1993

Fourier analysis of frequency domain discrete event simulation experiments

Mousumi Mitra Hazra
College of William & Mary - Arts & Sciences

Order Number 9429675

Fourier analysis of frequency domain discrete-event simulation experiments

Hazra, Mousumi Mitra, Ph.D.

The College of William and Mary, 1993

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

FOURIER ANALYSIS OF FREQUENCY DOMAIN DISCRETE-EVENT SIMULATION EXPERIMENTS

A Dissertation

Presented to

The Faculty of the Department of Computer Science

The College of William and Mary in Virginia

In Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

---

by

MOUSUMI MITRA HAZRA

1993

APPROVAL SHEET

This dissertation is submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

Mousumi Hazra

Author

Approved, November 1993

Stephen Park

Stephen K. Park

David Nicol

David M. Nicol

Rahul Simha

William L. Bynum

Lawrence Leemis

Lawrence L. Leemis

Department of Mathematics

To my daughter Dia,

my husband Rajeeb,

and my parents.

Table of Contents

Acknowledgements

I would first like to thank my dissertation advisor, Dr. Stephen K. Park. I am grateful to Dr. Park for his advice and support. Not only did he help me with my research, but his valuable advice helped my husband and me in making many important decisions in our lives. I shall always be indebted to Dr. Park and am honored to have had the opportunity of working with him.

I also would like to thank Dr. David M. Nicol, Dr. Rahul Simha, Dr. Lawrence M. Leemis and Dr. William Bynum for serving on my committee. Special thanks to Dr. Ernest E. Armstrong and Dr. Douglas Morrice—they were always very prompt in offering their help and advice.

Special thanks to Jean Bolot at INRIA, France. In the summer of 1992, after trying in vain for several months to obtain a solution for the probability of a free server ($P_0(t)$) in a non-stationary $M/M/1$ queue, I posted an e-mail to the Internet news group *comp.simulation*, enquiring if anyone knew of papers dealing with similar problems. It was a list of papers sent by Mr. Bolot that led me to the paper by Rider [32], which finally enabled me to obtain the results presented in chapter 5 and helped me to complete my dissertation.

I would like to thank Dr. Daniel D. Moerder, Dr. Douglas Price and my co-workers at NASA Langley Research Center's Space Craft Controls Branch for their support and encouragement when I was not making headway in my research.

My husband Rajeeb deserves special thanks for his support and understanding

throughout my dissertation, specially the period that he refers to as the "$P_0(t)$ days" when I was trying in vain to find an analytical solution for $P_0(t)$. Also, without his helpful editing comments, this dissertation would not have been completed. My thanks to my daughter Dia — although she made life harder during this period, her *joie de vie* gave me the required boost to continue with my research. I would like to thank my parents for their love and support throughout my life. Thanks to my father's encouragements I continued my education till the end; and without my mother's endless sacrifices for me I would not be where I am today.

Finally I want to thank Sree Ma of Sree Aurobindo Ashram, Pondicherry, India for always being with me and helping me through every moment of my life. I would like to end with one of her quotes that is my inspiration in life, "Forward ever forward, At the end of the tunnel is the Light, At the end of the Light is the Victory".

List of Tables

List of Figures

ABSTRACT

Frequency Domain Experiments (FDEs) were first used in discrete-event simulation to perform system parameter sensitivity analysis for factor screening in stochastic system simulations. FDEs are based on the intuitive assertion that if one or more system parameters are oscillated at fixed frequencies throughout a simulation run, then oscillations at the same frequencies will be induced in the system's response. Spectral (Fourier) analysis of these induced oscillations is then used to characterize and analyze the system. Since their introduction 12 years ago, significant work has been done to extend the applicability of FDEs to regression analysis, simulation optimization and gradient estimation. Two fundamental theoretical and data analysis FDE problems remain, however. Both problems are addressed in this dissertation.

To perform a FDE Fourier analysis, a sampled data sequence of response observations is used; i.e., the selected system response is sampled using a suitable oscillation (sampling) index. The choice of an appropriate oscillation index is an open problem in the literature known as the *FDE indexing problem.* This dissertation presents a solution to the FDE indexing problem. Specifically, a FDE Fourier data analysis algorithm is developed which uses the simulation clock as the oscillation index. This algorithm is based on the well-established theory of counting (Poisson) processes. The algorithm is implemented and tested on a variety of systems including several networks of nonstationary $M/G/1$ queues.

To justify the use of Fourier methods, a basic FDE model *assumption* is that if a particular system response statistic is sensitive to a system parameter, then sinusoidal variation of that system parameter at a fixed frequency will induce similar sinusoidal variations in the response statistic, at the same frequency. There is, however, a lack of theoretical support for this model assumption. This dissertation provides some of that theoretical support; i.e., the FDE Fourier data analysis algorithm developed in this dissertation is used to analyze the frequency response of a $M/M/1$ queuing system. An equation is derived which accurately characterizes the extent to which the departure process from a $M/M/1$ queuing system can be modeled as an amplitude-modulated, phase-shifted version of the oscillated arrival process.

FOURIER ANALYSIS OF FREQUENCY DOMAIN DISCRETE-EVENT SIMULATION EXPERIMENTS

INTRODUCTION

1.1 Motivation and Outline

A frequency domain experiment (FDE) is a discrete-event simulation experiment in which selected system parameters are oscillated sinusoidally to induce oscillations in one or more system response statistics of interest. The FDE output is a response sequence of observations corresponding to a system response statistic of interest, e.g., the waiting time in a system queue. The response sequence is used to estimate the response power spectral density (psd). Spectral analysis of the estimated response psd is then used to determine the sensitivity of the system response to variations in the selected system parameters. FDEs are based on the assumption that if a particular system response statistic is sensitive to a system parameter, then sinusoidal variation of that system parameter at a fixed frequency will induce similar sinusoidal variations in the response statistic, at the same frequency.

FDE's were introduced to discrete event simulation in 1981 by Schruben, et al. [40]. The objective was to perform input parameter sensitivity analysis for factor screening in complex discrete-event simulations. Since then, significant work has been done to develop FDE techniques. This work includes criteria for oscillation amplitude selection [11], driving frequency selection [10]; methods for generating and analyzing the response sequence [43], methods of flattening the noise spectrum [1] and methods for using the global simulation clock time as the FDE oscillation

index [26]. In addition, significant work has been done to extend the applicability of FDEs to regression analysis [37], [44], to simulation optimization [27], [39] and gradient estimation [12].

Although FDE techniques have been used successfully for some applications, two fundamental theoretical and data analysis FDE problems remain. As noted by Sargent et al. [38], these remaining problems "are the major road-blocks in the widespread use of the intuitively simple, yet powerful simulation tool." These two problems are highlighted in sections 1.1.1 and 1.1.2 respectively. Both problems are addressed in this dissertation.

1.1.1 FDE Indexing

To perform the FDE spectral analysis correctly, it is necessary to select a suitable *oscillation index* with respect to which all oscillations can be referenced. Until recently, the proper choice of the oscillation index has been an open problem in the FDE literature—the so-called "FDE indexing problem", see Sargent [38] and Buss [1]. The FDE indexing problem, presented in detail in **chapter 2**, can be briefly described as follows. In a FDE, all frequencies are measured relative to a common independent variable—an oscillation parameter.[1] The usual advice in [1], [10], [11], [27], [36]-[43] is to choose a discrete oscillation parameter, e.g., the job number

[1]In this dissertation, the common, independent variable with respect to which all FDE oscillations are referenced is called the oscillation parameter. The difference between the oscillation *parameter* and oscillation *index* is discussed in chapter 2.

in a single-server queue. For simple systems, choosing some discrete oscillation parameter yields satisfactory results. In more complex systems, however, e.g., an open network of queues, jobs do not necessarily leave the system in the order of arrival. Therefore, choosing a discrete oscillation parameter may require reordering of the response sequence before analysis. Other more complex systems, e.g., a closed network of queues, can be envisioned for which a natural discrete oscillation parameter may not exist. To solve this FDE indexing problem, what is required is an oscillation parameter that results in a sampled data sequence that is amenable to (discrete) Fourier analysis and can be generalized beyond a particular application.

Since the notion of time underlies all discrete-event simulations, an oscillation parameter based on the global simulation clock time is a natural choice for the FDE oscillation parameter. In **chapter 3** a correct way to use the global simulation clock time as the FDE oscillation parameter is developed. The development of this method is based on the established theory of counting processes. This new method, called the FDE Histogram method, can be used for FDE data analysis when the selected system response statistic is a rate. Most commonly used system response statistics are not rates, however, but are instead statistics like the expected wait in the system, the expected number in the system or the system utilization. In these cases, the *extended* FDE Histogram method, based on the FDE Histogram method and developed in **chapter 4**, should be used. The two methods are implemented and tested on several systems in chapters 3 and 4, respectively. The results show

that the FDE Histogram method and the extended FDE Histogram method are, in fact, correct solutions of the FDE indexing problem.

1.1.2 FDE System Model Assumption

To justify the use of Fourier methods, a basic FDE model assumption is that if a particular system response statistic is sensitive to a system parameter, then sinusoidal variation of that system parameter at a fixed frequency will induce similar sinusoidal variations in the response statistic, at the *same* frequency. This FDE system model assumption has been verified numerically by several FDE practitioners and queuing theory researchers interested in the analysis of queues with time-varying input processes. Unfortunately, the "verification" has been based, almost entirely, on empirical, application-specific experimental simulation studies. There is a serious lack of theoretical support for the FDE model assumption. The next part of this dissertation provides some of that theoretical support. That is, in **chapter 5** of this dissertation a simulation-clock based solution to the indexing problem is used to analyze the frequency response of a $M/M/1$ queuing system. An equation is derived which accurately characterizes the extent to which the departure process from a $M/M/1$ queuing system can be modeled as an amplitude-modulated, phase-shifted version of the oscillated arrival rate. In **chapter 6**, the FDE frequency response analysis is shown to be true for several networks of such queues.

CHAPTER II

TRADITIONAL FREQUENCY DOMAIN EXPERIMENTS

2.1 Traditional FDE Methods—Overview

Frequency domain simulation experiments were introduced by Schruben et al. [40] in 1981. A FDE is a discrete event simulation experiment in which selected system parameters are oscillated sinusoidally to induce oscillations in one or more system response statistics of interest. System parameters of interest $x_1, x_2, \ldots$ (e.g., the service rates for selected servers in a network of queues) are varied as

$$x_j(t) = x_j(0) + \alpha_j \sin(2\pi\omega_j t) \qquad j = 1, 2, \ldots \tag{2.1}$$

where $x_j(0)$ is the nominal value of the j^{th} system parameter, α_j is its amplitude of oscillation and ω_j is its frequency of oscillation, expressed in cycles per unit t. All frequencies $\omega_1, \omega_2, \ldots$ are measured relative to t, referred to in the traditional FDE literature [1], [10]–[12], [27]–[36], [40]–[44] as the oscillation index.[1] Although the notation suggests that t is time, as illustrated in examples 2.1, 2.2, 2.3 and 2.4 (to follow), in the traditional FDE literature a *discrete* simulation parameter is used as the oscillation index.

[1]The word "traditional" is used with some reservation. FDEs are a fairly recent development and thus are not in the same category chronologically as some more traditional simulation issues like event-list management and random variate generation.

The FDE output is a response sequence of observations $y_1, y_2, \ldots, y_n$ corresponding to a system response statistic of interest, e.g., the waiting time in the system. The response sequence is used to estimate the response power spectral density (psd) $\hat{y}(\omega)$. Spectral analysis of the (estimated) response psd is used to determine the sensitivity of the system response to variations in $x_1, x_2, \ldots$. Ideally, the response psd will have statistically significant "spikes" at $\omega_1, \omega_2, \ldots$ and be essentially zero at all other frequencies.

Schruben et al. [40] used the Blackman-Tukey approach to estimate the response psd as[2]

$$\hat{y}(\omega) = \sum_{k=-m}^{m} \lambda_k c_k \cos(2\pi\omega k) \qquad 0.0 < \omega < 0.5. \tag{2.2}$$

In equation 2.2 the auto-covariance is

$$c_k = \begin{cases} (1/n)\sum_{i=1}^{n-k}(y_i - \bar{y})(y_{i+k} - \bar{y}) & k = 0,\ 1,\ 2, \ldots,\ m \\ c_{-k} & k < 0 \end{cases} \tag{2.3}$$

where $\bar{y}$ denotes the average of $y_1, y_2, \ldots, y_n$ and m denotes the support of the window sequence λ_k. Different windows have been proposed including the Boxcar window, the Hanning window, the Hamming window, the Cosine Bell window and the Barlett window [8]. As in [40], the Cosine Bell window sequence

$$\lambda_k = \frac{1}{2}(1 + \cos(\pi k/m)), \qquad |k| \leq m \tag{2.4}$$

[2]If λ_k is an even sequence, i.e., $\lambda_k = \lambda_{-k}$ (this is the usual case), then equation 2.2 can be written as $\hat{y}(\omega) = \lambda_0 c_0 + 2\sum_{k=1}^{m} \lambda_k c_k \cos(2\pi\omega k)$.

is used with the Blackman-Tukey psd estimator in this dissertation. The (estimated) response psd is computed at a discrete set of frequencies in the range $0.0 < \omega < 0.5$. Frequencies outside this range are redundant because $\hat{y}(\omega)$ is real-valued, even and periodic with a (fundamental) period of 1.0.

The notion of "estimation error", or noise, in the psd estimate is important. The estimator defined by equation 2.2 is not exact in the sense that the estimated psd differs from the true (unknown) psd; the associated estimation error has both a deterministic and a stochastic component. The deterministic component, also known as estimation bias, is caused by the use of a finite-support window. A finite-support window is necessary because the amount of available data (n in equation 2.3) is finite. The stochastic component of the estimation error is due to the fact that the psd is estimated from a single (finite) realization of a stochastic process. Since the psd of a stochastic process is defined as an average of psd estimates over an ensemble of realizations, any estimate based on a single realization will differ (in a random manner) from the true psd.

The estimation error, or noise, in the psd estimate can be reduced by increasing the amount of data or averaging the individual psd estimates over additional replications. Equivalently, a single realization with a large amount of data can be sub-divided into blocks, with each block representing a different realization of the stochastic process and block averaging can be used to reduce noise [8].

The choice of m in equation 2.2 involves an empirical compromise between noise

and the magnitude of the spikes in $\hat{y}(\omega)$ at the frequencies $\omega_1, \omega_2, \ldots$ Increasing the value of m increases the total energy in the psd because of an increase in the number of points used to compute the psd estimate. This increased energy results in a proportional increase in the magnitude of the spikes; however, the magnitude of the noise may increase as well. While there is no "right" choice for m and n, the general philosophy is to collect as much data as possible—make n large. Then m can be made correspondingly large to increase the magnitude of the spikes. An empirical rule of thumb suggested in [26] is $m = 2\sqrt{(0.9)n}$.

The most common type of traditional FDE requires two simulation "runs", a *control run* and a *signal run.* The control run is a conventional discrete-event simulation run in which all system parameters are held constant at their nominal values and the *control spectrum* $\hat{y}_c(\omega)$ of the response sequence is calculated using equation 2.2. In the signal run, system parameters are oscillated sinusoidally during the run and the *signal spectrum* $\hat{y}_s(\omega)$ of the corresponding response sequence is calculated using equation 2.2. The spectral ratio

$$R(\omega) = \hat{y}_s(\omega)/\hat{y}_c(\omega) \qquad 0.0 < \omega < 0.5 \tag{2.5}$$

is examined for the presence of spikes at $\omega_1, \omega_2, \ldots$ A large spike in the spectral ratio at frequency ω_j indicates that the selected response statistic is sensitive to variations in x_j. If a statistically significant spike is not present at ω_j, it is concluded that the response statistic is not sensitive to variations in x_j. Schruben et al. [40] introduced the use of the spectral ratio, instead of the individual control and signal spectra, to

perform spectral analysis, claiming that the spectral ratio suppresses the noise in the individual spectra.

Som et al. [43] introduced a method of generating the response sequence by replication, instead of using the two-run process. The i^{th} value of the response sequence is the ensemble average of the i^{th} value of the response statistic for multiple independently-seeded replications. A discrete Fourier transform (DFT) of this ensemble averaged response sequence is computed; the square of the magnitude of the (complex-valued) result is the required psd estimate. No control run is required. Consistent with theory, Som et al. demonstrated that the estimated spectrum becomes less noisy as the number of replications is increased.

Several algorithms for computing the DFT exist in the literature. The Fast Fourier Transform (FFT), a well-known computationally efficient algorithm for computing the DFT, was introduced in the late 1960's [5]. With the use of the FFT, the DFT approach to estimating the psd became so computationally efficient that, in general, it replaced the Blackman-Tukey approach for most practical applications [8]. For the examples in section 2.4 and for the FDE histogram methods developed in chapters 3 and 4, the DFT approach is used for estimating the psd.

2.2 Traditional FDE Methods—Examples

Several examples illustrating typical (simple) applications of traditional FDE methods are presented in this section. In chapter 3, the same examples will be used to

highlight the FDE indexing problem (discussed in section 2.4) and motivate the development of the FDE histogram methods in chapters 3 and 4.

Unless otherwise stated, all traditional FDEs presented in this section and in section 3.1 generate 50000 response observations. The first 5000 observations are (arbitrarily) ignored to reduce the effect of initial transients. The spectral ratio of the resulting response sequence is calculated using equation 2.2 with $n = 45000$ and $m = 424$ (values of n and m originally proposed in [26]).

Example 2.1 As in [26], a FDE for a $M/M/1$ queue is performed. Customers arrive according to a Poisson process with arrival rate $\lambda = 0.5$ and join a FIFO queue before a single server. The service time of the t^{th} customer is sampled from an exponential distribution with service rate

$$\mu(t) = 1.0 + 0.4\sin(2\pi\omega_1 t) \qquad t = 1, 2, 3, \ldots \tag{2.6}$$

As indicated, the customer index is selected as the oscillation index. The frequency of oscillation is fixed at $\omega_1 = 0.03$ cycles per customer. The waiting time of customers in the system is the response statistic of interest.

The spectral ratio (figure 2.1a) exhibits a distinct spike at ω_1 suggesting that (as expected) the waiting time in a $M/M/1$ queue is sensitive to variations of the service rate. As explained in section 2.1, and as demonstrated by using the same $n = 45000$ but a larger $m = 1000$ (figure 2.1b), the magnitude of the spike can be increased by increasing the value of m. Increasing the value of m does, however, increase the required computational effort, often with only a marginal improvement

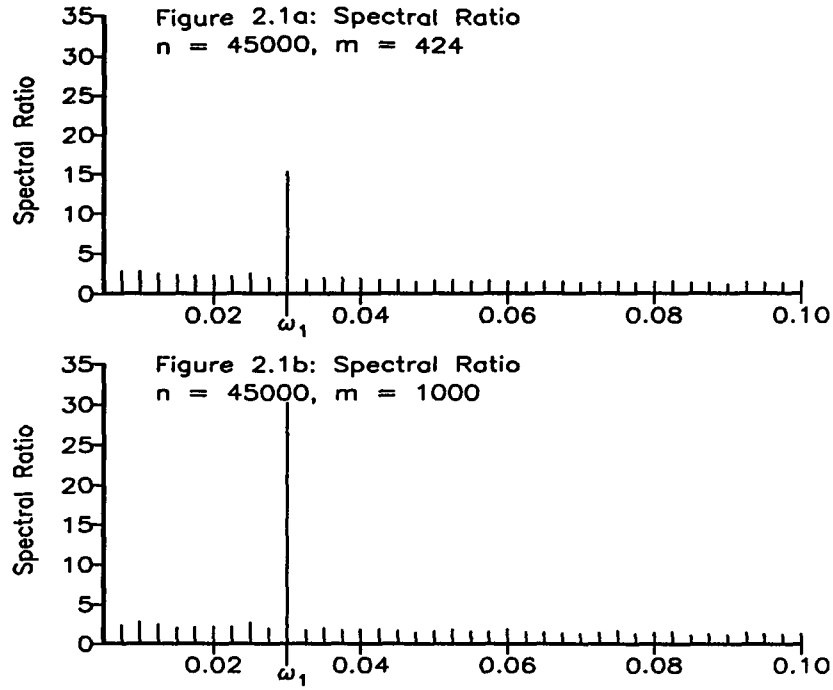


Figure 2.1: Spectral ratio for a M/M/1 queue using traditional FDE methods, $\omega_1 = 0.03$.

in one's ability to distinguish the spike from noise. A visual inspection of figure 2.1 indicates that in this case, $n = 45000$ and $m = 424$ is sufficient to produce a distinct spike at ω_1, as desired.

Example 2.2 As an extension of example 2.1 and as in [44], a traditional FDE for a feedback $M/M/1$ queuing system (figure 2.2) is performed. In this case, after receiving service a customer rejoins the end of the queue with probability $p = 0.25$ or leaves the system with probability $1 - p = 0.75$. As in example 2.1, the service time of the t^{th} customer *entering* the system is sampled from an exponential distribution with service rate $\mu(t) = 1.0 + 0.4\sin(2\pi\omega_1 t)$. Note that, if fed back, the t^{th} customer has the same *expected* service time but a different *actual* service time each time it

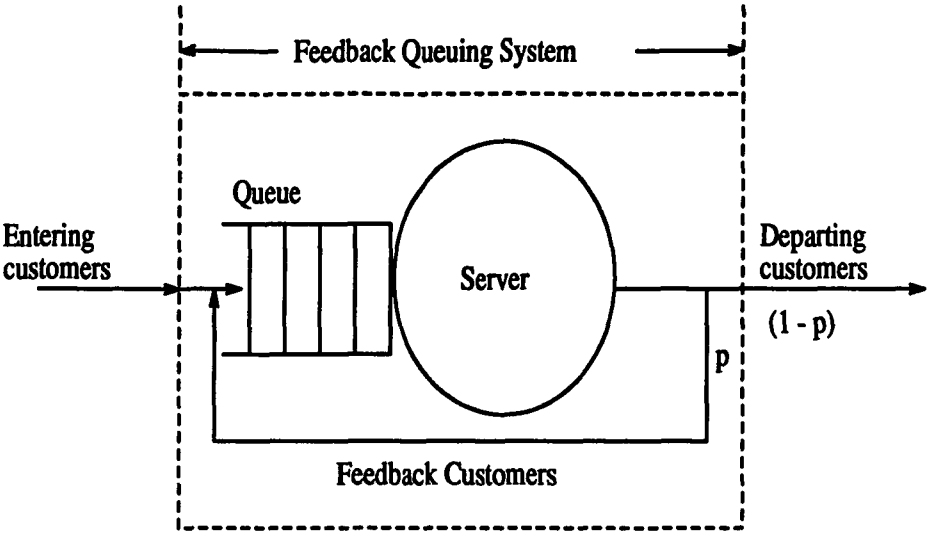


Figure 2.2: Feedback queuing system.

receives service. The total waiting time of the customers (including feedback) is the response statistic of interest.

A FDE complication with respect to example 2.1 is that the response sequence needs to be ordered by the index of arriving customers. That is, since customers do not necessarily leave the system in the order in which they arrive, either the response sequence has to be sorted before performing the spectral analysis or data gathering during the simulation has to be structured to ensure that the t^{th} value in the response sequence represents the total waiting time of the t^{th} customer entering the system. To illustrate, the spectral ratio in figure 2.3a indicates that *without sorting* the spike at ω_1 is "smeared"; however, the spectral ratio of the *sorted* response sequence in figure 2.3b has a much more distinct spike at $\omega = \omega_1$. In either case, however, the spike at ω_1 suggests that (as expected) the total waiting time in a feedback queue

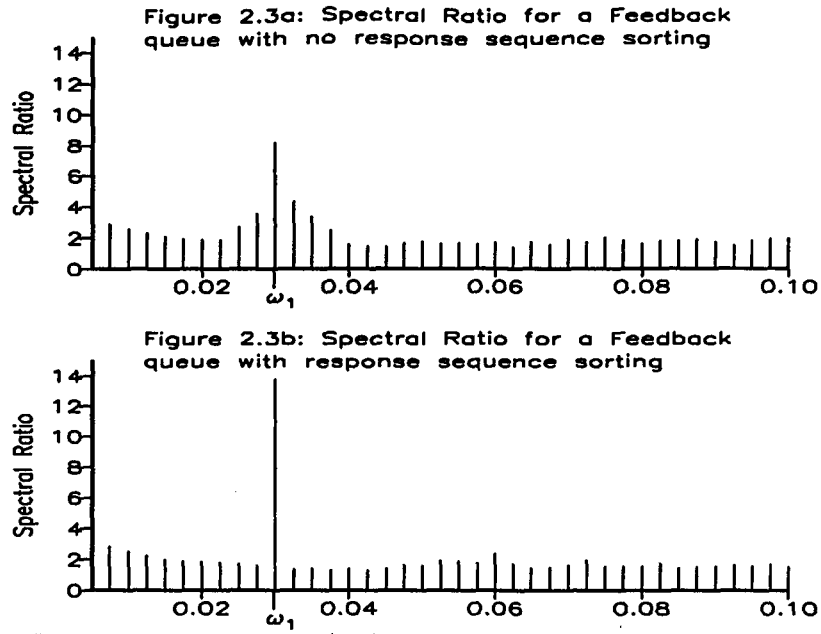


Figure 2.3: Spectral ratio for a $M/M/1$ feedback queue using traditional FDE methods, $\omega_1 = 0.03$ and $p = 0.25$.

is sensitive to variations of the service rate.

Example 2.3 As an extension of examples 2.1 and 2.2, the feedback queue FDE is repeated with a sinusoidally varying service rate *and* a sinusoidally varying probability of feedback. As before, the service time of the t^{th} customer entering the system is sampled from an exponential distribution and the service rate is $\mu(t) = 1.0 + 0.4\sin(2\pi\omega_1 t)$. After the t^{th} entering customer receives service (for the first time or after feedback) its probability of feedback is given by

$$p(t) = 0.25 + 0.15\sin(2\pi\omega_2 t) \qquad t = 1, 2, 3, \ldots \tag{2.7}$$

That is, the t^{th} customer has the same probability of feedback each time it leaves the server (if more than once). Both the service rate and feedback probability have

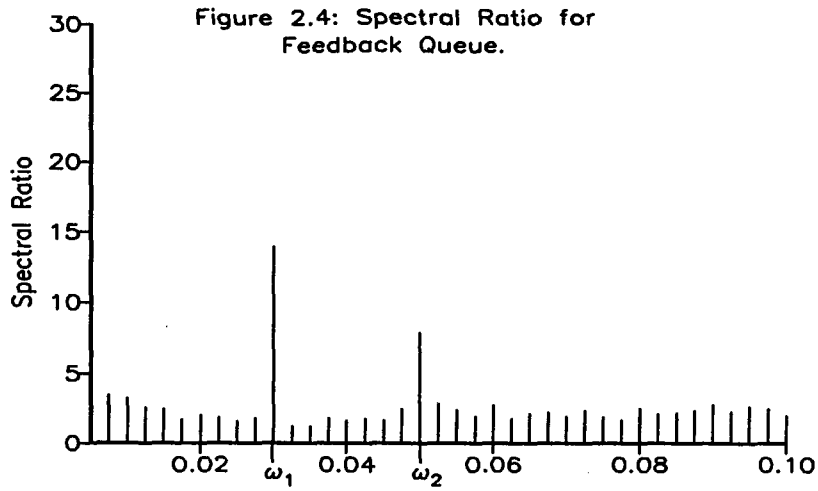


Figure 2.4: Spectral ratio for a $M/M/1$ feedback queue using traditional FDE methods, $\omega_1 = 0.03$, $\omega_2 = 0.05$.

the same index of oscillation. The total waiting time of the customers (including feedback) is the response statistic of interest.

A traditional FDE is performed with $\omega_1 = 0.03$ and $\omega_2 = 0.05$. Figure 2.4 represents the spectral ratio of the (sorted) response sequence. The spectral ratio has distinct spikes at both ω_1 and ω_2 indicating that (as expected) the total waiting time is sensitive to variations of both the service rate and the feedback probability. There is, however, a significant amount of noise and the possibility of a false spike at very low frequency.

Example 2.4 As in [26], a FDE for a simple manufacturing assembly station (figure 2.5) is performed. In the assembly, one job of type 1 is combined with two jobs of type 2, to form a type 3 job. For each type 3 job, the difference between its

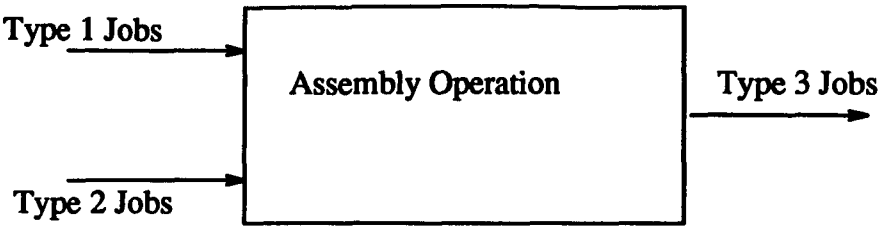


Figure 2.5: Simple manufacturing assembly station.

departure time and the arrival time of the latest component job (either the type 1 job or the second of the two type 2 jobs) is selected as the response. The arrival rate of the type 1 jobs is varied sinusoidally between 0.93 and 0.73 with frequency $\omega_1 = 0.01$, the arrival rate of the type 2 jobs is varied sinusoidally between 1.1 and 0.9 with frequency $\omega_2 = 0.03$ and the service rate of the type 3 jobs is varied sinusoidally between 1.0 and 0.8 with frequency $\omega_3 = 0.04$.

