Multiscale Representation and Estimation of Fractal Point Processes

Warren M. Lam and Gregory W. Wornell, *Member, IEEE*

***Abstract*—Fractal point processes have a potentially important role to play in the modeling of a wide range of natural and man-made phenomena. However, the lack of a suitable framework for their representation has frequently made their application in many problems difficult. We introduce natural multiscale representations for an important class of these processes based on mixtures of Poisson processes. In turn, this framework leads to efficient new algorithms for both the synthesis and the analysis of such processes. These include algorithms for optimal fractal dimension and interarrival time estimation that are of interest in a range of applications. Several aspects of the performance of these algorithms are also addressed.**

I. Introduction

An extraordinary range of natural and man-made phenomena lack any characteristic temporal or spatial scale. Such phenomena and their inherent statistical scale invariance are naturally modeled using the mathematics of self-similar and fractal geometry. One set of fractal random processes that are useful in signal modeling applications are those for which the associated waveforms are continuous-valued. Important classes of these processes are $1/f$ processes and fractional Brownian motions, and multiscale representations for these processes have proven to be extremely useful both conceptually and practically; see, e.g., [1].

Another set of fractal random processes that are useful in modeling applications are discrete-valued. Of particular interest are fractal point processes—collections of points or "events" randomly distributed temporally and/or spatially without a characteristic scale. Examples of phenomena well-modeled in this way are abundant, including the distributions of stars and planets in the universe, transmission errors on communications channels, and impulsive spikes in auditory neural signals [2], [3], [4], [5]. As we will develop in this paper, multiscale representations turn out to be an equally useful and efficient tool in the synthesis, analysis, and processing of these fractal random processes.

While no universal framework for modeling fractal point processes exists, a variety of approaches for such modeling have been pursued and developed in the literature. For example, Johnson *et al.* generate a point process with fractal characteristics from a doubly stochastic process—specifically, a nonhomogeneous Poisson process whose arrival rate function is a suitably chosen random process with scale-invariant characteristics [6]. This framework has proven useful in modeling several aspects of the spike trains observed in auditory neural signals [7]. However, the development of efficient signal processing algorithms using this framework has traditionally proved difficult.

In a somewhat different approach, Mandelbrot constructs a fractal point process from a renewal process with interarrival times governed by a Pareto (i.e., power law) distribution [4]. To accommodate the fact that Pareto random variables are not asymptotically integrable, the concept of conditional stationarity is introduced. This model, whose second-order statistics were subsequently explored in [8], has been used to model error clustering in the telephone network [9] among other applications. However, this model is generally more useful as a conceptual characterization for some fractal point processes than as a framework for the development of signal processing algorithms. As a consequence, the fractal point process definition we develop at the outset of this paper is closely related in spirit to this model.

As the main contribution of this work, we introduce and develop a multiscale framework for modeling fractal point processes. In particular, we consider the decomposition of a fractal point process into a mixture of constituent homogeneous Poisson processes, each of which contributes the features associated with a specific scale. As we shall see, because of the inherent scale invariance of the process, there is an important statistical scaling relationship among the constituents. As we will demonstrate, this framework is especially well-suited to the development of a variety of signal processing algorithms for such fractal processes.

The outline of this paper is as follows. In Section II, we define the particular class of fractal point processes of interest in this work, and develop some of their key characteristics. In Section III, we develop multiscale constructions for these processes and some of their important features in the context of signal synthesis. In Section IV, we demonstrate how these multiscale representations are useful in addressing some fundamental signal analysis and estimation problems involving fractal point processes. Finally, Section V contains some concluding remarks.

II. Conditionally-Renewing Fractal Point Processes

We begin by developing a useful and sufficiently formal definition of the class of fractal point processes of interest

Manuscript received July 21, 1995; revised April 10, 1995. This work was supported by the Advanced Research Projects Agency monitored by ONR under Contract No. N00014-93-1-0686, and the Air Force Office of Scientific Research under Grant No. AFOSR-91-0034. The associate editor coordinating the review of this paper and approving it for publication was Prof. Banu Onural.

The authors are with the Department of Electrical Engineering and Computer Science, and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

IEEE Log Number 9415084.

in this paper. For simplicity, we restrict our attention to the 1-D case, viewing the independent variable as time. Note, however, that multidimensional analogs can be constructed using straightforward extensions of the theory. For example, constructions similar to the Levy flight discussed in [2] and [3] can yield multidimensional point processes useful in a number of applications.

In general, a point process consists of a random collection of points distributed over the time axis, each of which we shall refer to as an "event," or "arrival." Furthermore, we shall choose our time origin to coincide with an arrival, and consider only $t \geq 0$. This choice of origin represents only a mild restriction that is for convenience of exposition here. While there are a variety of ways to characterize a point process, a particularly useful one is in terms of the interarrival intervals. In particular, we let $X[n]$ denote the time interval between the $(n-1)$st arrival and the nth arrival, with the zeroth arrival corresponding to the time origin. A related characterization of a point process is in terms of its event times, or arrival epochs, i.e., the time $S_X[n]$ at which the nth arrival occurs. Hence, we have

$$S_X[n] = \sum_{k=1}^{n} X[k].$$

An alternative and often useful characterization of a point process is in terms of the associated counting process $N_X(t)$, whose value at any time instant t is the total number of arrivals up to and including time t. Frequently, one is also interested in the generalized time-derivative of this process, which, when the arrivals are isolated, consists of a train of unit impulses located at the arrival epochs. Among the various relationships between these characterizations, we have, e.g.

$$N_X(t) = \sup\{n : S_X[n] \leq t\}.$$

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:

$$N_X(t) \stackrel{\mathcal{P}}{=} N_X(at) \tag{1}$$

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, (1) 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. Since renewal processes—i.e., processes with independent, identically-distributed interarrivals—are widely used to generate stationary point process models [10], it is tempting to restrict our attention to those self-similar point processes that are simultaneously renewal 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 $Y[n]$ is derived from a point process with interarrivals $X[n]$ via *conditioning on the event* $\mathcal{E}$ if $Y[n]$ is the subsequence of $X[n]$ 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 $X[n]$ 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 $\mathcal{F}_1, \mathcal{F}_2, \cdots, \mathcal{F}_L$ 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 paper we shall focus on self-similar point processes that 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.

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 $Y[n]$ of the resulting process is given by

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

where γ is real and σ_Y^2 is a normalization factor.

