Multiscale Methods for the Analysis and Application of Fractal Point Processes and Queues

Warren Michael Lam

RLE Technical Report No. 614

April 1997

This work has been supported in part by the Defense Advanced Research Projects Agency monitored by the U.S. Navy, Office of Naval Research under Contract N00014-93-1-0686; an associated AASERT monitored by the U.S. Army Research Office under Contract DAAH4-95-1-0473; the U.S. Air Force Office of Scientific Research under Grant F49620-96-1-0072; and the U.S. Navy, Office of Naval Research under Grant N00014-96-1-0930.

The Research Laboratory of Electronics
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139-4307

Multiscale Methods for the Analysis and Application of Fractal Point Processes and Queues

by

Warren Michael Lam

Submitted to the Department of Electrical Engineering and Computer Science
on January 31, 1997, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Fractal point processes are increasingly being viewed as important models for a host of natural and man-made phenomena. To adequately exploit such models, efficient techniques for processing, analyzing, and synthesizing fractal point processes in the context of such applications are required. This thesis develops a broad set of practical results for an important class of quasi-stationary fractal point processes we refer to as the fractal renewal processes. Starting from an engineering-oriented mathematical characterization of such processes, a novel multiscale framework, based on random mixture of Poisson constituents, is formulated for these processes, and serves as the foundation for analysis and algorithm development.

Using this framework, efficient signal processing algorithms are developed for fractal renewal processes, including a synthesis which requires a single Poisson process generator for its implementation. Included are a continuous-scale version for exact synthesis, and a discrete-scale version for arbitrarily accurate synthesis using a countable collection of constituents. Complementary multiscale analysis algorithms are also developed, aimed primarily at robust parameter and signal recovery in a noise-corrupted scenario. Specifically, a maximum-likelihood fractal dimension estimator and a Bayes' least-squares interarrival estimator are derived. Performance is evaluated using simulations and theoretical bounds.

Characterizations of fractal renewal processes in familiar discrete-event systems, particularly networks and queues, are also obtained using multiscale methods. The results, which include the interarrival density for a fractal point process subject to random erasure, and the counting process distribution for superposition of fractal point processes, suggest invariance of key fractal properties under traffic branching and merging. Steady-state customer distributions in a memoryless queue with fractal customer arrivals, and in the complementary system with power-law service and memoryless arrivals, are also derived and verified by simulations.

A number of problems of network design and management are also explored. Optimal multiscale server control policies are developed for queueing systems with fractal traffic input, which exploit past history of traffic to enhance performance. Comparison with policies which ignore past history suggests that our multiscale controller is superior in terms of average individual waiting time and service costs. Finally, optimal flow control policies are also formulated for power-law service of memoryless traffic.

Thesis Supervisor: Gregory W. Wornell
Title: Associate Professor of Electrical Engineering

Acknowledgments

I am deeply grateful to my mentor, my thesis supervisor Prof. Gregory Wornell, for not only his guidance in the technical aspects of my research, but also his patience, encouragement, and friendship, which have made my doctoral studies most rewarding and enjoyable. I have benefited greatly from all interactions with Greg, both formal and informal. Greg's influence on my career as an engineer and as a researcher is unquestionably unsurpassed.

I also feel deeply indebted to the members of my thesis committee, Prof. Alan Willsky and Prof. George Verghese, for the fruitful discussions which have resulted in numerous innovations for enhancing this thesis.

I express heartfelt gratitude to my parents, my brother Walter, and my sister-in-law Peggie for being constantly caring and supportive, in good times and bad. I am also sincerely thankful to my uncle and aunt, Mr. and Mrs. Ting-Fai Lam, who have been unchangingly rooting for me ever since I set foot in this country.

My colleagues in the DSPG must also be acknowledged, for their friendly interaction, as well as their constant recommendation to switch from `vi` to the superior `emacs`. Special thanks are due for my two office mates: Dr. Steven Isabelle for all his thoughtful advice, both in and out of the office; and Haralabos Papadopoulos for his valuable help and comments, which have made all the difference in the development of this thesis.

In addition, many have contributed greatly in my life, particularly in my years at M.I.T. To them I sincerely dedicate this thesis. In no particular order, I thank Grace Chung for her care and company, and constant reminder that I should work *really* hard; Yitwah Cheung for all his thought-provoking mathematical recreations as well as philosophical discussion of life in general; Shiu-Kei Cheung for his tasteful humor and for being the impeccable guide through the classical and contemporary music; and Walter Lee for keeping me up to date on the latest computer technology in general and athena hacks in particular. This list would be incomplete without Sunny Tam, Helen Meng, Shiufun and Josephine Cheung, and Albert Wong, whose friendships are much appreciated and cherished.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Efficient processing, analysis, and synthesis of fractal signals are increasingly important in many engineering applications. Generally speaking, a *fractal* is a mathematical construct which is *self-similar,* and hence possesses uniform structure over all scales. As such, the geometry of fractals is well-matched with a truly remarkable range of natural and man-made phenomena which lack a characteristic scale. As a few examples, landscape and meteorological features, turbulence formation, percolation transition in metals, and economic time series have all been adequately modeled as fractals (see, e.g., [14] [33] and the references therein). Naturally, advancement in the studies of fractals can further our understanding of these phenomena. Moreover, recent application of fractals in engineering studies has generated numerous innovative algorithms such as realistic rendering of natural images [38], efficient image compression [21], and signal design for communications [47].

While previous studies of fractals have focused primarily on fractal waveforms, self-similarity in discrete-event signals and systems, and the associated mathematical models of fractal point processes and queues, have recently sparked much research interest in the engineering community. Most notably, it has been widely observed that packet traffic over a variety of telecommunication networks possesses self-similar properties. This significant deviation from the traditional Poisson traffic model has much potential impact on the restructuring of existing network architectures and protocols, as well as on future networking design. Hence, there is a need for effective analysis of fractal traffic in networks. In addition to networking issues, important signal processing problems involving fractal point processes, such as synthesis, analysis, and estimation, have not been adequately solved. Our main

objective in this thesis is to address a number of these important problems.

Recently, multiscale paradigms have injected much insight into the study and application of fractal geometry. For example, Horowitz has developed filtering algorithms for a particular type of fractal noise model—the gyroscopic noise—using an infinite-dimensional state space model [17]. The multiscale ARMA-based models of Keshner have also been of interest, primarily in the synthesis of $1/f$ signals, as well as in the understanding of long-range memory inherent in these signals [24]. In another important direction, the wavelet transform has been formally shown to be a Karhunen-Loève-like basis for the $1/f$ family of fractal waveforms, and has given rise to a host of efficient modeling algorithms, including synthesis, estimation, and detection of these waveforms [47]. Motivated by the success of these multiscale paradigms, we introduce in this thesis a novel multiscale representation for a class of fractal point processes, which constitutes as the basis for our analysis and algorithm development. As will be apparent, there exist interesting connections between our multiscale representation and the aforementioned frameworks.

1.1 Outline of the Thesis

The problems addressed in this thesis fall coarsely into three categories. The first of these is composed of general signal processing problems such as synthesis and estimation of fractal point processes. Among other applications, the resulting algorithms are directly useful for signal classification and reconstruction. A second collection of problems focuses on the analysis of these point processes under random erasure, superposition, and queueing. Our results in this area offer much insight in the behavior of fractal traffic in various networking scenarios. Finally, we consider the design of optimal control policies for fractal traffic and fractal queues. The policies developed generally suggest the importance of past history for enhancing controller performance.

We begin by developing in Chapter 2 a sufficiently formal definition of fractal renewal processes, a class of fractal point processes in which we specialize throughout this thesis. In addition to its self-similarity, a key feature of this signal model is its quasi-stationarity, which is desirable for modeling phenomena with no preference for a space or time origin. In the same chapter, we reinforce the significance of this signal model by demonstrating the close agreement between its statistics and a wide range of physical phenomena observed

independently in remotely related areas, from communications engineering to biomedical research.

In Chapter 3, we introduce and develop a novel multiscale representation for the fractal renewal process. Based on the mixture of a multiscale family of Poisson processes, this framework allows exploitation of familiar results from the theory of Poisson processes. As an immediate application, this multiscale paradigm leads to efficient synthesis of fractal renewal processes, which includes a continuum version for exact synthesis, and a discretized version for synthesis with only a discrete or finite collection of constituents.

In Chapter 4, we apply our multiscale framework to a number of practical analysis and estimation problems involving fractal renewal processes. Specifically, we develop robust parameter and signal estimators for a realistic noisy scenario. We show that for the noise model considered, our fractal dimension estimation algorithm is more reliable than a number of existing estimators for the same parameter. On the other hand, our signal estimation algorithm is useful for recovery of fractal renewal processes from noise-corrupted measurements.

In Chapter 5, our emphasis shifts from general processing of fractal point processes to the analysis of these point processes in networking scenarios. Specifically, we concentrate on networking transformations typically experienced by traffic streams, including merging, branching, random erasure, and queueing. A main conclusion of our analysis is that interaction of multiple fractal point processes, such as merging and branching, often preserves key features of these signals. This offers additional evidence for self-similarity of the aggregate network traffic due to multiple users, which has been previously argued using mainly empirical observations.

While the queueing analysis of Chapter 5 is mainly descriptive, in Chapter 6 we shift emphasis to the synthesis of efficient queueing systems for manipulating fractal traffic. Our results in this chapter are generally insightful for the design of optimal networks for processing fractal point processes. Based on our multiscale framework, we apply dynamic programming and Markov decision techniques to formulate optimal server control and flow control algorithms. Although the problems we consider involve very specific objective functions and queueing set-ups, our results suggest that for the type of self-similar traffic and fractal servers considered, performance of queueing control can be enhanced by exploiting past history of the system, which is generally ignored in the optimal control of memoryless

queueing systems.

Finally, a number of concluding remarks are given in Chapter 7. Several directions for future research are also suggested.

Chapter 2

Self-Similar Point Phenomena and the Fractal Renewal Process

Numerous discrete-event distributions lack a characteristic scale. For example, in stellar and planetary distributions, the clustering of galaxies on coarse scales resembles the clustering of stars and planets on finer scales [40]. Likewise, the zero-crossings of a fractal waveform generally form a self-similar set. Recent studies of these phenomena have resulted in various mathematical characterizations for point processes having self-similar behavior, or fractal point processes. Models ranging from fractal doubly-stochastic processes of [22] to conditionally stationarity self-similar point processes of [32] have been developed for various phenomena ranging from auditory neural firing pattern to telephone channel error statistics. However, direct manipulations of many of these models are often difficult. The main purpose of this chapter is to develop a specific mathematical definition for a class of fractal point processes, sufficiently formal for the engineering-oriented discussion in the rest of this thesis. While we take an axiomatic approach in the definition, we show that this point process family exhibits many fractal properties observed in real phenomena.

This chapter is organized as follows. We begin by briefly reviewing the fundamentals of point process theory in Section 2.1, highlighting classical models such as the Poisson process and the renewal process. We present these basic models both for pedagogical purposes and because they will be central in our subsequent study of the fractal renewal process. In Section 2.2, we present background motivation for fractal point processes, citing results from recent case studies in various engineering and scientific areas. Our main technical

contribution in this chapter is a mathematical characterization of a class of fractal point processes—the fractal renewal processes—presented in Section 2.3. In the same section, we also present its key characterization in terms of the power-law interarrival density. Finally, additional statistical properties of fractal renewal processes are explored in Section 2.4.

2.1 Fundamental Point Process Theory

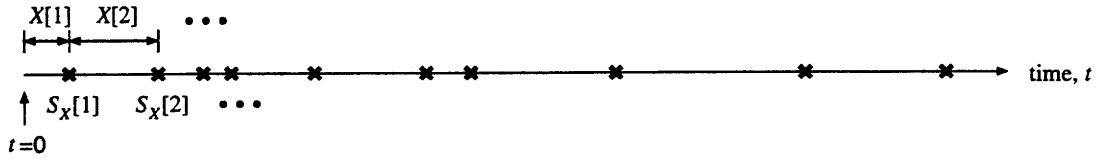
Roughly speaking, a point process is a collection of localized events, or "points," distributed randomly in space and/or time. While this thesis focuses on processes in a one-dimensional ambient space, extension of the theory to higher-dimensional cases is often immediate. Throughout, we refer to the ambient space as time, to emphasize the importance of these signal models in various queueing and scheduling contexts. A common graphical representation of a one-dimensional point process is shown in Fig. 2-1(a), where each point, or arrival, is marked with an "×" along the time axis. A useful mathematical characterization of a point process is in terms of the interarrivals $\{X[i]; i = 1, 2, \dots\}$, where $X[1]$ measures the gap between the time origin $t = 0$ and the first arrival, while $X[i]$ corresponds to the gap between the $(i-1)$st and ith arrivals, for $i > 1$. The location, or epoch, of the ith arrival is then clearly

$$S_X[i] = \sum_{i'=1}^{i} X[i'], \quad i = 1, 2, \dots.$$

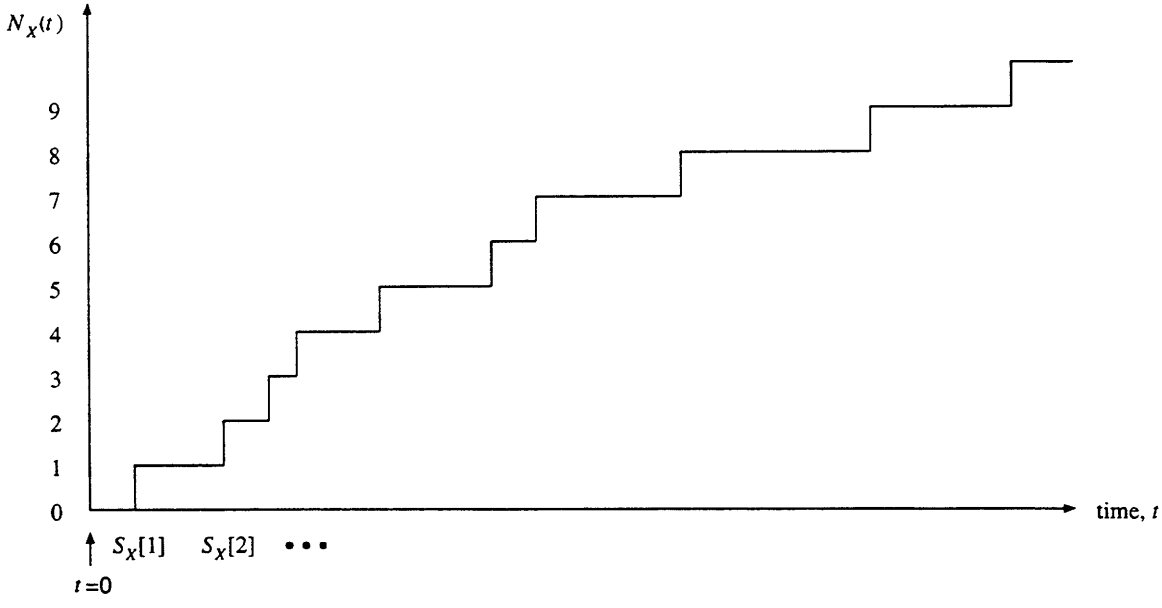
Alternatively, a point process can be characterized in terms of the associated counting process $N_X(t)$, formally defined as the number of arrivals in the interval $(0, t]$. As such, $N_X(t)$ is right-continuous. Fig. 2-1(b) depicts the counting process associated with the point process of Fig. 2-1(a). Among other relations, these quantities are related by

$$N_X(t) = \sup\{i : S_X[i] \leq t\}, \quad t > 0.$$

Because of its simplicity, the Poisson process traditionally plays a central role in point process studies. Formal definitions of this point process are typically given in terms of a single positive parameter λ, the arrival rate or the intensity of the process (see, e.g., [41] and [15]). A key feature of the Poisson process is its lack of memory: regardless of the activities elsewhere, the probability that an arbitrary infinitesimal interval of size dt contains zero



(a)



(b)

Figure 2-1: *Representations of a point process in a one-dimensional ambient space: (a) Arrivals are each denoted as an "×" along the horizontal axis. The sequence* $\{X[1], X[2], \dots\}$ *denotes the interarrival times, while the sequence* $\{S_X[1], S_X[2], \dots\}$ *denotes the arrival epochs. (b) Counting process corresponding to the point process of (a). The function* $N_X(t)$ *gives the number of points occurring in the interval* $(0, t]$ *and is thus continuous from the right.*

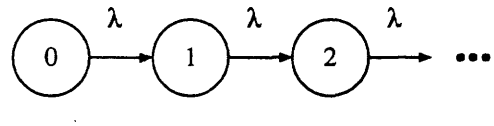


Figure 2-2: *The pure-birth process representing a Poisson counting process with arrival rate λ. Each state represents a certain number of births since the time $t = 0$. The departure rate from every state is λ.*

arrivals is $1 - \lambda\ dt$, while the probability it contains one arrival is $\lambda\ dt$. Probability of more arrivals is negligible. Several properties of the Poisson process are well known. First, its interarrivals $\{X[i], i = 1, 2, \dots\}$ are statistically independent, and are identically distributed according to the exponential probability density function

$$f_X(x) = \lambda \exp(-\lambda x), \quad x \geq 0. \tag{2.1}$$

Second, a closed-form probability distribution has been obtained for its counting process, which takes the form of a Poisson distribution,

$$\Pr\{N_X(t) = n\} = \frac{(\lambda t)^n \exp(-\lambda t)}{n!}, \quad n = 0, 1, \dots.$$

In addition, the Poisson counting process has independent increments, and is therefore Markov. A well-known continuous-time Markov representation of the Poisson counting process is the pure-birth model of Fig. 2-2 which, among other applications, is insightful for queueing analysis. Each state in the diagram represents a certain number of arrivals since the time $t = 0$, and the rate of departure from every state is λ. Adding to the list of well-known results, many signal processing problems such as shot-noise analysis, and various queueing problems have been adequately solved for the Poisson family [41] [26]. Finally, we remark that several variations of the Poisson process have also attracted much attention, including the inhomogeneous Poisson process whereby the intensity is a time-varying deterministic function $\lambda(t)$, and the doubly stochastic process, for which $\lambda(t)$ is itself stochastic. We refer the reader to [41] for in-depth studies of these Poisson extensions.

Poisson processes are members of a broader class of point processes known as renewal processes. A useful formal definition of the renewal process, sometimes more specifically

referred to as the ordinary renewal process [8], is a point process in which the interarrivals are independent, and identically distributed according to some common probability density function $f_X(x)$. Alternatively, if $X[1]$ is distributed differently, according to the density

$$f_{X[1]}(x) = \frac{1}{E\{X\}}\left(1 - \int_0^x f_X(u)du\right), \tag{2.2}$$

the resulting process will be stationary in the strict sense, and is sometimes referred to as an equilibrium renewal process [8]. Note that (2.2) is the waiting-time density upon random incidence, which accounts for a randomization of the time origin. Thus, the renewal process can be regarded as effectively stationary, or quasi-stationary.

A useful measure of the burstiness of a renewal process is its coefficient of variation C_X, formally defined as

$$C_X \triangleq \frac{\sigma_X}{\mu_X},$$

where μ_X and σ_X denote the mean and standard deviation of the interarrival distribution, respectively [9]. For evenly-spaced events, the coefficient of variation tends to be low, while for bursty arrivals, this quantity tends to be high. We remark that as the mean and standard deviation of the exponential density (2.1) are both $1/\lambda$, the coefficient of variation of every Poisson process is unity.

Except for the Poisson case, interarrivals of a renewal process are in general non-exponential, and the associated counting process non-Markov[1]. However, two classes of renewal processes are intimately related to the Poisson process, and are thus sufficiently captured within the Markov framework. An Lth-order Erlang process has Lth-order Erlang random variables for interarrivals, with the probability density function

$$f_X(x) = \frac{\lambda^L x^{L-1} e^{-\lambda x}}{(L-1)!}, \quad x > 0.$$

It is well known that this random variable is merely the sum of L independent, identically-distributed exponential random variables. Thus, an Lth-order Erlang process can be derived from a Poisson process by counting only every Lth arrival. This motivates the Markov process of Fig. 2-3. Generalizing the idea of a pure-birth process, this Markov process

[1]Generally speaking, however, a renewal process is sufficiently modeled as a semi-Markov process [15].

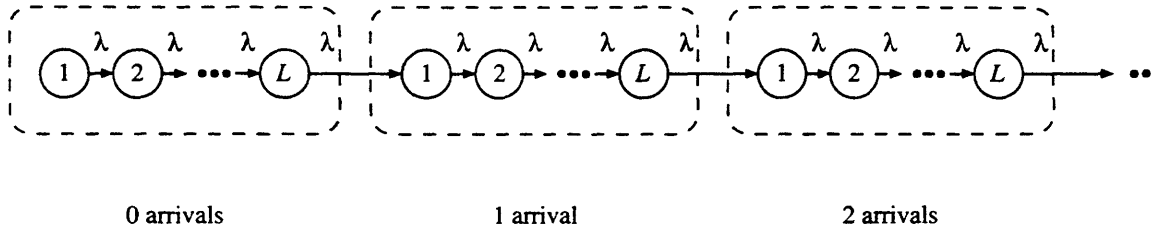


Figure 2-3: *A generalized pure-birth process corresponding to an Lth-order Erlang renewal process with parameter λ. Dashed boxes denote conceptual grouping into superstates. The sojourn time in each superstate is the sum of L independent exponential random variables, and is thus an Lth-order Erlang random variable.*

consists of a collection of "superstates," each of which represents a certain number of arrivals since the time $t = 0$. The arrival count is incremented only after L transitions of the underlying Poisson process.

As summation of well-behaved independent random variables generally reduces variability, Erlang interarrivals are expected to be more regular than Poisson interarrivals. Indeed, the coefficient of variation of an Lth-order Erlang process is

$$C_X = \frac{\sqrt{L} \cdot 1/\lambda}{L \cdot 1/\lambda} = \frac{1}{\sqrt{L}} \leq 1,$$

with equality only in the first-order case, where the process degenerates to a Poisson process. Hence, higher-order Erlang processes can be used to model more evenly-spaced event arrivals.

In the pure-birth representation of the Erlang process (Fig. 2-3), each superstate is a cascade of primitive states. Fig. 2-4 shows a parallel arrangement of the primitive states, which gives rise to the hyperexponential process. For an Lth-order hyperexponential process, each superstate consists of a collection of L states, with departure rates $\{\lambda_j; j = 0, 1, \ldots, L-1\}$. For meaningful discussion, λ_j will be assumed all distinct. Upon entering a superstate, a constituent is selected based on the distribution $\{p_j; j = 0, 1, \ldots, L-1\}$. Thus, the sojourn time in each superstate $X[i]$ is obtained as the random choice,

$$X[i] = X_j[i], \quad \text{with probability } p_j,$$

where $X_j[i]$ is an interarrival extracted from the Poisson process with rate λ_j. It follows that the probability density function of the resulting interarrival $X[i]$ assumes the form of

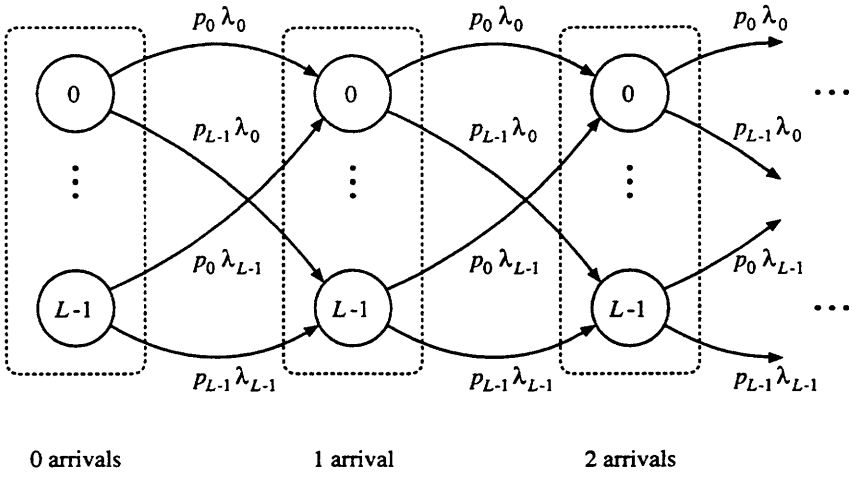


Figure 2-4: *A generalized pure-birth process corresponding to an Lth-order hyperexponential process. Dashed boxes denote conceptual grouping into superstates. The sojourn time in each superstate is the random choice among L independent exponential random variables, and is thus an Lth-order hyperexponential random variable.*

an Lth-order hyperexponential function

$$f_X(x) = \sum_{j=0}^{L-1} p_j \lambda_j \exp(-\lambda_j x), \quad x > 0.$$

It is straightforward to show that the coefficient of variation of the hyperexponential process is at least 1. To see this, first note that the mean and variance of the Lth-order hyperexponential interarrival density (2.1) are respectively

$$\mu_X = \sum_{j=0}^{L-1} p_j \frac{1}{\lambda_j},$$

and

$$\sigma_X^2 = E[X^2] - (E[X])^2 = \sum_{j=0}^{L-1} p_j \frac{2}{\lambda_j^2} - \left(\sum_{j=0}^{L-1} p_j \frac{1}{\lambda_j} \right)^2.$$

Next, we note the inequality

$$\sum_{j=0}^{L-1} p_j \frac{2}{\lambda_j^2} - 2\left(\sum_{j=0}^{L-1} p_j \frac{1}{\lambda_j}\right)^2 = 2\left(\sum_{j=0}^{L-1} p_j \frac{1}{\lambda_j^2} - \left(\sum_{m=0}^{L-1} p_m \frac{1}{\lambda_m}\right)^2\right)$$
$$= 2\left(\sum_{j=0}^{L-1} p_j \left[\frac{1}{\lambda_j} - \sum_{m=0}^{L-1} p_m \frac{1}{\lambda_m}\right]^2\right) \geq 0,$$

with equality if and only if all but one of the probabilities p_j are zero. So,

$$\sigma_X^2 = \sum_{j=0}^{L-1} p_j \frac{2}{\lambda_j^2} - \left(\sum_{j=0}^{L-1} p_j \frac{1}{\lambda_j}\right)^2 \geq \left(\sum_{j=0}^{L-1} p_j \frac{1}{\lambda_j}\right)^2 = \mu_X^2,$$

implying that the coefficient of variation C_X is greater than or equal to 1, with equality only in the first-order case. Thus, higher-order hyperexponential processes are well-suited for modeling bursty point phenomena. As will be apparent, the hyperexponential model has interesting connections with our multiscale representation for fractal point processes.

2.2 Self-Similar Point Process Models

While the Poisson process is desirable for its high analytic tractability, it is inadequate for capturing memory inherent in many phenomena, particularly various forms of clustering behavior. In the distribution of human population, for example, locations of individuals are highly correlated. In fact, a hierarchy of clustering is often present, from the gathering of households into neighborhoods, to the conglomeration of communities to form metropolises. Moreover, for many phenomena, the hierarchy of clustering exhibits much homogeneity. As an example, the clustering of galaxies on coarse length scales resembles stellar and planetary clustering on scales much finer [40].

Formal studies of self-similar point processes have been largely phenomenological, driven primarily by the ubiquity of such scale-invariant clustering activities. In the sequel, we present a series of case studies which reinforce the importance of these point processes as signal models. While these phenomena have been observed in remotely related areas, they have all been usefully modeled as fractal point processes.

Transmission Errors

In his seminal paper [32], Mandelbrot proposed a novel self-similar model for error occurrences in the telephone channel. While the observed burstiness suggests interdependence of error arrivals, the *intervals* $\{X[i]; i = 1, 2, \dots\}$ between errors are modeled as independent in Mandelbrot's model. Error correlation and self-similarity are built in via the fractional power-law intererror density function

$$f_X(x) = \begin{cases} (\gamma - 1)x^{-\gamma} & x \geq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

where $1 < \gamma < 2$. To accommodate for pathological peculiarities of this probability law, such as unbounded moments, Mandelbrot also introduced a notion of conditional stationarity to formalize discussion.

To motivate his point process model, Mandelbrot demonstrated its close agreement with real telephone data. Conditional and unconditioned first-order intererror histograms, as well as unconditioned higher-order intererror histograms are as predicted by the model, provided that correction factors are included to account for blocking effects [32]. Validity of Mandelbrot's model is further established by the studies in [42], conducted based on statistics of error-free transmission in no-correction and single-error-correction scenarios. As will be apparent, our fractal point process model has important connections to Mandelbrot's model, though we take an inherently different approach in the definition. For example, among other properties, we will show in Section 2.3 that the power-law interarrival distribution is an inevitable consequence of self-similarity in renewal processes.

Neural Activities

Recent studies of auditory neuron firing patterns have also uncovered self-similarity in these neural activities. In [43], Teich presented statistical analysis of auditory neural impulse trains based on pulse-number distributions (PND), which are in essence count histograms. Specifically, an observation of a spike train is partitioned into nonoverlapping windows each of duration T, to yield a sequence of counts $\{N[i]; i = 1, 2, \dots, n\}$ where n is the number of windows in the partition, and $N[i]$ is the number of spikes in the ith window. This sequence then naturally leads to estimates of the pulse-number distribution, its mean

$\mu_{N(T)}$, and variance $\sigma^2_{N(T)}$. To quantify the regularity of the distribution, a normalized variance estimate, frequently referred to as the Fano factor,

$$F(T) \triangleq \frac{\hat{\sigma}^2_{N(T)}}{\hat{\mu}_{N(T)}}, \tag{2.4}$$

may be used, where $\hat{\mu}_{N(T)}$ and $\hat{\sigma}^2_{N(T)}$ are respectively estimates of $\mu_{N(T)}$ and $\sigma^2_{N(T)}$ based on the partition. In the case of a Poisson process, the Fano factor tends to be independent of the partition size, while for periodic point arrivals, the Fano factor is zero [43]. For neural firing data, however, the Fano factor exhibits a dichotomy: while it remains constant for small windows due to refractoriness, it increases in a power-law fashion for larger windows [43]. Scale invariance of the Fano factor in the coarse-scale regime suggests fractal nature of the impulse arrivals. More importantly, based on the variation in the Fano factor, the fractal dimension of auditory neural impulse trains is estimated to range from 0.3 to 0.9, with dependence on the presence and level of stimuli. Self-similarity of neural firing is postulated to be optimal for sampling fractal stimuli so prominent in nature.

In [22], Johnson, et al., proposed a self-similar point process model to match the empirical observations of [43]. They established that for a general doubly stochastic process, the Fano factor has the asymptotic behavior

$$\lim_{T\to\infty} F(T) = 1 + \frac{1}{\mu_\lambda} \lim_{f\to 0} S_\lambda(f)$$

where μ_λ and $S_\lambda(f)$ are the mean and power spectral density of the intensity function $\lambda(t)$. Thus, an asymptotically scale-invariant Fano factor can be achieved with a self-similar intensity function. While this doubly stochastic model mimics the Fano factor behavior of neural pulses, it is analytically difficult to work with, and generates limited insight for signal processing and queueing problems involving fractal point processes. Moreover, describing a point process in terms of its intensity function obscures the discrete nature of the arrival process.

Network Transactions

In their study of local area network (LAN) traffic, Leland, et al., collected a remarkable set of Ethernet traffic data over time intervals as long as 27 hours, with resolution as fine as

10 ms [29]. This immense data set allowed study of traffic behavior over a range of time scales never before undertaken in network traffic study. The traffic data was represented as a time series $\{N[i]; i = 1, 2, \ldots, n\}$, where n is the number of 10ms time slots in the observation, while $N[i]$ denotes the number of packets recorded in the ith slot. The traffic data is said to be self-similar if its statistics are unchanged under different resolutions. Formally, it is required that for some parameter H, and for every $m = 1, 2, \ldots,$

$$\frac{1}{m^H}\Big(N[(i-1)m+1] + N[(i-1)m+2] + \cdots + N[im]\Big)$$

have the same covariance structure as $N[i]$, for every $m = 1, 2, \ldots$. A simple self-similar test known as the variance-time plot method then involves successive smoothing of $N[i]$, to obtain

$$N^{(m)}[i] \triangleq \frac{1}{m}\Big(N[(i-1)m+1] + N[(i-1)m+2] + \cdots + N[im]\Big), \quad m = 1, 2, \ldots, \quad (2.5)$$

followed by the estimation of the variance of $N^{(m)}[i]$ to check that

$$\mathrm{var}(N^{(m)}) \sim am^{-\beta}, \quad \text{as } m \to \infty,$$

where $\beta = 2(1 - H)$.

It is straightforward to relate the variance-time plot method of [29] is the Fano factor method of [43]. We note that from (2.4) and (2.5)

$$F(T) = \frac{m^2 \cdot \mathrm{var}(N^{(m)})}{\hat{\mu}_{N(T)}}$$

where T is m times the grid size of the finest partition. As $\hat{\mu}_{N(T)} \propto T$ for well-behaved point processes,

$$\log\left(F(T)\right) \sim \log T + \log\Big(\mathrm{var}(N^{(m)})\Big)$$

and the two tests are in essence equivalent. In addition to the variance-time plot analysis, statistical tests based on periodogram analysis are also employed in [29] to argue the self-similarity of LAN traffic.

To explain the fractal nature of LAN traffic, Willinger, et al., proposed an aggregated

on/off renewal process model [45]. An on/off renewal process is basically a two-state point process: points arrive at a fixed rate during the "on" periods, but are absent in the "off" periods. The duration of the "on" and "off" periods is each governed by a power-law density. While strictly speaking, the "on" and "off" periods are alternating, this property is not enforced in preliminary analyses. It is shown that superposition of a large number of such on/off renewal processes approaches a doubly stochastic process with a fractional Brownian motion intensity, and hence approaches the model of Johnson, et al., in [22]. The aggregated on/off renewal process model injects much insight into the fractal nature of network traffic as it suggests self-similarity of the overall effects of a large number of networking activities. However, this model again describes a point process indirectly in terms of its intensity.

Self-similarity has also been documented for wide area network (WAN) traffic. In their paper [37], Paxson, et al., thoroughly studied a variety of TCP protocols on wide-area networks, with main focus on connection and packet arrivals. Based on interarrival statistics, they concluded that while certain network activities, such as telnet and ftp connections, can be adequately modeled as inhomogeneous Poisson processes, packet traffic within each session is generally far too bursty for the Poisson process to capture. Packets due to a user in a telnet session, for example, are characterized by heavy-tailed interarrival histograms, which are decidedly different from the exponential distribution. Likewise, gaps between data sessions in a user-initiated ftp session are power-law distributed. In addition to interarrival statistics, statistical tools such as the variance-time plots were employed to argue self-similarity of the packet traffic.

Compressed Data

In addition to traffic over networks, various forms of variable-bit-rate (VBR) video data have been shown to possess self-similarity. In search of universal properties of variable-bit-rate video data, Beran, et al., studied the output of a number of different video codecs on video data of differing scenes. Using the same statistical methods as in [29], such as variance-time plot and periodogram methods, Beran, et al., concluded that arrival of variable-bit-rate video data is typically self-similar. This observation has far-reaching implications, as coded video data is expected to constitute as a main component in future integrated service networks such as B-ISDN.

2.3 A Mathematical Characterization of a Class of Fractal Point Processes

In this section, we develop a useful and sufficiently formal definition of the class of fractal point processes of central interest in this thesis. Our definition is based on a general, reasonable, broadly-accepted notion of self-similarity of point processes. Key consequences of this definition are also derived, particularly in terms of the resulting interarrival statistics. As will be apparent, while our definition is developed independently, there are important connections to the fractal point process models of [30] [32] [42] discussed earlier.

A point process is said to be *self-similar* when the associated counting process $N_X(t)$ is characterized by the following scale invariance relation

$$N_X(t) \stackrel{\mathcal{P}}{=} N_X(at) \tag{2.6}$$

for all $a > 0$, where the notation $\stackrel{\mathcal{P}}{=}$ denotes statistical equality, in particular in the sense of all finite-dimensional distributions. In essence, (2.6) is a statement that $N_X(t)$ is statistically indistinguishable from any temporally dilated or compressed version of the process—i.e., the process has no characteristic scale.

Many physical phenomena of interest exhibit no preference for a space or time origin. It is therefore natural to seek point process models possessing some stationary quality—processes whose behavior is, in some sense, independent of the time intervals in which they are viewed. Quasi-stationarity of renewal processes (see Section 2.1), therefore, advocates these processes as desirable prototypes for our self-similar point processes. However, as will become apparent in our development, no nontrivial self-similar point processes are bona fide renewal processes.

Fortunately, a weaker but still highly meaningful form of stationarity can be imposed by generalizing the notion of a renewal process. To develop this notion, we first introduce the following convenient terminology: we say that a point process with interarrivals $Y[i]$ is derived from a point process with interarrivals $X[i]$ via *conditioning on the event* $\mathcal{E}$ if $Y[i]$ is the subsequence of $X[i]$ formed by discarding those components $X[k]$ such that $X[k] \notin \mathcal{E}$. We are now ready to define our self-similar processes of interest.

Definition 1 *A self-similar point process with interarrivals $X[i]$ is said to be* conditionally-

renewing *if it satisfies the following conditions:*

1. *When conditioned on the event*

$$\mathcal{E}_1 = \{\underline{x} < X \leq \overline{x}\}$$

for some $0 < \underline{x} < \overline{x} < \infty$, *the resulting point process is a renewal process; and*

2. *When conditioned on each of any number of arbitrary, mutually exclusive events* $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_L$ *such that*

$$\mathcal{F}_l = \{\underline{x}_l < X \leq \overline{x}_l\}, \qquad 0 < \underline{x}_l < \overline{x}_l < \infty,$$

the resulting point processes are mutually independent.

For the remainder of the thesis we shall focus on self-similar point processes which are conditionally-renewing and shall refer to them as simply *fractal renewal processes.* As an immediate consequence of Definition 1, we have the following theorem, whose proof is contained in Appendix A.1.

Theorem 1 *A fractal renewal process, when conditioned on* any *event* $\mathcal{E}$ *of the form*

$$\mathcal{E} = \{x_L < X \leq x_U\},$$

where x_L *and* x_U *are arbitrary real numbers satisfying* $0 < x_L < x_U < \infty$, *yields a renewal process. Furthermore, the probability density function for the interarrival times* $Y[i]$ *of the resulting process is given by*

$$f_Y(y) = \begin{cases} \sigma_Y^2 / y^\gamma & x_L < y \leq x_U \\ 0 & \text{otherwise,} \end{cases} \tag{2.7}$$

where γ *is real and* σ_Y^2 *is a normalization factor. The power-law density of (2.7) is sometimes also referred to as the Pareto density.*

Several remarks regarding Theorem 1 are appropriate. First, we note that while Condition 1 in Definition 1 only requires *one* specific conditioning event $\mathcal{E}_1$ to yield a renewal process, Theorem 1 establishes that a renewal process is obtained when a fractal renewal

process is conditioned on *any* suitable event. As is apparent in the proof provided, this result is, in fact, a direct consequence of the statistical scale invariance of the process.

It is also important to emphasize that fractal renewal processes are not true renewal processes. Indeed, if they were renewal processes, then (2.7) implies that the common interarrival probability density for the $X[i]$ would be a power law for *all* $x > 0$. However, such a density is unnormalizable and, hence, not a valid density. Nevertheless, our terminology is convenient.

While power law probability densities on $x > 0$ are not valid densities, such behavior is, in fact, frequently observed in an extremely broad range of physical scenarios; see, e.g., [40]. Indeed, interarrival histograms generated from physical point processes often suggest power law behavior over many decades. Typically, however, finite resolution effects preclude measurement of very short interarrivals, while finite data length effects preclude measurement of very long interarrivals. Thus, Theorem 1 and the associated notion of event-conditioning provide a very natural means for interpreting these densities, and effectively captures what an observer can actually measure in practice.

The exponent γ in the density is a shape parameter that determines the skewness of the distribution and, hence, the relative frequencies of long versus short interarrivals. Not surprisingly, this parameter is directly related to the fractal dimension D of the point process, a useful measure of the extent to which arrivals cover the ambient space. In particular, we have [33]

$$D = \gamma - 1. \tag{2.8}$$

Note that as $\gamma \to 2$, $D \to 1$, in agreement with the predominance of very short interarrivals in this case. Meanwhile, as $\gamma \to 1$, we have $D \to 0$, consistent with the fact that longer interarrivals are favored.

In Fig. 2-5, we illustrate a typical sample function of a fractal renewal process with shape parameter $\gamma = 1.5$, viewed under successive magnification. For viewing convenience, the associated counting process is shown. In this example, interarrivals shorter than 1 time unit are discarded and the resulted first 50 000 points are shown. Note the hallmark scale-independent clustering behavior.

Although in practice we often observe values of γ in the range $1 < \gamma < 2$, it is important

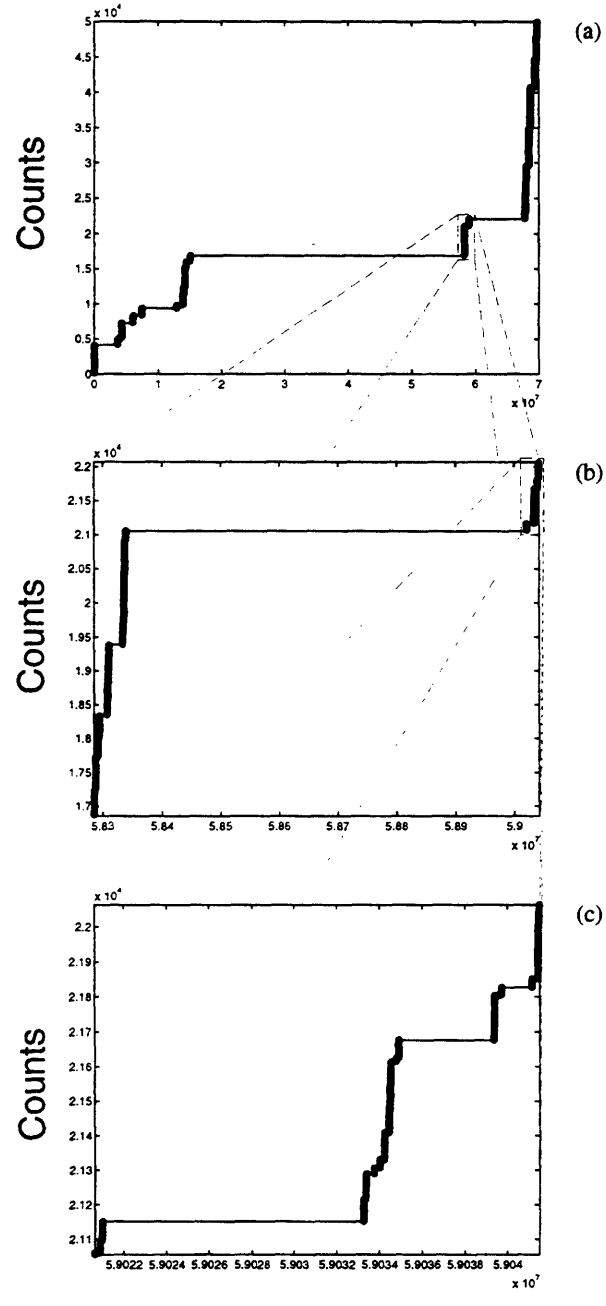


Figure 2-5: *Successive magnification of the counting process associated with a fractal renewal process; (a) the original process; (b) zoomed version of (a); (c) zoomed version of (b). Interarrivals were synthesize according to the power-law density function (2.7), with shape parameter* $\gamma = 1.5$

to note that values of γ outside this range are not uncommon. In this case, the corresponding processes are generalized point processes, and the fractal dimension (2.8) loses its direct physical interpretation. As an example, we note that the case $\gamma = 0$ corresponds to notion of uniformly distributed unbounded random variable, a form of observation dependent uncertainty that arises, for example, as a prior distribution in relating nonrandom and random parameter estimation.

Finally, we note that for $1 < \gamma < 2$, there is no preponderance of long interarrivals, and we may thus let $x_U \to \infty$ in Theorem 1 without introducing any technical difficulties. In particular, the interarrival density of the conditioned process is well-defined and normalizable—a characteristic we will sometimes exploit. Note also that this results in an interarrival density with unbounded support, as in Mandelbrot's model (see (2.3)).

2.4 Further Properties of the Fractal Renewal Process

As mentioned in Section 2.2, previous studies of fractal point processes have been largely phenomenological, driven primarily by empirical findings in various fields. In contrast, our fractal renewal process model has been axiomatically defined, rather than being motivated by any specific modeling application. We close this chapter with further analysis of the fractal renewal process, bridging our definition to the empirical results described in Section 2.2. In Section 2.3, we have already derived the signature power-law interarrival distribution of the fractal renewal process, which is well-matched with the transmission error statistics presented in [32] [42]. In this section, we address Fano factor and spectral properties of the fractal renewal process, and show that our model fulfils the criteria in [43] and [29].

Other than relating our model to previous work, we also review some existing results involving the fractal renewal process, as a preview for the work to come in this thesis. In particular, we briefly describe the counting process analysis pioneered by Sussman in [42], mainly for comparison with our work in Chapter 5. Finally, we end the chapter by highlighting, more broadly, peculiar properties of the power-law distribution, which motivates its modeling applications even beyond the context of point process theory.