A naive approach to choosing the oscillation index is to choose an individual discrete oscillation index $t = 1, 2, 3, \ldots$ for *each* job type. If this is done, then instead of a spike at $\omega = \omega_2$ (as desired), the spectral ratio exhibits a spike at $2\omega_2$ (see figure 2.6a). The reason the spike appears at $2\omega_2$ is because, on average, the oscillation index for a type 2 job increases twice as fast as the oscillation index for a type 1 or a type 3 job. This frequency doubling is a potential source of confusion caused by the fact that all oscillation frequencies are not based on a *common* oscillation index.

As in [26], to avoid the $2\omega_2$ confusion, the invoice number $t = 1, 2, 3, \ldots$ for type 3 jobs is used as the common oscillation index. The service time of the t^{th} type 3

job is sampled from an exponential distribution with service rate

$$\mu(t) = 0.9 + 0.1\sin(2\pi\omega_3 t). \tag{2.8}$$

Since there is one assembly and one type 1 job associated with each type 3 job (a one-to-one correspondence), the t^{th} interarrival time of type 1 jobs is drawn from an exponential distribution with rate

$$\lambda_1(t) = 0.83 + 0.1\sin(2\pi\omega_1 t). \tag{2.9}$$

and the oscillation index for the type 1 jobs is also $t = 1, 2, 3, \ldots$ Since there are two jobs of type 2 for each type 3 job, the oscillation index for the type 2 jobs is $t = 1, 1, 2, 2, 3, 3, \ldots$. That is, the $(2t-1)^{\text{th}}$ and the $(2t)^{\text{th}}$ interarrival times of type 2 jobs is drawn from an exponential distribution with rate

$$\lambda_2(t) = 1.0 + 0.1\sin(2\pi\omega_2 t). \tag{2.10}$$

This indexing scheme avoids the $2\omega_2$ confusion. Spikes in the spectral ratio at ω_1, ω_2 and ω_3 are expected.

Although the spectral ratio of the response sequence (figure 2.6b) is noisy and has some false spikes, smeared spikes are evident at ω_2 and ω_3 indicating that λ_2 and μ have a significant effect on the response but λ_1 does not. The FDE results suggest that the type 2 jobs are the bottleneck and, indeed, that is the case. That is, the (nominal) arrival rate (0.5) of the latest type 2 component job for each type 3 job is smaller than the (nominal) arrival rate (0.83) for the type 1 job. Thus, a type 2 job is usually the latest arriving component part and so type 2 jobs have

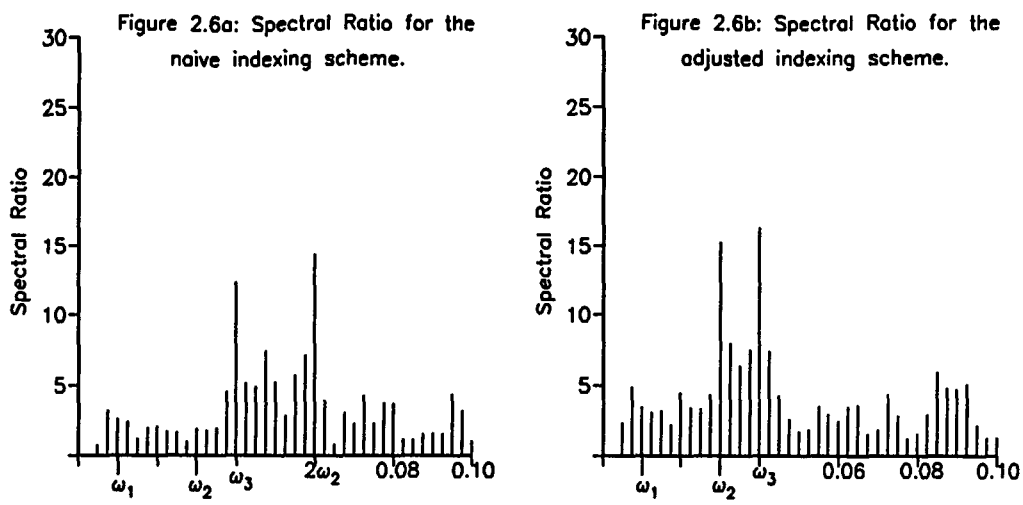


Figure 2.6: Spectral Ratio for an assembly operation using traditional FDE methods, $\omega_1 = 0.01, \omega_2 = 0.03$ and $\omega_3 = 0.04$.

more effect than type 1 jobs on the response. The service rate also has a significant effect on the response because a significant portion of each response observation is the time in service.

2.3 Traditional FDE Methods—Problems

The examples in section 2.2 demonstrate that traditional FDEs can be used with some success to perform parameter sensitivity analysis for simple systems. As pointed out by Buss [1] and Sargent et al. [38] however and as indicated by examples 2.1–2.4, there are several problems associated with FDE methods, two of which are addressed in this dissertation.

1. FDE Indexing.

   As illustrated by examples 2.1–2.4, traditional FDEs use a discrete oscillation index. There are no unambiguous guidelines about how to choose this index, however. Example 2.4 demonstrates that naively choosing an individual discrete oscillation index for each job type can result in a spike in the spectral ratio at the wrong frequency. Examples 2.2 and 2.3 demonstrate that even if an "obvious" discrete oscillation index is used, sorting of the response sequence may be required, thereby complicating the FDE analysis. The absence of unambiguous guidelines about the choice of an appropriate FDE oscillation index is called the FDE *indexing problem.* This problem is discussed in more detail in section 2.4 and a solution is presented in chapters 3 and 4. Related problems are noise, the smearing of spikes and false spikes in the response spectrum, as is evident in figures 2.3a and 2.6.

2. FDE Model Assumption.

   In each of examples 2.1–2.4, it is assumed that if the oscillation index is properly chosen then sinusoidal variations of selected system parameters at fixed frequencies will induce sinusoidal variations in the response statistic *at the same frequencies.* This assumption for $M/M/1$ queues has been verified numerically by several queuing theory researchers [3], [16], [17], [20], [21], [30], [32], [33], [35], [45], [48]. No theoretical support exists for this assumption however.

This dissertation addresses both these problems. A solution to the FDE indexing problem is provided in chapters 3 and 4. A partial solution to the FDE model assumption problem is presented in chapters 5 and 6.

2.4 FDE Indexing Problem

Morrice et al. [26] recognized the FDE indexing problem and suggested that a solution should be based on measuring frequencies with respect to the simulation clock time. Morrice et al. were right in their suggestion—unfortunately their implementation of an indexing scheme based on the simulation clock time was incorrect. Consequently they concluded, incorrectly, that the simulation clock time is not a good basis for a solution to the FDE indexing problem. This dissertation uses Morrice et al.'s original suggestion, but develops a correct way to use the simulation clock time to solve the FDE indexing problem.

As background for the solution to the FDE indexing problem developed in chapters 3 and 4, it is necessary to distinguish three important entities—the oscillation parameter, the oscillation index and the sampling index. The continuous, simulation clock time, t, which serves as the common reference for all oscillations, is the *oscillation parameter.* The discrete values of the oscillation parameter $t_1, t_2, t_3, \ldots$ that are associated with the (sampled) response sequence $y_i = y(t_i)$ for $i = 1, 2, 3, \ldots$, is the *oscillation index* and $i = 1, 2, 3, \ldots$ is the *sampling index.* In the traditional FDE literature, the term "oscillation index" is casually used to denote the (discrete)

oscillation parameter, the oscillation index and the sampling index. If a continuous oscillation parameter like the simulation clock time is used, however, then the response statistic has to be *sampled* at discrete times. In this case the oscillation parameter, the oscillation index and the sampling index are no longer the same.

To solve the FDE indexing problem, what is required is an oscillation index that results in a (sampled) response sequence that is amenable to (discrete) Fourier analysis. In particular, the FDE oscillation index should be selected so that the response psd has distinct spikes at the frequencies of input oscillations with minimal noise at all other frequencies. Two oscillation index requirements are evident.

1. The FDE oscillation index should be monotonically increasing; i.e., $t_i < t_{i+1}$ for $i = 1, 2, 3, \ldots$ If this condition is violated, the response sequence needs to be sorted, (as in example 2.2 and 2.3), prior to computing its psd.

2. The FDE oscillation index should be equally spaced; i.e., $t_{i+1} - t_i = \delta$ for $i = 1, 2, 3, \ldots$ and some constant $\delta > 0$. If this requirement is satisfied, then requirement 1 is automatically satisfied.

The following examples are used to illustrate why requirements 1 and 2 are critical. Example 2.5, which is analogous to examples 2.2 and 2.3, is used to demonstrate the need to sort the response sequence. Example 2.6 provides a graphic illustration of the smearing caused by the stochastic sampling of a deterministic function and example 2.7 demonstrates the uniform sampling of a stochastic function. These three examples are used to motivate the solution to the FDE indexing problem

presented in chapters 3 and 4.

Example 2.5 A deterministic function $x(t) = \sin(2\pi\omega_1 t)$ is sampled at times $t_i = i$ for $i = 1, 2, 3, 4, \ldots, n$ with $n = 1024$ and $\omega_1 = 0.0625$ cycles per unit time. For this example, $0 < t < \infty$ is the oscillation parameter. The oscillation index and the sampling index are the same because $t_i = i$. The uniformly sampled response sequence defined by $y_i = x(t_i)$ for $i = 1, 2, \ldots, n$ is shown in figure 2.7a.[3] The DFT of the response sequence is calculated and as in [43] the magnitude of the (complex) DFT is used to perform the spectral analysis.[4] Figure 2.7b indicates that (as expected) the magnitude spectrum has a spike at the frequency of oscillation. Even for small n ($n = 1024$ is small compared to the 45000 samples required for the Blackman-Tukey approach), the magnitude spectrum is virtually noise free and the spike at ω_1 is clearly discernible with no smearing.

To simulate shuffling similar to that which occurs in a feedback queue, the response sequence is shuffled based on a shuffling probability p. That is, for each index $i = 1, 2, \ldots, n$, the number of places M_i that the i^{th} data point is moved, is drawn from a Geometric distribution with parameter p. If $M_i = 0$, the i^{th} data point is not moved. If $M_i > 0$, then $M_i - 1$ data points to the right of i are each moved one place to the left and the i^{th} data point is moved to the $(i + M_i)^{\text{th}}$ position.

[3] Although the response sequence is discrete, for the purpose of generating visually meaningful plots, the discrete response sequence for this example and examples 2.6, 2.7 and 3.3 have been interpolated to a smooth curve. Also, only a portion of the response sequence is shown.

[4] The square of the magnitude of the complex DFT is a psd estimate.

The shuffling algorithm is presented, in pseudocode, at the end of this chapter.

Figures 2.7c and 2.7d represent the shuffled response sequence and the magnitude of the corresponding (complex) DFT for $p = 0.3$, respectively; figures 2.7e and 2.7f represent the shuffled response sequence and corresponding magnitude for $p = 0.6$ while figures 2.7g and 2.7h represent the shuffled response sequence and corresponding magnitude for $p = 0.9$. As p is increased the response sequence becomes more shuffled, increasing the noise and reducing the magnitude of the spike in the corresponding spectrum. The shuffling of the response sequence is similar to that produced by the feedback in examples 2.2 and 2.3). In both of these examples the response sequence had to be sorted before the psd could be computed to avoid smearing of the response spectrum. Example 2.5 emphasizes the need to sort the response sequence when performing a traditional FDE for the feedback queue, thereby increasing the computational complexity of the FDE data analysis.

The need for uniform sampling is demonstrated by example 2.6, which was presented in [24]. The example is based upon sampling a deterministic function at equal, slightly random and random time increments.

Example 2.6 The deterministic function $x(t) = \sin(2\pi\omega_1 t)$ is sampled at times defined by $t_i = t_{i-1} + \epsilon_i$, for $i = 1, 2, \ldots, n$ with $n = 1024$, $t_0 = 0$ and $\omega_1 = 0.0625$. As in example 2.5, $0 < t < \infty$ is the oscillation parameter; t_i is the oscillation index and i is the sampling index. Three cases are considered:

1. $x(t)$ is sampled at *equal* time increments, i.e., $\epsilon_i = 1$ for all i;

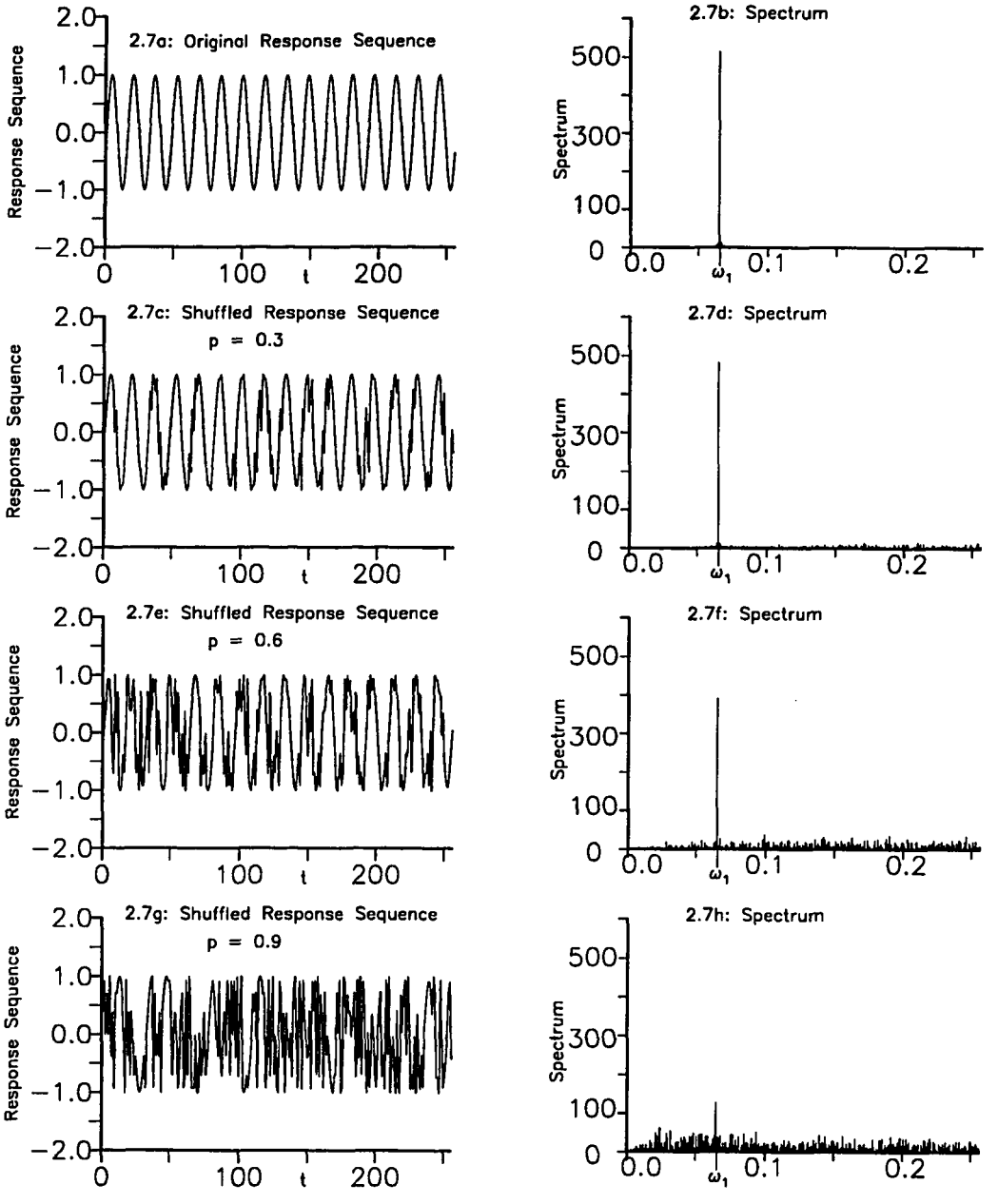


Figure 2.7: Response sequence and spectrum obtained by uniformly sampling a sine wave and then shuffling the resulting sequence with probability p, $\omega_1 = 0.0625$.

2. $x(t)$ is sampled at *slightly random* time increments, i.e., $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ is an independent, identically distributed (iid) sequence of random variables uniformly distributed between 0.0 and 2.0;

3. $x(t)$ is sampled at *random* time increments, i.e., $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ is an iid sequence of random variables exponentially distributed with mean 1.

In each case the DFT of the sampled data is calculated. The magnitude of the (complex-valued) result is the spectral estimate. Figures 2.8a, 2.8c and 2.8e represent the response sequence for cases 1, 2 and 3 respectively (only a portion of the response sequence is shown); figures 2.8b, 2.8d and 2.8f represent the corresponding spectrum. From these figures it is observed that the spike at ω_1 becomes more smeared as the randomness of the inter-sampling times increases.

Example 2.6 illustrates why equally-spaced samples are needed, even when the sampled function is *deterministic*. If the sampled function is *stochastic* the added randomness inherent in the samples increases the noise in the spectrum. Example 2.7 is used to demonstrate this.

Example 2.7 The stochastic function $x(t) = \sin(2\pi\omega_1 t) + F(t)$, where $F(t)$ is drawn from a Gaussian distribution with mean 0.0 and variance $\sigma^2 = 2.5$, is sampled at times $t_i = i$ for $i = 1, 2, 3, \ldots, n$ with $n = 1024$ and $\omega_1 = 0.0625$. The DFT of the sampled data is calculated and the magnitude of the (complex-valued) result is the resulting spectral estimate. Figure 2.9a indicates that, unlike the response

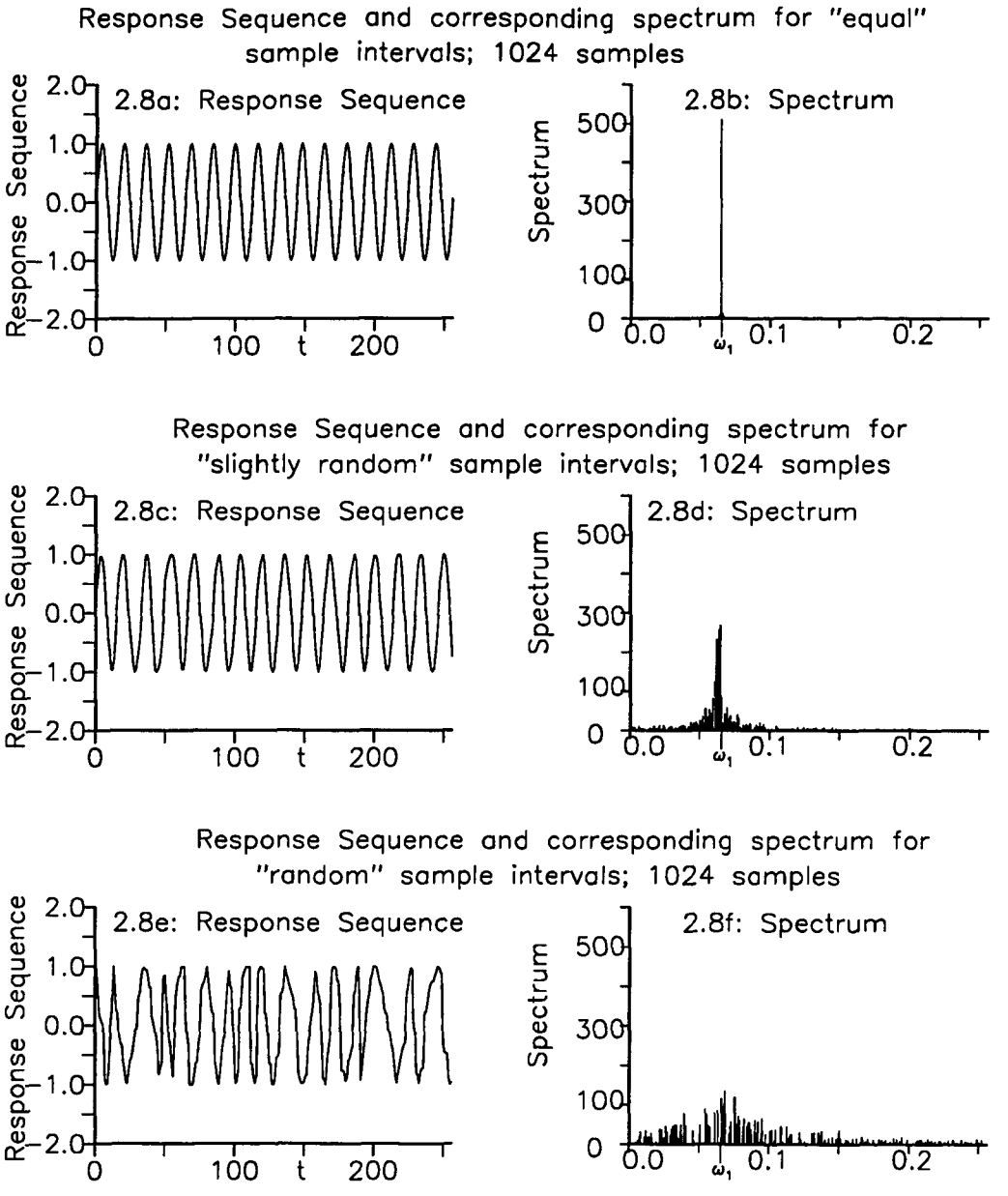


Figure 2.8: Response sequence and spectrum obtained by sampling a sine wave at equal, slightly random and random sample intervals, $\omega_1 = 0.0625$.

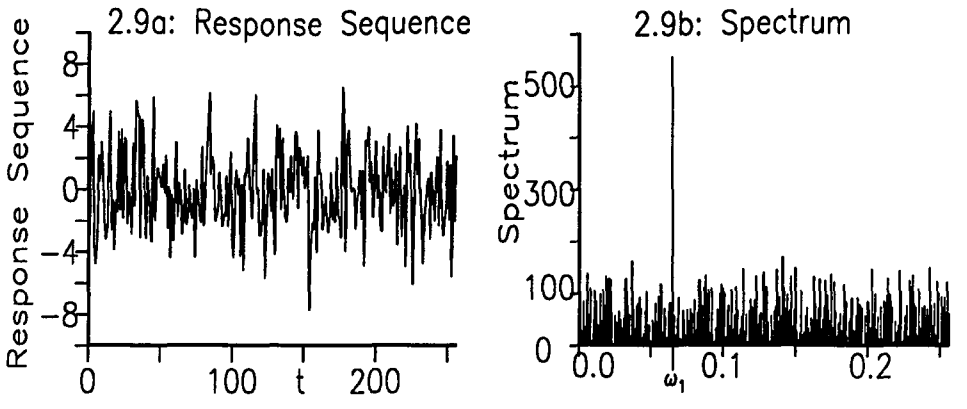


Figure 2.9: Response sequence and corresponding spectrum obtained by uniformly sampling a stochastic function at equal intervals; $\omega_1 = 0.0625$, $n = 1024$ samples.

sequence obtained when the deterministic function $\sin(2\pi\omega_1 t)$ is uniformly sampled (figure 2.8a), the response sequence of the stochastic function is noisy. The noise in the spectrum in this case is caused by the additive, uncorrelated Gaussian (white) noise.

The examples presented in this chapter show some typical FDE applications and demonstrate the traditional FDE approach. On one hand these examples provide an intuitive feel for the appeal of FDEs. But at the same time they also serve as examples of problems with traditional FDE methods. The rest of this dissertation deals with an examination of these problems and a search for their solutions.

Pseudo-code

```
type data_record  = record
                       value: double;
                       index: longint;
                    end;
var
  data : array [1..n] of data_record;
  hold : data_record;
begin
  for i := 1 to n do  { initialize response sequence }
  begin
    data[i].index := i;
    data[i].value := x(i);
  end
  i := 1;
  k := 1;
  while (i < n) do   { shuffle response sequence }
  begin
    M_i := Geometric(p); { drawn from a Geometric distribution }
    if (M_i > 0)
    begin
      hold.value := data[k].value;
      hold.index := data[k].index;
      for j:= k to (k+M_i-1) do
      begin
        data[j].value:= data[j+1].value;
        data[j].index:= data[j+1].index;
      end
      data[k+M_i].value := hold.value;
      data[k+M_i].index := hold.index;
    end
    i := i+1;
    while (data[k].index <> i)   { determine position of next data point }
      k := k+1;
  end
end
```

CHAPTER III

SIMULATION CLOCK TIME AS THE FDE OSCILLATION PARAMETER

3.1 Simulation Clock Time

As discussed in chapter 2, in a discrete-event FDE the (global) simulation clock time is a natural *continuous* variable with respect to which all dynamic variables are referenced and, for that reason, is an obvious choice for the FDE oscillation parameter. Morrice et al. [26] were the first to suggest that a solution to the FDE indexing problem should be based on measuring frequencies with respect to the simulation clock time. To investigate their suggestion, they performed two FDEs with the simulation clock time as the oscillation *parameter*. As demonstrated in this chapter their choice of the oscillation *index* was ambiguous, however, and they concluded incorrectly that the simulation clock time is not a good basis for a solution to the FDE indexing problem.

For reference, as in [26], a FDE for a $M/M/1$ queue and a simple manufacturing operation is performed using Morrice et al.'s approach. Example 3.3 is then used to demonstrate some reasons for the unsatisfactory results in examples 3.1, 3.2 and in [26].

Example 3.1 The FDE for a $M/M/1$ queue (example 2.1) is repeated with the simulation clock time used as the oscillation parameter. That is, the service time of a customer *entering service* at time t is sampled from an exponential distribution with service rate

$$\mu(t) = 1.0 + 0.4\sin(2\pi\omega_1 t) \tag{3.1}$$

and $\omega_1 = 0.03$. The waiting time in the system of the i^{th} customer, arriving at time t_i, is selected as the response statistic of interest, y_i. There is ambiguity in the definition of the oscillation index in this example. If an oscillation index were to be identified, it would be the arrival time t_i, somewhat consistent with the oscillation index definition in chapter 2. Unfortunately, the concept of an oscillation index was absent in [26] and hence the oscillation index is not used in any meaningful way for data analysis in this example; instead, only the sampling index $i = 1, 2, \ldots$ is used.

The spectral ratio of the response sequence $y_1, y_2, \ldots, y_n$ is calculated using the Blackman-Tukey approach for psd estimation, as discussed in section 2.1. Figure 3.1 depicts the result. Unlike the results in example 2.1 (see figure 2.1), in figure 3.1 the spike at ω_1 is smeared. Although the spike at ω_1 is consistent with expectations, the smearing is undesirable. An even more undesirable feature, however, is the spurious smeared spikes at $\omega = 0.06$ and $\omega = 0.085$.

Example 3.2 The FDE for a simple manufacturing assembly station (example 2.4) is repeated using the simulation clock time t as the oscillation parameter. The

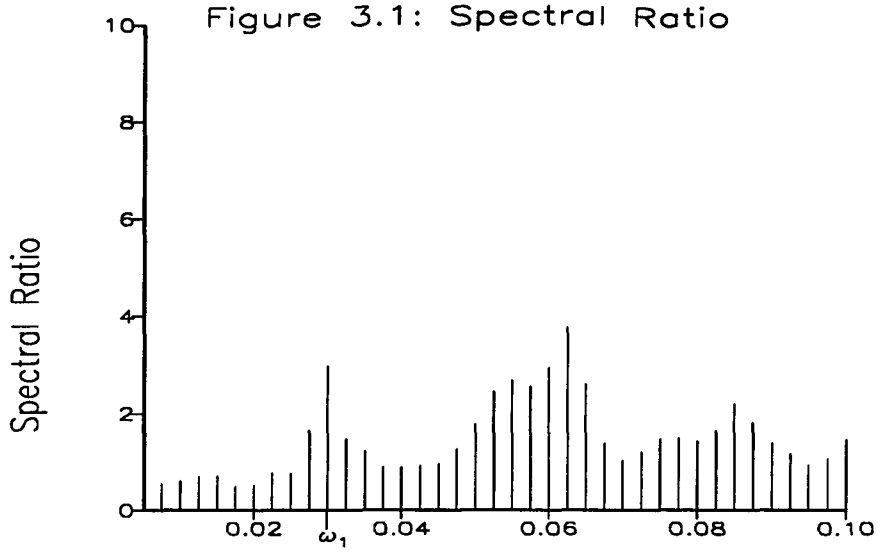


Figure 3.1: Spectral ratio for a M/M/1 queue using the simulation clock time incorrectly, $\omega_1 = 0.03$.

arrival rate of a type 1 job is

$$\lambda_1(t) = 0.83 + 0.1\sin(2\pi\omega_1 t) \tag{3.2}$$

with $\omega_1 = 0.01$. The arrival rate of a type 2 job is

$$\lambda_2(t) = 1.0 + 0.1\sin(2\pi\omega_2 t) \tag{3.3}$$

with $\omega_2 = 0.03$. The arrivals occur as a non-stationary Poisson process and the thinning method is used to simulate each arrival processes [18]. The service time of a type 3 job entering service at time t is drawn from an exponential distribution with service rate

$$\mu(t) = 0.9 + 0.1\sin(2\pi\omega_3 t) \tag{3.4}$$

and $\omega_3 = 0.04$. The difference between the departure time of the i^{th} type 3 job and the arrival time of its latest component job (either the type 1 job or the second

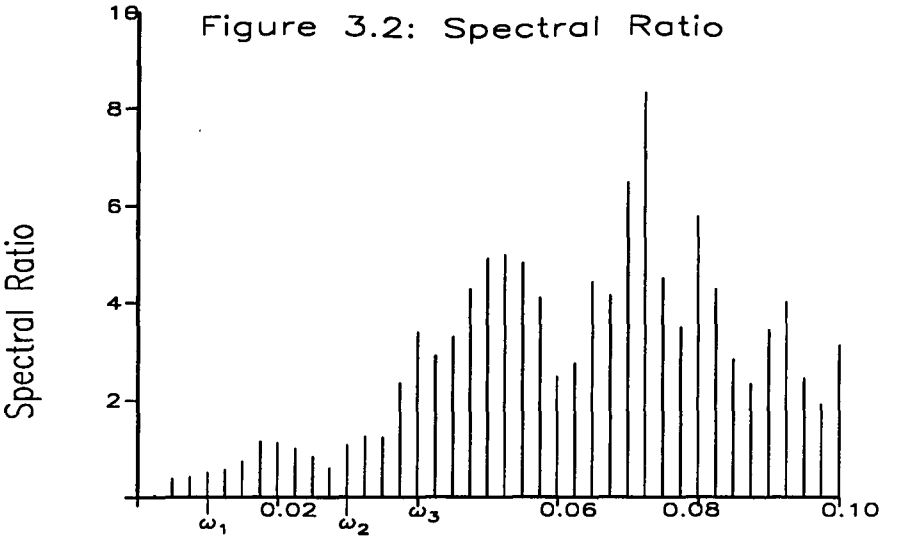


Figure 3.2: Spectral ratio for an assembly operation using the simulation clock time incorrectly, $\omega_1 = 0.01$, $\omega_2 = 0.03$ $\omega_3 = 0.04$.

of the two type 2 jobs) is selected as the i^{th} response statistic of interest. As in example 3.1 the definition of the oscillation index is also ambiguous in this example; only the sampling index is used.