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) implies that the common interarrival probability density for the $X[n]$ 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.,

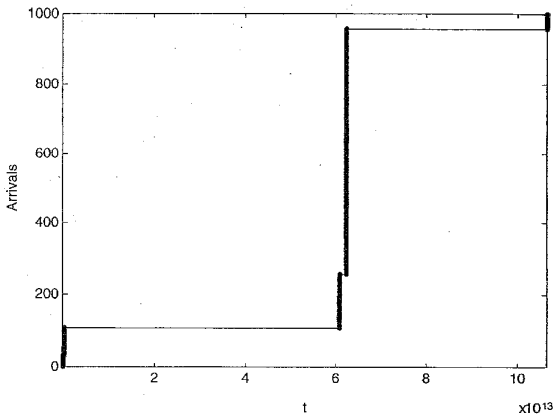


Fig. 1. Sample function showing the first 1000 arrivals of a fractal renewal process with shape parameter $\gamma = 1.2$.Note that the low fractal dimension ($D = 0.2$) is manifested in the strong conglomeration of arrivals into sparsely distributed clusters.

[3]. 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 capture 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 [2]

$$D = \gamma - 1. \tag{3}$$

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. 1, we illustrate a typical sample function of a fractal renewal process with shape parameter $\gamma = 1.2$. For viewing convenience, we have plotted the associated counting process. In this example, interarrivals shorter than one time unit are discarded and the resulted first 1000 points are shown. The strong clustering observed is typical of point processes having fractal dimension near 0. In such cases, arrivals are packed into sparsely located clusters and hence occupy little space. From the perspective of interarrival distributions, the strong clustering reflects an interarrival density that gives rise to an abundance of very short interarrivals separated by occasional very long interarrivals.

Although in practice we often observe values of γ in the range $1 < \gamma < 2$, it is important 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 (3) 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.

III. Multiscale Synthesis of Fractal Renewal Processes

In this section, we develop efficient algorithms for synthesizing fractal renewal processes from a multiscale family of Poisson processes. The use of Poisson processes as constituents is both natural and convenient—these memoryless point processes are highly tractable analytically and a wide range of algorithms exist for their processing. In particular, Poisson processes can be synthesized extremely efficiently; see, e.g., [11] and [12]. In the sequel, we present both an exact multiscale synthesis based on a continuum of constituent processes, and an approximate synthesis based on a discretization approach.

A. Exact Synthesis: The Continuous Mixture

We begin with a continuum collection of independent homogeneous Poisson processes that are statistically similar up to a dilation or compression. More specifically, denoting the collection of counting processes as $N_{W_A}(t)$, where A is an indexing parameter, we require that

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

where $N_{W_0}(t)$ is a prototype Poisson process whose mean arrival rate we denote by λ. Hence, it immediately follows that the mean arrival rate of $N_{W_A}(t)$ is

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

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

$$W_A \stackrel{\mathcal{P}}{=} e^{A}W_0.$$

To synthesize a fractal renewal process $N_X(t)$, our algorithm mainly involves a random mixture of these constituent Poisson processes. This is achieved with a sequence of independent, identically distributed random variables $A[n]$ that are independent of the processes $N_{W_A}(t)$, and are distributed according to the generalized-exponential density

$$f_A(a) = \begin{cases} \sigma_A^2\, e^{-(\gamma-1)a} & \underline{a} \le a \le \overline{a} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where σ_A^2 is a normalization constant. More specifically, the random variable $A[n]$ 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 $A[1]$, i.e.

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

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 $N_{W_{A[2]}}(t)$, i.e.

$$S_X[2] = S_{W_{A[2]}}[k]$$

where k is the smallest index such that

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

Subsequent arrivals are selected in a similar manner. Consequently, the sequence $A[n]$ can be interpreted as a sequence of scale indices that determine how the constituent at each scale contributes in the synthesis.

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

Theorem 2: 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

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

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 $x > 0$

$$\frac{f_X(x)}{\sigma_A^2} \to \frac{\sigma_0^2}{x^\gamma} \tag{5}$$

where

$$\sigma_0^2 = \frac{\Gamma(\gamma - 1)}{\lambda^{\gamma-1}}. \tag{6}$$

Several remarks are appropriate. First, we observe that the limit (5) is independent of λ, the rate of the prototype process. However, not surprisingly, the convergence is not uniform; specifically, (5) with (6) is a good approximation to the interarrival density for values of x satisfying

$$e^{\underline{a}}/\lambda \ll x \ll e^{\overline{a}}/\lambda \tag{7}$$

as is apparent from the derivation in Appendix B.

Finally, we note that the result (5) can also be obtained from two variations of this multiscale construction. In particular, 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

$$X[n] = W_{A[n]}[n] \tag{8}$$

and

$$X[n] = e^{A[n]} W_0[n]. \tag{9}$$

Note that while the characteristic feature of fractal renewal processes is their power law interarrival densities, (9) suggests a synthesis requiring only exponential random variables that, in practice, can be derived from a single Poisson process. As another interesting observation, the construction (9) 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.* [6].

B. Approximate Synthesis: The Discrete Mixture

While in Section III-A we constructed true fractal renewal processes by mixing Poisson processes on a continuum of scales, we now demonstrate that arbitrarily good approximations can, in fact, be realized using only a discrete set of constituents. As we will see, this discrete mixture is often very convenient in practice.

We denote our discrete collection of constituent Poisson processes by $N_{W_M}(t)$, where M is an integer-valued scale index. Again, we choose these processes to be statistically similar; specifically we let

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

where without loss of generality, the scale increment ρ will be assumed greater than one. Hence, the corresponding mean arrival rates are related according to

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

where λ can be chosen freely.

Based on these constituents, the synthesis is carried out in essentially the same manner as the continuous mixture of Section III-A, except that the sequence of scale indexes $M[n]$ are now independent identically-distributed random variables with a common generalized-geometric distribution of the form

$$p_M(m) \stackrel{\Delta}{=} \Pr[M = m] = \begin{cases} \sigma_M^2 \rho^{-(\gamma-1)m} & m = \underline{m}, \underline{m}+1, \cdots, \overline{m} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where σ_M^2 is a normalizing constant. This discrete synthesis is justified by the following theorem, a proof of which is contained in Appendix C.