2.4.1 Fano Factor Behavior

Due to the limited counting process results for the fractal renewal process, computation of the Fano factor is not straightforward. We proceed to study this statistic via simulations, using the 50 000-point sample function of Fig. 2-5. Specifically, we study the progression of the Fano factor over dyadically spaced partitions, with grid sizes

$$T = T_0 2^m, \quad m = 0, 1, \ldots, 9$$

where T_0 is the size of the finest partition. Based on each partition, we obtain the sample mean and sample variance of the corresponding counting histogram:

$$\hat{\mu}_{N(T)} = \frac{1}{n} \sum_{i=1}^{n} N[i],$$

and

$$\hat{\sigma}^2_{N(T)} = \frac{1}{n-1} \sum_{i=1}^{n} (N[i] - \hat{\mu}_{N(T)})^2,$$

where n is the number of windows in the partition, and $N[i]$ is the number of points in the ith window. The Fano factor is then computed via (2.4). The results are plotted in Fig. 2-6, alongside the dashed power-law plot included for reference. For comparison, the Fano factor of a Poisson sample function was obtained with identical methods, and the results are plotted in Fig. 2-7. From the plots, we see that the power-law growth of the Fano factor is in stark contrast to the flat graph in the Poisson case. The power-law behavior precisely matches the empirical results reported in [43], [29], [37], suggesting the viability of the fractal renewal process for modeling the corresponding phenomena.

2.4.2 Spectral Properties

Lowen, et al., contribute a set of frequency-domain results for renewal processes with truncated power-law interarrival density [30], which corresponds well with our fractal renewal process model under finite-resolution, finite-duration observations. Following the convention of (2.7), we use the symbols x_L and x_U to denote the low- and high-end truncation points, respectively. The key conclusion of the analysis of Lowen, et al., is the self-similarity of the

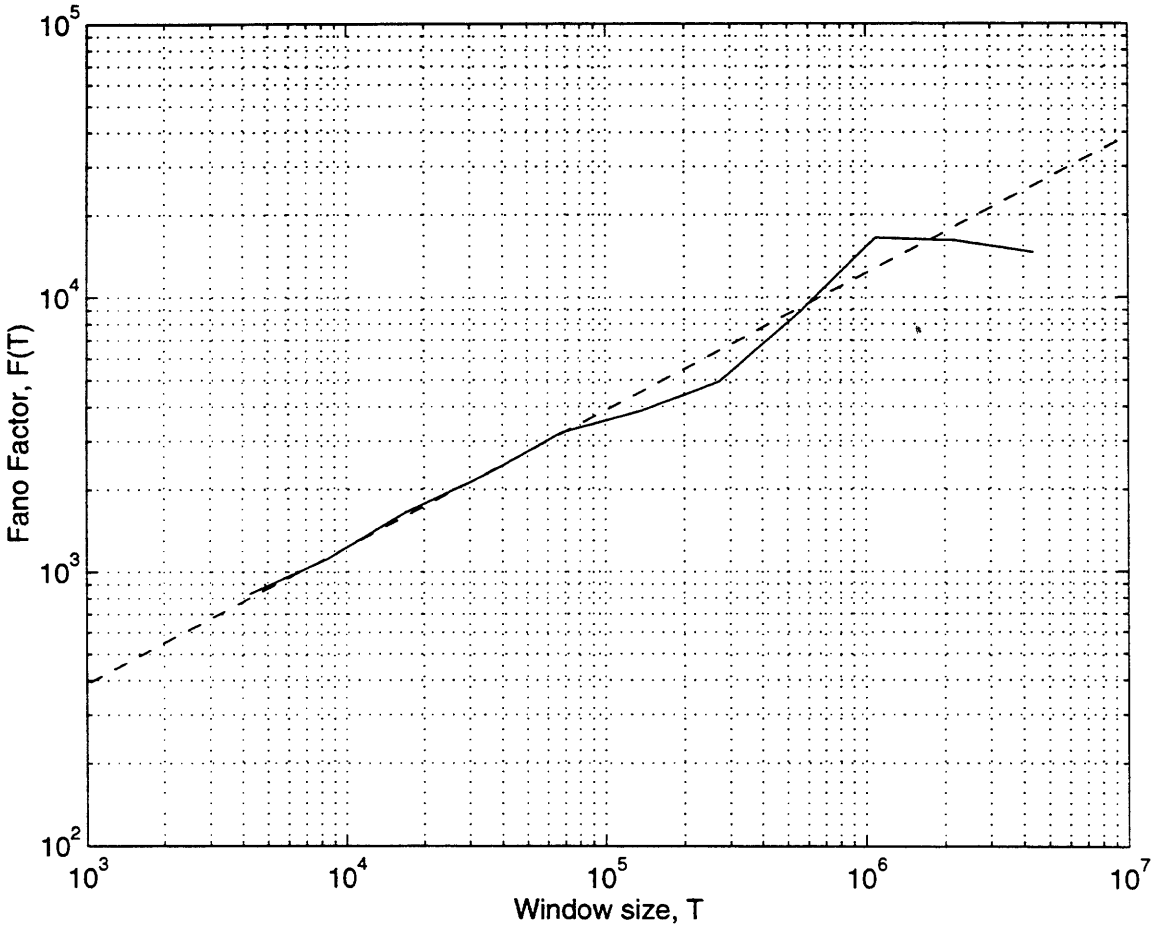


Figure 2-6: *Log-log plot of the Fano factor obtained over 10 dyadically-spaced grid sizes, for the fractal renewal process depicted in Fig. 2.5. The dashed plot represents a pure power-law function, included for comparison. The power-law behavior of the Fano factor in coarse-scale regime is in agreement with the empirical results obtained in [43] [29] [37]*

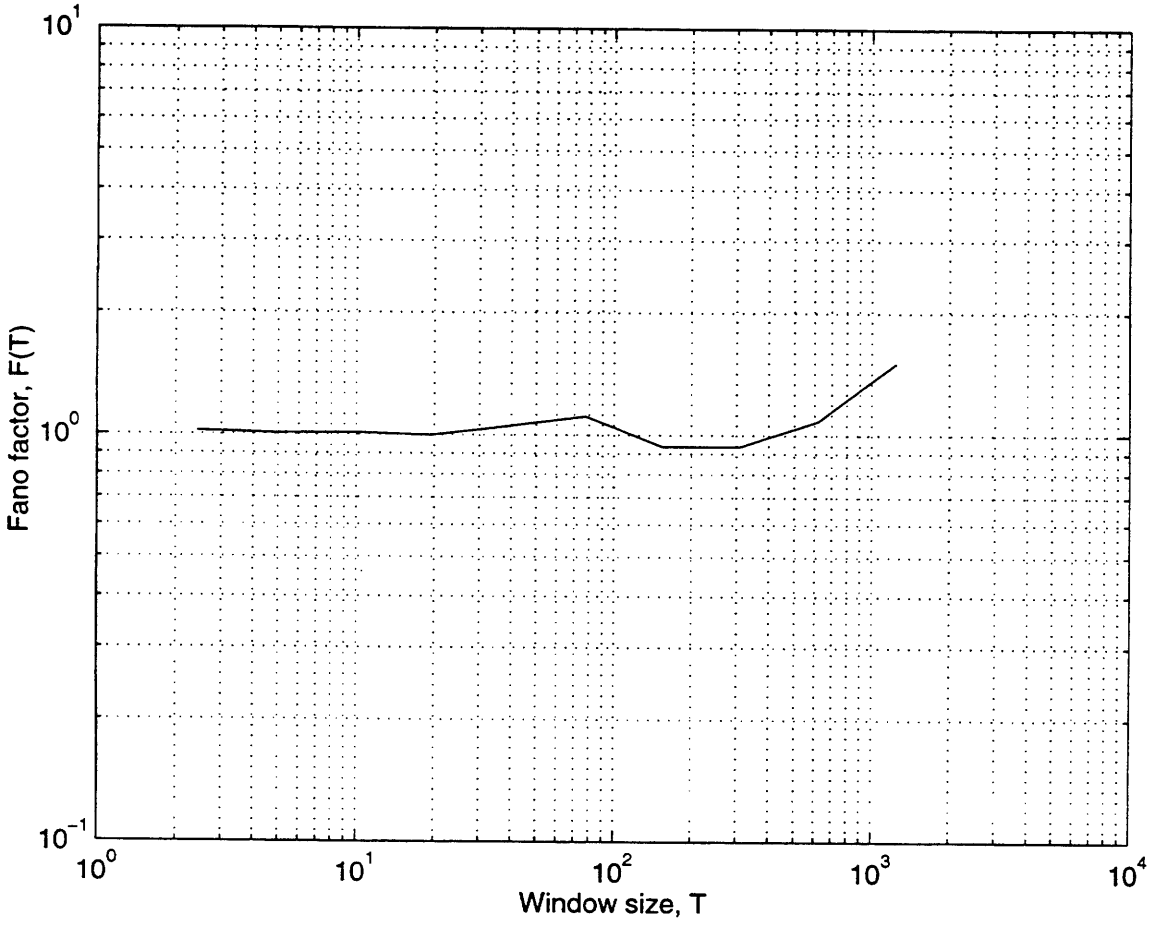


Figure 2-7: *Log-log plot of the Fano factor obtained over 10 dyadically-spaced grid sizes, for a Poisson process. The constant behavior over all grid sizes is in contrast to the fractal renewal process Fano factor depicted in Fig. 2.6.*

resulting spectrum, which suggests that the fractal renewal process has the periodogram property required in [29].

In general, the power spectral density of a stationary point process is defined as [10]

$$S_N(\omega) = \int_{-\infty}^{\infty} G_N(\tau) e^{-j\omega\tau}\, dt$$

where $G_N(\tau)$ is the coincidence rate of a stationary point process

$$G_N(\tau) = \lim_{\Delta \to 0} \frac{\Pr\{ \text{ at least 1 arrival in } (t, t+\Delta) \ \& \text{ at least 1 arrival in } (t+\tau, t+\tau+\Delta)\}}{\Delta^2}.$$

For a renewal process, it has been shown that [30]

$$S_N(\omega) = \mu^2 \delta\left(\frac{\omega}{2\pi}\right) + \mu \mathrm{Re}\left[\frac{1+\phi(\omega)}{1-\phi(\omega)}\right],$$

where $\phi(\omega)$ is the characteristic function of the interarrival density. For the case of a truncated power-law interarrival density, $\phi(\omega)$ is of the form of a power-law function [30]. Moreover, for frequencies in the range $x_U^{-1} \ll \omega \ll x_L^{-1}$, $1-\phi(\omega)$ also approximates a power law. Thus, overall, the power spectral density assumes the form of a power-law function. Finally, as another key observation, Lowen, et al., point out that this frequency-domain property is largely invariant under superposition of independent fractal renewal processes. As will be apparent, this has interesting connections with our results of Section 5.3.

2.4.3 Asymptotic Counting Process Analysis

In [42], Sussman presents an asymptotic characterization of the counting process associated with the Mandelbrot's fractal point process model of [32]. His approach mainly involves direct manipulation of the unbounded interarrival density given in (2.3). Convolving $f_X(x)$ with itself, the author first obtains a closed-form expression for the second-order interarrival density $f_{S_X[2]}(x)$. More importantly, the author characterizes its tail behavior as

$$f_{S_X[2]}(x) \approx 2\alpha x^{-\gamma}, \qquad \text{for large } x.$$

Iterating the same computations, asymptotic behavior of higher-order interarrival densities is shown to be of the form

$$f_{S_X[k]}(x) \approx k\alpha x^{-\gamma}, \quad \text{for large } x.$$

Using this set of results, the author also studies asymptotic behavior of the counting process distribution associated with the fractal renewal process. The first few terms are computed, for both the random incidence case, whereby the time origin $t = 0$ is selected randomly, independent of the point process, and the arrival-observed case, where the time $t = 0$ is an arrival. Moreover, it is demonstrated that evolution of these terms over time are asymptotically power-law functions. Specifically, for the arrival-observed case, the zeroth-order term in the distribution approaches

$$\Pr\{N_X(t) = 0\} \approx t^{-\gamma+1}, \quad \text{for large } t,$$

while the first-order term approaches

$$\Pr\{N_X(t) = 1\} \approx \Pr\{N_X(t) = 0\}, \quad \text{for large } t.$$

It is speculated that all higher-order terms have the same progression in the long run.

On the other hand, for the random incidence case, high-end truncation is first applied to the interarrival density, for otherwise, 0 arrivals will be encountered over any finite time interval with probability 1. While the computations are cumbersome, the author shows that the decay rate of zeroth-order term is much slower than higher-order terms. Specifically,

$$\Pr\{N_X(t) = 0\} \approx t^{2-\gamma}, \quad \text{for large } t,$$

while

$$\Pr\{N_X(t) = 1\} \approx t^{1-2\gamma}, \quad \text{for large } t,$$

and

$$\Pr\{N_X(t) = 2\} \approx t^{1-2\gamma}, \quad \text{for large } t.$$

This suggests the dominance of the 0th-order term.

While Sussman's direct approach yields mainly asymptotic results, it is only applicable in computation of lower-order interarrival densities, as well as lower-order terms in the counting process distribution. As we shall see in Chapter 5, the multiscale framework we develop in this thesis gives a systematic approach that allows derivation of results stronger than the asymptotics obtained in [42].

2.4.4 Peculiarities of the Power Law Distribution

As a heavy-tailed function, the power-law density exhibits a number of peculiar properties that are particularly useful for modeling certain human endeavors. As an example, this density elegantly captures a growing wait phenomenon: the longer one has waited for an event, the longer is the additional wait. This is not at all unfamiliar in many human activities, such as the completion of certain tasks: jobs either get done right away, or they will take forever.

To make this notion precise, let us consider the power-law random variable X, with complementary distribution

$$\Pr\{X > x\} = x^{-\gamma+1}, \quad x > 1,$$

for modeling interevent intervals. Given that no arrival has occurred for a duration of x_0 since the last arrival, the distribution of the total waiting time will be of the form

$$\begin{aligned}\Pr\{X \geq x_1 | X \geq x_0\} &= \frac{\Pr\{X \geq x_1, X \geq x_0\}}{\Pr\{X \geq x_0\}} \\ &= \frac{\Pr\{X \geq x_1\}}{\Pr\{X \geq x_0\}} \\ &= \frac{x_1^{-\gamma+1}}{x_0^{-\gamma+1}} = \left(\frac{x_1}{x_0}\right)^{-\gamma+1}\end{aligned}$$

for $x_1 \geq x_0$. Thus,

$$\Pr\{X \geq x_1 | X \geq x_0\} = \Pr\{X \geq kx_1 | X \geq kx_0\}$$

for every $k > 0$. In particular, for $k > 1$,

$$\begin{aligned}
&\Pr\{ \text{Remaining waiting time } \geq k(x_1 - x_0) | \text{ Having waited } x_0\} \\
&\leq \Pr\{ \text{Remaining waiting time } \geq x_1 - x_0 | \text{ Having waited } x_0\} \\
&= \Pr\{ \text{Remaining waiting time } \geq k(x_1 - x_0) | \text{ Having waited } kx_0\}.
\end{aligned}$$

Hence the remaining waiting time grows as one waits.

Finally, we remark that the heavy-tailed distribution has been desirable for modeling human usage of many resources, such as the telephone channel. Indeed, while there is a preponderence of telephone calls with very short duration, occasional extremely long phone calls are not unfamiliar. Thus, outside the context of point processes, the power-law function has long been argued as natural for server modeling in a number of queueing problems. We shall address such server behavior in our study of fractal queueing problems in Chapters 5 and 6.

To sum up, we have introduced and formally defined a class of fractal point processes in this chapter. In the next chapter, we will see that a natural multiscale representation exists for this class of fractal point processes, and will constitute as a powerful framework for the analysis and processing algorithms to come later in this thesis.

Chapter 3

Multiscale Representation and Synthesis of the Fractal Renewal Process

Given the ubiquity of self-similar point phenomena, there is a natural demand for efficient analysis and processing of fractal point processes. While practical problems involving these point processes are broad, from general signal processing such as synthesis and estimation, to specialized studies of fractal traffic over communications networks, much of the previous endeavor in these areas has been problem-specific, and the proposed techniques have often been difficult to generalize. Even in the case of the fractal renewal process, for which an explicit interarrival characterization is available, limited insight has been generated. In particular, direct manipulation of the power-law function has not appeared fruitful. The computations in [42], for example, only apply to lower-order terms in the counting process distribution, and lower-order interarrival densities. Even in these restrictive cases, the results are generally asymptotics with limited applicability. Likewise, the cumbersome spectral analysis of [30] appears unlikely to extend to other problems.

The core of this chapter is a novel framework for general studies of the fractal renewal process, which forms the basis for analysis and processing problems explored in subsequent chapters. The central idea of this framework is depicted in Fig. 3-1, and involves mainly a decomposition of the power-law function into a multiscale family of simple constituents. In this sense, our approach is reminiscent to the discrete wavelet transform analysis of $1/f$

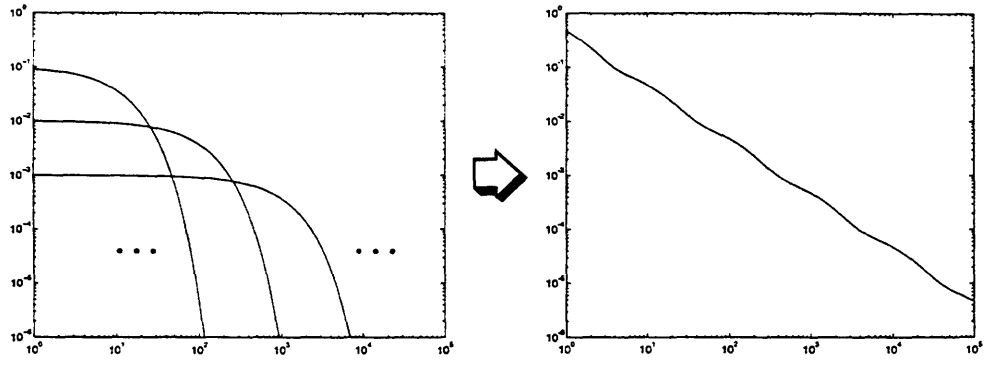


Figure 3-1: *Multiscale representation of the fractal renewal process: The power-law interarrival density is modeled as a weighted sum of a discrete family of exponential interarrival densities.*

waveforms treated in, for example, [47], as well as the infinite state space decomposition of fractal gyroscopic noise discussed in [17]. The constituents in our point process scenario are exponential density functions, which lead to a useful interpretation of the resulting process as a mixture of Poisson processes.

There are two aspects of this multiscale framework: multiscale synthesis and the complementary multiscale analysis. We begin with synthesis in this chapter, and defer the analysis to Chapter 4. A central component of our synthesis algorithms will be a Poisson process generator, for which efficient realizations are well known. Among various methods, one approach exploits the conditional independence and homogeneity of arrival times given the number of arrivals in an interval. Mathematically, for a Poisson process with arbitrary arrival rate λ, the unordered arrival times in an interval with prescribed number of arrivals are uniformly distributed

$$f_{\tilde{S}_X[1],\tilde{S}_X[2],\cdots,\tilde{S}_X[n]|N(T)=n}(s_1, s_2, \ldots, s_n) = \frac{1}{T^n} \quad n = 0, 1, \ldots,$$

where $\tilde{S}_X[i]$ is the ith unordered arrival [15]. Refined techniques for generating the Poisson random variable, such as those described in [2][3], make this an extremely efficient way to synthesize Poisson processes.

In the sequel, we present both an exact multiscale synthesis based on a continuum of constituent Poisson processes, and an approximate synthesis based on a discretization approach. Depending on applications, each may be more attractive.

3.1 Multiscale Synthesis of the Fractal Renewal Process

3.1.1 Approximate Synthesis: the Discrete Mixture

Using a Poisson process generator, we first construct a discrete collection of independent homogeneous Poisson processes via dilation and compression. Specifically, denoting the collection of counting processes by $N_{W_M}(t)$, where M is an indexing parameter, we require that

$$N_{W_M}(t) \stackrel{\mathcal{P}}{=} N_{W_0}(\rho^{-M}t), \tag{3.1}$$

where $N_{W_0}(t)$ is a prototype Poisson process with mean arrival rate λ, and, without loss of generality, the scale increment ρ is assumed greater than 1. It follows from (3.1) that the mean arrival rate of $N_{W_M}(t)$ is

$$\lambda_M = \rho^{-M}\lambda,$$

and that the associated interarrivals are statistically similar up to a scaling, i.e.,

$$W_M \stackrel{\mathcal{P}}{=} \rho^M W_0.$$

To synthesize a fractal renewal process $N_X(t)$, our algorithm involves a random mixture of these constituent Poisson processes. This is achieved via a sequence of independent, identically-distributed random variables $M[n]$ which are independent of the processes $N_{W_M}(t)$, and are distributed according to the generalized-geometric distribution

$$p_M(m) \stackrel{\Delta}{=} \Pr[M = m] = \begin{cases} \sigma_M^2 \rho^{-(\gamma-1)m} & m = \underline{m}, \underline{m}+1, \ldots, \overline{m} \\ 0 & \text{otherwise,} \end{cases} \tag{3.2}$$

where $\underline{m}$ and $\overline{m}$ denote the finest and coarsest scale indices, respectively, and

$$\sigma_M^2 = \frac{1-\rho^{-(\gamma-1)}}{\rho^{-(\gamma-1)\underline{m}}\left(1-\rho^{-(\gamma-1)(\overline{m}-\underline{m}+1)}\right)}$$

for normalization. More specifically, the random variable $M[n]$ identifies the Poisson process from which the nth arrival will be taken. Thus, the first arrival of $N_X(t)$ is chosen to be

the first arrival of the Poisson process corresponding to $M[1]$, i.e.,

$$S_X[1] = S_{W_{M[1]}}[1].$$

Next, the second arrival of $N_X(t)$ is chosen to be the first arrival occurring after $t = S_X[1]$ in the corresponding Poisson process $N_{W_{M[2]}}(t)$, i.e.,

$$S_X[2] = S_{W_{M[2]}}[k],$$

where k is the smallest index such that

$$S_{W_{M[2]}}[k] > S_X[1].$$

Subsequent arrivals are selected in a similar manner. An example of this synthesis is depicted in Fig. 3-2, where four constituents, $N_{W_1}(t)$ through $N_{W_4}(t)$, are used, and the sample sequence for $M[n]$ is:

$$2, 1, 1, 1, 2, 3, 4, 2, 1, 1, 3, 1, 1, 2, 1, 3, \ldots .$$

Consequently, $M[n]$ can be interpreted as a sequence of scale indices which determine how the constituent at each scale contributes in the synthesis. A simple architecture implementing this mixing procedure is shown in Fig. 3-3 for the case where the scale index M ranges from 0 to $L-1$. Note that the choice process is synchronized with the arrivals, and renews at every arrival.

This discrete synthesis is justified by the following theorem, a proof of which is contained in Appendix B.1.

Theorem 2 *Let us sequentially construct $N_X(t)$ from the $M[n]$ and $N_{W_M}(t)$ defined above as follows. Let the arrival times be generated iteratively as*

$$S_X[n] = \inf_{k:S_{W_{M[n]}}[k] > S_X[n-1]} S_{W_{M[n]}}[k],$$

where

$$S_X[0] = 0.$$

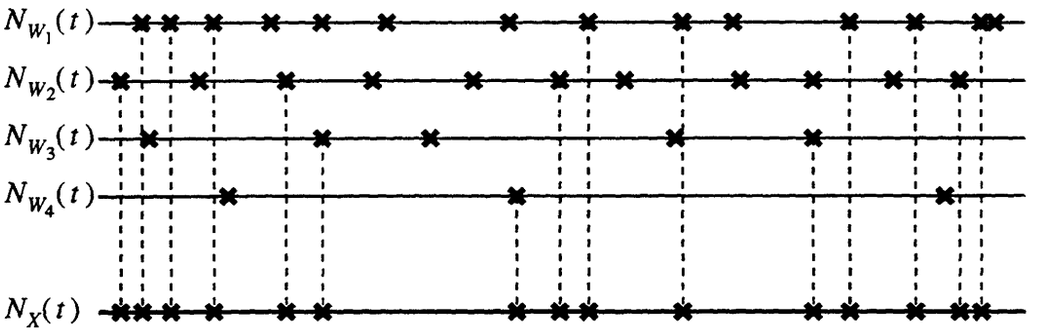


Figure 3-2: *Example of the discrete multiscale synthesis algorithm involving four constituents. $N_{W_1}(t)$ through $N_{W_4}(t)$ are independent dilations of Poisson process, while $N_X(t)$ denotes the output of the synthesis. The sample sequence of choice random variable $M[n]$ is 2, 1, 1, 1, 2, 3, 4, 2, 1, 1, 3, 1, 1, 2, 1, 3 ...*

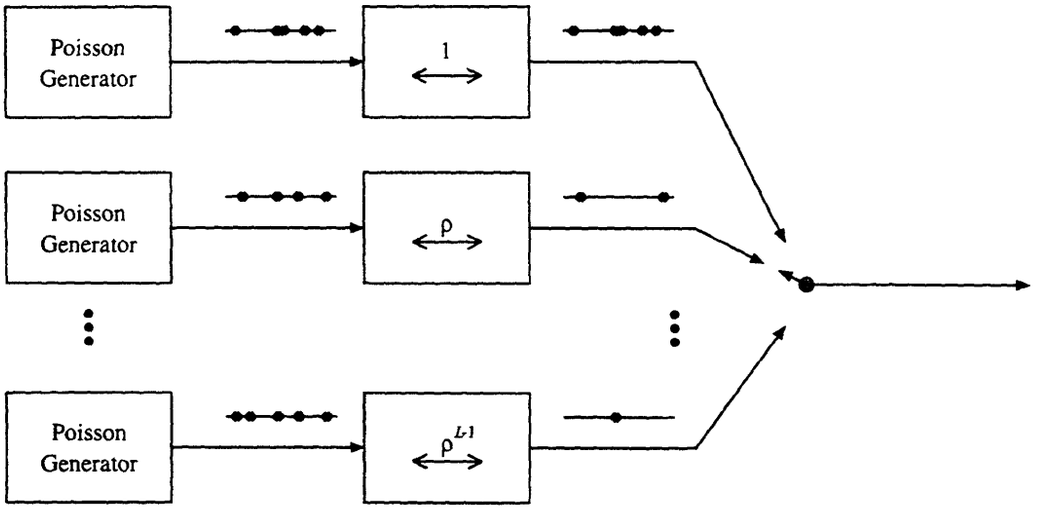


Figure 3-3: *Architecture implementing the discrete multiscale synthesis. The scale increment is ρ. The output of the Poisson generators, which are of the same rate, is dilated by the amount $\rho^0, \rho^1, \ldots, \rho^{L-1}$ by the modules in the second stage. The resulting point processes are combined by a random choice operation synchronized with the arrivals.*

Then $N_X(t)$ is a renewal process with interarrival density

$$f_X(x) = \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} \lambda \rho^{-\gamma m} \exp\left(-\lambda \rho^{-m} x\right) \tag{3.3}$$

for $x > 0$. Furthermore, this density has the property that, as $\underline{m} \to -\infty$ and $\overline{m} \to \infty$, for $\gamma > 0$, and for each $x > 0$,

$$\frac{\sigma_L^2}{x^\gamma} \leq \frac{f_X(x)}{\sigma_M^2} \leq \frac{\sigma_U^2}{x^\gamma}, \tag{3.4}$$

where σ_L^2 and σ_U^2 are constants such that $0 < \sigma_L^2 \leq \sigma_U^2 < \infty$.

A consequence of the proof of Theorem 2 is the following corollary, which gives a useful closed-form characterization of the bounding factors σ_L^2 and σ_U^2.

Corollary 1 *Let a point process be synthesized as described in Theorem 2. Then, the interarrival density satisfies (3.4) with*

$$\frac{\sigma_U^2}{\sigma_L^2} < \rho^{3\gamma}. \tag{3.5}$$

In particular,

$$\frac{\sigma_U^2}{\sigma_L^2} \to 1$$

as $\rho \to 1+$.

Several remarks regarding Theorem 2 are appropriate. First, while this theorem provides an asymptotic result involving a countably infinite collection of constituent scales, it is important to realize that in many practical cases we consider, a finite collection generally suffices to approximate any finite interarrival range of interest.

Second, we note that the selection of the scale increment ρ involves a tradeoff. On one hand, smaller values of ρ give rise to finer approximations, as suggested by Corollary 1. On the other hand, smaller values of ρ also mean that more constituent scales are required to approximate any given interarrival range of interest. Fig. 3-4 shows the interarrival density corresponding to the case $\rho = 10$ and $\gamma = 1.5$. As $\rho \to 1+$, ripple size—and, hence, approximation error—in fact decreases much more rapidly than suggested by (3.5). When

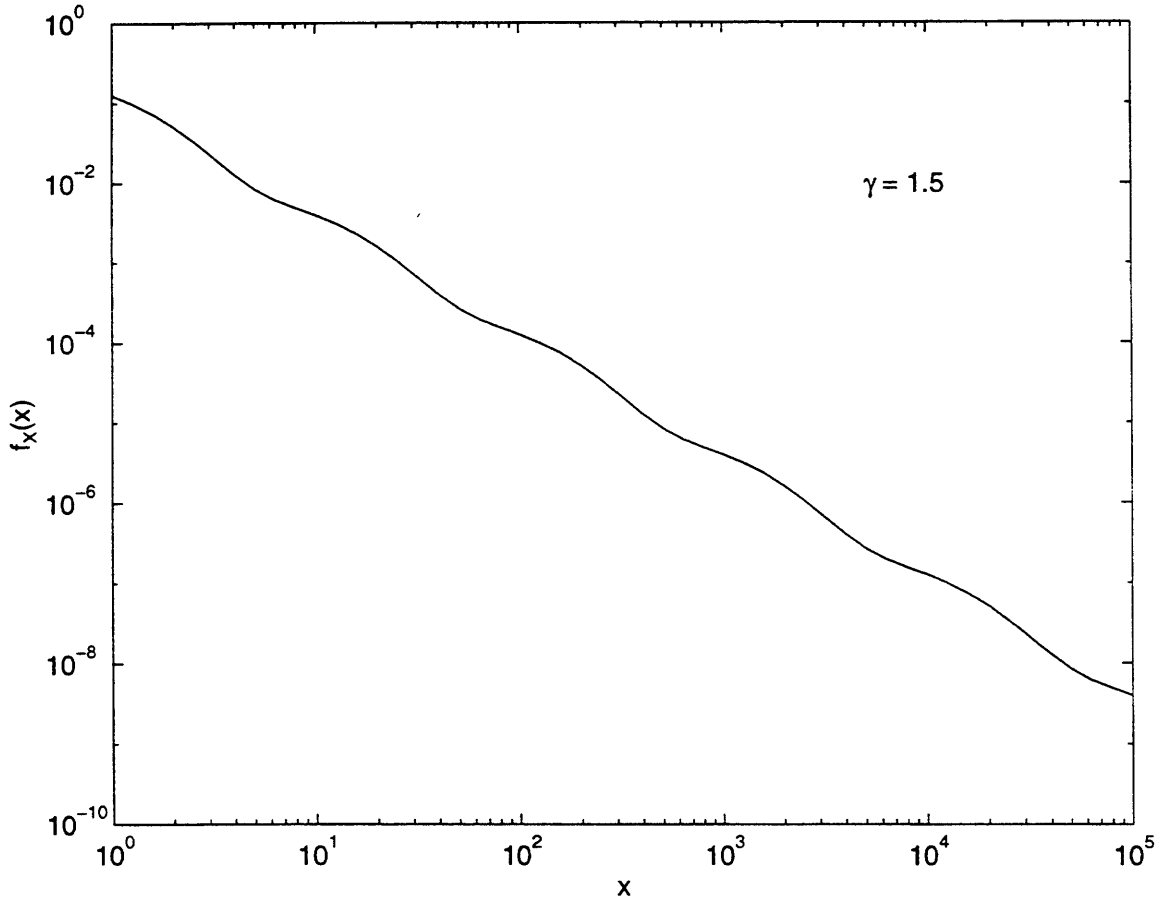


Figure 3-4: *Interarrival density of a point process constructed with the discrete multiscale synthesis algorithm. In this case, the shape parameter is* $\gamma = 1.5$, *and the scale increment is* $\rho = 10$. *On the log scale, the peak-to-peak ripple size is estimated to be* $\log 1.55$, *while the period of the ripple is* $\log \rho = \log 10$.

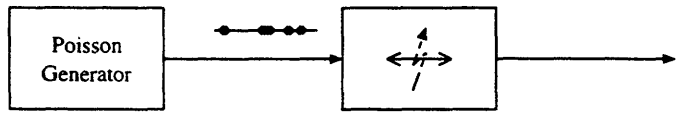


Figure 3-5: *An alternative implementation of the discrete-scale synthesis algorithm which employs a single Poisson generator. Each interarrival is independently dilated by the corresponding dilation factor $\rho^{M[n]}$, where $M[n]$ is a collection of generalized-geometric random variables.*

$\rho = 2$, for example, the approximation is already exceptionally good; numerical calculations yield $\log(\sigma_U^2/\sigma_L^2) = 3.2e{-}5$ for the case $\gamma = 1.5$!

It is also insightful to note from the proof in Appendix B.1 that, when plotted on a logarithmic scale, the approximation ripple has periodicity $\log \rho$, i.e., as $\underline{m} \to -\infty$ and $\overline{m} \to \infty$,

$$\frac{x^\gamma f_X(x)}{\sigma_M^2} = \frac{(\rho x)^\gamma f_X(\rho x)}{\sigma_M^2}.$$

This result is, of course, equivalent to observing that the limiting process is, in fact, self-similar, but not with respect to all dilations. In particular, as $\underline{m} \to -\infty$ and $\overline{m} \to \infty$, the limiting process satisfies (2.6) for all a of the form $a = \rho^m$ where m is an integer.

Finally, we note that the result (3.4) has other interpretations. For example, exploiting the memoryless property of the Poisson process, we have that two alternative but statistically equivalent constructions for $N_X(t)$ in Theorem 2 are

$$X[n] = W_{M[n]}[n]$$

and

$$X[n] = \rho^{M[n]} W_0[n]. \tag{3.6}$$

The algorithm based on (3.6) is depicted in Fig. 3-5, where independent random scaling is applied to each interarrival of a single Poisson process. As another interesting observation, this construction can be interpreted as a Poisson process in which the rate is selected randomly and independently after each arrival (and held constant between consecutive arrivals). As such, there are potentially useful connections between this model and the doubly stochastic process model of Johnson, et al. [22], discussed in Section 2.2.

3.1.2 Exact Synthesis: the Continuous Mixture

As we have seen in the previous section, the base of the discrete mixture is arbitrary. Moreover, we observe that as the scale increment ρ is varied, the approximation error can be made arbitrarily small. In this section, we show that formulation of an exact synthesis is in fact possible, and is achieved with a continuum collection of constituents.

We denote our continuum collection of constituent Poisson processes by $N_{W_A}(t)$ where A is an real-valued scale index. Again, we choose these processes to be independent and statistically similar; specifically we let

$$N_{W_A}(t) \stackrel{\mathcal{P}}{=} N_{W_0}(e^{-A}t).$$

Hence, the corresponding mean arrival rates are related according to

$$\lambda_A = e^{-A}\lambda,$$

where λ can be chosen freely.

Based on these constituents, the synthesis is carried out in essentially the same manner as the discrete mixture of Section 3.1.1, except that the sequence of scale indices $A[n]$ are now independent, identically-distributed random variables with a common generalized-exponential density

$$f_A(a) = \begin{cases} \sigma_A^2\, e^{-(\gamma-1)a} & \underline{a} \le a \le \overline{a} \\ 0 & \text{otherwise,} \end{cases} \tag{3.7}$$

where $\underline{a}$ and $\overline{a}$ denote the finest and coarsest scale indices, respectively, and

$$\sigma_A^2 = \frac{\gamma - 1}{e^{-(\gamma-1)\underline{a}} - e^{-(\gamma-1)\overline{a}}}$$

for normalization.

The following theorem, a proof of which is provided in Appendix B.2, describes the key statistical properties of this construction.

Theorem 3 *Let us sequentially construct a point process $N_X(t)$ from the $A[n]$ and the*

$N_{W_A}(t)$ defined above as follows. Let the arrival times be generated iteratively as

$$S_X[n] = \inf_{k:S_{W_{A[n]}}[k]>S_X[n-1]} S_{W_{A[n]}}[k],$$

where

$$S_X[0] = 0.$$

Then $N_X(t)$ is a renewal process whose interarrival density has the property that, as $\underline{a} \to -\infty$ and $\overline{a} \to \infty$ and for each $x > 0$,

$$\frac{f_X(x)}{\sigma_A^2} \to \frac{\sigma_0^2}{x^\gamma}, \tag{3.8}$$

where

$$\sigma_0^2 = \frac{\Gamma(\gamma-1)}{\lambda^{\gamma-1}}. \tag{3.9}$$

Several remarks are appropriate. First, we observe that the limit (3.8) is independent of λ, the rate of the prototype process. However, not surprisingly, the convergence is not uniform; specifically, (3.8) with (3.9) is a good approximation to the interarrival density for values of x satisfying

$$\frac{e^{\underline{a}}}{\lambda} \ll x \ll \frac{e^{\overline{a}}}{\lambda}, \tag{3.10}$$

as is apparent from the derivation in Appendix B.2. Fig. 3-6 illustrates the interarrival density arising from a continuous multiscale synthesis for the case $\gamma = 1.5$. The scales employed range from $\underline{a} = 0.1$ to $\overline{a} = 10$, while the rate of the prototype process is $\lambda = 1$. As suggested by the figure, exact power-law behavior can be achieved over a finite range of length scales for finite $\underline{a}$ and $\overline{a}$.

Finally, we again note that a statistically equivalent but more computationally efficient synthesis results by exploiting the memoryless property of Poisson processes. In particular, it suffices to randomly stretch each interarrival of a prototype Poisson process; i.e., analogous

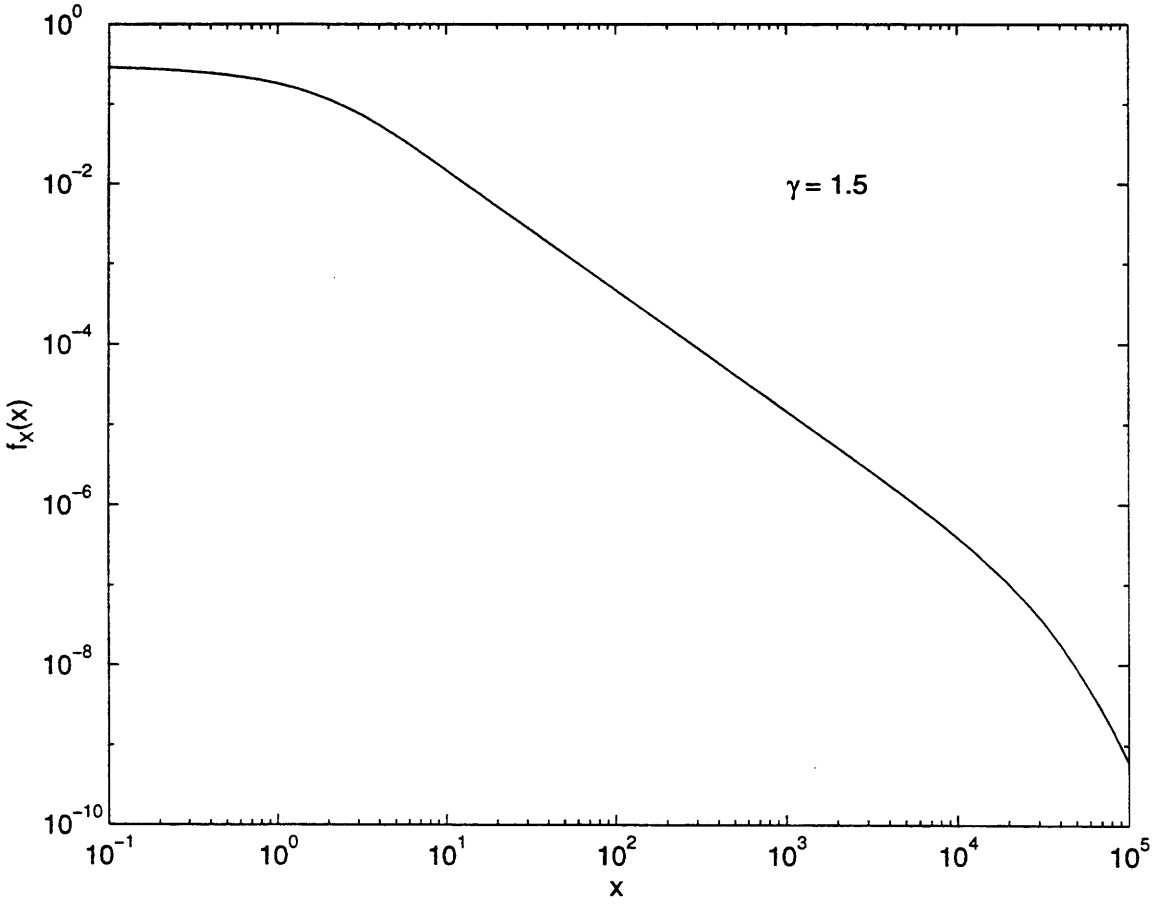


Figure 3-6: *Interarrival density of a point process constructed with the continuous-scale synthesis algorithm. In this case, the shape parameter is* $\gamma = 1.5$, *and the scales employed range from* $\underline{a} = 0.1$ *to* $\overline{a} = 10$. *Exact power-law behavior is observed over a range of near 3 decades, from 1e-1 to 1e-4.*

to (3.6), interarrivals of the form

$$X[n] = e^{A[n]} W_0[n] \tag{3.11}$$

are generated, where the $A[n]$ are the random sequence of scale indices and the $W_0[n]$ are the interarrivals of the prototype process. Note that while the characteristic feature of fractal renewal processes is their power law interarrival densities, (3.6) suggests a synthesis requiring only exponential random variables which, in practice, can be derived from a single Poisson process.

The synthesis results of this chapter, as summarized in Theorems 2 and 3 lay down the foundation of our multiscale framework for fractal renewal processes. In the following chapters, we shall explore the application of this framework, particularly the discrete-scale version, to various analysis problems such as estimation of these point processes as well as system and network analysis involving these signals.

Chapter 4

Multiscale Analysis and Estimation of the Fractal Renewal Process

In this chapter, we investigate the multiscale analysis relation complementary to the synthesis of Chapter 3. While the proposed synthesis is clearly not reversible, as will be apparent, meaningful analysis relations exist, particularly in the form of recovery of key aspects of the synthesis. In fact, we show in this chapter that this is often tractable even under adverse conditions. Of particular importance are the mixing probabilities, which essentially govern the synthesis process. Naturally, knowledge of these probabilities is equivalent to knowing the shape parameter of the interarrival density γ, and the fractal dimension of the point process D. In fact, we shall formulate this problem in terms of an estimation of γ. Contrary to existing methods for this purpose, our multiscale parameter estimator is robust under certain noise distortion. Our multiscale framework also leads to natural algorithms for interarrival estimation, which have importance in signal reconstruction and robust data acquisition in general. While the parameter and interarrival estimation will be dealt with separately, we shall see that these problems are actually closely coupled.

4.1 Robust Parameter Estimation for the Fractal Renewal Process

The need for robust methods for estimating the fractal dimension of fractal point processes arises in a host of applications. As discussed in Chapter 2, this parameter provides an

intrinsic measure of key properties of a self-similar point process, such as the space-filling efficiency. In many contexts, additional information can also be drawn from this parameter. For example, the fractal dimension of an auditory neural pulse train is useful for characterizing external stimuli [43]. Furthermore, knowledge of the fractal dimension facilitates many other detection and estimation problems involving self-similar point processes; one such problem will be the theme of Section 4.2.

In Chapter 2, we saw how the fractal dimension is manifested in key statistics of a self-similar point process, such as the Fano factor and periodogram. Various fractal dimension estimators have been developed based on these statistics and applied to network traffic data, variable-bit-rate video codec outputs, and neural activities [43] [29] [37] [5]. As these techniques are all formulated in terms of the intensity $\lambda(t)$, they are particularly well-suited to the doubly stochastic process framework of, say, [22].

For our fractal renewal process model, however, it is most natural to address this parameter estimation in terms of interarrival observations. More formally, we represent our data as a collection of N interarrival measurements $\{R[n]; n = 1, 2, \ldots, N\}$ taken from a fractal point process,

$$R[n] = X[n] + W[n], \quad n = 1, 2, \ldots, N, \tag{4.1}$$

where $\{X[n]; n = 1, 2, \ldots, N\}$ are the interarrival times and $\{W[n]; n = 1, 2, \ldots, N\}$ represent some form of additive distortion explicitly included for robustness of the estimation.

In Chapter 2, we showed that the interarrival density $f_X(x)$ assumes the form of a power-law function, with shape parameter γ directly related to the fractal dimension of the point process (see (2.8)). Thus, the fractal dimension estimation problem can be recast into a parameter estimation problem involving a power-law probability law. This is a familiar problem that has seen some prior attention in statistics, largely due to the prominence of the power-law distribution in engineering, the natural and social sciences. In general, this estimation problem is rather challenging due to the pathological nature of the power-law distribution, and many well-known, conventional estimation techniques have appeared inadequate for its solution. For example, while the method-of-moments and probability-weighted-moments estimators described by Hosking, et al., in [18] are useful in situations where $\gamma > 2$, they are valueless for our purposes since they require bounded moments,

unavailable in our case. To worsen the problem, substantial likelihood of outliers can be expected to decelerate convergence of the resulting estimators.

