Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond
by Jianbo Gao, Yinhe Cao, Wen-wen Tung and Jing Hu

CHAPTER 11

MODELS OF POWER-LAW-TYPE BEHAVIOR

Due to the ubiquity of heavy-tailed distribution as well as $1/f^{\beta}$ noise, modeling of such power-law-type behavior is important for two basic reasons: (1) A good model may shed new light on the mechanism for such power-law-type behavior and (2) an explicit, succinct account of such behavior may greatly simplify a pattern recognition/classification problem. In this chapter, we describe a few simple models for power-law-type behavior.

11.1 MODELS FOR HEAVY-TAILED DISTRIBUTION

In Chapter 7, we discussed stable distributions. We explained that each stable distribution can be considered an attractor, with the normal distribution being at the edge of the family of stable distributions. Any stable distribution can be a good model for a heavy-tailed distribution. In this section, we describe five more simple models.

11.1.1 Power law through queuing

In queuing theory, the simplest and most fundamental model is the Poisson traffic model, or the M/M/1 queuing system, where "M" stands for exponential distribu-

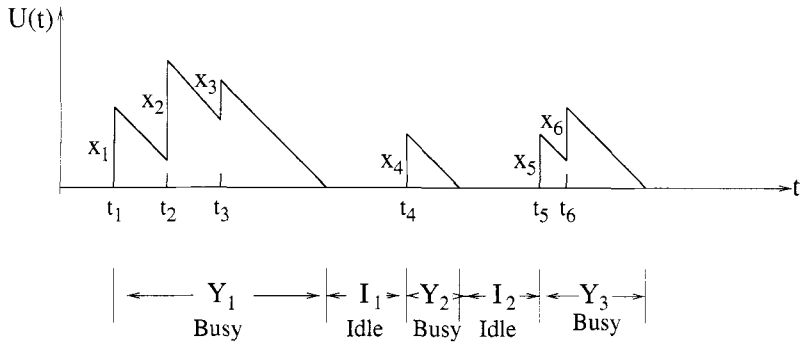

Figure 11.1. Schematic of alternating busy and idle periods for a queuing system with a constant service rate. $U(t)$ denotes an unfinished load of the system.

tion; the interarrival time follows an exponential distribution with rateλ, while the service time follows another exponential distribution with rateμ. More specifically, assume that at time instants t_1, $t_2, \cdots$, packets arrive at a queuing system. The interarrival times, defined as $T_i = t_{i+1} - t_i$, are exponentially distributed with CDF $1 - e^{-\lambda t}$. The packet size x_i, $i = 1, 2, \cdots$, in units of time measuring how soon the packet can be transmitted, follows another exponential distribution with CDF $1 - e^{-\mu t}$. The utilization level ρ is defined as λ/μ. It is well known that the busy period (which is schematically depicted in Fig. 11.1) of such a queuing system has the following distribution:

$$p(t) = \frac{\rho^{-1/2}}{t} e^{-(\lambda+\mu)t} I_1[2t\sqrt{\lambda\mu}], \tag{11.1}$$

where I_1 is the modified Bessel function of the first kind (of order 1). When$z \to \infty$, $I_1(z) \approx \frac{1}{\sqrt{2\pi}} z^{-1/2} e^z$. Therefore,

$$p(t) \sim t^{-3/2} e^{-\sqrt{\mu}(1-\sqrt{\rho})^2 t}, \quad t \to \infty. \tag{11.2}$$

That is, the tail of the distribution is a power law, followed by an exponential truncation. The truncation becomes less and less relevant when$\rho \to 1$. In fact, when $\rho = 1$, the tail becomes a pure power law. Comparing Eq. (11.2) with Eqs. (3.23) and (3.24), we note that $\alpha = 0.5$. The tail is in fact the same as that of the Levy distribution (see Eq. (7.4)).

11.1.2 Power law through approximation by log-normal distribution

Montroll and Shlesinger [316] have found that the power-law distribution with the parameter $\alpha = 0$ can be approximated by the log-normal distribution. Let$\log x$ have a normal distribution, with mean$\log \overline{x}$ and variance σ^2. Their idea is to find

the distribution for $x/\overline{x}$. Denote the PDF for $x/\overline{x}$ by $g(x/\overline{x})$. We have

$$\frac{e^{-[\log(x/\overline{x})]^2/2\sigma^2}}{(2\pi\sigma^2)^{1/2}} d\Big(\log\frac{x}{\overline{x}}\Big) = \frac{e^{-[\log(x/\overline{x})]^2/2\sigma^2}}{(2\pi\sigma^2)^{1/2}} \frac{dx/\overline{x}}{x/\overline{x}} = g\Big(\frac{x}{\overline{x}}\Big) d\Big(\frac{x}{\overline{x}}\Big). \quad (11.3)$$

Therefore,

$$\log[g(x/\overline{x})] = -\log(x/\overline{x}) - [\log(x/\overline{x})]^2/2\sigma^2 - \log(2\pi\sigma^2)/2. \quad (11.4)$$

When x is measured in multiples f of its mean $\overline{x}$, $x = f\overline{x}$,

$$\log[g(f)] = -\log f - (\log f)^2/2\sigma^2 - \log(2\pi\sigma^2)/2. \quad (11.5)$$

When the second term is negligible compared with the first term, the distribution is a power law, $g(f) \sim f^{-1}$. Comparing Eq. (11.5) with Eq. (3.23), we see that $\alpha = 0$. To neglect the second term, we may let $\sigma^2 \to \infty$. When σ^2 is large but finite, the power-law relation is valid in a finite range of f.