Theorem 3: 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

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

where

$$S_X[0] = 0.$$

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

$$f_X(x) = \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} \lambda \rho^{-\gamma m} \exp\left(-\lambda \rho^{-m} x\right) \tag{11}$$

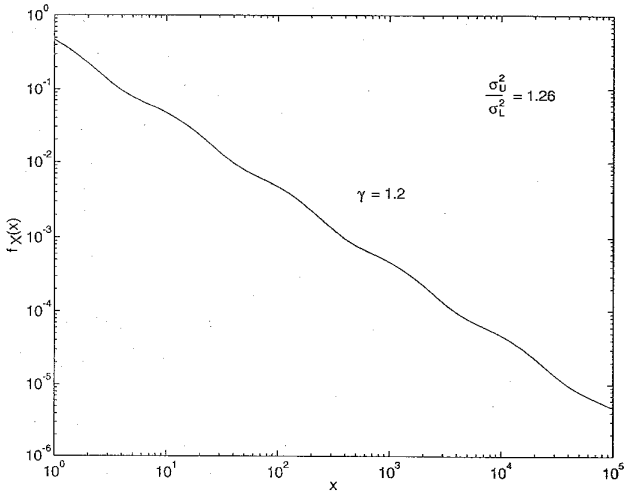


Fig. 2. Log-log plot of the interarrival density of a point process constructed with the discrete multiscale synthesis algorithm. In this case, the shape parameter is $\gamma = 1.2$ and the scale increment is $\rho = 10$. On the log scale, the peak-to-peak ripple size is estimated to be $\log 1.26$, while the period of the ripple is $\log \rho = \log 10$.

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 $x > 0$

$$\frac{\sigma_L^2}{x^\gamma} \leq \frac{f_X(x)}{\sigma_M^2} \leq \frac{\sigma_U^2}{x^\gamma} \tag{12}$$

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

Several remarks regarding Theorem 3 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 practice a finite collection suffices to approximate any finite interarrival range of interest.

Second, we note that the selection of the scale increment ρ involves a tradeoff. Smaller values of ρ give rise to finer approximations, i.e., empirical evaluation of the tightest bounding constants σ_L^2 and σ_U^2 suggests that

$$\sigma_U^2/\sigma_L^2 \to 1 \qquad \text{as } \rho \to 1+ \ .$$

However, smaller values of ρ also mean that more constituent scales are required to approximate any given interarrival range of interest. Fig. 2 shows the interarrival density corresponding to the case $\rho = 10$ and $\gamma = 1.2$. Ripple size—and, hence, approximation error—in fact decreases rapidly as $\rho \to 1+$. When $\rho = 2$, for example, the approximation is already exceptionally good; numerical calculations yield $\log(\sigma_U^2/\sigma_L^2) = 3.2e-5$ for the case $\gamma = 1.2$!

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

$$x^\gamma f_X(x)/\sigma_M^2 = (\rho x)^\gamma f_X(\rho x)/\sigma_M^2.$$

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 (1) for all a of the form $a = \rho^m$ where m is an integer.

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 to (9), interarrivals of the form

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

are generated, where the $M[n]$ are the random sequence of scale indexes and the $W_0[n]$ are the interarrivals of the prototype process.

IV. Multiscale Analysis of Fractal Renewal Processes

In this section, we demonstrate some of the ways in which the discrete multiscale framework of Section III-B is an equally important tool in the analysis of fractal renewal processes. In particular, we focus on the application of the framework to two practical estimation problems. First, estimation of the shape parameter is considered in Section IV-A, and a maximum-likelihood estimator is developed for the purpose. Next, an algorithm for obtaining minimum mean-square error interarrival estimates is presented in Section IV-B. While the estimation problems will be dealt with separately, we shall see that they are in fact rather intimately related.

A. Parameter Estimation for Fractal Renewal Processes

Estimation of the shape parameter γ is well motivated. As discussed in Section II, this parameter is directly related to the fractal dimension and, hence, often carries useful information about a point process. For example, the shape parameter associated with an auditory neural pulse train can be used to infer certain properties of the external stimulus [5]. In addition, knowledge of γ is required in intermediate stages of many other detection and estimation problems involving such processes. While parameter estimation techniques exist when perfect observations are available (see, e.g., [13]), we consider the more general case in which the observations are noise-corrupted. Such distortion is explicitly taken into account to improve estimator robustness.

The particular estimation problem we consider is as follows. Given a sequence of N observations $R[n]$ modeled as

$$R[n] = X[n] + W[n], \qquad n = 1, 2, \cdots, N$$

where the $X[n]$ are the first-order interarrival times of a fractal renewal process and the $W[n]$ represent some form of additive distortion, we seek an estimate of the shape parameter associated with the $X[n]$. 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

$$f_W(w) = \begin{cases} \alpha e^{-\alpha w} & w \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

with $\alpha > 0$. It is well known that such a random variable has second moment $2/\alpha^2$, that can be interpreted as a measure of noise strength. This noise component can be used for modeling

a variety of natural effects that arise in applications, such as a random processing delay in an interarrival measurement transducer.

We remark that the estimator in [13] 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 this case is difficult to manipulate. While our approach to this problem is also based on the method of maximum-likelihood (ML), we exploit the result of Section III-B and model the $X[n]$ with a finite-scale representation. For consistency, we shall follow the notational convention of Section III-B throughout the development of our estimator. In addition, we introduce, for convenience, a new parameter β defined as

$$\beta = \rho^{1-\gamma}$$

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 β via

$$\hat{\gamma}_{\mathrm{ML}} = 1 - \ln \hat{\beta}_{\mathrm{ML}} / \ln \rho$$
$$\hat{D}_{\mathrm{ML}} = -\ln \hat{\beta}_{\mathrm{ML}} / \ln \rho.$$

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 $\boldsymbol{\Theta} = (\lambda, \alpha, \beta)^T$.

