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 6

SELF-SIMILAR STOCHASTIC PROCESSES

In Chapter 5, we discussed the standard Cantor set and briefly touched on its randomized variants. In real-world signal processing applications, random fractals are far more useful. In this chapter, we discuss one of the prototypical random fractal models, the fractional Brownian motion (fBm) model. Much will be said about its simulation. In Chapter 7, we shall discuss a different type of random fractal, the Levy motions.

6.1 GENERAL DEFINITION

A continuous-time stochastic process, $X = \{X(t), t \geq 0\}$, is said to be self-similar if

$$X(\lambda t) \stackrel{d}{=} \lambda^H X(t), \quad t \geq 0 \tag{6.1}$$

for $\lambda > 0$, $0 < H < 1$, where $\stackrel{d}{=}$ denotes equality in distribution. H is called the self-similarity parameter or the Hurst parameter.

Before proceeding, we note that, more rigorously speaking, processes defined by Eq. (6.1) should be called self-affine processes instead of self-similar processes, since X and t have to be scaled differently to make the function look similar. A simple physical explanation for this is that the units for X and t are different. This

distinction is particularly important if one wishes to discuss the fractal dimensions of such processes, as we will see in Sec. 6.4. In general, however, we will continue to call such processes self-similar in order to follow the convention used in certain engineering disciplines, such as network traffic engineering.

The following three properties can be easily derived from Eq. (6.1):

$$E[X(t)] = \frac{E[X(\lambda t)]}{\lambda^H} \quad \text{Mean,} \tag{6.2}$$

$$Var[X(t)] = \frac{Var[X(\lambda t)]}{\lambda^{2H}} \quad \text{Variance,} \tag{6.3}$$

$$R_x(t,s) = \frac{R_x(\lambda t, \lambda s)}{\lambda^{2H}} \quad \text{Autocorrelation.} \tag{6.4}$$

Note that $Var[X(t)] = R_x(t,t)$; hence, Eq. (6.3) can be simply derived from Eq. (6.4); we have listed it as a separate equation for convenience of future reference. If we consider only second-order statistics, or if a process is Gaussian, Eqs. (6.2)–(6.4) can be used instead of Eq. (6.1) to define a self-similar process.

A very useful way of describing a self-similar process is by its spectral representation. Strictly speaking, the Fourier transform of $X(t)$ is undefined due to the nonstationary nature of $X(t)$. One can, however, consider $X(t)$ in a finite interval, say, $0 < t < T$:

$$X(t,T) = \begin{cases} X(t) & 0 < t < T \\ 0 & \text{otherwise} \end{cases} \tag{6.5}$$

and take the Fourier transform of $X(t,T)$:

$$F(f,T) = \int_0^T X(t) e^{-2\pi j f t} dt \tag{6.6}$$

$|F(f,T)|^2 df$ is the contribution to the total energy of $X(t,T)$ from those components with frequencies between f and $f + df$. The (average) power spectral density (PSD) of $X(t,T)$ is then

$$S(f,T) = \frac{1}{T}\left|F(f,T)\right|^2,$$

and the spectral density of X is obtained in the limit as $T \to \infty$

$$S(f) = \lim_{T\to\infty} S(f,T).$$

Noting that the PSD for $\lambda^{-H} X(\lambda t)$ is

$$\lambda^{-2H} \lim_{T\to\infty} \frac{1}{\lambda T}\left| \int_0^{\lambda T} X(\lambda t) e^{-2\pi j f t} dt \right|^2 = \lambda^{-2H-1} S(f/\lambda),$$

and that the PSD for $\lambda^{-H} X(\lambda t)$ and $X(t)$ are the same, we have

$$S(f) = S(f/\lambda)\lambda^{-2H-1}.$$

The solution to the above equation is

$$S(f) \sim f^{-\beta} \tag{6.7}$$

with

$$\beta = 2H + 1. \tag{6.8}$$

Processes with PSDs as described by Eq. (6.7) are called $\frac{1}{f}$ processes. Typically, the power-law relationships that define these processes extend over several decades of frequency. Such processes have been found in numerous areas of science and engineering. See the partial list of relevant literature in the Bibliographic Notes at the end of the chapter. The ubiquity of $\frac{1}{f}$ processes is one of the most stimulating puzzles in physics, astrophysics, geophysics, engineering, biology, and the social sciences, and many efforts have been made to discover the mechanisms for generating such processes. However, none of the mechanisms proposed so far has been considered universal. We shall return to this issue in Chapter 11.

6.2 BROWNIAN MOTION (BM)

As an example, let us consider the $\frac{1}{f}$ process known as Brownian motion (Bm). It is defined as a (nonstationary) stochastic process $B(t)$ that satisfies the following criteria:

1. All Brownian paths start at the origin: $B(0) = 0$.
2. For $0 < t_1 < t_2 < t_3 < t_4$, the random variables $B(t_2) - B(t_1)$ and $B(t_4) - B(t_3)$ are independent.
3. For all $(s, t) \geq 0$, the variable $B(t + s) - B(t)$ is a Gaussian variable with mean 0 and variance s.
4. $B(t)$ is a continuous function of t.

We may infer from the above definition that the probability distribution function of B is given by

$$P\{[B(t+s) - B(t)] \leq x\} = \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^{x} e^{-\frac{u^2}{2s}} du. \tag{6.9}$$

This function also satisfies the scaling property:

$$P\{[B(t+s) - B(t)] \leq x\} = P\{B(\lambda t + \lambda s) - B(\lambda t) \leq \lambda^{1/2} x\}. \tag{6.10}$$

In other words, $B(t)$ and $\lambda^{-1/2} B(\lambda t)$ have the same distribution. Thus, we see from Eq. (6.1) that Bm is a self-similar process with Hurst parameter 1/2.

Suppose we have measured $B(t_1)$, $B(t_2)$, $t_2 - t_1 > 0$. What can we say about $B(s)$, $t_1 < s < t_2$? The answer is given by the Levy interpolation formula:

$$B(s) = B(t_1) + \frac{(s - t_1)}{(t_2 - t_1)}[B(t_2) - B(t_1)] + \left[\frac{(t_2 - s)(s - t_1)}{t_2 - t_1}\right]^{1/2} W, \quad (6.11)$$

where W is a zero-mean and unit-variance Gaussian random variable. The first two terms are simply a linear interpolation. The third term gives the correct variances for $B(s) - B(t_1)$ and $B(t_2) - B(s)$, which are $s - t_1$ and $t_2 - s$, respectively.

