Y_{4+(j-1)6}(m_i), Y_{5+(j-1)6}(m_i), Y_{6+(j-1)6}(m_i)] : 1 \leq j \leq 16\}$

are quadrivariate normal random vectors;

if H_{mvn} is rejected, **then**

$m_{i+1} \leftarrow \lfloor \sqrt{2} m_i \rfloor$, $n_0 \leftarrow n_i + 1$, $n_{i+1} \leftarrow 96 m_i$, $k_{i+1} \leftarrow 96$, and $i \leftarrow i + 1$;

goto 1;

else

MVTestPassed $\leftarrow$ 'yes';

goto 6;

Figure 4.1. ASAP Algorithm continued.

6. Fit an ARMA process to the batch means $\{Y_3(m_i), \dots, Y_{k_i}(m_i)\}$;
 Compute variance estimators $\widehat{\text{Var}}[\overline{X}(m_i)]$ and $\widehat{\text{Var}}[\overline{X}(n_i)]$ from the ARMA fit;
 Calculate estimates $\hat{\kappa}_2$ and $\hat{\kappa}_4$ of the second and fourth cumulants of the t -statistic and

$$H \leftarrow \left[z_{1-\alpha/2} \left(1 + \frac{\hat{\kappa}_2 - 1}{2} - \frac{\hat{\kappa}_4}{8} \right) + \frac{\hat{\kappa}_4}{24} z_{1-\alpha/2}^3 \right] \sqrt{\frac{\widehat{\text{Var}}[\overline{X}(m)]}{k}},$$

construct the CI

$$\overline{X}(k_i^* m_i) \pm H$$

if RelPrec='yes' **then** $H^* \leftarrow r^* |\overline{X}(k_i^* m_i)|$.

if $(H \leq H^*)$ **or** $(r^* = 0 \text{ and } H^* = 0)$, **then**

    deliver CI $\overline{X}(k_i^* m_i) \pm H$ and **stop**;

else

goto 4.

Figure 4.1. ASAP Algorithm continued.

Chapter 5

Performance Evaluation of ASAP

We tested ASAP on many problems representing various types of stochastic processes. In this chapter we discuss the results of that experimentation. The steady-state mean is available analytically in all of these models. Therefore, we were able to evaluate the performance of ASAP in terms of actual coverage versus nominal coverage.

5.1 Suite of Test Problems

We chose a set of 20 problems, which includes some classic problems used for testing simulation output analysis procedures, some problems which more closely resemble real world problems, and some problems which display characteristics that we believe should stress any output analysis procedure, i.e., high dependency or markedly non-normal marginal distributions. We began the testing of ASAP with the six models used in Chapter 3 for the study of the convergence properties of batch means. We also tested ASAP using the 14 models that Law and Carson (1979) used in testing their batch-means algorithm (see Section 2.3.1).

5.1.1 Discrete Time Markov Chains

The functions defined on the three finite-state Discrete Time Markov Chains that we studied in Chapter 3 were included as test cases for ASAP. See Section 3.3.1 for a detailed description of these models. The first case is a real-valued function defined on a simple 2-state chain with one-step transition probability matrix (3.65) and with an alternating correlation structure (see Figure 3.1). The second case is a function

on a 2-state chain, with one-step transition probability matrix (3.67) and with high positive correlation between observations (see Figure 3.8). Since the matrix of one-step transition probabilities is doubly stochastic (that is, both the rows and columns sum to unity) for each of these two-state chains, the steady-state distribution vector for both of these 2-state chains is given by

$$\boldsymbol{\pi} = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}. \quad (5.1)$$

Since the cost vector $\mathbf{h}$ (see Equation (3.66)) is the same for both chains, their steady-state means both have the value $\mu_X = 7.5$.

The last case of a function defined on a DTMC was the cost function (3.69) defined on an 8-state DTMC with one-step transition probability matrix (3.68). In this case the correlation function resembles a damped sine wave (see Figure 3.15). The steady-state distribution vector for this model is given by

$$\boldsymbol{\pi} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0.1262 & 0.1255 & 0.1249 & 0.1244 & 0.1240 & 0.1237 & 0.1240 & 0.1273 \end{pmatrix}. \quad (5.2)$$

Using the cost function vector given in (3.69), we can calculate the steady-state mean $\mu_X = 17.4905$.

5.1.2 Autoregressive and Exponential Autoregressive Processes

The ASAP algorithm was also tested using an AR(1) process (3.75) with $\varphi = 0.9$, described in Section 3.3.3. This case also has decaying positive correlation structure. For our simulations, we chose $Z_0 \sim N(2.0, 5.263)$ and $\epsilon_i \sim N(0.2, 1)$. The steady-state mean μ_X is therefore 2.0.

We also simulated the EAR(1) process (3.78) with $\varphi = 0.9$ and ϵ_i distributed as exponential with rate 0.5 (see Section 3.3.4). The steady-state mean of this process is $\mu_X = 2.0$.

5.1.3 Queueing Systems

The last case that was used in studying the convergence properties of batch means was the sequence of waiting times in the M/M/1 queue with utilization $\tau = 0.9$.

This process is defined by equations (3.70)–(3.74). We also used this case as the first queueing model with which to test ASAP. This is a particularly difficult case in that the correlation of the observations is initially high and decays very slowly with increasing lags (see Figure 3.22). Also the marginal distribution (3.72) of the X_i 's has an exponential tail and is therefore markedly nonnormal.

The other queueing systems tested were the ones that Law and Carson (1979) used. Employing their notation, we let E_4 , M , and H_2 denote, respectively, the Erlang-4, exponential, and hyperexponential distributions with coefficients of variation 0.5, 1, and 2. Law and Carson tested eight queueing systems: $M/M/1$ ($\tau = 0.5$), $M/M/1$ ($\tau = 0.8$), $M/M/1$ LIFO ($\tau = 0.8$), $M/M/1$ with service in random order (SIRO) ($\tau = 0.8$), $E_4/M/1$ ($\tau = 0.8$), $M/H_2/1$ ($\tau = 0.8$), $M/M/2$ ($\tau = 0.8$), and $M/M/1/M/1$ ($\tau = 0.8$). The simulations of all queueing systems were started empty and idle. The steady-state expected waiting times are given in Table 5.1.

Table 5.1: Steady-State Expected Waiting Time in Selected Queueing Systems

System	Utilization τ	Expected Waiting Time
$M/M/1$	0.9	9.00
$M/M/1$	0.5	0.50
$M/M/1$	0.8	3.20
$M/M/1$ LIFO	0.8	3.20
$M/M/1$ SIRO	0.8	3.20
$E_4/M/1$	0.8	1.81
$M/H_2/1$	0.8	8.00
$M/M/2$	0.8	2.84
$M/M/1/M/1$	0.8	6.40

5.1.4 Computer Models

Law and Carson (1979) also tested their sequential output analysis procedure on queueing network models of computer systems. The first is a time-shared model with a single central processing unit (CPU) and N terminals. Each terminal “thinks” for a period of time that is exponentially distributed with rate μ_1 and then sends a job

to the CPU with a service time which is exponentially distributed with rate μ_2 . The jobs join a queue at the CPU, which allocates a maximum “time slice” of s^* time units to each job in FIFO order. If the remaining service time s of a job is less than s^* , then the job spends s time units plus a fixed overhead of h time units at the CPU and then returns to the terminal. If $s > s^*$, then the CPU spends $s^* + h$ time units processing the job, which then returns to the end of the queue. This process is continued until the job is finished, and then it returns to the terminal. The process of interest is the response time of the jobs $\{R_i : i \geq 1\}$ —that is, R_i is the elapsed time between the instant that job i joins the CPU queue at the end of a “thinking” period and the instant that job i completes its last time slice of service on the CPU. We chose the same parameters that Law and Carson used, i.e., $N = 35$, $\mu_1 = 1/25$, $\mu_2 = 5/4$, $s^* = 0.1$, and $h = 0.015$. (See Law and Carson 1979 for a more complete description of the model.)

The second computer model used by Law and Carson (1979) consists of a central server (CPU) and $M - 1$ peripheral units. The system has a fixed number jobs, N , in it. When a job is finished at the CPU, it leaves the system with probability p_1 and is immediately replaced with another job at the CPU queue. If the job does not leave the system, then it is routed to a peripheral unit. The probability that the job is routed to unit i from the CPU is p_i , $i = 2, \dots, M$. After getting service at one of the peripheral units, the job leaves the system and is immediately replaced by a job joining the CPU queue. The process of interest is the response time of a job, i.e., the time between its arrival at the CPU queue and its departure from the system. Law and Carson chose to simulate this model for four cases. Table 5.2 gives the four sets of parameters for which Law and Carson simulated this process. Also given are the steady-state utilizations of the CPU (unit 1) and the peripheral units. In each of the four cases p_1 , the probability that the job leaves the system after service at the CPU, is zero and the number of peripheral units is two.