11.1.3 Power law through transformation of exponential distribution

As we have seen in Chapter 3, many basic models of stochastic processes involve exponential distributions. For example, a Poisson process is defined through exponential interarrival times, while the sojourn times of Markov processes follow exponential distributions. Exponential distributions play an even more fundamental role in physics, since the basic laws in statistical mechanics and quantum mechanics are expressed as exponential distributions, while finite spin glass systems are equivalent to Markov chains. Given the prevalence of power-law-type behavior in complex interconnected systems, it is of fundamental importance to find a way to derive power-law distributions from exponential distributions. We develop a simple and elegant scheme here.

Recall that when a random variable T has an exponential distribution with parameter $\lambda > 0$, its complementary cumulative distribution function (CCDF) or tail probability is

$$P(T \geq t) = e^{-\lambda t}, \quad t \geq 0, \quad \lambda > 0.$$

In many situations, the parameter λ may also be a random variable with a PDF $p(\lambda)$. Then the tail probability for T becomes

$$P(T \geq t) = \int_0^\infty e^{-\lambda t} p(\lambda) d\lambda. \quad (11.6)$$

This is the very one-sided Laplace transform of $p(\lambda)$! With Eq. (11.6), the problem of transforming an exponential distribution into a power-law distribution involves finding Laplace transforms of PDFs $p(\lambda)$ with power-law-type behavior. This can be easily accomplished by looking up tables of Laplace transforms. There are many functions whose Laplace transform has a power-law tail. Such power-law behavior

may be said to come from the heterogeneity of the constituents of a complicated system.

For concreteness, in the following, we examine four simple functional forms for $p(\lambda)$.

1. λ is uniformly distributed within $[\lambda_{min}, \lambda_{max}]$:

$$\begin{aligned} p(T \geq t) &= \int_{\lambda_{min}}^{\lambda_{max}} e^{-\lambda t} \frac{1}{\lambda_{max} - \lambda_{min}} d\lambda \\ &= \frac{1}{t} \frac{e^{-\lambda_{min} t} - e^{-\lambda_{max} t}}{\lambda_{max} - \lambda_{min}}. \end{aligned} \tag{11.7}$$

We now consider three regimes of t:

(a) $t < 1/\lambda_{min}$: using Taylor series expansion, one readily finds that $p(T \geq t) \approx 1$. This corresponds to the power law with the exponent $\alpha = 0$.

(b) $1/\lambda_{max} < t < 1/\lambda_{min}$: in this case, we may drop the term $e^{-\lambda_{max} t}$ and approximate the term $e^{-\lambda_{min} t}$ by 1. Then $p(T \geq t) \sim t^{-1}$.

(c) $t > 1/\lambda_{min}$: in this regime, exponential decay dominates.

In conclusion, $p(T \geq t)$ follows a truncated power law.

2. λ is exponentially distributed within $[\lambda_{min}, \lambda_{max}]$:

$$\begin{aligned} p(T \geq t) &= c \int_{\lambda_{min}}^{\lambda_{max}} e^{-\lambda t} \gamma e^{-\gamma \lambda} d\lambda \\ &= \frac{c\gamma}{t + \gamma} [e^{-(t+\gamma)\lambda_{min}} - e^{-(t+\gamma)\lambda_{max}}]. \end{aligned} \tag{11.8}$$

c is the normalization constant satisfying

$$c \int_{\lambda_{min}}^{\lambda_{max}} \gamma e^{-\gamma \lambda} d\lambda = 1.$$

Thus

$$c = \frac{1}{e^{-\gamma \lambda_{min}} - e^{-\gamma \lambda_{max}}}.$$

Equation (11.8) then becomes

$$p(T \geq t) = \frac{\gamma}{t + \gamma} \frac{e^{-(t+\gamma)\lambda_{min}} - e^{-(t+\gamma)\lambda_{max}}}{e^{-\gamma \lambda_{min}} - e^{-\gamma \lambda_{max}}}. \tag{11.9}$$

Note that case 1 can be obtained from this case by taking $\gamma = 0$. Following the discussion of case 1, it is then clear that $p(T \geq t)$ here also follows a truncated power law.

3. λ follows the gamma distribution:
The PDF of the gamma distribution is given as

$$f(\lambda) = \frac{1}{\Gamma(\alpha)}\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}, \quad \lambda \geq 0, \ \alpha > 0, \ \beta > 0,$$

where $\Gamma(t)$ is the gamma function:

$$\Gamma(t) = \int_0^{\infty} y^{t-1}e^{-y}dy.$$

The Laplace transform for the gamma distribution can be readily found from a mathematics handbook. It is given by

$$P(T \geq t) = \left(1 + \frac{t}{\beta}\right)^{-\alpha}. \tag{11.10}$$

It is clear that when t is large, $p(T \geq t)$ follows a power law.

4. $1/\lambda$, which is the mean of T, follows the Pareto distribution of Eq. (3.24). Here it is more convenient to work with the PDF, which is given by

$$f(t) = \alpha b^{\alpha}t^{-\alpha-1}, \quad t \geq b > 0, \tag{11.11}$$

where α is the shape parameter and b is the location parameter. Recall that when $0 < \alpha < 2$, the variance is unbounded. When $0 < \alpha < 1$, the mean also diverges. With the distribution for $1/\lambda$ given by Eq. (11.11), one can easily find the distribution for λ to be

$$p(\lambda) = \alpha b^{\alpha}\lambda^{\alpha-1}, \quad \lambda \leq 1/b. \tag{11.12}$$

Then

$$p(T \geq t) = \int_0^{1/b} e^{-\lambda t}\alpha b^{\alpha}\lambda^{\alpha-1}d\lambda. \tag{11.13}$$

In simulations, we may choose b very close to 0; then Eq. (11.13) can be approximated as

$$\begin{aligned} p(T \geq t) &= \int_0^{\infty} e^{-\lambda t}\alpha b^{\alpha}\lambda^{\alpha-1}d\lambda \\ &= \alpha b^{\alpha}t^{-\alpha}\Gamma(\alpha), \end{aligned} \tag{11.14}$$

where $\Gamma(t)$ is the gamma function. It is interesting to note that T and $1/\lambda$ have the same type of heavy-tailed behavior (i.e., the exponent α is the same).