A popular way of generating Bm is by using the following random midpoint displacement method, which is essentially an application of the Levy interpolation formula. Suppose we are given $B(0) = 0$ and $B(1)$ as a sample of a Gaussian random variable with zero mean and variance σ^2. We can obtain $B(1/2)$ from the following formula:

$$B(1/2) - B(0) = \frac{1}{2}[B(1) - B(0)] + D_1,$$

where D_1 is a random variable with mean 0 and variance $2^{-2}\sigma^2$. D_1 is simply the third term of Eq. (6.11), and the coefficient 2^{-2} equals $(t_2 - s)(s - t_1)/(t_2 - t_1)$, with $t_1 = 0$, $t_2 = 1$, $s = 1/2$. Similarly, we have

$$B(1/4) - B(0) = \frac{1}{2}[B(1/2) - B(0)] + D_2,$$

where D_2 is a random variable with zero mean and variance $2^{-3}\sigma^2$. We can apply the same idea to obtain $B(3/4)$. In general, the variance for D_n is $2^{-(n+1)}\sigma^2$.

We note that a Brownian path is not a differentiable function of time. Heuristically, this can be readily understood: consider the variable $B(t + s) - B(t)$ with variance s. Its standard deviation, which is a measure of its order of magnitude, is $\sim \sqrt{s}$. Thus, the derivative of B at t behaves like the limit $\sqrt{s}/s = s^{-1/2}$ as $s \to 0$.

Although $B(t)$ is almost surely not differentiable in t, symbolically one still often writes

$$B(t) = \int_0^t w(\tau)d\tau, \quad (6.12)$$

where $w(t)$ is stationary white Gaussian noise, and extends the above equation to $t < 0$ through the convention

$$\int_0^{-t} \equiv -\int_t^0 .$$

It should be understood that integrals with respect to the differential element $w(t)dt$ should be interpreted more precisely as integrals with respect to the differential element $dB(t)$ in the Riemann-Stieltjes integral sense. This has profound consequences when numerically integrating Eq. (6.12).

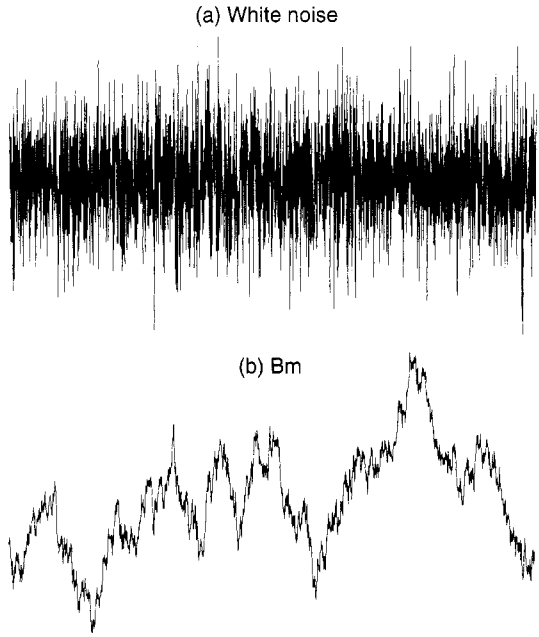

Figure 6.1. White noise (a) and Bm (b). Axes are arbitrary.

To see how this works, let us partition $[0, t]$ into n equally spaced intervals, $\Delta t = t/n$. Since $w(t)$ are Gaussian random variables with zero mean and unit variance, in order for $B(t)$ to have variance t, we should have

$$B(t) \approx \sum_{i=1}^{n} w(i\Delta t) \cdot (\Delta t)^{1/2}. \tag{6.13}$$

That is, the coefficient is $(\Delta t)^{1/2}$ instead of Δt, as one might have guessed. Typically, $\Delta t \ll 1$; hence, $(\Delta t)^{1/2} \gg \Delta t$. Thus, if one incorrectly uses Δt instead of $(\Delta t)^{1/2}$, one is severely underestimating the noise term.

When the time is genuinely discrete, or when the units of time can be arbitrary, one can take Δt to be 1 unit, and Eq. (6.13) becomes

$$B(t) = \sum_{i=1}^{n} w(i). \tag{6.14}$$

Equation (6.14) provides perhaps the simplest method of generating a sample of Bm. An example is shown in Fig. 6.1.

Equation (6.14) is also known as a random walk process. A more sophisticated random walk (or jump) process involves summing up an infinite number of jump

functions: $J_i(t) = A_i\beta(t - t_i)$, where $\beta(t)$ is a unit-step function

$$\beta(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{6.15}$$

and A_i, t_i are random variables with Gaussian and Poisson distributions, respectively.

6.3 FRACTIONAL BROWNIAN MOTION (FBM)

One-dimensional case: A normalized fractional Brownian motion (fBm) process, $Z(t)$, is defined as follows:

$$Z(t) \stackrel{d}{=} Wt^H \quad (t > 0), \tag{6.16}$$

where W is a Gaussian random variable of zero mean and unit variance and H is the Hurst parameter. When $H = 1/2$, the process reduces to the standard Bm. It is trivial to verify that this process satisfies the defining Eq. (6.1) for a self-similar stochastic process.

FBm $B_H(t)$ is a Gaussian process with mean 0, stationary increments, variance

$$E[(B_H(t))^2] = t^{2H}, \tag{6.17}$$

and covariance

$$E[B_H(s)B_H(t)] = \frac{1}{2}\left\{s^{2H} + t^{2H} - |s - t|^{2H}\right\}, \tag{6.18}$$

where H is the Hurst parameter. Due to its Gaussian nature, according to Eqs. (6.2)–(6.4), the above three properties completely determine its self-similar character. Figure 6.2 shows several fBm processes with different H.

Roughly, the distribution for an fBm process is

$$B_H(t) = \frac{1}{\Gamma(H + 1/2)} \int_{-\infty}^{t} (t - \tau)^{H-1/2} dB(\tau). \tag{6.19}$$

It is easy to verify that fBm satisfies the following scaling property:

$$P(B_H(t + s) - B_H(t) \leq x) = P\{B_H(\lambda t + \lambda s) - B_H(\lambda t) \leq \lambda^H x\}. \tag{6.20}$$

In other words, this process is invariant for transformations conserving the similarity variable x/t^H.