5.1.5 An Inventory System

We also considered the final model Law and Carson used for testing their algorithm, an (s, S) inventory system with zero delivery lag and backlogging. (See pp. 281–282 of Law and Kelton 1991 for a detailed description.) The process of interest is the operating cost per period, which has steady-state mean $\mu_X = 112.108$.

Table 5.2: Parameters for the Selected Central Server Models

Model	N	μ_2	μ_3	p_2	p_3	τ_1	τ_2	τ_3	Expected Response Time	Initial State
1	4	0.50	0.50	0.5	0.5	0.67	0.67	0.67	6.000	(1,1,2)
2	8	0.50	0.50	0.5	0.5	0.80	0.80	0.80	10.000	(1,1,6)
3	8	0.45	0.05	0.9	0.1	0.44	0.88	0.88	18.279	(5,1,2)
4	8	1.79	0.20	0.9	0.1	0.99	0.50	0.50	8.072	(7,1,0)

5.2 Summary of Experimental Results

We made 100 independent simulations of each system described in Section 5.1 and attempted to construct nominal 90% confidence intervals for three cases:

- (i) no precision requirement, i.e., we did *not* require that the CI have an absolute length, as in (1.22) or a length that was some fraction of the sample mean, as in (1.23), but instead terminated the procedure when a CI was constructed based on 94 batches of the size at which the batch means passed either the statistical test for independence or the test for multivariate normality;
- (ii) $\pm 15\%$ precision so that $r^* = 0.15$ in (4.41); and
- (iii) $\pm 7.5\%$ precision so that $r^* = 0.075$ in (4.41).

This enabled us to estimate the actual coverage of CIs constructed via the ASAP algorithm. We also tested the performance of the LATCH and ABATCH algorithms as described in Section 2.3.4 for comparison. Since LATCH and ABATCH do not explicitly determine a sample size, we passed to the LATCH and ABATCH algorithms the same data sets used by ASAP. Therefore, the CIs computed by the LATCH and ABATCH algorithms are based on the same sample sizes and data sets used for ASAP. We do not mean to imply that these results are what one could expect routinely from LATCH and ABATCH, but only what one could expect by applying LATCH and ABATCH to the data sets that resulted from first using the ASAP algorithm.

In this section Tables 5.3 through 5.22 display in detail the results of our tests. We included Law and Carson's (1979) results for those cases which were part of their study.

Tables 5.23 and 5.24 display the additional results obtained through standalone application of LBATCH and ABATCH to the 2-state DTMC with high correlation and to the $M/M/1$ queue with $\tau = 0.9$ when LBATCH and ABATCH operate with a stopping rule based on a user-specified precision requirement for the final confidence interval. We began the experiments for these systems with a sample size of 1536 (the same sample size required for the first iteration of ASAP). We then applied a stopping rule similar to the one used for ASAP. After the initial run with 1536 observations was made, the final CI constructed by LBATCH or ABATCH was examined to see if the precision requirement was met. If not, then we calculated an estimate of additional observations that would be required, we generated the additional observations, and we executed LBATCH or ABATCH again with all of the observations accumulated so far. This process was repeated until the final CI delivered by LBATCH or ABATCH met the precision requirement. We realize that LBATCH and ABATCH were not really designed to be used in this way, but using such stopping rules is a natural approach to planning steady-state simulations; and we believe that the results in Tables 5.23 and 5.24 provide a more complete perspective on the relative performance of LBATCH and ABATCH versus ASAP. No effort was made to analyze the convergence of the sample estimators from LBATCH and ABATCH, as is suggested in Fishman (1998). We only include these results to highlight the performance advantages achieved by ASAP without requiring analysis or manual intervention by the user.

Table 5.3: Performance of Batch-Means Procedures for the 2-State DTMC with Alternating Correlation Structure as Defined by (3.65) and (3.66) Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	NA			1580
coverage	NA	91%	91%	92%
avg. rel. precision	NA	0.0081	0.0082	0.0080
avg. CI half length	NA	0.061	0.061	0.060
var. CI half length	NA	0.00006	0.00007	0.00006
$\pm 15\%$ PRECISION				
average sample size	NA			1580
coverage	NA	91%	91%	92%
avg. rel. precision	NA	0.0081	0.0082	0.0080
avg. CI half length	NA	0.061	0.061	0.060
var. CI half length	NA	0.00006	0.00007	0.00006
$\pm 7.5\%$ PRECISION				
average sample size	NA			1580
coverage	NA	91%	91%	92%
avg. rel. precision	NA	0.0081	0.0082	0.0080
avg. CI half length	NA	0.061	0.061	0.060
var. CI half length	NA	0.00006	0.00007	0.00006

[†]No. of classical and corrected CIs generated by ASAP: 74 and 26, respectively.

Table 5.4: Performance of Batch-Means Procedures for the 2-State DTMC with High Positive Correlation Structure as Defined by (3.66) and (3.67) Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Relative Precision	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	NA			3036
coverage	NA	70%	85%	96%
avg. rel. precision	NA	0.069	0.086	0.159
avg. CI half length	NA	0.515	0.642	1.20
var. CI half length	NA	0.009	0.012	0.172
$\pm 15\%$ PRECISION				
average sample size	NA			5171
coverage	NA	72%	81%	96%
avg. rel. precision	NA	0.060	0.070	0.120
avg. CI half length	NA	0.045	0.053	0.906
var. CI half length	NA	0.011	0.010	0.023
$\pm 7.5\%$ PRECISION				
average sample size	NA			22711
coverage	NA	81%	86%	99%
avg. rel. precision	NA	0.034	0.038	0.059
avg. CI half length	NA	0.253	0.284	0.438
var. CI half length	NA	0.003	0.003	0.006

[†]No. of classical and corrected CIs generated by ASAP: 0 and 100, respectively.

Table 5.5: Performance of Batch-Means Procedures for the 8-State DTMC as Defined by (3.68) and (3.69) Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	NA			1553
coverage	NA	90%	89%	90%
avg. rel. precision	NA	0.022	0.023	0.022
avg. CI half length	NA	0.394	0.396	0.389
var. CI half length	NA	0.002	0.003	0.002
$\pm 15\%$ PRECISION				
average sample size	NA			1553
coverage	NA	90%	89%	90%
avg. rel. precision	NA	0.022	0.023	0.022
avg. CI half length	NA	0.394	0.396	0.389
var. CI half length	NA	0.002	0.003	0.002
$\pm 7.5\%$ PRECISION				
average sample size	NA			1553
coverage	NA	90%	89%	90%
avg. rel. precision	NA	0.022	0.023	0.022
avg. CI half length	NA	0.394	0.396	0.389
var. CI half length	NA	0.002	0.003	0.002

†No. of classical and corrected CIs generated by ASAP: 70 and 30, respectively.

Table 5.6: Performance of Batch-Means Procedures for the AR(1) Process (3.75) with $\varphi = 0.9$ and $\mu_X = 2.0$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	NA			1624
coverage	NA	84%	86%	93%
avg. rel. precision	NA	0.180	0.198	0.228
avg. CI half length	NA	0.351	0.387	0.446
var. CI half length	NA	0.002	0.005	0.006
$\pm 15\%$ PRECISION				
average sample size	NA			5862
coverage	NA	88%	88%	93%
avg. rel. precision	NA	0.104	0.107	0.123
avg. CI half length	NA	0.208	0.212	0.244
var. CI half length	NA	0.001	0.001	0.001
$\pm 7.5\%$ PRECISION				
average sample size	NA			24860
coverage	NA	87%	88%	91%
avg. rel. precision	NA	0.053	0.054	0.059
avg. CI half length	NA	0.106	0.108	0.118
var. CI half length	NA	0.0003	0.0003	0.0003

†No. of classical and corrected CIs generated by ASAP: 2 and 98, respectively.