11.1.4 Power law through maximization of Tsallis nonextensive entropy

An interesting way of obtaining a heavy-tailed distribution is to maximize the Tsallis entropy under a few simple constraints. The Tsallis entropy is a generalization of the Shannon (or Boltzmann-Gibbs) entropy. Mathematically, it is closely related to the Renyi entropy. Let us define all these entropies first.

Assume that there are m distinctive events, each occurring with probability p_i. The Shannon entropy is defined by

$$H = -\sum_{i=1}^{m} p_i \log p_i, \tag{11.15}$$

where the unit for H is a bit or baud corresponding to base 2 or e in the logarithm. Without loss of generality, we shall choose base e.

Next, we consider the Renyi entropy, defined by

$$H_q^R = \frac{1}{1-q} \log\Big(\sum_{i=1}^{m} {p_i}^q\Big). \tag{11.16}$$

H_q^R has a number of interesting properties:

- When $q = 1$, H_1^R is the Shannon entropy: $H_1^R = H$.
- $H_0^R = \log(m)$ is the topological entropy.
- If $p_1 = p_2 = \cdots = p_m = \frac{1}{m}$, then for all real valued q, $H_q^R = \log(m)$.
- In the case of unequal probability, H_q^R is a nonincreasing function of q. In particular, if we denote

$$p_{max} = \max_{1 \le i \le m} (p_i), \quad p_{min} = \min_{1 \le i \le m} (p_i),$$

then

$$\lim_{q \to -\infty} H_q^R = -\log(p_{min}), \quad \lim_{q \to \infty} H_q^R = -\log(p_{max}).$$

Finally, let us define the Tsallis entropy. It is given by

$$H_q^T = \frac{1}{q-1}\Big(1 - \sum_{i=1}^{m} {p_i}^q\Big). \tag{11.17}$$

In the continuous case, it can be written as

$$H_q^T = \frac{1}{q-1}\Big(1 - \int_{-\infty}^{\infty} d\Big(\frac{x}{\sigma}\Big)[\sigma p(x)]^q\Big). \tag{11.18}$$

The Tsallis entropy is often called extended entropy formalism. Tsallis and co-workers assert that this formalism has a number of remarkable mathematical

properties. For example, it preserves the Legendre transformation structure of thermodynamics and keeps the Ehrenfest theorem, von Neumann equation, and Onsgaer reciprocity theorem in the same form with all values of q.

The Renyi entropy and the Tsallis entropy are related by the following simple equation:

$$H_q^R = \frac{\ln\left[1+(1-q)H_q^T\right]}{1-q}, \qquad \lim_{q\to 1} H_q^R = \lim_{q\to 1} H_q^T = -\sum_{i=1}^{m} p_i \ln p_i.$$

Although mathematically the Renyi and Tsallis entropies seem to be equivalent, we note an important distinction. Under the Shannon entropy formalism, the object or situation under consideration is assumed to be fully random. The Renyi entropy aims to better characterize the situation by resorting to a spectrum of index q to emphasize the difference in p_i. While the Tsallis entropy also has this function, it focuses more on characterizing a type of motion whose complexity is neither regular nor fully chaotic/random by finding a specific q, often different than 1, that best describes the motion. Motions that are neither regular nor fully chaotic/random are ubiquitous. A distinguished class of examples are the fractal processes with long memory.

We now derive the Tsallis distribution under the following two constraints:

$$\int_{-\infty}^{\infty} p(x)dx = 1, \tag{11.19}$$

$$\sigma^2 = \frac{\int_{-\infty}^{\infty} x^2[p(x)]^q dx}{\int_{-\infty}^{\infty}[p(x)]^q dx}. \tag{11.20}$$

The first constraint simply says that the total probability is 1. The second constraint defines the second normalized moment in the framework of extended formalism. It can be written as

$$\int_{-\infty}^{\infty}[x^2-\sigma^2][p(x)]^q dx = 0.$$

To maximize the Tsallis entropy, we introduce two Lagrange multipliers, λ_1, λ_2, and write $y = p(x)$. We then have

$$\begin{aligned} S(y) &= \frac{1}{q-1}\Big(1-\int_{-\infty}^{\infty} d(\frac{x}{\sigma})[\sigma y]^q\Big) \\ &+\lambda_1\int_{-\infty}^{\infty} y dx + \lambda_2\int_{-\infty}^{\infty}(x^2-\sigma^2)y^q dx \\ &= \frac{1}{q-1} + \int_{-\infty}^{\infty}\Big[-\frac{\sigma^{q-1}}{q-1}y^q + \lambda_1 y + \lambda_2(x^2-\sigma^2)y^q\Big]dx. \end{aligned}$$

For simplicity, let us denote

$$L(x,y) = -\frac{\sigma^{q-1}}{q-1}y^q + \lambda_1 y + \lambda_2(x^2-\sigma^2)y^q.$$

Note that $S(y+\delta y) - S(y) = \int_{-\infty}^{\infty} \partial L/\partial y \delta y dx$. To maximize S, $\partial L/\partial y = 0$. The Tsallis distribution is then found to be

$$p(x) = \frac{1}{Z_q}[1 + \beta(q-1)x^2]^{1/(1-q)} \tag{11.21}$$

for $1 < q < 3$. When $q < 1$, the same formula holds when $|x| \leq [\beta(1-q)]^{-1/2}$ and $p(x) = 0$ for any other x. It is clear that Z_q is a normalization constant and β is related to the second moment.