The integral defined by Eq. (6.19) diverges. The more precise definition is

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H + 1/2)} \int_{-\infty}^{t} K(t - \tau) dB(\tau), \tag{6.21}$$

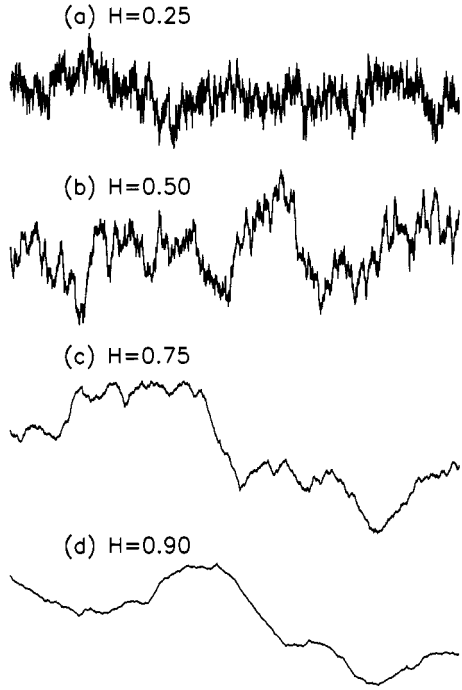

Figure 6.2. Several fBm processes with different H.

where the kernel $K(t-\tau)$ is given by

$$K(t-\tau) = \begin{cases} (t-\tau)^{H-1/2}, & 0 \le \tau \le t \\ (t-\tau)^{H-1/2} - (-\tau)^{H-1/2}, & \tau < 0. \end{cases} \tag{6.22}$$

The increment process of fBm, $X_i = B_H((i+1)\Delta t) - B_H(i\Delta t)$, $i \ge 1$, where Δt can be considered a sampling time, is called the fractional Gaussian noise (fGn) process. It is a zero mean stationary Gaussian time series. Noting that

$$E(X_i X_{i+k}) = E\{[B_H((i+1)\Delta t) - B_H(i\Delta t)][B_H((i+1+k)\Delta t) - B_H((i+k)\Delta t)]\},$$

by Eq. (6.18), one obtains the autocovariance function $\gamma(k)$ for the fGn process:

$$\gamma(k) = E(X_i X_{i+k})/E(X_i^2) = \frac{1}{2}\left\{(k+1)^{2H} - 2k^{2H} + |k-1|^{2H}\right\}, \; k \ge 0. \tag{6.23}$$

Notice that the expression is independent of Δt. Therefore, without loss of generality, one could take $\Delta t = 1$. In particular, we have

$$\gamma(1) = \frac{1}{2}\left(2^{2H} - 2\right).$$

Let us first note a few interesting properties of $\gamma(k)$:

1. When $H = 1/2$, $\gamma(k) = 0$ for $k \neq 0$. This is the well-known property of white Gaussian noise.

2. When $0 < H < 1/2$, $\gamma(1) < 0$.

3. When $1/2 < H < 1$, $\gamma(1) > 0$.

Properties 2 and 3 are often termed antipersistent and persistent correlations, respectively.

Next, we consider the behavior of $\gamma(k)$ when k is large. Noting that when $|x| \ll 1$,

$$(1+x)^\alpha \approx 1 + \alpha x + \frac{\alpha(\alpha-1)}{2}x^2,$$

we have, when $k \gg 1$,

$$\begin{aligned}(k+1)^{2H} + |k-1|^{2H} &= k^{2H}[(1+1/k)^{2H} + (1-1/k)^{2H}] \\ &= k^{2H}[2 + 2H(2H-1)k^{-2}].\end{aligned}$$

Hence,

$$\gamma(k) \sim H(2H-1)k^{2H-2} \quad \text{as } k \to \infty \tag{6.24}$$

when $H \neq 1/2$. When $H = 1/2$, $\gamma(k) = 0$ for $k \geq 1$, the X_i's are simply white noise.

High-dimensional case: Similar to the one-dimensional case, a fBm in a n-dimensional Euclidean space, $B_H(x_1, x_2, \cdots, x_n)$, is a scalar function of $(x_1, x_2, \cdots, x_n)$. It can be defined by the following property:

$$\begin{aligned}&\left\langle [B_H(x_1', x_2', \cdots, x_n') - B_H(x_1, x_2, \cdots, x_n)]^2 \right\rangle \\ &\propto \left[(x_1' - x_1)^2 + (x_2' - x_2)^2 + \cdots + (x_n' - x_n)^2\right]^H.\end{aligned} \tag{6.25}$$

The process B_H is Gaussian:

$$\begin{aligned}&Pr\{B_H(x_1', x_2', \cdots, x_n') - B_H(x_1, x_2, \cdots, x_n) < \bar{\zeta}\} \\ &\qquad = \frac{1}{\sqrt{2\pi}\sigma r^H} \int_{-\infty}^{\bar{\zeta}} \exp\left(\frac{\zeta^2}{2\sigma^2 r^{2H}}\right) d\zeta,\end{aligned} \tag{6.26}$$

where

$$r = \sqrt{(x_1' - x_1)^2 + (x_2' - x_2)^2 + \cdots + (x_n' - x_n)^2},$$

H is the Hurst parameter, and σ^2 measures the variance of the process in unit "distance."

6.4 DIMENSIONS OF BM AND FBM PROCESSES

Due to their self-affine nature, Bm and fBm have several different dimensions. We first discuss their box-counting dimension.

Suppose we only consider a Bm or fBm process in the unit interval [0,1]. We partition the unit interval into N bins, the length of each bin being $\epsilon = \Delta t = 1/N$. Since $B_H(\Delta t) \propto \Delta t^H > \Delta t$, when $0 < H < 1$, for each bin, we will need $B_H(\Delta t)/\Delta t = \Delta t^{H-1}$ square boxes to cover the function. Overall, we need $B_H(\Delta t)/\Delta t \cdot N = \Delta t^{H-2} = \epsilon^{-(2-H)}$ boxes to cover the function. By Eq. (2.2), we find that the box-counting dimension for fBm is $D = 2 - H$. For Bm, $H = 0.5$, $D = 1.5$. The dimension $D = 2 - H$ is sometimes called the graph dimension of the random function. The smaller the Hurst parameter is, the larger the dimension of the process. This can be easily appreciated from Fig. 6.2.