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

$$\ell(\boldsymbol{\Theta}) = \sum_{n=1}^{N} \ln f_R(r[n]; \boldsymbol{\Theta}) \tag{14}$$

where using (11) and (13) we have, for $r > 0$

$$f_R(r; \boldsymbol{\Theta}) = \int_0^r f_X(\tau)\, f_W(r-\tau)\, d\tau$$
$$= \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} \beta^m f_{R|M}(r \mid m; \boldsymbol{\Theta}) \tag{15}$$

with

$$f_{R|M}(r \mid m; \boldsymbol{\Theta}) = \begin{cases} \dfrac{\lambda_m \alpha}{\lambda_m - \alpha}\left[e^{-\alpha r} - e^{-\lambda_m r}\right] & \lambda_m \neq \alpha \\ \alpha^2 r e^{-\alpha r} & \text{otherwise} \end{cases} \tag{16}$$

where, as before, $\lambda_m = \rho^{-m}\lambda$. Without loss of generality we may set $\underline{m} = 1$, since λ may be scaled accordingly. The total number of scales required, which we denote by $\overline{m} = L$, is typically determined from the spread of the data. As will become apparent, overestimating L generally does not affect the estimation performance, although the corresponding algorithm is less efficient in terms of both computation and storage.

Calculation of the ML parameter estimates by directly maximizing (14) with (15) and (16) is difficult. However, these estimates can be efficiently computed using an iterative estimate-maximize (EM) algorithm [14]. As the parameter estimates are updated every iteration, we shall use the notation

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

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 D, 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., (17), as shown at the bottom of the page, where $\sigma^2_{M|R}$ and $\tilde{\sigma}^2_{M|R}$ are normalization terms.

M-step: Using the preceding table of probability estimates as weights, compute new estimates of the parameters via (18), as shown at the bottom of the following page, where

$$E\left[x[n] \,\middle|\, r[n], m; \hat{\boldsymbol{\Theta}}\right] = \begin{cases} r[n]\left[\dfrac{1}{(\hat{\lambda}_m - \hat{\alpha})r[n]} - \dfrac{1}{e^{(\hat{\lambda}_m - \hat{\alpha})r[n]} - 1}\right] & \hat{\lambda}_m \neq \hat{\alpha} \\ r[n]/2 & \text{otherwise.} \end{cases} \tag{19}$$

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.

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 (18b). This becomes more apparent when we explore the interarrival time estimation problem in Section IV-B.

$$p_{M|R}(m \mid r[n]; \hat{\boldsymbol{\Theta}}_{[k]}) = \begin{cases} \sigma^2_{M|R} \dfrac{\hat{\lambda}_{m[k]}\hat{\alpha}_{[k]}}{\hat{\lambda}_{m[k]} - \hat{\alpha}_{[k]}}\left[e^{-\hat{\alpha}_{[k]}r[n]} - e^{-\hat{\lambda}_{m[k]}r[n]}\right]\dfrac{1-\hat{\beta}_{[k]}}{1-\hat{\beta}^L_{[k]}}\hat{\beta}^{m-1}_{[k]} & \hat{\lambda}_{m[k]} \neq \hat{\alpha}_{[k]} \\ \tilde{\sigma}^2_{M|R}\hat{\lambda}^2_{m[k]} r[n] e^{-\hat{\lambda}_{m[k]}r[n]} \dfrac{1-\hat{\beta}_{[k]}}{1-\hat{\beta}^L_{[k]}}\hat{\beta}^{m-1}_{[k]} & \text{otherwise} \end{cases} \tag{17}$$

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 in a nondecreasing manner. In this particular application, empirical evidence suggests that over a nonpathological 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.

While analytic performance bounds for the estimator are in general difficult to obtain, in the special case of noise-free observations the Cramér–Rao bounds can be readily deduced using the exact power law density (2). In this case, we have, in particular

$$\operatorname{var}\hat{\gamma} \geq (\gamma - 1)^2/N. \tag{20}$$

Although this bound is somewhat loose, empirical studies have shown that in accordance with (20), variance in our estimate of γ generally decays for increasing sample size N. Fig. 3 summarizes the results of a set of Monte Carlo simulations conducted to explore the dependence of estimator performance on the sample size N, and true shape parameter γ. The experiments basically involved application of the estimator to simulated noise-corrupted interarrivals. To ensure that modeling error effects were included in the simulations, both the power-law random variables and exponential noise terms in the test data were synthesized via transformation of uniform random variables. Throughout this set of experiments, we used a dyadic scale representation for the interarrivals (i.e., $\rho = 2$), and set the signal and noise parameters λ and α such that $\alpha/\lambda = 1e-4$. The values plotted in the graph are the resulting RMS errors in the estimate $\hat{\gamma}_{\mathrm{ML}}$ as computed over 32 trials. While the simulation results confirm the decay of $\operatorname{var}\hat{\gamma}_{\mathrm{ML}}$ for increasing N, they also suggest a bias in the estimator that arises primarily due to the inherent difference between the discrete multiscale representation and the exact power law density.

The apparent relationship between the estimator performance and the true shape parameter γ can be understood as follows. When $\gamma \approx 1$, extremely long and extremely short interarrivals occur with comparable frequency. In this case, very few data points are sufficient to capture the behavior of the probability density function over a broad range. For $\gamma \approx 2$, however, short interarrivals predominate over longer ones. Hence, a narrower range of the distribution is generally observed, making it difficult to accurately estimate γ.

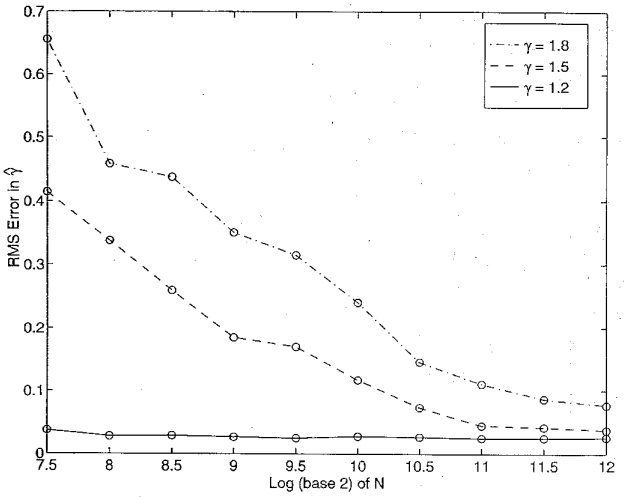


Fig. 3. Performance of the multiscale EM parameter estimator for a noise-corrupted fractal renewal process based on various numbers of interarrival measurements N.Each circle represents the RMS error in the shape parameter estimate $\hat{\gamma}$ as computed from 32 Monte Carlo trials on simulated data. The signal and noise parameters λ and α were chosen so that $\alpha/\lambda = 1e-4$.