Note that when $q = 1$, the derivation gives the normal distribution. When $q = 2$, the distribution reduces to the Cauchy distribution and, hence, is a stable distribution. When $5/3 < q < 3$, the distribution is heavy-tailed, with

$$0 \leq \alpha = 2/(q-1) - 1 = (3-q)/(q-1) \leq 2.$$

We may generalize the Tsallis distribution by replacing constraint expressed by Eq. (11.20) by

$$\sigma^{\alpha} = \frac{\int_{-\infty}^{\infty} x^{\alpha}[p(x)]^q dx}{\int_{-\infty}^{\infty}[p(x)]^q dx}. \tag{11.22}$$

Following the same procedure leading to Eq. (11.21), we have

$$p(x) = \frac{1}{Z_q}[1 + \beta(q-1)x^{\alpha}]^{1/(1-q)}, \tag{11.23}$$

where again Z_q is a normalization constant and β is related to the second moment. Equation (11.23) contains three parameters and is more flexible. To illustrate the power of Eq. (11.23), we apply it to describe sea clutter radar returns in Sec. 11.3.2.

11.1.5 Power law through optimization

Engineering systems are usually designed through constraint optimization. Doyle and Carlson [111] have proposed an interesting means of obtaining power-law behavior by minimizing a cost function of the form

$$J = \left\{ \sum p_i l_i \mid l_i = f(r_i), \quad \sum r_i \leq R \right\}, \tag{11.24}$$

where p_i is the probability of events with index i, $1 \leq i \leq N$, r_i is the resource allocation per unit loss, normalized to lie in the unit interval $0 \leq r_i \leq 1$, the relationship $l_i = f(r_i)$ describes how allocation of resources r_i limits the size l_i of events, and $\sum r_i \leq R$ is an overall constraint coarsely accounting for the connectivity and spatial structure of a real design problem. By assuming resource vs. loss function $l_i = f(r_i)$ having the following form,

$$f_{\beta}(r_i) = \begin{cases} -\log(r_i), & \beta = 0 \\ \frac{c}{\beta}(r_i^{-\beta} - 1), & \beta > 0, \end{cases} \tag{11.25}$$

they find the optimal solution to be

$$r_i = Rp_i^{\frac{1}{1+\beta}} \Big(\sum_j p_j^{\frac{1}{1+\beta}} \Big)^{-1} \tag{11.26}$$

$$l_i = \begin{cases} -\log(Rp_i) + \log(\sum_j p_j), & \beta = 0 \\ \frac{c}{\beta}\Big[\Big(Rp_i^{\frac{1}{1+\beta}}\Big)^{-\beta}\Big(\sum_j p_j^{\frac{1}{1+\beta}}\Big)^{\beta} - 1\Big], & \beta > 0 \end{cases} \tag{11.27}$$

$$l_i = \begin{cases} -\sum_i p_i \log(Rp_i) + (\sum_i p_i)\log(\sum_i p_i), & \beta = 0 \\ \beta^{-1}\Big[R^{-\beta}\Big(\sum_i p_i^{\frac{1}{1+\beta}}\Big)^{1+\beta} - \sum_i p_i\Big], & \beta > 0. \end{cases} \tag{11.28}$$

Assuming $r_i < 1$ and inverting Eq. (11.27), they obtain a power-law relation between p_i and l_i:

$$p_i(l_i) = c_1(l_i + c_2)^{-(1+1/\beta)}. \tag{11.29}$$

11.2 MODELS FOR $1/F^{\beta}$ PROCESSES

In the past few decades, considerable efforts have been made to find universal mechanisms for $1/f^{\beta}$ noise. While many excellent models have been proposed, it does not seem that a single universal mechanism exists for such noise. To provide a glimpse of this exciting field, we describe a few simple models here. Their use or abuse will also be briefly touched upon.

11.2.1 $1/f^{\beta}$ processes from superposition of relaxation processes

Superposition of relaxation processes was originally proposed to explain the $1/f^{\beta}$ noise in vacuum tubes. It can be considered one of the earliest mechanisms proposed for $1/f^{\beta}$ noise. The basic idea is that a random process

$$x(t) = \sum_i N_0 e^{-\lambda(t-t_i)},$$

where N_0 is a constant, t_i, $i = 1, 2, \cdots$ are random time instances when a relaxation process is initiated, and λ is another random variable having a PDF $p(\lambda)$, may possess a power-law decaying PSD when $p(\lambda)$ is chosen appropriately. In the language of vacuum tube noise, each relaxation process $N(t) = N_0 e^{-\lambda(t-t_i)}$ for $t \geq t_i$ and $N(t) = 0$ for $t < t_i$, which originates from cathode surface trapping sites, contributes to the vacuum tube current. The Fourier transform of a single exponential relaxation process is

$$F(\omega) = \int_{-\infty}^{\infty} N(t) e^{-j\omega t} dt = N_0 \int_0^{\infty} e^{-(\lambda + j\omega)t} dt = \frac{N_0}{\lambda + j\omega}. \tag{11.30}$$