The above argument can be easily extended to the n-dimensional case. To see how, let us replace Δt by Δr. We have $\epsilon = \Delta r = 1/N$, and $B_H(\Delta r) \propto \Delta r^H > \Delta r$. Now we need $B_H(\Delta r)/\Delta r \cdot N^n = \Delta r^{H-1-n}$ hypercubes of linear scale Δr to cover B_H. Therefore,

$$D = n + 1 - H.$$

Next, we consider the dimension obtained by measuring the length of the curve defined by Bm or fBm using a sequence of rulers. Consider a total time span of T_{total}, and partition it into equal time steps Δt. Then the length of the curve within each interval Δt is

$$\Delta l = \sqrt{\Delta t^2 + \Delta B_H^2} \propto \Delta t \sqrt{1 + \Delta t^{2H-2}}.$$

This value varies for different time intervals, but we assume that, for each Δt, it is bounded. When $\Delta t \ll 1$, $\Delta t^{2H-2} \gg 1$; hence, $\Delta l \propto \Delta t^H$. When $\Delta t \gg 1$, $\Delta l \propto \Delta t$. The number $N(\Delta l)$ of rulers of size Δl, which is $1/\Delta t$, is expected to vary as Δl^{-D}. Hence, for $\Delta t \gg 1$, $D = 1$, while for $\Delta t \ll 1$, $D = 1/H$. The latter is called the latent dimension.

We should point out that the above argument is also valid for an n-dimensional fBm process. Therefore, so far as the length of the fBm process is concerned, the latent dimension is always $D = 1/H$, regardless of the dimension of the space where fBm is defined.

To summarize, the same $B_H(t)$ can have an apparently self-similar dimension D of either 1, $1/H$, or $2 - H$, depending on the measurement technique and the choice of length scale. By now, the distinction between self-similarity and self-affinity should be clearer; self-affinity means that $B_H(t)$ and t can be scaled differently because they have different units.

One can obtain a number of interesting self-similar processes from a self-affine process. One simple process involves the so-called zerosets, defined by the time instants when $B_H(t) = 0$. A generalization of zerosets is the levelsets, obtained by the crossings of $B_H(t)$ with an arbitrary constant line, $B_H(t) = c$. Levelsets

have a fractal dimension of $1 - H$, which is simply one dimension lower than the box-counting dimension of $B_H(t)$. The gaps between successive points in a levelset have a power-law distribution, $P(r) \sim r^{-H}$, where r denotes the size of the gaps.

Another type of simple self-similar process can be obtained by using a number of independent fBm processes to form a vector process,

$$\mathbf{V}(t) = [B_H^{(1)}(t), B_H^{(2)}(t), ..., B_H^{(m)}(t)].$$

Now the units for all the components are the same. This is called an fBm trail or trajectory. Its fractal dimension is $1/H$, corresponding to the $\Delta t \ll 1$ case for the compass dimension (Now $\Delta t \gg 1$ case is no longer relevant.) $1/H$ is also its box-counting dimension if one realizes that the boxes used to cover the trace have to be square boxes with linear length $\Delta l (\propto \Delta t^H)$ instead of Δt.

The box-counting dimension for an fBm trail can be easily computed by the correlation dimension, originally developed by Grassberger and Procaccia for the characterization of chaotic systems. The algorithm works as follows. First, one samples $\mathbf{V}(t)$ to obtain $\mathbf{V}(1), \mathbf{V}(2), \cdots, \mathbf{V}(N)$. Then one computes the correlation integral,

$$C(N, r) = \frac{2}{N(N-1)} \sum_{i,j=1}^{N} \theta(r - \|\mathbf{V}(i) - \mathbf{V}(j)\|), \tag{6.27}$$

where θ is the Heaviside step function, N is the total number of points in the time series, and r is a prescribed small distance. The correlation dimension μ of the time series is then given by

$$C(N, r) \sim r^{\mu}, \quad r \to 0. \tag{6.28}$$

While in general the correlation dimension cannot be greater than the box-counting dimension, here it is the same as the box-counting dimension, $1/H$, since fBm is a monofractal.

In the chaos research community, for quite some time, it was thought that an estimated positive Lyapunov exponent or entropy, or a finite nonintegral correlation dimension would suffice to indicate that the time series under study is deterministically chaotic. $1/f$ processes, especially fBm processes, have been key counterexamples invalidating that assumption. Very recently, it has been clarified that the dimension estimated by the Grassberger and Procaccia algorithm corresponds to the box-counting dimension of the sojourn points in a neighborhood of an arbitrary point. In light of that result, we see that while the danger of misinterpreting a $1/f$ process to be deterministically chaotic can be readily avoided, one can in fact use the Grassberger and Procaccia algorithm to conveniently estimate the key parameter for the $1/f$ process.

The possibility of misinterpreting $1/f$ processes as being deterministic chaos never occurs if one works within the framework of power-law sensitivity to initial conditions that will be developed in Chapter 14. In that framework, one monitors the evolution of two nearby trajectories. If these trajectories diverge exponentially fast,

then the time series is chaotic; if the divergence increases in a power-law manner, then the trajectories belong to $1/f$ processes. These ideas can be easily expressed precisely. Let $||\mathbf{V}(s_1) - \mathbf{V}(s_2)||$ be the initial separation between two trajectories. This separation is assumed to be not larger than some small prescribed distance r. After time t, the distance between the two trajectories will be $||\mathbf{V}(s_1+t) - \mathbf{V}(s_2+t)||$. For truly chaotic systems,

$$||\mathbf{V}(s_1 + t) - \mathbf{V}(s_2 + t)|| \propto ||\mathbf{V}(s_1) - \mathbf{V}(s_2)||e^{\lambda t},$$

where $\lambda > 0$ is called the largest Lyapunov exponent. On the other hand, for $1/f$ processes,

$$||\mathbf{V}(s_1 + t) - \mathbf{V}(s_2 + t)|| \propto ||\mathbf{V}(s_1) - \mathbf{V}(s_2)||t^H.$$

The latter equation thus provides another simple means of estimating the Hurst parameter.

In practice, one often only has a scalar time series, possibly belonging to the family of $1/f$ processes. How may one get a vector time series so that one can use chaos theory to study the scalar time series? There is a very powerful tool called the time delay embedding technique, also developed for the study of chaotic systems, which is very easy to use. It will be discussed in detail in Chapter 13.