While a number of algorithms have been proposed for this problem, existing methods invariably assume noise-free data. One of these is the straightforward maximum-likelihood (ML) estimator due to Malik [31]. A two-parameter truncated power-law distribution is assumed, which is of the form

$$f_X(x) = \begin{cases} (\gamma - 1)a^{\gamma-1}x^{-\gamma} & x \geq a \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where a is the low-end cutoff, and γ is as usual the shape parameter. Malik models the two parameters as unknown, and develops estimation techniques for both. We denote the resulting estimators as $\hat{a}_{\text{Malik}}$ and $\hat{\gamma}_{\text{Malik}}$, respectively.

Using (4.2), the independence of interarrivals, and the absence of distortion, the likelihood function based on the observations $R[n]$ is

$$L(a, \gamma) = \prod_{n=1}^{N} f_R(r[n]; a, \gamma) = \prod_{n=1}^{N} (\gamma - 1)a^{\gamma-1}(r[n])^{-\gamma}, \tag{4.3}$$

which is defined for every $\gamma > 1$ and $a \leq r[1], r[2], \ldots, r[N]$. A powerful result of [31] is the decoupled computation of $\hat{a}_{\text{Malik}}$ and $\hat{\gamma}_{\text{Malik}}$. First, regardless of the true value of γ, the ML estimate of a is simply

$$\hat{a}_{\text{Malik}} = \min\{r[1], r[2], \ldots, r[N]\}. \tag{4.4}$$

This follows from the fact that the $a^{(\gamma-1)N}$ term in (4.3) is monotonically increasing in a.

Next, the ML estimate for γ can be obtained via differentiation of (4.3). It is straightforward to check that the resulting estimate is

$$\hat{\gamma}_{\text{Malik}} = \left[\frac{1}{N} \sum_{i=1}^{N} \ln \frac{r[i]}{a} \right]^{-1} + 1. \tag{4.5}$$

Malik's procedure for estimating γ therefore consists of the recovery of a using (4.4), followed by the estimation of γ based on (4.5).

In fact, as shown by Kern [23], a unique minimum variance unbiased (MVU) estimator

exists for this problem, and is closely related to Malik's maximum-likelihood estimator. The same truncated power-law distribution (4.2) is assumed for the random variables $X[n]$, and the data is again assumed noise-free. In this case, however, the parameter a is assumed known. In principle, this can be estimated according to (4.4). Although Kern's derivation of the MVU estimator is fundamentally different, distinction between the estimates is minimal:

$$\hat{\gamma}_{\text{Kern}} = \frac{N-1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\ln\frac{r[i]}{a}\right]^{-1} + 1$$

$$= \frac{N-1}{N}\,\hat{\gamma}_{\text{Malik}} + \frac{1}{N}. \tag{4.6}$$

Thus, the MVU technique also yields a convenient closed-form parameter estimate. Note also that Eq. (4.6), together with the fact that the MVU estimator is unbiased, implies that $\hat{\gamma}_{\text{Malik}}$ is asymptotically unbiased.

Hosking, et al., describe a maximum-likelihood estimation algorithm for the shape parameter based on the alternative generalized Pareto probability density function [18]

$$f_X(x) = \frac{1}{\alpha}\left(1 + \frac{x}{\alpha(\gamma-1)}\right)^{-\gamma}, \quad x > 0. \tag{4.7}$$

The shape parameter γ is assumed greater than 1. In contrast to the probability law (4.2), the generalized Pareto distribution does not suffer from an abrupt low-end cutoff, although it still adequately captures the persistent power-law decay. In particular,

$$f_X(x) \to \frac{1}{\alpha}, \quad \text{as } x \to 0+,$$

while

$$f_X(x) \to \frac{(\gamma-1)^{\gamma}}{\alpha^{1-\gamma}}x^{-\gamma}, \quad \text{for } x \gg \alpha(\gamma-1).$$

However, the likelihood function in this case cannot be maximized in closed form. Indeed,

the log-likelihood function assumes the form [18]

$$\ell(\alpha,\gamma) = \ln L(\alpha,\gamma) = \ln \prod_{n=1}^{N} f_R(r[n];\alpha,\gamma)$$
$$= -N\ln\alpha - \gamma\sum_{n=1}^{N}\ln\left(1+\frac{r[n]}{\alpha(\gamma-1)}\right). \tag{4.8}$$

Hosking, et al., propose a Newton-Raphson iteration technique for the numerical maximization of (4.8).

In contrast to the aforementioned approaches, we address the more general problem in which the observations are noise-corrupted. Thus, we adhere to the model (4.1) with the $W[n]$ nontrivial. We restrict our attention to the case where the $W[n]$ are both mutually independent and independent of the interarrivals $X[n]$, and are identically distributed according to an exponential probability density function

$$f_W(w) = \begin{cases} \alpha e^{-\alpha w} & w \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

with $\alpha > 0$. It is well known that such a random variable has second moment $2/\alpha^2$, which can be interpreted as a measure of noise strength. This noise component can be used for modeling a variety of natural effects which arise in applications, such as a random processing delay in an interarrival measurement transducer, as well as physical width of non-ideal spikes.

We remark that the maximum-likelihood estimators of Malik [31] and Hosking [18] cannot be usefully extended for our scenario of noisy observations. Essentially the convolution of a power law density and an exponential density, the likelihood function in these cases would be hard to manipulate. While our approach to this problem is also based on the method of maximum-likelihood, we exploit the result of Section 3.1.1 and model the $X[n]$ with a finite-scale representation. For consistency, we shall follow the notational convention of Section 3.1.1 throughout the development of our estimator. In addition, we introduce, for convenience, a new parameter q defined as

$$q = \rho^{1-\gamma},$$

keeping in mind that the ML estimates of the shape parameter γ and the fractal dimension D can, in turn, be obtained from the resulting ML estimate of q via

$$\begin{aligned} \hat{\gamma}_{\mathrm{ML}} &= 1 - \ln \hat{q}_{\mathrm{ML}} / \ln \rho \\ \hat{D}_{\mathrm{ML}} &= -\ln \hat{q}_{\mathrm{ML}} / \ln \rho. \end{aligned}$$

In addition to γ, the parameters λ and α are generally unknown *a priori*, and need to be estimated as well. Consequently, we represent the collection of parameters to be jointly estimated with the vector $\Theta = (\lambda, \alpha, q)^T$.

Based on the multiscale model, the log-likelihood function of the data can be computed in a straightforward manner, yielding

$$\ell(\Theta) = \sum_{n=1}^{N} \ln f_R(r[n]; \Theta), \tag{4.10}$$

where using (3.3) and (4.9) we have, for $r > 0$,

$$\begin{aligned} f_R(r; \Theta) &= \int_0^r f_X(\tau)\, f_W(r-\tau)\, d\tau \\ &= \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} q^m \int_0^r f_{X|M}(\tau|m)\, f_W(r-\tau)\, d\tau \\ &= \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} q^m f_{R|M}(r \mid m; \Theta), \end{aligned} \tag{4.11}$$

with

$$f_{R|M}(r \mid m; \Theta) = \begin{cases} \dfrac{\lambda_m \alpha}{\lambda_m - \alpha} \left[e^{-\alpha r} - e^{-\lambda_m r} \right] & \lambda_m \neq \alpha \\ \\ \alpha^2 r e^{-\alpha r} & \text{otherwise,} \end{cases} \tag{4.12}$$

where, as before, $\lambda_m = \rho^{-m}\lambda$. Without loss of generality we may set $\underline{m} = 0$, since λ may be scaled accordingly. The total number of scales required, which we denote by L, is typically determined from the spread of the data. As will become apparent, overestimating L generally does not affect the estimation performance, though the corresponding algorithm is less efficient in terms of both computation and storage.

Calculation of the ML parameter estimates by directly maximizing (4.10) with (4.11) and (4.12) is difficult. However, these estimates can be efficiently computed using an iterative Estimate-Maximize (EM) algorithm [27]. As the parameter estimates are updated every iteration, we shall use the notation

$$\hat{\Theta}_{[k]} = (\hat{\lambda}_{[k]}, \hat{\alpha}_{[k]}, \hat{q}_{[k]})^T$$

to denote the estimates obtained at iteration k. Also, for convenience, we shall use the shorthand notation $\hat{\lambda}_{m[k]}$ for $\rho^{-m}\hat{\lambda}_{[k]}$.

Each iteration of the resulting estimation algorithm, a detailed derivation of which is presented in Appendix C.1.1, consists of two steps.

E-Step Using the current set of parameter estimates, estimate for every valid pair of m and n, the probability that interarrival $x[n]$ was derived from scale m given the observation $r[n]$, i.e.,

$$p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]}) = \begin{cases} \sigma^2_{M|R} \dfrac{\hat{\lambda}_{m[k]}\hat{\alpha}_{[k]}}{\hat{\lambda}_{m[k]} - \hat{\alpha}_{[k]}} \left[e^{-\hat{\alpha}_{[k]}r[n]} - e^{-\hat{\lambda}_{m[k]}r[n]}\right] \dfrac{1-\hat{q}_{[k]}}{1-\hat{q}^L_{[k]}} \hat{q}^{m-1}_{[k]} & \hat{\lambda}_{m[k]} \neq \hat{\alpha}_{[k]} \\ \tilde{\sigma}^2_{M|R} \hat{\alpha}^2_{[k]} r[n] e^{-\hat{\alpha}_{[k]}r[n]} \dfrac{1-\hat{q}_{[k]}}{1-\hat{q}^L_{[k]}} \hat{q}^{m-1}_{[k]} & \text{otherwise,} \end{cases} \tag{4.13}$$

where $\sigma^2_{M|R}$ and $\tilde{\sigma}^2_{M|R}$ are normalization terms.

Also, for every valid pair of m and n, compute the conditional expected interarrival size $X[n]$, given the observation $r[n]$ and the scale of the interarrival m, i.e.,

$$E\left[X[n] \,\middle|\, r[n], m; \hat{\Theta}\right] = \begin{cases} r[n]\left[\dfrac{1}{(\hat{\lambda}_m - \hat{\alpha})r[n]} - \dfrac{1}{e^{(\hat{\lambda}_m - \hat{\alpha})r[n]} - 1}\right] & \hat{\lambda}_m \neq \hat{\alpha} \\ r[n]/2 & \text{otherwise.} \end{cases} \tag{4.14}$$

M-Step Using the preceding tables of probability and interarrival estimates as weights,

compute new estimates of the parameters via

$$\frac{1}{\hat{\lambda}_{[k+1]}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]})\, \rho^{-m}\, E\left[X[n] \,\middle|\, r[n], m; \hat{\Theta}_{[k]}\right] \quad (4.15a)$$

$$\frac{1}{\hat{\alpha}_{[k+1]}} = \frac{1}{N}\sum_{n=1}^{N}\left(r[n] - \sum_{m=1}^{L} p_{M|R}(m|r[n]; \hat{\Theta}_{[k]}) E\left[X[n] \,\middle|\, r[n], m; \hat{\Theta}_{[k]}\right]\right) \quad (4.15b)$$

$$\frac{1}{1-\hat{q}_{[k+1]}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]})\, m. \quad (4.15c)$$

It can be readily verified that straightforward variants of this algorithm apply when some of the parameters are known *a priori*. In particular, estimates of any known parameters in the algorithm are replaced with their true values in both the E- and M-steps, and the corresponding parameter update in the M-step is omitted. As would be expected, the convergence rate of the algorithm generally improves when some of the parameter values are known. The important special case of noise-free estimation will be the subject of Section 4.1.1.

It is useful to note that interarrival time estimation is an integral part of the parameter estimation process. For example, minimum mean-square error estimates of $X[n]$ based on the current parameter estimates and $r[n]$ are, in fact, constructed in updating $\hat{\alpha}$ in (4.15b). This becomes more apparent when we explore the interarrival time estimation problem in Section 4.2.

4.1.1 Noise-Free case: $W[n] \equiv 0$

In the case where the distortion terms $W[n]$ are *a priori* known to be zero, our EM estimation algorithm simplifies substantially. In particular, as we have one fewer parameter to estimate, the M-steps will only consist of update equations for λ and q. In addition, since we are now observing $X[n]$ directly, a number of terms, such as $E\left[X[n] \,\middle|\, r[n], m; \hat{\Theta}\right]$ simplify, or even become trivial. While the formal derivation of the simplified algorithm is developed in Appendix C.1.2, the key steps in the algorithm are

E-Step Using the current set of parameter estimates, estimate for every valid pair of m and n, the probability that interarrival $x[n]$ was derived from scale m given the clean

observation of $x[n]$, i.e.,

$$p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]}) = \sigma^2_{M|R} \hat{\lambda}_{m[k]} e^{-\hat{\lambda}_{m[k]} r[n]} \frac{1 - \hat{q}_{[k]}}{1 - \hat{q}^L_{[k]}} \hat{q}^{m-1}_{[k]}$$

where $\sigma^2_{M|R}$ is a normalization term. Note that the terms $E\left[X[n] \,\middle|\, r[n], m; \hat{\Theta}\right]$ now become trivial.

M-Step Using the preceding table of probability estimates as weights, compute new estimates of the parameters via

$$\frac{1}{\hat{\lambda}_{[k+1]}} = \frac{1}{N} \sum_{n=1}^{N} \sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]})\, \rho^{-m}\, r[n] \tag{4.16a}$$

$$\frac{1}{1 - \hat{q}_{[k+1]}} = \frac{1}{N} \sum_{n=1}^{N} \sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]})\, m. \tag{4.16b}$$

Properties of the Estimator

Like all EM algorithms, our estimator leads to a sequence of parameter estimates that, from an arbitrary set of initial estimates, ascend the likelihood function (4.10) in a nondecreasing manner [27]. In this particular application, empirical evidence suggests that over a non-pathological portion of parameter space the likelihood function is unimodal—i.e., there is a unique local maximum that is the global maximum. Thus, the algorithm is certain to converge to the ML estimates, although the rate of convergence is affected in part by the initial estimates. As with most EM algorithms, the convergence can often be accelerated by judiciously replacing EM iterations with Newton-Raphson iterations when in a neighborhood of the ML estimates.

Analytic performance bounds for our multiscale parameter estimator are in general difficult to obtain. Related to this problem is the computation of performance bounds for multiscale estimators for $1/f$ processes, based on the wavelet transform [46]. The techniques developed by Ninness in [36] can likely be adapted for our interarrival estimator. This remains an open problem for future investigation. On the other hand, in the special case of noise-free observations the Cramér-Rao bounds can be readily deduced using the exact

power-law density (2.7). In this case, we have, in particular,

$$\operatorname{var} \hat{\gamma} \geq (\gamma - 1)^2 / N. \tag{4.17}$$

Although this bound is somewhat loose, empirical studies have shown that in accordance with (4.17), variance in our estimate of γ generally decays for increasing sample size N.

To further investigate the performance of our parameter estimator, we have conducted a series of Monte Carlo tests on simulated data. To model interarrivals invariant under all dilations, we adhere to the interarrival distribution of the form (2.7), with $x_U \to \infty$. This coincides exactly with the probability law assumed in Malik's and Kern's problems. Thus, we expect the corresponding estimators to be optimal, at least in the noise-free case. However, as will be apparent, our multiscale EM estimator exhibits the most robustness under noise.

To ensure that modeling error effects were included in the simulations, both data and noise terms were synthesized via transformation of uniform random variables. Specifically, the power-law random variable X with probability law (2.7) was synthesized according to

$$X = x_L U^{1/(1-\gamma)},$$

where U is a uniform random variable distributed between 0 and 1, and x_L is a low-end cutoff. Using similar techniques, the exponential noise term W with probability law (4.9) was synthesized according to the prescription

$$W = -\frac{1}{\alpha} \ln U.$$

Fig. 4-1 summarizes the performance of the noise-free γ estimator of Section 4.1.1 for various sample size N and true shape parameter γ. Specifically, the plots show the bias in the dyadic ($\rho = 2$) multiscale estimator for the cases $\gamma = 1.8, 1.5, 1.2$, computed over 512 Monte Carlo trials. The errorbars denote the 1-standard-deviation confidence level in the simulation results. In general, the number of scales L can be set according to the spread in the data, and for our experiment,

$$L = \log_2 \frac{\max\{r[1], r[2], \ldots, r[N]\}}{\min\{r[1], r[2], \ldots, r[N]\}} + 10.$$

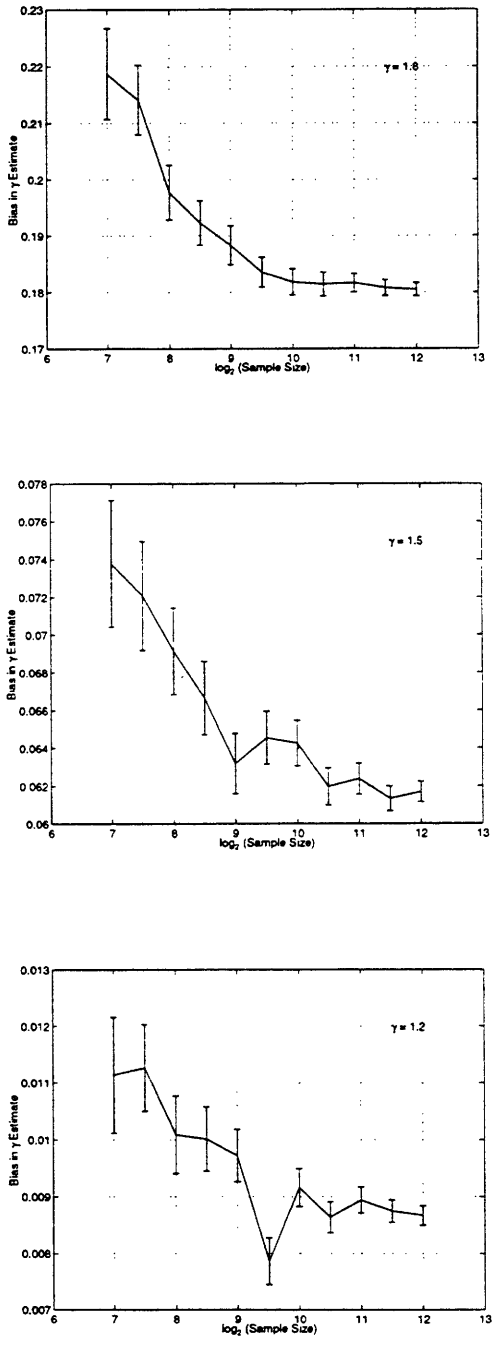


Figure 4-1: *Performance of the multiscale EM parameter estimator for noise-free fractal renewal process based on various numbers of interarrival measurements N. Each data point was computed from 512 Monte Carlo trials on simulated data. The plots correspond to the bias in the estimator for the cases $\gamma = 1.8, 1.5, 1.2$, with errorbars denoting 1-standard-deviation spread in the estimation results.*

Table 4.1: *Limiting bias in the noise-free EM estimator and Hosking's estimator for* γ

True Value	Limiting Bias	
	EM estimator	Hosking's estimator
1.8	+0.1790	+0.6018
1.5	+0.0612	+0.1878
1.2	+0.0074	+0.0244

To explore robustness and generality, the implementation is initialized with the estimates

$$\hat{\lambda}_{[1]} = 10/\min\{r[1], r[2], \ldots, r[N]\}$$
$$\hat{\gamma}_{[1]} = 1.1$$

throughout. The termination condition for the iteration is given in terms of the change in the γ estimate:

$$|\hat{\gamma}_{[k]} - \hat{\gamma}_{[k-1]}| < \epsilon = 1 \times 10^{-4}.$$

From Fig. 4-1, it is apparent that there exists bias in the estimates. This comes as a manifestation of the discrepancy between the exact power-law data and discrete-scale algorithm. The limiting bias of $\hat{\gamma}_{\mathrm{ML}}$ computed from the estimate corresponding to $N = 2^{12}$, are tabulated in Table 4.1, alongside the true parameter values. It is seen from the table that the estimate bias increases with γ. This is in agreement with degradation of our multiscale model in the high γ regime, as suggested by (B.8). We note also that the decay in the errorbars suggests promising convergence properties of the estimator, as the standard deviation is reduced monotonically with increasing sample size. Thus, for large sample size, estimation error will be primarily due to bias.

For comparison, Monte Carlo studies of the estimators of Malik [31], Kern [23], and Hosking [18] have also been performed, and the results are plotted in Figs. 4-2, 4-3, and 4-4 respectively. As the first two estimators are formulated based on the exact model, it is not surprising that little bias is observed in these cases. At the same time, the estimator performance improves as more data is used, as suggested by the standard deviation progression in both cases.

Our implementation of the estimator described by Hosking, et al., uses a simplex search

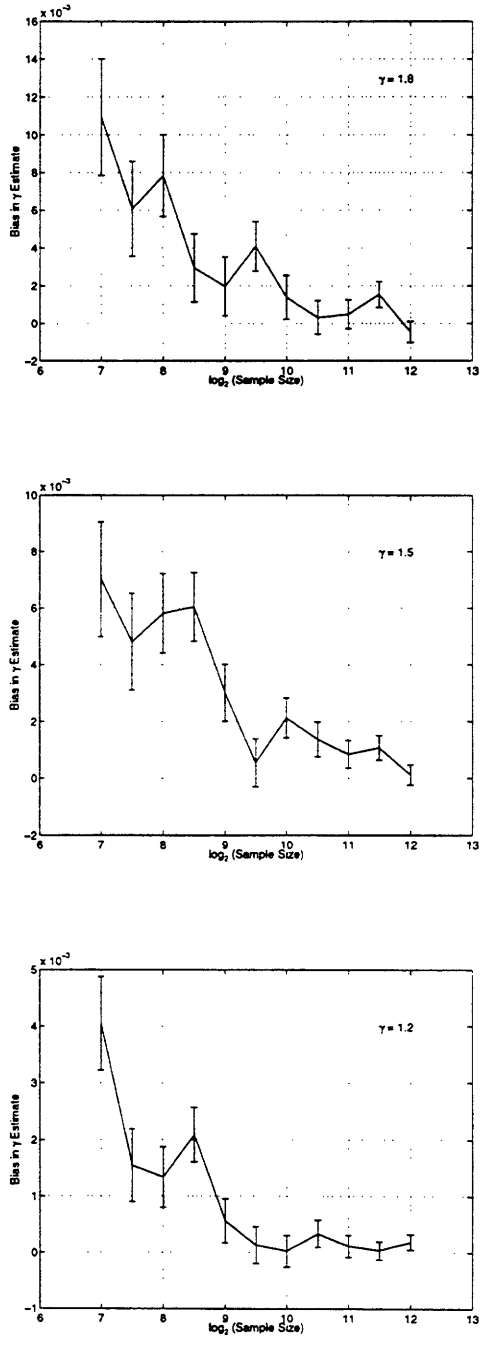


Figure 4-2: *Performance of Malik's parameter estimator for noise-free fractal renewal process based on various numbers of interarrival measurements N. Each data point was computed from 512 Monte Carlo trials on simulated data. The plots correspond to the bias in the estimator for the cases $\gamma = 1.8, 1.5, 1.2$, with errorbars denoting 1-standard-deviation spread in the estimation results.*

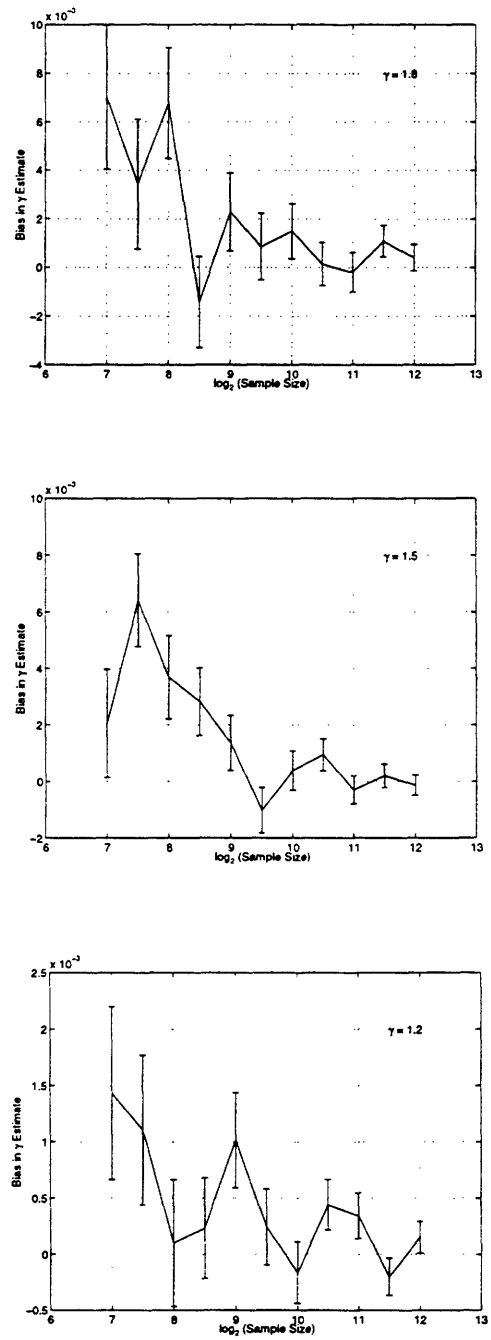


Figure 4-3: *Performance of Kern's parameter estimator for noise-free fractal renewal process based on various numbers of interarrival measurements N. Each data point was computed from 512 Monte Carlo trials on simulated data. The plots correspond to the bias in the estimator for the cases* $\gamma = 1.8, 1.5, 1.2$, *with errorbars denoting 1-standard-deviation spread in the estimation results.*

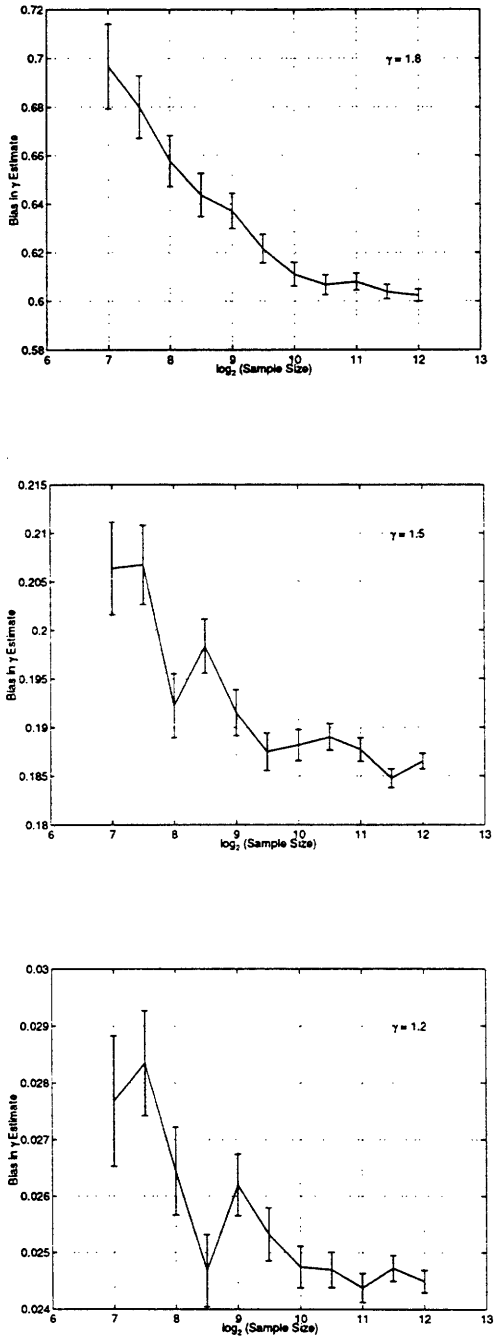


Figure 4-4: *Performance of Hosking's parameter estimator for noise-free fractal renewal process based on various numbers of interarrival measurements N. Each data point was computed from 512 Monte Carlo trials on simulated data. The plots correspond to the bias in the estimator for the cases $\gamma = 1.8, 1.5, 1.2$, with errorbars denoting 1-standard-deviation spread in the estimation results.*

method for the numerical maximization of the log-likelihood (4.8), which does not require any differentiation. The large bias observed in Fig. 4-4 is attributed primarily to the great discrepancy between the hard low-end truncation in the probability law for the data (4.2) and the soft low-end leveling in the model's probability density (4.7). Numerical values of the bias are tabulated in Table 4.1. These results contribute additional insight into the bias in our multiscale algorithm, as the underlying model possesses soft low-end cut-off similar to Hosking's model.

Additional simulations have also been conducted to explore the effects of noise. While the exact power-law density (2.7) lacks a well-defined second moment, a more reasonable measure of data strength is in terms of the sample quantiles. Specifically, all of our results are for the case where the mean of the exponential noise W is equal to the 75 percentile of the synthesized data. With this choice, roughly 3/4 of the interarrivals are significantly affected by the injected noise. While, as before, our implementation of our EM algorithm is initialized with the signal parameters

$$\hat{\lambda}_{[1]} = 10/\min\{x[1], x[2], \ldots, x[N]\}$$
$$\hat{\gamma}_{[1]} = 1.1,$$

the initial estimate for the noise parameter α is chosen to be

$$\hat{\alpha}_{[1]} = \alpha/10,$$

where α is the true value. Although α is generally unknown, this error margin in initial condition allows convenient implementation of the algorithm. Monte Carlo results for the EM parameter estimation are shown in Fig. 4-5, where as before, each point represents the bias for a certain combination of sample size N and shape parameter γ, computed from 512 trials. In these plots, we show the deviation of the γ estimates from the limiting estimates in the noise-free case. More precisely, the comparing factors are the true γ adjusted by the bias from the second column of Table 4.1. It is apparent from the plots that the estimation is nearly unaffected by the noise. As before, the standard deviation of the estimators suggests improved performance for increased sample size.

For comparison, we have applied Malik's, Kern's and Hosking's algorithms to scenarios

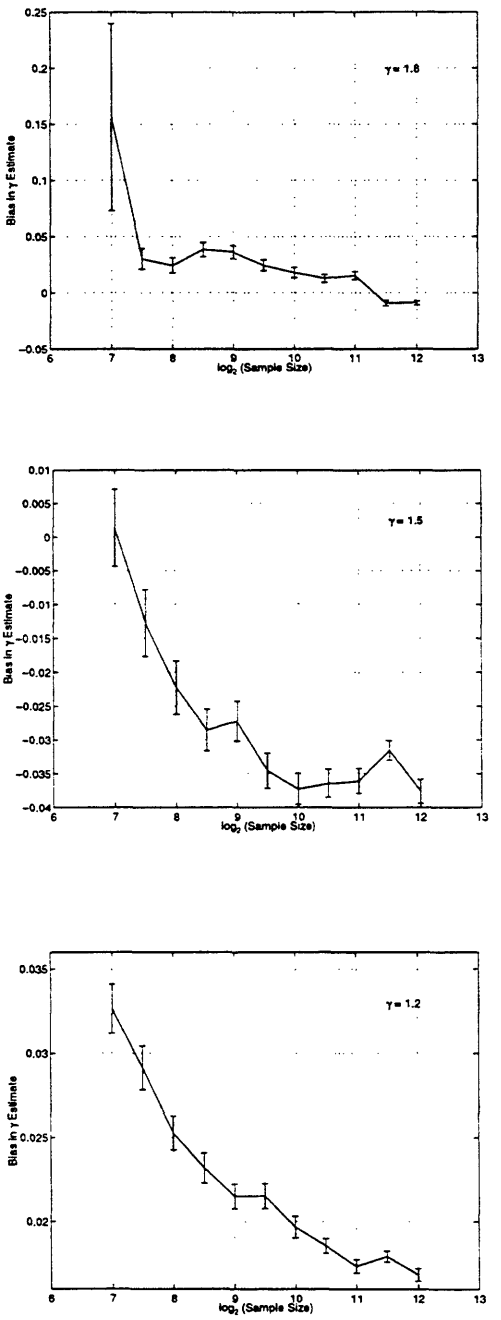


Figure 4-5: *Performance of the multiscale EM parameter estimator for noise-corrupted fractal renewal process based on various numbers of interarrival measurements N, and true γ values, 1.8, 1.5, 1.2. Each data point was computed from 512 Monte Carlo trials on simulated data. The noise term $W[n]$ has mean equal to the 75 percentile of sample interarrivals. The plots correspond to the deviation of the estimate of γ from the estimate in the noise-free case. The errorbars denote 1-standard-deviation spread in the estimation results.*

involving same noise strength. While these algorithms are developed with an assumption of clean data, one might argue that because of the high variance of the power-law function, the algorithms should be minimally affected by the noise terms. The results are shown in Figs. 4-6, 4-7, and 4-8, again in the form of deviation from the noise-free estimates. Despite the reliable estimates they generate for the noise-free cases, Malik's and Kern's algorithms suffer major degradation under such adverse conditions, as manifested by the great inconsistency of the estimates with the noise-free case. As is apparent from Fig. 4-8, the agreement between noise-free and noisy estimation is also poor in the case of Hosking's estimator. The effects of the noise were still very pronounced, seriously affecting the accuracy of the estimator. The robustness of these three algorithms thus appear much inferior to our EM parameter estimator in noisy scenarios.

Not surprisingly, the noise level also plays a part in determining the estimator performance. While, as stated earlier, the power in the exponential noise term $W[n]$ is proportional to $1/\alpha^2$, it is easy to show that assuming a finite-scale representation, the power in the $X[n]$ is proportional to $1/\lambda^2$ when the true shape parameter γ, and the lower and upper scale indices $\underline{m}$ and $\overline{m}$ are fixed. As such, the quantity α/λ provides a measure of the signal-to-noise ratio (SNR) in this case. We have conducted a set of simulations to explore the estimator performance under various values of α/λ and the results are summarized in Fig. 4-9. Throughout the experiments, N was fixed at 2^9, and $\underline{m}$ and $\overline{m}$ at 1 and 30, respectively. To allow easy specification of the true value of λ, $\underline{m}$, and $\overline{m}$, the synthesis of the power-law random variables $X[n]$ was based on the discrete multiscale framework. Each point in the graph represents the estimator variance computed from 32 Monte Carlo trials. As expected, the estimator performance improves as α/λ increases.

4.2 Interarrival Time Estimation

Given that the interarrival observations are modeled as noise-corrupted, a natural question is the recovery of the original interarrivals. In Section 4.1, we already saw that interarrival time estimation played an important role in the parameter estimation process. In this section, we explore the problem of interarrival time estimation in its own right. As will become apparent, parameter estimation, in turn, plays a key role in interarrival time estimation.

Our multiscale framework leads to an efficient procedure for computing the minimum

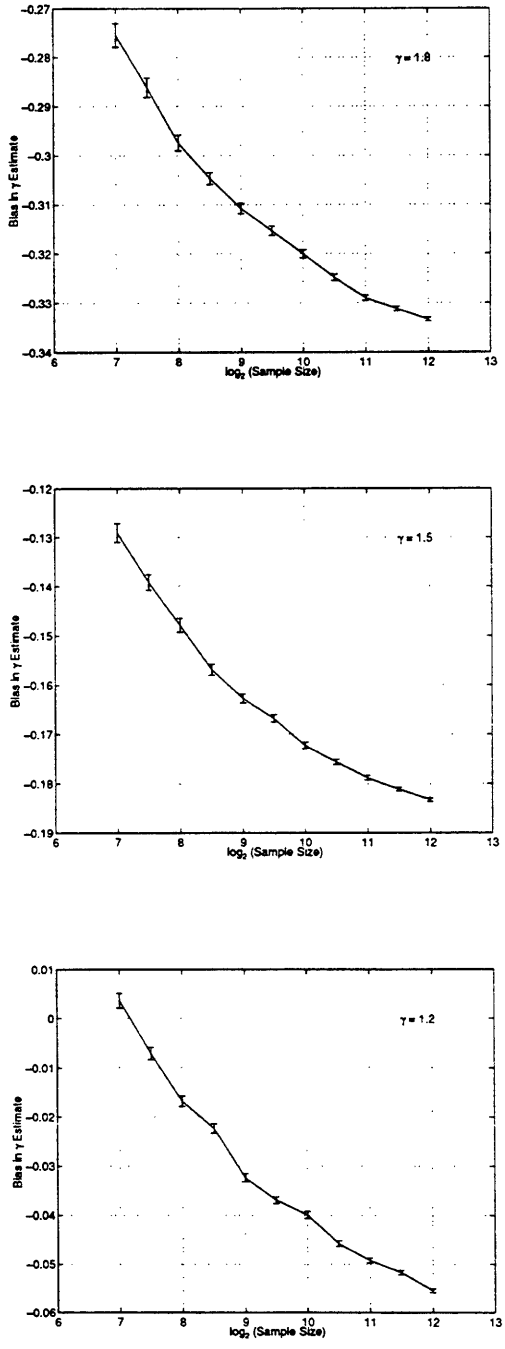


Figure 4-6: *Performance of Malik's EM parameter estimator for noise-corrupted fractal renewal process based on various numbers of interarrival measurements N, and true γ values, 1.8, 1.5, 1.2. Each data point was computed from 512 Monte Carlo trials on simulated data. The noise term $W[n]$ has mean equal to the 75 percentile of sample interarrivals. The plots correspond to the deviation of the estimate of γ from the estimate in the noise-free case. The errorbars denote 1-standard-deviation spread in the estimation results.*

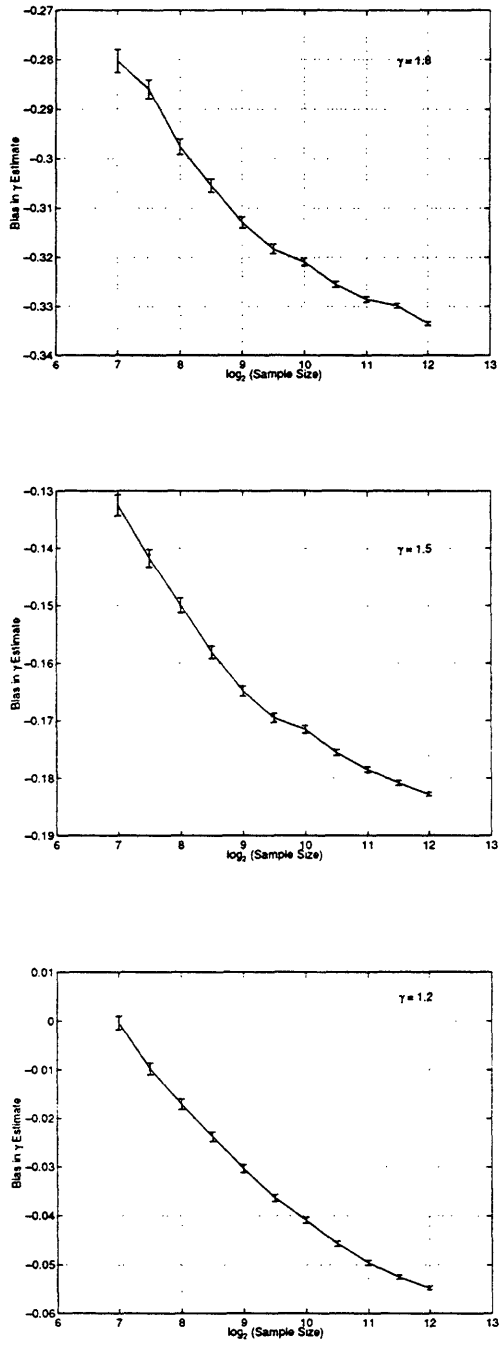


Figure 4-7: *Performance of Kern's EM parameter estimator for noise-corrupted fractal renewal process based on various numbers of interarrival measurements N, and true γ values, 1.8, 1.5, 1.2. Each data point was computed from 512 Monte Carlo trials on simulated data. The noise term $W[n]$ has mean equal to the 75 percentile of sample interarrivals. The plots correspond to the deviation of the estimate of γ from the estimate in the noise-free case. The errorbars denote 1-standard-deviation spread in the estimation results.*

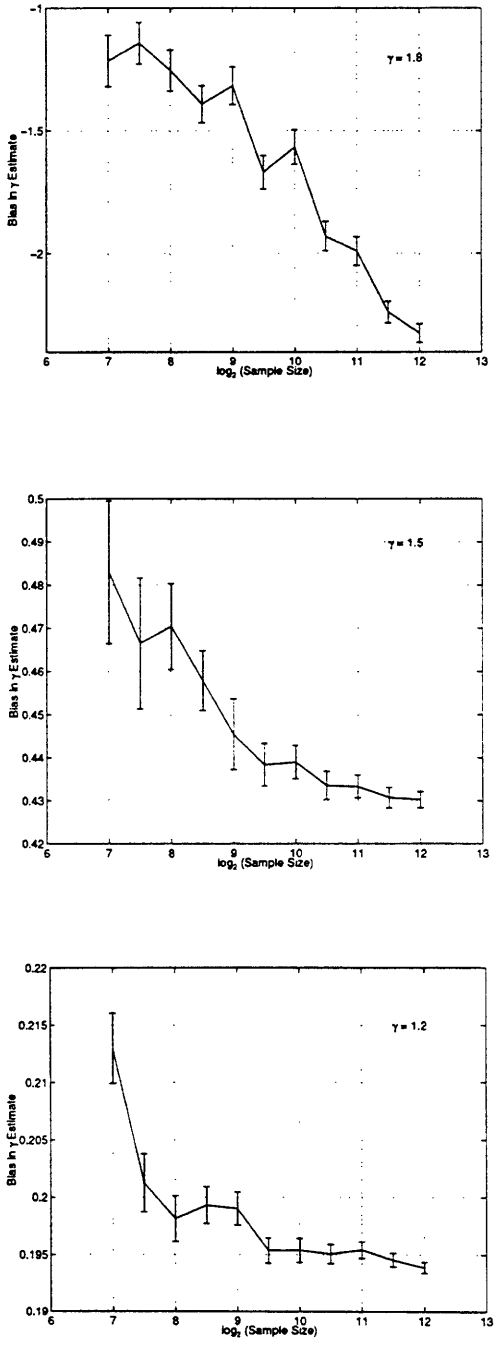


Figure 4-8: *Performance of Hosking's EM parameter estimator for noise-corrupted fractal renewal process based on various numbers of interarrival measurements N, and true γ values, 1.8, 1.5, 1.2. Each data point was computed from 512 Monte Carlo trials on simulated data. The noise term $W[n]$ has mean equal to the 75 percentile of sample interarrivals. The plots correspond to the deviation of the estimate of γ from the estimate in the noise-free case. The errorbars denote 1-standard-deviation spread in the estimation results.*

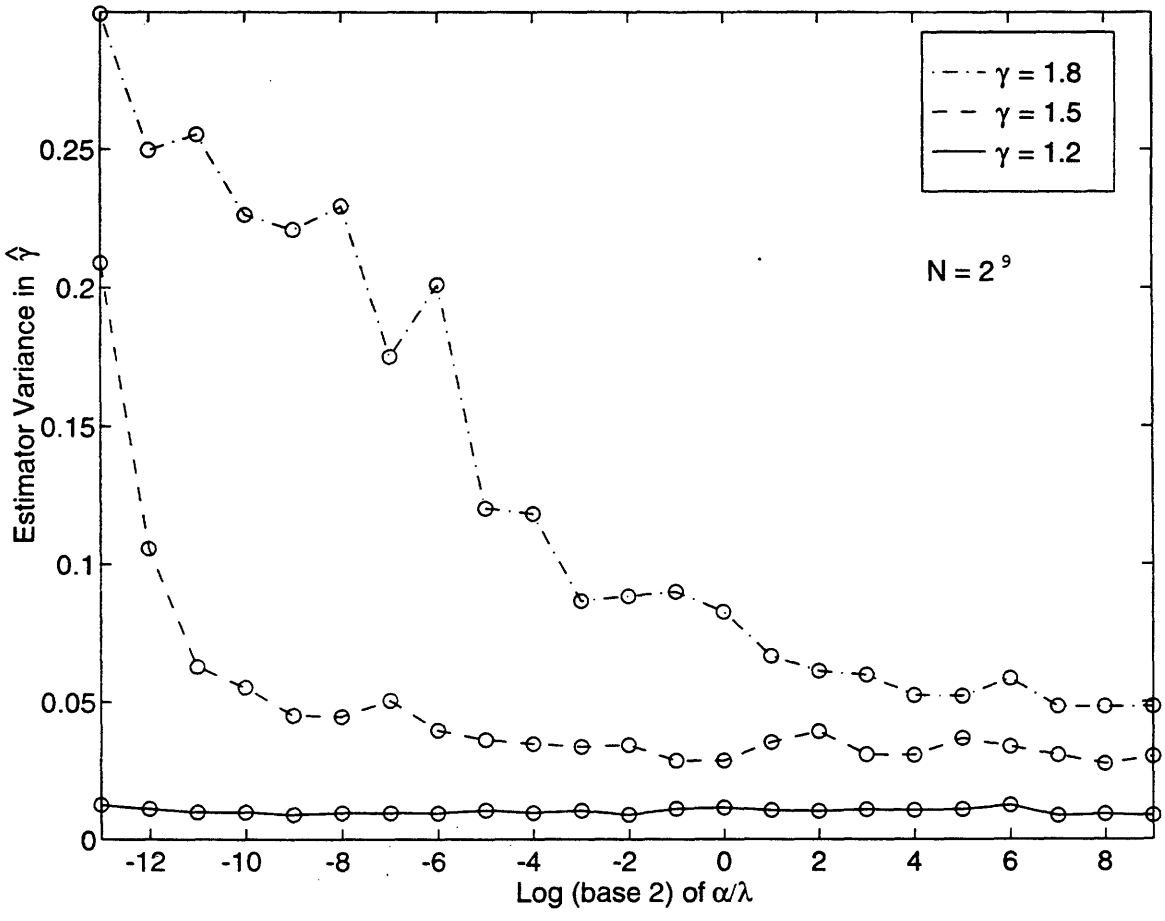


Figure 4-9: *Performance of the multiscale EM parameter estimator for corrupted fractal renewal process under various values of SNR. Each circle represents the variance in the shape parameter estimate $\hat{\gamma}$ as computed from 32 Monte Carlo trials on simulated data. The ratio between the noise and signal parameters, α/λ, is a measure of SNR. The sample size N was 2^9.*

mean-square error estimate of an interarrival interval given noisy measurements of the type considered in Section 4.1. In particular, we obtain

$$\begin{aligned}\hat{x}[n] = E\left[x[n] \mid r[n]\right] &= \int_0^{r[n]} \frac{x}{f_R(r[n])} \sum_{m=\underline{m}}^{\overline{m}} p_M(m) f_{X|M}(x \mid m) f_{R|X,M}(r[n] \mid x, m) dx \\ &= \sigma_M^2 \frac{\alpha e^{-\alpha r[n]}}{f_R(r[n])} \sum_{m=\underline{m}}^{\overline{m}} q^m \lambda_m \Phi_m(r[n]) \end{aligned} \tag{4.18}$$

where $f_R(r[n])$ is obtained via (4.11) with (4.12), and

$$\Phi_m(r) = \begin{cases} \dfrac{1}{(\lambda_m - \alpha)^2}\left[1 - e^{-(\lambda_m - \alpha)r}\left(1 + (\lambda_m - \alpha)r\right)\right] & \lambda_m \neq \alpha \\ \\ r^2/2 & \text{otherwise.} \end{cases} \tag{4.19}$$

It is important to emphasize that the resulting estimator is a highly nonlinear function of the data. This is, of course, not surprising given the highly non-Gaussian nature of the problem. Also, as expected, the interarrival estimator depends directly on the signal and noise parameters. In general, if their actual values are unknown *a priori,* estimates obtained with the algorithm in Section 4.1 can be used.