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. 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. As before, each point in the graph represents the RMS error computed from 32 Monte Carlo trials. As expected, the estimator performance improves as α/λ increases.

$$\frac{1}{\hat{\lambda}_{[k+1]}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\boldsymbol{\Theta}}_{[k]})\, \rho^{-m}\, E\Big[x[n] \;\Big|\; r[n], m; \hat{\boldsymbol{\Theta}}_{[k]}\Big] \tag{18a}$$

$$\frac{1}{\hat{\alpha}_{[k+1]}} = \frac{1}{N}\sum_{n=1}^{N}\left(r[n] - \sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\boldsymbol{\Theta}}_{[k]}) E\Big[x[n] \;\Big|\; r[n], m; \hat{\boldsymbol{\Theta}}_{[k]}\Big]\right) \tag{18b}$$

$$\frac{1}{1-\hat{\beta}_{[k+1]}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{L} p_{M|R}(m \mid r[n]; \hat{\boldsymbol{\Theta}}_{[k]})\, m \tag{18c}$$

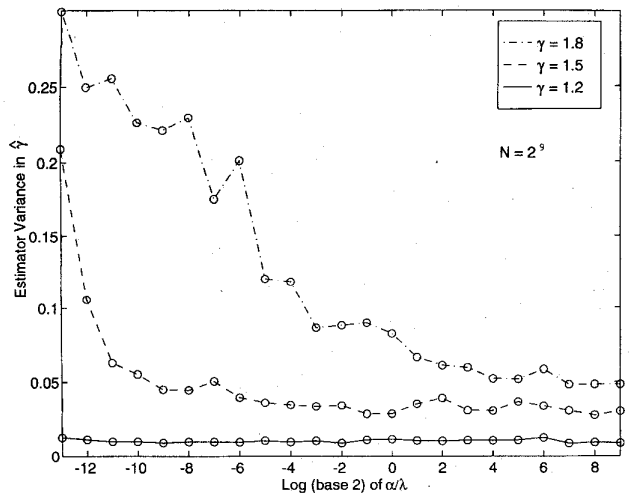


Fig. 4. Performance of the multiscale EM parameter estimator for a noise-corrupted fractal renewal process under various SNR. In this case, the ratio between the noise and signal parameters α/λ gives a measure of SNR. Each circle represents the variance in the shape parameter estimate $\hat{\gamma}$ as computed from 32 Monte Carlo trials on simulated data. The number of interarrival measurements used per trial was $N = 2^9$.

B. Interarrival Time Estimation

Parameter estimation and interarrival time estimation are closely related problems. In Section IV-A, we 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 mean-square error estimate of an interarrival interval given noisy measurements of the type considered in Section IV-A. In particular, we obtain (21), as shown at the bottom of the page, where $f_R(r[n])$ is obtained via (15) with (16), and

$$\Phi_m(r) = \begin{cases} \dfrac{1}{(\lambda_m-\alpha)^2}\left[1-e^{-(\lambda_m-\alpha)r}(1+(\lambda_m-\alpha)r)\right] & \lambda_m \neq \alpha \\ r^2/2 & \text{otherwise.} \end{cases} \tag{22}$$

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 IV-A can be used.

As a minimum mean-square error estimator, the algorithm specified by (21) with (22) is unbiased. Further, the variance of the estimator error can be computed in a straightforward manner. In particular, we have

$$E\left[(x[n]-\hat{x}[n])^2\right] = E\left[x[n]^2\right] - E\left[\hat{x}[n]^2\right] \tag{23}$$

where

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

and $E\left[\hat{x}[n]^2\right]$ can be computed numerically using (21) with (22) and (15). By comparing the initial noise variance $1/\alpha^2$ with the value obtained with (23), we can compute the theoretical achievable SNR gain, which we have plotted in Fig. 5. 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, a lower initial SNR and 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. 5 suggests that in scenarios of low initial SNR, the SNR gain is, to a good approximation, a function of α/λ.

V. Concluding Remarks

Using the multiscale framework introduced in this paper, we have developed several efficient new algorithms for the synthesis and analysis of fractal renewal processes that are potentially useful in a wide range of signal processing applications. While the basic viability of this framework and the associated algorithms was established through a series of preliminary simulations, many additional issues—both theoretical and practical—remain to be explored. First, while we have pointed out some potentially interesting connections between our multiscale representation and the point process models of Mandelbrot [4] and Johnson *et al.* [6], investigation of deeper and clearer relations with existing fractal point process models remains an important topic in the future study of our model. In addition, there are many opportunities where additional analysis could provide useful insights into the algorithms we have proposed. For example, in the discrete synthesis theorem, the derived bounding constants are not particularly tight, and, in particular, do not establish the property $\sigma_U^2/\sigma_L^2 \to 1$ as

$$\hat{x}[n] = E[x[n] \mid r[n]] = \int_0^{r[n]} \frac{x}{f_R(r[n])} \sum_{m=\underline{m}}^{\overline{m}} p_M(m) f_{X|M}(x \mid m) f_{R|X,M}(r[n] \mid x, m)\,dx$$

$$= \sigma_M^2 \frac{\alpha e^{-\alpha r[n]}}{f_R(r[n])} \sum_{m=\underline{m}}^{\overline{m}} \beta^m \lambda_m \Phi_m(r[n]) \tag{21}$$

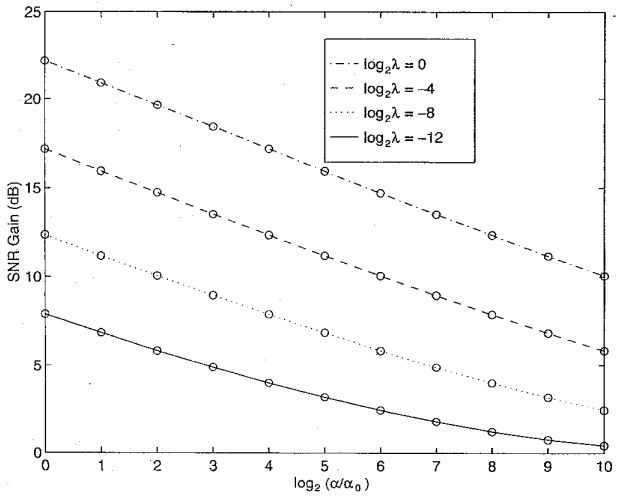


Fig. 5. Performance of the interarrival time estimator for a noise-corrupted fractal renewal process under various combinations of signal and noise parameters λ and α. Each circle represents the achievable SNR gain computed by comparing the initial error variance that is $1/\alpha^2$,and the theoretical error variance in the estimates, which is determined numerically. In this case, the shape parameter is $\gamma = 1.5$ and the normalization constant is $\alpha_0 = 2.1e\text{-}7$.