6.5 WAVELET REPRESENTATION OF FBM PROCESSES

In Chapter 4, Sec. 2, we introduced wavelet MRA. Using the scaling function and the wavelet function, fBm can be expressed as

$$B_H(t) = \sum_{k=-\infty}^{\infty} \phi_H(t-k) S_k^{(H)} + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} 2^{-jH} \psi_H(2^j t - k) \epsilon_{j,k} - b_0, \quad (6.29)$$

where ϕ_H and ψ_H are defined through their Fourier transform

$$\hat{\phi}_H(x) = \left(\frac{1 - e^{-ix}}{ix}\right)^{H+1/2} \hat{\phi}(x),$$

$$\hat{\psi}_H(x) = (ix)^{-(H+1/2)} \hat{\psi}(x),$$

where $\hat{\phi}(x)$ and $\hat{\psi}(x)$ are the Fourier transforms of a chosen scaling function $\phi(x)$ and a wavelet $\psi(x)$, respectively. $S_k^{(H)}$, $k \in Z$, is a partial sum process of a $FARIMA(0, H-1/2, 0)$ sequence, $\epsilon_{j,k}, j \geq 0, k \in Z$, are independent Gaussian $N(0,1)$ random variables, and b_0 is a random constant such that $B_H(0) = 0$.

The term $\psi_H(2^j t - k)$ in Eq. (6.29) seems to contradict the term $\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k)$ in Eq. (4.21), given the same j. To understand this, we note that in practical implementation of wavelet decomposition, one usually chooses a maximum scale resolution level J for analysis. If one redefines $j^* = J - j$,

then Eqs. (6.29) and (4.21) become consistent. The transformation $j^* = J - j$ amounts to viewing the wavelet multiresolution decomposition by two different ways: bottom-up or top-down.

The summands in the second term of Eq. (6.29) can be expressed as $d(j,k)\tilde{\psi}_{j,k}$, where

$$d(j,k) = 2^{-j(H+1/2)}\epsilon_{j,k} = 2^{j^*(H+1/2)}2^{-J(H+1/2)}\epsilon_{j,k} \tag{6.30}$$

and

$$\tilde{\psi}_{j,k} = 2^{j/2}\psi_H(2^j t - k). \tag{6.31}$$

For a specific scale j^*, let us denote the variance for the coefficients $d(j,k)$ by Γ_{j^*}. Then by Eq. (6.30), we have

$$\log_2 \Gamma_{j^*} = (2H+1)j^* + c_0, \tag{6.32}$$

where c_0 is some constant.

6.6 SYNTHESIS OF FBM PROCESSES

In the past several decades, a number of algorithms have been developed for synthesizing the fBm processes. We will describe some of them here, based on the consideration that they are either computationally very fast or will deepen our understanding of the fBm processes.

Random midpoint displacement (RMD) method. This method is similar to that for a Bm process. Suppose we are given $B_H(0) = 0$ and $B_H(1)$ as a sample of a Gaussian random variable with mean 0 and variance σ^2. To obtain $B_H(1/2)$, we also write

$$B_H(1/2) - B_H(0) = \frac{1}{2}[B_H(1) - B_H(0)] + D_1.$$

In order to have

$$Var(B_H(1/2) - B_H(0)) = Var(B_H(1) - B_H(1/2)) = (1/2)^{2H}\sigma^2,$$

D_1 must be a Gaussian random variable with zero mean and variance

$$(1/2)^{2H}\sigma^2(1 - 2^{2H-2}).$$

Applying this idea recursively, we find that D_n must have the variances

$$\Delta_n^2 = \sigma^2\left(\frac{1}{2}\right)^{2Hn}(1 - 2^{2H-2}). \tag{6.33}$$

It turns out that the method does not yield the true fBm when $H \neq 1/2$. One can easily verify that $Var(B_H(3/4) - B_H(1/4)) \neq (1/2)^{2H}\sigma^2$. Hence, the process does not have stationary increments. This defect causes the graphs of $B_H(t)$ to show some visible traces of the first few stages in the recursion, especially when H is close to 1. Visually, such features are not very pleasant.

Successive random addition (SRA) method. To overcome the shortcomings of the RMD method, Voss developed the SRA method. It has become quite popular due to its speed, efficiency, and flexibility in generating various random fractal surfaces and processes. The idea is to add a random variable to every point instead of only to the middle points at any stage. More concretely, the method works as follows. Start at $t_i = 0,\ 1/2,\ 1$. First, let $B_H(t_i) = 0$ be the initial condition. Now add a Gaussian random variable of zero mean and unit variance, $\sigma_1^2 = 1$, to all three points. We refer to these preparations as stage 1. At the next stage, we consider $t_i = 0,\ 1/4,\ 1/2,\ 3/4,\ 1$. $B_H(1/4)$ and $B_H(3/4)$ are first approximated by linear interpolations from $B_H(t_i)$, $t_i = 0,\ 1/2,\ 1$. Then a Gaussian random variable of zero mean and variance $\sigma_2^2 = (1/2)^{2H}\sigma_1^2$ is added to all five points. This is stage 2. At stage 3, the midpoints of these five points are again obtained by interpolation. Then Gaussian random variables of zero mean and variance $\sigma_3^2 = (1/2)^{2H}\sigma_2^2$ are added to all nine points. The method is applied recursively. At stage n, we have $\sigma_n^2 = (1/2)^{2H}\sigma_{n-1}^2 = (1/2)^{2H(n-1)}\sigma_1^2$.

It should be noted however, that the random processes or fields simulated by SRA are not truly statistically homogeneous. To overcome the difficulty, Elliott et al. and Majda et al. have recently introduced hierarchical Monte Carlo methods, which are capable of generating self-similarity over 12 decades of scales with about 100 realizations.

Fast Fourier transform (FFT) filtering. This is a simple and efficient method. By Eqs. (6.7) and (6.8), we have, for an fBm process, $S(f) \propto f^{-(2H+1)}$. One thus can proceed as follows: Start with a "white noise" sequence $W(t)$, whose spectral density is a constant, and filter the sequence with a transfer function $T(f) \propto f^{-(H+1/2)}$; then the output, $B_H(t)$, has the desired spectral density, $S_B(f) \propto |T(f)|^2 S_W(f) \propto f^{-(2H+1)}$. In computer simulations, we want to obtain a discrete sequence, B_n, defined at discrete times $t_n = n\Delta t,\ n = 0, 1, \cdots, N-1$. We can write:

$$B_n = \sum_{m=0}^{N-1} v_m e^{2\pi i f_m t_n}, \tag{6.34}$$

where

$$f_m = \frac{m}{N\Delta t} \quad \text{for } m = 0 \text{ to } N-1,$$

and

$$< |v_m|^2 > \propto f^{-(2H+1)} \propto m^{-(2H+1)}. \tag{6.35}$$

There are several ways to obtain v_m: (1) multiply the Fourier coefficients of a white noise sequence by $f^{-(H+1/2)}$; (2) directly choose complex random variables with mean square amplitude given by Eq. (6.35) and random phases; (3) simply set

$$|v_m|^2 = m^{-(2H+1)}$$

and randomize phases. The third method is perhaps the most popular.