As a minimum mean-square error estimator, the algorithm specified by (4.18) with (4.19) is unbiased. Further, the variance of the estimator error can be computed in a straightforward manner. In particular, we have

$$E\left[(x[n] - \hat{x}[n])^2\right] = E\left[x[n]^2\right] - E\left[\hat{x}[n]^2\right] \tag{4.20}$$

where

$$E\left[x[n]^2\right] = \frac{2\sigma_M^2}{\lambda^2} \sum_{m=1}^{L} q^m \rho^{2m}$$

and $E\left[\hat{x}[n]^2\right]$ can be computed numerically using (4.18) with (4.19) and (4.11). By comparing the initial noise variance $1/\alpha^2$ with the value obtained with (4.20), we can compute the theoretical achievable SNR gain, which we have plotted in Fig. 4-10. Note that the horizontal axis is normalized with a constant α_0 chosen such that the initial SNR corresponding to $\lambda = 1$ and $\alpha = \alpha_0$ is 0 dB. In the case considered, $\gamma = 1.5$, $\underline{m} = 1$, and $\overline{m} = 30$. As

expected in a typical signal estimation problem, lower initial SNR, correspondingly smaller value of α/λ, generally leads to higher SNR gain. However, it should be noted that while the performance depends on α and λ, in general this dependence cannot be summarized in terms of a simple SNR-type quantity such as α/λ, as in the case of the parameter estimation algorithm. Nevertheless, Fig. 4-10 suggests that in scenarios of low initial SNR, the SNR gain is, to a good approximation, a function of α/λ.

This concludes our investigation of analysis and estimation problems involving fractal renewal processes, and also our study of these point processes in isolation. In the remaining chapters, we focus on the behavior of these point processes in familiar discrete-event systems, particularly, networks and queues.

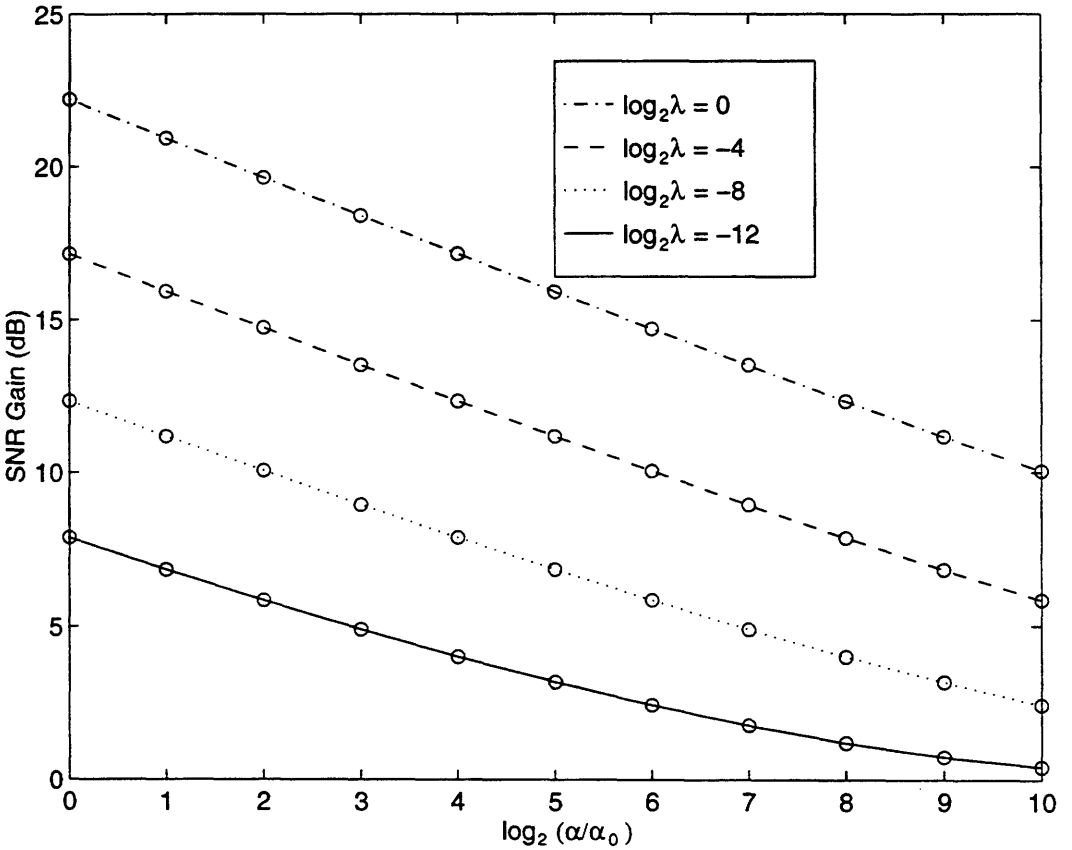


Figure 4-10: *Performance of the interarrival time estimator for fractal renewal processes under various combinations of signal and noise parameters λ and α. Each circle represents the achievable SNR gain computed by comparing the initial error variance which is $1/\alpha^2$, and the theoretical error variance in the estimates, which is determined numerically. In this case, the shape parameter $\gamma = 1.5$. The normalization constant α_0 ($\approx 2.1e-7$) is chosen such that the initial SNR corresponding to $\lambda = 1$ and $\alpha = \alpha_0$ is 0 dB.*

Chapter 5

Networking Problems involving the Fractal Renewal Process

The problems addressed in Chapters 3 and 4 involve processing of fractal point processes in isolation. In many cases, there is a need to analyze such point processes in various systems. Of particular interest are discrete-event systems such as networks and queues. From a communications engineering viewpoint, typical activities within a network are merging and branching of traffic. The corresponding mathematical representations of superposition and random erasure of point processes are therefore of importance. Given that fractal point processes are well-matched with network traffic (as suggested by, e.g., [32] [29] [37] [5]), there are great incentives to understand the superposition and branching of these point processes.

In addition to interactions among multiple traffic streams, an equally important class of networking problems pertains to individual streams at switches and links in a network. In a packet-switched environment, these networking resources are allocated according to the need of randomly arriving requests [7]. As such, queueing theory furnishes the appropriate tools for their modeling. In fact, together with superposition and branching, queueing theory provides a complete model for a very broad class of networks. In addition, queueing theory by itself is central for modeling various scheduling problems such as resource sharing in a computer system. Motivated by self-similar nature of various forms of network traffic, significance of queueing problems involving fractal inputs is unquestionable. As will be apparent, closely related are another set of important queueing systems characterized by

their power-law holding times. Both classes of queueing problems will be addressed in this chapter.

As in the last chapter, the multiscale framework will remain the basis of our algorithm development and analysis in this chapter. However, while previous chapters relied heavily on interarrival statistics, we now shift our emphasis more to the counting statistics of these point processes, which will prove to be insightful for many of the new problems. With the aid of the multiscale pure-birth process, an alternative perspective of our multiscale representation, we first derive in Section 5.1 a collection of counting process results, the most notable of which is a z-transform characterization. As an immediate application, these transform results lead to a time-domain characterization of the counting process distribution associated with a fractal renewal process, both as observed by an arrival within the process and upon random incidence. In Sections 5.2 and 5.3, we apply these counting process results to obtain characterizations of fractal renewal processes under random erasure and superposition, which, among other applications, are appropriate models for traffic branching and merging, respectively. Finally, in Sections 5.4 and 5.5, we introduce generalizations of the multiscale pure-birth process, and demonstrate their usefulness in queueing analysis involving fractal traffic as well as self-similar service statistics. These queueing models will play an even more important role in the development of queueing policies in Chapter 6. As in our formulation of the multiscale estimators of Sections 4.1 and 4.2, our discussion of the upcoming problems will be founded on the finite-scale framework, although extensions to the infinite- or semi-infinite-scale versions are feasible in a number of cases.

5.1 Counting Statistics of the Fractal Renewal Process

For renewal processes, counting statistics generally provide a useful alternative characterization to the interarrival characterization. From the viewpoint of a service provider such as a data network, for example, intelligent allocation of resources is often founded on accurate estimate of the number of requests over given time intervals. From the perspective of each member of a traffic stream, the number of other customers present is often critical in determining expected waiting time as well as the throughput available. Thus, the counting statistics under these two modes of observation are of particular importance. As is well known, these observations are identical for the Poisson process, mainly because of its mem-

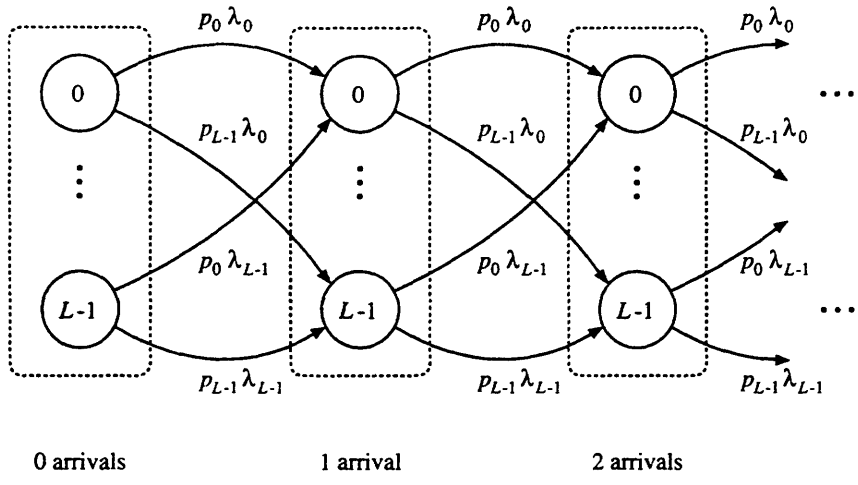


Figure 5-1: *The multiscale pure-birth process based on a finite-scale representation; dashed boxes denote conceptual grouping into superstates. The rate of departure from each state is a function the scale,* $\lambda_j = \lambda/\eta^j$; *the probability of entering scale* j' *of succeeding superstate is* $p_{j'} = \sigma^2 q^{j'}$.

oryless nature. On the other hand, we shall see that these two modes of observation are drastically different for the fractal renewal process.

Traditionally, the pure-birth process (see Fig. 2-2) plays a central role in the study of the Poisson counting process [26]. Indeed, with births corresponding to arrivals, this Markov model has led to standard procedures for computing the counting statistics of the Poisson process, particularly the Poisson probability distribution, which governs the number of arrivals over arbitrary time intervals. In this section, we introduce a generalized pure-birth process which we refer to as the multiscale pure-birth process. While it is mathematically equivalent to our multiscale framework of the previous chapters, the multiscale pure-birth process readily leads to new counting statistics results for the fractal renewal process, the most significant of which being a characterization of the counting process distribution.

5.1.1 The Multiscale Pure-Birth Process

Fig. 5-1 depicts the multiscale pure-birth process which forms the basis of our current exposition. Generalizing the well-known pure-birth process (see,e.g.,[26]), the state space of this Markov process is naturally partitioned into "superstates," each of which represents a certain number of births. A superstate is in turn composed of a set of states which correspond to the scales in our finite-scale framework. Hence, we index the states with a pair

of integers (i,j), where the superstate index i is nonnegative, while the scale index j ranges from 0 to $L-1$ for an L-scale representation. We also denote by $P_{i,j}(t)$ the probability that the process is in state (i,j) at time t. For convenience, we collect the probabilities within each superstate to form the vector $\mathbf{p}_i(t) \triangleq (P_{i,0}(t), P_{i,1}(t), \ldots, P_{i,L-1}(t))$. With this notation, the probability of i births at time t is $\pi_i(t) = \mathbf{p}_i(t)\mathbf{1}^{\mathrm{T}}$, where $\mathbf{1}$ is a row vector where the entries are all 1's. The ensemble $\{\pi_i(t); i = 0, 1, \ldots\}$ then constitutes the counting process distribution.

Every transition in the multiscale pure-birth process results in an increment in the superstate index, and thus a birth. In our model, the birth rate is assumed independent of the number of births already occurred. Thus, the mean departure rate from each state (i,j) is only a function of the scale j, of the specific form

$$\lambda_j \triangleq \lambda/\eta^j,$$

where as before, $\eta > 1$ is the scale increment, and $\lambda > 0$ is the reference rate. Upon a birth, every state of the succeeding superstate can be immediately reached, with probability dependent only on the scale of the target state. Specifically, the choice probability is of the form

$$p_{j'} \triangleq \sigma_M^2 q^{j'},$$

where j' is the scale index of the target state, $q \triangleq \eta^{1-\gamma}$ governs the relative weighting on the scales, while σ_M^2 is a normalization factor. Thus, this collection of probabilities is identical to those of (3.2). Moreover, it can be shown that the interval between consecutive births is a random choice among a multiscale family of exponential random variables, with a density of the form of (3.3), and hence approximates the power law over a finite range of length scales in the sense of (3.4). As an aside, we remark that an equivalent description of this Markov process is as shown in Fig. 5-1, where the transition rate from state (i,j) to state $(i+1, j')$ is equal $p_{j'}\lambda_j$. Comparing with Fig. 2-4, we see that our multiscale pure-birth process can be viewed as a special form of the hyperexponential pure-birth process, with geometric weighting probabilities p_j, and logarithmically spaced constituent rates λ_j.

5.1.2 The Counting Process Distribution

From the state-transition diagram of Fig. 5-1, it is straightforward to write the forward Kolmogorov equations [26] for the multiscale pure-birth process. Specifically, lumping the scalar equations, we have that

$$\frac{1}{\lambda}\frac{d}{dt}\mathbf{p}_0(t) = -\mathbf{p}_0(t)\mathbf{B} \tag{5.1a}$$

$$\frac{1}{\lambda}\frac{d}{dt}\mathbf{p}_i(t) = -\mathbf{p}_i(t)\mathbf{B} + \mathbf{p}_{i-1}(t)\mathbf{b}^{\mathrm{T}}\mathbf{q}, \qquad i \geq 1 \tag{5.1b}$$

where $\mathbf{q} \triangleq (p_0, p_1, \ldots, p_{L-1})$ is a vector of the choice probabilities, $\mathbf{b} \triangleq (\eta^0, \eta^1, \ldots, \eta^{L-1})$ is a vector of the multiscale dilation factors, and $\mathbf{B} \triangleq \mathrm{diag}(\mathbf{b})$ is a diagonal matrix with the dilation factors along its main diagonal. We proceed to solve the system of equations (5.1) in the transform domain, with the z-transform of the sequence $\{\mathbf{p}_i(t); i = 0, 1, \ldots\}$ defined as

$$\hat{\mathbf{p}}(z;t) \triangleq \sum_{i=0}^{\infty} z^i \mathbf{p}_i(t).$$

We remark that the z-transform of the counting process distribution $\{\pi_i(t); i = 0, 1, \ldots\}$ is given simply by $\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}}$. Now, multiplying the ith equation in (5.1) by z^i and summing the equations, we get that

$$\frac{1}{\lambda}\frac{d}{dt}\hat{\mathbf{p}}(z;t) = \hat{\mathbf{p}}(z;t)\left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right), \tag{5.2}$$

which can be readily solved to yield

$$\hat{\mathbf{p}}(z;t) = \hat{\mathbf{p}}(z;0)\exp\left(\left[-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right]\lambda t\right), \quad t \geq 0. \tag{5.3}$$

Thus, in contrast to the interarrival characterization, the counting process distribution will be dependent on initial conditions. Two choices of the initial condition $\hat{\mathbf{p}}(z;0)$ are of special importance, and these basically correspond to the two modes of observation discussed earlier. From the perspective of an arrival, it is natural to set the point of reference $t = 0$ at the arrival itself. In the context of data network analysis, for example, the resulting arrival-observed distribution is useful for describing a traffic stream as perceived by a packet in the stream. On the other hand, for observation upon random incidence, the reference point

$t = 0$ is randomized independently of the arrival process. This mode of observation is useful for describing packet traffic as viewed by the network, for example. In the sequel, we will address the two scenarios in detail.

In the arrival-observed case, where the point of reference $t = 0$ is a renewal, the statistics of the first interarrival are identical to every other interarrival. Hence, the scales within the zeroth superstate are chosen with the probabilities $\{p_j; j = 0, 1, \ldots, L-1\}$, and the initial condition in (5.3) is $\hat{\mathbf{p}}(z;0) = \mathbf{q}$. Inverting the resulting transform $\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}}$, we can obtain the time-domain characterization of the arrival-observed counting process distribution, which we denote by $\left\{\pi_i^{(\mathrm{a})}(t); i = 0, 1, \ldots\right\}$. In particular, evaluating the transform at $z = 0$, we have that

$$\pi_0^{(\mathrm{a})}(t) = \mathbf{p}_0(t)\mathbf{1}^{\mathrm{T}} = \hat{\mathbf{p}}(0;t)\mathbf{1}^{\mathrm{T}} = \mathbf{q}\exp(-\mathbf{B}\lambda t)\mathbf{1}^{\mathrm{T}}. \tag{5.4}$$

Expanding (5.4), we get the closed-form expression

$$\pi_0^{(\mathrm{a})}(t) = \sum_{i=0}^{L-1} p_i \exp\left(-\lambda_i t\right), \tag{5.5}$$

which agrees with well-known results for general renewal processes (see, e.g., [9]). Higher-order terms can be obtained numerically from the Taylor series expansion of the transform $\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}}$, which is of the form

$$\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}} = \mathbf{q}\left(\mathbf{I} + \left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)(\lambda t) + \frac{\left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)^2(\lambda t)^2}{2!} + \cdots\right)\mathbf{1}^{\mathrm{T}}, \tag{5.6}$$

where $\mathbf{I}$ denotes the identity matrix. While detailed derivations are given in Appendix D.1.1, the main results are that for $i \geq 1$,

$$\pi_i^{(\mathrm{a})}(t) = a_i^{(i)}\frac{(\lambda t)^i}{i!} - a_{i+1}^{(i)}\frac{(\lambda t)^{i+1}}{(i+1)!} + \cdots,$$

with the first-order coefficients given by

$$a_k^{(1)} = \sum_{l=1}^{k} M_l M_{k-l} \tag{5.7}$$

and the higher-order coefficients obtained via the recurrence relation

$$a_k^{(i)} = \sum_{l=1}^{k-i+1} M_l a_{k-l}^{(i-1)}, \tag{5.8}$$

where M_l is the lth moment of the random variable R distributed according to

$$\Pr\{R = \eta^{-j}\} = p_j, \quad j = 0, 1, \ldots, L-1. \tag{5.9}$$

For random incidence, we assume observation begins at a random time, with the point process already in equilibrium. With this assumption, the scale of the first interarrival is selected with the steady-state marginal distribution over scales. As a function of time t, this marginal distribution is $\sum_{i=0}^{\infty} \mathbf{p}_i(t)$, or equivalently, the transform $\hat{\mathbf{p}}(z;t)$ evaluated at $z = 1$. From (5.2), the governing equation for $\hat{\mathbf{p}}(1;t)$ is

$$\frac{1}{\lambda}\frac{d}{dt}\hat{\mathbf{p}}(1;t) = \hat{\mathbf{p}}(1;t)\left(-\mathbf{B} + \mathbf{b}^{\mathrm{T}}\mathbf{q}\right). \tag{5.10}$$

To investigate the steady-state marginal distribution over scales, i.e., $\lim_{t\to\infty} \hat{\mathbf{p}}(1;t)$, we set the derivative in (5.10) to zero. Now, it can be shown that the left null space of $(-\mathbf{B}+\mathbf{b}^{\mathrm{T}}\mathbf{q})$ is spanned by $\mathbf{q}\mathbf{B}^{-1}$; an argument is sketched in Appendix D.1.2. The initial condition for the random incidence scenario is therefore given by $\tilde{\sigma}^2\mathbf{q}\mathbf{B}^{-1}$, where the normalization constant $\tilde{\sigma}^2$ ensures that the elements of the vector sum to 1. Using the same techniques as before, we can obtain a time-domain characterization of the random incidence counting process distribution which we denote by $\left\{\pi_i^{(\mathrm{r})}(t); i = 0, 1, \ldots\right\}$. For example, it is straightforward to get the closed-form expression

$$\pi_0^{(\mathrm{r})}(t) = \tilde{\sigma}^2 \sum_{i=0}^{L-1} \frac{p_i}{\lambda_i} \exp\left(-\lambda_i t\right). \tag{5.11}$$

Also, considering the Taylor series expansion of the transform $\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}}$, which is,

$$\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}} = \tilde{\sigma}^2\mathbf{q}\mathbf{B}^{-1}\left(\mathbf{I} + \left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)(\lambda t) + \frac{(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q})^2(\lambda t)^2}{2!} + \cdots\right)\mathbf{1}^{\mathrm{T}}, \tag{5.12}$$

we conclude that higher-order terms of the probability distribution are of the form

$$\pi_i^{(r)}(t) = r_i^{(i)} \frac{(\lambda t)^i}{i!} - r_{i+1}^{(i)} \frac{(\lambda t)^{i+1}}{(i+1)!} + \cdots .$$

where the series coefficients can be obtained from those in the arrival-observed distribution via

$$r_k^{(i)} = (a_{k-1}^{(i)} + a_k^{(i-1)})\tilde{\sigma}^2. \tag{5.13}$$

Details of the derivation are given in Appendix D.1.2.

Figs. 5-2 and 5-3 show log-log plots of the time evolution of the first four terms in the arrival-observed and random incidence counting process distributions, respectively. Because these results are computed with a finite-scale representation, a characteristic scale is inherent. To adjust for this, the time axis has been normalized with respect to the finest-scale arrival rate λ. Empirically, we have observed that very few scales are required to obtain a good approximation to the probabilities over a finite time interval, although more scales are generally required in small γ situations due to the more persistent tail of the distribution.

The plots in Fig. 5-2 and 5-3 are obtained via computation with twenty dyadic scales, i.e., with $L = 20$, and $\eta = 2$, with the shape parameter set at $\gamma = 1.8$ in both cases. While the zeroth-order terms $\pi_0(t)$ are as given in the closed-form expressions (5.5) and (5.11), higher-order terms are generated numerically using Taylor series expansions with coefficients specified in (5.7), (5.8), and (5.13). Several features of the plots are worth pointing out. First, the marked qualitative difference between the two sets of plots reflects memory in the fractal renewal process. Contrast this with the memoryless Poisson process, for which the counting statistics are unaffected by the mode of observation.

Next, in the arrival-observed case of Fig. 5-2, the probability curves all tend to peak near the origin $t = 0$. More specifically, as we model the interarrival density with increasingly fine scales, i.e., with $\lambda \to \infty$, the peaks will tend to the time origin. This is in agreement with the heavy clustering property of fractal point processes in general: occurrence of an arrival at $t = 0$ boosts the likelihood of arrivals nearby. On the other extreme, our results suggest slow power-law decay in each plot in the long run, apparently much more persistent than the exponential tail of the Poisson distribution, for example. This heavy-tail behavior is attributed to the vast spacing between clusters (see, e.g., Fig. 2-5). Indeed, substantial

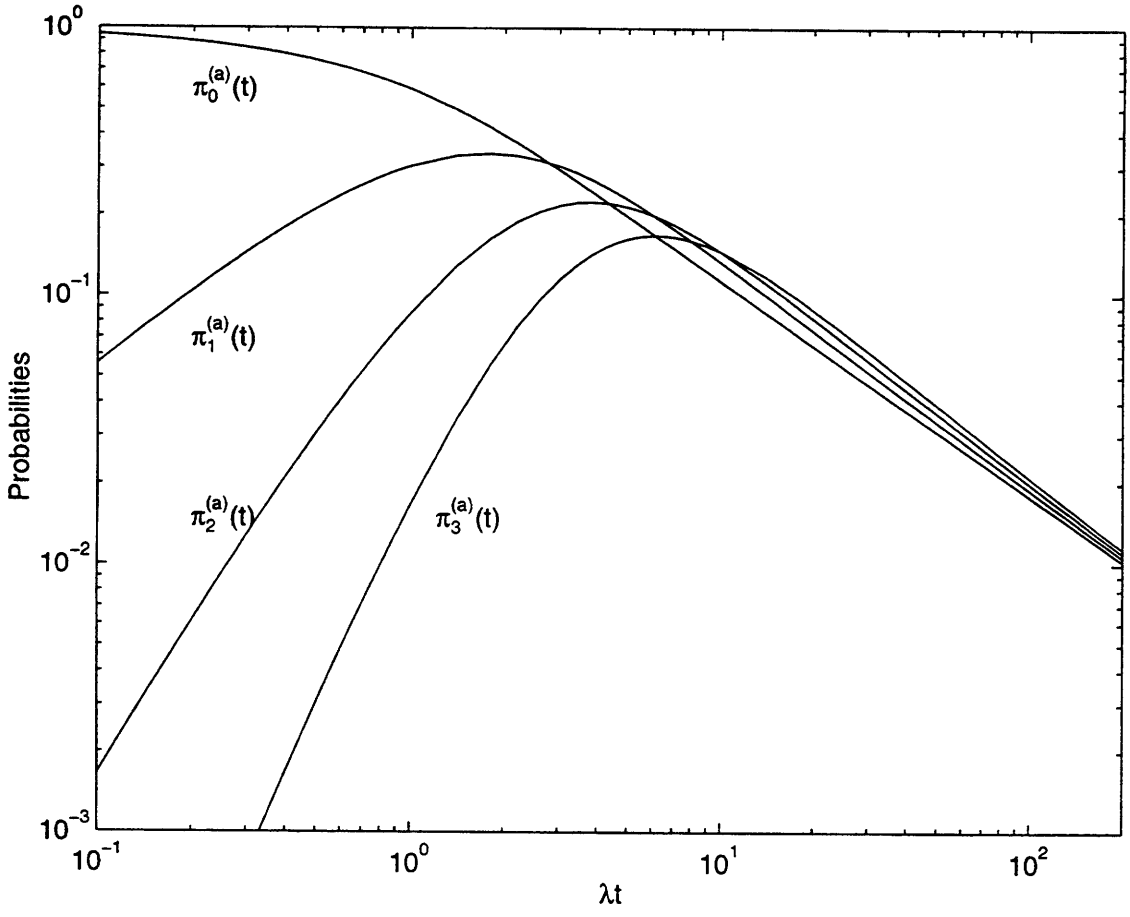


Figure 5-2: *First 4 terms in the arrival-observed counting process probability distribution for a fractal renewal process, computed with a 20-scale dyadic representation; the shape parameter of the interarrival distribution is* $\gamma = 1.8$. *The time axis is normalized with respect to the finest-scale arrival rate* λ.

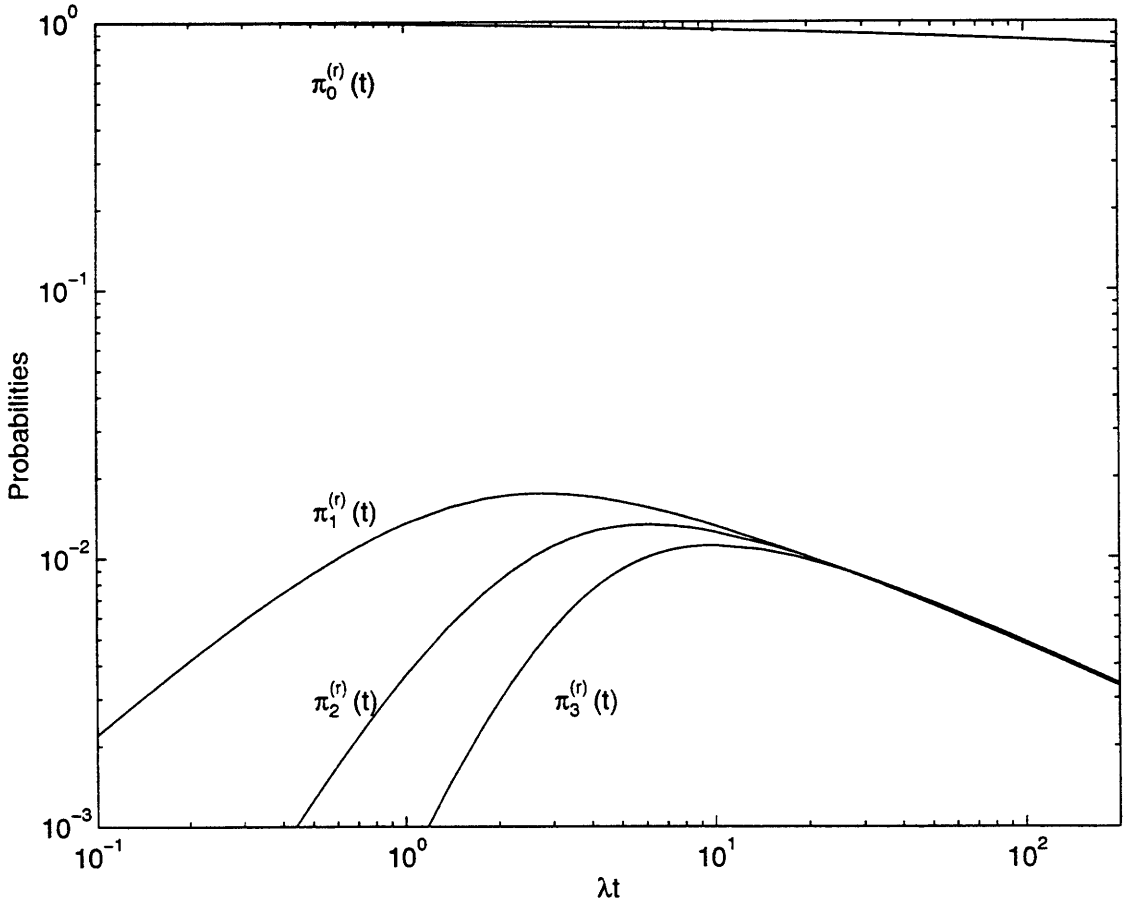


Figure 5-3: *First 4 terms in the random incidence counting process probability distribution for a fractal renewal process, computed with a 20-scale dyadic representation; the shape parameter of the interarrival distribution is* $\gamma = 1.8$. *The time axis is normalized with respect to the finest-scale arrival rate* λ.

probabilities of few arrivals over long gaps correspond closely to the scarceness of points over long quiescent gaps. Moreover, the power-law asymptotes are also in close agreement with the computational results of [42], based on asymptotic studies of convolution of multiple power-law functions.

In the random incidence case of Fig. 5-3, an outstanding property is the dominance of the zeroth-order term. Indeed, the probability of zero arrivals is consistently one to two orders of magnitude higher than the remaining terms. This dominance again agrees with the unusually long quiescent periods between clusters found in fractal point processes in general. As will be apparent, dominance of the zeroth-order term will have much implication in the superposition of fractal renewal processes.

5.2 Random Erasure of Fractal Renewal Processes

Random erasure, or random deletion of arrivals, is a point process transformation of general importance. In networking, for example, it is clearly a natural model for random packet losses. In addition, packet streams resulting from traffic branching can be individually modeled as an erasure of the original stream. As a preliminary investigation, we consider in this section the behavior of fractal renewal processes under the simple yet realistic Bernoulli erasure. While direct analysis based on the power-law interarrival density would be quite formidable, our counting process results of Section 5.1 readily yield a closed-form interarrival density for an erased fractal renewal process. As will be apparent, the key result of this analysis is the invariance of coarse-scale structures under this mode of erasure.

5.2.1 Interarrival Density Results

Under Bernoulli erasure, each arrival of a point process is deleted independently, with some common probability p. It is well known that under this transformation, a Poisson process with arrival rate λ remains a Poisson process, but with a modified rate $(1-p)\lambda$ (see, e.g., [25]). More generally, Bernoulli erasure of a renewal process yields a renewal process, with interarrivals W each equal to the sum of a random number of the original interarrivals X, where the number of terms summed is geometrically distributed [26]. In the transform

domain, the relation between the interarrivals is simple:

$$\hat{f}_W(s) = \frac{1-p}{1-p\hat{f}_X(s)}\hat{f}_X(s), \tag{5.14}$$

where $\hat{f}_W(s)$ is the Laplace transform of the interarrival density of the erased process $f_W(w)$, i.e.,

$$\hat{f}_W(s) \triangleq \int_0^\infty f_W(w)e^{-sw}\,dw,$$

and $\hat{f}_X(s)$ is the Laplace transform of the original interarrival density $f_X(x)$, defined analogously. Based on (5.14), general statements can be made about the asymptotic behavior of $f_W(w)$. Specifically, combining (5.14) with the identity $\hat{f}_W(0) = 1$ suggests that for well-behaved $\hat{f}_X(s)$,

$$\hat{f}_W(s) \to \hat{f}_X(s), \text{ as } s \to 0,$$

correspondingly in the density domain,

$$f_W(w) \to f_X(w), \text{ as } w \to \infty.$$

Thus, general renewal theory suggests preservation of structures, but only in an asymptotic sense.

Because fractal renewal processes are not true renewal processes, and in any case, the Laplace transform of the power-law density is unavailable in closed form, the relation (5.14) is not directly useful. Moreover, even with the multiscale representation, simplification of (5.14) does not appear to be straightforward. Thus, little additional insight is attainable from the transform relation (5.14). On the other hand, our counting process results of Section 5.1 provide a straightforward solution to this interarrival characterization problem. Now, with probability $1-p$, an arrival of the original process contributes a count of unity to the counting process resulting from Bernoulli erasure. Similarly, with probability p, it contributes a count of zero. In the transform domain, this corresponds to replacing z by $p+(1-p)z$ in the z-transform of the original counting process distribution. In particular, using (5.3), the counting process distribution of an erased fractal renewal process has the

z-transform

$$\hat{\mathbf{p}}(p+(1-p)z;t)\mathbf{1}^{\mathrm{T}} = \hat{\mathbf{p}}(z;0)\exp\left(\left[-\mathbf{B}+(p+(1-p)z)\mathbf{b}^{\mathrm{T}}\mathbf{q}\right]\lambda t\right)\mathbf{1}^{\mathrm{T}}. \tag{5.15}$$

Setting the initial condition to $\hat{\mathbf{p}}(z;0) = \mathbf{q}$, and evaluating (5.15) at $z = 0$, we have for the erased process,

$$\Pr\left\{\text{zero arrivals in } (0,w]\middle|\ \text{an arrival at } 0\ \right\} = \mathbf{q}\exp\left(\left[-\mathbf{B}+p\mathbf{b}^{\mathrm{T}}\mathbf{q}\right]\lambda w\right)\mathbf{1}^{\mathrm{T}}. \tag{5.16}$$

But this event is equivalent to the event $\{W \geq w\}$, where W is the interarrival beginning at 0. Upon differentiation, (5.16) leads to the interarrival density

$$\begin{aligned} f_W(w) = -\frac{d}{dx}\Pr\{W \geq w\} &= \mathbf{q}\exp\left(\left[-\mathbf{B}+p\mathbf{b}^{\mathrm{T}}\mathbf{q}\right]\lambda w\right)\left(\mathbf{B}-p\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)\mathbf{1}^{\mathrm{T}} \\ &= \lambda(1-p)\mathbf{q}\exp\left(\left[-\mathbf{B}+p\mathbf{b}^{\mathrm{T}}\mathbf{q}\right]\lambda w\right)\mathbf{b}^{\mathrm{T}}, \quad w \geq 0. \end{aligned} \tag{5.17}$$

Using (5.17), we have plotted in Fig. 5-4 the interarrival density of a fractal renewal process with shape parameter $\gamma = 1.8$, subject to various erasure probabilities. The computation was performed with a 20-scale dyadic representation. As before, the plots have been normalized to account for finite scales effects. Specifically, we have plotted $f_W(w)/\lambda$ against λw. These plots suggest that up to a renormalization, the erased interarrival density retains much of the power-law characteristics of the original density. In fact, our empirical studies suggest that as $\lambda \to \infty$, $L \to \infty$,

$$f_W(w) \sim \frac{1}{w^{\gamma}}$$

for every $w > 0$. Thus, in comparison with the Laplace transform results, our multi-scale analysis suggests a much stronger sense of invariance of fractal point processes under Bernoulli erasure.

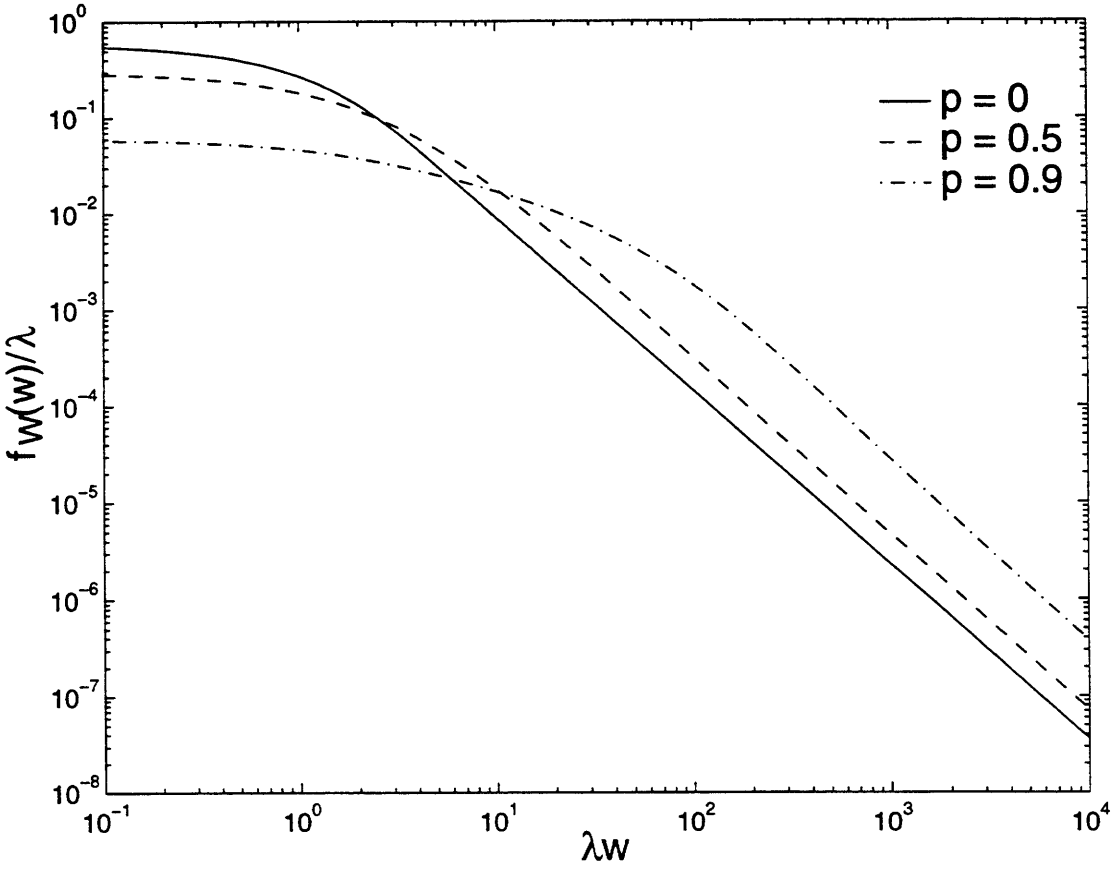


Figure 5-4: *Interarrival density function of a fractal renewal process under Bernoulli erasure, with erasure probability p. The shape parameter of the original process is* $\gamma = 1.8$. *These computations were performed with a 20-scale dyadic representation. The plots are normalized with respect to the finest-scale arrival rate* λ.

5.3 Superposition of Fractal Renewal Processes

In many situations, the overall characteristics of coexisting point processes are as important as those of the individual constituents. As a prime example, network design is typically concerned with the aggregate usage of multiple users, rather than any individual. While it is well known that the Poisson family constitutes a domain of attraction for superposition (see, e.g., [25]), self-similarity in data network traffic suggests that fractal point processes possess similar attraction behavior. As a preliminary study, we explore in this section a simpler but necessary condition, whether fractal renewal processes regenerate under superposition. Since superposition of point processes—even renewal processs—in general does not yield a renewal process, our discussion will be conducted from a counting statistics perspective.

Superposition of point processes corresponds simply to addition of counting processes. Using our results from Section 5.1, we can obtain a counting process characterization of the superposition of independent fractal renewal processes, either as observed by an arrival in one of the constituents or upon random incidence. Now, from an arrival's viewpoint, the structure of the two counting process distributions are inherently different. On one hand, the process to which it belongs is governed by the arrival-observed distribution; we denote this distribution by $\left\{\pi_i^{(1,\mathrm{a})}(t); i = 0, 1, \ldots\right\}$. On the other hand, since the constituents are independent, the arrival observes the random incidence counting processing distribution, denoted as $\left\{\pi_i^{(2,\mathrm{r})}(t); i = 0, 1, \ldots\right\}$, for the other point process. Thus, the overall counting process distribution will be a discrete convolution of $\pi_i^{(1,\mathrm{a})}(t)$ and $\pi_i^{(2,\mathrm{r})}(t)$. If, in addition, the two constituents have the same fractal dimension, the resulting counting process distribution will be expressed more simply as

$$\tilde{\pi}_i^{(\mathrm{a})}(t) = \sum_{j=0}^{i} \pi_j^{(\mathrm{a})}(t)\pi_{i-j}^{(\mathrm{r})}(t), \quad i = 0, 1, \ldots. \tag{5.18}$$

By a similar argument, the random incidence distribution of the superposition is the convolution of the two constituent random incidence counting process distributions, which for the case of equal fractal dimension, reduces to

$$\tilde{\pi}_i^{(\mathrm{r})}(t) = \sum_{j=0}^{i} \pi_j^{(\mathrm{r})}(t)\pi_{i-j}^{(\mathrm{r})}(t), \quad i = 0, 1, \ldots. \tag{5.19}$$

Figs. 5-5 and 5-6 show the arrival-observed and random incidence counting process distributions corresponding to the superposition of two independent fractal renewal processes. The shape parameters of the constituents are both $\gamma = 1.8$, and the computations were performed according to (5.18) and (5.19), respectively, using the counting process distribution results of Section 5.1. Comparing this set of plots with those for a single process (Figs. 5-2 and 5-3), we observe that key features such as the asymptotic power-law decay in the arrival-observed distribution, and dominance of zeroth-order term in the random incidence distribution, are largely preserved under superposition, suggesting invariance of fractal renewal processes under this transformation. The invariance comes about mainly because of the dominance of the zeroth-order term in the random incidence counting process probability distribution. Intuitively, this suggests that local structures are unaffected since the sparsely distributed clusters are not likely to collide. As a final remark, our counting process results provide additional verification of the invariance of fractal point proceses under superposition, complementing the spectral evidence given by Lowen, et al., in [30].

5.4 Queueing of Fractal Renewal Processes

Random erasure and superposition naturally model interaction among traffic streams in a communication network. To model activities at individual sites and links, queueing theory often provides the appropriate framework. More generally, in a resource-sharing environment, queueing policies often govern allocation and scheduling of resource usage, and are thus chiefly responsible for assurance of performance. As such, queueing theory is fundamental in operations research and industrial engineering, as well as various other engineering disciplines, with important applications in manufacturing, computer multitasking, air and vehicular traffic regulation.