For a train of such pulses $N(t,t_k) = N_0 e^{-\lambda(t-t_k)}$ for $t \geq t_k$ and $N(t,t_k) = 0$ for $t < t_k$, we have

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{\infty} \sum_k N(t,t_k) e^{-j\omega t} dt \\ &= N_0 \sum_k e^{j\omega t_k} \int_0^{\infty} e^{-(\lambda+j\omega)t} dt = \frac{N_0}{\lambda + j\omega} \sum_k e^{j\omega t_k} \end{aligned} \tag{11.31}$$

and the spectrum is

$$\begin{aligned} S(\omega) &= \lim_{T\to\infty} \frac{1}{T} \langle |F(\omega)|^2 \rangle = \frac{N_0^2}{\lambda^2 + \omega^2} \lim_{T\to\infty} \frac{1}{T} \left\langle \left| \sum_k e^{j\omega t_k} \right|^2 \right\rangle \\ &= \frac{N_0^2 n}{\lambda^2 + \omega^2}, \end{aligned} \tag{11.32}$$

where n is the average pulse rate and the triangle brackets denote an ensemble average. This spectrum is nearly flat near the origin, and after a transition region it becomes proportional to $1/\omega^2$ at high frequency. $1/\omega^\beta$-type behavior can be obtained by choosing appropriate forms of the PDF $p(\lambda)$. For example, if the relaxation rate is uniformly distributed between two values λ_1 and λ_2 and the amplitude of each pulse remains constant, we find the spectrum

$$\begin{aligned} S(\omega) &= \frac{1}{\lambda_2 - \lambda_1} \int_{\lambda_1}^{\lambda_2} \frac{N_0^2 n}{\lambda^2 + \omega^2} d\lambda = \frac{N_0^2 n}{\omega(\lambda_2 - \lambda_1)} \left[\arctan \frac{\lambda_2}{\omega} - \arctan \frac{\lambda_1}{\omega} \right] \\ &= \begin{cases} N_0^2 n, & 0 < \omega \ll \lambda_1 \ll \lambda_2 \\ \frac{N_0^2 n \pi}{2\omega(\lambda_2 - \lambda_1)}, & \lambda_1 \ll \omega \ll \lambda_2 \\ \frac{N_0^2 n}{\omega^2}, & \lambda_1 \ll \lambda_2 \ll \omega. \end{cases} \end{aligned} \tag{11.33}$$

As another example, assume that $dP(\lambda) = \frac{A}{\lambda^\gamma} d\lambda$ occurs in the range $\lambda_1 < \lambda < \lambda_2$. In this case, we obtain

$$\begin{aligned} S(\omega) \propto \int_{\lambda_1}^{\lambda_2} \frac{1}{\lambda^2 + \omega^2} \frac{d\lambda}{\lambda^\gamma} &= \frac{1}{\omega^{1+\gamma}} \int_{\lambda_1}^{\lambda_2} \frac{1}{1 + \lambda^2/\omega^2} \frac{d(\lambda/\omega)}{(\lambda/\omega)^\gamma} \\ &= \frac{1}{\omega^{1+\gamma}} \int_{\lambda_1/\omega}^{\lambda_2/\omega} \frac{1}{1+x^2} \frac{dx}{x^\gamma}. \end{aligned}$$

The above integral can be approximated as

$$S(\omega) \approx \frac{1}{\omega^{1+\gamma}} \int_0^{\infty} \frac{1}{1+x^2} \frac{dx}{x^\gamma} \propto \frac{1}{\omega^{1+\gamma}}. \tag{11.34}$$

Note that the superposition of relaxation processes may be transformed into an ON/OFF train. The question of whether the noise may have a spectrum of $1/\omega^\beta$

instead of $1/\omega^2$ is then equivalent to the question of whether ON/OFF trains may have power-law tails. Our discussions in Sec. 11.1.3 provide the general framework for finding a suitable PDF $p(\lambda)$ such that the spectrum is $1/\omega^{\beta}$.

11.2.2 $1/f^{\beta}$ processes modeled by ON/OFF trains

In Sec. 8.7, we saw that power-law distributed ON/OFF trains possess $1/f^{2H+1}$ spectra. When $1 \leq \alpha \leq 2$, it can be rigorously proven that the parameter H and the parameter α are related by Eq. (8.27). In numerical simulations, we have found that when ON and OFF periods have the same heavy-tailed distribution, Eq. (8.27) is still valid even if $0 \leq \alpha \leq 1$. In this subsection, we discuss other means of obtaining ON/OFF trains without prescribing their distributions.

One simple way is to start from ON/OFF trains with exponential distributions with rate λ. As we have seen in Sec. 11.1.3, when λ follows a suitable distribution, the ON and OFF periods can become heavy-tailed. While a power-law distributed λ or $1/\lambda$ is an interesting case, to obtain heavy-tailed ON/OFF periods, the distribution of λ may be light-tailed. This is clear from the in-depth discussion in Sec. 11.1.3; therefore, we shall say no more about this here.

Another way of modeling the ON/OFF trains is through chaotic maps. Our in-depth discussion of chaos will be presented in Chapter 13. Here, we shall use the basic understanding of chaos as described in Chapter 2 to give a glimpse of the idea. For this reason, it would be helpful to review Sec. 2.2 before reading the following presentation.