Since the FFT-based algorithm is periodic in time, $B_n = B_{n+N}$, at most half of the sequence is usable. Typically, one generates a long sequence and only retains a portion with length $N/4$ to $N/2$. Another drawback of the above algorithms is that the autocorrelation function for short time lags will not exactly match that for the fBm.

Weierstrass-Mandelbrot function–based method. The Fourier series of Eq. (6.34) involves a linear progression of frequencies. There is an interesting function, called the Weierstrass-Mandelbrot function,

$$f(t) = \sum_{k=1}^{\infty} r^{(2-D)k} \sin(2\pi r^{-k} t), \tag{6.36}$$

where D is the box-counting dimension of the curve, which involves a geometric progression of frequencies. This function is continuous but nowhere differentiable. Since $H + D = 2$, we find that the amplitude for the frequency $f = r^{-k}$ is f^{-H}. The square amplitude is thus $\propto f^{-2H}$. In a bandwidth of $\Delta f \propto f$, we thus have the spectral density $S(f) \propto amplitude^2/\Delta f \propto f^{-(2H+1)}$. In order to build an fBm generator based on the Weierstrass-Mandelbrot function, one only needs to randomize the function in a certain fashion. A general form was given by Mandelbrot:

$$V_{WM}(t) = \sum_{k=-\infty}^{\infty} A_n r^{kH} \sin(2\pi r^{-k} t + \phi_k), \tag{6.37}$$

where A_n is a Gaussian random variable with the same variance for all n and ϕ_k is a random phase uniformly distributed on $[0, 2\pi]$. In numerical simulations, after one specifies the low and high cutoff frequencies, the series becomes finite. To speed up computation, one can simply set $\phi_k = 0$.

Inverse Fourier transform method. In applications where the exact autocovariance function of fBm or fGn (Eq. (6.18) or (6.23)) is required, one may use the FFT-based approach proposed by Yin. This method involves first computing an fGn process, and then obtaining an fBm process by summing the fGn. The fGn process is obtained by noting that its autocovariance function and spectral density form a Fourier transform pair. Since the former is known, its spectral density can be exactly computed. Generating a sequence from a known spectral density is then an easy matter.

Wavelet-based method. Starting from Eq. (6.29), Abry and Sellan proposed a fast algorithm to synthesize an fBm process. It works as follows. First, one generates a Gaussian $FARIMA(0, s, 0)$ sequence of finite length as an approximation of fBm at some coarse scale, where $s = H + 1/2$. Then one applies, recursively, a fast wavelet transform, which involves two filters, one called a fractional low-pass filter, and the other a fractional high-pass filter, to obtain a much longer $FARIMA(0, s, 0)$ sequence. Finally, the $FARIMA(0, s, 0)$ sequence is properly

normalized and taken for an approximation of fBm at some finer scale. Abry and Sellan have made their implementation available on the World Wide Web.

6.7 APPLICATIONS

In recent years, the fBm and related models have found numerous applications. In this section, we discuss network traffic modeling by fBm in some depth and then briefly describe modeling of rough surfaces by two-dimensional fBm or related processes.

6.7.1 Network traffic modeling

Network traffic modeling is an important yet challenging problem. Traditionally, traffic is modeled as a Poisson or Markovian process. A significant discovery in the early 1990s was that network traffic often has very long temporal correlations. In this section, we briefly describe how network traffic is specified; quantitatively show how miserably the Poisson traffic model fails to describe the measured traffic trace; and finally show that the fBm model is much more successful than the Poisson process model for network traffic.

6.7.1.1 Description of traffic patterns Network traffic is often measured by collecting interarrival times, $\{T_j\}$, where T_j denotes the jth interarrival time between two successive packet arrivals, and the packet-length sequence, $\{B_j\}$, where B_j represents the length of the jth packet. This is a point-process description. See Fig. 6.3(a). Sometimes one may only be concerned with the number of packets. Then one could set $B_j = b$, where b is the average size of the packets. This is shown in Fig. 6.3(b). The Poisson traffic model is a point-process approach.

Sometimes network traffic is specified in terms of aggregated traffic flows measured at a network node. This is the counting process description. Corresponding to Figs. 6.3(a,b), we have Figs. 6.3(c,d), respectively, designating the total number of bytes and packets in each time window Δt. The fBm model is a counting-process approach.

It is found that as long as the time window used for aggregating network traffic is not longer than the average message delay time, the point and counting processes are equivalent in terms of network performance.

6.7.1.2 Fractal behavior of network traffic Until about the late 1980s, researchers in the field of traffic engineering were quite happy with Poisson or Markovian traffic modeling. Since then, however, a number of traffic traces have been collected, first by the Bellcore (now Telcordia) people. Researchers were surprised to find out that real traffic differs greatly from Poisson traffic. The aggregation of LAN traffic data at different time scales is shown in the left panel of Fig. 6.4. The corresponding Poisson traffic, also aggregated at the same time scales, is shown in

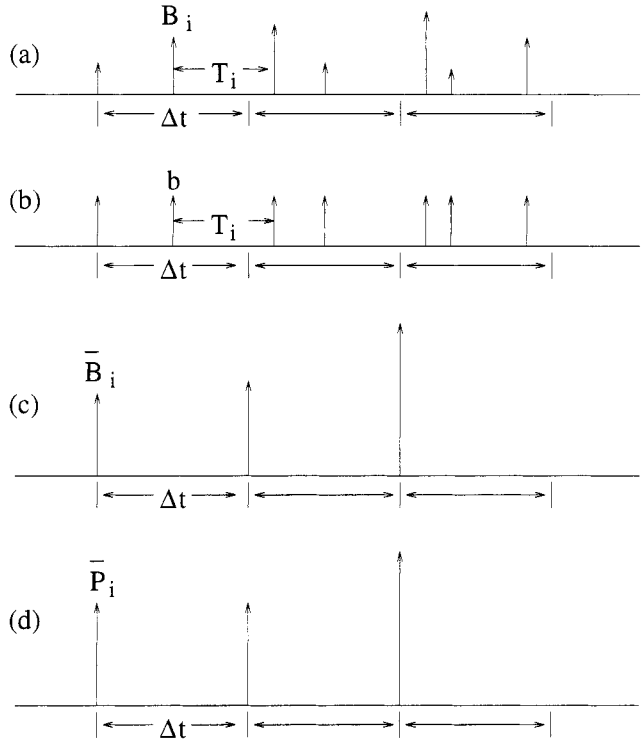

Figure 6.3. Schematics for the pattern of (a) Traffic B, (b) Traffic P, (c) Traffic $\overline{B}$, and (d) Traffic $\overline{P}$.