While queueing problems involving memoryless service and memoryless traffic streams have been thoroughly addressed (see, e.g., [26]), these simple models are inadequate for many real scenarios. In particular, self-similarity of network traffic suggests the relevance of queueing problems involving fractal arrivals. Indeed, in the context of queueing analysis involving fractal point processes, many important questions arise in applications involving manipulation of opportunities are available, since variable-bit-rate video, Ethernet and WAN traffic with fractal characteristics. Most of the previous approaches have focused on

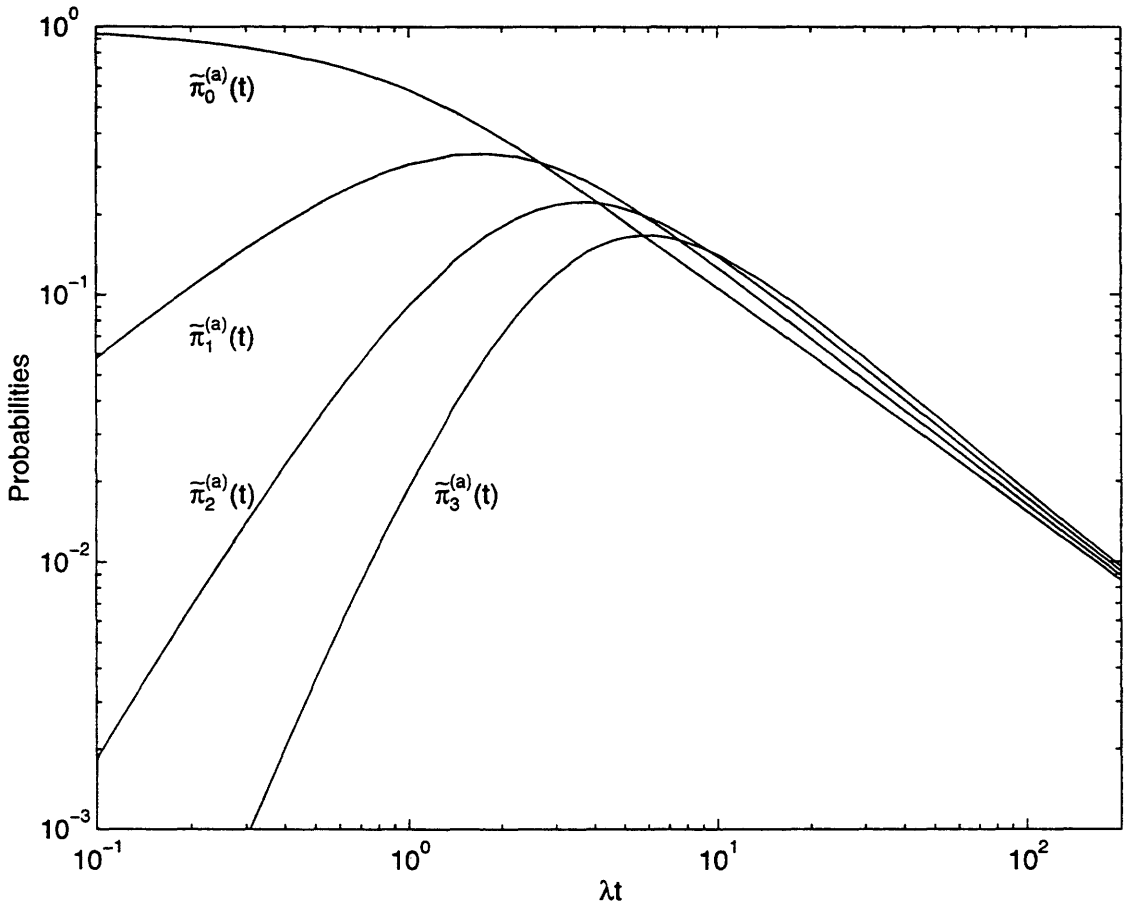


Figure 5-5: *Arrival-observed counting process distribution of the superposition of two independent fractal renewal processes, generated with the counting process results of Figs. 5.2 and 5.3. The shape parameter of both constituents is* $\gamma = 1.8$. *The time axis is normalized with respect to the finest-scale arrival rate* λ.

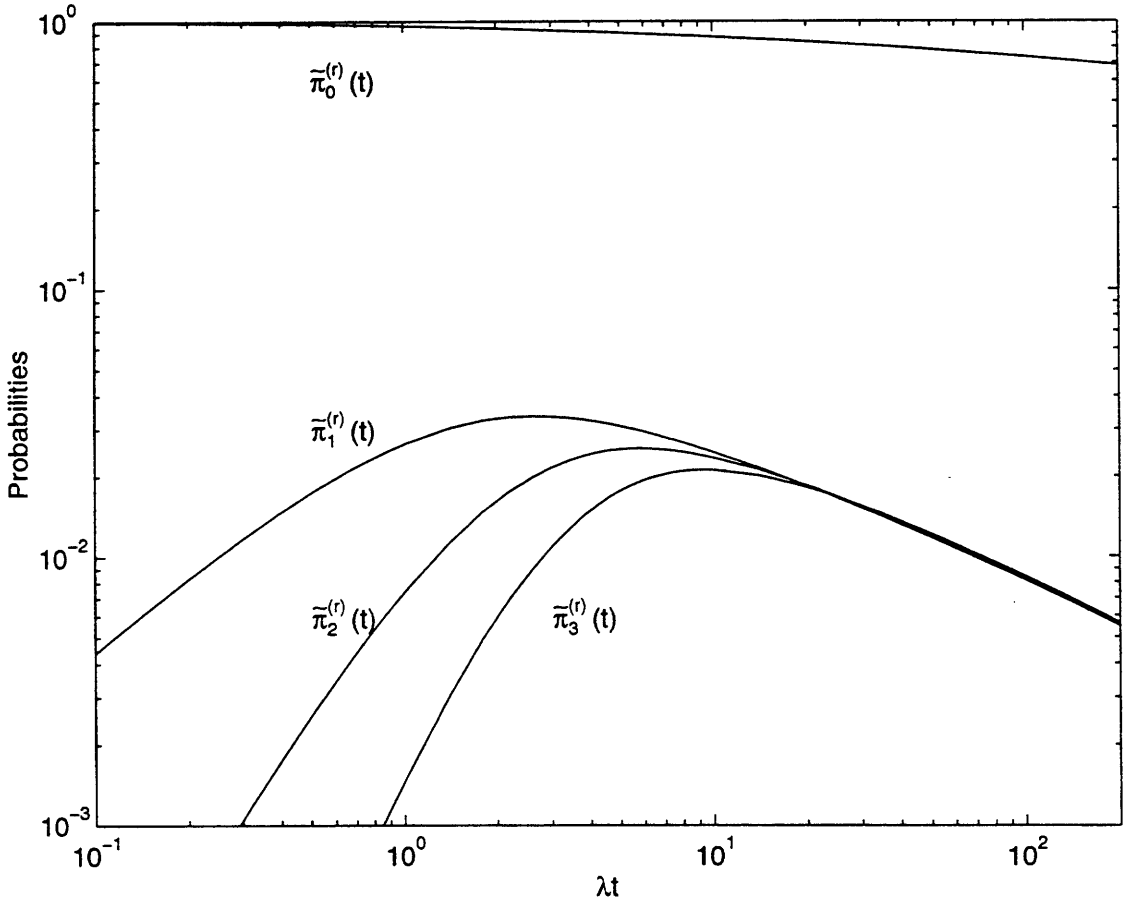


Figure 5-6: *Random incidence counting process distribution of the superposition of two independent fractal renewal processes, generated with the counting process results of Figs. 5.3. The shape parameter of both constituents is* $\gamma = 1.8$. *The time axis is normalized with respect to the finest-scale arrival rate* λ.

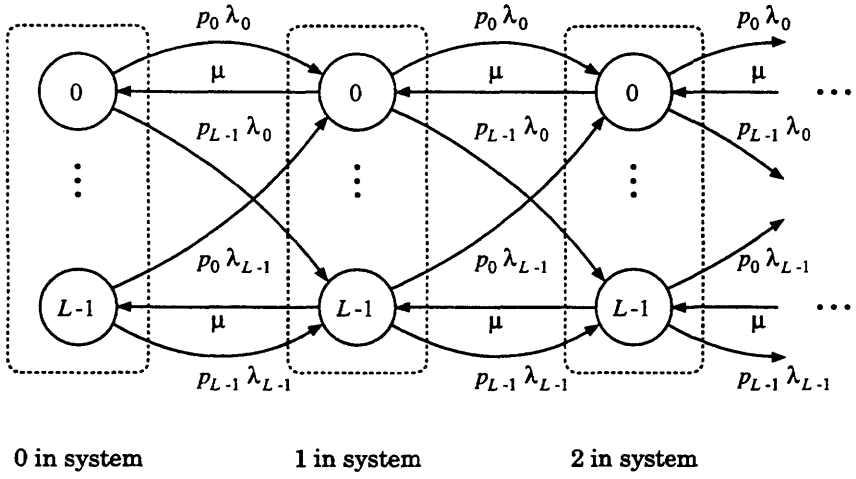


Figure 5-7: *The multiscale birth-and-death process based on a finite-scale representation. This is obtained by addition of death transitions to the multiscale pure-birth process of Fig. 5.1. To model a single-server memoryless queueing system, all death transitions occur with equal rate μ.*

queueing systems involving deterministic service, and analyses are mainly based on fluid models. In what follows, we consider queueing scenarios most naturally related to our multiscale birth-and-death process of the previous section. In particular, we shall see that simple modification of our multiscale pure-birth process leads to a framework for studying memoryless service of fractal traffic. In addition, a dual system allows us to explore the properties of queueing systems with memoryless input and power-law holding time. As we shall see, while both cases can be analyzed very readily, the latter requires somewhat more involved computations.

5.4.1 The Multiscale Birth-and-Death Process

The multiscale pure-birth process of Section 5.1.1 can be readily extended to suit a number of modeling applications. One such extension is the multiscale birth-and-death process depicted in Fig. 5-7, obtained by adding death transitions of the form $(i+1, j) \rightarrow (i, j)$, with rates all equal to μ. As another useful application, this process is useful for the study of single-server memoryless queueing systems servicing self-similar customer requests, with births modeling customer arrivals and deaths modeling service completion and customer departures. Because of the memoryless nature of the service process, this multiscale birth-and-death model is applicable to arbitrary service disciplines, as long as the queueing system is work-conservative, i.e., the server is never idle when there are customers.

It is worth noting that in addition to our multiscale birth-and-death process, several other variations of the multiscale pure-birth process are also of much interest. First of all, by varying the rates of the death transitions, the multiscale birth-and-death process can be generalized for multi-server scenarios. As another variation, by reversing the roles of the birth and death transitions, we can obtain a dual model well suited for memoryless customer arrivals with job sizes distributed as a power-law distribution—a reasonable model for typical phone usage, for example [40]. Generalizing further, using multiscale representation for both births and deaths, we can model self-similar arrivals with power-law job size distribution. In the present section, we shall focus primarily on the scenarios captured by the model of Fig. 5-7, while in the following section, we shall consider the dual situation. Our analysis techniques can be extended for most parts to many of the generalizations mentioned.

5.4.2 Steady-State Customer Distribution

The dynamics of the process in Fig. 5-7 are governed by the system of forward Kolmogorov equations

$$\frac{d}{dt}\mathbf{p}_0(t) = -\mathbf{p}_0(t)\lambda\mathbf{B} + \mathbf{p}_1(t)\mu\mathbf{I} \tag{5.20a}$$

$$\frac{d}{dt}\mathbf{p}_i(t) = -\mathbf{p}_i(t)(\lambda\mathbf{B} + \mu\mathbf{I}) + \mathbf{p}_{i-1}(t)\lambda\mathbf{b}^{\mathrm{T}}\mathbf{q} + \mathbf{p}_{i+1}(t)\mu\mathbf{I}, \quad i \geq 1. \tag{5.20b}$$

An interesting inquiry is the behavior of the queue in equilibrium, specifically the steady-state customer probability distribution $\{\pi_i; i = 0, 1, \dots\}$. To this end, we could seek the equilibrium state distribution $\{\mathbf{p}_i; i = 0, 1, \dots\}$, which could be computed by setting the derivatives in (5.20) to zero, and solving the resulting system of equations. Alternatively, an efficient algorithm for this computation is the matrix-geometric methods of [35]. By this, we first obtain a positive semi-definite solution $\mathbf{R}$ to the quadratic equation

$$\mathbf{0} = \rho\mathbf{b}^{\mathrm{T}}\mathbf{q} - \mathbf{R}(\rho\mathbf{B} + \mathbf{I}) + \mathbf{R}^2, \tag{5.21}$$

where ρ denotes λ/μ. This can be achieved via successive approximation. Specifically, beginning with $\mathbf{R}_{[0]} = \mathbf{0}$, we iterate the equation

$$\mathbf{R}_{[i]} \leftarrow (\rho \mathbf{b}^{\mathrm{T}} \mathbf{q} + (\mathbf{R}_{[i-1]})^2)\,(\rho \mathbf{B} + \mathbf{I})^{-1} \tag{5.22}$$

until convergence. The limit of the sequence $\{\mathbf{R}_{[i]}; i = 0, 1, \dots\}$ will be a positive semi-definite solution $\mathbf{R}$ of (5.21) [35]. Once $\mathbf{R}$ is obtained, the steady-state probability distribution is given by

$$\mathbf{p}_i = \mathbf{x}\mathbf{R}^i, \quad i = 0, 1, \dots, \tag{5.23}$$

where $\mathbf{x}$ is the left null vector of $\rho\mathbf{B} - \mathbf{R}$, normalized by the constraint

$$\mathbf{x}(\mathbf{I} - \mathbf{R})^{-1}\mathbf{1}^{\mathrm{T}} = 1.$$

Finally, the steady-state customer distribution can be obtained via $\pi_i = \mathbf{p}_i \mathbf{1}^{\mathrm{T}}$.

We have plotted the steady-state customer distribution in Fig. 5-8, for the case $\mu = 0.8$, $\lambda = 1.0$, and $\gamma = 1.8$, computed with a 20-scale dyadic representation. Empirically, we have found that the computations only require very few scales to stabilize. In the case considered, scales beyond the 20-scale representation contributed little visible change in the plot. To compute $\mathbf{R}$, we applied the iteration (5.22) until the absolute change in each coefficient was less than $\epsilon = 1 \times 10^{-13}$. For computation of $\mathbf{x}$, we numerically computed the eigenvector corresponding to the eigenvalue $\rho\mathbf{B} - \mathbf{R}$ of the smallest absolute value. A most notable feature of the distribution is the dominance of the probability of zero customers, equivalently, the idle probability. We attribute this to the unusually long gaps between clusters of customer arrivals.

Another outstanding aspect of the plot is that conditioned on 1 or more customers, it approaches a geometric distribution. That this is the case is attributed to the fact that the solution to (5.21) is of rank 1. Denoting the rank 1 solution to (5.21) via

$$\mathbf{R} \stackrel{\triangle}{=} \mathbf{v}^{\mathrm{T}}\mathbf{w}, \tag{5.24}$$

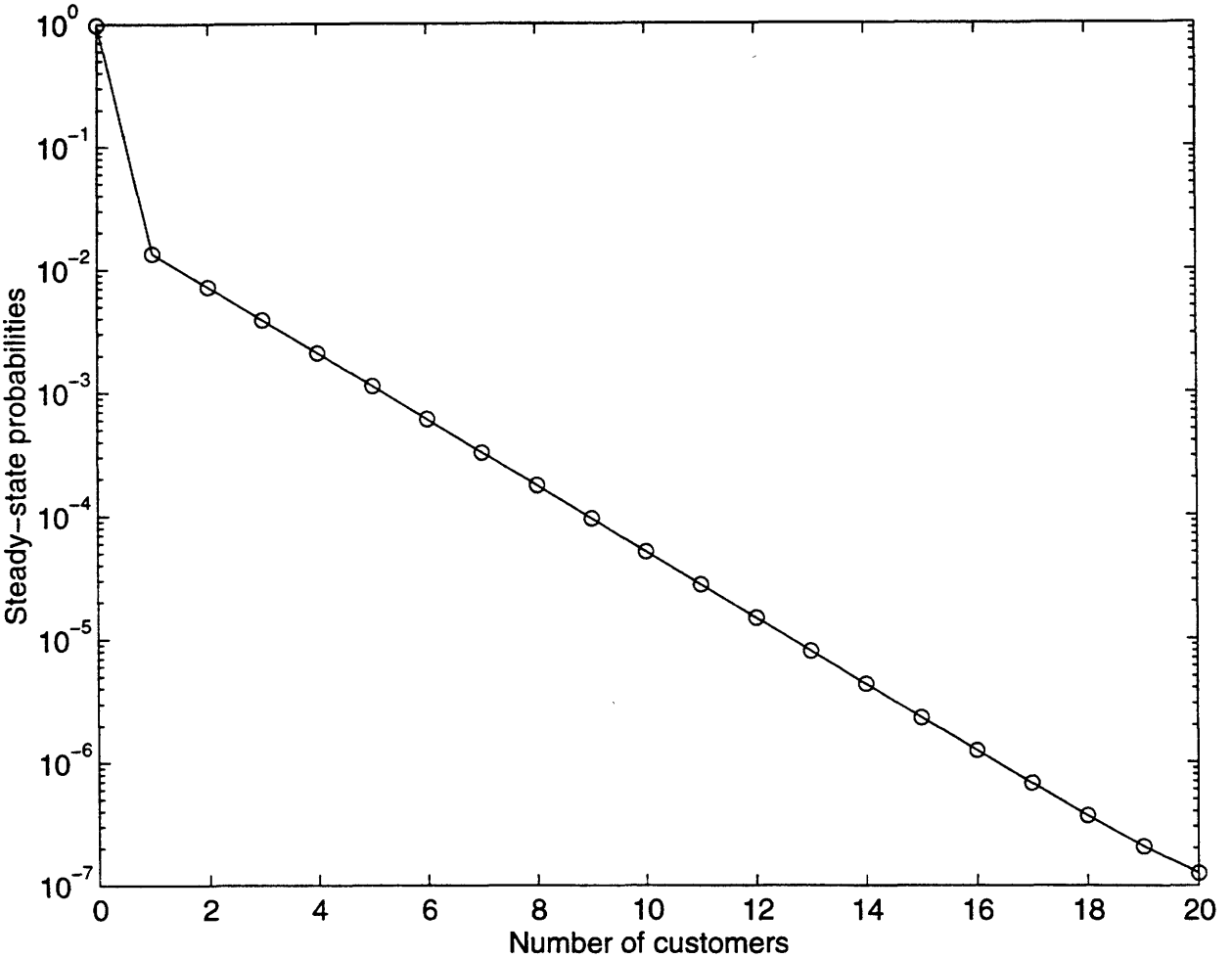


Figure 5-8: *First 21 terms in the steady-state customer distribution corresponding to a memoryless single-server queueing system with fractal customer arrivals. The input process is modeled with a 20-scale dyadic representation, with* $\gamma = 1.8$, $\lambda = 1$, *while the service rate is* $\mu = 0.8$.

we obtain from (5.23) that the steady-state customer distribution satisfies

$$\frac{\pi_{i+1}}{\pi_i} = k, \quad i = 1, 2, \ldots, \tag{5.25}$$

where k denotes the inner product $\mathbf{w}\mathbf{v}^{\mathrm{T}}$. We proceed to obtain a characterization of k.

From (5.21), it is clear that $\mathbf{v}$ must be in the direction of $\mathbf{b}$. Setting $\mathbf{v} = \mathbf{b}$, we get that $\mathbf{w}$ must satisfy

$$\rho\mathbf{q} = \mathbf{w}\left(\rho\mathbf{B} + \mathbf{I}\right) - \mathbf{w}k = \mathbf{w}\left[\rho\mathbf{B} + (1-k)\,\mathbf{I}\right].$$

Right multiplication of both sides by $\left[\rho\mathbf{B} + (1-k)\mathbf{I}\right]^{-1}\mathbf{b}^{\mathrm{T}}$ leads to

$$\sigma^2\left[\frac{\rho}{\rho + (1-k)} + \frac{q\rho/\eta}{\rho/\eta + (1-k)} + \cdots + \frac{q^{L-1}\rho/\eta^{L-1}}{\rho/\eta^{L-1} + (1-k)}\right] = k. \tag{5.26}$$

Eq. (5.26) can be solved iteratively using bisection search techniques, for instance. While closed-form solution of this equation is difficult, a useful bound can be obtained for k. By Jensen's inequality, the left hand side is less than or equal to

$$\frac{\rho E\left[R\right]}{\rho E\left[R\right] + (1-k)}. \tag{5.27}$$

where R is the random variable defined in (5.9). Simplifying, we get that

$$k(1-k) \leq \rho E\left[R\right](1-k).$$

Finally, since for an ergodic queue, $k < 1$, we obtain that

$$k \leq \rho E\left[R\right]. \tag{5.28}$$

The solid curves in Fig. 5-9 show values of k for various combinations of service rate and arrival statistics, obtained via a bisection search based on (5.26). The search was terminated when the absolute change in k was less than $\epsilon = 5 \times 10^{-10}$. Again, a 20-scale dyadic representation was used in the computation, with $\lambda = 1$. In general, we see that k decreases for fast service rates, implying a sharper decay in the steady-state customer distribution, or essentially, a less crowded system. Also, k is smaller for lower γ since the

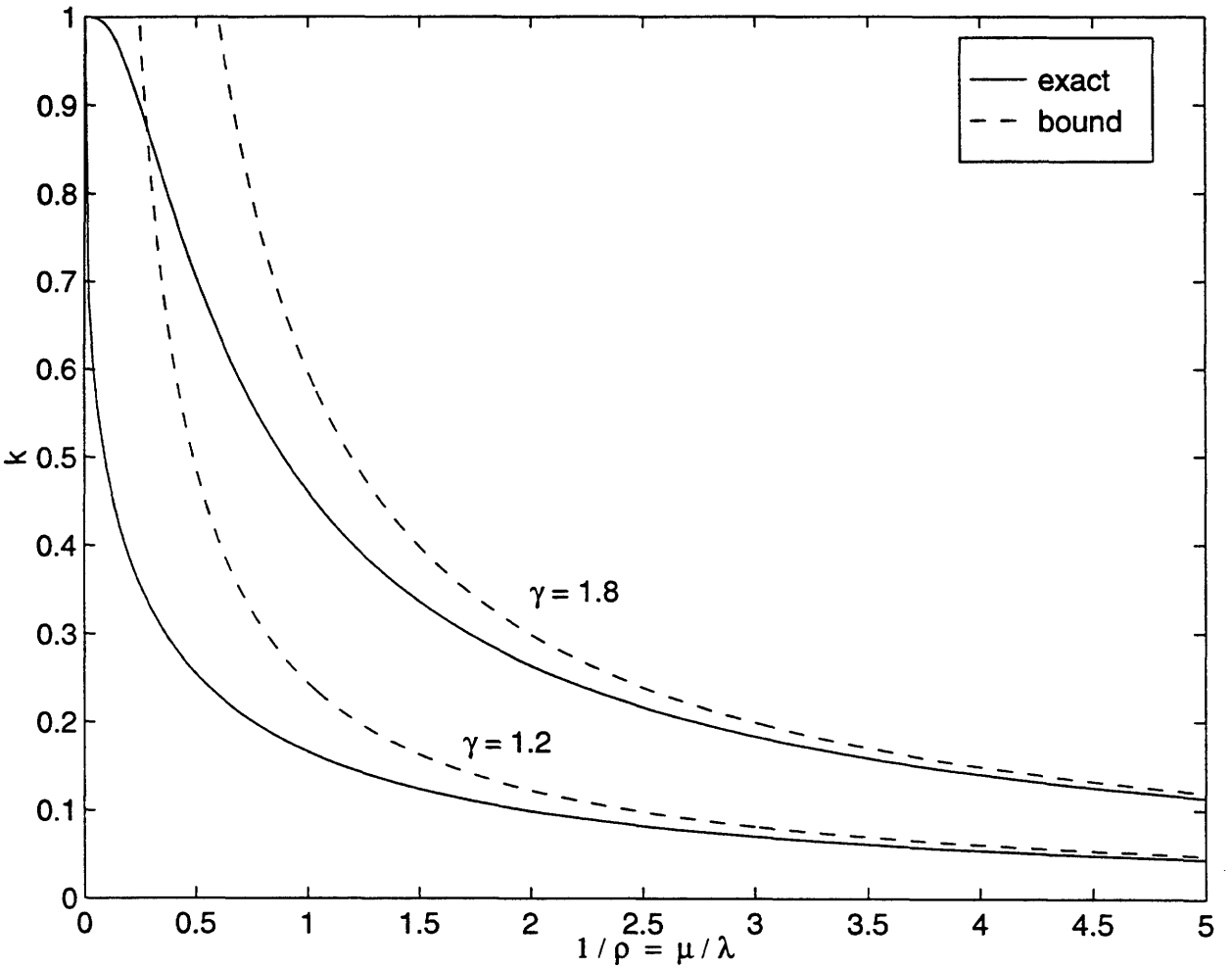


Figure 5-9: *Decay rate of the steady-state customer distribution in a single-server queueing system with fractal renewal process input. The solid curve is obtained via bisection search method based on (5.26). The dashed curve is the closed-form upperbound prescribed by (5.28). In this computation, the parameter values were* $\lambda = 1$, $L = 20$, $\eta = 2$.

arrivals are spaced further apart in these cases, leading to less congestion. The dashed plots in the same figure are bounds as prescribed by (5.28). While the bound is loose for slow-service scenarios, the plots suggest that (5.28) is in fact a potentially useful closed-form approximation for k in the fast-service regime.

To demonstrate the usefulness of our techniques for predicting queueing behavior, we have performed simulations of a memoryless queue servicing fractal renewal process inputs. Our simulation was based on the next-event time advance method for simulating discrete-event processes [28]. Cases where the input process had shape parameter $\gamma = 1.8$ and $\gamma = 1.5$, were considered, with the service rate set at $\mu = 0.15$ throughout. To include the effects of modeling errors, an exact power-law random variable, with density of the form (2.7) was used to synthesize the interarrivals. For service durations, an exponential random variable was used. Both random variables were synthesized by mapping uniform random variables. In all our simulations, 50 000 arrivals were used, and the system was observed until the last of these arrivals departed the system. The steady-state customer distribution was then obtained by time-integration. Specifically, following the ideas of [28], the probability estimates $\{\hat{\pi}_i; i = 0, 1, \dots\}$ were obtained via

$$\hat{\pi}_i = \frac{\int_0^T \chi_i(t)dt}{T},$$

where $\chi_i(t)$ is the indicator function for times when i customers were present in the system, and T represents the total time of the simulation. The estimated steady-state distributions are plotted as the solid curves in Fig. 5-10.

For analytic derivation of the steady-state queueing distributions, we first obtained a multiscale representation for the input process. To account for the bias in our multiscale model, we employed the noise-free parameter estimator of Section 4.1 to construct this representation. The resulting representation was then employed in the analysis described above. The resulting distributions are then plotted as the dashed curves in Fig. 5-10.

Extremely close agreement between the simulated data and predicted results is obtained. As Fig. 5-10 reflects, Plotted on a semilog scale, the slope of the linear portion from queueing simulation is $-.0412$ for the case $\gamma = 1.8$, while the theoretical slope is $-.0416$. Likewise, for the case of $\gamma = 1.5$, the simulation yielded a semilog plot with a slope of $-.1969$, while the theoretical slope is $-.2034$.

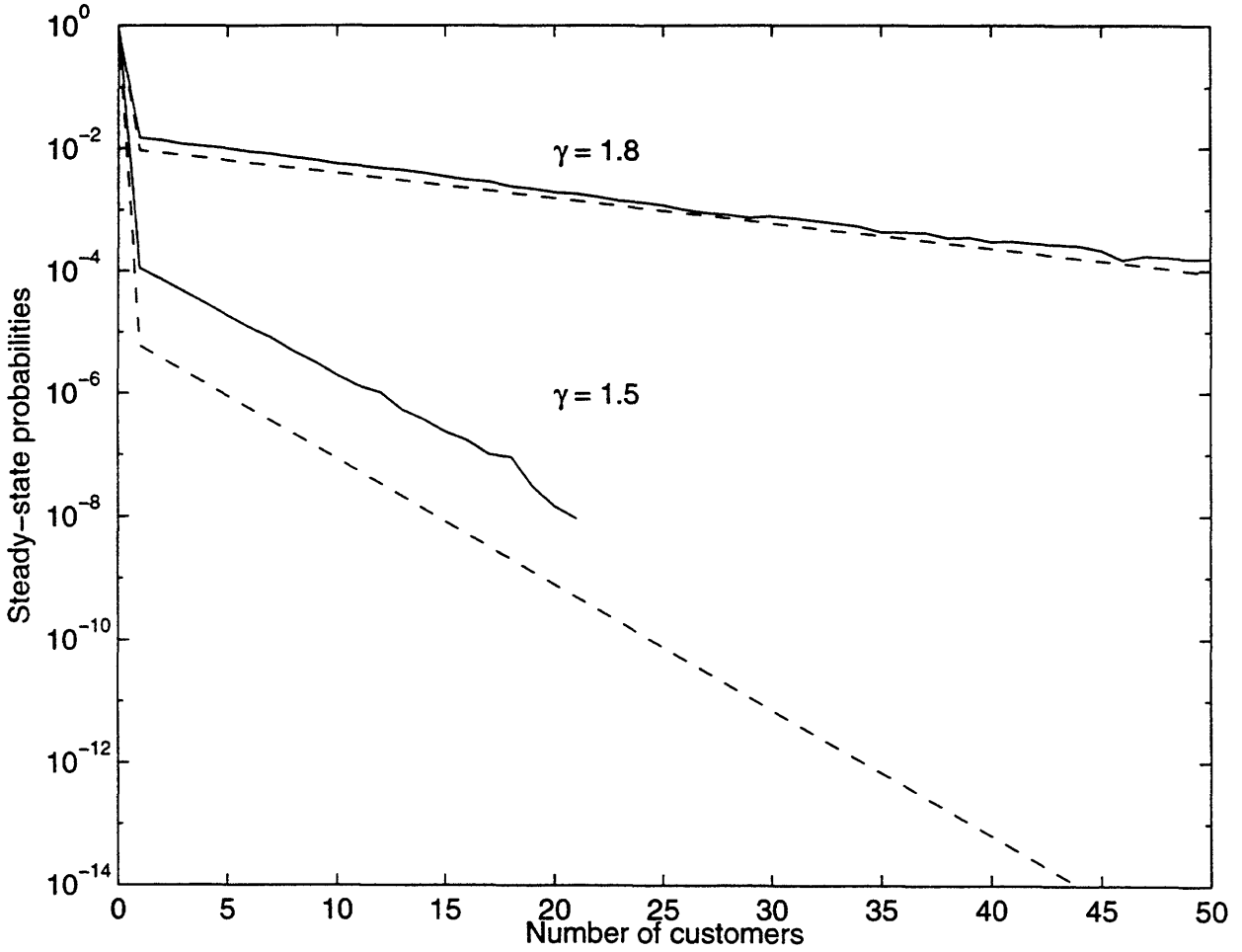


Figure 5-10: *Comparison of simulated and theoretical steady-state customer distribution for a memoryless queueing system servicing fractal renewal process input. The solid curves represent results of discrete-event simulation of the queueing system using 50 000 customers. The dashed curves represent theoretically-predicted distributions for the corresponding queueing scenarios. For both cases where input shape parameter* $\gamma = 1.8$ *and* $\gamma = 1.5$, *the service rate is* $\mu = 0.15$.

It is worth noting that in the broader framework of G/M/1 queues, the steady-state queue-length distribution could be computed using semi-Markov methods. In particular, it is proved that for G/M/1 queues in general, a geometric tail is observed:

$$\Pr\{N = n\} \sim \omega^n, \quad n = 1, 2, \ldots. \tag{5.29}$$

To retrieve ω, however, the general technique requires solution of the equation

$$\hat{f}_X(1 - \omega) = \omega \tag{5.30}$$

where $\hat{f}_X(s)$ denotes the Laplace transform of the interarrival density. Eq. (5.30) is intractable in the case of the power-law interarrival distribution. On the other hand, without invoking semi-Markov process theory, our framework yields straightforward analysis techniques which lead to the same results, as well as a closed-form bounding factor which is tight for fast-service scenarios. As will be apparent in the following chapter, our Markov framework will be extremely valuable in the formulation of queueing control policies.

5.5 Power-Law Service of Memoryless Input

The dual of the problem in the previous section, obtained by reversing the roles of customer arrivals and departures, is also of interest in a variety of applications. Specifically, customer arrivals are now modeled as memoryless. This assumption is compatible with a number of networking activities, such as TCP/IP connection arrivals discussed in [37]. This arrival model is also suggested by classical convergence theorems on superposition of well-behaved point processes [25]. On the other hand, the servicing statistics are now governed by a power-law random variable. As mentioned earlier, this heavy-tailed holding time is desirable for modeling many aspects of human activities such as telephone usage, among other applications: while typical jobs are short, substantial amount of outliers request lengthy service. As we did in Section 5.4, we shall now investigate the steady-state behavior of this queueing system.

Using our multiscale framework, we first model the service time as a mixture of a multiscale family of exponential functions. Specifically, for the power-law service time U,

with probability density function

$$f_U(u) = \begin{cases} \sigma_U^2 u^{-\gamma} & \underline{u} < u < \overline{u} \\ 0 & \text{otherwise} \end{cases} \tag{5.31}$$

where γ is again referred to as the shape parameter and σ_U^2 is for normalization, we employ a multiscale family of exponential functions obtained via dilation of a reference prototype U_0. More specifically, the prototype U_0 is governed by the probability density function

$$f_{U_0}(u) = \mu \exp(-\mu u), \quad u > 0,$$

where μ is the prototype service rate. Adhering to the finite-scale framework, service rates of the form $\mu_j \triangleq \mu/\eta^j$ are obtained for $j = 0, 1, \ldots, L-1$, where, as before, η is the scale increment. The choice probabilities will be of the form $p_i \triangleq \rho^{(1-\gamma)}$ where γ is the shape parameter in (5.31).

Using this multiscale representation for the power-law service duration, a generalized birth-and-death process can again be constructed for this queueing system. For simplicity, we consider the scenario where the input process is memoryless. The resulting Markov model is shown in Fig. 5-11. In contrast to the queueing system of Section 5.4, the service offerred by this system is not memoryless. Thus, the model in Fig. 5-11 is restricted to service disciplines whereby each customer is served until completion. In this sense, our model precludes various forms of time-slicing, for example. Nevertheless, many important disciplines are still captured, such as first-in-first-out, last-in-first-out, service in random order, and round robin.

From the state-transition diagram, the dynamics of this process are governed by

$$\frac{d}{dt}\mathbf{p}_0(t) = -\mathbf{p}_0(t)\lambda\mathbf{I} + \mathbf{p}_1(t)\mu\mathbf{B} \tag{5.32a}$$

$$\frac{d}{dt}\mathbf{p}_i(t) = -\mathbf{p}_i(t)(\lambda\mathbf{I} + \mu\mathbf{B}) + \mathbf{p}_{i-1}(t)\lambda\mathbf{I} + \mathbf{p}_{i+1}(t)\mu\mathbf{b}^{\mathrm{T}}\mathbf{q}, \quad i \geq 1, \tag{5.32b}$$

where λ denotes the arrival rate of the Poisson input, while $\mathbf{q}$, $\mathbf{b}$, and $\mathbf{B}$ are as in the previous case, and represent the row vector of choice probabilities, the row vector of dilation factors, and the diagonal matrix consisting of the dilation factors, respectively.

We proceed to find the steady-state customer distribution of this queueing system. As

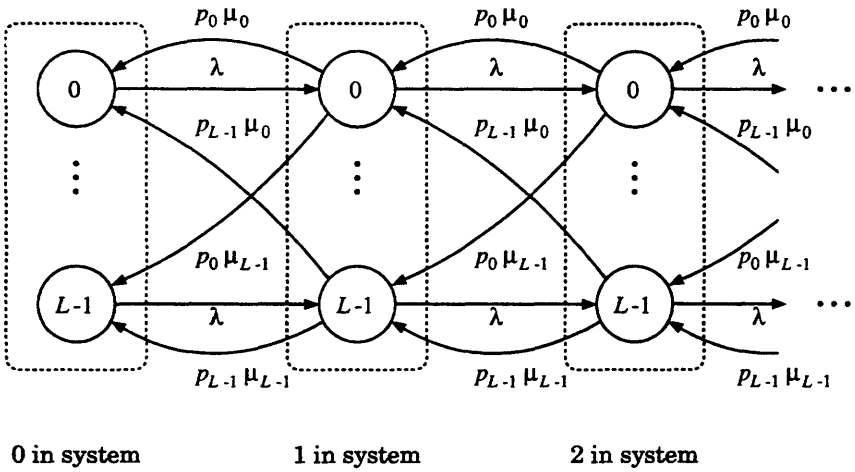


Figure 5-11: *The transposed multiscale birth-and-death process for memoryless input serviced by power-law server. This process is obtained by reversal of the roles of births and deaths in the process of Fig. 5.7.*

before, the matrix-geometric method of [35] will be of much importance. We first seek the solution to the matrix quadratic equation

$$\mathbf{0} = \rho\mathbf{I} - \mathbf{R}(\rho\mathbf{I} + \mathbf{B}) + \mathbf{R}^2\mathbf{b}^{\mathrm{T}}\mathbf{q}, \tag{5.33}$$

where as before $\rho \stackrel{\Delta}{=} \lambda/\mu$. Analogous to the method used in Section5.4, this equation can be solved iteratively, with the initial condition $\mathbf{R}_{[0]} = \mathbf{0}$, and successive updates as

$$\mathbf{R}_{[i]} \leftarrow (\rho\mathbf{I} + (\mathbf{R}_{[i-1]})^2\mathbf{b}^{\mathrm{T}}\mathbf{q})\,(\rho\mathbf{I} + \mathbf{B})^{-1}.$$

Once the solution to (5.33) is obtained, the steady-state probabilities can be computed via

$$\mathbf{p}_i = \mathbf{x}\mathbf{R}^i, \quad i = 0, 1, \ldots,$$

where $\mathbf{x}$ is the left null vector of $\rho\mathbf{I} - \mathbf{R}\mathbf{B}$, normalized by the constraint

$$\mathbf{x}(\mathbf{I} - \mathbf{R})^{-1}\mathbf{1}^{\mathrm{T}} = 1. \tag{5.34}$$

Finally, the steady-state customer distribution can be obtained via $\pi_i = \mathbf{p}_i\mathbf{1}^{\mathrm{T}}$.

In contrast to the memoryless service of fractal input, further closed form analysis is difficult because the solution to the matrix quadratic equation (5.33) is not rank 1. Nev-

ertheless, using the above techniques, we can numerically obtain the steady-state customer distribution for a queueing system consisting of a heavy-tail power-law server, servicing memoryless input. Results for the specific case where the input arrival is $\lambda = 1 \times 10^{-5}$, the shape parameter of the service is $\gamma = 1.8$. and the reference service rate is $\mu = 1.0$ are depicted in Fig. 5-12, which plots the resulting steady-state distribution on a log-log scale Forty dyadic scales were used in the computation. In contrast to the results of Section 5.4, the most notable feature of this set of results is the slow power-law decay, which suggests high likelihood of buffer overflow in this system. Empirically, we have observed similar trends for other combinations of service shape parameters and input rates. A reason for this is the substantial fraction of customers requesting long services, which leads to accumulation of customers in the systems, and ultimately to congestion. Flow control for such queueing systems, therefore appears to be an important problem, and is a main focus of Chapter 6.

Finally, to verify the validity of our analysis, we compare simulations of these queueing systems to compare against our theoretical predictions. We again employ the next-event time advance method of [28] for the simulations. As we did in Section 5.4, we use exact power-law random variables and exponential random variables for simulations. For queueing analysis, we first obtain a finite-scale representation for the power-law service, with the parameters extracted via the noise-free parameter estimator of Section 4.1. The analysis methods of this section were then applied using the estimated multiscale model.

Two sets of simulation results are plotted in Fig. 5-13. For the top curves, the input arrival rate was set at $\lambda = 2 \times 10^{-7}$. while the shape parameter of the service duration was $\gamma = 1.5$. The solid curve denotes the steady-state probabilities obtained from simulations, which employed 2×10^6 arrivals, and the dashed curve represents the theoretical prediction. For the bottom curves, the arrival rate was set at $\lambda = 1 \times 10^{-5}$, while the shape parameter of the service time was $\gamma = 1.8$. The simulation was conducted using 5×10^6 arrivals. Again, the simulation results are shown as the solid curve while the theoretical one is the dashed curve. Agreement between our prediction and simulation results is again close.

Thus, we have demonstrated our multiscale framework as a feasible basis for certain queueing analysis. We will pursue this framework further in the following chapter to formulate queueing design and management strategies for a set of fractal queues.

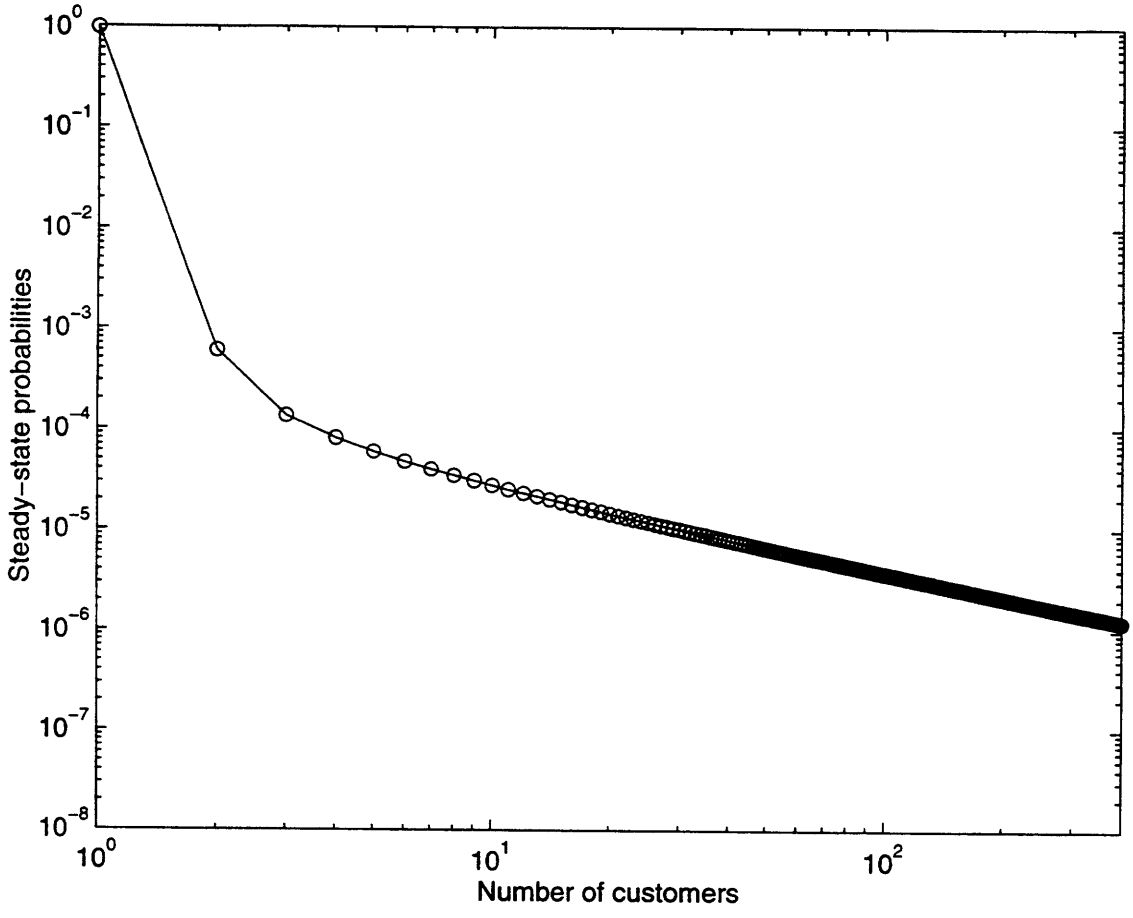


Figure 5-12: *First 400 terms in the steady-state customer distribution corresponding to a single-server queueing system with memoryless input and power-law holding time. The power-law service is modeled with a 40-scale dyadic representation, with* $\gamma = 1.8$, $\mu = 1$, *while the service rate is* $\mu = 0.00001$.