Although the spectral ratio obtained in example 2.4 (see figure 2.6) was noisy, spikes were at least discernible at ω_2 and ω_3. As illustrated in figure 3.2, however, when the simulation clock time is used (incorrectly), the psd does not exhibit distinct spikes at ω_2 or ω_3. Instead the spectral ratio has spurious smeared spikes at $\omega = 0.0525$ and $\omega = 0.0725$. The FDE for the assembly operation performed by Morrice et al. [26] yielded similar unsatisfactory results.

Based on unsatisfactory results like those in figures 3.1 and 3.2, Morrice et al. [26] concluded that the simulation clock time is not a suitable oscillation parameter

for FDEs. The following example, partially demonstrates why the attempt in [26] failed and what must be done to ensure that the simulation clock can be successfully used as the oscillation parameter.

Example 3.3 Consider a system whose input is $x(t) = \sin(2\pi\omega t)$. The system output response is a sequence $y_1, y_2, \ldots, y_n$ defined by

$$y_i = \int_{t_i}^{t_i+w_i} \sin(2\pi\omega_1 t)dt \qquad i = 1, 2, \ldots, n \tag{3.5}$$

with $\omega_1 = 0.0625$ and $n = 1024$. The DFT of the response sequence is calculated and the magnitude of the (complex-valued) result is used to perform the spectral analysis. This system model is an integration process defined by the times $t_1,\ t_2, \ldots, t_n$ and integration widths $w_1,\ w_2, \ldots, w_n$. For this example, consistent with the definitions in chapter 2, t is the oscillation parameter, t_i is the oscillation index and i is the sampling index. Four cases are considered:

1. $t_i = i$ for $i = 1, 2, \ldots, n$ and $w_i = 0.5$ for $i = 1, 2, \ldots, n$.

   Figure 3.3a shows the response sequence and 3.3b the corresponding spectrum. The spectrum exhibits a distinct spike at ω_1. The magnitude of the spike is diminished, however, compared to the magnitude of the spike in example 2.5 (figure 2.7b). The decrease in the magnitude of the spectrum at ω_1 is caused by the integration operation in equation 3.5, which is equivalent to a low-pass filtering operation.

2. $t_i = i$ for $i = 1, 2, \ldots, n$ and $w_1,\ w_2, \ldots, w_n$ is an iid sequence of exponential

random variables with mean 0.5.

Figure 3.3c shows the response sequence and 3.3d the corresponding spectrum. Similar to the previous case, the resulting spectrum exhibits a distinct spike at ω_1. However the randomness in $w_1, w_2, \ldots, w_n$ creates some distortion in the response sequence and corresponding noise in the spectrum.

3. $t_i = t_{i-1} + \epsilon_i$ for $i = 1, 2, \ldots, n$ where $\epsilon_1, \ \epsilon_2, \ldots, \epsilon_n$ is an iid sequence of random variables exponentially distributed with mean 1 and $w_i = 0.5$ for $i = 1, 2, \ldots, n$.

   Figure 3.3e shows the response sequence and 3.3f the corresponding spectrum. The response sequence is badly distorted and the spectrum exhibits a badly smeared spike at ω_1. If figures 3.3d and 3.3f are compared, it is seen that although the randomness in $w_1, \ w_2, \ldots, w_n$ creates noise in the spectrum, the spectrum still has a distinct spike at ω_1; randomness in $t_1, \ t_2, \ldots, t_n$, however, creates noise in the spectrum *and* results in a badly smeared spike in the spectrum.

4. $t_i = t_{i-1} + \epsilon_i$ for $i = 1, 2, \ldots, n$ where $\epsilon_1 \ \epsilon_2, \ldots, \epsilon_n$ is an iid sequence of random variables exponentially distributed with mean 1 and $w_1, \ w_2, \ldots, w_n$ is an iid sequence of random variables exponentially distributed with mean 0.5.

   Figure 3.3g shows the response sequence and figure 3.3h is the corresponding spectrum. As in case 3, the response sequence is badly distorted and the spectrum has a badly smeared spike about ω_1. In this case, however, the

noise in the spectrum and smearing of the spike is even worse than that in case 3.

Figure 3.3 indicates that randomness in the integration start times $t_1, t_2, \ldots, t_n$ is more critical than randomness in the integration widths $x_1, x_2, \ldots, x_n$. Randomness in the integration width increases the noise in the spectrum, but the spike in the spectrum at ω_1 is still distinct. On the other hand randomness in the integration start times results in a smeared spectrum with no distinct spike at ω_1.

The response sequences in examples 3.1 and 3.2 are analogous to the sequence obtained in case 4 of example 3.3, where both t_i and w_i are stochastic. The stochastic integration start time t_i is analogous to the arrival time of the i^{th} customer for the $M/M/1$ queue or the arrival time of the latest component job of the i^{th} type 3 job for the simple manufacturing assembly operation. The stochastic integration width w_i is analogous to the total wait in the system for the i^{th} customer for the $M/M/1$ queue or the difference between the departure time of the i^{th} type 3 job and the arrival time of its latest component job in the simple manufacturing assembly operation.

Example 3.3 demonstrates that when using the simulation clock time as the oscillation parameter, naively sampling at stochastic processing times results in a smeared and noisy response spectrum.[1] Instead, when the simulation clock time

[1]Example 3.3 does not explain the spurious spectral spikes seen in examples 3.1 and 3.2. The cause of the spurious spikes is not yet known.

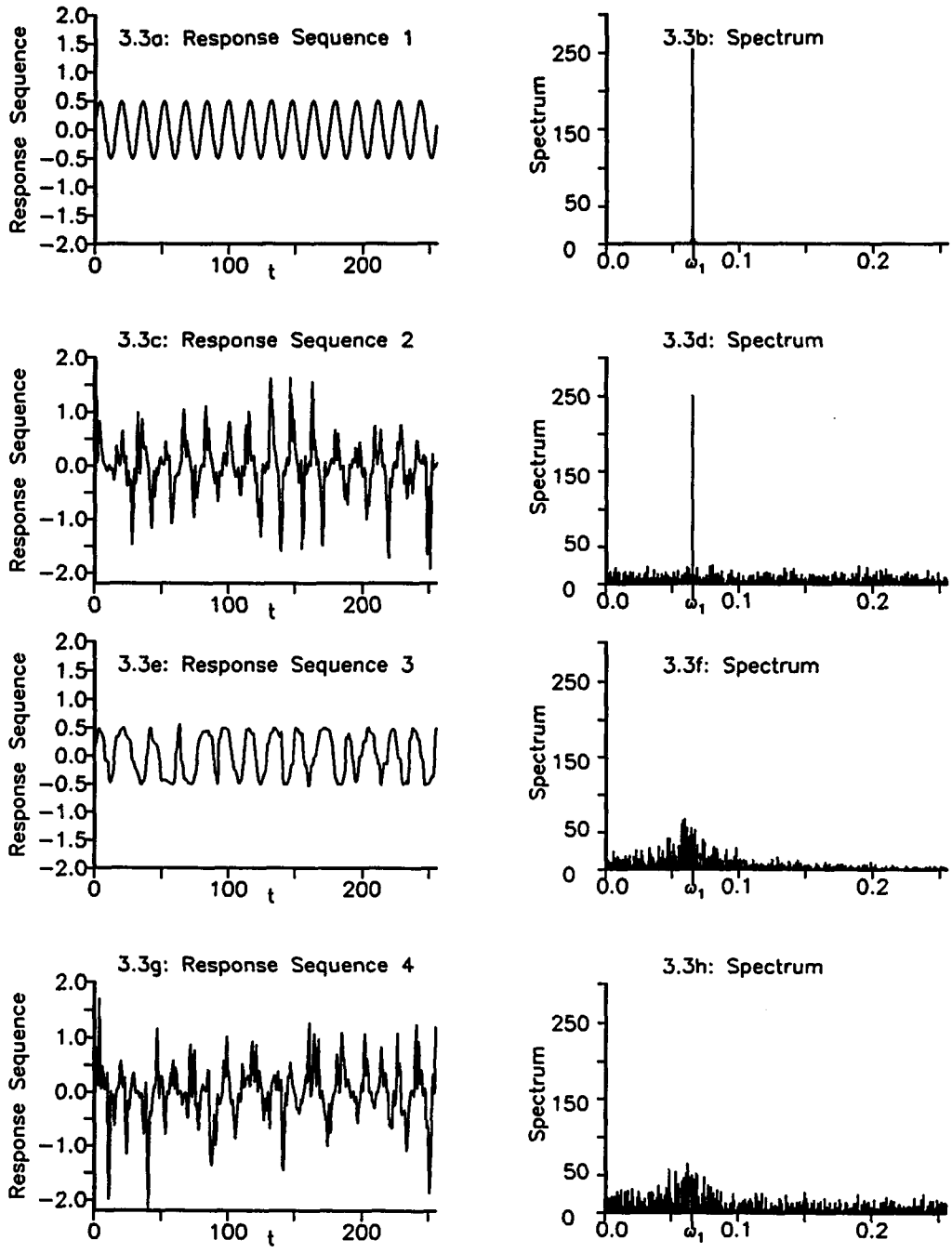


Figure 3.3: Response sequence and spectra for the experiments in example 3.3.

is used as the oscillation parameter, care should be taken to generate an *equally spaced* response sequence. Because Morrice et al. [26] used the stochastic sampling approach when using the simulation clock time, they were unable to get satisfactory results.

3.2 FDE Histogram (Integration) Method

In this section a correct way to use the simulation clock time as the FDE oscillation parameter is developed. This development is based on the established theory of counting processes.

Let $\{N(t),\ t \geq 0\}$ be a nonstationary counting process with rate $\lambda(t)$. That is, $N(t)$ represents the number of events that occur in $[0, t]$. Let $m(t) = E[N(t)]$ be the *expectation* function for the counting process. If $m(t)$ is differentiable for all t then

$$m(t+\delta) - m(t) = \int_t^{t+\delta} \lambda(y)dy \tag{3.6}$$

provided $dm(t)/dt$ is bounded on $[t, t+\delta]$ and is continuous for all but finitely many points in $[t, t+\delta]$ [18]. If δ is small then from equation 3.6 it follows that

$$\lambda(t) \approx \frac{m(t+\delta) - m(t)}{\delta} = \frac{\bar{n}(t,\delta)}{\delta} \tag{3.7}$$

where $\bar{n}(t,\delta) = m(t+\delta) - m(t)$ is the expected number of events in the interval $[t, t+\delta]$.

Equation 3.7 can be used to estimate any rate function, e.g., an arrival rate, service rate or departure rate, by counting the associated number of events. The

FDE Histogram method is a data analysis method that uses equation 3.7 to estimate the value of a rate function in intervals $(i\delta, (i+1)\delta]$ for $i = 0, 1, 2, \ldots, n-1$ and thereby defines an *equally-spaced* response sequence. This method can be used for FDE data analysis when the selected system response statistic is a rate function. In traditional FDEs the expected wait in the system, the utilization or the expected number in the system is usually the selected response statistic of interest. The *extended* FDE Histogram method, developed in chapter 4 and based on the FDE Histogram method, should be used in this case.

An outline of the FDE Histogram method is presented next, followed by a pseudo-code implementation.

1. **Setup:** Select the simulation stopping time T, the number of histogram bins n and the histogram bin width δ. The parameters T, n and δ should be selected so that $T = T_1 + n\delta$, where T_1 is a warm-up time and $n\delta$ is the time over which data is collected. Initialize the histogram. Select the number of replications S.

2. **Simulation:** Select a rate function as the response statistic of interest. For each replication, vary the selected system parameters x_1, $x_2, \ldots$ sinusoidally with frequency ω_1, $\omega_2, \ldots$ and build a histogram of the number of events corresponding to the response rate of interest.

3. **Postprocess:** After S replications, estimate the response rate for each bin by dividing the number of events in each bin by $S\delta$; the result defines the

response sequence. Then calculate the DFT of the response sequence and compute the magnitude of the complex DFT to perform the spectral analysis.

Pseudo-code

```
for i:= 0 to n-1 do { Setup }
  sum[i] := 0.0;
for q:= 1 to S do { Perform S replications }
begin
  t := 0.0;
  while (t <= T) do { Simulate and build the histogram }
  begin
    t := Time of next event e;
    Process  next event e;
    if (e = event of interest) then
    begin
      i := t div (bin width);
      sum[i] := sum[i] + 1
    end
  end
end;
for i:=0 to n-1 do { Determine the response sequence }
  rate[i] := sum[i]/(S*(bin width));
DFT(rate[0..n-1]) { Calculate the DFT of the response sequence }
PSD(rate[0..n-1]) { Compute the magnitude of the complex DFT }
```

To implement the FDE Histogram method effectively, the parameters S, T, n, δ, T_1 and the oscillation frequencies ω_1, ω_2,... need to be selected properly. Although exact mathematical justification for particular choices of these parameters is difficult, some guidelines can be provided. The following discussion provides these guidelines.

- (Choose n) The integer n should be a power of two so that a standard FFT algorithm can be used for calculating the DFT of the response sequence.

- (Choose ω_1, ω_2,...) The frequencies ω_1, ω_2,... should be selected so that

$\omega_j = C_j/n < 0.5$ for all j, where C_j is an integer constant. If $\omega_j \neq C_j/n$, something other than integer multiples of the oscillation period is sampled, thereby causing smearing and noise in the spectrum. The constraint $\omega_j < 0.5$ ensures that under-sampling, which can result in spurious spikes (aliasing) in the spectrum due to frequency folding, is avoided.

- (Choose δ) The choice of δ is important but somewhat arbitrary. Some general guidelines can be provided however. Let $\omega_0 = \max\{\omega_1, \omega_2, \ldots\}$. Assuming all system parameters $x_1, x_2, \ldots$ are rates and are measured in the same units, let $\bar{x}_{\min} = \min\{x_1(0), x_2(0), \ldots\}$. If $\delta \geq 1/\bar{x}_{\min}$ and S is large then on average the histogram bins corresponding to each of $x_1, x_2, \ldots$ are not empty. If $\delta \leq 1/2\omega_0$ then aliasing is avoided. (This is a restatement of Shannon's sampling theorem [42].) Hence the ideal solution is the choice of bin size such that

$$1/\bar{x}_{\min} \leq \delta \leq 1/2\omega_0.$$

  These requirements for selecting δ also implicitly define constraints for the selection of the nominal values of the system parameters and the oscillation frequencies. Other constraints for FDE oscillation frequency selection are presented in [10].

- (Choose T_1) As mentioned earlier, T_1 is a warm-up time. For the examples presented in section 3.3, the results indicate that warm-up is not necessary, hence $T_1 = 0$ is used.

- (Choose S) Consistent with the discussion in section 2.1, to reduce noise in the response spectrum, S replications are performed and the ensemble average of the histogram counts is used as the response sequence. A compromise between the desired noise sensitivity and the number of replications is needed. For the examples presented later in this section and in section 3.3, $S = 100$ is sufficient to produce a distinct spike in the response spectrum at ω_1.

For completeness, it should be mentioned that, an alternate approach for building the response sequence histograms is to introduce a new sampling event into the simulation. The sampling event is invoked at times $\delta, 2\delta, 3\delta \ldots$ An event counter is maintained for the event of interest and the event counter is sampled to determine the number of events of interest that occurred in the interval $((i-1)\delta, i\delta]$ for $i = 1, 2, \ldots, n$. The number of events in each interval, tallied over S replications, is determined and the ensemble average is divided by δ to yield the desired response rate sequence. Because this alternate approach requires significant modification of an existing simulation program, it is not simple to implement.

Although the software engineering implications of this statement are not investigated in this dissertation, it is important to note that unlike the alternate approach, the FDE Histogram method is very simple to incorporate into an existing discrete-event simulation experiment. Therefore for this dissertation the simpler FDE Histogram method is used for building the histograms.

As a simple illustration of the FDE Histogram method, a non-stationary Poisson

arrival process is simulated using the thinning method. Customers arrive at a service facility according to a non-stationary Poisson process with arrival rate

$$\lambda(t) = 1.00 + 0.10 \sin(2\pi\omega_1 t) \tag{3.8}$$

The thinning method is used to simulate the arrival process. The arrivals constitute the events of interest. The histogram bin size is $\delta = 1$ and $T = n = 4096$. $T_1 = 0$ is used because the resulting spectra indicate that warmup is not necessary. The number of replications is $S = 100$.

Figures 3.4a, 3.4c, 3.4e and 3.4g illustrate a portion of the resulting ensemble averaged histogram corresponding to four different values of ω_1. In each case, the histogram is a noisy estimate of $\lambda(t)$. The noise level can be reduced, if desired, by averaging additional replications. However, as figures 3.4b, 3.4d, 3.4f and 3.4h illustrate, additional noise suppression is not necessary because, as desired, the corresponding spectral estimates exhibit a distinct spike at ω_1. Four different frequencies are chosen to demonstrate that the FDE Histogram method works well over a wide frequency band.[2]

3.3 FDE Histogram Method—Examples

Several examples are presented in this section to demonstrate the effectiveness of the FDE Histogram method. Example 3.4 is a FDE for a $M/G/1$ queue. Example

[2]The horizontal scale for figures 3.4f and 3.4h is different from that of figures 3.4b and 3.4d to clearly exhibit the spike at ω_1 in each case.

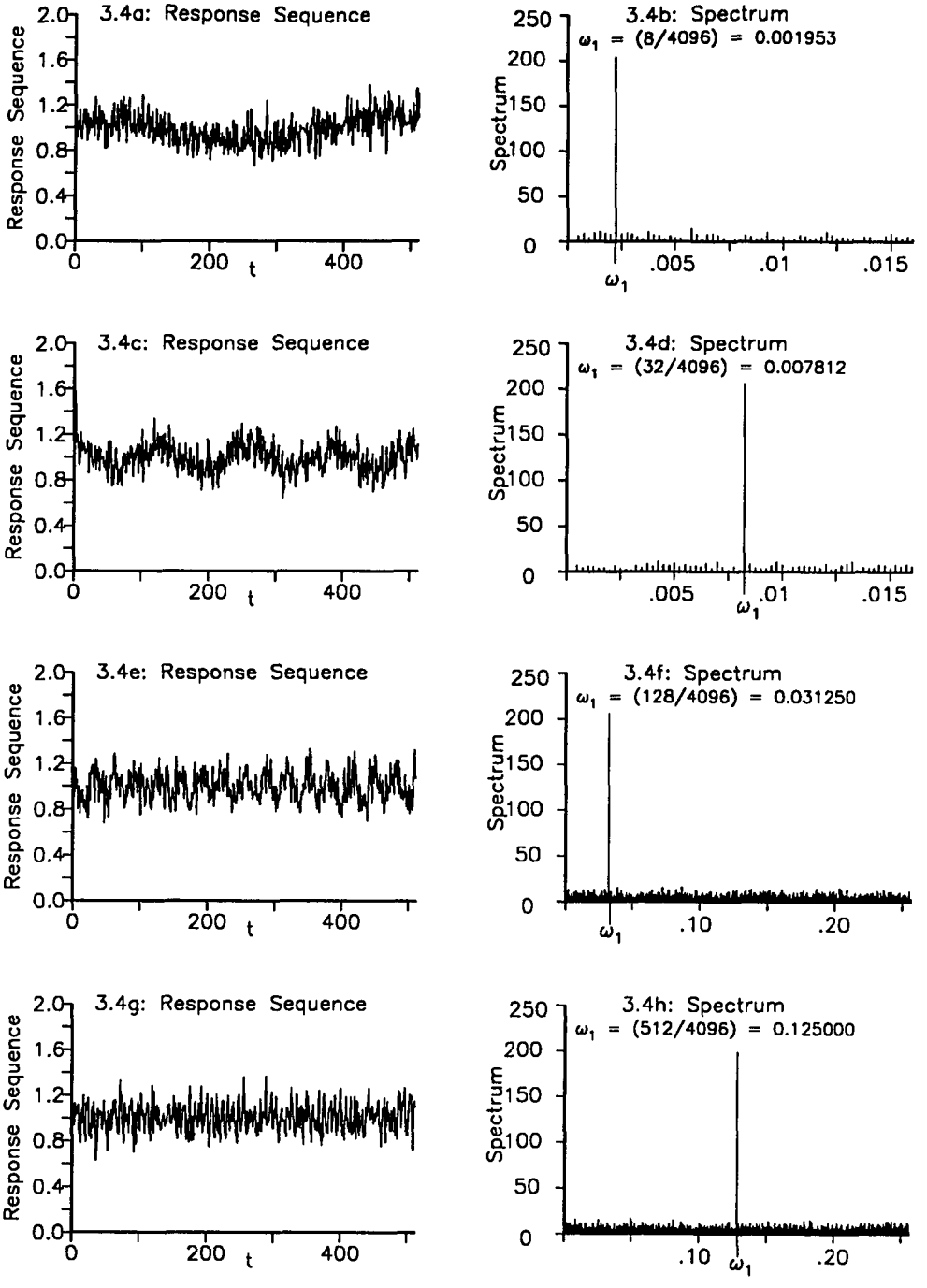


Figure 3.4: An arrival process response sequence and spectra using the FDE Histogram method.

3.5 is a FDE for a tandem network of single-server queues. Examples 3.6 and 3.7 present FDEs for two general networks of single-server queues.

Example 3.4 A FDE for a $M/G/1$ queue is performed using the FDE Histogram method. Customers arrive as a non-stationary Poisson process, with an arrival rate

$$\lambda(t) = 1.0 + 0.1 \sin(2\pi\omega_1 t). \tag{3.9}$$

The thinning method is used to simulate the arrival process. As in section 3.2, four different oscillation frequencies, $\omega_1 = 0.001953,\ 0.007812,\ 0.031250$ and 0.125000 are considered to demonstrate that the FDE Histogram method works well for a range of frequencies. The service rate is fixed at $\mu = 2$. Four service time distributions are considered

1. Exponential with mean 0.5 (standard deviation 0.5);
2. Uniform between 0.35 and 0.65 (standard deviation 0.0866);
3. Erlang with shape parameter 4.0 and scale parameter 0.125 (standard deviation 0.25);
4. Lognormal with scale parameter -0.698 and shape parameter $\sqrt{0.1} = 0.31622$ (standard deviation 0.05).

A FDE for the $M/G/1$ queue is performed to determine the sensitivity of the departure rate to variations of the arrival rate. Figures 3.5a–3.5d illustrate the departure

Figure 3.5a: ω_1=0.001953

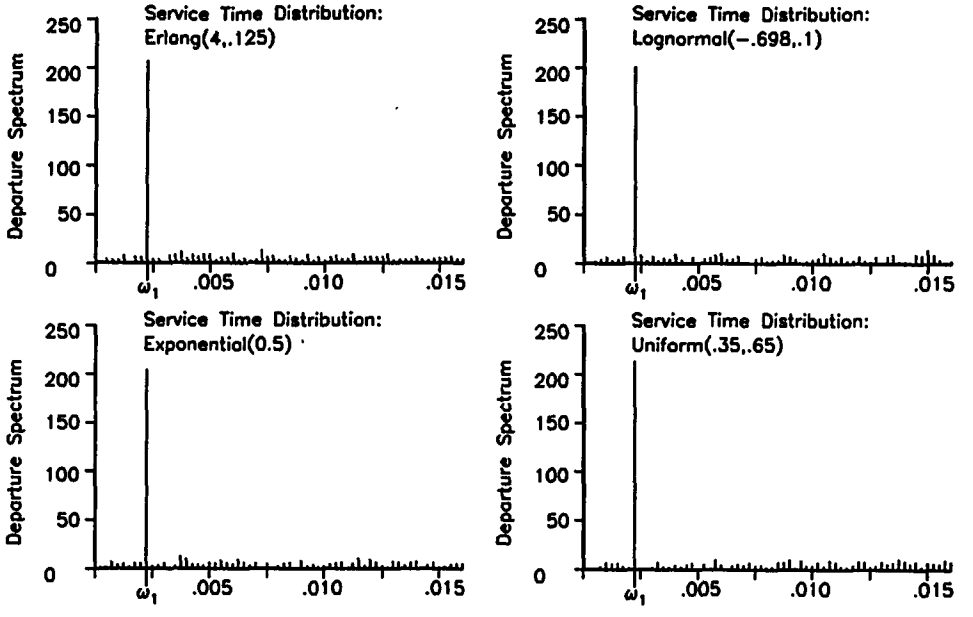


Figure 3.5b: ω_1=0.007812

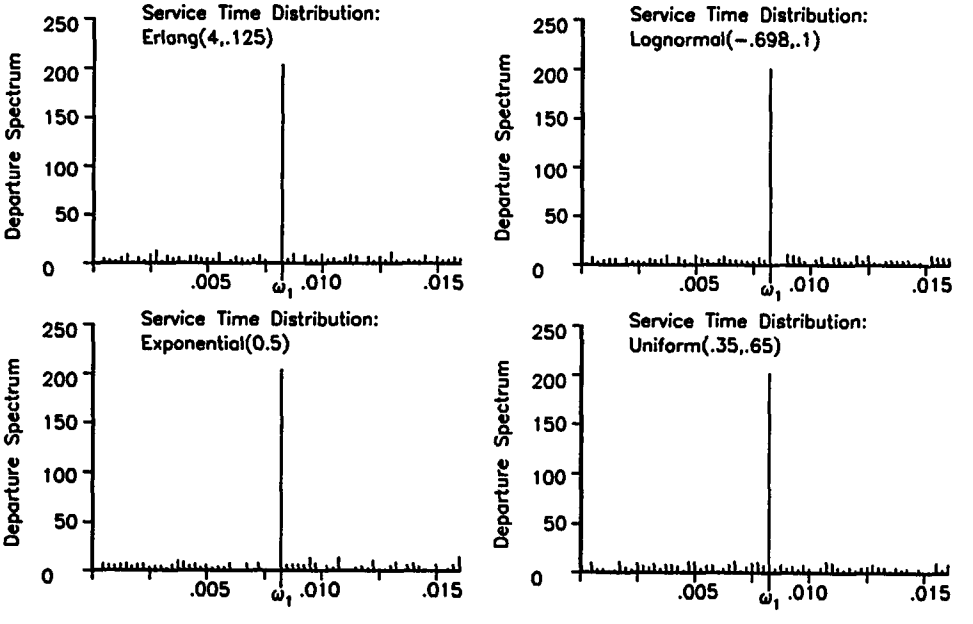


Figure 3.5: Departure rate spectra for M/G/1 queues, $\omega_1 = 0.001953$ and $\omega_1 = 0.007812$.