Fig. 6.4 in the right panel. We notice that the real traffic varies over the time scales from 0.01 to 100 s. However, the Poisson traffic varies only on time scales smaller than 1 s. The "little smoothing" behavior of real traffic can be characterized by the Hurst parameter, as we shall explain in Chapter 8.

6.7.1.3 Failure of the Poisson model In a Poisson model, interarrival times between two successive messages follow one exponential distribution; the size of the messages follows another exponential distribution. In queuing theory, this model is denoted as $M/M/1$. If a measured traffic trace is indeed Poissonian, then the two parameters, the mean of the message size and the interarrival times, can be easily estimated from the measured data.

To show quantitatively how bad the Poisson traffic model can be, let us study a single server queuing system with first-in-first-out (FIFO) service discipline and an infinite buffer, driven by both measured and Poisson traffic, and focus on three of the most important performance measures — throughput (which is also called utilization level), delay, and packet loss probability. Figure 6.5 shows a typical result of queue size tail distribution. To understand the result, the x axis can be

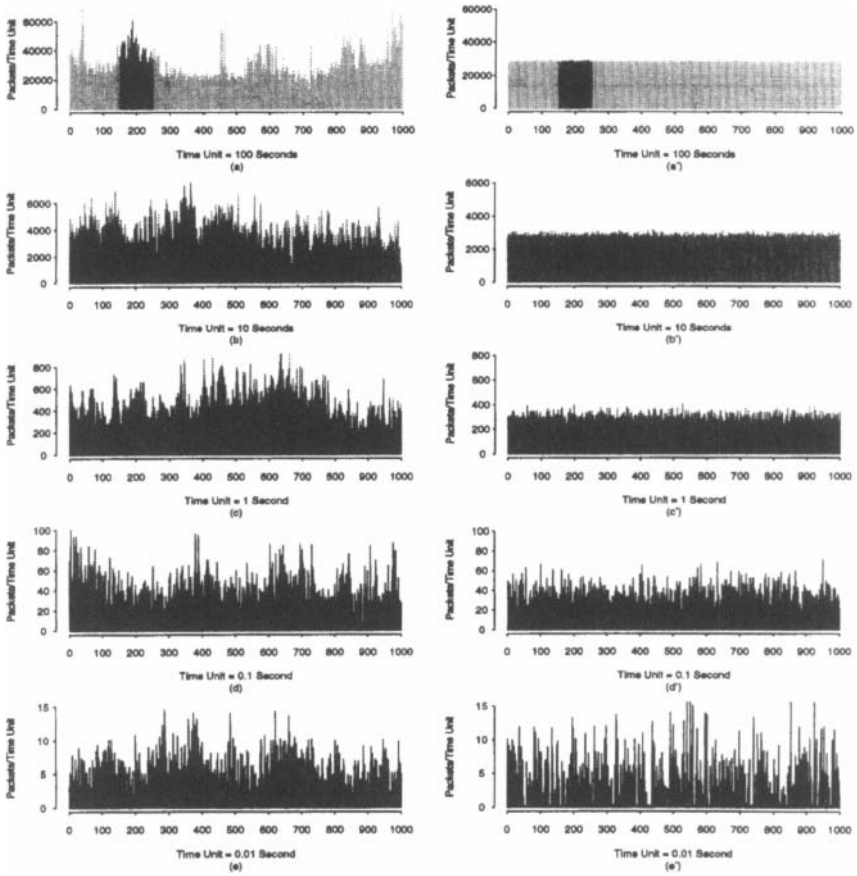

Figure 6.4. Left panel, (a)-(e), measured LAN traffic trace pAug.TL aggregated on different time scales; right panel, (a')-(e'), the corresponding (compound) Poisson traffic aggregated on the same time scales. Courtesy of Leland et al. [281].

considered the buffer size. Let us fix it at 600 Kbytes. Then for a given utilization level ρ, say, 0.5, the tail probability measures packet loss probability, which is above 1% for real traffic but much smaller than 10^{-4}% for the Poisson model. This means that Poisson modeling underestimates packet loss probability, delay time, and buffer size by several orders of magnitude.

6.7.1.4 Performance of the fBm model FBm is one of the most intensively studied models exhibiting the long-range dependence property in traffic engineering. After Norros introduced fBm as a traffic model, Erramilli et al. checked the complementary queue length distribution formula of Norros and found excellent agreement with simulations for a single server queuing system driven by some measured traffic trace operating at utilization $\rho = 0.5$. Gao and Rubin extended Erramilli

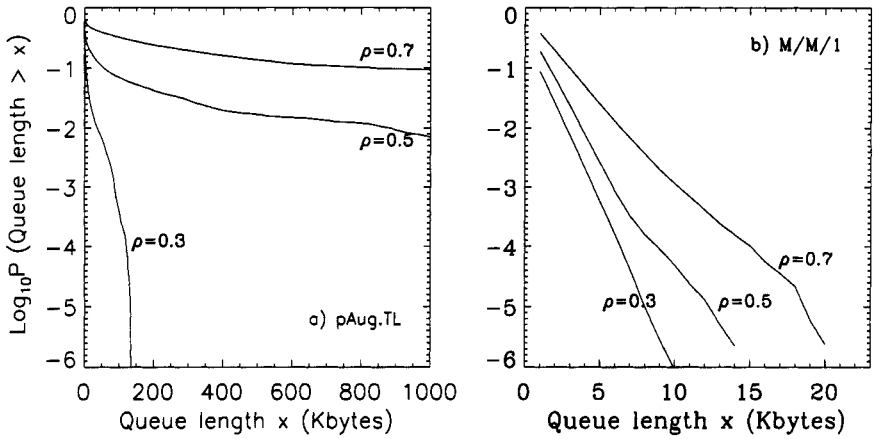

Figure 6.5. The complementary queue size distributions obtained when (a) a measured trace and (b) its corresponding Poisson traffic were used to drive a single-server queuing system. The three curves, from bottom to top, correspond to utilization levels $\rho = 0.3, 0.5$, and 0.7. The complementary queue size distribution estimates the packet loss probability when the buffer is of finite size. It also gives the delay time statistics when the queue length is normalized by the service rate. Notice the huge difference in the x-axis ranges. This means that Poisson modeling underestimates packet loss probability, delay time, and buffer size by several orders of magnitude.