$\rho \to 1+$ that is observed in practice. As another example, in the estimation algorithms, useful analytic performance bounds would be valuable for predicting the behavior in various scenarios. In addition, many additional experiments involving these algorithms with both simulated and real data are needed to further explore their properties.

Finally, there are many remaining aspects of fractal renewal processes in general that remain to be explored. Many characteristics of these processes have proven difficult or awkward to describe from traditional representations. Some examples are the probability distribution of the counting process and the statistics of the process under random incidence or with erasures. Multiscale representations may well provide important insights into these issues as well.

Appendix A
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

$$f_{Y_a}(y) = (1/a)\, f_{Y_1}(y/a). \tag{24}$$

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 $n_1, n_2, \cdots, n_K$ denote an arbitrary collection of interarrival indexes and $x_1, x_2, \cdots, x_K$ an arbitrary collection of constants. Then, (25)–(29) as shown at the bottom of the page follow, where (25) is a consequence of the definition of $Y_a[n]$, (26) is a consequence of (1), and (29) is a consequence of the fact that the $Y_1[n]$ constitute a renewal process. From (29) we can immediately conclude that the the $Y_a[n]$ are independent and identically-distributed. Finally, differentiating both sides of (29) we get (24). □

We are now ready for the main proof.

Let M_L and M_U be integers such that

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

where $\alpha = \overline{x}/\underline{x}$.

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

$$f_{\tilde{Y}}(y) = \sum_m f_{Y_{\alpha^m}}(y) \cdot \Pr\left[\mathcal{E}_{\alpha^m} \,\middle|\, \tilde{\mathcal{E}}\right]. \tag{30}$$

Finally, since

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

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

$$f_Y(y) = \begin{cases} f_{\tilde{Y}}(y)/\Pr\left[\mathcal{E} \,\middle|\, \tilde{\mathcal{E}}\right] & x_L < y \leq x_U \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

$$\begin{aligned}
&\Pr[Y_a[n_k] < x_k, k = 1, 2, \cdots, K] \\
&\quad = \Pr\left[X[n_k] < x_k, k = 1, 2, \cdots, K \mid a\underline{x} < X[n_k] \leq a\overline{x}, k = 1, 2, \cdots, K\right] && (25)\\
&\quad = \Pr\left[aX[n_k] < x_k, k = 1, 2, \cdots, K \mid a\underline{x} < aX[n_k] \leq a\overline{x}, k = 1, 2, \cdots, K\right] && (26)\\
&\quad = \Pr\left[X[n_k] < x_k/a, k = 1, 2, \cdots, K \mid \underline{x} < X[n_k] \leq \overline{x}, k = 1, 2, \cdots, K\right] && (27)\\
&\quad = \Pr[Y_1[n_k] < x_k/a, k = 1, 2, \cdots, K] && (28)\\
&\quad = \prod_{k=1}^{K} \Pr[Y_1 < x_k/a] && (29)
\end{aligned}$$

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 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 $a\underline{x} < y \leq \overline{x}$

$$f_{Y_*}(y) = f_{Y_1}(y)/\Pr[\mathcal{E}_* \mid \mathcal{E}_1] \tag{32}$$

and

$$f_{Y_*}(y) = f_{Y_a}(y)/\Pr[\mathcal{E}_* \mid \mathcal{E}_a]. \tag{33}$$

Equating (32) and (33), cross-multiplying, and exploiting (24) we get, for $a\underline{x} < y \leq \overline{x}$

$$f_{Y_1}(y)\int_{a\underline{x}}^{\overline{x}} f_{Y_1}(x/a)\,dx = f_{Y_1}(y/a)\int_{a\underline{x}}^{\overline{x}} f_{Y_1}(x)\,dx. \tag{34}$$

Differentiating (34) with respect to a and letting $a \to 1+$ we get, for $\underline{x} < y \leq \overline{x}$

$$f_{Y_1}(y)\int_{\underline{x}}^{\overline{x}} x\,f'_{Y_1}(x)\,dx = y\,f'_{Y_1}(y)\int_{\underline{x}}^{\overline{x}} f_{Y_1}(x)\,dx. \tag{35}$$

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

$$\frac{y f'_{Y_1}(y)}{f_{Y_1}(y)} = \gamma \tag{36}$$

where γ is the constant obtained by integrating (35) by parts

$$\gamma = \frac{\overline{x}f_{Y_1}(\overline{x}) - \underline{x}f_{Y_1}(\underline{x})}{\displaystyle\int_{\underline{x}}^{\overline{x}} f_{Y_1}(x)\,dx} - 1. \tag{37}$$

Now, all regular, positive solutions to (36) on $\underline{x} < y \leq \overline{x}$ can be obtained by separation of variables and are of the form [15]

$$f_{Y_1}(y) = \sigma_1^2/y^\gamma \tag{38}$$

where σ_1^2 is a normalization constant. In addition, it can be readily verified by substitution of (38) into (37) that γ is a free parameter.

Finally, combining (38) with (24), and substituting the result into (30) and, in turn, into (31), we obtain (2). □

Appendix B
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 $N_{W_A}(t)$, 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 $A[n] = a$, $X[n]$ is the waiting time until the next arrival in $N_{W_a}(t)$ after $t = S_X[n-1]$. Because a Poisson process is memoryless, the density for this waiting time is

$$f_{X[n]|A[n]}(x \mid a) = \begin{cases} e^{-a}\lambda \exp[-e^{-a}\lambda x] & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{39}$$

which we note is independent of n. Using (4), (39), and a change of variables we obtain

$$\begin{aligned} f_X(x) &= \int_{\underline{a}}^{\overline{a}} f_A(a)\,f_{X|A}(x \mid a)\,da \\ &= \frac{1}{x^\gamma}\frac{\sigma_A^2}{\lambda^{\gamma-1}}\int_{e^{-\overline{a}}\lambda x}^{e^{-\underline{a}}\lambda x} u^{\gamma-1}\,e^{-u}\,du \end{aligned} \tag{40}$$

from which we see that for each $x > 0$, when $\underline{a} \to -\infty$ and $\overline{a} \to \infty$, we obtain (5) with (6). Furthermore, note that (40) also establishes (7).