Table 5.7: Performance of Batch-Means Procedures for the EAR(1) Process (3.78) with $\varphi = 0.9$ and $\mu_X = 2.0$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	NA			2030
coverage	NA	88%	93%	93%
avg. rel. precision	NA	0.152	0.160	0.182
avg. CI half length	NA	0.305	0.322	0.366
var. CI half length	NA	0.0038	0.0067	0.0097
$\pm 15\%$ PRECISION				
average sample size	NA			4320
coverage	NA	89%	90%	92%
avg. rel. precision	NA	0.107	0.110	0.121
avg. CI half length	NA	0.214	0.219	0.243
var. CI half length	NA	0.00112	0.00162	0.00132
$\pm 7.5\%$ PRECISION				
average sample size	NA			16714
coverage	NA	92%	93%	94%
avg. rel. precision	NA	0.057	0.058	0.062
avg. CI half length	NA	0.114	0.116	0.123
var. CI half length	NA	0.0004	0.0004	0.0002

†No. of classical and corrected CIs generated by ASAP: 10 and 90, respectively.

Table 5.8: Performance of Batch-Means Procedures for the M/M/1 Queue Waiting Time Process with $\tau = 0.9$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	NA			7719
coverage	NA	44%	60%	83%
avg. rel. precision	NA	0.202	0.301	1.088
avg. CI half length	NA	1.70	2.67	11.8
var. CI half length	NA	0.683	3.92	523.0
$\pm 15\%$ PRECISION				
average sample size	NA			298950
coverage	NA	79%	80%	88%
avg. rel. precision	NA	0.061	0.069	0.089
avg. CI half length	NA	0.543	0.613	0.783
var. CI half length	NA	0.027	0.039	0.082
$\pm 7.5\%$ PRECISION				
average sample size	NA			815755
coverage	NA	88%	90%	94%
avg. rel. precision	NA	0.039	0.043	0.046
avg. CI half length	NA	0.353	0.382	0.413
var. CI half length	NA	0.012	0.039	0.018

†No. of classical and corrected CIs generated by ASAP: 4 and 96, respectively.

Table 5.9: Performance of Batch-Means Procedures for the M/M/1 Queue Waiting Time Process with $\tau = 0.5$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	3768			2259
coverage	91%	81%	81%	80%
avg. rel. precision	0.136	0.168	0.168	0.166
avg. CI half length	NA	0.163	0.163	0.162
var. CI half length	NA	0.001	0.001	0.001
$\pm 15\%$ PRECISION				
average sample size	5448			4326
coverage	90%	85%	84%	82%
avg. rel. precision	0.131	0.127	0.128	0.120
avg. CI half length	NA	0.124	0.126	0.116
var. CI half length	NA	0.0006	0.0006	0.0003
$\pm 7.5\%$ PRECISION				
average sample size	19472			15828
coverage	87%	89%	91%	84%
avg. rel. precision	0.064	0.071	0.072	0.063
avg. CI half length	NA	0.070	0.072	0.063
var. CI half length	NA	0.0001	0.0002	0.00006

†No. of classical and corrected CIs generated by ASAP: 87 and 13, respectively.

Table 5.10: Performance of Batch-Means Procedures for the M/M/1 Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	32960			8482
coverage	85%	67%	73%	85%
avg. rel. precision	0.096	0.169	0.196	0.286
avg. CI half length	NA	0.682	0.811	1.31
var. CI half length	NA	0.088	0.025	5.21
$\pm 15\%$ PRECISION				
average sample size	32960			38911
coverage	85%	80%	83%	87%
avg. rel. precision	0.096	0.093	0.100	0.112
avg. CI half length	NA	0.367	0.392	0.441
var. CI half length	NA	0.010	0.013	0.009
$\pm 7.5\%$ PRECISION				
average sample size	75648			136491
coverage	87%	88%	92%	90%
avg. rel. precision	0.065	0.053	0.055	0.056
avg. CI half length	NA	0.210	0.219	0.222
var. CI half length	NA	0.0027	0.0032	0.0019

[†]No. of classical and corrected CIs generated by ASAP: 33 and 67, respectively.

Table 5.11: Performance of Batch-Means Procedures for the M/M/1 LIFO Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	3120			5025
coverage	64%	72%	75%	72%
avg. rel. precision	0.236	0.209	0.223	0.210
avg. CI half length	NA	0.645	0.693	0.652
var. CI half length	NA	0.070	0.113	0.074
$\pm 15\%$ PRECISION				
average sample size	13944			14317
coverage	76%	80%	81%	77%
avg. rel. precision	0.131	0.135	0.143	0.119
avg. CI half length	NA	0.426	0.451	0.372
var. CI half length	NA	0.009	0.013	0.004
$\pm 7.5\%$ PRECISION				
average sample size	74624			57539
coverage	84%	86%	89%	82%
avg. rel. precision	0.064	0.073	0.075	0.062
avg. CI half length	NA	0.231	0.239	0.196
var. CI half length	NA	0.002	0.002	0.0006

[†]No. of classical and corrected CIs generated by ASAP: 92 and 8, respectively.

Table 5.12: Performance of Batch-Means Procedures for the M/M/1 SIRO Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	33600			6481
coverage	92%	66%	76%	88%
avg. rel. precision	0.095	0.168	0.199	0.264
avg. CI half length	NA	0.534	0.652	0.901
var. CI half length	NA	0.053	0.196	0.800
$\pm 15\%$ PRECISION				
average sample size	33600			32544
coverage	92%	81%	82%	92%
avg. rel. precision	0.095	0.093	0.102	0.115
avg. CI half length	NA	0.299	0.328	0.370
var. CI half length	NA	0.0064	0.0092	0.0062
$\pm 7.5\%$ PRECISION				
average sample size	78464			116925
coverage	92%	86%	88%	86%
avg. rel. precision	0.065	0.054	0.054	0.056
avg. CI half length	NA	0.171	0.180	0.179
var. CI half length	NA	0.0015	0.0017	0.0009

[†]No. of classical and corrected CIs generated by ASAP: 29 and 71, respectively.

Table 5.13: Performance of Batch-Means Procedures for the $E_4/M/1$ Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	20160			8702
coverage	84%	77%	81%	80%
avg. rel. precision	0.107	0.157	0.171	0.185
avg. CI half length	NA	0.274	0.297	0.321
var. CI half length	NA	0.006	0.008	0.020
$\pm 15\%$ PRECISION				
average sample size	20160			19653
coverage	84%	84%	84%	83%
avg. rel. precision	0.107	0.111	0.117	0.119
avg. CI half length	NA	0.197	0.208	0.211
var. CI half length	NA	0.002	0.002	0.002
$\pm 7.5\%$ PRECISION				
average sample size	58752			73155
coverage	86%	91%	93%	89%
avg. rel. precision	0.064	0.062	0.064	0.059
avg. CI half length	NA	0.112	0.116	0.107
var. CI half length	NA	0.0008	0.0009	0.0003

†No. of classical and corrected CIs generated by ASAP: 55 and 45, respectively.

Table 5.14: Performance of Batch-Means Procedures for the M/H₂/1 Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	86144			16716
coverage	88%	64%	70%	76%
avg. rel. precision	0.106	0.201	0.244	0.539
avg. CI half length	NA	1.53	1.91	4.76
var. CI half length	NA	0.683	1.57	43.73
$\pm 15\%$ PRECISION				
average sample size	86144			148820
coverage	88%	84%	86%	88%
avg. rel. precision	0.106	0.081	0.090	0.102
avg. CI half length	NA	0.633	0.709	0.802
var. CI half length	NA	0.042	0.060	0.055
$\pm 7.5\%$ PRECISION				
average sample size	229632			405854
coverage	90%	90%	93%	93%
avg. rel. precision	0.067	0.050	0.053	0.053
avg. CI half length	NA	0.398	0.422	0.421
var. CI half length	NA	0.015	0.018	0.012

[†]No. of classical and corrected CIs generated by ASAP: 24 and 76, respectively.

Table 5.15: Performance of Batch-Means Procedures for the M/M/2 Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP [†]
NO PRECISION				
average sample size	35008			10392
coverage	85%	75%	78%	80%
avg. rel. precision	0.104	0.174	0.201	0.314
avg. CI half length	NA	0.487	0.564	0.985
var. CI half length	NA	0.066	0.108	3.78
$\pm 15\%$ PRECISION				
average sample size	35008			54682
coverage	85%	83%	84%	86%
avg. rel. precision	0.104	0.093	0.100	0.107
avg. CI half length	NA	0.263	0.281	0.302
var. CI half length	NA	0.006	0.008	0.005
$\pm 7.5\%$ PRECISION				
average sample size	102528			189205
coverage	89%	87%	87%	88%
avg. rel. precision	0.065	0.055	0.057	0.055
avg. CI half length	NA	0.154	0.160	0.156
var. CI half length	NA	0.002	0.002	0.001

[†]No. of classical and corrected CIs generated by ASAP: 35 and 65, respectively.