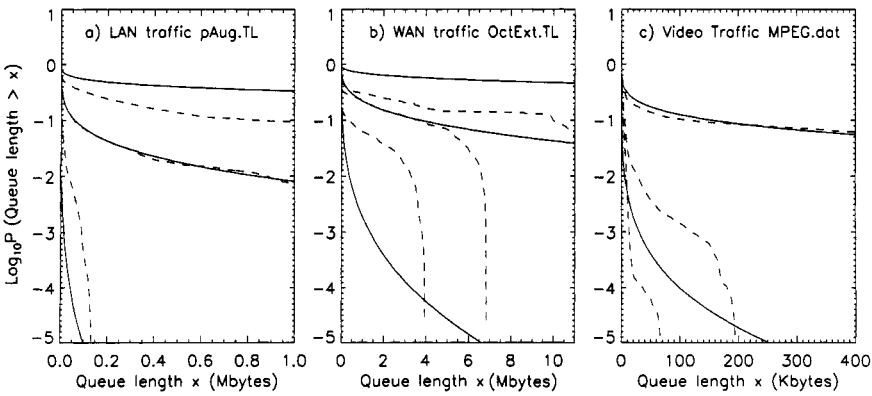

Figure 6.6. Comparison of complementary queue length distributions of single-server FIFO queuing systems driven by measured data (a) local area network (LAN) traffic, (b) wide area network (WAN) traffic, and (c) video traffic (dashed lines) and corresponding fBm traffic processes (solid lines). Three curves, from top to bottom, correspond to $\rho = 0.7$, 0.5, and 0.3, respectively. Note that for video traffic MPEG.data, the solid curve for $\rho = 0.3$ is too close to the y-axis to be seen.

et al.'s study by considering several utilization levels simultaneously. Figure 6.6 shows a typical result for three different types of traffic processes. When it is compared with Fig. 6.5, it is clear that fBm is indeed a much improved model. The key reason is that fBm has captured the long-range correlation properties of the traffic processes. We shall discuss in detail how to estimate long-range correlation properties in real data in Chapter 8.

6.7.2 Modeling of rough surfaces

One of the most vivid examples of high-dimensional fractals is the beautiful, irregular mountainous massifs. It is thus no wonder that two-dimensional fBm has been used to produce beautiful images. Recently, it has been found that scattering from natural rough surfaces modeled by two-dimensional fBm fits measurement data better than when the surface is modeled by nonfractal classic models. In related work, rough sea surface has been modeled by a Weierstrass-Mandelbrot function augmented with parameters describing wave motion on the sea surface. As we have explained, the fractal dimension of a rough surface and the Hurst parameter are related by a simple equation, $D = 3 - H$. It is thus clear that an important problem in such studies is to accurately estimate the Hurst parameter. In later chapters, we shall discuss in depth many methods for estimating the Hurst parameter; therefore, we shall not go into any of the details of rough surface modeling here.

6.8 BIBLIOGRAPHIC NOTES

For more in-depth treatment of Bm, we refer to [76]. Early studies of $1/f$ processes can be found, for example, in Press [353], Bak [22], and Wornell [483]. Some of the more recently discovered $1/f$ processes are found in traffic engineering [44,89,281,337], DNA sequences [284,340,463], genomes of DNA copy number changes associated with tumors [225], human cognition [188], ambiguous visual perception [151,496], coordination [71], posture [81], dynamic images [50,51], the distribution of prime numbers [479], and carbon nanotube devices [82,406], among many others. Mechanisms for proposed $1/f$ processes range from the superposition of many independent relaxation processes [199,315] to self-organized criticality [23,24]. A mechanism for fractal traffic is proposed by Willinger et al. [476], based on the superposition of many independent heavy-tailed ON/OFF sources, somewhat similar to the relaxation processes of [199,315]. This approach will be briefly reviewed in Chapter 11, where we shall also show how the observed traffic can be generated by a conventional communication network driven by Poisson traffic. Readers interested in the study of $1/f$ processes using state space and chaos theory are referred to [145,149,155,158,178–180,328,334,357,421,431]. Relevant references for the representation and synthesis of fBm processes using wavelets include [3,307,347,394]. References [295,387,462] discuss the RMD method, [130,461] discuss the SRA method, [118,291] discuss multi-wavelet expansion (MWE) Monte

Carlo methods, [387, 462, 488] discuss Fourier transform–based approaches for synthesizing fBm processes, and [48, 294] introduce the Weierstrass-Mandelbrot function. Finally, for traffic modeling, we refer to [119, 171–173, 321, 322], and for rough surface modeling and scattering on rough surfaces, we refer to [139, 207, 302].

6.9 EXERCISES

1. Let
$$Y = \sum_{i=1}^{n} X_i,$$
where X_i, $i = 1, 2, \cdots$ are independent uniformly distributed random variables on the unit interval [0,1]. For each fixed $n = 2, 3, 4, \cdots$, check to see if Y follows the normal distribution. What is the "minimal" n such that Y is almost normally distributed?

2. Imagine you are tossing a coin, and set $X = 1$ when you get a head and $X = -1$ when you get a tail. Assume that head and tail occur with an equal probability of 0.5. Now assume that at time $t = 0, 1, 2, \cdots$ you toss the coin. Form the sum $Y = \sum_{i=1}^{n} X_i$. Simulate a path of $Y(t)$. This is the standard Bm, often used to model how a complete drunk walks (in one-dimensional space).

3. Rework problem 2 with $p = 0.6$ for a head and $q = 1 - p = 0.4$ for a tail.

4. Develop a two-dimensional version of problems 2 and 3.

5. Prove Eq. (6.7) by starting from the defining property of Eq. (6.4).

6. Implement one or a few of the methods described in the text for synthesizing the fBm model. This will be handy in future.