Appendix C
Proof of Theorem 3

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

$$\begin{aligned} f_X(x) &= \sum_{m=\underline{m}}^{\overline{m}} p_M(m)\,f_{X|M}(x \mid m) \\ &= \sigma_M^2 \sum_{m=\underline{m}}^{\overline{m}} \rho^{-(\gamma-1)m}\,f_{W_m}(x) \end{aligned} \tag{41}$$

where

$$f_{W_m}(x) = \lambda\rho^{-m}\exp(-\lambda\rho^{-m}x) \tag{42}$$

for $x \geq 0$. Substituting (42) into (41) immediately yields (11).

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

$$x = \rho^{m_0}x_0 \tag{43}$$

where m_0 is an integer, and $1 \leq x_0 < \rho$. Using (43) in (42), it is then straightforward to show that

$$f_{W_m}(x) = \rho^{-m_0}f_{W_{m-m_0}}(x_0). \tag{44}$$

Taking limits and substituting (44) into (41) with (42), we get that

$$\begin{aligned} \tilde{f}_X(x) &\triangleq \lim_{\substack{\underline{m}\to-\infty \\ \overline{m}\to\infty}} f_X(x)/\sigma_M^2 \\ &= \sum_{m=-\infty}^{\infty} \rho^{(1-\gamma)m}f_{W_m}(x) \end{aligned} \tag{45}$$

$$= \rho^{-m_0\gamma} \sum_{m=-\infty}^{\infty} \rho^{(1-\gamma)m - m_0 + m_0\gamma} f_{W_{m-m_0}}(x_0)$$
$$= \rho^{-m_0\gamma} \tilde{f}_X(x_0). \quad (46)$$

Multiplying both sides of (46) by x^γ and again exploiting (43), we get

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

and, in turn

$$\frac{1}{x^\gamma} \cdot \left[\inf_{1 \le x < \rho} \tilde{f}_X(x) \right] \le \tilde{f}_X(x) \le \frac{\rho^\gamma}{x^\gamma} \cdot \left[\sup_{1 \le x < \rho} \tilde{f}_X(x) \right]. \quad (47)$$

Thus, it remains only to bound the two terms in brackets in (47).

We begin with the lower bound. Since $\rho^{(1-\gamma)m} f_{W_m}(x) > 0$ for every m, we have from (45) that, for $1 \le x < \rho$

$$\tilde{f}_X(x) > \rho^{(1-\gamma)0} f_{W_0}(x)$$
$$> \lambda e^{-\rho\lambda}.$$

Hence

$$\inf_{1 \le x < \rho} \tilde{f}_X(x) \ge \lambda e^{-\rho\lambda} > 0.$$

To derive the upper bound, we begin by noting that

$$\tilde{f}_X(x) = \xi_1 + \xi_2$$

where

$$\xi_1 = \sum_{m=0}^{\infty} \rho^{(1-\gamma)m} f_{W_m}(x) \quad (48)$$

$$\xi_2 = \sum_{m=1}^{\infty} \rho^{-(1-\gamma)m} f_{W_{-m}}(x). \quad (49)$$

Thus, to show

$$\sup_{1 \le x < \rho} \tilde{f}_X(x) < \infty$$

it suffices to show that (48) and (49) are convergent for $x \ge 1$.

We have immediately that $\xi_1 < \infty$ since, for $x > 0$ and $\gamma > 0$

$$\sum_{m=0}^{\infty} \rho^{(1-\gamma)m} f_{W_m}(x) \le \lambda \sum_{m=0}^{\infty} \rho^{-\gamma m} = \frac{\lambda}{1 - \rho^{-\gamma}}.$$

To see that $\xi_2 < \infty$ we first note that, for $x \ge 1$

$$\xi_2 = \lambda \sum_{m=1}^{\infty} \rho^{\gamma m} \exp(-\lambda \rho^m x)$$
$$\le \lambda \sum_{m=1}^{\infty} \rho^{\gamma m} \exp(-\lambda \rho^m). \quad (50)$$

Finally, applying the ratio test for series [16] to (50), we see that $\xi_2 < \infty$ because

$$\lim_{m \to \infty} \frac{\rho^{\gamma(m+1)} \exp(-\lambda \rho^{m+1})}{\rho^{\gamma m} \exp(-\lambda \rho^m)}$$
$$= \lim_{m \to \infty} \rho^\gamma \exp\left(-\lambda(\rho - 1)\rho^m\right) = 0.$$

□

Appendix D
Derivation of the EM Parameter Estimation Algorithm

Each iteration of the EM algorithm involves the computation and maximization of the function

$$\Psi(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}}) \triangleq E\left[\ln f_{\mathbf{R},\mathbf{S}}(\mathbf{r}, \mathbf{s}; \boldsymbol{\Theta}) \,\middle|\, \mathbf{r}; \hat{\boldsymbol{\Theta}} \right] \quad (51)$$

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

$$\mathbf{r} = \{r[n],\ n = 1, 2, \cdots N\}$$
$$\mathbf{s} = \{x[n], m[n],\ n = 1, 2, \cdots N\}$$

leads to

$$\ln f_{\mathbf{R},\mathbf{S}}(\mathbf{r}, \mathbf{s}; \boldsymbol{\Theta}) = \sum_{n=1}^{N} \Big\{ \ln(1 - \beta) + (m[n] - 1) \ln \beta + \ln\left(\lambda \rho^{-m[n]} \exp(-\lambda \rho^{-m[n]} x[n]) \right) + \ln\left(\alpha \exp(-\alpha(r[n] - x[n])) \right) \Big\}$$
$$= N(\ln(1 - \beta) - \ln \beta + \ln \lambda + \ln \alpha) - \lambda \sum_{n=1}^{N} \rho^{-m[n]} x[n] + (\ln \beta - \ln \rho) \cdot \sum_{n=1}^{N} m[n] - \alpha \sum_{n=1}^{N} (r[n] - x[n]) \quad (52)$$

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 (52) into (51) we obtain, in turn

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

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.