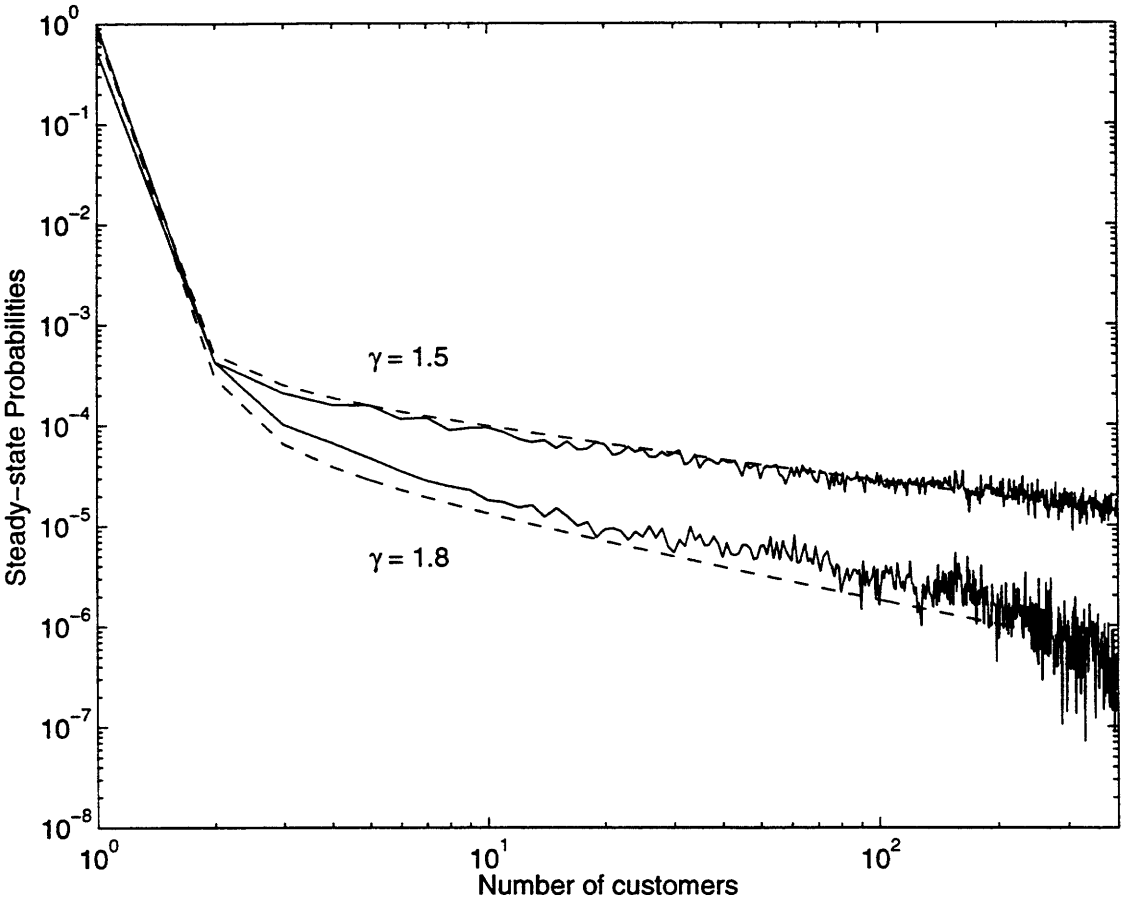


Figure 5-13: *Comparison of simulated and theoretical steady-state customer distribution for a queueing system with power-law holding time servicing memoryless input. The solid curves represent results of discrete-event simulation of the queueing system. For the case* $\gamma = 1.5$, 2×10^6 *arrivals were used in the simulation, while for the case* $\gamma = 1.8$, 5×10^6 *arrivals were used. The dashed curves represent theoretically-predicted distributions for the corresponding queueing scenarios. For the case* $\gamma = 1.5$, *the input rate is* $\lambda = 2 \times 10^{-7}$, *while for the case* $\gamma = 1.8$, *the input rate is* $\lambda = 1 \times 10^{-5}$.

Chapter 6

Optimal Control of Fractal Queueing Systems

Queueing analysis results of the type presented in Sections 5.4 and 5.5 are fundamental for assessing and predicting performance of predetermined queueing setups. In many realistic scenarios, however, various aspects of a general queueing system are typically controllable, often in real time. For example, controllability of service rate is quite feasible. In various service industries, manpower can be allocated rather flexibly. In the context of communications engineering, a number of schemes have been proposed for versatile allocation of bandwidth, ranging from the fast packet switching networks [1] to the flexible assignment of multilevel trunks and trunk groups [20]. In another direction, control of customer arrival rate, or flow control, is also often possible. Throttling of input, or reduction of arrival rate, can be implemented via admission toll, service rejection, or input re-routing.

By manipulating these controls, a queueing policy generally attempts to achieve several typical objectives. One of the more important goals is the assurance of service quality, which is often measured in terms of the waiting time experienced by each customer. Under a first-in-first-out (FIFO) discipline, the expected waiting time of an incoming customer is proportional to the queue length seen. Thus, it is reasonable to quantify service quality in terms of a holding cost, which is a function of the queue length. While a linear holding cost corresponds appropriately to the cost of waiting, several other functional forms might also be natural. For example, where buffering space is precious, the holding cost might be defined to severely penalize long queues. In addition to the holding cost, operation of

the server might induce a cost. Typically, more efficient service requires higher costs as more or better servers are involved. Likewise, input throttling often induces a penalty for rejecting, or redirecting input. Thus optimization is often formulated in terms of overall cost minimization.

In Sections 5.4 and 5.5, we have established that our multiscale framework facilitates the analysis of queueing systems involving fractal input or service. In the present chapter, we show that the same framework is convenient for systematic and rigorous design of queueing control policies as well. In particular, our multiscale birth-and-death process of Fig. 5-7 and its transpose of Fig. 5-11 furnish the appropriate foundation for dynamic programming and Markov decision analysis involving fractal queues, which formalizes policy development for these systems. For the sake of discussion, we shall apply our techniques to two problems, one on optimal server control and the other on optimal flow control. While these problems involve very specific modes of control and optimization objectives, our methodologies appear to be more broadly applicable to a variety of other queueing control problems as well. As will be apparent, in contrast to optimal control of memoryless queues driven by Poisson inputs, a key feature of the algorithms developed here is that they heavily exploit past history of the system to enhance performance.

6.1 Background

Queueing theory has traditionally specialized in descriptive studies of systems, and has generated a taxonomy of results for various queueing arrangements. Queueing control, however, has been left largely ad hoc until the last two decades, which witnessed the involvement of dynamic programming techniques in formulating mathematically-sound queueing policies. Generally speaking, the field of dynamic programming or Markov decision analysis was pioneered by the work of researchers such as Howard [19] and Bellman [4]. Central to a dynamic programming problem is a controllable system whose dynamics are Markovian in nature, and whose operation either requires a certain cost and/or generates a certain reward. The typical goal of the programming is then the long-range cost-minimization or reward-maximization via manipulation of the system dynamics. Dynamic programming techniques generally exploit the Markovian nature of the system to recast the long-run optimization problem into one-step optimizations.

In the particular application of queueing control, the first works have appeared in the 1970s. Two directions have since been of particular interest—dynamic flow control and server control. In [13], Ephremides, et al., present an optimal solution to a problem of the former class, although the problem more closely resembles a primitive routing problem. Two identical memoryless queueing systems are present in their simple network, and the problem involves optimal routing through the network, where optimality is defined in terms of the minimization of residual work load in the network, accumulated throughout its operation. Ephremides, et al., are able to give closed-form flow control policies for both the perfect-information and zero-information cases. They show that provided the states of the two queues are completely known, the optimal policy is not surprisingly the send-to-shorter-queue policy. On the other hand, if the states of the queues are completely unknown, the optimal policy is again not surprisingly a round robin approach, which alternates arrivals between the queues.

An equally important problem of flow control, involving optimal admission to a memoryless queueing system, is considered by Stidham in his classic paper [39]. The problem is formulated in a G/M/1 type setting[1], whereby the server is memoryless while the input has general statistics. Hard decisions for arrival admission are sought. Specifically, for every customer arriving at the system, the request is either admitted and granted service, or rejected entirely. The objective function in this problem consists of two components, a cost for holding and a reward for each admitted job. Stidham shows that the admission strategy which maximizes the expected long-run net benefit is a form of thresholding, where customers are admitted if and only if there are at most a certain number of customers in the queue. A similar problem is considered by Doshi in [11], where the queueing system now provides general service, while the arrival stream is memoryless. Again, at least among hard decision strategies, a thresholding scheme is proved optimal.

While flow control is particularly useful in networking scenarios where input throttling is acceptable, server control is more suitable for many multiplexed environments, where process cycles can be allocated with agility. Mitchell considers the problem of service rate control for a M/G/1 queueing system [34], whose analysis is subsequently refined by Doshi

[1]We adopt notations from general queueing theory: in the symbol $\cdot/\cdot/\cdot$, the first entry describes the input statistics, the second entry describe service statistics, and the third denotes the number of servers. For the first two entries, "G" represents general statistics, while "M" represents memoryless arrival or service

in [12]. As in many other queueing control problems, the objective function in this problem is composed of two parts, the holding cost incurred by buffer occupancy, and the service rate cost. The key result is that under very general conditions, the optimal service rate is a monotonic function of residual workload [12], implying that the busier system requires more efficient service.

Fundamental results for M/M/1 queueing control have also been discussed in depth in [6], where both server and flow control are addressed. For server control, the arrival rate is viewed as exogenous, while the service rate is the design parameter. The control objective is then centered around minimization of a realistic long-term cost of two components, a holding cost and a service rate cost. To account for inflation of cost over time, a discount factor is also included in the objective function. The strategy for solving this problem involves two main steps. First, a rate uniformization synchronizes all transitions to a common clock. Next, the uniformized continuous-time Markov decision problem is converted into a discrete-time problem to which powerful techniques such as the value iteration method can be applied. As will be apparent, our treatment of the fractal queueing control problems will be modeled closely after this solution technique. For the M/M/1 server control problem, it is proved that the optimal policy depends only on the current state of the queue, i.e., the number of customers in system [6]. As a key result, it is also shown that with convexity assumed for the holding cost, and continuity for the service rate cost, the optimal service rate for the M/M/1 queue increases monotonically with the number of customers in the system.

In contrast to the policies developed in [39] and [11], soft decisions are pursued for the flow control problem of [6]. In this case, the service rate is assumed fixed, while arrival rate can be arbitrarily selected over a continuous range of values. Thus, one interpretation is that instead of accepted or rejected with certainty, each customer is admitted with some probability. The cost function is defined similarly as in the server control problem, and is made up of a holding cost and a penalty for input throttling. As in the server control problem, optimal strategy for this problem depends only on the current state of the system, and the results agree with intuition: with convexity assumed for the holding cost and continuity assumed for the throttling cost, customer arrivals are increasingly deterred as the queue becomes more crowded.

6.2 A Dynamic Server Control Problem

In Section 5.4, we studied the equilibrium customer distribution of a fixed-rate memoryless queueing system servicing fractal renewal process input. We now allow the service rate to vary, and apply dynamic programming techniques to determine the best strategy for selecting this service rate for the system. To capture long-term behavior of the system, we shall formulate this server control problem as an infinite-horizon dynamic programming problem [6]. As a preliminary observation, we expect substantial differences in behavior from the Poisson case. As just one example, due to the peculiar growing wait phenomenon described in Section 2.4.4, we expect the demand on the server to continuously lower as long as no new customers emerge.

Formally, we model the service rate as a variable μ, which can be arbitrarily chosen from the range $[0, \overline{\mu}]$ at all times, where the maximum achievable rate $\overline{\mu}$ is chosen sufficiently large that overloading is avoided. Intelligent server control is then founded on a realistically-defined cost function of two components. The first of these is associated with the operation of the server, and is defined as a continuous, nondecreasing function of the service rate, $c(\mu)$. This cost may come about due to machine constraints, bandwidth limitations, depending on the application at hand. For convenience, an idle server is assumed to inflict zero costs, i.e., $c(0) = 0$. A second component of the cost function is a holding cost $h(i)$, which is a nondecreasing function of the number of customers i in the system. This cost is realistic in systems with limited buffering resources, and, in addition, reflects the cost of delay under the first-in-first-out discipline. For convenience, an empty system is assumed cost-free, i.e., $h(0) = 0$. The overall cost is accumulated over time, with a discount rate β that weights future costs relative to the present. More precisely, a dollar expended after a period of dt will cost $1 - \beta dt$ dollars now. Combining these components, the cost to be minimized assumes the form

$$J = E\left[\int_0^\infty e^{-\beta t}\left[h(i(t)) + c(\mu(t))\right] dt\right] \tag{6.1}$$

where $i(t)$ is the number of customers in the queueing system at time t, and $\mu(t)$ is the service rate at time t.

Our approach to this problem is again based on our multiscale framework. In particular, the state space for this queueing system has a form almost identical to the multiscale birth-

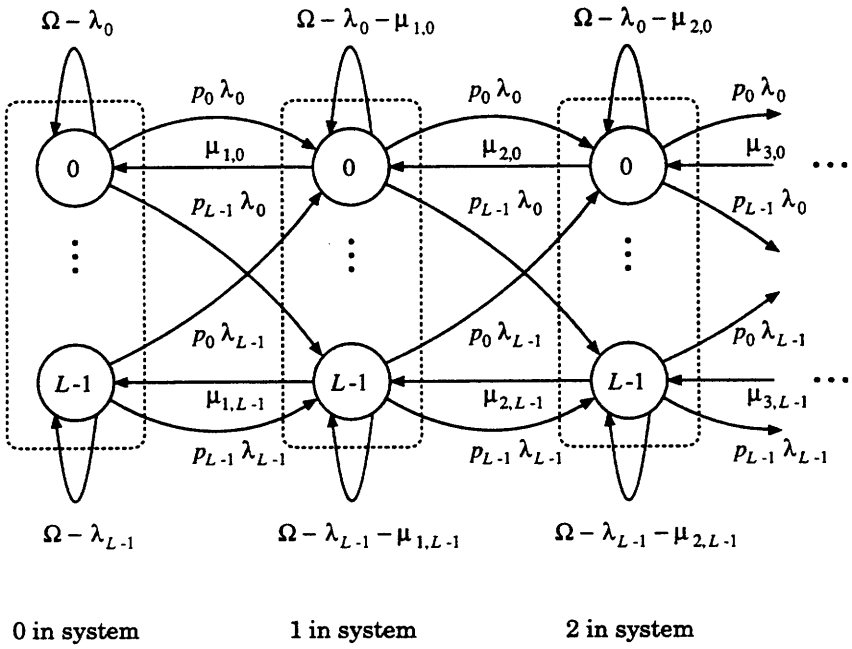


Figure 6-1: *The continuous-time Markov process employed in server control policy development. This process is obtained from the multiscale birth-and-death process of Fig. 5.7 via rate uniformization. Self-transitions are added such that the transition rate from each state is* Ω, *while preserving system dynamics. As another change from Fig. 5.7, the service rates are now state-dependent.*

and-death process depicted in Fig. 5-7, with the exception of the controllable service rate μ. For simplicity, the service rate in our design will be dependent only on the state and will be denoted by $\mu_{i,j}$; more general scenarios are conceivable but do not seem more useful. Thus, our control policy is multiscale in nature, having dependence on scales. While it may be unrealistic to assume the controller has access to perfect scale information, we shall see that reasonable scale estimates can often be obtained.

While transition rate of this Markov process is state-dependent, a preliminary simplification can be achieved via a rate uniformization step which, in essence, synchronizes all transitions to a common clock. This yields the equivalent Markov process shown in Fig. 6-1, for which the common transition rate is $\Omega \stackrel{\Delta}{=} \lambda + \overline{\mu}$, the maximum transition rate in the original Markov process. Note that self-transitions have to be added to preserve system dynamics.

Based on the uniformized process, we can rewrite the cost in (6.1) as

$$J = E\left[\sum_{n=0}^{\infty}\int_{t_n}^{t_{n+1}} e^{-\beta t}dt\,(h(i_n)+c(\mu_n))\right], \tag{6.2}$$

where $\{t_n; n=0,1,\dots\}$ are the transition epochs of the uniformized process, and i_n and μ_n are respectively the number of customers and the service rate during the nth interval, $(t_n, t_{n+1}]$. Note that here we have exploited the assumption that the service rate is dependent only on the state. Carrying out the integration in (6.2) and taking expectation, we get that the cost is

$$\begin{aligned} J &= \sum_{n=0}^{\infty} E\left[\frac{e^{-\beta t_n}-e^{-\beta t_{n+1}}}{\beta}\right] E\left[h(i_n)+c(\mu_n)\right] \\ &= \frac{1}{\beta+\Omega}\sum_{n=0}^{\infty}\frac{\Omega^n}{(\beta+\Omega)^n}\mathrm{E}\{h(i_n)+c(\mu_n)\}, \end{aligned} \tag{6.3}$$

where Ω is the common transition rate in the uniformized process. Hence, the continuous-time dynamic programming problem has been converted to a discete-time problem. Specifically, the underlying Markov chain of this discrete-time dynamic programming problem has the same connectivity as the continuous-time process of Fig. 6-1, with probability of each transition equal to r/Ω, where r is the rate of the corresponding transition in the continuous-time process. The holding and service rate costs are respectively, $h(\cdot)/(\beta+\Omega)$ and $c(\cdot)/(\beta+\Omega)$, while the discount rate is $\Omega/(\beta+\Omega)$.

Because of the Markovian nature of the discretized system, the long-term cost (6.3) is most easily accumulated step-by-step. This idea is formalized in terms of the Bellman equations, which govern the optimal stationary decision-making [6]. Specifically, for the problem at hand, the Bellman equations are

$$V_{0,j} = \frac{1}{\beta+\Omega}\left\{\frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{1,j'} + \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right)V_{0,j}\right\} \tag{6.4}$$

$$\begin{aligned} V_{i,j} = \min_{\mu\in[0,\bar{\mu}]}\frac{1}{\beta+\Omega}\Bigg\{&h(i)+c(\mu)+\frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i+1,j'}+\mu V_{i-1,j} \\ &+\left(\Omega-\frac{\lambda}{\eta^{j-1}}-\mu\right)V_{i,j}\Bigg\}, \quad i=1,2,\dots \end{aligned} \tag{6.5}$$

where $V_{i,j}$ is the total accumulated cost if the system commences at state (i,j). The optimal

stationary service rate $\mu^*_{i,j}$ when the system is in state (i,j) is then the minimizing rate for the corresponding equation of (6.5). If the minimum is achieved at multiple values of μ, we pick the slowest rate. Note that from (6.4), it is clear that in the zeroth superstate, the choice of service rate has no influence on the future. Hence the decision should be made to optimize $c(\mu)$, implying that $\mu_{0,j} = 0$, for every j. To obtain the optimal service rates for the other cases, we seek solution to (6.5). While this system of equations is analytically intractable, they can be solved numerically via the value iteration method [6]. More precisely, we iterate the system

$$V_{0,j}(k+1) = \frac{1}{\beta+\Omega}\left\{\frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{1,j'}(k) + \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right)V_{0,j}(k)\right\} \tag{6.6}$$

$$\begin{aligned} V_{i,j}(k+1) = \min_{\mu\in[0,\bar{\mu}]}\frac{1}{\beta+\Omega}\Bigg\{h(i) + c(\mu) + \frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i+1,j'}(k) + \mu V_{i-1,j}(k) \\ + \left(\Omega - \frac{\lambda}{\eta^{j-1}} - \mu\right)V_{i,j}(k)\Bigg\}, \quad i = 1,2,\ldots, \end{aligned} \tag{6.7}$$

beginning with $V_{i,j}(0) = 0$ for every i and j.

Figs. 6-2 and 6-3 show examples of optimal stationary service rates $\mu^*_{i,j}$ obtained with the value iteration method. Specifically, we iterate (6.6) and (6.7) until the absolute change in $\mu^*_{i,j}$ is small, in this case less than $\epsilon = 1 \times 10^{-9}$ for all i,j. Various combinations of the parameters are considered, and each case is depicted in one of the graphs. Each curve in the graphs represents the optimal service rate as a function of scales, with the number of customers fixed. In these examples, a simple quadratic service rate cost $c(\mu)$ is chosen, i.e.,

$$c(\mu) = c_0\mu^2$$

where c_0 is a positive constant. For the examples of Fig. 6-2, the holding cost is of the linear form

$$h(i) = h_0 i,$$

while for Fig. 6-3, this cost is quadratic

$$h(i) = h_0 i^2.$$

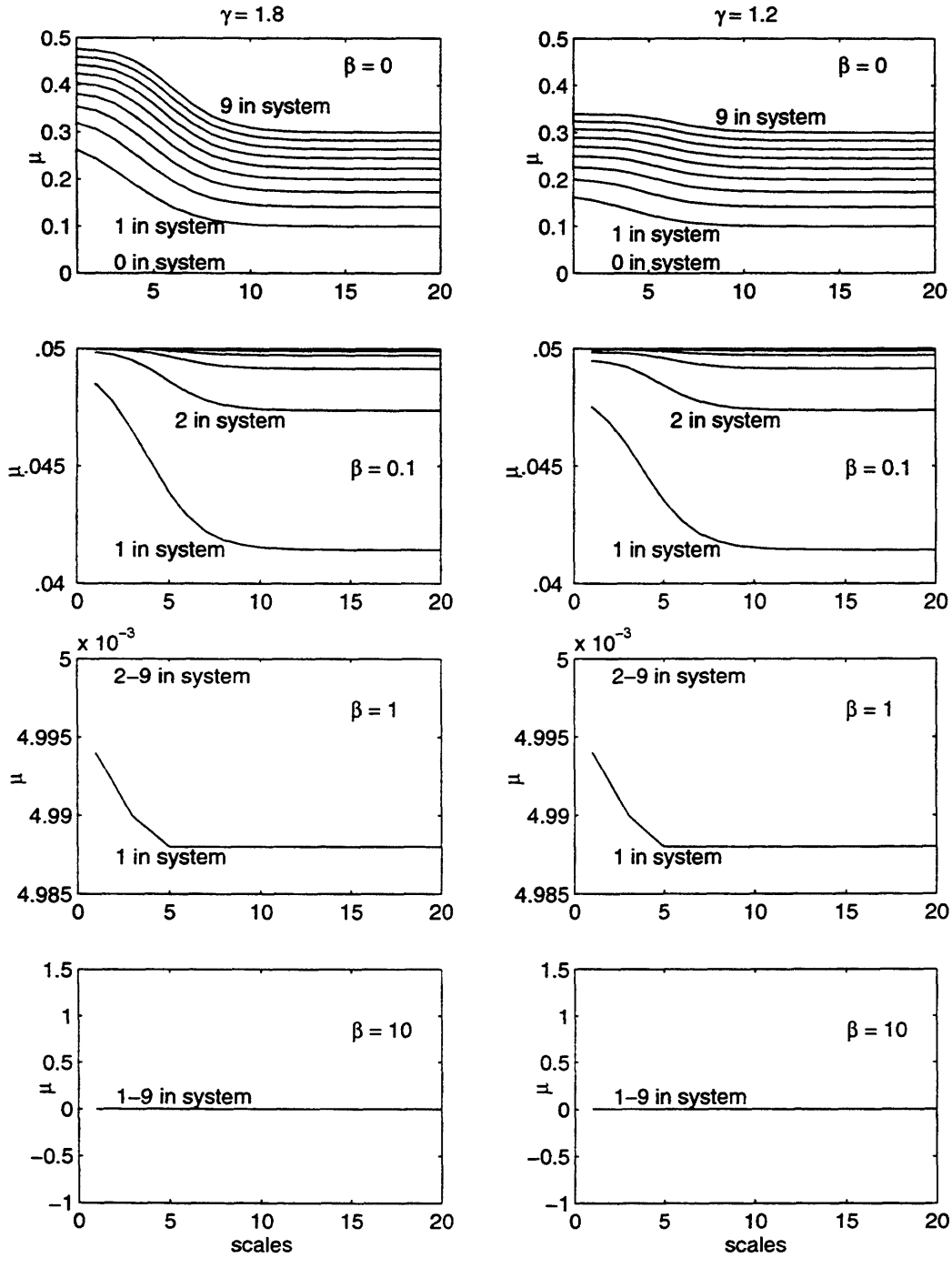


Figure 6-2: *Perfect-information optimal stationary policy for queueing system with fractal customer arrivals. Optimal service rate obtained via the value iteration method (Eqns. (6.5),(6.6)), based on a twenty-scale dyadic multiscale birth-and-death framework:* $\lambda = 1$, $L = 20$, $\eta = 2$. *Cases where* $\gamma = 1.8$ *and* $\gamma = 1.2$ *are considered, with discount rate* $\beta = 0, 0.1, 1, 10$. *The holding costs was* $h(i) = 0.01i$, *and the service rate cost was* $c(\mu) = \mu^2$. *The optimal service rates for 0-9 customers in the system are shown. Note that for 0 customer in system, optimal service rate is identically 0.*

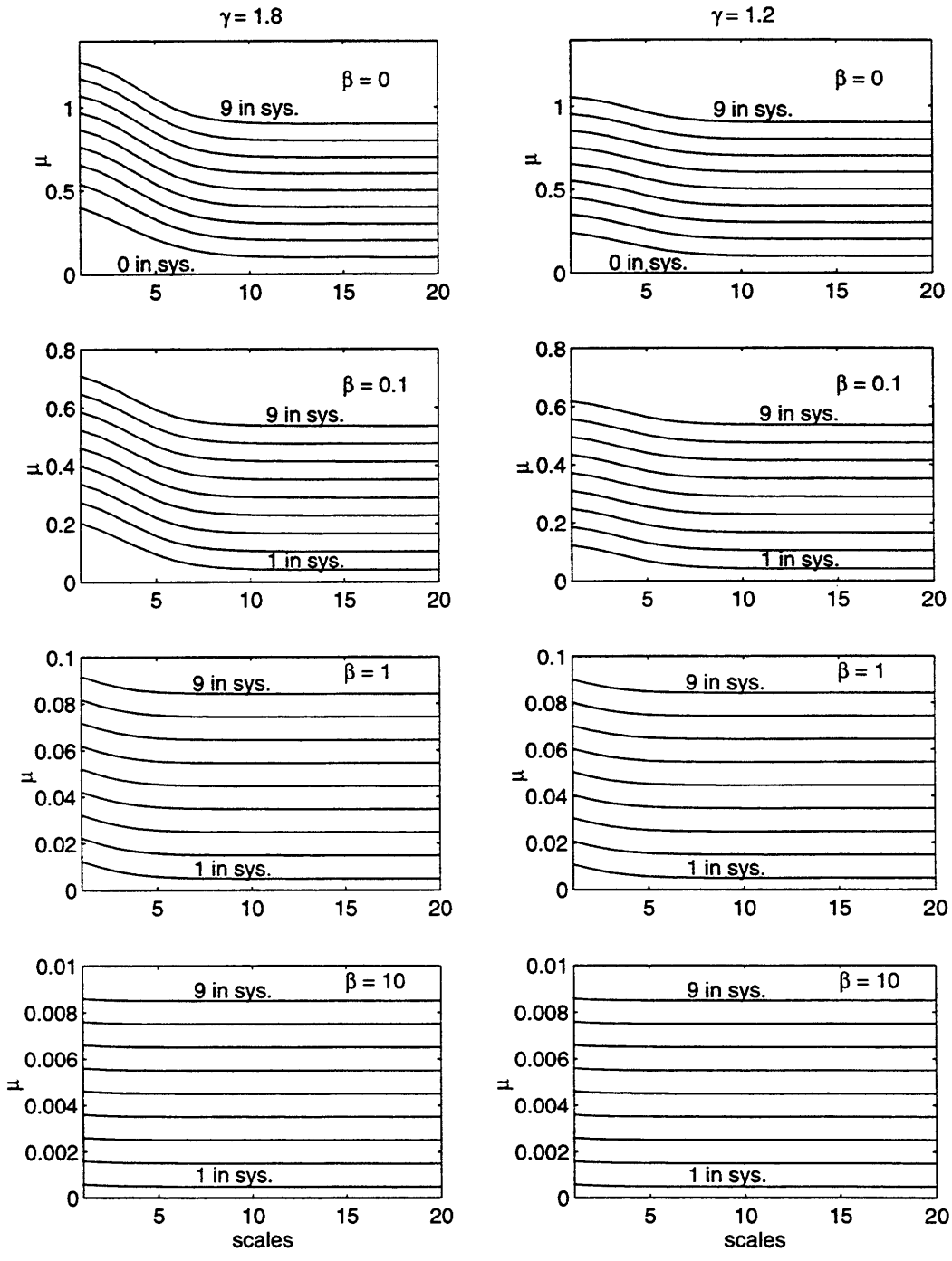


Figure 6-3: *Perfect-information optimal stationary policy for queueing system with fractal customer arrivals. Optimal service rate obtained via the value iteration method (Eqns. (6.5),(6.6)), based on a twenty-scale dyadic multiscale birth-and-death framework:* $\lambda = 1$, $L = 20$, $\eta = 2$. *Cases where* $\gamma = 1.8$ *and* $\gamma = 1.2$ *are considered, with discount rate* $\beta = 0, 0.1, 1, 10$. *The holding costs was* $h(i) = 0.01i^2$, *and the service rate cost was* $c(\mu) = \mu^2$. *The optimal service rates for 0-9 customers in the system are shown. Note that for 0 customer in system, optimal service rate is identically 0.*

Other than varying the algebraic form of the cost functions, simple tuning of c_0 and h_0 can effect tradeoffs between the two costs. For our computations, we set these constants at $c_0 = 1, h_0 = 0.01$. In the left columns of Figs. 6-2 and 6-3, we have plotted the policies for an input process with shape parameter $\gamma = 1.8$, while on the right, the shape parameter is $\gamma = 1.2$. Several values of the discount rate β are also considered. As before, the computations are performed for the 20-scale dyadic case, with $\lambda = 1$.

Several features are common to the cases considered. First, for a fixed scale, the optimal service rate is monotonic in the number of customers: the busier the system, the higher the optimal service rate. In fact, we can show that this holds more broadly for any monotonic convex holding cost $h(i)$; Appendix E.1 furnishes a formal proof. On the other hand, for a fixed queue length, we observe that the optimal service rate decreases for higher scale indices. Thus, if the next arrival is expected to be distant, the service rate can be reduced without burdening the future. Also, the discount rate plays a crucial role in determining the shape of these plots. In particular, if the future is important, i.e., the discount rate is low, then the strategies are sensitive to the scale of the current interarrival. Also, in these cases, the service rates tend to be higher in general. In words, this reflects an expenditure on current service as an investment for a less costly future. To see the effects of high discount rates, we note that for the case $\gamma = 1.8$, $\beta = 10.0$, and $h(i) = 0.01i$ (bottom-left graph in Fig. 6-2), the optimal service rates are a mere 5×10^{-4} throughout all states (except the zeroth superstate)!

Comparing the policies for the different shape parameters, we note that the lower γ cases appear to be less demanding on the server. We attribute this to the fact that heavier tails of the corresponding interarrival densities suggest that prospective customers are more distant, and hence, have less impact on the present decision-making.

The policies given in Figs. 6-2 and 6-3 require perfect information. In general, however, the queue controller does not have knowledge of the scale of the next interarrival, but has to estimate it. Based on our multiscale representation, we can formulate a maximum-a-posteriori (MAP) estimator for the scale based on the time elapsed since the last arrival, t. Formally, given that the interarrival is at least t, we choose the scale j which maximizes

$$L(j) = p_j \exp(-\lambda_j t).$$

Now, we observe that

$$\frac{L(j+1)}{L(j)} = q \exp\left(-(\lambda_{j+1} - \lambda_j)t\right),$$

which is monotonically increasing in t. Thus, $L(j+1) > L(j)$ if and only if

$$t > -\eta^j \log_e(q) \frac{\eta}{(\eta-1)\lambda},$$

where η is, as before, the scale increment. Thus, our scale estimator is of the form

$$\hat{j}(t) = j,$$

where j is the unique integer such that

$$-\eta^{j-1} \log_e(q) \frac{\eta}{(\eta-1)\lambda} < t < -\eta^j \log_e(q) \frac{\eta}{(\eta-1)\lambda}.$$

Hence, the optimal policy based on partial information is related to the perfect-information controller via a simple warping of the horizontal axis. Fig. 6-4 shows such a warping of the top left graph in Fig. 6-2. To apply the prescription of Fig. 6-4, a controller follows the same curve as time progresses, until a customer enters or leaves the system. For a customer arrival, the controller will promote to the beginning of next higher level curve. For a departure, the controller will just drop vertically to the next lower level curve. From the shape of these curves, it is apparent that for servicing of fractal renewal input, the optimal queueing policy involves continuous decrease of service rate whenever no new customer arrives. It is important to stress that this behavior is fundamentally different from the optimal server control for the M/M/1 queue given in [6], where the optimal service rate remains constant between customer arrivals or departures.

6.2.1 Simulations

As noted earlier, the performance of our multiscale server controller is enhanced by exploiting the history of the arrival process. To see the gain in this approach, we compare our control algorithm with schemes that do not make use of past information, particularly, an optimal controller for an M/M/1 system [6], as well as some of its variations.

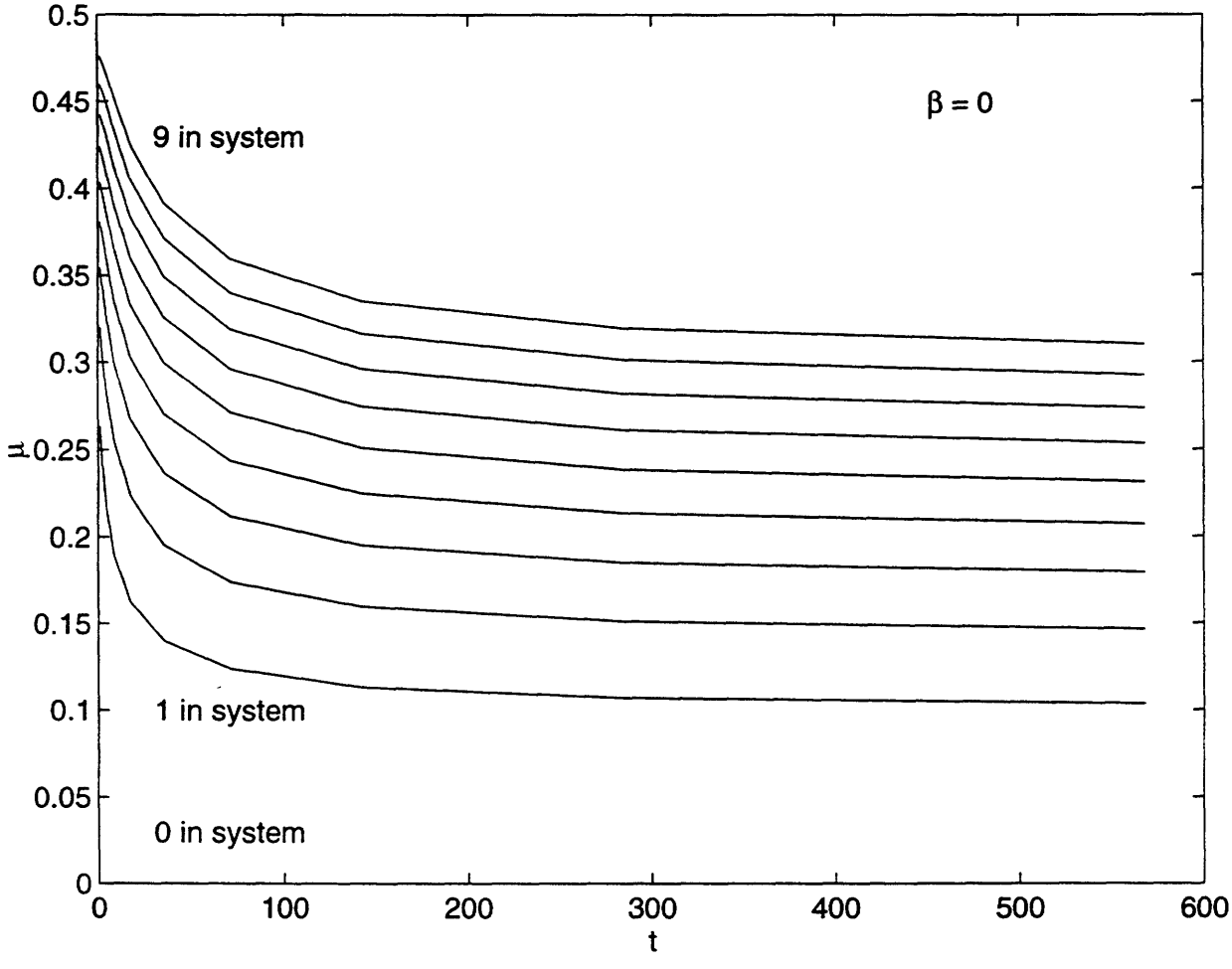


Figure 6-4: *Partial-information control policy for queueing system driven by fractal customer arrivals; the scale information is withheld from the controller. The depicted policy is for the case* $\gamma = 1.8, c(\mu) = \mu^2, h(i) = 0.01i$, *and is obtained via a rewarping of the horizontal axis in the top-left graph of Fig. 6.2.*

The distinction between the two schemes is most pronounced in cases where the optimal multiscale server rate $\mu^*_{i,j}$ varies greatly over scales. From Figs. 6-2 and 6-3, sensitivity of the server rates is greatest for non-discounted cases, i.e., $\beta = 0$. Moreover, the server rate also changes rapidly along scales for large γ. Finally, for cases where the service rate cost is weighted much more heavily than the holding cost, the change in optimal server rate is also more apparent. For our experimentation, therefore, we study the case where $\gamma = 1.8$, $C_0 = 10.0$, $H_0 = 0.01$, and $\beta = 0$.

As in Sections 5.4 and 5.5, we simulate the queue controller with the next-event time advance method of [28]. The input to the queueing system is a fractal renewal process with $\gamma = 1.8$, consisting of 500 000 arrivals. To compute the optimal multiscale queueing policy for servicing this input stream, we first extract the parameters for the arrival process, using the noise-free EM parameter estimation algorithm of Section 4.1. The resulting parameters, γ, λ, and the number of scales are then employed in the value iteration developed in this section. The iteration is terminated when the maximum absolute change in $\mu^*_{i,j}$ is less than $\epsilon = 1 \times 10^{-9}$. For comparison, we also compute the optimal service policy for an M/M/1 queue servicing a traffic stream with the same arrival rate, which is estimated according to $500\,000/S_X[500\,000]$, where $S_X[500\,000]$ denotes the arrival epoch of the 500 000th customer. We follow the value iteration method of [6] to obtain this controller. The optimal service policies for the two cases are shown in Fig. 6-5. As before, we only show the level curves for 0 through 9 customers in the system. The solid curves correspond to the multiscale controller, while the dashed curves are for the M/M/1 controller. As is apparent from the figure, the service rate for the M/M/1 controller is much less than the multiscale service rate for fine scales, while it is clearly higher for coarse scale cases.

The value of the objective function (6.1) is estimated by accumulating the service rate cost and the holding cost until the 500 000th customer leaves the system. As another measure of performance, the average waiting time under a first-in-first-out discipline is also estimated. Both quantities are tabulated in Table 6.1 for various queueing controllers. In terms of the objective function, the M/M/1 controller is substantially inferior. In addition, this controller gives rise to higher delay and a more congested queue. From Fig. 6-5, it is apparent that this controller attempts to lower the total cost by reducing the service rate cost, at the expense of a much higher holding cost. Fig. 6-6 shows the estimated customer distribution for the two cases. The higher holding cost is an indication that the M/M/1

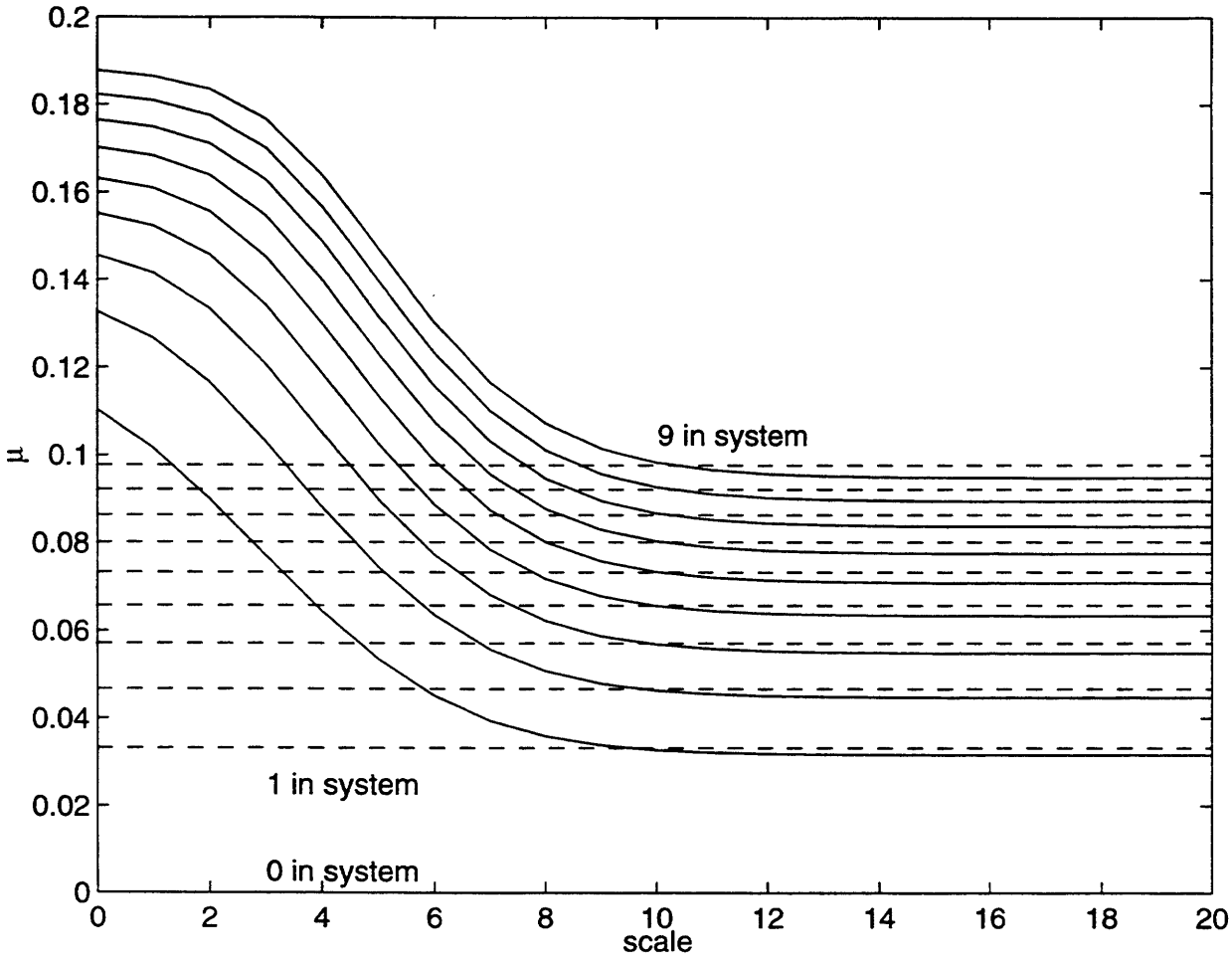


Figure 6-5: *Stationary server control policies for queueing systems servicing fractal renewal process input, for the case* $\gamma = 1.8$, $c(\mu) = 10\mu^2$, $h(i) = 0.01i$, $\beta = 0.0$. *The solid curves are the level graphs for a multiscale server computed with a 20-scale dyadic representation. The dashed curves denote the optimal policy for an M/M/1 queueing controller computed based on the input arrival rate.*

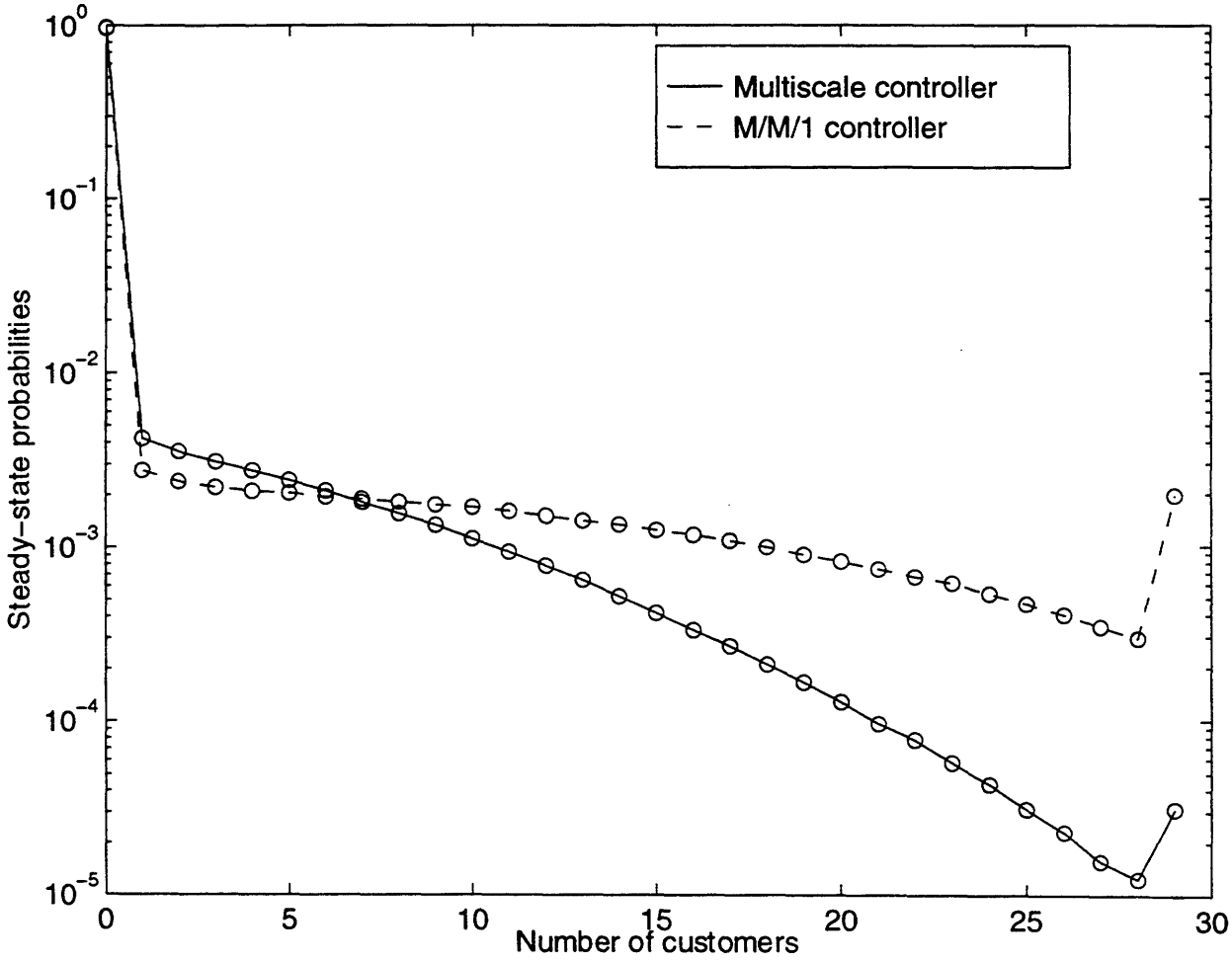


Figure 6-6: *First 30 terms of estimated steady-state customer distribution for a queueing system arising from the optimal multiscale controller and the M/M/1 controller, for the case* $\gamma = 1.8$, $c(\mu) = 10\mu^2$, $h(i) = 0.01i$, $\beta = 0.0$. *500 000 arrivals were used in forming the estimates.*

Table 6.1: *Performance of various queueing server controllers servicing fractal renewal process input*

	Multiscale Controller	M/M/1 Controller	Modified M/M/1 Controller	Modified Fine-Scale Controller
Overall Cost	0.9963e6	1.161e6	1.009e6	1.017e6
Ave. Waiting Time	44.4	115.0	46.8	49.0

controller is more susceptible to congestion. Also, if the service discipline is first-in-first-out, individual customers will suffer longer waiting time, as is apparent from Table 6.1.