One simple way of obtaining an ON/OFF train from a chaotic time series $x_1, x_2, \cdots$ is through thresholding,

$$y_n = \begin{cases} 0 & 0 \leq x_n < d_y, \\ 1, & d_y \leq x_n \leq 1, \end{cases} \tag{11.35}$$

where the chaotic time series $x_1, x_2, \cdots$ are generated by the following iterations:

$$x_{n+1} = f(x_n) = \begin{cases} f_1(x_n), & 0 \leq x_n < d, \\ f_2(x_n), & d \leq x_n \leq 1, \end{cases} \tag{11.36}$$

where $f_1(x_n)$ and $f_2(x_n)$ are two chaotic maps. The challenge is to find suitable $f_1(\cdot)$ and $f_2(\cdot)$ such that y_n have desired properties. One map that has proven to be useful in modeling long-range-dependent network traffic is the following:

$$x_{n+1} = \begin{cases} f_1(x_n) = x_n + \frac{1-d}{d^{m_1}}(x_n)^{m_1}, & 0 \leq x_n < d, \\ f_2(x_n) = x_n - \frac{d}{(1-d)^{m_2}}(1-x_n)^{m_2}, & d \leq x_n \leq 1, \end{cases} \tag{11.37}$$

where $0 < x_n, d < 1$ and $1 < m_1, m_2 < 2$. This system has a number of interesting properties. For example, the ON/OFF trains have heavy-tailed distributions, and the Hurst parameter is given by

$$H = \frac{3m_i - 4}{2(m_i - 1)},$$

where $m_i = \max(m_1, m_2)$.

It should be noted that in the physical literature, there are types of maps called ON/OFF intermittency maps. Thresholding a time series generated by such maps is straightforward. One generic form of such maps is the following:

$$x_{n+1} = z_n f(x_n), \tag{11.38}$$

with $f(0) = 0$, $f'(0) = 1$, and $z_n = a\eta_n \geq 0$ a random or chaotic process. Given $\{\eta_n\}$, there is a critical value of a denoted as a_c such that for $a < a_c$, the origin is asymptotically stable. If a is increased slightly above a_c, the variable x then exhibits ON/OFF intermittency. An example of such a map is given by

$$x_{n+1} = z_n x_n e^{-bx_n}, \tag{11.39}$$

where $x \geq 0$ and $b > 0$. When $z_n = a > 1$ is a constant, the fixed point is found to be $x^* = \ln a/b$. To understand this process, one should realize that when $x > x^*$, the system is in an ON state; otherwise, it is in an OFF state. When z_n is a random process (such as a fGn process) or chaotic process, the x_n signal is intermittent.

11.2.3 $1/f^{\beta}$ processes modeled by self-organized criticality

It is common to find that complex systems tend to organize or disorganize, depending on whether they start from highly disorganized or organized states. The ubiquity of $1/f^{\beta}$ processes in nature motivated Bak and co-workers [23, 24] to suggest that the preferred state in a complex system is neither regular nor fully chaotic/random, but something similar to the "edge of chaos" in the sense that there is no characteristic scale in space or time. This idea is so appealing to scientists in vastly different fields that now the term self-organized criticality (SOC) has become one of the most popular in the literature.

Technically, the theory of SOC refers to the behavior of the sand pile model. The model consists of an $N \times N$ square grid of cells, each characterized by a positive integer $Z(i, j)$, which can be considered the local slope (or height difference between adjacent cells) of the pile. Initial conditions are chosen randomly in the range $1 \leq Z \leq K$, where $K \geq 3$ is a threshold value. At each time step, a randomly chosen cell (i_0, j_0) is incremented by 1:

$$Z_{n+1}(i_0, j_0) = Z_n(i_0, j_0) + 1. \tag{11.40}$$

This amounts to adding a grain of sand to the sand pile. To simulate an avalanche phenomenon, a local instability condition is introduced, which indicates that any cells with $Z(i, j) > K$ and their four nearest neighbors are readjusted according to

$$Z_{n+1}(i, j) = Z_n(i, j) - 4 \tag{11.41}$$

$$Z_{n+1}(i \pm 1, j) = Z_n(i \pm 1, j) + 1 \tag{11.42}$$

$$Z_{n+1}(i, j \pm 1) = Z_n(i, j \pm 1) + 1. \tag{11.43}$$

The cells outside the boundary $Z(0, j)$, $Z(i, 0)$, $Z(N + 1, j)$, $Z(i, N + 1)$ are kept at zero. This condition makes the model an open system, allowing sand grains to leave the system. Equations (11.41)–(11.43) are iterated until no cells with $Z > K$ remain.

There are many types of power-law behavior in the sand pile model. For example, the avalanche size, defined as the number of cells that topple at each time step, follows a power-law distribution; the PSD of the total size of the pile

$$Z_{total} = \sum_{i=1}^{N} \sum_{j=1}^{N} Z(i, j)$$

also follows a power law.

Being discrete and nonlinear, the sand pile model is hard to tackle analytically. One therefore turns to numerical simulations. The drawback of such an approach is that when calculating the PSD, some researchers mistake the magnitude of the Fourier transform for the PSD. By doing so, they incorrectly interpret $1/f^2$ as $1/f$. It turns out that the sand pile model of Bak et al. really produces a $1/f^2$ spectrum instead of $1/f$. Therefore, the SOC model is not quite able to explain the $1/f^\beta$ process. Unfortunately, quite often, when the $1/f^\beta$ spectrum is found from certain data, some researchers attribute it to SOC.

11.3 APPLICATIONS

The materials presented in this chapter have found numerous applications in many areas of science and engineering. To illustrate their usefulness, in this section we consider two important problems, the mechanism for long-range-dependent network traffic and modeling of sea clutter.