Figure 3.5c: ω_1=0.031250

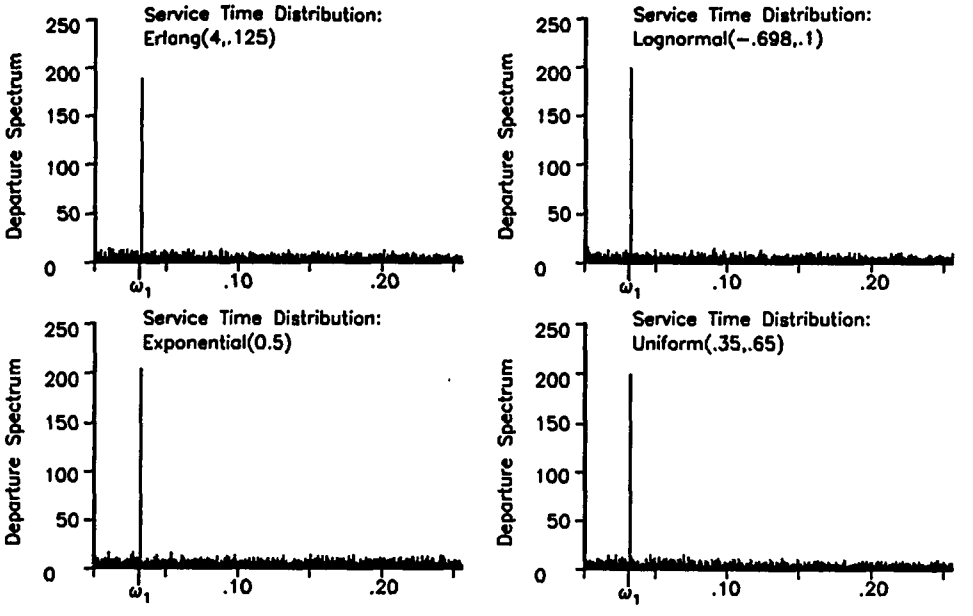


Figure 3.5d: ω_1=0.125000

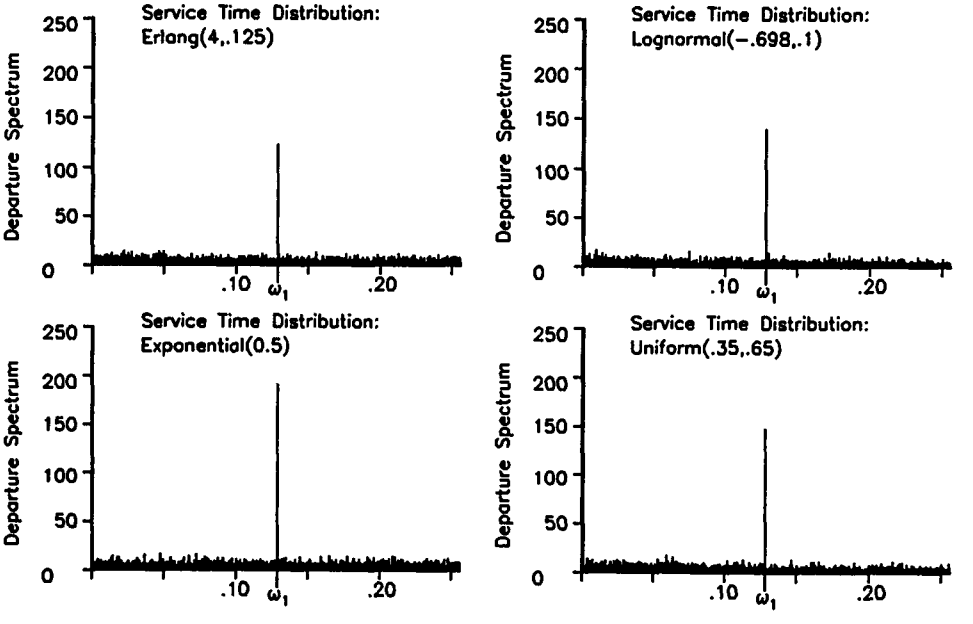


Figure 3.5: Departure rate spectra for M/G/1 queues, $\omega_1 = 0.031250$ and $\omega_1 = 0.125000$.

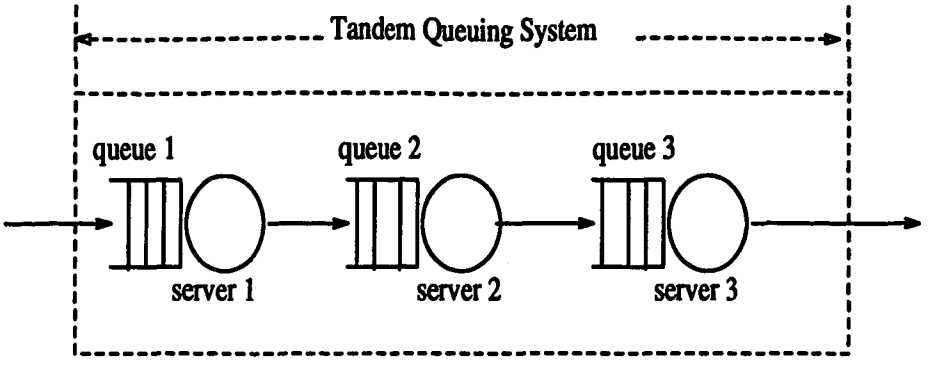


Figure 3.6: Tandem of queues.

rate spectra obtained by using the FDE Histogram method with the histogram bin width $\delta = 1$, $S = 100$ replications, $T = n = 4096$ and $T_1 = 0$.

In vivid contrast to the smeared spectrum obtained when naively using the simulation clock time (figure 3.1), each of figures 3.5a–3.5d have a distinct spike in the departure rate spectrum at ω_1. The distinct spikes at ω_1 demonstrate that unlike traditional FDEs, the FDE Histogram method effectively and correctly uses the simulation clock time as the FDE oscillation parameter.[3]

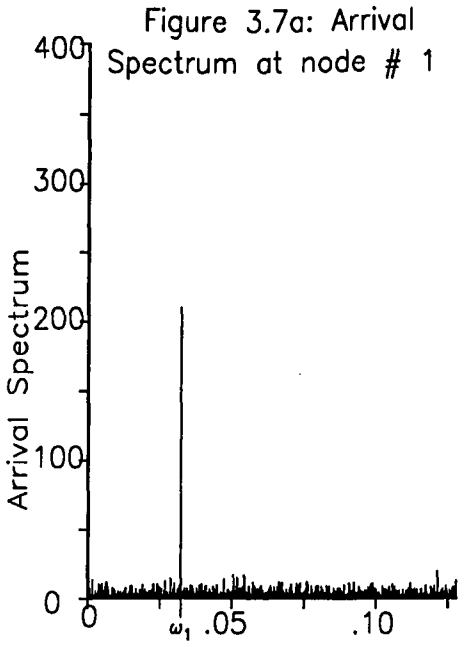
Example 3.5 A FDE for a tandem network of three infinite capacity $M/G/1$ queues (figure 3.6) is performed using the FDE Histogram method. External arrivals take place at node 1 according to a non-stationary Poisson process, with the arrival rate given by equation 3.9 with $\omega_1 = 0.031250$. After exiting one node cus-

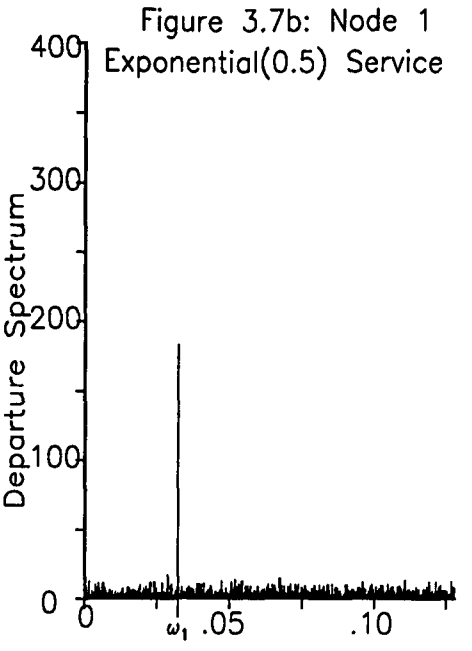
[3]The height of the spike at ω_1 is proportional to the standard deviation of the service time distribution; i.e., the spike corresponding to the service time distribution with the smallest standard deviation (Lognormal) is the shortest while that corresponding to the service time distribution with the largest standard deviation (Exponential) is the tallest.

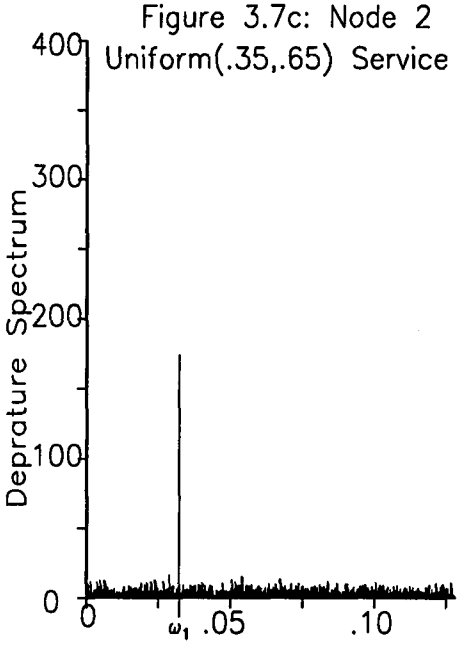
tomers move to the next node in series and finally leave the system from node 3. Server 1 has an exponentially distributed service time with mean 0.5. The mean service time distribution of server 2 is Uniform between 0.35 and 0.65 and that of server 3 is Erlang with shape parameter 4.0 and scale parameter 0.125. The service rate at each node is 2.

The departure rate from each node is determined using the FDE Histogram method with $T = n = 4096$, $\delta = 1.0$, $T_1 = 0$ and $S = 100$. Figure 3.7a depicts the arrival rate spectrum at node 1. Figures 3.7b–3.7d depict the departure rate spectra for nodes 1, 2 and 3 respectively. For nodes 2 and 3 the arrival process is no longer Poisson. The departure rate spectrum from each node still exhibits a distinct spike at the frequency of input oscillation, however, indicating that (as expected) the departure rate at each node in a tandem network of $M/G/1$ queues is sensitive to variations in the external arrival rate. The distinct spike in each spectrum at ω_1 demonstrates the effectiveness of the FDE Histogram method in performing a FDE for a tandem network of $M/G/1$ queues.

Example 3.6 A FDE for a *feed-forward* network of queues (figure 3.8) is performed using the FDE Histogram method. Each node is a single-server $M/G/1$ FIFO queue with mean service rate $\mu = 2$. Server 1 has exponentially distributed service times with mean 0.5; the service time distribution of server 2 is Uniform between 0.35 and 0.65, the service time distribution of server 3 is Erlang with shape parameter 4.0 and scale parameter 0.125 and that of server 4 is Lognormal with scale pa-







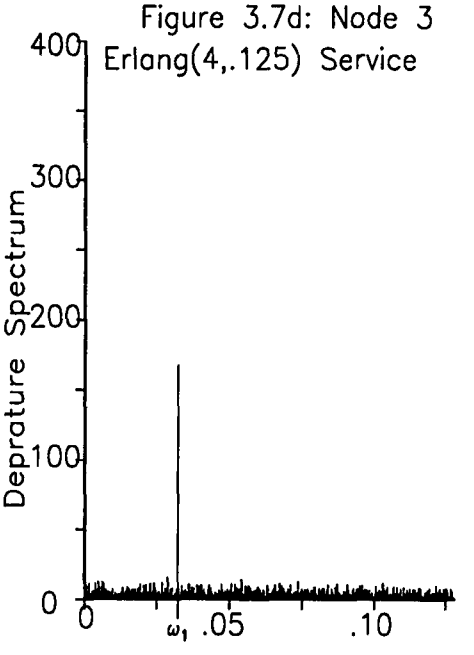


Figure 3.7: Arrival and departure rate spectra for a tandem network of three $M/G/1$ queues; $\lambda(0) = 1.0$; $\alpha = 1.0$; $\omega_1 = 0.031250$.

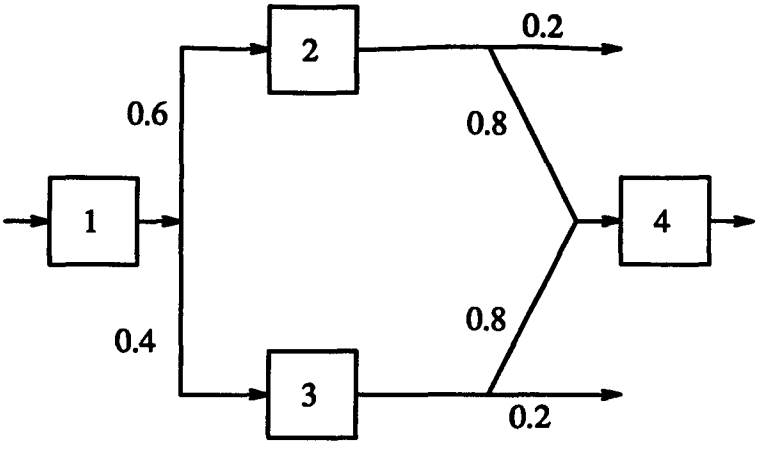


Figure 3.8: Feed-forward network of single server FIFO queues.

rameter -0.698 and shape parameter $\sqrt{0.1} = 0.31622$. External arrivals occur at node 1 according to a non-stationary Poisson process with the arrival rate given by equation 3.9 with $\omega_1 = 0.031250$. After exiting one node, a customer selects its next destination (another node or the outside world) consistent with the indicated transition probability associated with each path. The departure rate at each node is selected as the response statistic of interest and the FDE Histogram method is used to obtain the departure rate spectrum at each node.

Figure 3.9 is the arrival rate spectrum and figure 3.10 is the departure rate spectra for each node. As in the much simpler case of a single-server queue, the response spectra at each node exhibits a distinct spike at the frequency of input oscillation, despite the fact that the arrival process to each of the internal nodes in the network is no longer Poisson. Let $\hat{y}_1$, $\hat{y}_2$, $\hat{y}_3$, $\hat{y}_4$, $\hat{y}_5$ denote the height of the spikes at ω_1 in the departure rate spectra for node $1, 2, 3, 4, 5$ respectively. It is

observed that

$$\hat{y}_2 \approx 0.6\hat{y}_1 \tag{3.10}$$

where 0.6 is the transition probability from node 1 to node 2,

$$\hat{y}_3 \approx 0.4\hat{y}_1 \tag{3.11}$$

where 0.4 is the transition probability from node 1 to node 3, and

$$\hat{y}_4 \approx 0.8\hat{y}_2 + 0.8\hat{y}_3 \tag{3.12}$$

where 0.8 is the transition probability from node 2 to node 4 and from node 3 to node 4. Equations 3.10–3.12 are used in chapter 6 to show analytically that the height of the spike at ω_1 in the departure rate spectrum of each node is related to that of the other departure rate spectra via the transition probabilities of the network.

Example 3.7 A FDE for a *feedback* network of $M/G/1$ queues (figure 3.11) is performed using the FDE Histogram method. As in the feed-forward network, each node is a single-server $M/G/1$ FIFO queue with mean service rate $\mu = 2$. Servers 1 and 5 have exponentially distributed service times; the service time distribution of server 2 is Uniform between 0.35 and 0.65, the service time distribution of server 3 is Erlang with shape parameter 4.0 and scale parameter 0.125 while that of server 4 is Lognormal with scale parameter -0.698 and shape parameter $\sqrt{0.1} = 0.31622$. External arrivals occur at node 1 according to a non-stationary Poisson process with rate given by equation 3.9 with $\omega_1 = 0.031250$. After exiting one node, a customer

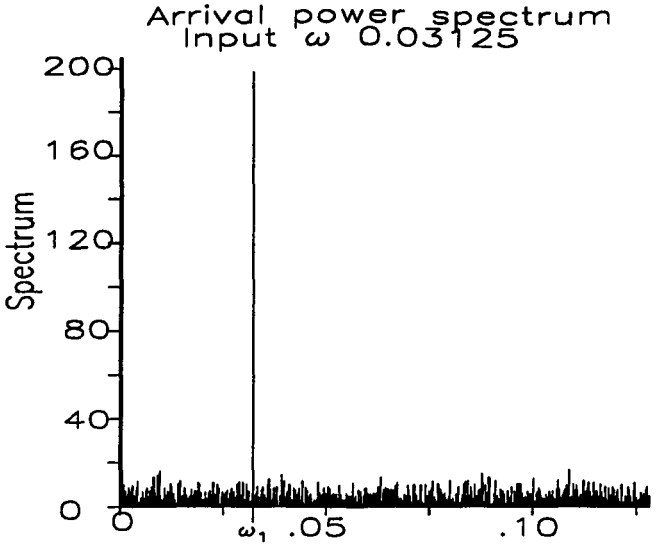


Figure 3.9: Arrival rate spectrum for the feed-forward network in figure 3.8; $\lambda(0) = 1.0; \alpha = 1.0; \omega_1 = 0.031250.$

selects its next destination (another node or the outside world) consistent with the indicated transition probability associated with each path. The departure rate at each node is selected as the response statistic of interest and the FDE Histogram method is used to obtain the departure rate spectrum at each node.

Figure 3.12 is the external arrival rate spectrum and 3.13 the departure rate spectra for each node, respectively. As in the case of the feed-forward network (example 3.6), the departure rate spectra at each node is seen to exhibit a distinct spike at the frequency of input oscillation. In chapter 6 it is demonstrated analytically that the height of the spike at ω_1 in the departure rate spectrum of each node is related to that of the other departure rate spectra via the transition probabilities of the network.

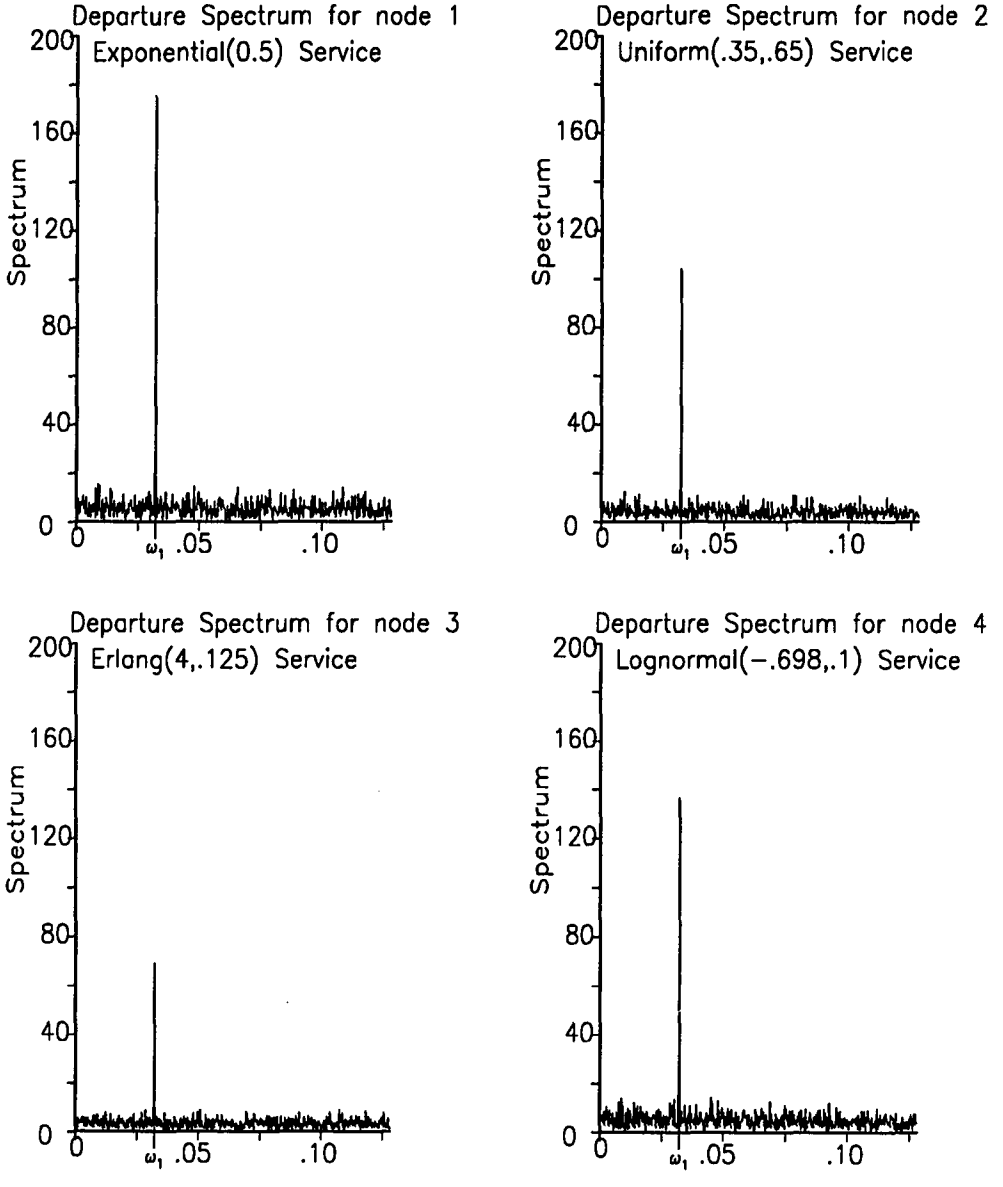


Figure 3.10: Departure rate spectra for the feed-forward network in figure 3.8; $\lambda(0) = 1.0; \alpha = 0.1; \omega_1 = 0.031250.$

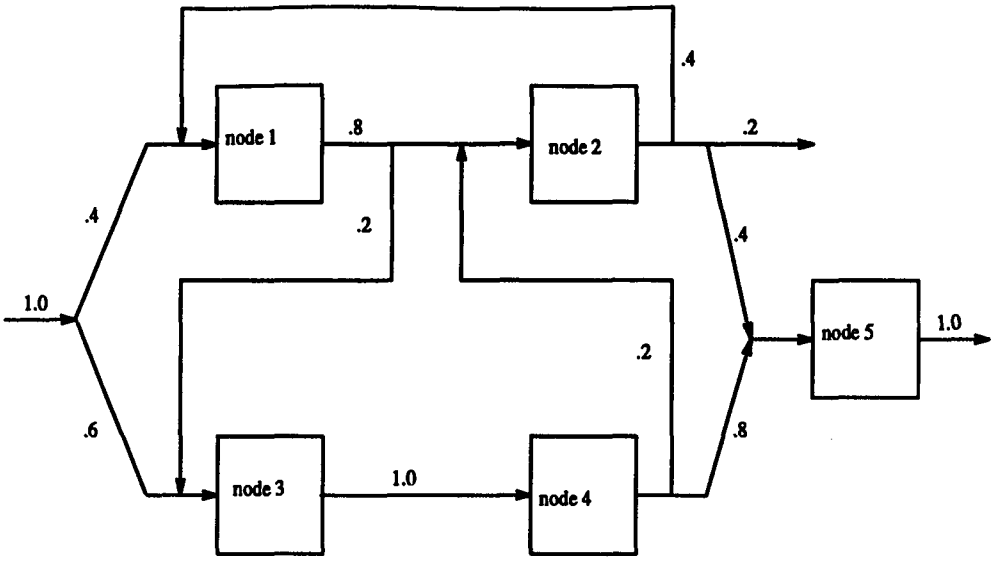


Figure 3.11: Feedback network of single server FIFO queues.

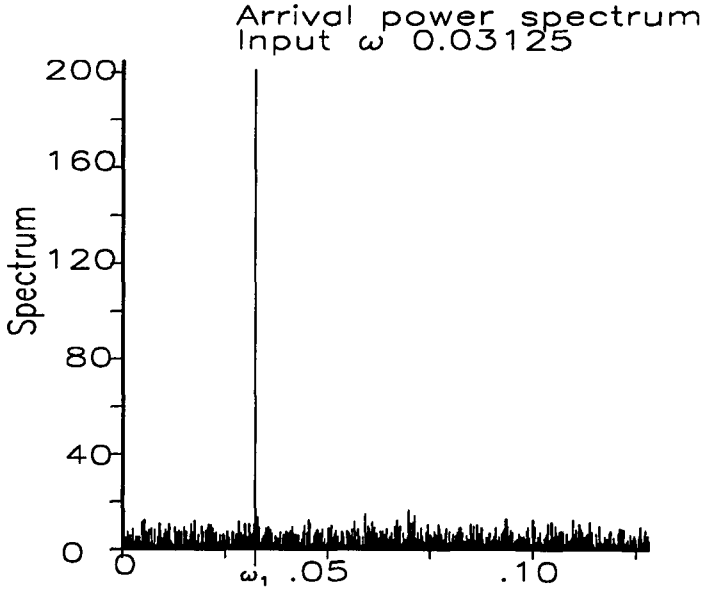


Figure 3.12: Arrival rate spectrum for the feedback network in figure 3.11; $\lambda(0) = 1.0; \alpha = 1.0; \omega_1 = 0.031250.$

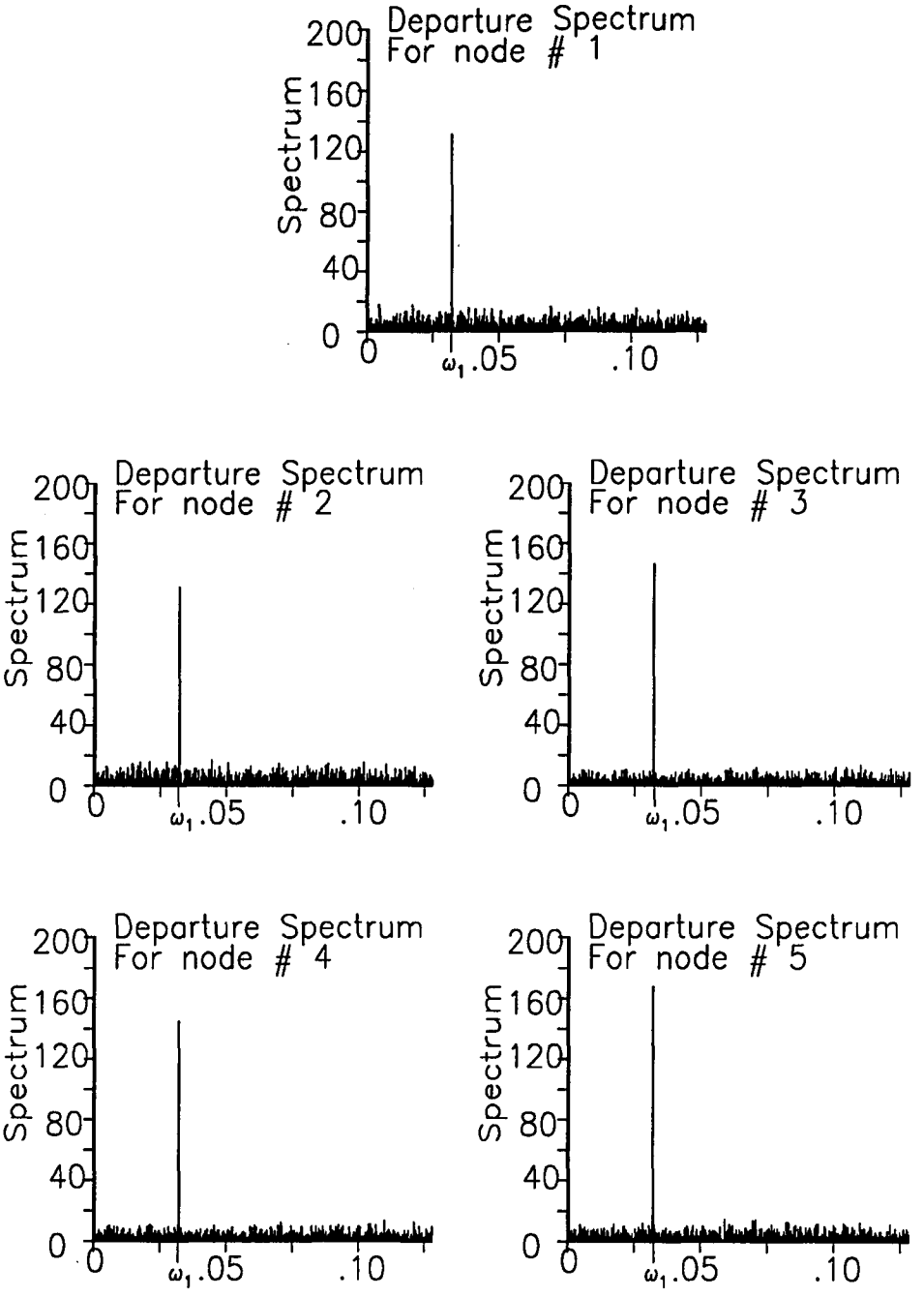


Figure 3.13: Departure rate spectra for the feedback network in figure 3.11; $\lambda(0) = 1.0$; $\alpha = 0.1$; $\omega_1 = 0.031250$; Average service rate per node = 2.0.

The FDE Histogram method derived in this chapter solves the existing FDE indexing problem by using the simulation clock time as the oscillation parameter and by generating uniformly-spaced samples for the response sequence. The FDE Histogram method can be effectively used when the selected response statistic is a rate function. The *extended* FDE Histogram method presented in chapter 4 can be used when the selected response statistic is not a rate function. Correct implementations of examples 2.2–2.4 are presented in chapter 4, using the extended FDE Histogram method.

Similar to traditional FDEs, the FDE Histogram method presented in this chapter is based on the assumption that sinusoidal oscillation of a system parameter at a selected frequency induces oscillations at the same frequency in a system response statistic, provided that the response statistic is sensitive to the oscillated parameter. This frequency-invariance assumption has been empirically verified by several FDE practitioners, but never theoretically defended. Theoretical support for this assumption is presented in chapter 5 for $M/M/1$ queues. In **chapter 6**, the frequency-invariance assumption is shown to be true for networks of such queues.

CHAPTER IV

EXTENDED FDE HISTOGRAM METHOD

4.1 Extended FDE Histogram Method

The FDE Histogram method presented in chapter 3 is conditioned on the selected response statistic being a rate function; e.g., an arrival rate or a departure rate. Unfortunately, most commonly used system response statistics are not rate functions, a fact (implicitly) recognized by early FDE practitioners who devoted their attention to (non-rate) response statistics like the expected wait in the system or the system utilization. The FDE Histogram method is thus of limited direct use. As demonstrated in this chapter, the FDE Histogram method can be easily extended, however, to system response statistics that are not rate functions.

The mathematics used to extend the FDE Histogram method to estimate non-rate system response statistics is presented in section 4.1.1. An outline of the extended FDE Histogram method is presented in section 4.1.2, followed by a pseudo-code implementation in section 4.1.3. Specific examples of estimating the expected wait and utilization for a $M/G/1$ queue using the extended FDE Histogram method are given in sections 4.1.4 and 4.1.5 respectively. In section 4.2, a method for estimating the expected number in the system is outlined. The effectiveness of the extended FDE Histogram method is further demonstrated in section 4.3 by

performing a FDE for the feedback queuing systems and simple manufacturing system discussed in chapter 2.