The optimal M/M/1 controller requires a low service cost at the expense of very high holding costs. By appropriately increasing the service rates, the holding cost, and possibly the overall cost, may be reduced. A simple way to realize this trade-off is by uniformly scaling up the service rates of the optimal M/M/1 controller. By trial and error, we have tuned the controller so that its service rate cost in the same simulation is comparable with that of the multiscale controller. The resulting controller is depicted in Fig. 6-7, alongside the optimal multiscale controller. Although it is the best controller obtainable by scaling the M/M/1 controller the arrangement is nevertheless still suboptimal. Specifically, from the third column of Table 6.1, it is apparent that the average waiting time is still higher. Also, as is apparent from Fig. 6-7, the modified M/M/1 controller is more demanding to implement; the peak service rate required is substantially higher than the multiscale controller.

Fig. 6-5 suggests that the M/M/1 controller approaches the multiscale controller in the coarse scales. As an alternative, we could devise a system that operates according to the fine-scale service rate. Similar to the modified M/M/1 controller, we scale the service rates to obtain the best controller in terms of the objective function. By trial and error, we have tuned this controller to yield a service rate cost nearly equal the optimal multiscale controller, and have plotted the resulting policy in Fig. 6-8. Nevertheless, the performance of this controller is still inferior to the multiscale scheme. From Fig. 6-8, it is clear that in order to lower the overall service rate cost, fine scale interarrivals are not served as efficiently as possible. This is reflected in the longer average waiting time. However, this server does have a number of desirable properties such as its simplicity for implementation as well as low peak-rate requirement.

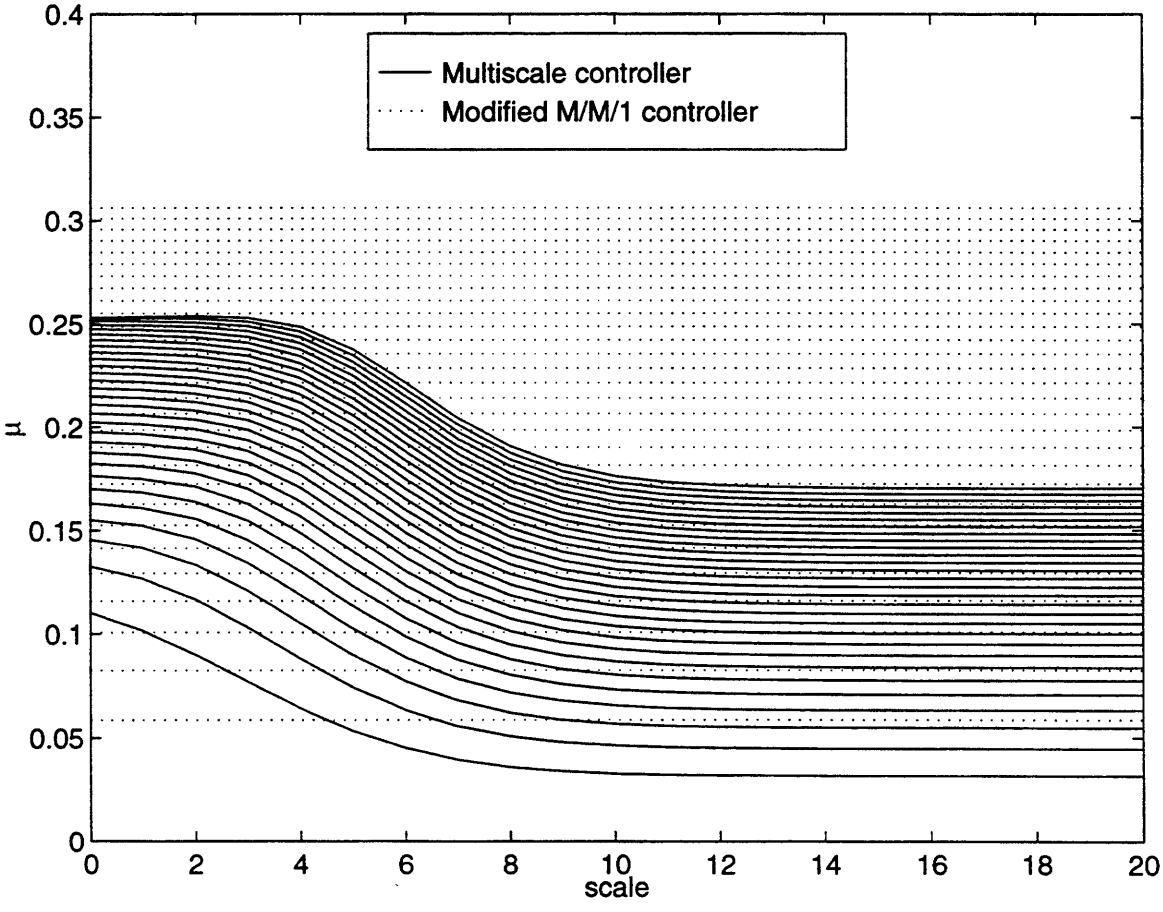


Figure 6-7: *Stationary server control policies for queueing systems servicing fractal renewal process input, for the case* $\gamma = 1.8$, $c(\mu) = 10\mu^2$, $h(i) = 0.01i$, $\beta = 0.0$. *The solid curves are the level graphs for a multiscale server computed with a 20-scale dyadic representation. The dotted curves denote the policy for the optimized variation of the M/M/1 queueing controller obtained by scaling.*

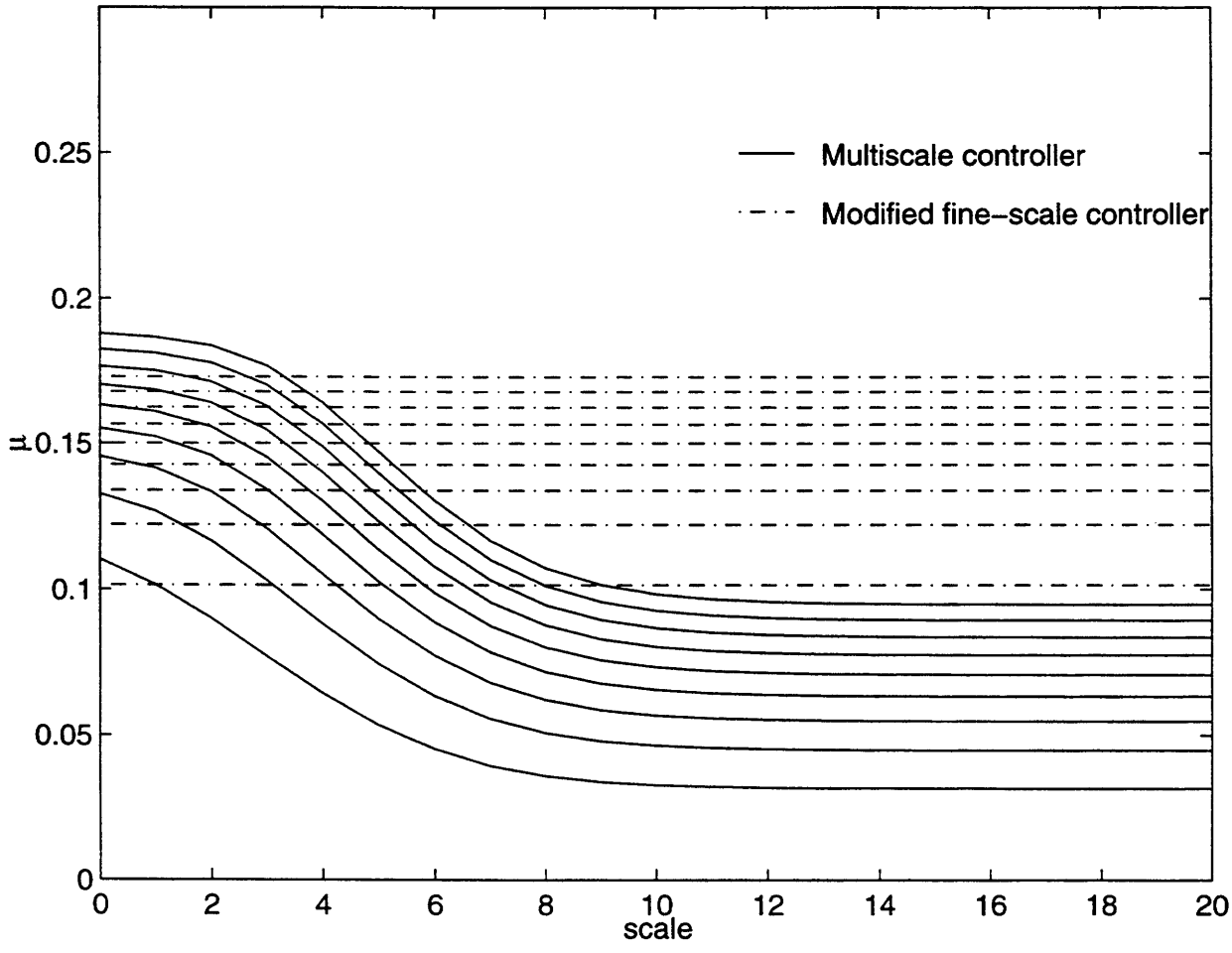


Figure 6-8: *Stationary server control policies for queueing systems servicing fractal renewal process input, for the case* $\gamma = 1.8$, $c(\mu) = 10\mu^2$, $h(i) = 0.01i$, $\beta = 0.0$. *The solid curves are the level graphs for a multiscale server computed with a 20-scale dyadic representation. The dot-dashed curves denote the optimized policy obtained by scaling the fine-scale service rates taken from the multiscale policy.*

6.3 A Dynamic Flow Control Problem

In Section 5.5, we observed that a potential problem in a queueing system with power-law holding time is the heavy-tailed customer distribution, and high likelihood of congestion or buffer overflow. In this section, we explore flow control remedies for these problems. In particular, for a queueing system with fixed power-law service statistics, we formulate policies to regulate the Poisson input traffic to optimize performance.

Formally, our flow control problem is formulated as follows. We consider a queueing system with power-law holding time servicing Poisson input with arrival rate $\overline{\lambda}$. This input can be dynamically throttled so that the actual arrival rate admitted for service is λ, chosen in the range $[0, \overline{\lambda}]$. The cost to be minimized again consists of two components, a holding cost $h(i)$ which reflects buffer occupancy, and a throttling cost $c(\lambda)$, due to loss of customers. We assume that the holding cost is monotonically nondecreasing in the number of customers, with $h(0) = 0$, while the throttling cost is monotonically nonincreasing in the arrival rate, with $c(\overline{\lambda}) = 0$. This total cost is accumulated over time, with a discount factor β included. Specifically, the cost to be optimized assumes the form

$$E\left[\int_0^\infty e^{-\beta t}\left[h(i(t)) + c(\lambda(t))\right] dt\right] \tag{6.8}$$

where $i(t)$ is the number of customers in the queueing system at time t, and $\lambda(t)$ is the customer arrival rate at time t.

Our solution to this problem will again make use of our multiscale framework, in particular, the generalized birth-and-death process of Section 5.5. As before, the control will be state-dependent, hence $\lambda(t)$ reduces to a function of the current state. As another change, we have lumped the zeroth superstate into a single state to obtain a more realistic model; scale of the next service in general cannot be deduced from an empty queue.

Upon rate uniformization, we obtain the equivalent Markov process of Fig. 6-9. Based on the uniformized process, we rewrite the cost (6.8) as

$$J = \mathrm{E}\left\{\sum_{n=0}^{\infty} \int_{t_n}^{t_{n+1}} e^{-\beta t} dt \left(h(i_n) + c(\lambda_n)\right)\right\},$$

where $\{t_n; n = 0, 1, \dots\}$ are the transition instants of the uniformized process, and i_n and λ_n are the number of customers and the arrival rate during the nth interval, $(t_n, t_{n+1}]$.

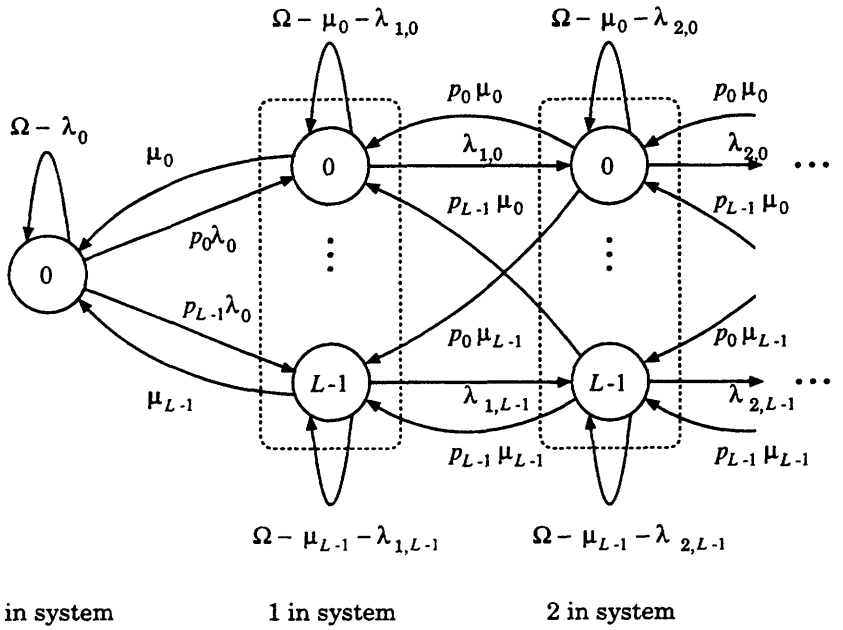


Figure 6-9: *The continuous-time Markov process used in flow control policy development. This process is obtained from the generalized birth-and-death process of Fig. 5.11 via rate uniformization. Self-transitions are added such that the transition rate from each state is* Ω. *As another modification, the arrival rates are now state-dependent.*

Carrying out the integration and taking expectation, we get that the cost is

$$J = \frac{1}{\beta + \Omega} \sum_{n=0}^{\infty} \frac{\Omega^n}{(\beta + \Omega)^n} \mathrm{E}\{h(i_n) + c(\lambda_n)\}.$$

Thus, the problem has again been transformed into a discrete-time Markov decision problem, with the corresponding Bellman equations

$$V_0 = \min_{\lambda \in [0, \bar{\lambda}]} \frac{1}{\beta + \Omega} \left\{ c(\lambda) + \lambda \sum_{j'=1}^{L} \sigma^2 q^{j'-1} V_{1,j'} + (\Omega - \lambda) V_0 \right\} \tag{6.9}$$

$$\begin{aligned} V_{i,j} = \min_{\lambda \in [0, \bar{\lambda}]} \frac{1}{\beta + \Omega} \Bigg\{ & h(i) + c(\lambda) + \frac{\mu}{\eta^{j-1}} \sum_{j'=1}^{L} \sigma^2 q^{j'-1} V_{i-1,j'} + \lambda V_{i+1,j} \\ & + \left(\Omega - \frac{\mu}{\eta^{j-1}} - \lambda \right) V_{i,j} \Bigg\}, \quad i = 1, 2, \ldots \end{aligned} \tag{6.10}$$

where $V_{i,j}$ is the total accumulated cost if the system commences at state (i, j), and the optimal stationary arrival rate $\lambda^*_{i,j}$ when the system is in state (i, j) is the minimizing rate for the corresponding equation of (6.9) and (6.10). If the minimum is achieved at multiple

values of $\lambda_{i,j}$, we pick the fastest rate. In contrast to the optimal server control problem of Section 6.2, the zeroth-order equation (6.9) is no longer trivial, since the throttling decision in the zeroth superstate does have an impact on the future. While this set of equations (6.9) (6.10) are analytically intractable, we can again solve them via the value iteration method.

Figs. 6-10 and 6-11 shows the optimal stationary flow control policy which we have obtained via the value iteration method. Specifically, we iterate the value iteration equations until the maximum absolute change in $\lambda^*_{i,j}$ is less than $\epsilon = 1 \times 10^{-5}$ for all i, j. Various combinations of the parameters are considered, and each case is depicted in one of the graphs. To simplify the minimization at each step, a quadratic throttling cost $c(\lambda)$ is chosen in our examples, i.e.,

$$c(\lambda) = c_0(\overline{\lambda} - \lambda)^2$$

where c_0 is a positive constant. Corresponding to Fig. 6-10, the holding cost $h(i)$ is of the linear form

$$h(i) = h_0 i,$$

while for Fig. 6-11, this cost is quadratic

$$h(i) = h_0 i^2.$$

For our computations, we set these constants at $c_0 = 1, h_0 = 0.01$. In the first columns of Figs. 6-10 and 6-11, we have plotted the policies for power-law service with shape parameter $\gamma = 1.8$, while in the second columns, the shape parameter is $\gamma = 1.2$. Several values of the discount rate β are also considered. As before, the computations are performed for the 20-scale dyadic case, with $\lambda = 1$.

Note again that for a fixed scale, the optimal arrival rate is monotonic in the number of customers; the busier system the higher degree of throttling required. As before, we can analytically establish this monotonicity for convex holding cost $h(i)$. A proof is given in Appendix E.2. On the other hand, for a fixed queue length, the optimal arrival rate decreases for higher scale indices. Thus, if the current job is expected to require long service, the arrival rate should be reduced to prevent high costs for the future. Also, as in

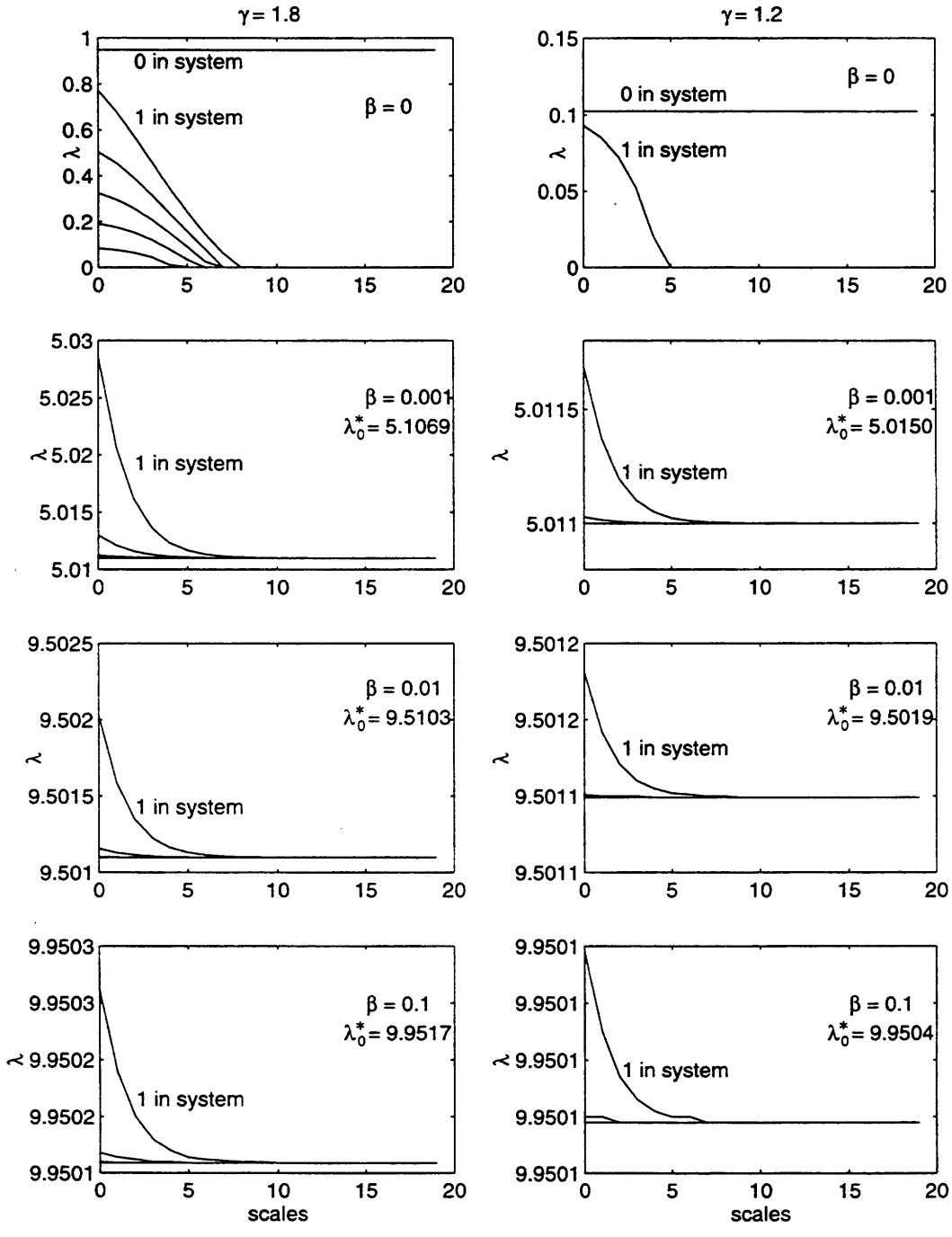


Figure 6-10: *Perfect-information optimal stationary policy for queueing system with power-law holding time and Poisson customer arrivals. Optimal arrival rate obtained via the value iteration method, based on a twenty-scale dyadic multiscale birth-and-death framework:* $\lambda = 1$, $L = 20$, $\eta = 2$. *The holding cost was* $h(i) = 0.01i$, *and the service rate cost was* $c(\lambda) = (\overline{\lambda} - \lambda)^2$.

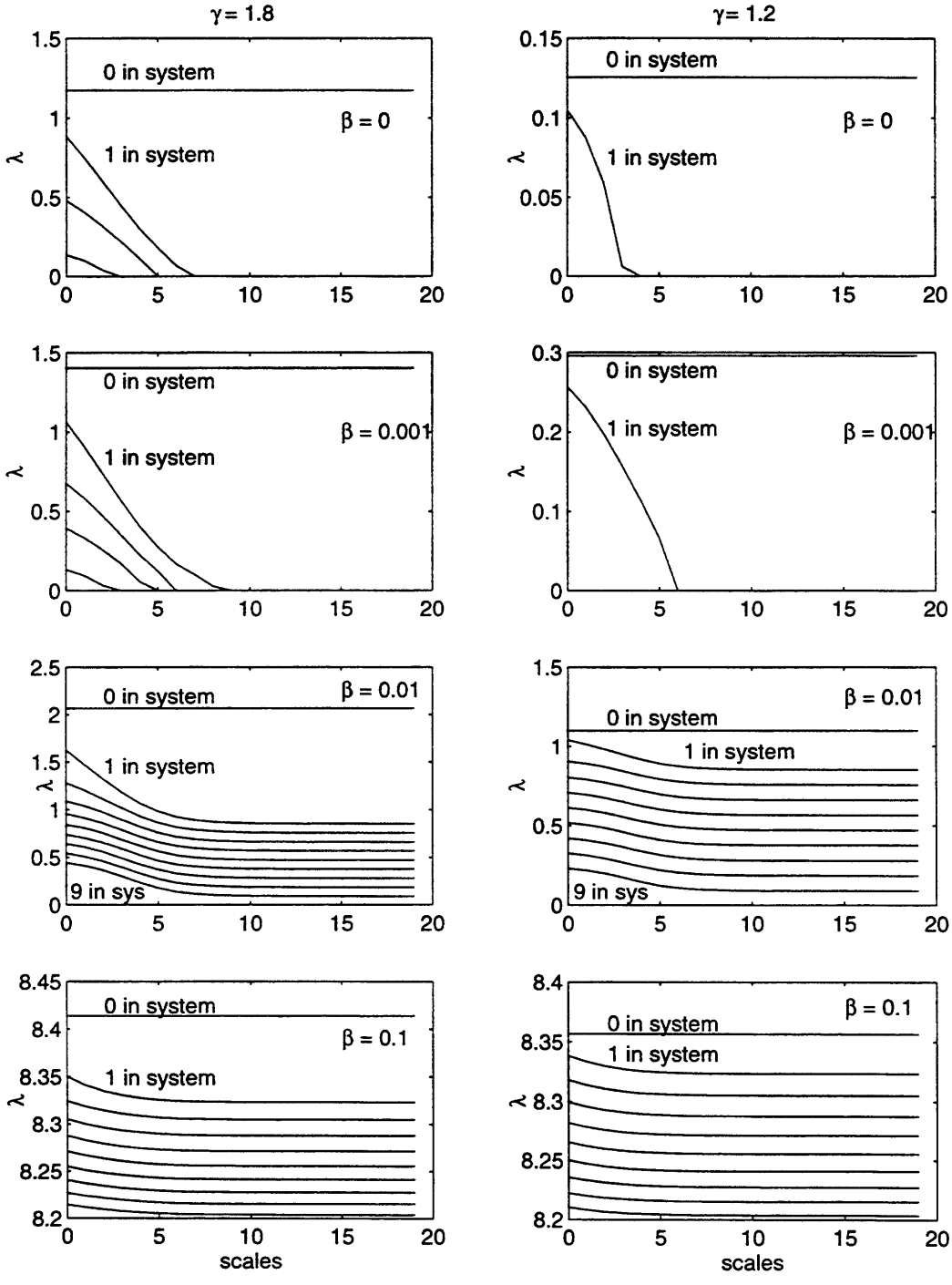


Figure 6-11: *Perfect-information optimal stationary policy for queueing system with power-law holding time and Poisson customer arrivals. Optimal arrival rate obtained via the value iteration method, based on a twenty-scale dyadic multiscale birth-and-death framework:* $\lambda = 1$, $L = 20$, $\eta = 2$. *The holding cost was* $h(i) = 0.01i^2$, *and the service rate cost was* $c(\lambda) = (\overline{\lambda} - \lambda)^2$.

the server control problem of Section 6.2, the discount rate plays a crucial role in determining the shape of these plots. In particular, if the future is important, i.e., the discount rate is low, then the strategies are very sensitive to the scale of the current service. On the other hand, if the discount rate is high, very small difference is observed over the scales. Finally, as the discount rate increases, the future becomes less important, and the input rate can be raised to minimize throttling cost, without burdening the future.

Again, as in the server control problem, practical flow control algorithms must take into account the lack of perfect state information, in particular, the scale of the current service duration is rarely known. Taking the same approach as before, we can formulate an MAP estimator for the scale based on the time elapsed since the last job completion, t. Given that the service time is greater than t, we choose the scale j which maximizes

$$L(j) = p_j \exp(-\mu_j t).$$

We observe that

$$\frac{L(j+1)}{L(j)} = q \exp(-(\mu_{j+1} - \mu_j)t),$$

which is monotonic in t. Thus, $L(j+1) > L(j)$ if and only if

$$t > -\eta^j \log_e(q) \frac{\eta}{(\eta-1)\mu}.$$

Thus, our scale estimator is of the form

$$\hat{j}(t) = j, \quad \text{for } -\eta^{j-1} \log_e(q) \frac{\eta}{(\eta-1)\mu} < t < -\eta^j \log_e(q) \frac{\eta}{(\eta-1)\mu}.$$

This again leads to a simple warping of the horizontal axis. Since monotonicity of the curves is preserved, the optimal flow control policy involves continuously decreasing arrival rate whenever no customer departs. This is in contrast to the optimal flow control for Poisson inputs arriving at a memoryless queueing system, whereby the input rate is held constant between customer arrivals or departures. Thus, like the server control strategies of Section 6.2, past history plays a central role in this queueing scenario as well.

To sum up, in this chapter we have demonstrated that by combining Markov decision

techniques with our multiscale framework, intelligent queueing control algorithms can be obtained in a very straightforward manner. A general conclusion is the importance of the history of the system in determining the present decision. This results are expected to have much potential importance and implications in the design of future networks.

Chapter 7

Contributions and Future Directions

7.1 Contributions of the Thesis

Fractal point processes have long been recognized as important signal models. Indeed, previous studies have identified a host of physical phenomena well captured as fractal point processes. However, fundamental processing and analysis problems involving these signal models have remained largely open, mainly due to the lack of an adequate mathematical framework for addressing them.

As a main contribution of this thesis, we have introduced in Chapter 3 a novel multiscale representation for a class of fractal point processes, known as fractal renewal processes. As it is composed of Poisson constituents, this framework facilitates the studies of fractal renewal processes by allowing exploitation of familiar results from the theory of Poisson processes.

We have applied our multiscale framework to two major areas. First, in the general processing of fractal renewal processes, we have demonstrated in Chapter 3 that this framework readily leads to efficient synthesis of fractal renewal processes. Such synthesis algorithms have potential importance in various shot noise generation, as well as Monte Carlo queueing studies. For analysis of these signals, we have developed several practical estimation algorithms in Chapter 4. Among these is an iterative expectation-maximization (EM) algorithm for the estimation of the fractal dimension. Based on Monte Carlo simulations, we have demonstrated that this algorithm is robust even under rather adverse conditions. In

the same chapter, we have developed an unbiased interarrival estimation algorithm which is optimal in the sense of minimum mean-square error. In addition to providing a means for classification and reconstruction of fractal renewal processes, the parameter and signal estimation algorithms also pave way for the application of algorithms developed later in the thesis, allowing calibration of our model according to real or simulated data.

In the area of networking studies, our multiscale framework has given rise to analysis techniques for fractal traffic undergoing typical networking transformations. Our analysis results for superposition and random erasure of fractal renewal processes from Chapter 5, for example, furnish additional evidence for self-similarity of aggregated traffic arising from multiple users, which have been previously argued both empirically and analytically (through spectral analysis [30], for example). Our multiscale framework also allows us to address queueing problems involving fractal traffic, which capture the activities at individual sites and links within a network. The usefulness of our framework has been demonstrated via computation of the steady-state customer distribution of a memoryless queueing system servicing fractal renewal process traffic, which has been verified by simulated study of the same system. A related queueing scenario which involves the power-law service of memoryless input has also been analyzed using our multiscale framework. Such queueing systems are of importance in the modeling of heavy-tail holding behavior such as in typical telephone usage.

Finally, in Chapter 6, we have applied our multiscale framework together with dynamic programming techniques to generate optimal multiscale queueing policies for systems related to fractal renewal processes, which are of directly practicable value. For servicing of fractal renewal traffic, we obtained multiscale algorithms which allow trade-offs of server rate requirements and holding expenses. Also, for power-law servicing of memoryless input, we have developed optimal multiscale strategies for flow control which allows trade-offs of throttling and holding costs.

7.2 Future Directions

While a broad collection of problems have been addressed in this work, our investigation has been highly preliminary, and many of these problems warrants further, more detailed study. Characterization of the statistical properties of estimation algorithms developed in

Chapter 4, for example, provides a stimulating future direction of exploration. In Chapter 4, we have argued via simulations that the parameter estimator of Section 4.1 exhibits desirable properties such as robustness under noise corruption, and vanishing variance for large samples. In addition, we have observed dependence of the estimation performance on the fractal dimension, or shape parameter γ. It would be valuable to have a theoretical characterization of the performance of this estimator. It is plausible that the analysis in [36] can provide significant insight into this problem.

In our study of networking problems involving fractal renewal processes, a number of our assumptions have been somewhat simplistic. While we have considered memoryless erasure and superposition of independent point processes, bulk erasures, or erasures in bursts, and merging of dependent packet streams are also realistic in many networking scenarios. In addition, while our analysis has furnished strong qualitative evidence of invariance of fractal point processes under random erasure and superposition, there is much research opportunity in rigorous formulation of a notion of invariance.

Our analysis of fractal queueing problems has been quite fundamental, involving either memoryless service (as in Section 5.4), or memoryless arrivals (as in Section 5.5). More sophisticated queueing models involving fractal renewal processes could also be formulated using our multiscale framework. For example, power-law service of fractal arrivals should be a straightforward extension of our models. In addition, while we have focused exclusively on computations of the steady-state customer distributions, many other quantities are just as relevant and insightful. These include waiting-time distribution, as well as busy period distribution. Finally, fractal queueing problems involving priority, polling, and various behavior of the server, such as vacationing, all have their importance.

Just like queueing analysis, the queueing control problems we have addressed have natural extensions to priority and polling scenarios. Even for the queueing set-ups that we have considered, cost functions other than the ones we chose could be more appropriate in many real situations. Extension of our techniques to other cost functions should be straightforward, and in general, for very reasonable objectives, it is observed that our multiscale framework allows convenient employment of past history of the system to improve performance. As a broader question, the existence of a universal optimal queueing strategy remains to be explored, which may itself exhibit self-similarity. We expect multiscale paradigms similar to the one proposed here to have a potentially significant role to play in

these open problems.

As a closing remark, this thesis has demonstrated the usefulness of a particular form of multiscale framework for studying fractal point processes. We note that other multiscale frameworks are also under active investigation. Of particular interest we cite the work of Teich, et al., [44], which employs the wavelet transforma to formulate generalized notions of the Fano factor for the study of fractal point processes. This framework has appeared promising for a number of estimation problems involving fractal point processes. It is conceivable that appropriate combination of multiscale paradigms could have potential importance in furthering our knowledge of the intriguing family of fractal point processes.

Appendix A

A.1 Proof of Theorem 1

We first establish the following lemma.

Lemma 1 *When a fractal renewal process with interarrivals $X[n]$ is conditioned on the event*

$$\mathcal{E}_a = \{a\underline{x} < X \leq a\overline{x}\}$$

for any $a > 0$, the resulting process is a renewal process with interarrivals $Y_a[n]$ distributed according to the density

$$f_{Y_a}(y) = (1/a)\, f_{Y_1}(y/a). \tag{A.1}$$

Proof:

First, we note as an immediate consequence of Definition 1 we have that the interarrivals $Y_1[n]$ constitute a renewal process. Next, for some positive integer K, let $n_1, n_2, \ldots, n_K$ denote an arbitrary collection of interarrival indices and $x_1, x_2, \ldots, x_K$ an arbitrary collection of constants. Then

$$\begin{aligned}
&\Pr[Y_a[n_k] < x_k; k = 1, 2, \ldots, K] \\
&\quad = \Pr[X[n_k] < x_k; k = 1, 2, \ldots, K \mid a\underline{x} < X[n_k] \leq a\overline{x}; k = 1, 2, \ldots, K] && \text{(A.2)} \\
&\quad = \Pr[aX[n_k] < x_k; k = 1, 2, \ldots, K \mid a\underline{x} < aX[n_k] \leq a\overline{x}; k = 1, 2, \ldots, K] && \text{(A.3)} \\
&\quad = \Pr[X[n_k] < x_k/a; k = 1, 2, \ldots, K \mid \underline{x} < X[n_k] \leq \overline{x}; k = 1, 2, \ldots, K] && \text{(A.4)} \\
&\quad = \Pr[Y_1[n_k] < x_k/a; k = 1, 2, \ldots, K] && \text{(A.5)} \\
&\quad = \prod_{k=1}^{K} \Pr[Y_1 < x_k/a], && \text{(A.6)}
\end{aligned}$$

where (A.2) is a consequence of the definition of $Y_a[n]$, (A.3) is a consequence of (2.6), and (A.6) is a consequence of the fact that the $Y_1[n]$ constitute a renewal process. From (A.6) we can immediately conclude that the the $Y_a[n]$ are independent and identically-distributed. Finally, differentiating both sides of (A.6) we get (A.1).

■

We are now ready for the main proof.

Let α denote the separation between the endpoints, $\overline{x}/\underline{x}$. Also, let M_L and M_U be integers such that

$$\alpha^{M_L}\underline{x} < x_L < x_U < \alpha^{M_U}\overline{x}.$$

Let $\tilde{Y}[n]$ be the interarrivals generated from the $X[n]$ by conditioning on the event

$$\tilde{\mathcal{E}} = \{\alpha^{M_L}\underline{x} < X \leq \alpha^{M_U}\overline{x}\}.$$

Since $\tilde{\mathcal{E}}$ can be expressed as

$$\tilde{\mathcal{E}} = \bigcup_{m=M_L}^{M_U} \mathcal{E}_{\alpha^m},$$

and $\mathcal{E}_{\alpha^m}$ are mutually exclusive events, we have by Definition 1 and Lemma 1 that the interarrivals $\tilde{Y}[n]$ are independent. Furthermore, we have that the $\tilde{Y}[n]$ are identically distributed with density

$$f_{\tilde{Y}}(y) = \sum_m f_{Y_{\alpha^m}}(y) \cdot \Pr\left[\mathcal{E}_{\alpha^m} \,\middle|\, \tilde{\mathcal{E}}\right]. \tag{A.7}$$

Finally, since

$$\mathcal{E} \Rightarrow \tilde{\mathcal{E}},$$

we have that the interarrivals $Y[n]$ are also independent and identically distributed with density

$$f_Y(y) = \begin{cases} f_{\tilde{Y}}(y)/\Pr\left[\mathcal{E} \,\middle|\, \tilde{\mathcal{E}}\right] & x_L < y \leq x_U \\ 0 & \text{otherwise.} \end{cases} \tag{A.8}$$

It remains only to derive the density of $Y[n]$. We begin by defining

$$\mathcal{E}_* = \{a\underline{x} < X < \overline{x}\},$$

where $1 < a < \overline{x}/\underline{x}$, and letting $Y_*[n]$ be the interarrivals resulting from conditioning the process with interarrivals $X[n]$ on $\mathcal{E}_*$. Then

$$\mathcal{E}_* \Rightarrow \mathcal{E}_1$$

and

$$\mathcal{E}_* \Rightarrow \mathcal{E}_a$$

respectively imply, for $a\underline{x} < y \leq \overline{x}$,

$$f_{Y_*}(y) = f_{Y_1}(y) / \Pr[\mathcal{E}_* \mid \mathcal{E}_1] \tag{A.9}$$

and

$$f_{Y_*}(y) = f_{Y_a}(y) / \Pr[\mathcal{E}_* \mid \mathcal{E}_a]. \tag{A.10}$$

Equating (A.9) and (A.10), cross-multiplying, and exploiting (A.1) we get, for $a\underline{x} < y \leq \overline{x}$,

$$f_{Y_1}(y) \int_{a\underline{x}}^{\overline{x}} f_{Y_1}(x/a)\, dx = f_{Y_1}(y/a) \int_{a\underline{x}}^{\overline{x}} f_{Y_1}(x)\, dx. \tag{A.11}$$

Differentiating (A.11) with respect to a and letting $a \to 1+$ we get, for $\underline{x} < y \leq \overline{x}$,

$$f_{Y_1}(y) \int_{\underline{x}}^{\overline{x}} x\, f'_{Y_1}(x)\, dx = y\, f'_{Y_1}(y) \int_{\underline{x}}^{\overline{x}} f_{Y_1}(x)\, dx. \tag{A.12}$$

Rearranging terms we get, for $\underline{x} < y \leq \overline{x}$ such that $f_{Y_1}(y) \neq 0$,

$$\frac{y f'_{Y_1}(y)}{f_{Y_1}(y)} = \gamma, \tag{A.13}$$

where γ is the constant obtained by integrating (A.12) by parts

$$\gamma = \frac{\overline{x} f_{Y_1}(\overline{x}) - \underline{x} f_{Y_1}(\underline{x})}{\int_{\underline{x}}^{\overline{x}} f_{Y_1}(x)\, dx} - 1. \tag{A.14}$$

Now all regular, positive solutions to (A.13) on $\underline{x} < y \leq \overline{x}$ can be obtained by separation of variables and are of the form [16]

$$f_{Y_1}(y) = \sigma_1^2 / y^\gamma, \tag{A.15}$$

where σ_1^2 is a normalization constant. In addition, it can be readily verified by substitution of (A.15) into (A.14) that γ is a free parameter.

Finally, combining (A.15) with (A.1), and substituting the result into (A.7) and, in turn, into (A.8) we obtain (2.7).

■

Appendix B

B.1 Proof of Theorem 2

First, observe that mutual independence of $X[n]$ follows as an immediate consequence of the mutual independence of the constituent processes $N_{W_M}(t)$, the independent selection from among these processes, and the independent-increments property of Poisson processes.

To derive the density of the interarrivals, we begin by noting that, conditioned on $M[n] = m$, $X[n]$ is the waiting time until the next arrival in $N_{W_m}(t)$ after $t = S_X[n-1]$. Because a Poisson process is memoryless, the density for this conditional waiting time is

$$f_{W_m}(x) = f_{X[n]|M[n]}(x \mid m) = \begin{cases} \lambda\rho^{-m}\exp(-\lambda\rho^{-m}x) & \\ 0 & \text{otherwise,} \end{cases} \tag{B.1}$$

which we note is independent of n. Weighting (B.1) with (3.7), and summing, we get the unconditional density

$$\begin{aligned} f_X(x) &= \sum_{m=\underline{m}}^{\overline{m}} p_M(m)\, f_{X|M}(x \mid m) \\ &= \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} \rho^{-(\gamma-1)m}\, \lambda\rho^{-m}\exp(-\lambda\rho^{-m}x), \end{aligned} \tag{B.2}$$

where for $x \geq 0$. This is just (3.3).

Now, every $x > 0$ can be uniquely expressed in the form

$$x = \rho^{m_0} x_0, \tag{B.3}$$

where m_0 is an integer, and $1 \leq x_0 < \rho$. Using (B.3) in (B.1) it is then straightforward to

show that

$$f_{W_m}(x) = \rho^{-m_0} f_{W_{m-m_0}}(x_0). \tag{B.4}$$

Taking limits and substituting (B.4) into (B.2) with (B.1), we get that

$$\begin{aligned} \tilde{f}_X(x) \triangleq \lim_{\substack{m\to-\infty \\ \overline{m}\to\infty}} f_X(x)/\sigma_M^2 &= \sum_{m=-\infty}^{\infty} \rho^{(1-\gamma)m} f_{W_m}(x) && \text{(B.5)} \\ &= \rho^{-m_0\gamma} \sum_{m=-\infty}^{\infty} \rho^{(1-\gamma)m-m_0+m_0\gamma} f_{W_{m-m_0}}(x_0) \\ &= \rho^{-m_0\gamma} \tilde{f}_X(x_0). && \text{(B.6)} \end{aligned}$$

Multiplying both sides of (B.6) by x^γ and again exploiting (B.3), we get

$$x^\gamma \tilde{f}_X(x) = x_0^\gamma \tilde{f}_X(x_0)$$

and, in turn,

$$\frac{1}{x^\gamma} \cdot \left[\inf_{1\le x<\rho} \tilde{f}_X(x) \right] \le \tilde{f}_X(x) \le \frac{\rho^\gamma}{x^\gamma} \cdot \left[\sup_{1\le x<\rho} \tilde{f}_X(x) \right]. \tag{B.7}$$

Thus, it remains only to bound the two terms in brackets.