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

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

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

$$\hat{\lambda}_{[k+1]} = \frac{N}{\sum_{n=1}^{N} E\left[\rho^{-m[n]} x[n] \,\middle|\, r[n]; \hat{\boldsymbol{\Theta}}_{[k]} \right]} \quad (53a)$$

$$\hat{\alpha}_{[k+1]} = \frac{N}{\sum_{n=1}^{N} E\left[r[n] - x[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]} \tag{53b}$$

$$\hat{\beta}_{[k+1]} = \frac{\sum_{n=1}^{N} \left(E\left[m[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]\right) - N}{\sum_{n=1}^{N} E\left[m[n] \,\middle|\, r[n]; \hat{\Theta}_{[k]}\right]}. \tag{53c}$$

Evaluating the expectations in (53) iteratively by first conditioning on $m[n]$, we then obtain (18).

For the computation of $E\left[x[n] \,\middle|\, r[n], m; \hat{\Theta}\right]$, it is straightforward to show that this quantity can be expressed in the form

$$E\left[x[n] \,\middle|\, r[n], m; \hat{\Theta}\right] = \frac{\int_0^{r[n]} x\, f_{R|X,M}(r[n] \mid x, m; \hat{\Theta})\, f_{X|M}(x \mid m; \hat{\Theta})\, dx}{\int_0^{r[n]} f_{R|X,M}(r[n] \mid x, m; \hat{\Theta})\, f_{X|M}(x \mid m; \hat{\Theta})\, dx} \tag{54}$$

where

$$f_{R|X,M}(r \mid x, m) = f_W(r - x) \tag{55a}$$

$$f_{X|M}(x \mid m) = f_{W_m}(x). \tag{55b}$$

Using (54) with (55), we then obtain (19).

Having now obtained $\hat{\lambda}_{[k+1]}$, $\hat{\alpha}_{[k+1]}$, and $\hat{\beta}_{[k+1]}$ in terms of $p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]})$ and previous estimates, it remains only to determine the update equation for $p_{M|R}(m \mid r[n]; \hat{\Theta}_{[k]})$. This is readily obtained from Bayes Rule, i.e.

$$p_{M|R}(m \mid r[n]; \hat{\Theta}) = \frac{f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}{\sum_{m=1}^{L} f_{R|M}(r[n] \mid m; \hat{\Theta}) p_M(m; \hat{\Theta})}. \tag{56}$$

Using (56) together with

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

we obtain (17).

REFERENCES

[1] G. W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proc. IEEE*, vol. 81, pp. 1428–1450, Oct. 1993.

[2] B. B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA: Freeman, 1982.

[3] M. Schroeder, *Fractals, Chaos, Power Laws*. New York, NY: W. H. Freeman, 1991.

[4] B. B. Mandelbrot, "Self-similar error clusters in communication systems and the concept of conditional stationarity," *IEEE Trans. Commun.*, vol. COM-13, pp. 71–90, Mar. 1965.

[5] M. C. Teich, "Fractal character of the auditory neural spike train," *IEEE Trans. Biomed. Eng.*, pp. 150–160, Jan. 1989.

[6] D. H. Johnson and A. R. Kumar, "Modeling and analyzing fractal point processes," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, vol. 3, 1990, pp. 1353–1356.

[7] M. C. Teich, D. H. Johnson, A. R. Kumar, and R. G. Turcott, "Rate fluctuations and fractional power-law noise recorded from cells in the lower auditory pathway of the cat," *Hearing Res.*, vol. 46, pp. 41–52, June 1990.

[8] S. B. Lowen and M. C. Teich, "Fractal renewal processes generate $1/f$ noise," *Phys. Rev. E*, vol. 47, pp. 992–1001, Feb. 1993.

[9] S. M. Sussman, "Analysis of the Pareto model for error statistics on telephone circuits," *IEEE Trans. Commun.*, vol. CS-11, no. 2, pp. 213–221, June 1963.

[10] S. M. Ross, *Stochastic Processes*. New York, NY: Wiley, 1983.

[11] J. H. Ahrens and U. Dieter, "Sampling from binomial and Poisson distributions: A method with bounded computation times," *Computing*, vol. 25, pp. 193–208, 1980.

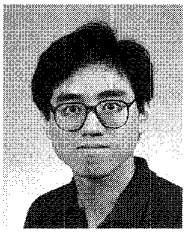
[12] A. C. Atkinson, "The computer generation of Poisson random variables," *Appl. Stat.*, vol. 28, no. 1, pp. 29–35, 1979.

[13] H. J. Malik, "Estimation of the parameters of the Pareto distribution," *Metrika*, vol. 15, pp. 126–132, 1970.

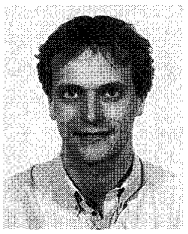
[14] N. M. Laird, A. P. Dempster, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Ann. Roy. Stat. Soc.*, pp. 1–38, Dec. 1977.

[15] I. M. Gelfand, G. E. Shilov, N. Y. Vilenkin, and M. I. Graev, *Generalized Functions*. New York: Academic, 1964.

[16] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 3rd ed., 1976.



Warren M. Lam was born in Monroe, Louisiana, in 1966. He received the B.Sc. degree from the University of California, Berkeley, in 1989, and the S.M. degree from the Massachusetts Institute of Technology, in 1992, both in electrical engineering. He is currently a doctoral student in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology, where he holds a NSF Graduate Fellowship. His research interests include signal processing and applied probability.



Gregory W. Wornell (S'83–M'85) was born in Montreal, Canada, in 1962. He received the B.A.Sc. degree (with honors) from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering, in 1985, 1987, and 1991, respectively. Since 1991, he has been on the faculty at MIT, where he is currently ITT Career Development Assistant Professor in the Department of Electrical Engineering and Computer Science. During the 1992–93 academic year, he was on leave at AT&T Bell Laboratories, Murray Hill, NJ, and during 1990 he was a Visiting Investigator at the Woods Hole Oceanographic Institution, Woods Hole, MA. His current research interests include signal processing, broadband and wireless communications, and applications of fractal geometry and nonlinear dynamical system theory in these areas. He is a consultant to industry and is an inventor on one patent in the area of communications; two others are pending.

From 1985 to 1989, he held a 1967 Science and Engineering Scholarship from the Natural Sciences and Engineering Research Council of Canada, and, in 1991, he received the MIT Goodwin Medal for "conspicuously effective teaching." In 1995, he received an NSF Faculty Early Career Development Award and an AT&T New Research Fund Award.

Dr. Wornell is a member of Tau Beta Pi and Sigma Xi.