4.1.1 Mathematical Basis

Corresponding to the non-stationary counting process $\{N(t), t \geq 0\}$ defined in section 3.2, let $\bar{n}(t, \delta)$ be the expected number of events occurring in the interval $(t, t + \delta]$. That is, $\bar{n}(t, \delta) = m(t + \delta) - m(t)$, where $m(t) = E[N(t)]$. The FDE Histogram method estimates $\bar{n}(t, \delta)$ via replication—the time line $t > 0$ is divided into intervals of length δ and during the q^{th} replication of the process the total number of events $n_q(t, \delta)$ that occur in the interval $(t, t + \delta]$ is determined. If the number of replications S is large, then $\bar{n}(t, \delta)$ can be estimated by

$$\bar{n}(t, \delta) \simeq \frac{1}{S} \sum_{q=1}^{S} n_q(t, \delta). \tag{4.1}$$

Let $u(t)$ be the expected value of the system response statistic that is to be estimated, e.g., the expected wait in a $M/G/1$ queue. Let $u_q(t, \delta)$ represent the sum of the responses (e.g., waiting times) associated with the $n_q(t, \delta)$ events that occur in the interval $(t, t + \delta]$ during the q^{th} replication of the system. Let $\bar{u}(t, \delta)$ represent the sum of the responses accumulated in the interval $(t, t + \delta]$ averaged over S replications. Then $\bar{u}(t, \delta)$ can be estimated by

$$\bar{u}(t, \delta) \simeq \frac{1}{S} \sum_{q=1}^{S} u_q(t, \delta). \tag{4.2}$$

The expected response in each interval is the ratio of the sum of the responses associated with all the events occurring in the interval to the total number of events

occurring in the interval. Equivalently, in each interval $(t, t+\delta]$ the expected response is

$$u(t,\delta) \simeq \frac{\bar{u}(t,\delta)}{\bar{n}(t,\delta)} \tag{4.3}$$

Provided S is large and δ is small, the system response statistic $u(t)$ is approximately equal to $u(t,\delta)$.

4.1.2 Outline

The outline for the extended FDE Histogram method is:

1. **Setup:** Select the simulation stopping time T, the number of histogram bins n and the histogram bin width δ. As in chapter 3, the parameters T, n and δ should be selected so that $T = T_1 + n\delta$, where T_1 is a warm-up time and $n\delta$ is the time over which data is collected. Initialize the histogram. Select the number of replications S.

2. **Simulation:** Select a system response statistic of interest. For each replication, vary the selected system parameters $x_1,\ x_2, \ldots$ sinusoidally with frequency $\omega_1,\ \omega_2, \ldots$, determine the number of associated events of interest occurring in the intervals $(0,\ \delta], (\delta,\ 2\delta], (2\delta,\ 3\delta], \ldots, ((n-1)\delta,\ n\delta]$ and sum the corresponding response in each interval.

3. **Post process:** After S replications, estimate the average response in each interval by using equation 4.3; the resulting histogram values define the response

sequence. Then calculate the DFT of the response sequence and compute the magnitude of the complex DFT to perform the spectral analysis.

4.1.3 Pseudo-code

```
type
   hist_record  = record
                     u: double;
                     n: longint;
                  end;
var
  hist       : array [0..n-1] of hist_record;

for i:= 0 to n-1 do  { Setup }
begin
  hist[i].u := 0.0;
  hist[i].n := 0;
end
for q:= 1 to S do   { Perform S replications }
begin
  t := 0.0;
  while (t <= T) do  { Simulate and build histogram }
  begin
    t := Time of next event e;
    Process next event e;
    if (e = event of interest) then
    begin
        i := t div (bin width);            { determine bin for t }
        hist[i].n := hist[i].n + 1;        {  Build histogram }
        hist[i].u := response for event e  { Sum response }
    end
  end
 end;
 for i:= 0 to n-1 do       { Determine the response sequence }
   average[i] := hist[i].u/hist[i].n;
 DFT(average[0..n-1]); { Calculate the DFT of the response sequence }
 PSD(average[0..n-1]); { Compute the magnitude of the complex DFT }
```

The guidelines for the selection of T, K, δ, S and ω_1, ω_2,... for the FDE Histogram method hold for the extended FDE Histogram method as well.

4.1.4 Estimating the Expected Wait

To estimate the expected wait in a system, $n_q(t, \delta)$ is defined as the number of arrivals in the interval $(t, t + \delta]$ during the q^{th} replication and $u_q(t, \delta)$ is defined as the sum of the waits experienced by these customers. The expected wait (response) sequence can then be obtained by using equation 4.3.

Example 4.1 A FDE for a single-server $M/G/1$ FIFO queue is performed using the extended FDE Histogram method. The service time of a customer entering service at time t is drawn from an Erlang distribution with shape parameter 4.0 and sinusoidally varying scale parameter $0.25 + 0.025\sin(2\pi\omega_1 t)$ where $\omega_1 = 0.031250$. The average service rate is 1.0. The arrival rate is fixed at $\lambda = 0.5$. The waiting time in the system of the i^{th} customer arriving at time t_i is selected as the response statistic of interest; i.e., t_i is the oscillation index and i is the sampling index.[1] The extended FDE Histogram method is used to build a histogram of the time-dependent expected waiting time in the system. A histogram bin size of $\delta = 1$ is used and $T = n = 4096$. The resulting spectrum indicates that warmup is not necessary, so $T_1 = 0$. The number of replications is $S = 100$.

Figure 4.1a depicts the expected wait response sequence and figure 4.1b shows the corresponding spectrum. Sinusoidal variation of the service time at frequency ω_1 induces similar oscillations in the expected wait in the system at the same frequency.

[1]The arrival time t_i of the i^{th} customer is stored and at the departure time of the i^{th} customer, the wait in the system of the i^{th} customer is determined and binned using t_i.

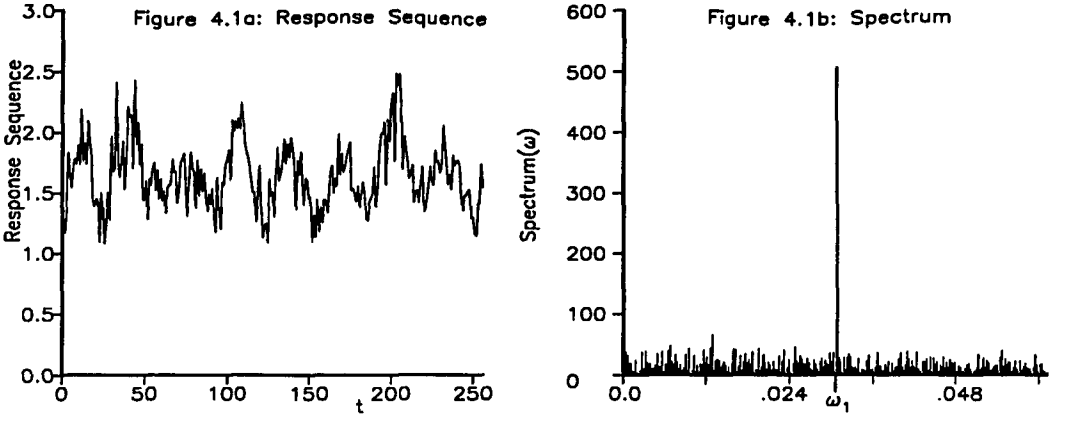


Figure 4.1: Response sequence and corresponding spectrum for the expected wait in a M/G/1 queue, $\omega_1 = 0.031250$.

As expected and as indicated by the spectrum, the expected wait is sensitive to variations of the service rate. Unlike example 3.1, where Morrice et al.'s incorrect FDE approach resulted in a noisy spectrum with spurious spikes, figure 4.1b has a distinct spike at ω_1 with no spurious spikes. The results in figure 4.1b demonstrate the effectiveness of the extended FDE Histogram method in performing FDE data analysis for a single-server queue if the simulation clock time is used as the oscillation parameter.

4.1.5 Estimating the Expected Utilization

For the purpose of the extended FDE Histogram method, the traditional definition of system utilization is written in an equivalent form, as the ratio of the total number of departures that leaves the server busy to the total number of departures [18]. Therefore, to estimate the expected system utilization, the number of departures

that occur in the interval $(t, t+\delta]$ during the q^{th} replication is selected to be $n_q(t,\delta)$ and of these departures, the number that leave the server busy is selected to be $u_q(t,\delta)$. The utilization (response) sequence can then be obtained by using equation 4.3.

Example 4.2 A FDE for a $M/G/1$ queue is performed using the extended FDE Histogram method. The system utilization is the response statistic of interest. Customers arrive as a non-stationary Poisson process, with arrival rate

$$\lambda(t) = 1.0 + 0.1\sin(2\pi\omega_1 t) \tag{4.4}$$

and $\omega_1 = 0.031250$. The thinning method is used to simulate the non-stationary Poisson arrival process. Customers join a FIFO queue before a single server. The service time distribution is Lognormal with scale parameter -0.698 and shape parameter $\sqrt{0.1} \approx 0.31622$. The service rate is fixed at $\mu = 2$.

The extended FDE Histogram method is used to analyze the system utilization. As before $\delta = 1$, $T = n = 4096$, $T_1 = 0$ and $S = 100$ is used. Figure 4.2a is the response sequence corresponding to the system utilization, figure 4.2b is the corresponding spectrum. Sinusoidal variation of the arrival rate at frequency ω_1 induces similar oscillations in the system utilization at the same frequency. The spectrum has a distinct spike at ω_1 indicating that the system utilization is sensitive to variations of the arrival rate. Therefore, the extended FDE Histogram method is effective in performing a FDE if the simulation clock time is used as the oscillation parameter.

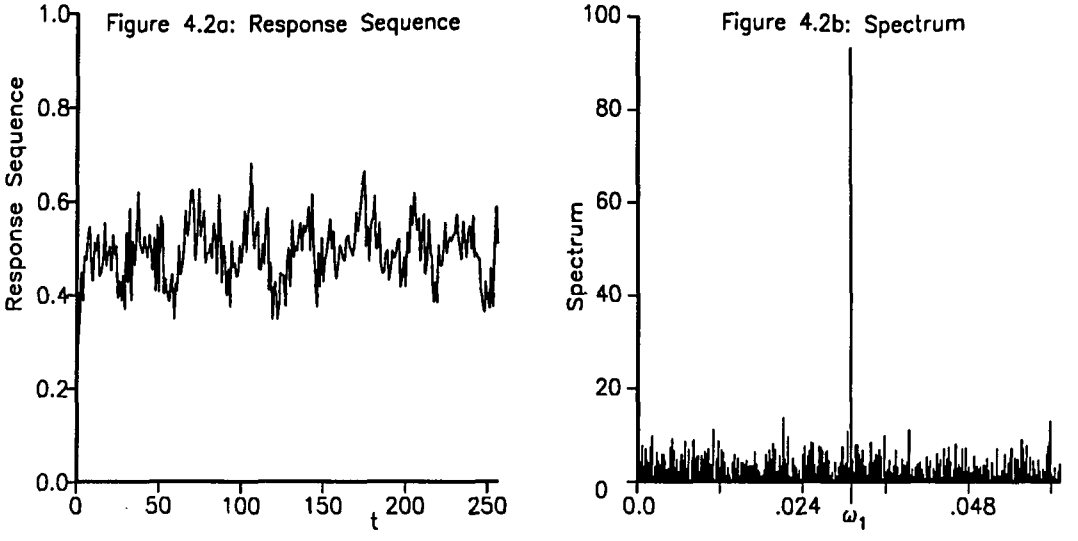


Figure 4.2: Response sequence and corresponding spectrum for the utilization in a M/G/1 queue, $\omega_1 = 0.031250$.

4.2 Estimating the Expected Number in the System

Unlike the expected wait in the system, but similar to the utilization, the number in the system is an instantaneous statistic. Unlike the utilization, however, the number in the system cannot be (naturally) modeled as a counting process. Therefore, instead of using a stochastic counting process, an equally spaced response sequence corresponding to the expected number in the system can be generated by sampling the number in the system at regular time intervals.[2]

To estimate the expected number in the system, a new event is introduced into the simulation. The new event is invoked at times $\delta, 2\delta, 3\delta \ldots$ For each replication,

[2]This method of generating the sequence of expected number in the system was advocated by Dr. Douglas J. Morrice and was first presented in a paper by Mitra, et al. [25].

at time $i\delta$, the number in the system is determined and accumulated over S replications. Provided S is large, the average number in the system at times $\delta, 2\delta, 3\delta \ldots$ constitute the expected number in the system response sequence. The same method can be used to estimate the expected number in the queue.

Example 4.3 A FDE for a $M/G/1$ queue is performed with the expected number in the system as the response statistic of interest. Customers arrive as a non-stationary Poisson process, with arrival rate

$$\lambda(t) = 1.0 + 0.1 \sin(2\pi\omega_1 t) \tag{4.5}$$

and $\omega_1 = 0.031250$. As before, the thinning method is used to simulate the non-stationary Poisson arrival process. Customers join a FIFO queue before a single server. The service time distribution is Uniform between 0.35 and 0.65. The service rate is fixed at $\mu = 2$.

As before $\delta = 1$, $T = n = 4096$, $T_1 = 0$ and $S = 100$ is used. Figure 4.3a is the response sequence corresponding to the expected number in the system and figure 4.3b is the corresponding spectrum. Sinusoidal variation of the arrival rate at frequency ω_1 induces similar oscillations in the expected number in the system at the same frequency. The spectrum exhibits a distinct spike at ω_1 indicating that the expected number in the system is sensitive to variations of the arrival rate. Therefore, the extended FDE Histogram method is effective in performing a FDE if the simulation clock time is used as the oscillation parameter.

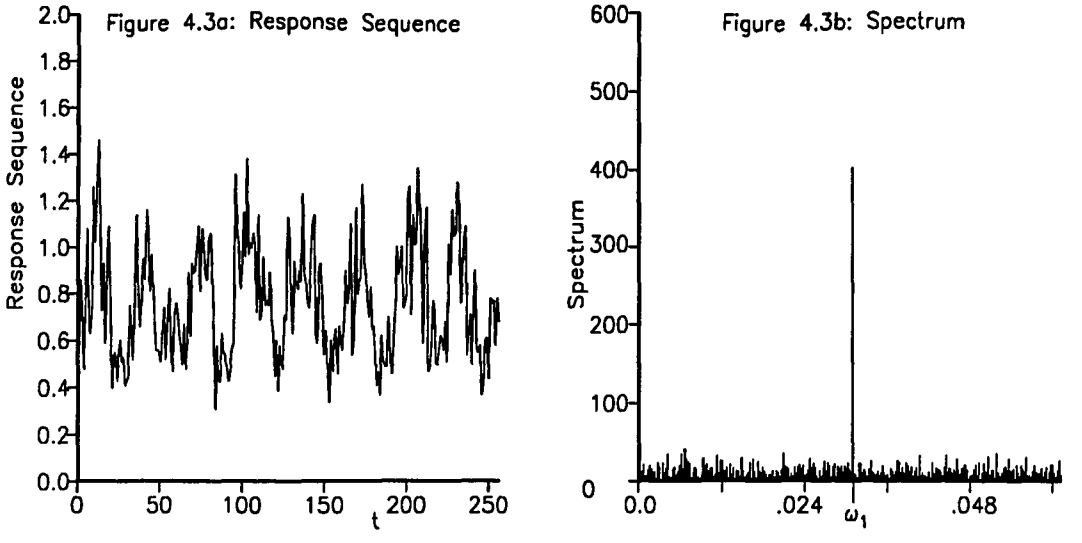


Figure 4.3: Response sequence and corresponding spectrum for the expected number in a $M/G/1$ queuing system, $\omega_1 = 0.03125$.

4.3 Application of the Extended FDE Histogram Method

In section 2.2 it was demonstrated that when traditional FDE methods are used, the oscillation index has to be selected in an application dependent manner. In vivid contrast to traditional FDE methods the extended FDE Histogram method is elegantly simple—all that is required is the variation of the system parameters using the simulation clock time and the sampling of the response statistic using the well-defined oscillation index. In this section, the extended FDE Histogram method is applied to feedback queuing systems (examples 2.2 and 2.3) and a manufacturing system (example 2.4). These examples emphasize the simplicity, correctness and elegance of the extended FDE Histogram method.

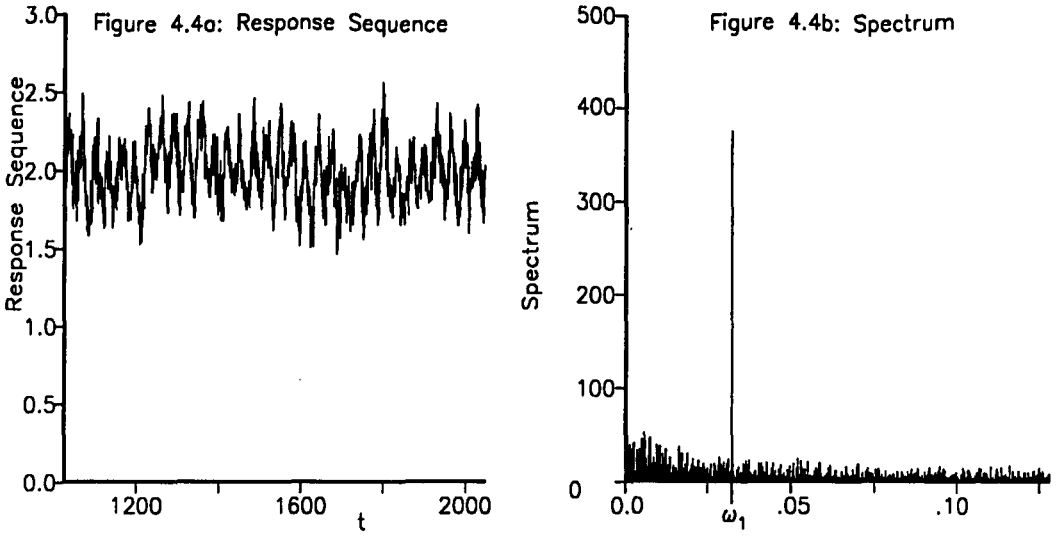


Figure 4.4: Response sequence and spectrum of the expected wait in a feedback queue, $\omega_1 = 0.031250$.

Example 4.4 A FDE for the feedback $M/M/1$ queuing system described in example 2.2 is performed using the extended FDE Histogram method. Customers arrive as a Poisson process with fixed arrival rate $\lambda = 0.5$ and join a FIFO queue before a single server. The service time of a customer entering service at time t is sampled from an exponential distribution with service rate

$$\mu(t) = 1.0 + 0.4\sin(2\pi\omega_1 t) \tag{4.6}$$

where $\omega_1 = 0.031250$. After receiving service a customer rejoins the end of the queue with probability $p = 0.25$ or leaves the system with probability $1 - p = 0.75$.

A simulation stopping time of $T = 5120$ is selected. The system is allowed to warm up for $T_1 = 1024$ time units (selected empirically); warm up is necessary in this case to reduce the low frequency noise in the spectrum caused by the slow

rise in the expected wait from zero to the nominal steady-state value of 2.0. The expected waiting time in the system (including feedback) of customers entering the system (after T_1) is the selected response statistic of interest. The histogram bin size is $\delta = 1$, the number of bins is $n = 4096$ and $S = 100$.

Figure 4.4a depicts a portion of the response sequence and figure 4.4b the corresponding spectrum. The spectrum has a distinct spike at the frequency of input oscillation indicating (as expected) that the expected waiting time in a feedback queue is sensitive to variations of the service rate. Note that when the extended FDE Histogram method is used there is no need to sort the response sequence by the index of arriving customers, as is necessary when the traditional FDE methods are used.

Example 4.5 As a variation of example 2.3, a FDE for the feedback $M/M/1$ queuing system is performed with a sinusoidally varying arrival rate and a sinusoidally varying probability of feedback, using the extended FDE Histogram method. Customers arrive as a non-stationary Poisson process with arrival rate

$$\lambda(t) = 1.0 + 0.1\sin(2\pi\omega_1 t) \tag{4.7}$$

and $\omega_1 = 0.031250$. The thinning method is used to simulate the non-stationary Poisson arrival process. Customers join a FIFO queue before a single server, which has an exponentially distributed random service time and fixed service rate $\mu = 2.0$. After receiving service a customer rejoins the end of the queue with probability

$$p(t) = 0.25 + 0.1\sin(2\pi\omega_2 t) \tag{4.8}$$

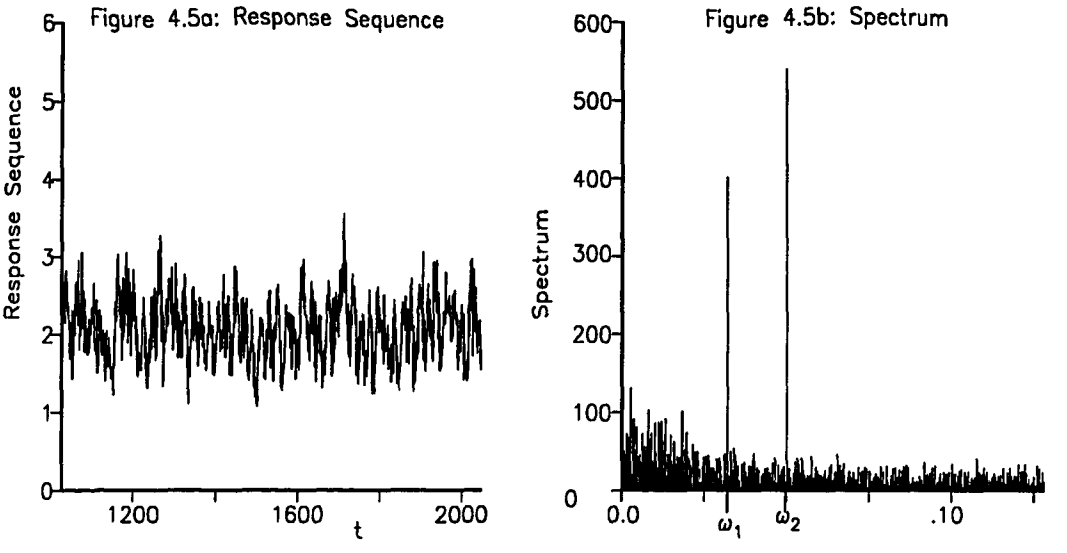


Figure 4.5: Response sequence and spectrum of the expected wait in a feedback queue, $\omega_1 = 0.031250$ and $\omega_2 = 0.50048$.

and $\omega_2 = 0.050048$. As before, a simulation stopping time of $T = 5120$ is selected and the system is allowed to warm up with $T_1 = 1024$ (selected empirically). The expected waiting time in the system (including feedback) of customers entering the system (after T_1) is the selected response statistic of interest. Again $\delta = 1, n = 4096$ and $S = 100$ is used. Figure 4.5a depicts a portion of the response sequence and figure 4.5b the corresponding spectrum. As expected, the spectrum has distinct spikes at both ω_1 and ω_2 indicating that the wait in the system is sensitive to sinusoidal variations of the arrival rate and the feedback probability. Note that, unlike example 2.3 there is no need to sort the response sequence by the index of arriving customers, thereby demonstrating the simplicity and effectiveness of the extended FDE Histogram method.

Example 4.6 A FDE for the simple manufacturing assembly station in example 2.4 is repeated using the extended Histogram method. The arrival rate of a type 1 job is given by equation 3.2, that of a type 2 job is given by equation 3.3 while the service rate of a type 3 job is given by equation 3.4. The thinning method is used to simulate the non-stationary Poisson arrival processes. For the i^{th} type 3 job departing the system at time t_i, the difference between its departure time and the arrival time of the latest component job (either the type 1 job or the second of the two type 2 jobs) is selected as the response statistic of interest. The extended FDE Histogram method is used to generate the corresponding response sequence. A stopping time of $T = 4096$ is selected and no warm-up is used. The number of bins is $n = 4096$ with $\delta = 1$ and $S = 100$. Figure 4.6a is a portion of the response sequence, figure 4.6b represents the corresponding spectrum. The spectrum exhibits distinct spikes at ω_2 and ω_3. The spike at ω_1 is not very distinct because, as discussed in example 2.4, for the given nominal parameter settings, type 2 jobs are the bottleneck and therefore λ_2 has more effect on the response than λ_1; the service rate has a significant effect on the response also since a significant portion of each response observation is the time in service.

To further illustrate that the extended FDE Histogram method can detect bottleneck jobs, the FDE is repeated with

$$\lambda_1(t) = 0.45 + 0.05\sin(2\pi\omega_1 t) \tag{4.9}$$

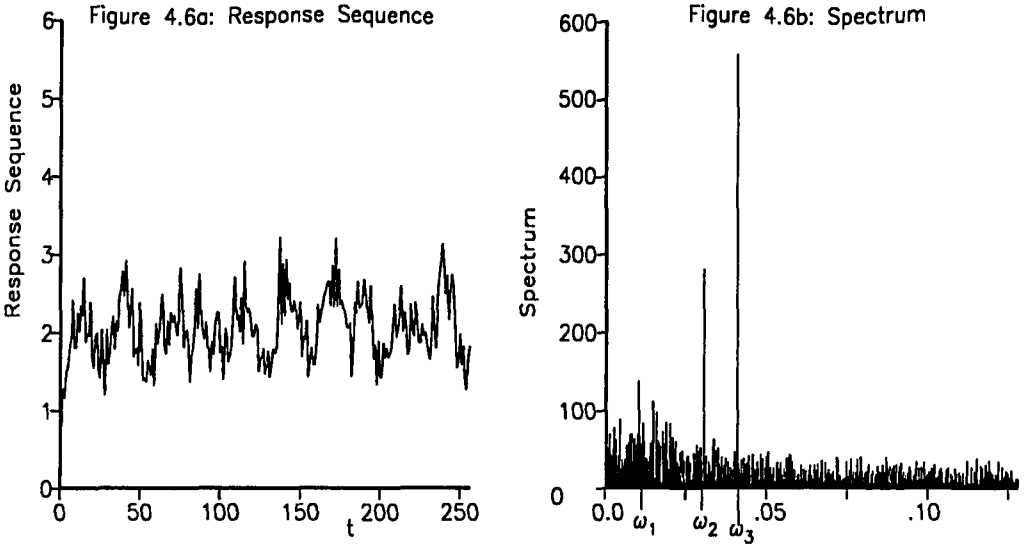


Figure 4.6: Response sequence and spectrum for the manufacturing example with type 2 jobs as the bottleneck.

and $\omega_1 = 0.01$; $\lambda_2(t)$ and $\mu(t)$ are unchanged. A stopping time of $T = 8192$ is selected and a warm-up of $T_1 = 1024$ is used. As before, $n = 4096$, $\delta = 1$ and $S = 100$ is selected. Figure 4.7a is a portion of the response sequence, figure 4.7b is the corresponding spectrum. In this case the type 1 jobs are the bottleneck because, on average, the arrival rates for type 1 jobs are less than half the arrival rates for type 2 jobs. As before the assembly rate has a significant effect on the response. Consistent with expectations, the response spectrum exhibits distinct spikes at ω_1 and ω_3. The results indicate that unlike Morrice et al.'s incorrect methods for using the simulation clock time as the oscillation parameter, the extended FDE Histogram Method is an efficient and effective tool for performing a FDE for the manufacturing example.

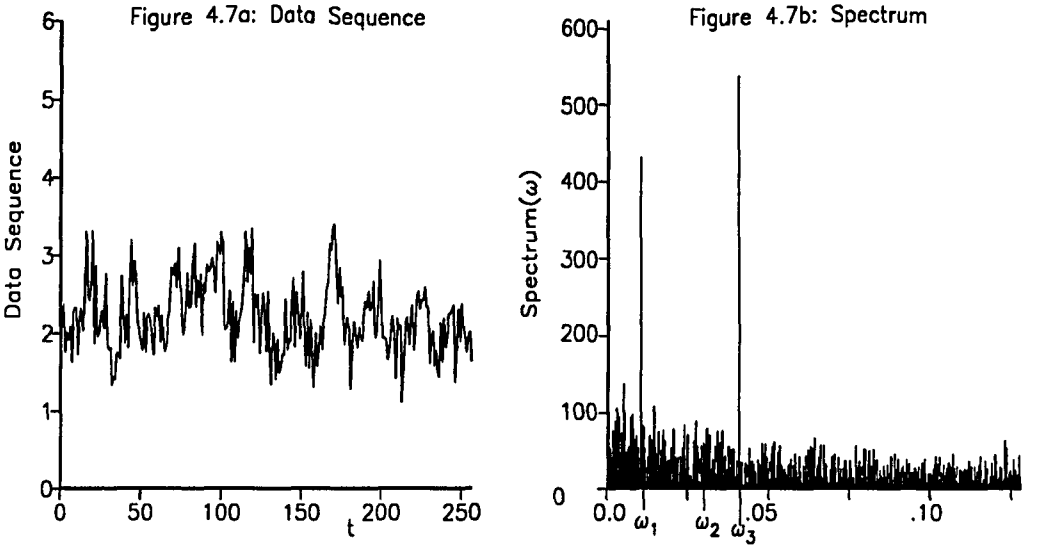


Figure 4.7: Response sequence and spectrum for the manufacturing example with type 1 jobs as the bottleneck ($S = 100$).

To demonstrate how increasing the number of replications decreases the spectral noise, the above experiment is repeated with $S = 1000$ and all other parameters unchanged. Figure 4.8a is a portion of the response sequence and figure 4.8b is the corresponding spectrum. As in the case when $S = 100$, the spectrum has spikes at ω_1 and ω_3. As discussed in chapter 1, the noise in the response sequence and the corresponding spectrum is significantly reduced by increasing the number of replications from $S = 100$ (see figure 4.7) to $S = 1000$ (see figure 4.8).