We begin with the lower bound. Since for every m, $f_{W_m}(x)$ is monotonically decreasing in x,

$$\begin{aligned} \inf_{1\le x<\rho} \tilde{f}_X(x) &= \tilde{f}_X(\rho) \\ &= \sum_{m=-\infty}^{\infty} \lambda \rho^{-\gamma m} \exp(-\lambda \rho^{-m} \rho). \end{aligned}$$

Regarding the summand as a function of a continuous variable r

$$g(r) = \lambda \rho^{-\gamma r} \exp(-\lambda \rho^{-r} \rho),$$

it is clear that $g(r)$ has monotonically decreasing slope:

$$\begin{aligned}\frac{d}{dr}\left(\ln g(r)\right) &= \frac{d}{dr}\left(\ln\lambda - \gamma r \ln\rho - \lambda\rho^{-r}\rho\right) \\ &= -\gamma\ln\rho + \lambda\rho\rho^{-r}\ln\rho,\end{aligned}$$

and has one global maximum r_0. So, for $m' \leq m < r_0$,

$$g(m') \leq g(m)$$

while for $r_0 \leq m \leq m'$,

$$g(m') \leq g(m).$$

Thus, defining the piecewise constant function

$$h(r) = g(m), \quad m \leq r < m+1,$$

we have that

$$h(r) \geq g(r) = \lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r}\rho)$$

for every $r > r_0$, and

$$\begin{aligned}h(r) \geq g(r-1) &= \lambda\rho^{-\gamma(r-1)}\exp(-\lambda\rho^{-(r-1)}\rho) \\ &= \rho^{\gamma}\lambda\rho^{-\gamma r}\exp(-\lambda\rho^{2}\rho^{-r})\end{aligned}$$

for every $r < r_0$. Thus, for every r,

$$\begin{aligned}h(r) &\geq \min(\rho^{\gamma}\lambda\rho^{-\gamma r}\exp(-\lambda\rho^{2}\rho^{-r}), \lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r}\rho)) \\ &\geq \lambda\rho^{-\gamma r}\exp(-\lambda\rho^{2}\rho^{-r}).\end{aligned}$$

Thus,

$$\int_{-\infty}^{\infty} h(r)\,dr \geq \int_{-\infty}^{\infty} \lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r+2})\,dr.$$

But the left hand side is just $\tilde{f}(\rho)$. Applying the change of variables $u = \lambda\rho^{-r+2}$, we have that

$$\begin{aligned}\int_{-\infty}^{\infty} \lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r+2})\,dr &= \lambda\int_0^{\infty}\left(\frac{u}{\lambda\rho^2}\right)^{\gamma}\frac{1}{u\ln\rho}\exp(-u)\,du \\ &= \Gamma(\gamma-1)\frac{\lambda}{\ln\rho(\lambda\rho^2)^{\gamma}} > 0.\end{aligned}$$

Similar computations can be applied to establish the upper bound. More specifically,

$$\begin{aligned}\sup_{1\le x<\rho}\tilde{f}_X(x) &= \tilde{f}_X(1) \\ &= \sum_{m=-\infty}^{\infty}\lambda\rho^{-\gamma m}\exp(-\lambda\rho^{-m}).\end{aligned}$$

Defining a piecewise function corresponding to the summand, we get

$$h(r) = \lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r}), \quad m \le r < m+1$$

which is bounded above by

$$\rho^{\gamma}\lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r+1}).$$

Thus,

$$\begin{aligned}\tilde{f}_X(1) &= \int_{-\infty}^{\infty} h(r)\,dr \\ &\le \int_{-\infty}^{\infty}\rho^{\gamma}\lambda\rho^{-\gamma r}\exp(-\lambda\rho^{-r+1})\,dr \\ &= \rho^{\gamma}\Gamma(\gamma-1)\frac{\lambda}{\ln\rho(\lambda\rho)^{\gamma}} < \infty.\end{aligned}$$

With this set of bounds, (B.7) suggests a bound for the ratio σ_U^2/σ_L^2 of (3.4),

$$\sigma_U^2/\sigma_L^2 < \rho^{3\gamma}. \tag{B.8}$$

While this bound is in general loose, it does require that

$$\sigma_U^2/\sigma_L^2 \to 1$$

as $\rho \to 1+$, in agreement with our empirical observations.

■

B.2 Proof of Theorem 3

Using arguments exactly analogous to those applied in the discrete case, we obtain that the interarrivals are independent and identically-distributed with common density

$$f_X(x) = \int_{\underline{a}}^{\overline{a}} f_A(a)\, f_{X|A}(x \mid a)\, da, \tag{B.9}$$

$$= \int_{\underline{a}}^{\overline{a}} f_A(a) e^{-a} \lambda \exp[-e^{-a}\lambda x]\, da \quad x \geq 0. \tag{B.10}$$

Via the change of variables $u = e^{-a}\lambda x$, we rewrite

$$f_X(x) = \int_{\underline{a}}^{\overline{a}} f_A(a)\, e^{-a} \lambda \exp[-e^{-a}\lambda x]\, da \quad x \geq 0.\, da$$

$$= \frac{1}{x^\gamma} \frac{\sigma_A^2}{\lambda^{\gamma-1}} \int_{e^{-\overline{a}}\lambda x}^{e^{-\underline{a}}\lambda x} u^{\gamma-1}\, e^{-u}\, du, \tag{B.11}$$

from which we see that for each $x > 0$, when $\underline{a} \to -\infty$ and $\overline{a} \to \infty$, we obtain (3.8) with (3.9). Furthermore, note that (B.11) also establishes (3.10).

Appendix C

C.1 Derivation of the EM Parameter Estimation Algorithm

C.1.1 General case

We begin by considering the general form of our parameter estimation problem. Specifically, we assume that the observations are noisy, of the form

$$R[n] = X[n] + W[n], \quad n = 1, 2, \ldots, N,$$

where $X[n]$ are the fractal renewal process interarrivals and $W[n]$ are the exponential distortion terms.

Following [27], each iteration of the EM algorithm involves the computation and maximization of the function

$$\Psi(\Theta, \hat{\Theta}) \stackrel{\Delta}{=} E\left[\ln f_{\mathbf{R},\mathbf{S}}(\mathbf{r}, \mathbf{s}; \Theta) \;\middle|\; \mathbf{r}; \hat{\Theta}\right], \tag{C.1}$$

where $\mathbf{r}$ denotes the incomplete data, and $(\mathbf{r}, \mathbf{s})$ the complete data. Choosing

$$\begin{aligned} \mathbf{r} &= \{r[n];\ n = 1, 2, \ldots, N\} \\ \mathbf{s} &= \{x[n], m[n];\ n = 1, 2, \ldots, N\} \end{aligned}$$

leads to

$$\begin{aligned}\ln f_{\mathbf{R},\mathbf{S}}(\mathbf{r},\mathbf{s};\boldsymbol{\Theta}) &= \sum_{n=1}^{N}\Big\{\ln(1-\beta)+(m[n]-1)\ln\beta+\ln\left(\lambda\rho^{-m[n]}\exp(-\lambda\rho^{-m[n]}x[n])\right)\\ &\qquad +\ln\left(\alpha\exp(-\alpha(r[n]-x[n]))\right)\Big\}\\ &= N(\ln(1-\beta)-\ln\beta+\ln\lambda+\ln\alpha)\\ &\qquad -\lambda\sum_{n=1}^{N}\rho^{-m[n]}x[n]+(\ln\beta-\ln\rho)\sum_{n=1}^{N}m[n]-\alpha\sum_{n=1}^{N}(r[n]-x[n]),\end{aligned} \tag{C.2}$$

where for convenience, we have made the reasonable assumption that the number of scales L is large so that $1-\beta^L \approx 1$. Substituting (C.2) into (C.1) we obtain, in turn,

$$\begin{aligned}\Psi(\boldsymbol{\Theta},\hat{\boldsymbol{\Theta}}) &= N(\ln(1-\beta)-\ln\beta+\ln\lambda+\ln\alpha)-\lambda\sum_{n=1}^{N}E\left[\rho^{-m[n]}x[n]\,\middle|\,\mathbf{r};\hat{\boldsymbol{\Theta}}\right]\\ &\quad+(\ln\beta-\ln\rho)\sum_{n=1}^{N}E\left[m[n]\,\middle|\,\mathbf{r};\hat{\boldsymbol{\Theta}}\right]-\alpha\sum_{n=1}^{N}E\left[r[n]-x[n]\,\middle|\,\mathbf{r};\hat{\boldsymbol{\Theta}}\right]\\ &= N(\ln(1-\beta)-\ln\beta+\ln\lambda+\ln\alpha)-\lambda\sum_{n=1}^{N}E\left[\rho^{-m[n]}x[n]\,\middle|\,r[n];\hat{\boldsymbol{\Theta}}\right]\\ &\quad+(\ln\beta-\ln\rho)\sum_{n=1}^{N}E\left[m[n]\,\middle|\,r[n];\hat{\boldsymbol{\Theta}}\right]-\alpha\sum_{n=1}^{N}r[n]-\alpha\sum_{n=1}^{N}E\left[x[n]\,\middle|\,r[n];\hat{\boldsymbol{\Theta}}\right],\end{aligned} \tag{C.3}$$

where the second equality follows from the mutual independence of the observations. The E-step of iteration k of the EM algorithm, then, simply involves the computation of this quantity with $\hat{\boldsymbol{\Theta}}$ set to $\hat{\boldsymbol{\Theta}}_{[k]}$, the current parameter estimates.

We compute the expected values in (C.3) as iterated expectations. More specifically, we use the relation

$$\begin{aligned}E\left[\cdot\,\middle|\,r[n];\hat{\boldsymbol{\Theta}}\right] &= E\left[E\left[\cdot\,\middle|\,r[n],m[n];\hat{\boldsymbol{\Theta}}\right]\,\middle|\,r[n];\hat{\boldsymbol{\Theta}}\right]\\ &= \sum_{m=1}^{L}E\left[\cdot\,\middle|\,r[n],m;\hat{\boldsymbol{\Theta}}\right]p_{M|R}(m\mid r[n];\hat{\boldsymbol{\Theta}})\end{aligned}$$

to simplify the calculations.

Thus, the only terms we need to compute are $E\left[x[n] \,\middle|\, r[n], m; \hat{\Theta}\right]$ and $p_{M|R}(m \mid r[n]; \hat{\Theta})$. Now, by definition,

$$E\left[x[n] \,\middle|\, r[n], m; \hat{\Theta}\right] = \frac{\displaystyle\int_0^{r[n]} x\, f_{R|X,M}(r[n] \mid x, m; \hat{\Theta})\, f_{X|M}(x \mid m; \hat{\Theta})\, dx}{\displaystyle\int_0^{r[n]} f_{R|X,M}(r[n] \mid x, m; \hat{\Theta})\, f_{X|M}(x \mid m; \hat{\Theta})\, dx}, \tag{C.4}$$

where, for the particular problem at hand,

$$f_{R|X,M}(r \mid x, m) = f_W(r - x) \tag{C.5a}$$

$$f_{X|M}(x \mid m) = f_{W_m}(x). \tag{C.5b}$$

The integrals in (C.4) can be evaluated in a straightforward manner. First, we consider the case $\lambda_m \neq \alpha$. The denominator in this case is

$$\begin{aligned}\int_0^{r[n]} f_{R|X,M}(r[n] \mid x, m; \hat{\Theta})\, f_{X|M}(x \mid m; \hat{\Theta})\, dx &= \lambda_m \alpha e^{-\alpha r[n]} \int_0^{r[n]} e^{-(\lambda_m - \alpha)x}\, dx \\ &= \frac{\lambda_m \alpha}{\lambda_m - \alpha} e^{-\alpha r[n]} \left(1 - e^{-(\lambda_m - \alpha) r[n]}\right).\end{aligned} \tag{C.6}$$

The numerator, on the other hand, can be integrated by parts, to yield

$$\begin{aligned}&\int_0^{r[n]} x\, f_{R|X,M}(r[n] \mid x, m; \hat{\Theta})\, f_{X|M}(x \mid m; \hat{\Theta})\, dx \\ &\quad = \lambda_m \alpha e^{-\alpha r[n]} \int_0^{r[n]} x e^{-(\lambda_m - \alpha)x}\, dx \\ &\quad = \lambda_m \alpha e^{-\alpha r[n]} \left(-\frac{x e^{-(\lambda_m - \alpha)x}}{\lambda_m - \alpha}\bigg|_0^{r[n]} + \frac{1}{\lambda_m - \alpha} \int_0^{r[n]} e^{-(\lambda_m - \alpha)x}\, dx \right) \\ &\quad = \frac{\lambda_m \alpha e^{-\alpha r[n]}}{(\lambda_m - \alpha)^2} \left(-r[n] e^{-(\lambda_m - \alpha) r[n]} (\lambda_m - \alpha) + 1 - e^{-(\lambda_m - \alpha) r[n]} \right).\end{aligned} \tag{C.7}$$

Dividing (C.7) by (C.6), we get the first case in (4.14).

For $\lambda_m = \alpha$, the denominator becomes

$$\alpha^2 e^{-\alpha r[n]} \int_0^{r[n]} dx = \alpha^2 r[n] e^{-\alpha r[n]}, \tag{C.8}$$

while the numerator becomes

$$\alpha^2 e^{-\alpha r[n]} \int_0^{r[n]} x\, dx = \frac{\alpha^2 (r[n])^2 e^{-\alpha r[n]}}{2}. \tag{C.9}$$

Dividing (C.9) by (C.8), we get the second case in (4.14).

To compute $p_{M|R}(m \mid r[n]; \hat{\Theta})$, we invoke Bayes Rule, i.e.,

$$p_{M|R}(m \mid r[n]; \hat{\Theta}) = \frac{f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}{\sum_{m=1}^{L} f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}. \tag{C.10}$$

Using (C.10) together with

$$f_{R|M}(r \mid m) = \int_0^{\tau} f_{W_m}(\tau)\, f_W(r - \tau)\, d\tau$$

$$= \begin{cases} \lambda_m \alpha e^{-\alpha r} \int_0^{\tau} e^{-(\lambda_m - \alpha)\tau}\, d\tau & \lambda_m \neq \alpha \\ \alpha^2 e^{-\alpha r} \int_0^{\tau} d\tau & \text{otherwise} \end{cases}$$

we obtain (4.13), with

$$\sigma^2_{M|R} \triangleq \frac{1}{\sum_{m=1}^{L} f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}.$$

In the M-step, we seek the parameter estimates for the next iteration $\hat{\Theta}_{[k+1]}$ by solving

$$\hat{\Theta}_{[k+1]} = \arg\max_{\Theta} \Psi(\Theta, \hat{\Theta}_{[k]}).$$

To perform this maximization, we set the partial derivatives $\partial\Psi/\partial\lambda$, $\partial\Psi/\partial\alpha$ and $\partial\Psi/\partial\beta$ to zero. Conveniently, the resulting equations may be solved independently (and uniquely)

for the individual parameter estimates, from which we obtain

$$\hat{\lambda}_{[k+1]} = \frac{N}{\sum_{n=1}^{N} E\left[\rho^{-m[n]}x[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]}, \tag{C.11a}$$

$$\hat{\alpha}_{[k+1]} = \frac{N}{\sum_{n=1}^{N} E\left[r[n] - x[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]}, \tag{C.11b}$$

$$\hat{\beta}_{[k+1]} = \frac{\sum_{n=1}^{N}\left(E\left[m[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]\right) - N}{\sum_{n=1}^{N} E\left[m[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]}. \tag{C.11c}$$

Evaluating the expectations in eqs. (C.11) iteratively by first conditioning on $m[n]$, we then obtain eqs. (4.15).

C.1.2 Noise-free case; $W[n] \equiv 0$

In the event that $W[n]$ is identically zero, we choose our complete data as $(\mathbf{r}, \mathbf{s})$, with the incomplete data

$$\mathbf{r} = \{r[n];\ n = 1, 2, \ldots N\}$$

as before, and

$$\mathbf{s} = \{m[n];\ n = 1, 2, \ldots N\}$$

to complement $\mathbf{r}$. Thus,

$$\begin{aligned}\ln f_{\mathbf{R},\mathbf{S}}(\mathbf{r}, \mathbf{s}; \Theta) &= \sum_{n=1}^{N}\left\{\ln(1-\beta) + (m[n]-1)\ln\beta + \ln\left(\lambda\rho^{-m[n]}\exp(-\lambda\rho^{-m[n]}r[n])\right)\right\} \\ &= N(\ln(1-\beta) - \ln\beta + \ln\lambda) - \lambda\sum_{n=1}^{N}\rho^{-m[n]}r[n] + (\ln\beta - \ln\rho)\sum_{n=1}^{N} m[n],\end{aligned} \tag{C.12}$$

and it follows that

$$\Psi(\Theta, \hat{\Theta}) = N(\ln(1-\beta) - \ln\beta + \ln\lambda) - \lambda \sum_{n=1}^{N} r[n] E\left[\rho^{-m[n]} \,\middle|\, r[n]; \hat{\Theta}\right]$$
$$+ (\ln\beta - \ln\rho) \sum_{n=1}^{N} E\left[m[n] \,\middle|\, r[n]; \hat{\Theta}\right], \tag{C.13}$$

similar to the noise-corrupted case.

The only piece missing at this point is $p_{M|R}(m \mid r[n]; \hat{\Theta})$. Again, invoking Bayes Rule, we get that

$$p_{M|R}(m \mid r[n]; \hat{\Theta}) = \frac{f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}{\sum_{m=1}^{L} f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}.$$

But since the observations are now clean, $f_{R|M}(r|m)$ is simply $\lambda_m \exp(-\lambda_m r)$. This completes the derivation of the noise-free EM algorithm.

Appendix D

D.1 Derivation of the Counting Process Distribution Coefficients

D.1.1 Arrival-Observed Case

We begin with the arrival-observed counting process distribution $\left\{\pi_i^{(a)}(t); i = 0, 1, \ldots\right\}$. First, as given in (5.5), we have a closed-form expression for the 0th-order term $\pi_0^{(a)}(t)$. To obtain Taylor's series coefficients for higher-order terms, we consider the z-transform of the probability distribution, which is given in (5.6)

$$\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}} = \mathbf{q}\left(\mathbf{I} + \left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)(\lambda t) + \frac{\left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)^2(\lambda t)^2}{2!} + \cdots\right)\mathbf{1}^{\mathrm{T}}.$$

Extracting the coefficient of z^1, we get that

$$\pi_1^{(a)}(t) = a_1^{(1)}(\lambda t) - a_2^{(1)}\frac{(\lambda t)^2}{2!} + \cdots \tag{D.1}$$

where for $k \geq 1$,

$$a_k^{(1)} = \mathbf{q}\mathbf{B}^{k-1}\mathbf{b}^{\mathrm{T}}\mathbf{q}\mathbf{1}^{\mathrm{T}} + \mathbf{q}\mathbf{B}^{k-2}\mathbf{b}^{\mathrm{T}}\mathbf{q}\mathbf{B}\mathbf{1}^{\mathrm{T}} + \cdots + \mathbf{q}\mathbf{B}\mathbf{b}^{\mathrm{T}}\mathbf{q}\mathbf{B}^{k-2}\mathbf{1}^{\mathrm{T}} + \mathbf{q}\mathbf{b}^{\mathrm{T}}\mathbf{q}\mathbf{B}^{k-1}\mathbf{1}^{\mathrm{T}}. \tag{D.2}$$

But $\mathbf{b} = \mathbf{1}\mathbf{B}$. Thus,

$$a_k^{(1)} = \mathbf{q}\mathbf{B}^{k}\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{1}^{\mathrm{T}} + \mathbf{q}\mathbf{B}^{k-1}\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{B}\mathbf{1}^{\mathrm{T}} + \cdots + \mathbf{q}\mathbf{B}^{2}\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{B}^{k-2}\mathbf{1}^{\mathrm{T}} + \mathbf{q}\mathbf{B}\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{B}^{k-1}\mathbf{1}^{\mathrm{T}}. \tag{D.3}$$

This is simply the sum of all terms of the form

$$\mathbf{q}\mathbf{B}^m\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{B}^n\mathbf{1}^{\mathrm{T}}$$

such that $m \geq 1$, $n \geq 0$, and $m+n=k$. Moreover, we recognize $\mathbf{q}\mathbf{B}^l\mathbf{1}^{\mathrm{T}}$ as the lth moment of the random variable R, with distribution given in (5.9). Thus, we have

$$a_k^{(1)} = \sum_{l=1}^{k} M_l M_{k-l}, \tag{D.4}$$

with M_l denoting the lth moment of R.

Next we turn to the expansion of $\pi_i^{(\mathrm{a})}(t)$ for arbitrary i. First, it is clear that the coefficients of t^k are zero for $k < i$. Thus, we have

$$\pi_i^{(\mathrm{a})}(t) = a_i^{(i)}\frac{(\lambda t)^i}{i!} - a_{i+1}^{(i)}\frac{(\lambda t)^{i+1}}{(i+1)!} + \cdots . \tag{D.5}$$

Using the same argument as before, we have that each coefficient $a_k^{(i)}$ is the sum of all terms of the form

$$\mathbf{q}\mathbf{B}^{m_1}\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{B}^{m_2}\mathbf{1}^{\mathrm{T}}\cdots\mathbf{q}\mathbf{B}^{m_i}\mathbf{1}^{\mathrm{T}}\mathbf{q}\mathbf{B}^{m_{i+1}}\mathbf{1}^{\mathrm{T}}, \tag{D.6}$$

with $m_1, m_2, \ldots, m_i \geq 1$, $m_{i+1} \geq 0$, and $m_1+m_2+\cdots+m_{i+1}=k$. Such terms with $m_1 = l$, where $1 \leq l \leq k-i+1$, are those with $m_2, m_3, \ldots, m_i \geq 1$, $m_{i+1} \geq 0$, $m_2+m_3+\cdots+m_{i+1} = k-l$. But these are exactly the terms making up the sum $a_{k-l}^{(i-1)}$. Thus, we have

$$a_k^{(i)} = \sum_{l=1}^{k-i+1} M_l a_{k-l}^{(i-1)}. \tag{D.7}$$

D.1.2 Random Incidence Case

We first argue that the left null space of the matrix $\left(-\mathbf{B}+\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)$ is spanned by the vector $\mathbf{q}\mathbf{B}^{-1}$. Suppose the vector $\mathbf{v}$ is in the left null space. In other words, we have that

$$\mathbf{v}\left(-\mathbf{B}+\mathbf{b}^{\mathrm{T}}\mathbf{q}\right) = \mathbf{0}$$

where $\mathbf{0}$ is a zero row vector. So,

$$\mathbf{vB} = \mathbf{vb}^{\mathrm{T}}\mathbf{q} = \kappa\mathbf{q}$$

where κ is the inner product of $\mathbf{v}$ and $\mathbf{b}$. Right-multiplying both sides by $\mathbf{B}^{-1}$, then, we get that

$$\mathbf{v} = \kappa\left(\mathbf{qB}^{-1}\right).$$

We proceed to determine the coefficients in the counting process probability distribution $\left\{\pi_i^{(\mathrm{r})}(t); i = 0, 1, \ldots\right\}$. Again, as given in (5.11), we have a closed-form expression for the 0th-order term $\pi_0^{(\mathrm{r})}(t)$. To obtain higher-order terms, we again turn to the z-transform of the distribution given in (5.12),

$$\hat{\mathbf{p}}(z;t)\mathbf{1}^{\mathrm{T}} = \tilde{\sigma}^2\mathbf{qB}^{-1}\left(\mathbf{I} + \left(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q}\right)(\lambda t) + \frac{(-\mathbf{B} + z\mathbf{b}^{\mathrm{T}}\mathbf{q})^2(\lambda t)^2}{2!} + \cdots\right)\mathbf{1}^{\mathrm{T}}, \qquad \text{(D.8)}$$

As before, it is clear that the coefficients for t^k will be zero for $k < i$. Thus, we have

$$\pi_i^{(\mathrm{r})}(t) = r_i^{(i)}\frac{(\lambda t)^i}{i!} - r_{i+1}^{(i)}\frac{(\lambda t)^{i+1}}{(i+1)!} + \cdots. \qquad \text{(D.9)}$$

Focusing on the kth power of t in the coefficient of z^i, we see that this is the sum of all terms of the form

$$\tilde{\sigma}^2\mathbf{qB}^{-1}\mathbf{B}^{m_1}\mathbf{1}^{\mathrm{T}}\mathbf{qB}^{m_2}\mathbf{1}^{\mathrm{T}}\cdots\mathbf{qB}^{m_i}\mathbf{1}^{\mathrm{T}}\mathbf{qB}^{m_{i+1}}\mathbf{1}^{\mathrm{T}}, \qquad \text{(D.10)}$$

with $m_1, m_2, \ldots, m_i \geq 1$, $m_{i+1} \geq 0$, and $m_1 + m_2 + \cdots + m_{i+1} = k$. For those terms with $m_1 = 1$, the product reduces to

$$\tilde{\sigma}^2\mathbf{qB}^{m_2}\mathbf{1}^{\mathrm{T}}\cdots\mathbf{qB}^{m_i}\mathbf{1}^{\mathrm{T}}\mathbf{qB}^{m_{i+1}}\mathbf{1}^{\mathrm{T}}, \qquad \text{(D.11)}$$

since $\mathbf{q1}^{\mathrm{T}} = 1$. But the sum of all terms of this form is just $\tilde{\sigma}^2 a_k^{(i-1)}$. On the other hand, terms with $m_1 > 1$ can be written as

$$\tilde{\sigma}^2\mathbf{qB}^{\tilde{m}_1}\mathbf{1}^{\mathrm{T}}\mathbf{qB}^{m_2}\mathbf{1}^{\mathrm{T}}\cdots\mathbf{qB}^{m_i}\mathbf{1}^{\mathrm{T}}\mathbf{qB}^{m_{i+1}}\mathbf{1}^{\mathrm{T}}, \qquad \text{(D.12)}$$

with $\tilde{m}_1, m_2, \ldots, m_i \geq 1$, $m_{i+1} \geq 0$, and $\tilde{m}_1 + m_2 + \cdots + m_{i+1} = k - 1$. But this is precisely $\tilde{\sigma}^2 a_{k-1}^{(i)}$. Thus, we have shown the relation

$$r_k^{(i)} = (a_{k-1}^{(i)} + a_k^{(i-1)})\tilde{\sigma}^2. \tag{D.13}$$

Appendix E

E.1 Proof of Monotonicity of the Optimal Fractal Queue Server

In this section, we show that for servicing a fractal renewal process input, the optimal service rate $\mu^*_{i,j}$ is monotonic in the number of customers, given that the holding cost function $h(i)$ is convex, i.e.,

$$h(i) - h(i-1) \leq h(i+1) - h(i)$$

for every $i = 1, 2, \ldots$. More precisely, for every $i = 1, 2, \ldots,$

$$\mu^*_{i-1,j} \leq \mu^*_{i,j}.$$

We first show that if the first difference of $V_{i,j}$ is nondecreasing in i, then so is $\mu^*_{i,j}$. More precisely, we define the first difference of $V_{i,j}$ as

$$\Delta_{i,j} \triangleq V_{i,j} - V_{i-1,j}.$$

Now, rewriting the Bellman equations (6.5) for $i > 0$, we get that

$$\begin{aligned} V_{i,j} = \frac{1}{\beta + \Omega} \Bigg\{ h(i) + \frac{\lambda}{\eta^{j-1}} \sum_{j'=1}^{L} \sigma^2 q^{j'-1} V_{i+1,j'} + \left(\Omega - \frac{\lambda}{\eta^{j-1}} \right) V_{i,j} \\ + \min_{\mu \in [0,\overline{\mu}]} \Big[c(\mu) - \mu \Delta_{i,j} \Big] \Bigg\}, \quad i = 1, 2, \ldots. \end{aligned} \tag{E.1}$$

Now, suppose that $\Delta_{i,j}$ is indeed nondecreasing in i, and let i_1 be some nonnegative integer.

By definition, we have that

$$c(\mu^*_{i_1,j}) - \mu^*_{i_1,j}\Delta_{i_1,j} < c(\mu) - \mu\Delta_{i_1,j}$$

for every $\mu < \mu^*_{i_1,j}$. Thus, for $i_2 > i_1$,

$$\begin{aligned} c(\mu^*_{i_1,j}) - \mu^*_{i_1,j}\Delta_{i_2,j} &= c(\mu^*_{i_1,j}) - \mu^*_{i_1,j}\Delta_{i_1,j} - \mu^*_{i_1,j}(\Delta_{i_2,j} - \Delta_{i_1,j}) \\ &< c(\mu) - \mu\Delta_{i_1,j} - \mu^*_{i_1,j}(\Delta_{i_2,j} - \Delta_{i_1,j}) \\ &< c(\mu) - \mu\Delta_{i_1,j} - \mu(\Delta_{i_2,j} - \Delta_{i_1,j}) = c(\mu) - \mu\Delta_{i_2,j} \end{aligned}$$

for every $\mu < \mu^*_{i_1,j}$. Thus,

$$\mu^*_{i_1,j} \leq \mu^*_{i_2,j}.$$

We proceed to show that $\Delta_{i,j}$ is monotonic in i. For this, it suffices to show that at each stage k of the value iteration method, the first difference of $V_{i,j}(k)$ has this property. We define this first difference as

$$\Delta_{i,j}(k) \triangleq V_{i,j}(k) - V_{i-1,j}(k), \tag{E.2}$$

with $\Delta_{0,j}(k)$ set to 0 for all j and k. It is clear that $\Delta_{i,j}(0)$ is nondecreasing in i, since $V_{i,j}(0)$ is identically zero. We assume this holds for k, and analyze

$$\Delta_{i,j}(k+1) - \Delta_{i-1,j}(k+1). \tag{E.3}$$

To this end, we rewrite the value iteration equation (6.7) as

$$\begin{aligned} V_{i,j}(k+1) = \frac{1}{\beta+\Omega}\Bigg\{ & h(i) + \frac{\lambda}{\eta^{j-1}} \sum_{j'=1}^{L} \sigma^2 q^{j'-1} V_{i+1,j'}(k) + \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right) V_{i,j}(k) \\ & + \min_{\mu\in[0,\bar{\mu}]} \Big[c(\mu) - \mu\Delta_{i,j}(k) \Big] \Bigg\}, \quad i = 1, 2, \ldots. \end{aligned} \tag{E.4}$$

Defining the optimal service rate at each stage of the value iteration method as $\mu^*_{i,j}(k)$, i.e.,

$$\mu^*_{i,j}(k) \triangleq \arg\min_{\mu\in[0,\bar{\mu}]} \Big[c(\mu) - \mu\Delta_{i,j}(k) \Big], \tag{E.5}$$

for every i, j, k, we have that

$$\begin{aligned}
\Delta_{i+1,j}(k+1) &= V_{i+1,j}(k+1) - V_{i,j}(k+1) \\
&\geq \frac{1}{\beta+\Omega}\Bigg\{h(i+1) + \frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i+2,j'}(k) + \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right)V_{i+1,j}(k) \\
&\qquad + c(\mu^*_{i+1,j}(k)) - \mu^*_{i+1,j}(k)\Delta_{i+1,j}(k) \\
&\qquad - h(i) - \frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i+1,j'}(k) - \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right)V_{i,j}(k) \\
&\qquad - c(\mu^*_{i+1,j}(k)) + \mu^*_{i+1,j}(k)\Delta_{i,j}(k)\Bigg\} \\
&= \frac{1}{\beta+\Omega}\Bigg\{h(i+1) - h(i) + \frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\Delta_{i+2,j'}(k) \\
&\qquad + \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right)\Delta_{i+1,j}(k) - \mu^*_{i+1,j}(k)\left(\Delta_{i+1,j}(k) - \Delta_{i,j}(k)\right)\Bigg\}.
\end{aligned} \tag{E.6}$$

Similarly, we get that

$$\begin{aligned}
\Delta_{i,j}(k+1) \leq \frac{1}{\beta+\Omega}\Bigg\{&h(i) - h(i-1) + \frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\Delta_{i+1,j'}(k) + \left(\Omega - \frac{\lambda}{\eta^{j-1}}\right)\Delta_{i,j}(k) \\
&- \mu^*_{i-1,j}(k)\left(\Delta_{i,j}(k) - \Delta_{i-1,j}(k)\right)\Bigg\}.
\end{aligned} \tag{E.7}$$

Subtracting these two inequalities, we get that

$$\begin{aligned}
(\beta+\Omega)(\Delta_{i+1,j}(k+1) - \Delta_{i,j}(k+1)) \geq \Bigg(&(h(i+1) - h(i)) - (h(i) - h(i-1)) \\
&+ \frac{\lambda}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}[\Delta_{i+2,j'}(k) - \Delta_{i+1,j'}(k)] \\
&+ \left(\Omega - \frac{\lambda}{\eta^{j-1}} - \mu^*_{i+1,j}(k)\right)[\Delta_{i+1,j}(k) - \Delta_{i,j}(k)] \\
&+ \mu^*_{i-1,j}(k)[\Delta_{i,j}(k) - \Delta_{i-1,j}(k)]\Bigg).
\end{aligned} \tag{E.8}$$

By the induction hypothesis, convexity of $h(i)$, and the fact that $\Omega \geq \lambda/\eta^{j-1} - \mu^*_{i+1,j}(k)$,

we get that $\Delta_{i+1,j}(k+1) \geq \Delta_{i,j}(k+1)$. Our assertion therefore follows by mathematical induction.

E.2 Proof of Monotonicity of the Optimal Flow Control Policy

In this section, we show that for a power-law queueing system servicing a Poisson input, the optimal rate of admission $\lambda^*_{i,j}$ is monotonic in the number of customers, given that the holding cost function $h(i)$ is convex, i.e.,

$$h(i) - h(i-1) \leq h(i+1) - h(i)$$

for every $i = 1, 2, \ldots$. More precisely, for every $i = 1, 2, \ldots,$

$$\lambda^*_{i-1,j} \geq \lambda^*_{i,j}.$$

Using exactly the same argument as in Section E.1, we can show that $\lambda^*_{i,j}$ is nonincreasing in i, if the first difference of $V_{i,j}$ is nondecreasing in i. We proceed to show that the first difference $\Delta_{i,j}$, defined as in Section E.1, is monotonically nondecreasing. Our approach will exploit the value iteration method, and prove by mathematical induction that the first difference of $V_{i,j}(k)$ at each stage of the iteration, is monotonic in i. Again, $\Delta_0(k)$ is set to be zero for all k. Also, it is clear that $\Delta_{i,j}(0)$ is nondecreasing in i, since $V_{i,j}(0)$ is identically zero. We assume the assertion holds for k, and analyze

$$\Delta_{i,j}(k+1) - \Delta_{i-1,j}(k+1).$$

For each k, we also define $V_{0,j}(k) \triangleq V_0(k)$, and $\Delta_{0,j}(k) \triangleq \Delta_0(k)$.

To this end, we write the value iteration equations as

$$V_0(k+1) = \frac{1}{\beta+\Omega}\left\{\Omega V_0(k) + \min_{\lambda\in[0,\overline{\lambda}]}\left[c(\lambda) + \lambda\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\left(V_{1,j'}(k) - V_0(k)\right)\right]\right\} \tag{E.9}$$

$$\begin{aligned} V_{i,j}(k+1) = \frac{1}{\beta+\Omega}\Bigg\{h(i) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i-1,j'}(k) + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)V_{i,j}(k) \\ + \min_{\lambda\in[0,\overline{\lambda}]}\left[c(\lambda) + \lambda\Delta_{i+1,j}(k)\right]\Bigg\}, \quad i = 1,2,\ldots \end{aligned} \tag{E.10}$$

We define the optimal service rate at each stage of the value iteration method as $\lambda^*_{i,j}(k)$, i.e.,

$$\lambda^*_{i,j}(k) \stackrel{\Delta}{=} \arg\min_{\lambda\in[0,\overline{\lambda}]}\left[c(\lambda) + \lambda\Delta_{i+1,j}(k)\right]. \tag{E.11}$$

Now, for $i > 1$, we get that

$$\begin{aligned} V_{i+1,j}(k+1) = \frac{1}{\beta+\Omega}\Bigg\{h(i+1) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i,j'}(k) + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)V_{i+1,j}(k) \\ + c(\lambda^*_{i+1,j}) + \lambda^*_{i+1,j}\Delta_{i+2,j}(k)\Bigg\}, \end{aligned}$$

and

$$\begin{aligned} V_{i,j}(k+1) = \frac{1}{\beta+\Omega}\Bigg\{h(i) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}V_{i-1,j'}(k) + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)V_{i,j}(k) \\ + c(\lambda^*_{i,j}) + \lambda^*_{i,j}\Delta_{i+1,j}(k)\Bigg\}. \end{aligned}$$

Subtracting, we get that

$$\begin{aligned} \Delta_{i+1,j}(k+1) &= V_{i+1,j}(k+1) - V_{i,j}(k+1) \\ &\geq \frac{1}{\beta+\Omega}\Bigg\{h(i+1) - h(i) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\Delta_{i,j'}(k) \\ &\quad + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)\Delta_{i+1,j}(k) + \lambda^*_{i+1,j}(k)\left(\Delta_{i+2,j}(k) - \Delta_{i+1,j}(k)\right)\Bigg\}. \end{aligned}$$

Similarly,

$$\begin{aligned}\Delta_{i,j}(k+1) &= V_{i,j}(k+1) - V_{i-1,j}(k+1) \\ &\leq \frac{1}{\beta+\Omega}\Bigg\{h(i) - h(i-1) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\Delta_{i-1,j'}(k) + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)\Delta_{i,j}(k) \\ &\qquad + \lambda^*_{i-1,j}(k)\left(\Delta_{i+1,j}(k) - \Delta_{i,j}(k)\right)\Bigg\}.\end{aligned}$$

Thus,

$$\begin{aligned}(\beta+\Omega)(\Delta_{i+1,j}(k+1) - \Delta_{i,j}(k+1)) \geq \Bigg(&(h(i+1) - h(i)) - (h(i) - h(i-1)) \\ &+ \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}[\Delta_{i,j'}(k) - \Delta_{i-1,j'}(k)] \\ &+ \left(\Omega - \frac{\mu}{\eta^{j-1}} - \lambda^*_{i-1,j}(k)\right)[\Delta_{i+1,j}(k) - \Delta_{i,j}(k)] \\ &+ \lambda^*_{i+1,j}(k)[\Delta_{i+2,j}(k) - \Delta_{i+1,j}(k)]\Bigg).\end{aligned} \tag{E.12}$$

Thus, it follows from the induction hypothesis, convexity of $h(i)$, and the fact that $\Omega \geq \frac{\mu}{\eta^{j-1}} + \lambda^*_{i-1,j}(k)$, that $\Delta_{i,j}(k+1)$ is nondecreasing for $i \geq 2$. To complete the proof, we now show that $\Delta_{2,j}(k+1) \geq \Delta_{1,j}(k+1)$. We note that

$$\begin{aligned}\Delta_{2,j}(k+1) &= V_{2,j}(k+1) - V_{1,j}(k+1) \\ &\geq \frac{1}{\beta+\Omega}\Bigg\{h(2) - h(1) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\Delta_{1,j'}(k) + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)\Delta_{2,j}(k) \\ &\qquad + \lambda^*_{2,j}(k)\left(\Delta_{3,j}(k) - \Delta_{2,j}(k)\right)\Bigg\},\end{aligned}$$

and

$$\begin{aligned}\Delta_{1,j}(k+1) &= V_{1,j}(k+1) - V_{0,j}(k+1) \\ &\leq \frac{1}{\beta+\Omega}\left\{h(1) + \frac{\mu}{\eta^{j-1}}\sum_{j'=1}^{L}\sigma^2 q^{j'-1}\Delta_{0,j'}(k) + \left(\Omega - \frac{\mu}{\eta^{j-1}}\right)\Delta_{1,j}(k)\right. \\ &\qquad \left. + \lambda^*_{0,j}(k)\left(\Delta_{2,j}(k) - \Delta_{1,j}(k)\right)\right\}.\end{aligned}$$

Thus, again, our assertion follows for the case $i = 1$, for the same reasons.

Bibliography

[1] R. G. Addie and R. E. Warfield. Bandwidth switching and new network architectures. *Teletraffic Sci.*, ITC-12:665–671, 1989.

[2] J. H. Ahrens and U. Dieter. Sampling from binomial and Poisson distributions: A method with bounded computation times. *Computing*, 25:193–208, 1980.

[3] A. C. Atkinson. The computer generation of Poisson random variables. *Appl. Stat.*, 28(1):29–35, 1979.

[4] R. E. Bellman. *Dynamic programming and modern control theory.* Academic Press, New York, NY, 1965.

[5] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Trans. Communications*, 43(2/3/4):1566–1579, 1995.

[6] D. P. Bertsekas. *Dynamic Programming and Optimal Control.* Athena Scientific, Belmont, Massachusetts, 1995.

[7] D. P. Bertsekas and R. G. Gallager. *Data Networks.* Prentice-Hall International, London, 1992.

[8] D. R. Cox. *Renewal Theory.* Wiley, London, 1962.

[9] D. R. Cox. *Point Processes.* Chapman and Hall, London, 1980.

[10] D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events.* Methuen, London, 1966.

[11] B. T. Doshi. Continuous time control of the arrival process in an M/G/1 queue. *Stoch. Proc. and their Applic.*, 5:265–284, 1977.

[12] B. T. Doshi. Optimal control of the service rate in an M/G/1 queueing system. *Adv. Appl. Prob.*, 10:682–701, 1978.

[13] A. Ephremides, P. Varaiya, and J. Walrand. A simple dynamic routing problem. *IEEE Trans. Auto. Cont.*, 25:690–693, 1980.

[14] M. Fleischmann, D.J. Tildesley, and R.C. Ball. *Fractals in the Natural Sciences : a Discussion.* Princeton University Press, Princetone, NY, 1989.

[15] R. G. Gallager. *Discrete Stochastic Processes.* Kluwer Academic Publishers, Boston, Massachusetts, 1996.

[16] I. M. Gel'fand, G. E. Shilov, N. Ya. Vilenkin, and M. I. Graev. *Generalized Functions.* Academic Press, New York, NY, 1964.

[17] L. L. Horowitz. *Optimal Filtering of Gyroscopic Noise.* PhD thesis, M. I. T., Cambridge, MA, 1974.

[18] J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometics*, 29(3):339–349, 1987.

[19] R. A. Howard. *Dynamic programming and Markov processes.* Technology Press of Massachusetts Institute of Technology, Cambridge, Ma, 1960.

[20] J. Y. Hui. Resource allocation for broadband networks. *IEEE Trans. Communications*, 6(9):1598–1608, 1988.

[21] A. E. Jacquin. Fractal image coding: A review. *Proc. IEEE*, 81(10):1451–1465, October 1993.

[22] D. H. Johnson and A. R. Kumar. Modeling and analyzing fractal point processes. In *Proc. Int. Conf. Acoust. Speech, Signal Processing*, 1990.

[23] D. M. Kern. Minimum variance unbiased estimation in the Pareto distribution. *Metrika*, 30:15–19, 1983.

[24] M. S. Keshner. $1/f$ noise. *Proc. IEEE*, 70:212–218, 1982.

[25] A. Y. Khintchine. *Mathematical Methods in the Theory of Queueing.* Griffin, London, 1960.

[26] L. Kleinrock. *Queueing Systems*, volume I. John Wiley & Sons, New York, 1975.

[27] N. M. Laird, A. P. Dempster, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm". *Ann. Roy. Stat. Soc.*, pages 1–38, December 1977.

[28] A. M. Law and W. D. Kelton. *Simulation modeling and analysis.* McGraw-Hill, New York, NY, 1991.

[29] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Networking*, 2(1):1–15, 1994.

[30] S. B. Lowen and M. C. Teich. Fractal renewal processes generate $1/f$ noise. *Physical Review E*, 47:992–1001, Feb. 1993.

[31] H. J. Malik. Estimation of the parameters of the Pareto distribution. *Metrika*, 15:126–132, 1970.

[32] B. Mandelbrot. Self-similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Trans. Comm. Tech.*, 13:71–90, Mar. 1965.

[33] B. B. Mandelbrot. *The Fractal Geometry of Nature.* W. H. Freeman, San Francisco, 1983.

[34] B. Mitchell. Optimal service-rate selection in an M/G/1 queue. *SIAM Jour. Appl. Math.*, 24:19–35, 1973.

[35] M. Neuts. *Matrix-Geometric Solutions in Stochastic Models.* Johns Hopkins University Press, Baltimore, Maryland, 1981.

[36] B. Ninness. Analysis of wavelet based maximum likelihood estimation of $1/f$ noise. Technical report, University of Newcastle, Australia, 1996.

[37] V. Paxson and S. Floyd. Wide area traffic: The failure of Poisson modeling. *IEEE/ACM Trans. Networking*, 3(3):226–244, 1995.

[38] H.-O. Peitgen and D. Saupe. *The Science of Fractal Images.* Springer Verlag, New York, NY, 1988.

[39] Jr. S. Stidham. Optimal control of admission to a queueing system. *IEEE Trans. Auto. Cont.*, 30:705–713, 1985.

[40] M. Schroeder. *Fractals, Chaos, Power Laws.* W. H. Freeman, New York, NY, 1991.

[41] D. L. Snyder. *Random Point Processes.* John Wiley & Son, Inc., New York, New York, 1975.

[42] S. M. Sussman. Analysis of the Pareto model for error statistics on telephone circuits. *IEEE Trans. Comm. Sys.*, 11:213–221, Jun. 1963.

[43] M. C. Teich. Fractal character of the auditory neural spike train. *IEEE Trans. Bio. Eng.*, 36:150–160, Jan. 1989.

[44] M. C. Teich, C. Heneghan, S. B. Lowen, and R. G. Turcott. Estimating the fractal exponent of point processes in biological systems using wavelet- and fourier-transform methods. In A. Aldroubi and M. Unser, editors, *Wavelets in Medicine and Biology*, pages 383–412. CRC Press, 1996.

[45] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson. Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *Computer communication review*, 25(4):100–113, 1995.

[46] G. W. Wornell. Wavelet-based representations for the $1/f$ family of fractal processes. *Proc. IEEE*, 81(10):1428–1450, October 1993.

[47] G. W. Wornell. *Signal Processing with Fractals : a Wavelet-Based Approach.* Prentice Hall PTR, Upper Saddle River, NJ, 1996.