Table 5.16: Performance of Batch-Means Procedures for the M/M/1/M/1 Queue Waiting Time Process with $\tau = 0.8$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	3120			3152
coverage	64%	65%	75%	85%
avg. rel. precision	0.236	0.162	0.222	0.454
avg. CI half length	NA	1.053	1.476	3.250
var. CI half length	NA	0.119	0.585	14.06
$\pm 15\%$ PRECISION				
average sample size	13944			46610
coverage	76%	80%	80%	93%
avg. rel. precision	0.131	0.070	0.074	0.103
avg. CI half length	NA	0.438	0.465	0.649
var. CI half length	NA	0.016	0.018	0.030
$\pm 7.5\%$ PRECISION				
average sample size	49920			117339
coverage	87%	85%	87%	90%
avg. rel. precision	0.065	0.042	0.044	0.050
avg. CI half length	NA	0.266	0.281	0.318
var. CI half length	NA	0.005	0.005	0.008

†No. of classical and corrected CIs generated by ASAP: 6 and 93, respectively.

Table 5.17: Performance of Batch-Means Procedures for the Time-Shared Model
Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	17696			1765
coverage	92%	73%	79%	92%
avg. rel. precision	0.049	0.113	0.139	0.183
avg. CI half length	NA	0.900	1.102	1.468
var. CI half length	NA	0.016	0.050	0.139
$\pm 15\%$ PRECISION				
average sample size	17696			4496
coverage	92%	83%	84%	94%
avg. rel. precision	0.049	0.088	0.095	0.121
avg. CI half length	NA	0.722	0.772	0.985
var. CI half length	NA	0.016	0.021	0.021
$\pm 7.5\%$ PRECISION				
average sample size	17696			18747
coverage	92%	86%	87%	98%
avg. rel. precision	0.049	0.046	0.048	0.061
avg. CI half length	NA	0.380	0.392	0.500
var. CI half length	NA	0.003	0.003	0.003

†No. of classical and corrected CIs generated by ASAP: 0 and 100, respectively.

Table 5.18: Performance of Batch-Means Procedures for the Central Server Model 1
Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	828			1559
coverage	87%	92%	94%	92%
avg. rel. precision	0.050	0.037	0.039	0.037
avg. CI half length	NA	0.222	0.231	0.220
var. CI half length	NA	0.0009	0.0016	0.0009
$\pm 15\%$ PRECISION				
average sample size	828			1559
coverage	87%	92%	94%	92%
avg. rel. precision	0.050	0.037	0.039	0.037
avg. CI half length	NA	0.222	0.231	0.220
var. CI half length	NA	0.0009	0.0016	0.0009
$\pm 7.5\%$ PRECISION				
average sample size	828			1559
coverage	87%	92%	94%	92%
avg. rel. precision	0.050	0.037	0.039	0.037
avg. CI half length	NA	0.222	0.231	0.220
var. CI half length	NA	0.0009	0.0016	0.0009

†No. of classical and corrected CIs generated by ASAP: 77 and 23, respectively.

Table 5.19: Performance of Batch-Means Procedures for the Central Server Model 2
Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	1220			1580
coverage	87%	87%	90%	86%
avg. rel. precision	0.045	0.039	0.040	0.039
avg. CI half length	NA	0.388	0.400	0.387
var. CI half length	NA	0.002	0.005	0.003
$\pm 15\%$ PRECISION				
average sample size	1220			1580
coverage	87%	87%	90%	86%
avg. rel. precision	0.045	0.039	0.040	0.039
avg. CI half length	NA	0.388	0.400	0.387
var. CI half length	NA	0.002	0.005	0.003
$\pm 7.5\%$ PRECISION				
average sample size	1220			1580
coverage	87%	87%	90%	86%
avg. rel. precision	0.045	0.039	0.040	0.039
avg. CI half length	NA	0.388	0.400	0.387
var. CI half length	NA	0.002	0.005	0.003

†No. of classical and corrected CIs generated by ASAP: 67 and 33, respectively.

Table 5.20: Performance of Batch-Means Procedures for the Central Model 3 Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	1152			2277
coverage	87%	75%	79%	78%
avg. rel. precision	0.097	0.073	0.076	0.074
avg. CI half length	NA	1.33	1.40	1.35
var. CI half length	NA	0.107	0.163	0.135
$\pm 15\%$ PRECISION				
average sample size	1152			2277
coverage	87%	75%	79%	78%
avg. rel. precision	0.097	0.073	0.076	0.074
avg. CI half length	NA	1.33	1.40	1.35
var. CI half length	NA	0.107	0.163	0.135
$\pm 7.5\%$ PRECISION				
average sample size	3740			3389
coverage	88%	75%	76%	79%
avg. rel. precision	0.059	0.060	0.062	0.058
avg. CI half length	NA	1.08	1.11	1.05
var. CI half length	NA	0.037	0.047	0.028

†No. of classical and corrected CIs generated by ASAP: 76 and 24, respectively.

Table 5.21: Performance of Batch-Means Procedures for the Central Server Model 4
Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	1496			1553
coverage	92%	88%	89%	91%
avg. rel. precision	0.045	0.041	0.043	0.042
avg. CI half length	NA	0.329	0.344	0.336
var. CI half length	NA	0.002	0.003	0.002
$\pm 15\%$ PRECISION				
average sample size	1496			1553
coverage	92%	88%	89%	91%
avg. rel. precision	0.045	0.041	0.043	0.042
avg. CI half length	NA	0.329	0.344	0.336
var. CI half length	NA	0.002	0.003	0.002
$\pm 7.5\%$ PRECISION				
average sample size	1496			1553
coverage	92%	88%	89%	91%
avg. rel. precision	0.045	0.041	0.043	0.042
avg. CI half length	NA	0.329	0.344	0.336
var. CI half length	NA	0.002	0.003	0.002

†No. of classical and corrected CIs generated by ASAP: 70 and 30, respectively.

Table 5.22: Performance of Batch-Means Procedures for the (s, S) Inventory System
Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	Law and Carson	LBATCH	ABATCH	ASAP†
NO PRECISION				
average sample size	800			1618
coverage	98%	96%	96%	93%
avg. rel. precision	0.013	0.007	0.007	0.007
avg. CI half length	NA	0.824	0.823	0.781
var. CI half length	NA	0.015	0.012	0.014
$\pm 15\%$ PRECISION				
average sample size	800			1618
coverage	98%	96%	96%	93%
avg. rel. precision	0.013	0.007	0.007	0.007
avg. CI half length	NA	0.824	0.823	0.781
var. CI half length	NA	0.015	0.012	0.014
$\pm 7.5\%$ PRECISION				
average sample size	800			1618
coverage	98%	96%	96%	93%
avg. rel. precision	0.013	0.007	0.007	0.007
avg. CI half length	NA	0.824	0.823	0.781
var. CI half length	NA	0.015	0.012	0.014

†No. of classical and corrected CIs generated by ASAP: 33 and 67, respectively.

Table 5.23: Performance of LBATCH and ABATCH Algorithms under a Relative Precision Requirement for 2-state DTMC with High Correlation as Defined by (3.66) and (3.67) Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	LBATCH	ABATCH
NO PRECISION		
average sample size	1536	1536
coverage	68%	87%
avg. rel. precision	0.072	0.113
avg. CI half length	0.536	0.843
var. CI half length	0.002	0.013
$\pm 15\%$ PRECISION		
average sample size	1536	1623
coverage	68%	88%
avg. rel. precision	0.072	0.110
avg. CI half length	.536	0.823
var. CI half length	0.002	0.110
$\pm 7.5\%$ PRECISION		
average sample size	2434	6884
coverage	70%	85%
avg. rel. precision	0.066	0.061
avg. CI half length	0.500	0.464
var. CI half length	0.003	0.005

Table 5.24: Performance of LBATCH and ABATCH Algorithms under a Relative Precision Requirement for $M/M/1$ Queue $\tau = 0.9$ Based on 100 Independent Replications of Nominal 90% Confidence Intervals

Precision Requirement	LBATCH	ABATCH
NO PRECISION		
average sample size	1536	1536
coverage	35%	54%
avg. rel. precision	0.204	0.338
avg. CI half length	1.648	2.882
var. CI half length	0.552	4.250
$\pm 15\%$ PRECISION		
average sample size	34349	50910
coverage	65%	77%
avg. rel. precision	0.121	0.125
avg. CI half length	.1.071	1.080
var. CI half length	0.0513	0.0336
$\pm 7.5\%$ PRECISION		
average sample size	227987	397387
coverage	80%	81%
avg. rel. precision	0.062	0.062
avg. CI half length	0.551	0.553
var. CI half length	0.005	0.007

5.3 Discussion of Results

In this section, we analyze and explain the results of the experimental testing of the ASAP algorithm and compare its results with those of Law and Carson's (1979) procedure and LBATCH and ABATCH.