Examples 4.1–4.6 indicate that the extended FDE Histogram method can be effectively used to perform FDE data analysis using the simulation clock time as the FDE oscillation parameter when the response statistic of interest is not a rate function. Unlike the smeared spectra obtained in section 3.1, the spectra obtained in

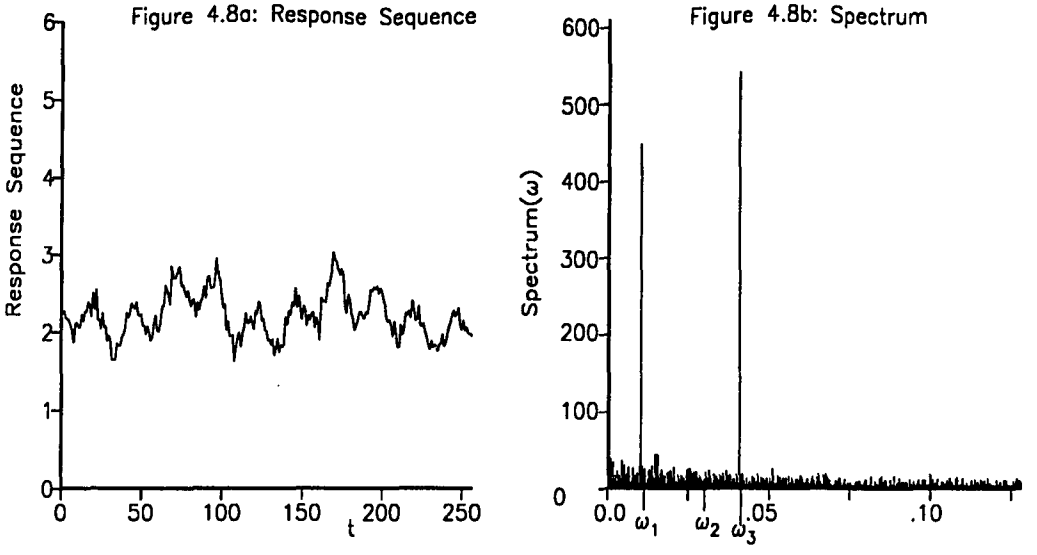


Figure 4.8: Response sequence and spectrum for the manufacturing example with type 1 jobs as the bottleneck ($S = 1000$).

these examples using the extended FDE Histogram method exhibit distinct spikes at the frequencies of input oscillation, clearly indicating the sensitivity of the response statistics to variations in the system parameters. Also, because the extended FDE Histogram method uses the simulation clock time as the oscillation parameter, it obviates the need in traditional FDE's to select a discrete common oscillation index for each system parameter. Therefore, the extended FDE Histogram method is a simple, unifying approach to FDEs.

CHAPTER V

FREQUENCY RESPONSE OF A $M/M/1$ QUEUE

5.1 FDE Model Assumption

As discussed in chapter 2, FDEs are based on a "ω-in/ω-out" or frequency-invariance assumption. That is, if a particular system response statistic is sensitive to a system parameter, then sinusoidal variation of that system parameter at a fixed frequency will induce similar sinusoidal variations in the response statistic, at the same frequency. Most of the systems on which a FDE has been performed have a $M/G/1$ queuing system as the basic unit. The frequency-invariance assumption for $M/G/1$ queues has been verified numerically by several queuing theory researchers [3], [16], [17], [20], [21], [30], [32], [33], [35], [45], [48]. No theoretical support exists for this assumption, however. In this chapter the mathematical theory to support the FDE model assumption for a $M/M/1$ queue with sinusoidally varying arrival rate and fixed service rate is derived.[1]

For a stationary $M/M/1$ queuing system with fixed arrival rate λ_0 and fixed

[1] Using the mathematical developments provided in this chapter, the frequency response for a $M/M/1$ queue with sinusoidally varying arrival and *service rates* or for other $M/G/1$ queues (e.g., with Uniform or Erlang service time distributions) could (probably) be derived. Such derivations have been identified for future research.

service rate μ ($\lambda_0/\mu < 1$), the steady-state departure rate is λ_0. The system can be converted to a non-stationary $M/M/1$ queue by introducing small sinusoidal variations into the arrival rate such that[2]

$$\lambda(t) = \lambda_0 + \sum_{h=1}^{r} \alpha_h \cos(2\pi\omega_h t). \tag{5.1}$$

The departure rate (as given by Cohen [4]) is

$$\xi(t) = \mu\big(1 - P_0(t)\big) \tag{5.2}$$

where $P_0(t)$ is the probability of a free server at time t. Therefore, an expression for $P_0(t)$ is necessary to derive an expression for $\xi(t)$. An expression for $P_0(t)$ for a $M/M/1$ queuing system with arrival rate given by equation 5.1 and fixed service rate is derived in section 5.2. In section 5.3, this expression for $P_0(t)$ is used to establish that $\xi(t)$ is a phase shifted, amplitude modulated version of the arrival rate, thereby validating the frequency-invariance FDE model assumption for a $M/M/1$ queue.

5.2 Solutions for $P_0(t)$

$P_0(t)$ can be obtained by solving the well-known Kolmogrov equations, given by

$$\frac{dP_0(t)}{dt} = -\lambda(t)P_0(t) + \mu(t)P_1(t) \tag{5.3}$$

$$\frac{dP_l(t)}{dt} = -(\lambda(t) + \mu(t))P_l(t) + \lambda(t)P_{l-1}(t) + \mu(t)P_{l+1}(t) \quad l = 1, 2, \ldots \tag{5.4}$$

with the initial conditions:

$$P_l(0) = p_l \qquad l = 0, 1, 2, \ldots \tag{5.5}$$

[2]For $r = 1$, equation 5.1 corresponds to the equation for $\lambda(t)$ used in chapters 2–4.

where $P_l(t)$ is the probability of l jobs in the system at time t. For the purpose of the derivations in this chapter, a constant service rate $\mu(t) \equiv \mu$ is considered.

An *approximate analytical expression* for $P_0(t)$ in terms of λ_0, μ, α_1, $\alpha_2, \ldots, \alpha_r$, ω_1, $\omega_2, \ldots, \omega_r$ and $2r$ unknown parameters is derived in section 5.2.1 using a numerical solution of the Kolmogrov equations. In section 5.2.2 a *differential equation* relating $P_0(t)$, $\lambda(t)$ and μ is derived and in section 5.2.3 the analytical expression and the differential equation are used to solve for the unknown parameters in the analytical expression for $P_0(t)$, thereby giving an analytical solution for $P_0(t)$.

5.2.1 Numerical Solution for $P_0(t)$

Single Frequency of Oscillation

First an arrival rate

$$\lambda(t) = \lambda_0 + \alpha_1 \cos(2\pi\omega_1 t) \tag{5.6}$$

is considered. To solve the Kolmogrov equations numerically, it is necessary to assume that the queue is effectively truncated. If a large truncation level is selected, however, then truncation does not have any significant effect on the numerical results. For the range of values of λ_0, μ, α_1 and ω_1 considered, experimental results indicate that a truncation level of $l \leq 30$ is sufficient. The Kolmogrov differential equations for $P_l(t)$ are solved numerically for $l = 0, 1, \ldots, 30$ using fourth and fifth order Runge Kutta formulas provided by the Matlab [22] software package.

Figure 5.1 is a typical plot of $P_0(t)$ versus t obtained by solving equations 5.3

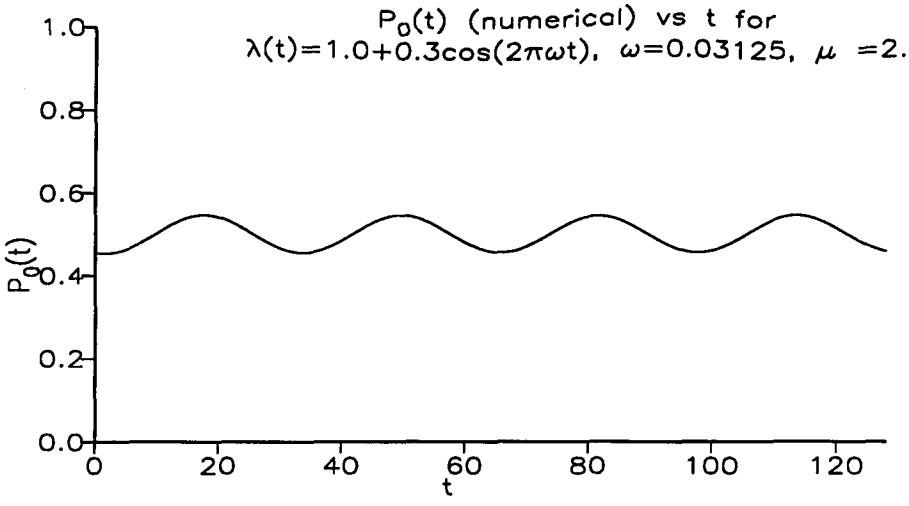


Figure 5.1: $P_0(t)$ versus t; $\lambda_0 = 1.0$, $\mu = 2.0$, $(\alpha_1, \omega_1) = (0.3, 0.03125)$.

and 5.4 using $p_0 = 0.456$, $p_l = p_0(1-p_0)^l$ for $l = 1, 2, 3, \ldots, 30$ and $\lambda(t) = 1.0 + 0.3\cos(2\pi\omega_1 t)$ with $\omega_1 = 0.03125$ and $\mu = 2.0$. Plots of $P_0(t)$ versus t indicate that $P_0(t)$ is periodic with period $1/\omega_1$ and thus has a Fourier series representation of the form

$$P_0(t) = 1 - \rho_0 - \sum_{\nu=1}^{\infty} \rho_\nu \cos(2\pi\nu\omega_1 t - \phi_\nu) \tag{5.7}$$

where the magnitude ρ_ν and the phase ϕ_ν are to be determined for $\nu = 1, 2, \ldots$ The Fourier series is written in a somewhat non-standard form because the departure rate $(1 - P_0(t))\mu$ is of primary interest and

$$1 - P_0(t) = \rho_0 + \sum_{\nu=1}^{\infty} \rho_\nu \cos(2\pi\nu\omega_1 t - \phi_\nu) \tag{5.8}$$

where $\rho_0 = \lambda_0/\mu$ is expected to be the $\alpha_1 = 0$ utilization of the system.

Standard least-squares techniques are used to fit equation 5.7 to numerical values of $P_0(t)$ and thereby obtain values for ρ_0, ρ_ν and ϕ_ν, for a small range of ν values. Tables 5.1–5.4 show the resulting values of ρ_0, ρ_ν for different parameter settings. The results indicate that, as expected, $\rho_0 = \lambda_0/\mu$ and $\rho_1 = \alpha_1 k_1/\mu$ where k_1 is a

Table 5.1: $\rho_0, \rho_1, \rho_2, \ldots$ in the Fourier series representation of $P_0(t)$ for different ω_1; $\lambda_0 = 0.5$, $\alpha_1 = 0.1$ and $\mu = 4.00$.

ω_1	Fourier Coefficients						
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
0.007812	0.875000	0.024996	0.000011	0.000000	0.000000	0.000000	0.000000
0.015625	0.875000	0.024984	0.000023	0.000000	0.000000	0.000000	0.000000
0.031250	0.875000	0.024937	0.000045	0.000001	0.000000	0.000000	0.000000
0.062500	0.874999	0.024750	0.000082	0.000000	0.000001	0.000001	0.000001
0.125000	0.874998	0.024061	0.000128	0.000002	0.000002	0.000002	0.000002

function of λ_0, μ and ω_1. Moreover, for $\nu = 2, 3, \ldots$, the coefficients ρ_ν are effectively zero relative to ρ_1. That is, if $\lambda(t)$ is given by equation 5.6 with μ constant, then

$$P_0(t) \approx 1 - \frac{\lambda_0}{\mu} - k_1 \frac{\alpha_1}{\mu} \cos(2\pi\omega_1 t - \phi_1) \tag{5.9}$$

provided $\lambda_0 + \alpha_1 < \mu$.

Multiple Frequencies of Oscillation

Next equation 5.1 is considered for $r > 1$. Figures 5.2a–5.2h are $r = 2$ plots of $P_0(t)$ verses t and the magnitude of the corresponding (complex-valued) DFT for different values of μ obtained by solving the Kolmogrov equations. Figures 5.3 and 5.4 show $P_0(t)$ versus t and the corresponding spectrum for $r = 3$ and $r = 7$. All three figures indicate that when $\lambda(t)$ is given by equation 5.1, then $P_0(t)$ has a Fourier series representation with r distinct spikes in its spectrum at $\omega_1, \omega_2, \ldots, \omega_r$.

Table 5.2: $\rho_0, \rho_1, \rho_2, \ldots$ in the Fourier series representation of $P_0(t)$ for different α_1;

$\lambda_0 = 1.0$, $\mu = 4.00$ and $\omega_1 = 0.015625$.

α_1	Fourier Coefficients						
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
0.100000	0.750000	0.024964	0.000036	0.000001	0.000000	0.000000	0.000000
0.200000	0.750000	0.049928	0.000144	0.000007	0.000000	0.000000	0.000000
0.300000	0.750000	0.074889	0.000325	0.000024	0.000001	0.000000	0.000000
0.400000	0.750000	0.099847	0.000581	0.000056	0.000005	0.000000	0.000000
0.500000	0.750000	0.124801	0.000916	0.000111	0.000012	0.000001	0.000000
0.600000	0.750000	0.149750	0.001334	0.000193	0.000025	0.000003	0.000000
0.700000	0.750000	0.174691	0.001838	0.000310	0.000046	0.000007	0.000001

Table 5.3: $\rho_0, \rho_1, \rho_2, \ldots$ in the Fourier series representation of $P_0(t)$ for different μ;

$\lambda_0 = 1.0$, $\alpha_1 = 0.1$ and $\omega_1 = 0.015625$.

μ	Fourier Coefficients						
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
4.000000	0.750000	0.024964	0.000036	0.000001	0.000000	0.000000	0.000000
2.000000	0.500000	0.048260	0.000680	0.000028	0.000001	0.000000	0.000000
1.333333	0.250948	0.048673	0.002239	0.000181	0.000155	0.000128	0.000111

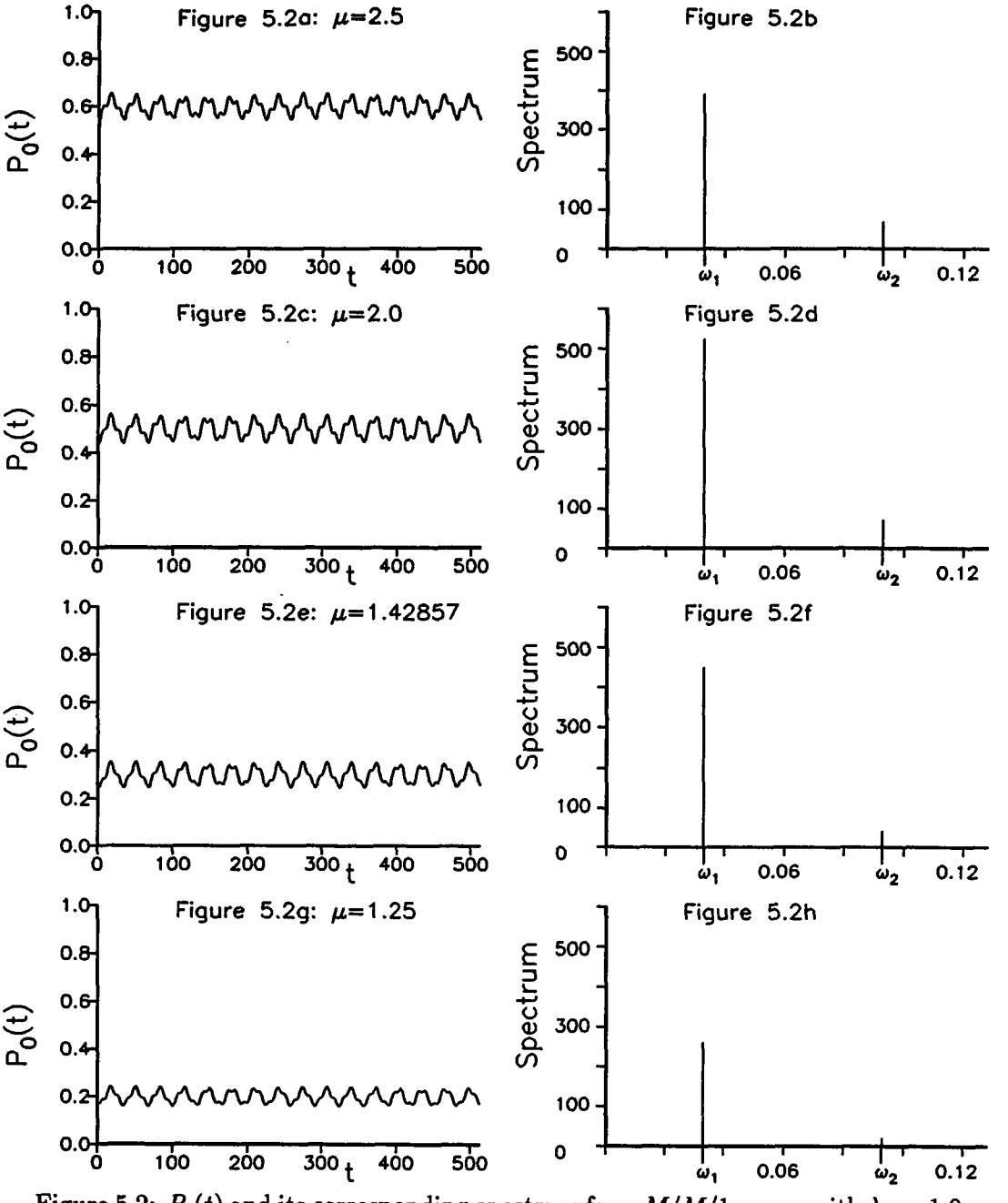


Figure 5.2: $P_0(t)$ and its corresponding spectrum for a $M/M/1$ queue with $\lambda_0 = 1.0$, $(\alpha_1, \omega_1) = (0.1, 0.03125)$, $(\alpha_2, \omega_2) = (0.05, 0.08984)$ and different values of μ.

Table 5.4: $\rho_0, \rho_1, \rho_2, \ldots$ in the Fourier series representation of $P_0(t)$ for different λ_0; $\alpha_1 = 0.1$, $\mu = 4.0$ and $\omega_1 = 0.015625$.

λ_0	Fourier Coefficients						
	ρ_0	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6
1.000000	0.750000	0.024964	0.000036	0.000001	0.000000	0.000000	0.000000
1.500000	0.625000	0.024914	0.000060	0.000002	0.000000	0.000000	0.000000
2.000000	0.500000	0.024765	0.000110	0.000003	0.000000	0.000000	0.000000
2.500000	0.374990	0.024229	0.000205	0.000007	0.000001	0.000001	0.000001

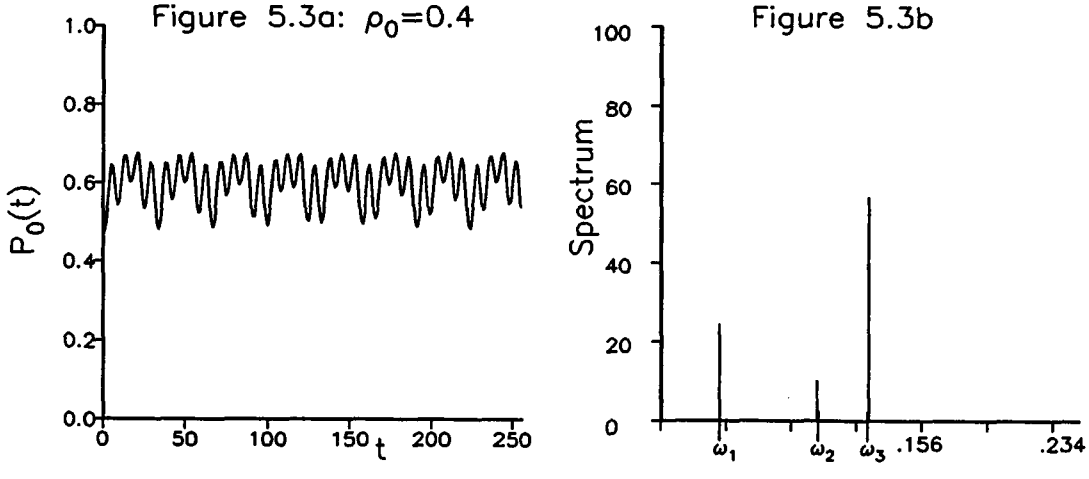


Figure 5.3: $P_0(t)$ and its spectrum for a $M/M/1$ queue with $(\alpha_1, \omega_1) = (0.1, 0.031250)$, $(\alpha_2, \omega_2) = (0.05, 0.08984)$, $(\alpha_3, \omega_3) = (0.2, 0.12109375)$, $\lambda_0 = 1.0$ and $\mu = 2.5$.

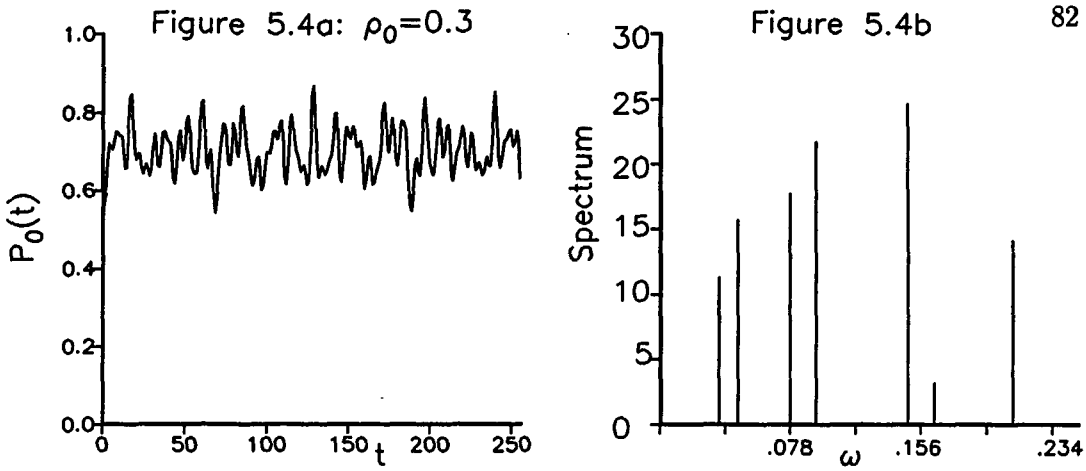


Figure 5.4: $P_0(t)$ and its spectrum for a $M/M/1$ queue; $(\alpha_1, \omega_1) = (0.05, 0.031250)$, $(\alpha_2, \omega_2) = (0.06, 0.04296)$, $(\alpha_3, \omega_3) = (0.07, 0.07421)$, $(\alpha_4, \omega_4) = (0.08, 0.08984)$, $(\alpha_5, \omega_5) = (0.04, 0.16015)$, $(\alpha_6, \omega_6) = (0.09, 0.20703)$, $(\alpha_7, \omega_7) = (0.1, 0.14453)$, $\lambda_0 = 1.0$ and $\mu = 2.5$.

Standard least-square techniques are used to estimate the Fourier coefficients in the Fourier series representation of $P_0(t)$. The least-square fits indicate that, provided the maximum amplitude of oscillation of the arrival rate is less then μ, $P_0(t)$ can be approximated by

$$P_0(t) \approx 1 - \frac{\lambda_0}{\mu} - \sum_{h=1}^{r} \frac{\alpha_h}{\mu} k_h \cos(2\pi\omega_h t - \phi_h). \tag{5.10}$$

Equation 5.10 is a key, albeit empirical, result that has been verified with many numerical simulations.

5.2.2 Differential Equation For $P_0(t)$

In this section, a differential expression relating $P_0(t)$, $\lambda(t)$ and μ is derived. The differential equation and equation 5.10 are used to derive an analytical expression

for $P_0(t)$ in section 5.2.3. For a time-dependent $M/M/1$ queue, let $Q(t)$ represent the expected number in the system. Therefore by definition

$$Q(t) = \sum_{l=1}^{\infty} lP_l(t). \tag{5.11}$$

Differentiating both sides of equation 5.11 with respect to t yields

$$\begin{aligned} Q'(t) &= \sum_{l=1}^{\infty} lP_l'(t) \\ &= \sum_{l=1}^{\infty} l\Big(\lambda(t)P_{l-1}(t) - (\lambda(t)+\mu(t))P_l(t) + \mu(t)P_{l+1}(t)\Big) \end{aligned} \tag{5.12}$$

where $\lambda(t)$ and $\mu(t)$ represent the time-varying arrival and service rates, respectively.[3] Simplifying equation 5.12 yields

$$Q'(t) = \lambda(t) - \mu(t)\Big(1 - P_0(t)\Big) \tag{5.13}$$

In [32], Rider derives an approximation expression for $P_0(t)$ as

$$P_0(t) = (1 - e^{-\mu(t)T})(1 - \rho(t)) + \frac{e^{-\mu(t)T}}{1 + Q(t)} \tag{5.14}$$

where T is an unknown small positive parameter. Rider shows that simulation results for $Q(t)$ obtained using equations 5.13 and 5.14 with T set to zero are close to the numerical results obtained by solving the Kolmogrov equations and then using equation 5.11 to computer $Q(t)$. When $T = 0$, equation 5.14 reduces to

$$P_0(t) = \frac{1}{1 + Q(t)}. \tag{5.15}$$

[3] A time-varying service rate $\mu(t)$ is used to derive the differential equation for $P_0(t)$. Substituting $\mu \equiv \mu(t)$ in the resulting differential equation will yield the required differential equation when the service rate is constant.

Tipper et al.[47] validated the accuracy of equation 5.15 via simulation and numerical methods.

As an independent validation, in this dissertation the accuracy of equation 5.15 is tested numerically for a range of parameter values of interest. Equation 5.6 is used to represent $\lambda(t)$. Figures 5.5a–5.5f show $P_0(t)$ obtained by numerically integrating equations 5.3 and 5.4 to give $P_l(t)$ for $l = 0, 1, 2, \ldots, 30$ and $P_0(t)$ obtained by (i) numerically integrating equations 5.3 and 5.4, (ii) using the numerical solutions of $P_l(t)$ for $l = 0, 1, 2, \ldots, 30$ to give $Q(t)$ using equation 5.11 and finally (iii) substituting numerical value of $Q(t)$ into equation 5.15. The two $P_0(t)$ versus t figures are indistinguishable, except when $\lambda_0 + \alpha_1$ approaches μ. That is, equation 5.15 is a good approximation for $P_0(t)$ provided $\lambda_0 + \alpha_1 < \mu$.

Provided $\lambda_0 + \alpha_1 < \mu$, $Q(t)$ can be eliminated from equations 5.13 and 5.15 to obtain an equation relating $P_0(t)$, $\lambda(t)$ and $\mu(t)$. That is, equation 5.15 gives

$$1 + Q(t) = \frac{1}{P_0(t)}$$

or,

$$Q'(t) = \frac{-P_0'(t)}{P_0^2(t)}. \tag{5.16}$$

Combining equations 5.13 and 5.16 yields

$$\frac{-P_0'(t)}{P_0^2(t)} = \lambda(t) - \mu(t)\big(1 - P_0(t)\big)$$

or,

$$-P_0'(t) = (\lambda(t) - \mu(t))P_0^2(t) + \mu(t)P_0^3(t). \tag{5.17}$$

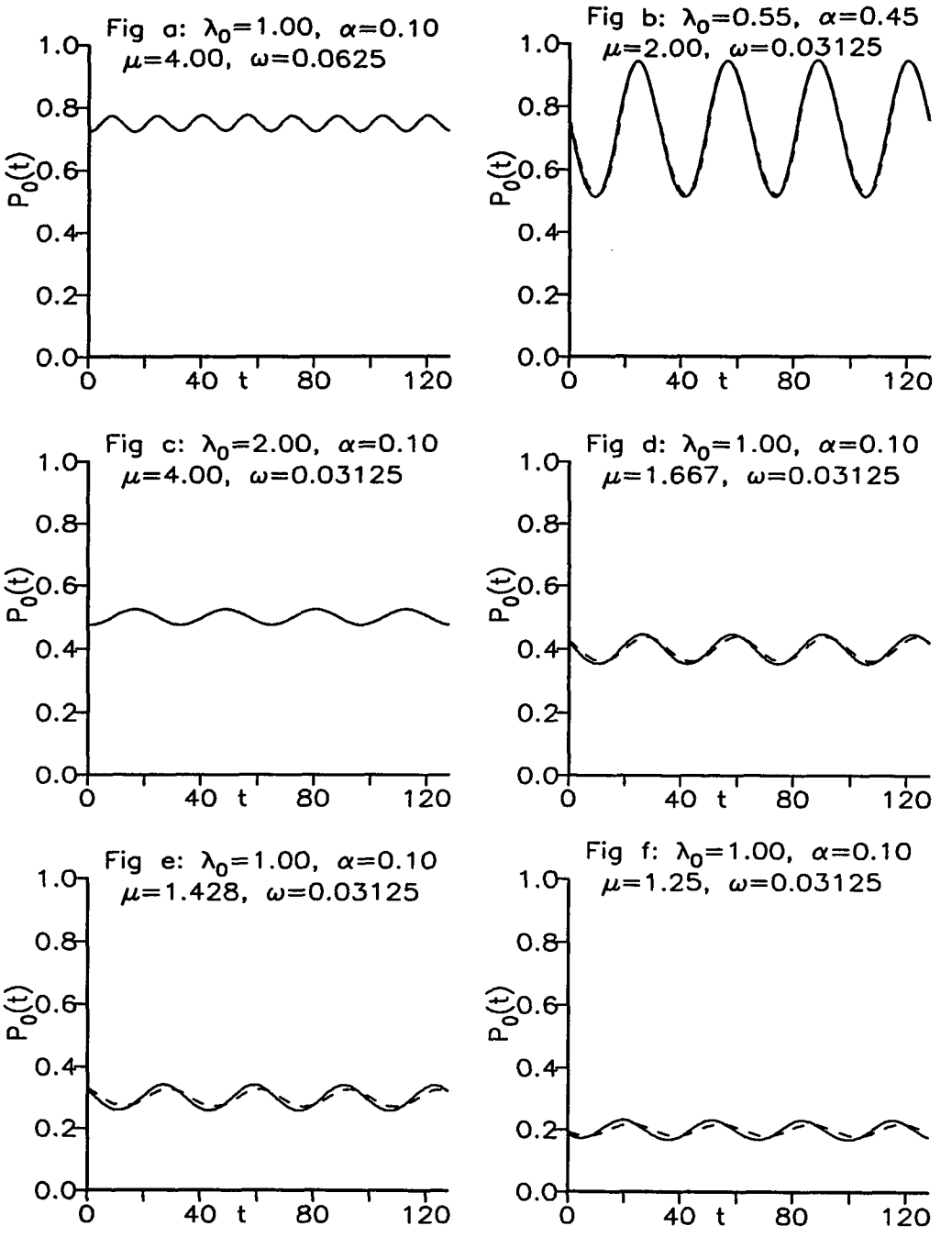


Figure 5.5: $P_0(t)$ obtained by solving the Kolmogrov equations (continuous plot) and by using equation 5.15 (dashed plot).