11.3.1 Mechanism for long-range-dependent network traffic

Long-range dependence is a prevalent feature of network traffic. Willinger et al. [476] have modeled traffic between a single source and destination by an ON/OFF process, where ON and OFF periods are independent (see Fig. 11.2). The ON periods are identically distributed; so are the OFF periods. Furthermore, at least one of them follows heavy-tailed distributions, and the heavy tails come solely from user behavior. Willinger et al. have further proved that superposition of such processes can be approximated by fBm. This model has now been accepted as the most plausible mechanism for LRD network traffic. However, a number of fundamental questions remain unanswered. For example, where do ON/OFF trains come from? Can heavy-tailed distributions come from mechanisms other than user behavior? Amazingly, by addressing such questions, one arrives at a universal, protocol-independent mechanism for LRD traffic.

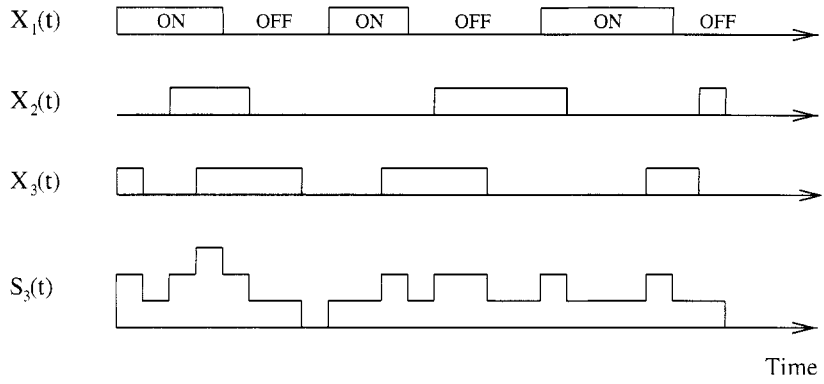

Figure 11.2. $N = 3$ ON/OFF sources, $X_1(t)$, $X_2(t)$, $X_3(t)$, and their summation $S_3(t)$.

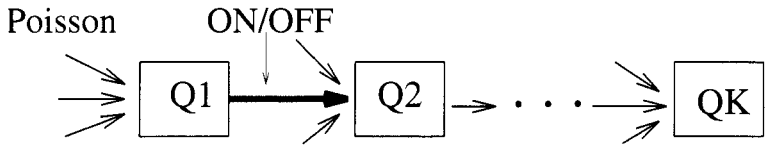

Figure 11.3. Schematic of a tandem of queuing system.

To address the above issues, we first note that network traffic processes represent aggregate traffic streams collected at some backbone network nodes. Before they are measured, most of them must have been transmitted across several links. This suggests that the most relevant scenario to study is the simple tandem network presented in Fig. 11.3. We now ask a question: If the input traffic to Q1 is the simplest Poisson traffic (i.e., the M/M/1 queuing system), can we observe heavy-tailed ON/OFF sources to Q2, Q3, $\cdots$, and so on? The answer is yes, as we shall explain below.

Our first crucial observation is that a busy period of Q1 is an ON period on the link to Q2, while an idle period of Q1 is an OFF period on the link to Q2, as shown in Fig. 11.3. This simple observation shows that ON/OFF processes naturally arise in a network. Realizing the equivalence of an ON episode and a busy period, and noticing (as discussed in Sec. 11.1.1) that a busy period has a heavy-tailed distribution when utilization approaches 1, we immediately see that the power-law behavior does not depend solely on user behavior. It can be generated through transformation of Poisson traffic in a tandem network.

Let us consider transmission of packets in a network. Assume that a long ON period is generated by Q1. Typically, it may not be passed through Q2 as a whole when the utilization level at Q2 is not too low. It has to be packetized and transmitted piece by piece so that the bandwidth can also be used by other users to ensure fairness. While the first piece would be fairly large, other pieces may be considerably smaller. The time intervals between the successive transmissions of these pieces may be quite random. These pieces may even be delivered out of order. In short,

the output traffic component originating from this single long ON episode may constitute a very bursty traffic component to Q3. For this reason, the exponent α characterizing the heavy-tailed distributions of the ON periods may become quite different from that of a M/M/1 queuing system, which is 1/2. In fact, the OFF periods can also become heavy-tailed.

11.3.2 Distributional analysis of sea clutter

We have discussed sea clutter in some depth in a number of previous chapters. In this subsection, we consider distributional analysis of sea clutter. The details of sea clutter data are presented in Sec. A.2 of Appendix A.

As we noted in Chapter 1, sea clutter data are often highly non-Gaussian. Much effort has been made to fit various distributions to the observed amplitude data of sea clutter, including Weibull, log-normal, K, and compound-Gaussian distributions. However, the fitting of those distributions to real sea clutter data is not excellent, and quite often using parameters estimated from those distributions is not very effective for distinguishing sea clutter data with and without targets. Only recently have we realized that the ineffectiveness of conventional distributional analysis may be due to the fact that sea clutter data are highly nonstationary. This nonstationarity motivates us to perform distributional analysis on the data obtained by differencing amplitude data of sea clutter. Denote the sea clutter amplitude data by $y(n), n = 1, 2, \cdots$. The differenced data of sea clutter are denoted as

$$x(n) = y(n+1) - y(n), \quad n = 1, 2, \cdots .$$

Specifically, we fit the differenced data by the Tsallis distribution and compare the fitting with the best distribution for sea clutter, the K distribution,

$$f(x) = \frac{\sqrt{2\nu}}{\sqrt{\mu}\Gamma(\nu)2^{\nu-1}} \left(\sqrt{\frac{2\nu}{\mu}}x\right)^{\nu} K_{\nu-1}\left(\sqrt{\frac{2\nu}{\mu}}x\right), \quad x \geq 0 \tag{11.44}$$

where ν and μ are parameters (μ is in fact equal to half of the second moment), $\Gamma(\nu)$ is the usual gamma function, and $K_{\nu-1}$ is the modified Bessel function of the third kind of order $\nu - 1$. Note, however, that conventionally, the K distribution is used to fit the original sea clutter data, not the differenced data.