5.3.1 Discrete Time Markov Chains

These cases were not part of Law and Carson's (1979) work. Therefore this discussion includes results from only our testing of ASAP, LBATCH and ABATCH.

ASAP, LBATCH and ABATCH all performed well on both the 2-state DTMC (3.65) with alternating positive and negative correlations and the 8-state DTMC (3.68) with the correlation function that resembles a damped sine wave. No additional batches were required in these cases to satisfy the precision requirements, so the final number of batches used was 96 in all replications of all cases. For both models in the majority of the replications of ASAP (74 replications using the 2-state chain and 70 replications using the 8-state chain), this procedure constructed CIs based on a nonsignificant result from the independence test (that is, the batch means passed the test for independence, described in Section 2.2.2). This is what we expect since, as can be seen from Figures 3.2 and 3.16, the lag-1 correlations of the batch means become negligible at small batch sizes. These cases may be ones in which a smaller initial batch size would be adequate.

ASAP showed somewhat better coverage than LBATCH in the case of the 2-state DTMC (3.67) with high positive correlation, especially in the cases of no precision requirement and $\pm 15\%$ precision requirement. For this model, ASAP constructed adjusted CIs based on a nonsignificant result on the test for multivariate normality (that is, the batch means passed the Shapiro-Wilk test for multivariate normality) in all 100 cases. The CIs from ASAP are wider than those from LBATCH and ABATCH, which is necessary for the improved coverage. However, the coefficient of variation of the CI half-lengths are smaller than those from LBATCH and ABATCH.

When LBATCH and ABATCH were repeatedly run until their final CIs met the same precision requirements used for ASAP, coverages from the LBATCH and ABATCH algorithms were slightly degraded below those experienced when the sample size was first determined by ASAP. See Table 5.23 for details. Sample sizes were smaller than those determined by ASAP.

5.3.2 Autoregressive and Exponential Autoregressive Processes

Again, we cannot include Law and Carson's procedure in this discussion since these cases were not included in their study.

The three algorithms, ASAP, LBATCH and ABATCH, all performed well on both the AR(1) and EAR(1) processes with $\varphi = 0.9$. These processes have the same correlation function. ASAP constructed adjusted CIs based on a nonsignificant result for multivariate normality test in 98 replications of the AR(1) process and in 90 replications of the EAR(1) process. In the case of no precision requirement, the average sample size is smaller for the AR(1) process than for the EAR(1), signifying that multivariate normality was accepted at smaller batch sizes for the AR(1) process as compared to the EAR(1) process. This is what we expect based on our knowledge of the shapes of the marginal distributions of the two processes.

5.3.3 Queueing Systems

As stated previously the waiting time process in the $M/M/1$ queue with $\tau = 0.9$ is a difficult case in that the lag-1 correlation of the observations is close to one, the correlation function decays slowly, and the marginal distribution of X_i has an exponential tail. Because of these characteristics, we can expect slow convergence to both conditions for ensuring the validity of the batch means method, i.e., that the batch means are independent and identically normally distributed. Figure 3.23 makes clear that large batch sizes are required before the batch means will appear independent. This case most dramatically displays one of the advantages of the ASAP algorithm, i.e., it does not rely solely on the von Neumann test for independence. In fact, in 96 replications, ASAP constructed adjusted CIs based on a nonsignificant result from the test for multivariate normality.

As can be seen from Table 5.8, ASAP substantially outperforms LBATCH and ABATCH for the case of no precision requirement. (Law and Carson did not include the $M/M/1$ queue with $\tau = 0.9$ in their work.) As we demand more precision, we are of course forced to perform more sampling. At the precision requirement of $\pm 7.5\%$ the three algorithms give similar results. This implies that LBATCH and ABATCH will give satisfactory results if supplied with an adequate amount of data. However, LBATCH and ABATCH provide no mechanism for assessing the amount

of data which should be used. This emphasizes a desirable feature of ASAP—that it determines a sample size that gives acceptable results, even when no precision is specified.

From Table 5.24 we see that in the $M/M/1$ queue with $\tau = 0.9$, if LBATCH and ABATCH are run until a certain precision requirement is met, coverage is severely degraded, especially when the precision requirement is so “loose” that it leads to relatively little additional sampling. Note that the sample sizes are much smaller than those required by ASAP to achieve the same precision. For example, the average sample size used by ABATCH for the waiting time process in the $M/M/1$ queue with utilization $\tau = 0.9$ and $\pm 7.5\%$ precision is approximately 397,387. This is considerably less than the average sample size of 815,755 used by ASAP. In order to get some idea of what sample size should be required to construct a nominal 90% CI having $\pm 7.5\%$ precision, we used Whitt’s (1989) approximation for estimating required run lengths of queueing simulations. Whitt (1989) contends that his heuristic is “quite accurate for high traffic intensities, e.g., $\tau \geq 0.8$ ”. For $\pm 7.5\%$ precision and 90% confidence, Whitt’s (1989) approximation yields an estimated sample size of 855,238 for the waiting time process in the $M/M/1$ queue with $\tau = 0.9$. Applying Whitt’s heuristic to the queueing systems selected for the current experimental performance evaluation strongly suggests that ASAP yields adequate sample sizes when a precision requirement is specified.

As the utilization τ in the $M/M/1$ queue decreases, fewer ASAP CIs are based on a nonsignificant result from the multivariate normality test. In the case of $\tau = 0.8$, 67 replications constructed CIs based on a nonsignificant result from the multivariate normality test; while for $\tau = 0.5$, only 13 replications built adjusted CIs based on multivariate normality. In cases where more ASAP CIs are based on assumed independence of the batch means, we see similar results for ASAP, LBATCH and ABATCH. In the case of no precision requirement for both the models where $\tau = 0.5$ and $\tau = 0.8$, Tables 5.9 and 5.10 and show that ASAP is using smaller sample sizes as compared to Law and Carson’s procedure. ASAP shows some under-coverage in the case of $\tau = 0.5$ with no precision requirement. In the case of a $\pm 15\%$ precision requirement and $\tau = 0.5$, ASAP’s coverage is slightly worse than for the same precision requirement and $\tau = 0.8$. This leads us to believe that ASAP might give better coverage in more dependent systems. Only in the case where $\tau = 0.8$ and the precision requirement is $\pm 7.5\%$ does ASAP require significantly more data than Law

and Carson's procedure.

ASAP, Law and Carson's procedure, LBATCH, and ABATCH performed similarly on the $M/M/1$ LIFO queue with $\tau = 0.8$, showing some evidence of under-coverage at the no precision and $\pm 15\%$ precision requirements. All the algorithms showed adequate coverage at the $\pm 7.5\%$ precision requirement. Here in 98 replications, CIs from ASAP were based on a nonsignificant result on the independence test.

For the case of $M/M/1$ SIRO queue with $\tau = 0.8$, ASAP and Law and Carson's procedure displayed better coverages than LBATCH and ABATCH in the cases with no precision requirement and with $\pm 15\%$ precision.

The most noticeable difference between ASAP and Law and Carson's procedure in the cases of the $E_4/M/1$ queue with $\tau = 0.8$ and the $M/M/2$ queue $\tau = 0.8$ is in the sample sizes required for the no-precision case. ASAP required far less sampling than Law and Carson's procedure. The coverage results were similar.

Results for the $M/H_2/1$ queue with $\tau = 0.8$ show that Law and Carson's procedure has better coverage than ASAP, LBATCH, and ABATCH in the case of no precision. However, Law and Carson's procedure used more sampling. When a precision requirement is specified, then ASAP, LBATCH, and ABATCH perform as well as the Law and Carson procedure, but of course they also require more sampling.

ASAP performed better at the three different precision requirements than the other three algorithms in the case of the $M/M/1/M/1$ queue with $\tau = 0.8$. For the case of $\pm 7.5\%$ precision, Law and Carson's procedure, LBATCH, and ABATCH all experienced coverages close to nominal.