Therefore eliminating $Q(t)$ from equations 5.13 and 5.15 results in the following Abel's equation [28] of the first kind relating the arrival rate $\lambda(t)$, the service rate $\mu(t)$ and $P_0(t)$

$$P_0'(t) + \mu(t)P_0^3(t) + (\lambda(t) - \mu(t))P_0^2(t) = 0. \tag{5.18}$$

When $\mu(t) \equiv \mu$, the Abel's equation for $P_0(t)$ is

$$P_0'(t) + \mu P_0^3(t) + (\lambda(t) - \mu)P_0^2(t) = 0. \tag{5.19}$$

5.2.3 Analytical Solution for $P_0(t)$

Single Frequency of Oscillation

When the arrival rate is given by equation 5.6, the corresponding expression for $P_0(t)$ is given by equation 5.9 in terms of two unknowns k_1 and ϕ_1. Equations for k_1 and ϕ_1 are now derived.

Because equations 5.3 and 5.4 are linear, if $\lambda(t)$ is allowed to be complex-valued then the probabilities $P_l(t)$ are complex-valued as well. The real parts of the probabilities correspond to the solution associated with the real part of $\lambda(t)$. Therefore, a solution for $P_0(t)$ can be obtained by using a complex variable form of $\lambda(t)$ and equation 5.9 and then considering the real part of the resulting expression for $P_0(t)$.[4]

[4]Although the FDEs are performed with sinusoidal inputs, the mathematical derivations presented here are based on cosinusoidal inputs for the sake of simplicity of dealing with the complex variables. Equivalent expressions can be derived if sinusoidal inputs are considered—at the expense of some mathematical simplicity.

That is, it is assumed that

$$\lambda(t) = \lambda_0 + \alpha_1 e^{i2\pi\omega_1 t} \tag{5.20}$$

$$P_0(t) = (1 - \frac{\lambda_0}{\mu}) - \frac{z_1 \alpha_1}{\mu} e^{i2\pi\omega_1 t} \tag{5.21}$$

where z_1 is a complex number. Substituting (5.20) and (5.21) in (5.19) yields

$$-\left(\frac{\alpha_1^2 e^{i6\pi\omega_1 t}}{\mu^2}\right) z_1^3 + \left(\frac{\alpha_1^2 e^{i6\pi\omega_1 t}}{\mu^2} - \frac{2\alpha_1 e^{i4\pi\omega_1 t}\lambda_0}{\mu^2} + \frac{2\alpha_1 e^{i4\pi\omega_1 t}}{\mu}\right) z_1^2$$
$$+\left(-e^{i2\pi\omega_1 t} + \frac{2\alpha_1 e^{i4\pi\omega_1 t}\lambda_0}{\mu^2} - \frac{e^{i2\pi\omega_1 t}\lambda_0^2}{\mu^2} - \frac{2\alpha_1 e^{i4\pi\omega_1 t}}{\mu} + \frac{2e^{i2\pi\omega_1 t}\lambda_0}{\mu} - \frac{2ie^{i2\pi\omega_1 t}\pi\omega_1}{\mu}\right) z_1$$
$$+e^{i2\pi\omega_1 t}\left(1 - \frac{2\lambda_0}{\mu} + \frac{\lambda_0^2}{\mu^2}\right) = 0. \tag{5.22}$$

Provided α_1 is sufficiently small, terms of order α_1 or smaller can be ignored, in which case equation 5.22 simplifies to

$$e^{i2\pi\omega_1 t}\left(-1 - \frac{\lambda_0^2}{\mu^2} + \frac{2\lambda_0}{\mu} - \frac{2i\pi\omega_1}{\mu}\right) z_1 + e^{i2\pi\omega_1 t}\left(1 - 2\frac{\lambda_0}{\mu} + \frac{\lambda_0^2}{\mu^2}\right) = 0 \tag{5.23}$$

Because $e^{i2\pi\omega_1 t} \neq 0$ for all t, solving for z_1 yields

$$z_1 = \frac{(\lambda_0 - \mu)^2}{(\lambda_0 - \mu)^2 + i2\pi\omega_1\mu} = \frac{1}{1 + ib_1} \tag{5.24}$$

where

$$b_1 = \frac{2\pi\omega_1\mu}{(\lambda_0 - \mu)^2}. \tag{5.25}$$

Therefore, the real part of equation 5.25 gives

$$P_0(t) = \left(1 - \frac{\lambda_0}{\mu}\right) - \frac{\alpha_1}{\mu} k_1 \cos(2\pi\omega_1 t - \phi_1) \tag{5.26}$$

where

$$k_1 = \frac{1}{\sqrt{1 + b_1^2}} \tag{5.27}$$

and

$$\phi_1 = \tan^{-1}(b_1). \tag{5.28}$$

Validating the Analytical Expression of $P_0(t)$

The approximate analytical expressions for $P_0(t)$ (equation 5.26) is obtained by solving the differential equation for $P_0(t)$ (equation 5.19) using an empirically determined form for $P_0(t)$ (equation 5.9). To determine the validity of the empirical form, the analytical solution for $P_0(t)$ (equation 5.26) is compared to the numerical solution of $P_0(t)$ obtained by solving the Kolmogrov equations (equations 5.3 and 5.4).

Figures 5.6a–5.6e show $P_0(t)$ versus t obtained by numerically integrating equations 5.3 and 5.4 as discussed in section 5.2.2 and $P_0(t)$ obtained analytically using equation 5.26. Figure 5.6 indicates that the solutions given by equation 5.26 match the numerical solution for $P_0(t)$ provided $\lambda_0 + \alpha_1 < \mu$. As $\lambda_0 + \alpha_1 \rightarrow \mu$, the match becomes less accurate.

Multiple Frequencies of Oscillation

Equation 5.10 gives an analytical expression for $P_0(t)$ for a $M/M/1$ queue with arrival rate given by equation 5.1 and fixed service rate μ. The following theorem specifies the values of k_h and ϕ_h for $h = 1, 2, \ldots, r$ in equation 5.10.

Theorem 5.2.1 *If the arrival rate $\lambda(t)$ to a $M/M/1$ queue with service rate μ is*

$$\lambda(t) = \lambda_0 + \sum_{h=1}^{r} \alpha_h \cos(2\pi\omega_h t) \tag{5.29}$$

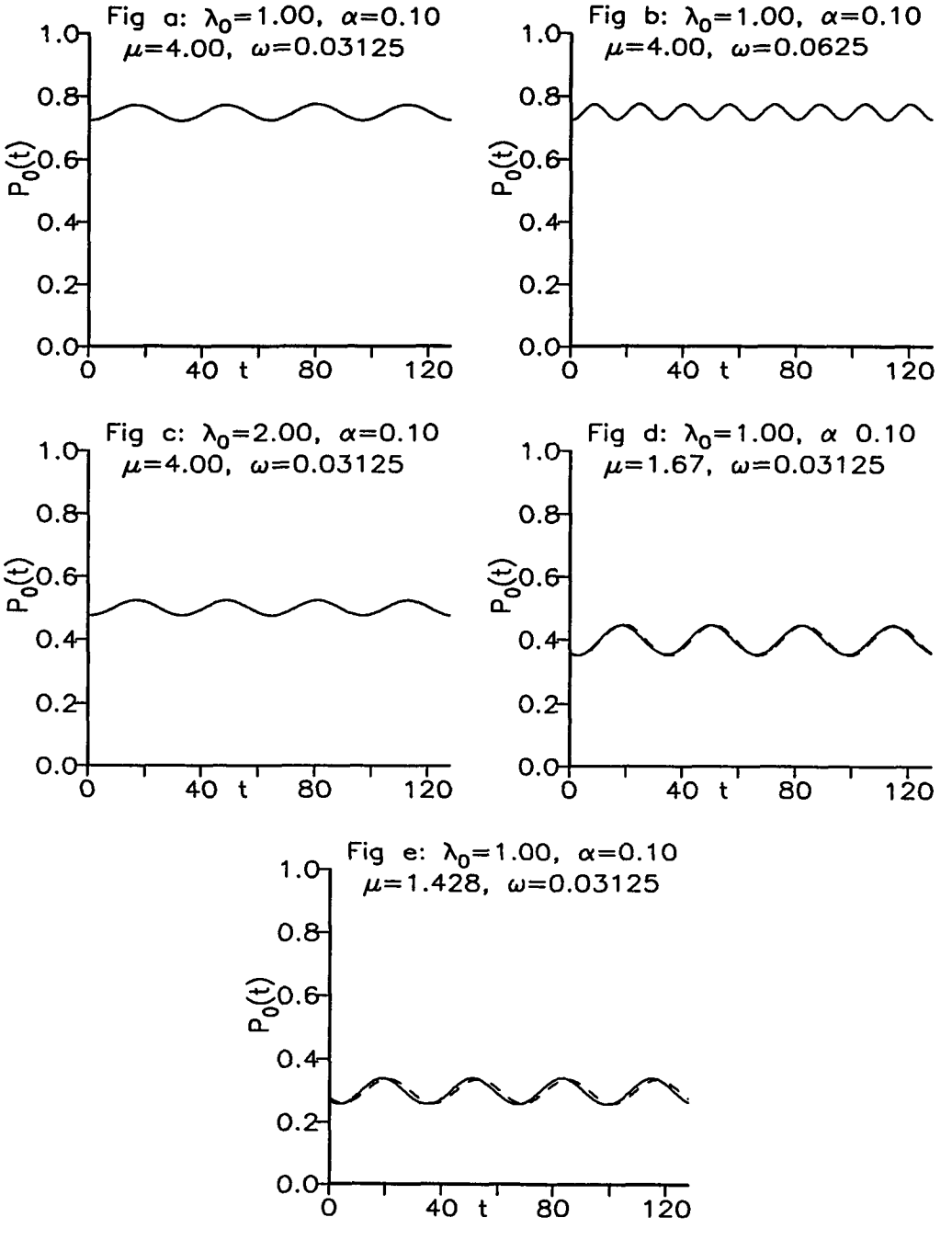


Figure 5.6: $P_0(t)$ obtained by solving the Kolmogrov equations (continuous plot) and by using equation 5.26 (dashed plot).

and the corresponding probability of a free server is

$$P_0(t) = 1 - \frac{\lambda_0}{\mu} - \sum_{h=1}^{r} \frac{\alpha_h}{\mu} k_h \cos(2\pi\omega_h t - \phi_h) \tag{5.30}$$

then for all $h = 1, 2, \ldots, r$,

$$k_h \simeq 1/\sqrt{(1 + b_h^2)} \tag{5.31}$$

$$\phi_h \simeq \tan^{-1}(b_h) \tag{5.32}$$

where

$$b_h = \frac{2\pi\omega_h\mu}{(\lambda_0 - \mu)^2}$$

provided $\lambda_0 + \alpha_1 + \alpha_2 + \ldots + \alpha_r < \mu$.

Proof: Expressions for k_h and ϕ_h, for $h = 1, 2, \ldots, r$ can be obtained by combining equations (5.19) and (5.30). As in section 5.2.3, these expressions are obtained by using a complex variable form of $\lambda(t)$ and equation (5.30). That is, the equations

$$\lambda(t) = \lambda_0 + \sum_{h=1}^{r} \alpha_h e^{i2\pi\omega_h t} \tag{5.33}$$

$$P_0(t) = 1 - \frac{\lambda_0}{\mu} - \sum_{h=1}^{r} \frac{\alpha_h}{\mu} z_h e^{i2\pi\omega_h t} \tag{5.34}$$

are used, where $z_1, z_2, \ldots, z_r$ are unknown, complex-valued quantities. Equations (5.34) indicates that

$$P_0'(t) = -i2\pi \sum_{h=1}^{r} \alpha_h z_h \frac{\omega_h}{\mu} e^{i2\pi\omega_h t} \tag{5.35}$$

and

$$P_0^2(t)(\mu P_0(t) + \lambda(t) - \mu) = \left(1 - \frac{\lambda_0}{\mu} - \sum_{h=1}^{r} \frac{\alpha_h}{\mu} z_h e^{i2\pi\omega_h t}\right)^2 \sum_{h=1}^{r} e^{i2\pi\omega_h t} \alpha_h (1 - z_h). \tag{5.36}$$

Assuming that $\alpha_1, \alpha_2, \ldots, \alpha_r$ are sufficiently small so that all terms of quadratic or higher order are approximately zero, equation (5.36) reduces to:

$$P_0^2(t)(\mu P_0(t) + \lambda(t) - \mu) \simeq \left(1 - \frac{\lambda_0}{\mu}\right)^2 \sum_{h=1}^{r} e^{i2\pi\omega_h t}\alpha_h(1 - z_h). \tag{5.37}$$

Equations (5.19), (5.35) and (5.37) give

$$-i2\pi \sum_{h=1}^{r} \alpha_h z_h \frac{\omega_h}{\mu} e^{i2\pi\omega_h t} + \left(1 - \frac{\lambda_0}{\mu}\right)^2 \sum_{h=1}^{r} e^{i2\pi\omega_h t}\alpha_h(1 - z_h) \simeq 0$$

or,

$$\sum_{h=1}^{r} \alpha_h e^{i2\pi\omega_h t}\Big((1 - \frac{\lambda_0}{\mu})^2 - z_h((1 - \frac{\lambda_0}{\mu})^2 + i2\pi\frac{\omega_h}{\mu})\Big] \simeq 0. \tag{5.38}$$

Provided $\omega_1, \omega_1, \ldots, \omega_r$ are distinct, $e^{i2\pi\omega_h t}$ for $h = 1, 2, \cdots, r$ are linearly independent functions and the only way equation (5.38) can be valid for all t is when

$$(1 - \frac{\lambda_0}{\mu})^2 - z_h\Big((1 - \frac{\lambda_0}{\mu})^2\mu + i2\pi\frac{\omega_h}{\mu}\Big) \simeq 0 \qquad h = 1, 2, \ldots, r. \tag{5.39}$$

Equivalently,

$$(1 - \frac{\lambda_0}{\mu})^2 \simeq z_h\Big((1 - \frac{\lambda_0}{\mu})^2\mu + i2\pi\frac{\omega_h}{\mu}\Big) \qquad h = 1, 2, \ldots, r$$

or,

$$z_h \simeq \frac{1}{1 + ib_h} \qquad h = 1, 2, \ldots, r \tag{5.40}$$

where

$$b_h = \frac{2\pi\omega_h}{\mu(1 - \frac{\lambda_0}{\mu})^2} \qquad h = 1, 2, \ldots, r. \tag{5.41}$$

Substituting in equation (5.34) and using the real part of the resulting expression for $P_0(t)$,

$$P_0(t) = 1 - \frac{\lambda_0}{\mu} - \sum_{h=1}^{r} \frac{\alpha_h}{\mu} k_h \cos(2\pi\omega_h t - \phi_h) \tag{5.42}$$

where

$$k_h \simeq 1/\sqrt{1+ib_h} \qquad h = 1,2,\ldots,r$$

and

$$\phi_h \simeq \tan^{-1}(b_h) \qquad h = 1,2,\ldots,r$$

which proves theorem 5.2.1.

5.2.4 High Utilization

Figure 5.6 indicates that if $\lambda_0 + \alpha_1$ is close to μ, then the exact numerical solution differs from the (approximate) analytical solution, thereby requiring a correction to be introduced to equation 5.26 to match the numerical results. To determine the correction term, $P_0(t)$ is derived from equations 5.13 and 5.14 instead of equations 5.13 and 5.15. For $\mu(t) \equiv \mu$, equation 5.14 can be rewritten as

$$P_0(t) = g(t) + \frac{e^{-\mu T}}{1+Q(t)} \tag{5.43}$$

where

$$g(t) = (1-e^{-\mu T})(1-\rho(t)). \tag{5.44}$$

From equation 5.43 it follows that

$$Q'(t) = e^{-\mu T}\frac{-P_0'(t)+g'(t)}{(P_0(t)-g(t))^2}. \tag{5.45}$$

Elimination of $Q'(t)$ from equations 5.13 and 5.45 results in

$$P_0'(t) + AP_0^3(t) + BP_0^2(t) + CP_0(t) + B = 0 \tag{5.46}$$

where,

$$\begin{aligned} A &= e^{\mu T}\mu, \\ B &= e^{\mu T}(-2\mu g(t) + \lambda(t) - \mu), \\ C &= e^{\mu T}(\mu g^2(t) - 2g(t)(\lambda(t) - \mu)), \\ D &= e^{\mu T}(g^2(t)(\lambda(t) - \mu)) - g'(t). \end{aligned}$$

Note that for $T = 0$, $g(t) = 0$, $B = \lambda(t) - \mu$, $C = D = 0$ and equation 5.46 reduces to equation 5.19.

Using equation 5.21 as an approximation for $P_0(t)$ and substituting equations 5.20 and 5.21 in equation 5.46 gives

$$z_1 = \frac{1 + ib_1(e^{\mu T} - 1)}{1 + ib_1 e^{\mu T}} \tag{5.47}$$

where b_1 is given by equation 5.25. Equation 5.47 yields

$$\begin{aligned} z_1 &= 1 - \frac{b_1^2 e^{\mu T}}{1 + b_1^2 e^{2\mu T}} - \frac{ib_1}{1 + b_1^2 e^{2\mu T}} \\ &= d - ie \end{aligned} \tag{5.48}$$

where

$$d = 1 - \frac{b_1^2 e^{\mu T}}{1 + b_1^2 e^{2\mu T}} \tag{5.49}$$

and

$$e = \frac{b_1}{1 + b_1^2 e^{2\mu T}}. \tag{5.50}$$

The real part of equation 5.48 gives

$$P_0(t) = \left(1 - \frac{\lambda_0}{\mu}\right) - \frac{\alpha_1}{\mu}\sqrt{d^2 + e^2}\cos(2\pi\omega_1 t - \tan^{-1}(e/d)) \tag{5.51}$$

where d and e are given by equations 5.49 and 5.50 respectively.

In equation 5.51, T is a correction factor which is not known. If analytical solutions approximately matched to the numerical solutions are desired, however, T can be adjusted to get a good approximation. Consider a $M/M/1$ queue with high average utilization. Specifically $\lambda_0 = 1.0$, $\alpha_1 = 0.1$, $\mu = 1.25$ and $\omega_1 = 0.03125$. These parameter settings are used with equations 5.3 and 5.4 to get a numerical solution for $P_0(t)$ which is then compared with equation 5.51 for various values of T. Figure 5.7 illustrates the results for four different values of T. For $T = 0.0$, equation 5.51 yields a solution smaller than the numerical results (see figure 5.7a); for $T = 0.8/\mu$ the equation 5.51 solution overshoots the numerical results (see figure 5.7b); $T = 0.5/\mu$ gives a closer match (see figure 5.7c) and $T = 0.47/\mu$ gives an almost exact match (see figure 5.7d). Thus by manipulating T a good approximation for the analytical solution of $P_0(t)$ can be obtained from equation 5.51 when $\lambda_0 + \alpha$ is close to μ, is desired.

5.3 Departure Rate from a $M/M/1$ Queue

The departure rate for a $M/M/1$ queue with arrival rate given by equation 5.1 and fixed service rate μ is obtained by substituting equation 5.30 in equation 5.2. That is

$$\xi(t) = \lambda_0 + \sum_{h=1}^{r} \alpha_h k_h \cos(2\pi\omega_h t - \phi_h) \tag{5.52}$$

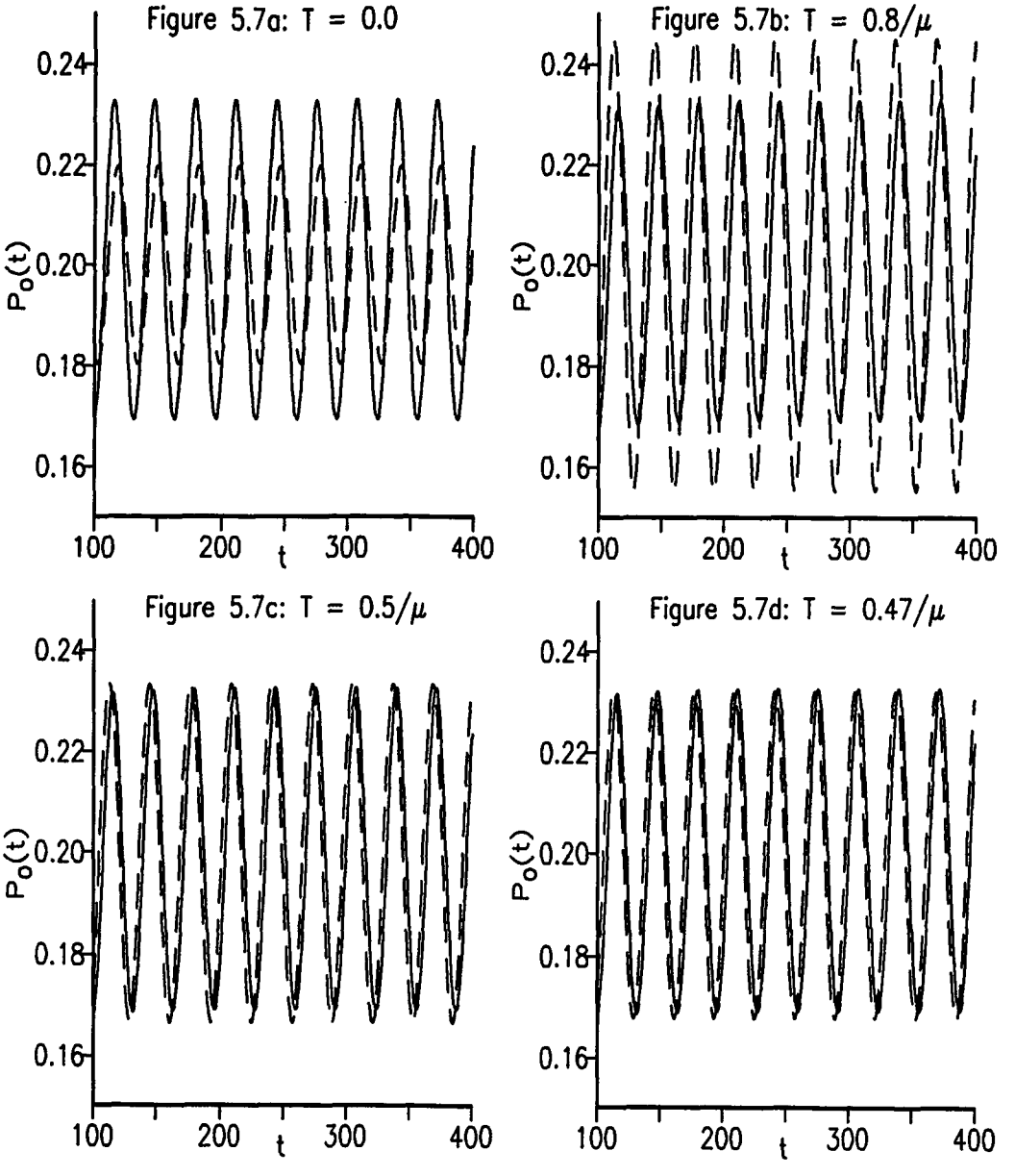


Figure 5.7: $P_0(t)$ versus t (dashed plot: analytical and continuous plot: numerical) for different values of T, with $\lambda_0 = 1.0$, $\mu = 1.25$, $\alpha_1 = 0.1$, $\omega = 0.03125$.

where k_h and ϕ_h are given by equations 5.31 and 5.32 respectively. Provided $\lambda_0 + \alpha_1 + \alpha_2 + \ldots + \alpha_r < \mu$, equation 5.52 indicates that the departure rate from a $M/M/1$ queue with sinusoidally varying arrival rate and fixed service rate is an amplitude modulated, phase shifted version of the arrival rate. Thus, equation 5.52 validates the frequency-invariance FDE model assumption. This mathematical derivation is applied in chapter 6 to determine the frequency response of a $M/M/1$ queuing system and to determine the frequency response of a network of such queues.

CHAPTER VI

FREQUENCY RESPONSE OF $M/M/1$ QUEUING NETWORKS

In the first section of this chapter, the mathematical derivation of chapter 5 is used to determine the frequency response of a $M/M/1$ queuing system. Then, in section 6.2 a FDE for two networks of $M/M/1$ queues is used to demonstrate that the frequency-invariance FDE model assumption is valid for a network of $M/M/1$ queues, provided the maximum amplitude of oscillation of the arrival rate at each node is less than its service rate. The frequency response characterization of a network of $M/M/1$ queues opens up avenues of FDE-based modeling and Fourier analysis of systems common in a wide variety of applications such as computer and communication networks.

6.1 $M/M/1$ Queue Frequency Response

In section 5.3 it was demonstrated that when the input (arrival rate) to the $M/M/1$ queuing system is varied according to

$$\lambda(t) = \lambda_0 + \alpha \cos(2\pi\omega t) \tag{6.1}$$

then the corresponding system output (departure rate) is

$$\xi(t) = \lambda_0 + \frac{\alpha}{\sqrt{1+b^2}} \cos(2\pi\omega t - \tan^{-1}(b)) \tag{6.2}$$

where

$$b = \frac{2\pi\omega\mu}{(\lambda_0 - \mu)^2}. \tag{6.3}$$

By definition the frequency response of a system is defined as the ratio of the output spectrum to the input spectrum, at every frequency. That means that if the $M/M/1$ queue is regarded as a linear system, then the frequency transfer function of the system is

$$H(\omega) = \frac{1}{\sqrt{1 + \omega/\omega_c}} e^{i \tan^{-1}(\omega/\omega_c)} \tag{6.4}$$

where

$$\omega_c = \frac{(\lambda_0 - \mu)^2}{2\pi\mu} \tag{6.5}$$

Usually, the system frequency response is represented by two plots—the amplitude $|H(\omega)|$ versus ω and the phase angle $\angle H(\omega)$ versus ω. A plot of $|H(\omega)|$ versus ω and $\angle H(\omega)$ versus ω together is known as a *Bode plot.*

Bode plots for a $M/M/1$ queue can be obtained by using equation 6.4. Bode plots also can be obtained by performing a FDE for a $M/M/1$ queue—the magnitude of the ratio of the DFTs of the departure rate and the arrival rate at the frequencies of input oscillation gives $|H(\omega)|$ while the difference between the phase angles give $\angle H(\omega)$. Bode plots obtained by using the FDE Histogram method for different values of utilization are compared with the frequency response given by equation 6.4; figure 6.1 shows the amplitude plots and 6.2 the phase plots for different values of $\lambda_0/\mu = 0.1,\ 0.3,\ 0.5,\ 0.7$, with $\alpha_1 = 0.1$ and $\lambda_0 = 1.0$. The figures indicate that the analytical solution for $|H(\omega)|$ given by equation 6.4 matches the simulation

results even for high average utilization, while the analytical solution for $\angle H(\omega)$ matches the simulation results only for small utilization and ω.

6.2 Frequency Response of a Complex Network of Queues

This section investigates the frequency response of an open network of s $M/M/1$ service nodes. Each service node has an Exponential service time distribution and infinite capacity. For $v = 1, 2, \ldots, s$, the (stationary) service rate of node v is μ_v. If there is a network link between nodes v and v', then $p(v, v')$ represent the transition probability associated with the link. The 0^{th} node denotes the exterior of the network; the "source" of external arrivals and the "sink" for network departures. With this convention, the network transition matrix is written as:

$$\begin{pmatrix} p(0,0) & p(0,1) & p(0,2) & \ldots & p(0,s) \\ p(1,0) & p(1,1) & p(1,2) & \ldots & p(1,s) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p(s,0) & p(s,1) & p(s,2) & \ldots & p(s,s) \end{pmatrix}$$

The 0^{th} row of this matrix represents the external arrival probabilities; the 0^{th} column represents the network departure probabilities. External arrivals to the network occur according to a non-stationary Poisson process with rate

$$\lambda(t) = \lambda_0 + \alpha_0 \sin(2\pi\omega_0 t) \tag{6.6}$$

That means that, for $v' = 1, 2, \ldots, s$, external arrivals occur at node v' with probability $p(0, v')$ and therefore $\lambda(t)p(0, v')$ is the external arrival rate at node v'.