Figures 11.4(a,b) show typical results of using the Tsallis distribution to fit the differenced sea clutter data without and with a target, respectively. We observe that the Tsallis distribution fits the differenced data very well. For comparison, Figs. 11.4(c,d) show the results of using the K distribution to fit the same amplitude data that were used to generate Figs. 11.4(a,b). We observe that the fitting to these data using the K distribution is worse than that using the Tsallis distribution.

Note that the goodness of fit of a distribution can be quantified by the Kolmogorov-Smirnov (KS) test. We have found that the Tsallis distribution indeed fits sea clutter data much better than the K distribution. We also note that for detecting low observ-

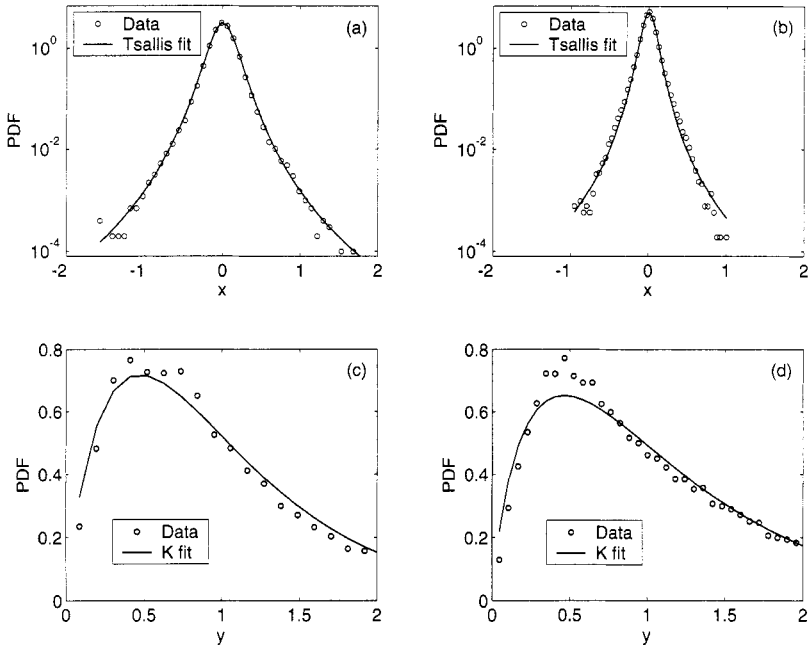

Figure 11.4. Representative results of using (a,b) the Tsallis distribution to fit the differenced data and (c,d) the K distribution to fit the amplitude data of sea clutter. (a,c) are for the sea clutter data without a target, while (b,d) are for the data with a target. Circles and solid lines denote the raw and fitted PDFs, respectively.

able targets within sea clutter, the parameters estimated from the Tsallis distribution are more effective than the ones from K or other distributions.

To understand why the Tsallis distribution provides a better fit to the sea clutter data, we note two sources of complexity for sea clutter: the roughness of the sea surface, largely due to wave-turbulence interactions on the sea surface and ocean sprays, and the multipath propagation. Either source of complexity suggests that sea clutter may be considered as a superposition of signals massively reflected from the ocean surface. Therefore, the central limit theorem or the generalized central limit theorem has to play a crucial role. Consequently, the Tsallis distribution can be expected to fit sea clutter data well.

11.4 BIBLIOGRAPHIC NOTES

This chapter contains a lot of background material that cannot be fully explained here. We refer readers to the classic book by Kleinrock [263] on the background of queuing theory, to an exquisite paper by Montroll and Shlesinger on the relation between log-normal distribution and power-law distribution [315,316], to a number of recent papers on the Tsallis distribution [38,449] and the distribution obtained

by Doyle and Carlson [62, 111], to a few classic papers [46, 60, 243, 391, 456] as well as an entertaining survey article [310] (downloadable at
http://www.nslij-genetics.org/wli/1fnoise/1fnoise_review.html)
on the superposition of relaxation processes, to [216,366,426,476] on the correlation structure of ON/OFF models, to [109, 120, 210, 215, 312–314, 350] on modeling ON/OFF trains by chaotic maps, to [23, 24, 98, 129, 217, 237, 240] on SOC, to [279, 457, 476] on the mechanism of LRD network traffic, and finally, to [69, 128, 189, 190, 238, 318, 320, 447, 466] on distributional analysis of sea clutter.

11.5 EXERCISES

1. Numerically verify
   (a) Eq. (11.7),
   (b) Eq. (11.9), and
   (c) Eq. (11.14).
   (Hint: when solving (a), be aware that in order to observe the power-law tail, λ_{min} and λ_{max} cannot be too close together. Similar care has to be taken for (b) and (c)).

2. Fill in the details in deriving Eq. (11.29).

3. Simulate $1/f^{\beta}$ processes by superposing relaxation processes.

4. Simulate exponentially distributed ON/OFF trains with the rate λ taking various distributions discussed in Sec. 11.1.3. Estimate the Hurst parameter of those simulated processes.

5. Generate time series by Eq. (11.39) and estimate the Hurst parameter.