5.3.4 Computer Models

In the case of the time-shared computer model, ASAP exhibits almost ideal behavior. The sample size for the no-precision case is small, but coverage is acceptable. With increasingly stringent precision requirements, the CIs become smaller and less variable and the coverage improves. For this model, ASAP built adjusted CIs based on a nonsignificant result from the test for multivariate normality in all 100 replications. Again, at the less stringent precision requirements Law and Carson's procedure requires much more sampling than ASAP does.

All four algorithms displayed good coverages for central server models 1, 2, and 4. ASAP and Law and Carson's procedure both achieved the $\pm 15\%$ and the $\pm 7.5\%$

precision requirements without increasing the sample size over that which was used for the no-precision case. ASAP required a larger average sample size for model 1 than did Law and Carson’s procedure. The average sample size for ASAP was 1550, which is only slightly larger than the minimum initial sample size of 1536. This implies that this is a case where the initial sample size required by ASAP may be more than is really necessary. We feel that the benefits seen in other cases from starting with an initial sample size of 1536 outweigh the relatively small inconvenience of possible oversampling in cases such as this. Although in a few cases an initial sample size of 1536 may be more than necessary, in practice it is a modest sample size.

The central server model 3 is the only case where Law and Carson’s procedure seems to significantly outperform ASAP. The under-coverages by ASAP here are a concern but not really to be viewed as catastrophic. We also ran a case requesting $\pm 2\%$ precision and achieved 85% coverage for the CIs constructed by ASAP. Law and Carson’s procedure achieved near nominal coverage for the three original precision requirements in this case. LBATCH and ABATCH perform similarly to ASAP. It is not entirely clear what conclusions can be drawn from Table 5.20.

5.3.5 Inventory Model

For the (s, S) inventory system no additional sampling above what was taken for the no-precision case was required for the precision requirements in both the ASAP and Law and Carson’s procedures. All four algorithms gave good coverage results. ASAP required slightly more than its initial sample size and about twice what Law and Carson’s procedure required. Even Law and Carson’s minimal sample size of 800 seems to be more than enough since their coverage was 98%. This may be another case where ASAP is requiring more data than is necessary, but here again an average sample size of 1618 is in no way unreasonable.

Chapter 6

Conclusions and Recommendations

6.1 Main Conclusions of the Research

In Chapter 1 we discussed the need for an automated method of analyzing simulation output. Such a procedure should yield a valid confidence interval for the steady-state expected response that is centered on the sample mean of a portion of the series of simulation-generated responses and should satisfy a user-specified absolute or relative precision requirement. In Chapter 2 we surveyed the main methods for steady-state simulation output analysis; and we reviewed the theory of *weakly dependent* processes, explaining how this theory is employed in constructing CIs with the nonoverlapping batch means method.

In Chapter 3, we established a central limit theorem that characterizes the asymptotic distribution of the vector of nonoverlapping batch means computed from a stationary ϕ -mixing process when the number of batches is fixed and the batch size tends to infinity. We also analyzed the convergence properties of the numerator and squared denominator of the classical NOBM t -ratio, and we formulated moment conditions on these components which we proposed as surrogates for the distributional requirements necessary to ensure the validity of the classical NOBM confidence interval. Using graphical displays, we examined the behavior of the batch means method in six selected stochastic processes, including both theoretical convergence to the prescribed moment conditions as well as empirical convergence to the associated limiting distributions of the two components of the NOBM t -ratio. Finally in Chapter 3 we derived a matrix-analytic expression for the steady-state variance constant, σ^2 , for a

process defined as a function of a finite-state, discrete-time Markov chain.

In Chapter 4 we formulated ASAP, a procedure for constructing confidence intervals on the expected response of a steady-state simulation. ASAP is based on sequential testing for independence or multivariate normality of the batch means. This algorithm qualifies as a fixed-number-of-batches method since the number of batches remains constant and the batch size is increased until a nonsignificant result is obtained from one of these tests. If the batch means pass the test for independence on some iteration of the procedure, then ASAP delivers a classical NOBM confidence interval. If the batch means fail the test for independence but pass the test for multivariate normality on some iteration of the procedure, then a correction is made to the classical NOBM confidence interval to compensate for any dependence of the batch means. A detailed derivation of this correction is given in Chapter 4, including

- a) calculation of the cumulants of the NOBM t -ratio under certain simplifying assumptions based on approximate multivariate normality of the batch means;
- b) use of these cumulants to calculate corrected limits for the classical NOBM confidence interval based on an inverted Cornish-Fisher expansion for the NOBM t -ratio; and
- c) estimation of these cumulants by fitting an autoregressive-moving average time series model to the sequence of batch means.

After an adequate batch size has been determined via sequential testing for independence or multivariate normality of the batch means, ASAP computes the number of additional batches needed to satisfy any user-specified absolute or relative precision requirement on the final confidence interval.

Chapter 5 includes descriptions of 20 stochastic systems that we used in the experimental performance evaluation to compare ASAP and other batch-means procedures. Tables of test results display the performance of ASAP together with the LBATCH and ABATCH algorithms; and in 14 cases the performance of Law and Carson's procedure is also summarized. In general, the results of the experimentation show that ASAP performs well in a variety of problems and compares favorably with the other algorithms that were tested. In particular, ASAP seems to outperform the other algorithms in processes that have a high degree of stochastic dependence as represented by the correlation structure of the process.

Batching schemes to date have ignored the question of normality based on the assumption that if the batch size is large enough for the batch means to be approximately independent, then the batch size is large enough for the batch means to be approximately normally distributed. These schemes have focused on selecting a batch size large enough to achieve near independence of the batch means. The method of determining whether the batch means are independent varies from scheme to scheme. LBATCH and ABATCH, for instance, rely on the von Neumann test for independence. On the other hand, Law and Carson's procedure basically relies on estimating the correlation between batches and on assuming the batch means are independent if the magnitude of the correlation is less than some arbitrary value. (The procedure of Law and Carson is more complex than this simplistic description, but this is the essence of their method.) ASAP is the first method to recognize the frequently occurring phenomenon of approximate multivariate normality being achieved at smaller batch sizes than approximate independence; and ASAP exploits this phenomenon when it is detected so as to compensate for any remaining dependence between the batch means.

Our experimentation with ASAP indicates that it may occasionally require larger-than-necessary sample sizes. In most of these cases, the average final sample sizes are close to the initial sample size of 1536. This suggests that in these cases an initial sample size smaller than 1536 might be adequate. However, a sample size of 1500 to 1600 is usually easy and inexpensive to generate; and we feel that the good results achieved in the more difficult cases justify the default initial sample size required by ASAP.

The main advantage of ASAP is that it performs well in highly dependent processes. In these cases, ASAP requires sample sizes that are effective in achieving adequate CI coverage but that are not excessively large. Another important feature of ASAP is that it does not rely entirely on the requirement that the batch means be approximately independent. In particular, ASAP is not based wholly on a nonsignificant result from the von Neumann test for independence, which lacks adequate power for detecting dependence in small samples. However, the independence test in ASAP is based on a sample of 94 batch means, which should be a large enough sample to give the von Neumann test adequate power to detect significant dependence between the batch means. In many cases the CI constructed via ASAP is based on a nonsignificant result from the Shapiro-Wilk test for multivariate normality. Although we

know no way of comparing the power of the von Neumann test for randomness with the power of the Shapiro-Wilk test for multivariate normality, we know the univariate Shapiro-Wilk test is fairly sensitive to departures from normality (Shapiro, Wilk and Chen 1968).

6.2 Recommendations for Future Research

While the results of our experimentation with ASAP are generally good, there remain some unresolved issues. For instance, this research has concentrated on stationary ϕ -mixing processes, and the theory that supports the algorithm is based on this property. There are stationary processes which are known not to be ϕ -mixing—for example, certain infinite-state Markov chain models of a one-dimensional random walk that are strong mixing but not ϕ -mixing (Glynn 1985); and yet there is substantial experimental evidence to show that ASAP performs well on such strong mixing processes. If we can establish an analogue of Theorem 3.1 for processes that are strong mixing, then we will substantially expand the class of stochastic systems to which ASAP may be applied.