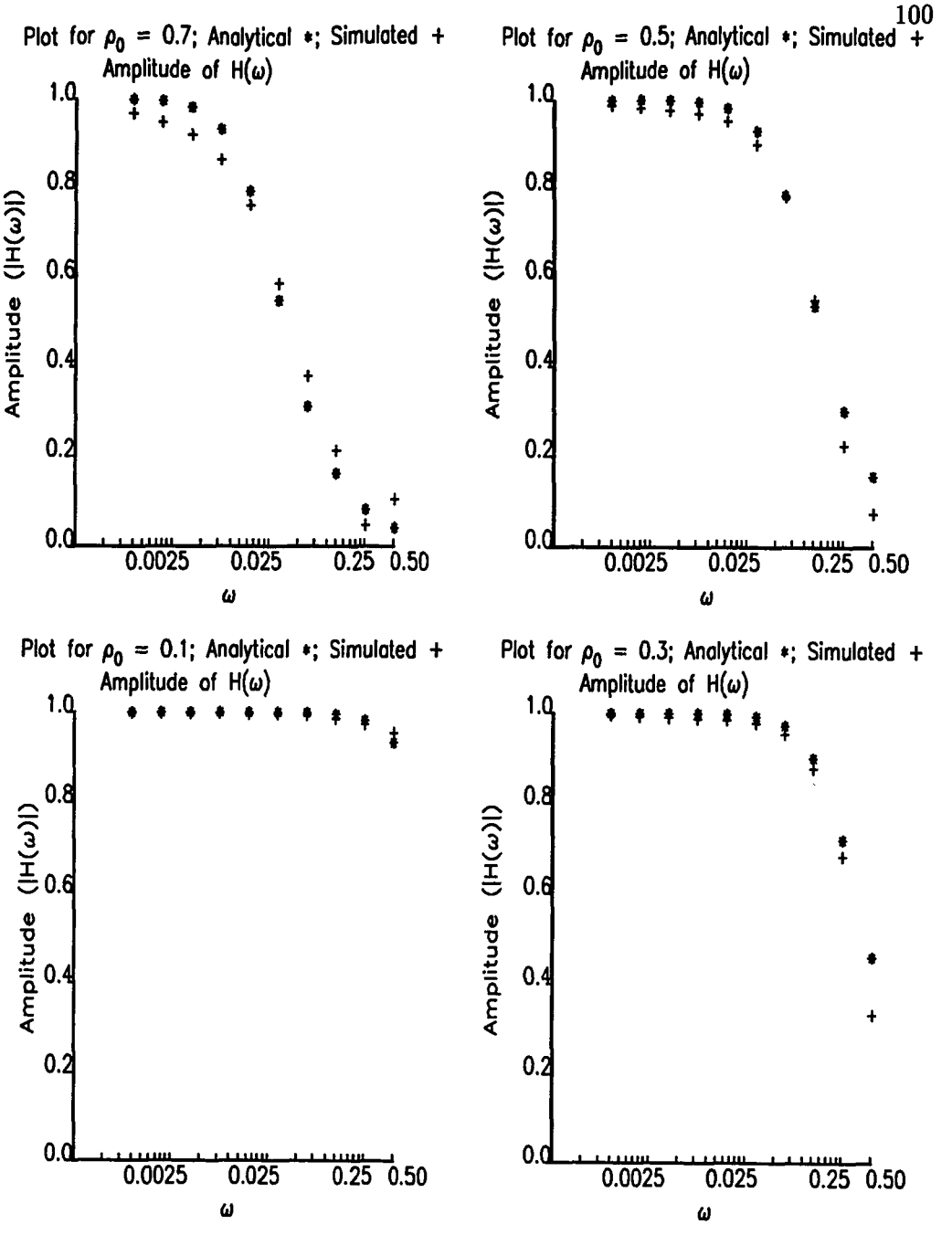


Figure 6.1: Amplitude plots for the frequency response of a $M/M/1$ queue comparing simulated and analytical results.

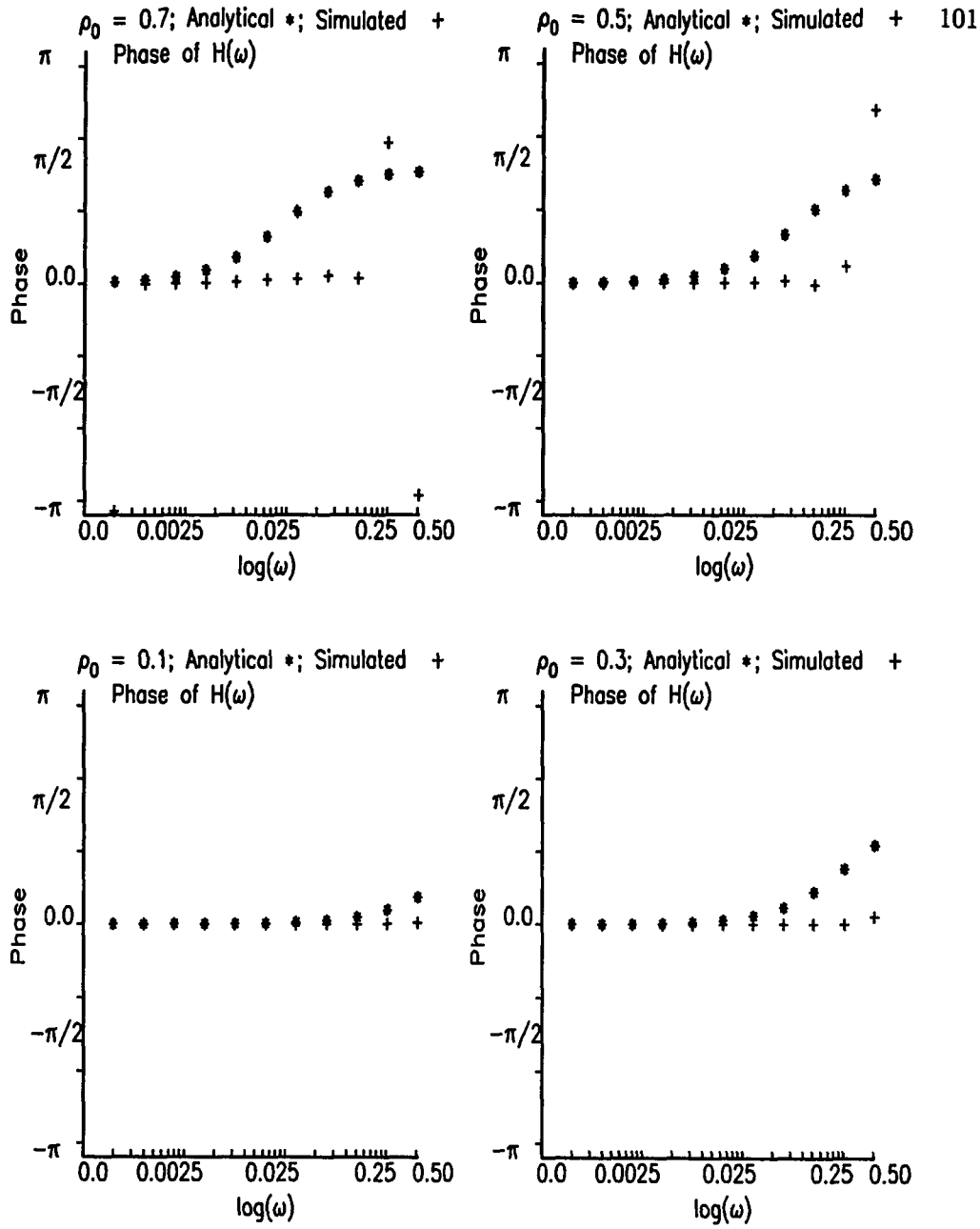


Figure 6.2: Phase plots for the frequency response of a $M/M/1$ queue comparing simulated and analytical results.

Let $\hat{\lambda}(\omega_0)$ represent the magnitude of the DFT of the external arrival sequence and let $\hat{\xi}_v(\omega_0)$ represent the magnitude of the DFT of the departure rate sequence from node v at ω_0.[1]. For each node v, let λ_v and α_v represent the sum of the nominal values of all inputs to the node and the sum of the amplitudes of oscillation for all inputs to the node respectively; let $k_v(\omega_0) = |H_v(\omega_0)| = 1/\sqrt{(1+b_v^2)}$ represent the magnitude of the transfer function from node v' where $b_{v'} = 2\pi\omega_0/\mu_{v'}(\lambda_{v'} - \mu_{v'})^2$. Provided the maximum amplitude of oscillation of the arrival rate at each node is less than its service rate, the "ω-in/ω-out" assumption for each node holds and thus, the "flow out power spectrum equals flow in power spectrum times transfer function of the node". The flow balance can be characterized by

$$\hat{\xi}_{v'}(\omega_0) = k_{v'}(\omega_0)\Big(\hat{\lambda}(\omega_0)p(0,v') + \sum_{v=1}^{s}\hat{\xi}_v(\omega_0)p(v,v')\Big) \qquad v' = 1,2,\ldots,s \qquad (6.7)$$

A unique solution of these equations can be obtained to express each $\hat{\xi}_{v'}(\omega_0)$ as a function of $\hat{\lambda}(\omega_0)$.

To demonstrate the application of equation 6.7, the "feed-forward" network in

[1]Only the magnitude of the DFTs at the frequency of oscillation of the external arrival rate $\omega = \omega_0$ is considered, because the magnitude of the DFTs at other frequencies is essentially negligible.

figure 3.8 is considered. For this network $s = 4$ and the transition matrix is

$$\begin{pmatrix} 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.6 & 0.4 & 0.0 \\ 0.2 & 0.0 & 0.0 & 0.0 & 0.8 \\ 0.2 & 0.0 & 0.0 & 0.0 & 0.8 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}$$

The only variation from the network in example 3.6 is that each node exponential server. External arrivals occur only at node 1 according to a non-stationary Poisson process with rate $\lambda(t)$ given by equation 6.6.

From equation 6.7, the departure rate spectrum from each node is

$$\hat{\xi}_1(\omega_0) = k_1(\omega_0)\hat{\lambda}(\omega_0) \tag{6.8}$$

$$\hat{\xi}_2(\omega_0) = k_2(\omega_0)(0.6\hat{\xi}_1(\omega_0)) \tag{6.9}$$

$$\hat{\xi}_3(\omega_0) = k_3(\omega_0)(0.4\hat{\xi}_1(\omega_0)) \tag{6.10}$$

$$\hat{\xi}_4(\omega_0) = k_4(\omega_0)(0.8\hat{\xi}_2(\omega_0) + 0.8\hat{\xi}_3(\omega_0)) \tag{6.11}$$

provided that the maximum amplitude of the arrival rate at each node is less than its service rate.

The (unique) solution to equations 6.8–6.11 is

$$\hat{\xi}_1(\omega_0) = k_1(\omega_0)\hat{\lambda}(\omega_0) \tag{6.12}$$

$$\hat{\xi}_2(\omega_0) = 0.6k_1(\omega_0)k_2(\omega_0)\hat{\lambda}(\omega_0) \tag{6.13}$$

$$\hat{\xi}_3(\omega_0) = 0.4k_1(\omega_0)k_3(\omega_0)\hat{\lambda}(\omega_0) \tag{6.14}$$

$$\hat{\xi}_4(\omega_0) = 0.48k_1(\omega_0)k_2(\omega_0)k_4(\omega_0)\hat{\lambda}(\omega_0) + 0.32k_1(\omega_0)k_3(\omega_0)k_4(\omega_0)\hat{\lambda}(\omega_0) \tag{6.15}$$

Table 6.1: λ_v, α_v, and k_v for each node v of the network.

v	λ_v	α_v	k_v
1	1.0	0.100	0.999
2	0.6	0.058	0.998
3	0.4	0.039	0.999
4	0.8	0.077	0.997

Example 6.1 A FDE for the network in figure 3.8, with $\lambda(t) = 1.0+0.1\sin(2\pi\omega_0 t)$, $\omega_0 = 0.031250$ and $\mu_v = 4.0$ for $v = 1,2,3,4$ yields $\hat{\lambda}(\omega_0) = 203$, $\hat{\xi}_1(\omega_0) = 202$, $\hat{\xi}_2(\omega_0) = 120$, $\hat{\xi}_3(\omega_0) = 80$ and $\hat{\xi}_4(\omega_0) = 160$. Table 6.1 gives the nominal values of the arrival rate, the maximum amplitude of oscillation and the magnitude of the transfer function at each node for the network i.e., for each node v, $\lambda_v + \alpha_v < \mu_v$. Substituting $\hat{\lambda}(\omega_0)$ and k_v for $v = 1,\ldots,4$ in equations 6.12–6.15 gives

$$\begin{aligned}
\hat{\xi}_1(\omega_0) &= (0.99)(203) = 202 \\
\hat{\xi}_2(\omega_0) &= (0.6)(0.99)(0.99)(203) \approx 121 \\
\hat{\xi}_3(\omega_0) &= (0.4)(0.99)(0.99)(203) \approx 81 \\
\hat{\xi}_4(\omega_0) &= 0.8((0.6)(0.99)(0.99) + 0.4(0.99)(0.99))(0.99)(203) \approx 161. \quad (6.16)
\end{aligned}$$

Thus the FDE results are (essentially) identical to the results obtained by using equation 6.7. That is, provided the maximum amplitude of oscillation of the arrival rate to each node is less than the service rate, the input frequency is invariant and

spectral energy in the network is flow balanced. Equation 6.7 also validates the claim made in chapter 3 that the height of the spike in the departure rate spectrum of each node is related to that of the other departure rate spectra via the transition probabilities of the network.

Example 6.2 A FDE for the feedback network given in figure 3.11, with $s = 5$ and the corresponding transition matrix

$$\begin{pmatrix} 0.0 & 0.4 & 0.0 & 0.6 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.8 & 0.2 & 0.0 & 0.0 \\ 0.2 & 0.4 & 0.0 & 0.0 & 0.0 & 0.4 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.0 & 0.0 & 0.8 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}$$

is performed with $\lambda(t) = 1.0 + 0.1\sin(2\pi\omega_0 t)$ where $\omega_0 = 0.031250$ and $\mu_v = 4.0$ for $v = 1, 2, 3, 4$. The FDE yields $\hat{\lambda}(\omega_0) = 203$, $\hat{\xi}_1(\omega_0) = 134$, $\hat{\xi}_2(\omega_0) = 136$, $\hat{\xi}_3(\omega_0) = 148$, $\hat{\xi}_4(\omega_0) = 148$ and $\hat{\xi}_5(\omega_0) = 171$. From equation 6.7, the flow-balance equations for the feed-back network are:

$$\hat{\xi}_1(\omega_0) = (0.4\hat{\lambda}(\omega_0) + 0.4y_2(\omega_0))k_1(\omega_0) \tag{6.17}$$

$$\hat{\xi}_2(\omega_0) = (0.8\hat{\xi}_1(\omega_0) + 0.2\hat{\xi}_4(\omega_0))k_2(\omega_0) \tag{6.18}$$

$$\hat{\xi}_3(\omega_0) = (0.6\hat{\lambda}(\omega_0) + 0.2\hat{\xi}_1(\omega_0))k_3(\omega_0) \tag{6.19}$$

$$\hat{\xi}_4(\omega_0) = \hat{\xi}_3(\omega_0)k_4(\omega_0) \tag{6.20}$$

$$\hat{\xi}_5(\omega_0) = (0.4\hat{\xi}_2(\omega_0) + 0.8\hat{\xi}_4(\omega_0))k_5(\omega_0). \tag{6.21}$$

The nominal value of the arrival rate, the maximum amplitude of oscillation and the magnitude of the frequency response at each node in the network are such that $\lambda_v + \alpha_v < \mu_v$ for $v = 1, 2, \ldots, s$ and $k_v \approx 1.0$. Substituting $\hat{\lambda}(\omega_0)$ and k_v for $v = 1, \ldots, 5$ in equations 6.17–6.21 gives

$$\begin{aligned}
\hat{\xi}_1(\omega_0) &= 0.4\hat{\lambda}(\omega_0) + 0.4\hat{\xi}_2(\omega_0) \\
\hat{\xi}_2(\omega_0) &= 0.8\hat{\xi}_1(\omega_0) + 0.2\hat{\xi}_4(\omega_0) \\
\hat{\xi}_3(\omega_0) &= 0.6\hat{\lambda}(\omega_0) + 0.2\hat{\xi}_1(\omega_0) \\
\hat{\xi}_4(\omega_0) &= \hat{\xi}_3(\omega_0) \\
\hat{\xi}_5(\omega_0) &= 0.4\hat{\xi}_2(\omega_0) + 0.8\hat{\xi}_4(\omega_0). \qquad (6.22)
\end{aligned}$$

The (unique) solution to these five equations is:

$$\begin{aligned}
\hat{\xi}_1(\omega_0) &= (560/830)\hat{\lambda}(\omega_0) \\
\hat{\xi}_2(\omega_0) &= (570/830)\hat{\lambda}(\omega_0) \\
\hat{\xi}_3(\omega_0) &= (610/830)\hat{\lambda}(\omega_0) \\
\hat{\xi}_4(\omega_0) &= (610/830)\hat{\lambda}(\omega_0) \\
\hat{\xi}_5(\omega_0) &= (716/830)\hat{\lambda}(\omega_0). \qquad (6.23)
\end{aligned}$$

Substituting $\hat{\lambda}(\omega_0) = 203$ in equation 6.23 yields

$$\hat{\xi}_1(\omega_0) = (560/830)(203) = 138$$

$$\hat{\xi}_2(\omega_0) = (570/830)(203) = 140$$

$$\hat{\xi}_3(\omega_0) = (610/830)(203) = 149$$

$$\hat{\xi}_4(\omega_0) = (610/830)(203) = 149$$

$$\hat{\xi}_5(\omega_0) = (716/830)(203) = 175. \tag{6.24}$$

The values of $\hat{\xi}_v(\omega_0)$ in equation 6.24 are approximately equal to the values generated by the FDE for the network. This example further demonstrates that (*i*) provided the maximum amplitude of oscillation of the arrival rate to each node is less than the service rate, each node in a network of $M/M/1$ queues obeys the frequency invariance property (*ii*) spectral energy in the network are flow balanced (*iii*) the height of the spike in the departure rate spectrum of each node at the frequency of oscillation of the external arrival rate is related to that of the other departure rate spectra via the transition probabilities of the network.

CHAPTER VII

Conclusions

7.1 Conclusions

Since their introduction in the early 80's, significant work has been done to extend the applicability of FDEs to regression analysis, simulation optimization and gradient estimation. Two fundamental theoretical and data analysis FDE problems remain, however. These problems are a roadblock to the widespread acceptance and use of FDEs. The objective of this dissertation has been to investigate and solve these problems.

To perform a FDE spectral analysis correctly, it is necessary to select a suitable oscillation parameter and corresponding oscillation index. Until recently, the proper choice of the oscillation parameter and index has been an open problem in the FDE literature—the so-called "FDE indexing problem". Solutions to the FDE indexing problem using the simulation clock time, the FDE Histogram method and the extended FDE Histogram method, were developed in chapters 3 and 4 of this dissertation, respectively. The FDE Histogram method is used for FDE data analysis when the selected system response statistic is a rate function. The *extended* FDE Histogram method is used for the FDE data analysis of non-rate response statistics. Several discrete-event simulations were provided to demonstrate

the effectiveness of the two methods.

FDEs are based upon the assumption that if a particular system response statistic is sensitive to a system parameter, then sinusoidal variation of that system parameter at a fixed frequency will induce similar sinusoidal variations in the response statistic, at the same frequency. Chapter 5 of this dissertation presented numerical and analytical support for this FDE model assumption for a $M/M/1$ queue. In chapter 6 the validity of this FDE model assumption was demonstrated for two networks of $M/M/1$ queues.

7.2 Future Research

7.2.1 Frequency Response of a $M(t)/M(t)/1$ queue

In this dissertation the frequency response for a $M/M/1$ queuing system with time-varying arrival rate and fixed service rate is derived. Further research is necessary to determine whether the derivation can be extended to $M/M/1$ queues with sinusoidally varying arrival and *service rates*, or for other non-stationary $M/G/1$ queues. Also, using the analytical expression for $P_0(t)$ for a $M/M/1$ queue, the analytical expressions for other system statistic like the expected number in the system and expected wait need to be derived.

7.2.2 Gradient Estimation

Recent efforts have been made by Jacobson [12] to apply traditional FDE methods to perform gradient estimation. The use of the FDE Histogram method and extended Histogram method for harmonic gradient estimation needs to be investigated.

7.2.3 Network Decomposition

This dissertation demonstrates that the frequency response of each network node in a network of queues is related to the arrival spectrum via the transition probabilities of the paths leading to that node. Hence the spectral estimates could be used as "estimators" for the transition probabilities in the network, for a fixed-topology network. Estimating the transition probabilities using a FDE would thus yield a method for performing the decomposition of complex networks or to identify network bottlenecks.

7.2.4 Performance Analysis of "Connected" Systems

Complex stochastic queuing networks occur in a wide variety of applications and their perturbation (sensitivity) analysis is a daunting task, except in the simplest of cases. Analytical methods for analyzing the performance of such systems is limited, if not non-existent in most cases. What is required is a simulation-based analysis technique that allows the models to be realistic (as opposed to being overly simplistic for tractable analysis) and at the same time provides the capability to

extract meaningful performance metrics from the simulation data. FDEs have the potential of providing a solution to this problem. In particular, the Fourier-based performance analysis of a network of time-dependent queues needs to be investigated for actual networks.

7.2.5 FDE for Terminating Simulations

As pointed by Sargent et. al. in [38], in most of the available FDE literature, the system is allowed to reach a *quasi* steady-state before the data is collected for spectral analysis. The Histogram method needs to be extended to perform data analysis for terminating simulations.

Bibliography

[1] Buss, A. H. Some extensions and limitations of frequency domain experiments. *1988 Winter Simulation Conference Proceedings*, pages 549–557, 1988.

[2] Chen H. and Schmeiser B. W. Simulation of Poisson Processes with Trignometric rates. *1992 Winter Simulation Conference Proceedings*, pages 609–617, 1992.

[3] Clarke A. B. A Waiting Time Process of Markov Type. *Ann. Math. Statist.*, vol. 27, pages 452-459, 1956.

[4] Cohen, J. W. *The Single Server Queue.* John Wiley and Sons, Inc., New York, 1969.

[5] Cooley J. W. and Tukey J. W. An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, vol. 19, pages 297-301, 1965.

[6] Gnedenko, B. V. and Kovalenko, I. N. *Introduction to Queuing Theory.* Translated from Russian by Eng. R. Kondor, Translation edited by D. Louvish, 1966.

[7] Gross, D. and Harris, C. M. *Fundamentals of Queuing Theory.* 2nd edition Wiley, 1985.

[8] Hardin, J. C. *Introduction to Time Series Analysis.* NASA Reference Publication 1145, March, 1986.

[9] Ho, Y.C. Eyler, M.A. and Chien, T.T. A Gradient Technique for General Buffer Storage Design in a Production Line. *International Journal of Production Research*, Volume 17, Number 3, 557-580

[10] Jacobson, S. H., Buss, A. and Schruben, L.W. Driving frequency selection for frequency domain simulation experiments. Technical Report 714, School of Operations Research and Industrial Engg., Cornell University, September 1987.

[11] Jacobson, S. H. Oscillation amplitude considerations in frequency domain experiments. *1989 Winter Simulation Conference Proceedings*, pages 406–410, 1989.

[12] Jacobson, S.H. Input/output model extension for gradient estimation on the frequency domain. *29th. TIMS/ORSA Joint National Meeting, Las Vegas*, 1990.

[13] Jacobson, S.H. and Schruben, L.W. A Simulation Optimization Procedure using Harmonic Analysis. Working Paper, Dept. of Operations Research, Weatherland School of Management, Case Western Reserve University, September, 1991.

[14] Jacobson, S.H. Variance and Bias Reduction Techniques for the Harmonic Gradient Estimator. *Applied Mathematics and Communication 55*, pages 153-186, Elsevier Science Publishing Co. Inc, 1993.

[15] Jacobson, S.H. Convergence Results for Harmonic Gradient Estimators. Working Paper, Dept. of Operations Research, Weatherland School of Management, Case Western Reserve University, 1991.

[16] Kolesar, P. J., Rider, K. L., Crabill, T. B. and Walker, W. E. A Queuing-Linear Programming Approach to Scheduling Police Patrol Cars. *Operations Research*, vol. 23, no. 6, pages 1045-1062, (Nov-Dec) 1975.

[17] Koopman, B. O. Air Terminal Queues under Time-Dependent Conditions. *Operations Research*, vol. 20, no. 6, pages 1089-1114, (Nov-Dec) 1972.

[18] Law, A. M. and Kelton, W. D. *Simulation Modeling and Analysis.* Second Edition, McGraw-Hill Book Company, 1991.

[19] Liu, C.L. and Liu J. W. S. *Linear System Analysis.* McGraw-Hill Book Company, 1975.

[20] Massey, W. A. Asymptotic Analysis of the Time Dependent M/M/1 Queue. *Math. of Operations Research*, vol. 10, no. 2, pages 305-327, May 1985.

[21] Luchak, G. The Solution of the Single-channel Queuing Equation Characterized by a Time-dependent Poisson-distributed Arrival Rate and a General Class of Holding Times. *Operations Research*, vol. 4, pages 711-732, 1956.

[22] Pro-Matlab, The MathWorks, Inc. 1984-1991, Version 3.5i, 23-Jul-1991

[23] Mathematica 2.1 for SPARC Wolfram Research, Inc., 1988-1992.

[24] Mitra, M. and Park, S. K. Solution to the Indexing Problem of Frequency Domain Simulation Experiments. *1991 Winter Simulation Conference Proceedings*, pages 907–915, 1991.

[25] Mitra, M., Morrice, D. and Park, S. K. On the Selection and Implementation of the Frequency Domain Experiment Oscillation Parameter. Technical Report 92/93-3-5, Department of Management Science and Information Systems, University of Texas at Austin, Graduate School of Business, Austin, Texas.

[26] Morrice, D., Jacobson, S.H. and Schruben, L.W. The global simulation clock as the frequency domain experiment index. *1988 Winter Simulation Conference Proceedings*, pages 558–563, 1988.

[27] Morrice, D. J. and Schruben, L. W. Simulation sensitivity analysis using frequency domain experiments. *1989 Winter Simulation Conference Proceedings*, pages 367–363, 1989.

[28] Murphy, G. M. *Ordinary Differential Equations and their Solutions.* Princeton, N.J., Van Nostrand, 1960

[29] Myers, R.H. *Response Surface Methodology* Allyn and Bacon, Boston, 1971 (reprinted by Edwards Bros., Ann Arbor, Michigan.)

[30] Newell, G. F. Queues with Time-Dependent Arrival Rates I, II, III. *Journal of Applied Probability*, May 1968, pages 436-451, 579-590 and 591-606.

[31] Papoulis, A. *Probability, Random Variables and Stochastic Processes.* McGraw Hill Book Company, 1965.

[32] Rider, K. K. A Simple Approximation to the average queue size in the Time-Dependent M/M/1 Queue. *Journal of the ACM*, April 1976, Volume 23, Number 3, pages 361–367.

[33] Rothkopf, M. H. and Oren, S. S. A Closure Approximation for the Nonstationary M/M/s Queue. *Management Science*, June 1979, Volume 25, Number 6, pages 522–534.

[34] Ross, Sheldon M. *Stochastic Processes.* Hohn Wiley & Sons, 1983.

[35] Saaty, T. L. *Elements of Queuing Theory.* McGraw-Hill Book Company, Inc., 1961.

[36] Sanchez, P. J. and Schruben, L. W. Simulation factor screening using frequency domain methods: An illustrative example. Submitted for publication, 1986.

[37] Sanchez P.J. and Buss, A. H. A model for frequency domain experiments. *1987 Winter Simulation Conference Proceedings*, pages 424–427, 1987.

[38] Sargent, R. G. and Som, T. K. A Different View of Frequency Domain Experiments. *Management Science*, 38(5), 667-687, 1992.

[39] Schruben, L. W. Simulation optimization using frequency domain methods. *1986 Winter Simulation Conference Proceedings*, pages 366–369, 1986.

[40] Schruben, L. W. and Cogliano, V. J. Simulation sensitivity analysis : A frequency domain approach. *1981 Winter Simulation Conference Proceedings*, pages 455–459, 1981.

[41] Schruben, L. W. and Cogliano, V. J. An Experimental Procedure for Simulation Response Surface Model Identification. *Communications of the ACM*, August 1987, Volume 30, Number 8, pages 716–730.

[42] Shannon, C. E. Communication in the presence of noise. *Proc IRE*, Volume 37, Number 1, January 1949, pages 10–21.

[43] Som, T.K. and Sargent, R.G. Alternate methods for generating and analyzing the output series of frequency domain experiments. *1988 Winter Simulation Conference Proceedings*, pages 564–567, 1988.

[44] Sargent, R.G. Som, T.K. and Schruben, L.W. Frequency domain modeling of a feedback queue. *1987 Winter Simulation Conference Proceedings*, pages 419–423, 1987.

[45] Taaffe, M. R. and Ong, K. L. *Approximating Nonstationary Ph(t)/M(t)/s/c queuing systems. Ann. of Operations Research*, Vol 8, pages 103-116, 1987.

[46] Takacs, L. *Introduction to the Theory of Queues.* Oxford University Press, New York, 1962.

[47] Tipper, D. and Sudareshan, M. K. Numerical Methods for Modeling Computer Networks Under Nonstationary Conditions. *IEEE Journal on Selected Areas in Communications*, vol. 8, No. 9, December 1990, pages 1682-1695

[48] Zhang, J. and Coyle, E. J. The Transient Solution of Time-Dependent M/M/1 Queues. *IEEE Transactions on Information Theory*, Vol 37, no. 6, pages 1690-1696, November, 1991.

VITA

Mousumi Mitra Hazra

The author was born in Calcutta, India on the 19th day of September, 1964. She received her Bachelor's Degree in Computer Science and Engineering from Jadavpur University, Calcutta, in 1987. She then entered the graduate program in the Department of Computer Science at the College of William & Mary in the Fall of 1987, and was awarded a Master of Science in Computer Science in December of 1989. She was admitted into the Ph.D. program in August 1988, and has been a Senior Programmer Associate for Lockheed Engineering and Sciences Company since January, 1990.