Another avenue of future research is in the area of possible improvements to the ASAP algorithm. Results of the experimentation presented in Chapter 5 indicate that when ASAP constructs CIs based on a nonsignificant result from the multivariate normality test, coverages are close to or better than nominal. In comparing ASAP's results with those of the Law and Carson procedure as well as the LBATCH and ABATCH algorithms, we found that when ASAP delivers a classical NOBM confidence interval based on a nonsignificant result from the independence test, the resulting coverage probability is slightly lower than that achieved with the other algorithms. Possible topics related to future improvements in ASAP include

- modification of the batching scheme geared to improving coverage when the CI is based on the acceptance of the hypothesis of independence of the batch means;
- formulation of alternative tests or correlation-detection procedures to replace or augment the von Neumann test for independence;
- development of an improved scheme for identifying the best-fitting ARMA representation of the batch-means process;

- investigation of alternative schemes (such as spectral analysis) for estimating the terms of the inverted Cornish-Fisher expansion;
- extension of ASAP to handle multiple output processes; and
- extension of ASAP to handle continuous-time processes.

Bibliography

- Alexopoulos, C., and A. F. Seila. 1996. Implementing the batch means method in simulation experiments. In *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 214–221. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Anderson, T. W. 1971. *The statistical analysis of time series*. New York: John Wiley & Sons, Inc.
- Billingsley, P. 1968. *Convergence of probability measures*. New York: John Wiley & Sons, Inc.
- Box, G. E. P., and G. M. Jenkins. 1976. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, Inc.
- Bradley, R. C. 1986. Basic properties of strong mixing conditions. *Dependency in probability and statistics*, ed. Eberlein, E. and M. S. Taquu, Boston: Birkhauser.
- Bratley, P., B. L. Fox, and L. E. Schrage. 1987. *A guide to simulation*. 2nd ed. New York: Springer-Verlag.
- Brillinger, D. R. 1973. Estimation of the mean of a stationary time series by sampling. *Journal of Applied Probability* 10:419–431.
- Carlstein, E. 1984. Asymptotic normality for a general statistic from a stationary sequence. *Annals of Probability* 14 (4): 1371–1379.
- Chien, C. 1989. Small sample theory for steady state confidence intervals. Technical Report No. 37, Department of Operations Research, Stanford University, Stanford, California. U.S. Army Research Contract DAAL-**-K-0063
- Chow, Y. S., and H. Robbins. 1965. On the asymptotic theory of fixed-width sequential confidence for the mean. *Annals of Mathematical Statistics* 36:457–462.
- Damerdji, H. 1991. Strong consistency and other properties of the spectral variance estimator. *Management Science* 35 (11): 1424–1440.

- Fishman, G. S. 1971. Estimating sample size in computer simulation experiments. *Management Science* 18:21–38.
- Fishman, G. S. 1973. *Concepts and methods in discrete event digital simulation*. New York: John Wiley.
- Fishman, G. S. 1977. Achieving specific accuracy in simulation output analysis. *Communications of the Association for Computing Machinery* 20:310–315.
- Fishman, G. S. 1978a. Grouping observations in digital simulation. *Management Science* 24:510–521.
- Fishman, G. S. 1978b. *Principles of discrete event simulation*. New York: John Wiley.
- Fishman, G. S. 1996. *Monte Carlo: Concepts, algorithms, and applications*. New York: Springer-Verlag.
- Fishman, G. S. 1998. LABATCH.2: Software for statistical analysis of simulation sample path data. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 131–139. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Fishman, G. S., and L. S. Yarberrry. 1997. An implementation of the batch means method. *INFORMS Journal on Computing* 9 (3): 296–310.
- Glynn, P. W. 1984. Some asymptotic formulas for Markov chains with applications to simulation. *Journal of Statistical Computation and Simulation* 19:97–112.
- Glynn, P. W. 1985. Regenerative processes and the ϕ -mixing property. *Newsletter of the TIMS College on Simulation and Gaming* 9 (1): 6–7. Providence, Rhode Island: The Institute of Management Sciences.
- Glynn, P. W., and D. L. Iglehart. 1990. Simulation output analysis using standardized time series. *Mathematics of Operations Research* 15 (1): 1–16.
- Glynn, P. W., and W. Whitt. 1992. The asymptotic validity of sequential stopping rules. *Annals of Applied Probability* 2:180–198.
- Hadley, G. 1961. *Linear algebra*. Reading, Massachusetts: Addison-Wesley.
- Hall, P. 1983. Inverting an Edgeworth expansion. *Annals of Statistics* 11 (2): 560–576.
- Hanselman, D., and B. Littlefield. 1997. *The student edition of MATLAB: Version 5, user's guide/The MathWorks, Inc.* Upper Saddle River, New Jersey: Prentice-Hall.

- Hazen, G. B., and A. A. B. Pritsker. 1980. Formulas for the variance of the sample mean in finite state Markov processes. *Journal of Statistical Computation and Simulation* 12:25–40.
- Heidelberger, P., and P. D. Welch. 1981a. A spectral method for confidence interval generation and run length control in simulations. *Communications of the Association for Computing Machinery* 24:233–245.
- Heidelberger, P., and P. D. Welch. 1981b. Adaptive spectral methods for simulation output analysis. *IBM Journal of Research and Development* 25:860–876.
- Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in simulations. *Operations Research* 31:1109–1144.
- IMSL Problem-Solving Software Systems. 1987. *User's manual: STAT/LIBRARY*. Houston, TX: IMSL.
- Johnson, N. J. 1978. Modified t test and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association* 73 (363): 536–544.
- Karlin, S. 1966. *A first course in stochastic processes*. New York: Academic Press.
- Kemeny, J.G., and J.L. Snell. 1960. *Finite Markov chains*. Princeton, New Jersey: D. Van Nostrand and Co.
- Kendall, M., A. Stuart, and J.K. Ord. 1987. *Kendall's advanced theory of statistics*. New York: Oxford University Press.
- Lavenberg, S. S., and C. H. Sauer. 1977. Sequential stopping rules for regenerative method of simulation. *IBM Journal of Research and Development* 21:545–558.
- Law, A. M., and J. S. Carson. 1979. A sequential procedure for determining the length of a steady-state simulation. *Operations Research* 27 (5): 1011–1025.
- Law, A. M., and W. D. Kelton. 1991. *Simulation modeling and analysis*. 2nd ed. New York: McGraw-Hill.
- Malkovich, J. F., and A. A. Afifi. 1973. On tests for multivariate normality. *Journal of the American Statistical Association* 68:176–179.
- Meketon, M. S., and B. Schmeiser. 1984. Overlapping batch means: something for nothing? In *Proceedings of the 1984 Winter Simulation Conference*, ed. S. Sheppard, U. Pooch, and D. Pegden, 227–230. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Nádas, A. 1969. An extension of a theorem of Chow and Robbins on sequential confidence intervals for the mean. *Annals of Mathematical Statistics* 40 (2): 667–671.

- Nakayama, M. K. 1994. Two-stage stopping procedures based on standardized time series. *Management Science* 40:1889–1206.
- Priestley, M. B. 1981. *Spectral analysis and time series*. New York; London: Academic Press.
- Royston, J. P. 1982. An extension of Shapiro and Wilk’s W test for normality to large samples. *Applied Statistics* 31 (2): 115–124.
- Sargent, R. G., K. Kang, and D. Goldsman. 1992. An investigation of finite-sample behavior of confidence interval estimators. *Operations Research* 40:898–913.
- Schmeiser, B. 1982. Batch size effects in the analysis of simulation output. *Operations Research* 30:556–568.
- Schruben, L. W. 1983. Confidence interval estimation using standardized time series. *Operations Research* 31:1090–1108.
- Shapiro, S. S., M. B. Wilk, and H. J. Chen. 1968. A comparative study of various tests for normality. *American Statistical Association Journal* 63:1343–1372.
- Shedler, G. S. 1993. *Regenerative stochastic simulation*. Boston: Academic Press.
- Song, W. T., and B. W. Schmeiser. 1995. Optimal mean-squared-error batch sizes. *Management Science* 41:110–123.
- Tew, J. D., and J. R. Wilson. 1992. Validation of simulation analysis methods for the Schruben-Margolin correlation-induction strategy. *Operations Research* 40 (1): 87–103.
- von Neumann, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* 12:367–395.
- Welch, P. D. 1983. The statistical analysis of simulation results. In *The computer performance modeling handbook*, ed. S. S. Lavenburg, 268–328. New York: Academic Press.
- Whitt, W. 1989a. Planning queueing simulations. *Management Science* 35:1341–1366.
- Whitt, W. 1989b. Simulation run length planning. In *Proceedings of the 1989 Winter Simulation Conference*, ed. E. A. MacNair, K. J. Musselman, P. Heidelberger, 106–112. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.
- Wilson, J. R. 1979. Variance reduction techniques for the simulation of queueing networks. Ph.D. dissertation, Purdue University, West Lafayette, Indiana.

Appendix A

Derivation of Mixed Moments of the Batch Means

If we assume that the vector of batch means $(Y_1, Y_2, \dots, Y_k)$ has a multivariate normal distribution with null mean vector $\mathbf{0}_k$ and $k \times k$ covariance matrix $\Sigma = [\sigma_{ij}]$ so that $\sigma_{ij} = \text{Cov}(Y_i, Y_j)$ for $1 \leq i, j \leq k$, then their joint moment generating function is given by

$$M(\boldsymbol{\theta}) = \exp\left(\frac{1}{2}\boldsymbol{\theta}\Sigma\boldsymbol{\theta}^T\right) \quad \text{for all } \boldsymbol{\theta} \in \Re^k. \quad (\text{A.1})$$

We may obtain mixed moments of the form $E[Y_i^a Y_j^b Y_u^c Y_\ell^d]$ by taking partial derivatives of the moment generating function (A.1) and evaluating the resulting expressions at $\boldsymbol{\theta} = \mathbf{0}_k$. For $\ell = 1, \dots, k$, the first order partial derivative is

$$\frac{\partial M}{\partial \theta_\ell} = \underbrace{\left(\sum_{i=1}^k \theta_i \sigma_{i\ell}\right)}_{L_\ell(\boldsymbol{\theta})} \underbrace{\left(\exp\left[\frac{1}{2}\boldsymbol{\theta}\Sigma\boldsymbol{\theta}^T\right]\right)}_{M(\boldsymbol{\theta})}. \quad (\text{A.2})$$

Differentiating again with respect to θ_ℓ gives the second order partial derivative

$$\frac{\partial^2 M}{\partial \theta_\ell^2} = L_\ell \frac{\partial M}{\partial \theta_\ell} + M \frac{\partial L_\ell}{\partial \theta_\ell} = L_\ell^2 M + \sigma_{\ell\ell} M, \quad (\text{A.3})$$

where $\sigma_{\ell\ell}$ is the variance of the batch mean Y_ℓ . Similarly, we can obtain the 3rd and 4th order partial derivatives:

$$\frac{\partial^3 M}{\partial \theta_\ell^3} = L_\ell^3 M + 3\sigma_{\ell\ell} L_\ell M \quad (\text{A.4})$$

and

$$\frac{\partial^4 M}{\partial \theta_\ell^4} = L_\ell^4 M + 6\sigma_{\ell\ell} L_\ell^2 M + 3\sigma_{\ell\ell}^2 M. \quad (\text{A.5})$$

Additional terms involving partial derivatives of the moment generating function which are necessary to evaluate $E[S_{n,k}^2]$ follow.

The second order mixed partial derivatives of $M(\boldsymbol{\theta})$ are given by

$$\frac{\partial^2 M}{\partial \theta_u \partial \theta_\ell} = L_\ell L_u M + \sigma_{\ell u} M \quad \text{for } 1 \leq \ell, u \leq k. \quad (\text{A.6})$$

The third order mixed partial derivatives of $M(\boldsymbol{\theta})$ are given by

$$\frac{\partial^3 M}{\partial \theta_u^2 \partial \theta_\ell} = L_\ell L_u^2 M + \sigma_{uu} L_\ell M + 2\sigma_{\ell u} L_u M \quad (\text{A.7})$$

and

$$\frac{\partial^3 M}{\partial \theta_j \partial \theta_u \partial \theta_\ell} = L_j L_u L_\ell M + \sigma_{uj} L_l M + \sigma_{\ell j} L_u M + \sigma_{\ell u} L_j M$$

for $1 \leq u, j, \ell \leq k$.

The fourth order mixed partial derivatives of $M(\boldsymbol{\theta})$ are given by

$$\begin{aligned} \frac{\partial^4 M}{\partial \theta_u^3 \partial \theta_\ell} &= L_u^3 L_\ell M + 3\sigma_{\ell u} L_u^2 M + 3\sigma_{uu} L_\ell L_u M + 3\sigma_{uu} \sigma_{\ell u} M \\ \frac{\partial^4 M}{\partial \theta_u^2 \partial \theta_\ell^2} &= L_u^2 L_\ell^2 M + \sigma_{\ell\ell} L_u^2 M + \sigma_{uu} L_\ell^2 M + 4\sigma_{\ell u} L_u L_\ell M + \sigma_{uu} \sigma_{\ell\ell} M + 2\sigma_{\ell u}^2 M \\ \frac{\partial^4 M}{\partial \theta_u^2 \partial \theta_\ell \partial \theta_j} &= L_u^2 L_\ell L_j M + \sigma_{\ell j} L_u^2 M + 2\sigma_{uj} L_\ell L_u M + \sigma_{uu} \sigma_{\ell j} M + \sigma_{uu} L_\ell L_j M \\ &\quad + 2\sigma_{\ell u} \sigma_{uj} M + 2\sigma_{lu} L_u L_j M \\ \frac{\partial^4 M}{\partial \theta_i \partial \theta_j \partial \theta_u \partial \theta_\ell} &= L_i L_j L_u L_\ell M + \sigma_{ij} L_u L_\ell M + \sigma_{iu} L_j L_\ell M + \sigma_{il} L_j L_u M + \sigma_{ju} L_i L_\ell M \\ &\quad + \sigma_{il} \sigma_{ju} M + \sigma_{j\ell} L_i L_u M + \sigma_{j\ell} \sigma_{iu} M + \sigma_{ij} \sigma_{ul} M + \sigma_{ul} L_i L_j M \end{aligned} \quad (\text{A.8})$$

for $1 \leq i, j, u, \ell \leq k$. At $\boldsymbol{\theta} = \mathbf{0}$, $M = 1$ and $L_\ell = 0$ for $\ell = 1, \dots, k$. Therefore evaluating the partial derivatives at $\boldsymbol{\theta} = \mathbf{0}$ yields the following values for the mixed

moments:

$$\begin{aligned}
\frac{\partial M}{\partial \theta_\ell}(\mathbf{0}) &= \mathbb{E} Y_\ell = 0 \\
\frac{\partial^2 M}{\partial \theta_\ell^2}(\mathbf{0}) &= \mathbb{E} Y_\ell^2 = \sigma_{\ell\ell} \\
\frac{\partial^3 M}{\partial \theta_\ell^3}(\mathbf{0}) &= \mathbb{E} Y_\ell^3 = 0 \\
\frac{\partial^4 M}{\partial \theta_\ell^4}(\mathbf{0}) &= \mathbb{E} Y_\ell^4 = 3\sigma_{\ell\ell}^2 \\
\frac{\partial^2 M}{\partial \theta_u \partial \theta_\ell}(\mathbf{0}) &= \mathbb{E} Y_u Y_\ell = \sigma_{u\ell} \\
\frac{\partial^3 M}{\partial \theta_u^2 \partial \theta_\ell}(\mathbf{0}) &= \mathbb{E} Y_u^2 Y_\ell(\mathbf{0}) = 0 \\
\frac{\partial^3 M}{\partial \theta_u \partial \theta_\ell \partial \theta_j}(\mathbf{0}) &= \mathbb{E} Y_u Y_\ell Y_j = 0 \\
\frac{\partial^4 M}{\partial \theta_u^3 \partial \theta_\ell}(\mathbf{0}) &= \mathbb{E} Y_u^3 Y_\ell = 3\sigma_{uu}\sigma_{u\ell} \\
\frac{\partial^4 M}{\partial \theta_u^2 \partial \theta_\ell^2}(\mathbf{0}) &= \mathbb{E} Y_u^2 Y_\ell^2 = \sigma_{uu}\sigma_{\ell\ell} + 2\sigma_{u\ell}^2 \\
\frac{\partial^4 M}{\partial \theta_u^2 \partial \theta_\ell \partial \theta_j}(\mathbf{0}) &= \mathbb{E} Y_u^2 Y_\ell Y_j = \sigma_{uu}\sigma_{\ell j} + 2\sigma_{u\ell}\sigma_{uj}. \\
\frac{\partial^4 M}{\partial \theta_u \partial \theta_\ell \partial \theta_j \partial \theta_i}(\mathbf{0}) &= \mathbb{E} Y_u Y_\ell Y_j Y_i = \sigma_{i\ell}\sigma_{uj} + \sigma_{ui}\sigma_{\ell j} + \sigma_{ji}\sigma_{u\ell}
\end{aligned} \tag{A.9}$$

for $1 \leq i, j, u, \ell, \leq k$.