



Copyrighted Material



The Nature of Mathematical Modeling

NEIL GERSHENFELD

$\mu_{\text{rowcluster}}(\exp(-\text{dot}(x - \mu_{\text{rowcluster}})))$

$(\text{rowcluster} * (x - \mu_{\text{rowcluster}}) \text{transpose}) / 2.0$
 $/ (\text{math.fabs}(\text{rowcluster})) * \text{sqrt}(\text{pow}(2.0 * \text{pi}, \text{dim} / 2.0)))$

$$\nabla_{\vec{\mu}_m} L = \nabla_{\vec{\mu}_m} \log \prod_{n=1}^N p(\vec{x}_n)$$



$$\sum_{n=1}^N \bar{x}$$

Copyrighted Material

5 Random Systems

So far we have been studying deterministic systems. But the world around us is not very deterministic; there are fluctuations in everything from a cup of coffee to the global economy. In principle, these could be described from first principles, accounting for the actions of every molecule in the cup or every person on the planet. This is hopeless of course, for even if we could determine the initial condition of the whole system, any tiny errors we made would rapidly grow so that our model would no longer be exact. Even if that wasn't a problem, such an enormous model would offer little insight beyond observing the original system.

Fortunately, random (or *stochastic*) systems can be as simple to understand as deterministic systems, if we're careful to ask the right questions. It's a mistake to try to guess the value of the toss of a fair coin, because it is completely uncertain. But we can quite precisely answer questions about the likelihood of events such as seeing a particular string of heads and tails, and easily model a typical sequence of tosses. In fact, simple stochastic systems are as straightforward as simple deterministic ones. Just as deterministic systems get harder to understand as they become more complex, stochastic systems become harder to understand as they become less ideally random, and approximation methods are needed to describe this deviation.

Different problems lie at different ends of this continuum. For a bridge builder, the bridge had better remain standing with near-certain probability, so the uncertainty is $\approx 1 - \epsilon$, where ϵ is very small. For a financial trader, markets are nearly random, and so any (legal) trading strategy must start with the assumption of near-random uncertainty. There, the probabilities are $\approx 0.5 + \epsilon$. The representations appropriate to $1 - \epsilon$ are different from the ones aimed at $0.5 + \epsilon$, but they do share the central role of variables that are random rather than deterministic.

Our first step will be to look at how to describe the properties of a random variable, and the inter-relationships among a collection of them. Then we will turn to stochastic processes, which are random systems that evolve in time. The chapter closes by looking at algorithms for generating random numbers.

5.1 RANDOM VARIABLES

The most important concept in all of stochastic systems is the idea of a *random variable*. This is a fluctuating quantity, such as the hiss from an amplifier, that is described by governing equations in terms of probability distributions. The crucial distinction between

random variables and ordinary deterministic variables is that it is not possible to predict the value of a random variable, but it is possible to predict the probability for seeing a given event in terms of the random variable (such as the likelihood of observing a value in a particular range, or the functional form of the power spectrum). An ensemble of identical stochastic processes will generate different values for the random variables, but each member of the ensemble (or *realization*) will obey these same distribution laws. It is important to distinguish between the random variables that appear in distribution functions and the particular values that are obtained in a single realization of a stochastic process by drawing from the distribution.

The simplest random system consists of values x taken from a distribution $p(x)$. For example, in a coin toss x can be heads or tails, and $p(\text{heads}) = p(\text{tails}) = 1/2$. In this case x takes on discrete values; it is also possible for a random variable to come from a continuous distribution. For a continuous variable, $p(x) dx$ is the probability to observe a value between x and $x + dx$, and more generally

$$\int_a^b p(x) dx \quad (5.1)$$

is the probability to observe x between a and b . For a continuous variable, remember that $p(x)$ must always appear with a dx if you want to make a statement about the likelihood of an observable event – a common mistake is to try to evaluate the likelihood of an event by using the value of the density $p(x)$ alone without integrating it over an interval. That is meaningless; among many problems it can easily lead to the impossible result of probabilities greater than 1.

Now consider a string (x_1, x_2, \dots, x_N) of N x 's drawn from a distribution $p(x)$ (such as many coin tosses). The average value (or *expectation value*) of a function $f(x)$ is defined by

$$\begin{aligned} \langle f(x) \rangle &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N f(x_i) \\ &= \int_{-\infty}^{\infty} f(x) p(x) dx \quad . \end{aligned} \quad (5.2)$$

(in the statistical literature, the expectation is usually written $E[f(x)]$ instead of $\langle f(x) \rangle$). The equivalence of these two equations can be taken as an empirical definition of the probability distribution $p(x)$. This is for a continuous variable; here and throughout this chapter, for a discrete variable the integral is replaced by a sum over the allowed values. If $p(x)$ is not defined from $-\infty$ to ∞ then the integral extends over the values for which it is defined (this is called the *support*). A trivial example of an expectation value is

$$\langle 1 \rangle = \int_{-\infty}^{\infty} p(x) dx = 1 \quad (5.3)$$

(since probability distributions must be normalized). An important expectation is the *mean value*

$$\bar{x} \equiv \langle x \rangle = \int_{-\infty}^{\infty} x p(x) dx \quad . \quad (5.4)$$

The mean value might never actually occur (for example $p(x)$ might be a *bimodal* dis-

tribution with two peaks that vanish at $p(\bar{x})$). Therefore, another useful quantity is the value of x at which $p(x)$ is a maximum, the *maximum likelihood* value.

The mean value tells us nothing about how big the fluctuations in x are around it. A convenient way to measure this is by the *variance* σ_x^2 , defined to be the average value of the square of the deviation around the mean

$$\begin{aligned}\sigma_x^2 &\equiv \langle (x - \bar{x})^2 \rangle \\ &= \langle x^2 - 2x\bar{x} + \bar{x}^2 \rangle \\ &= \langle x^2 \rangle - 2\langle x \rangle \bar{x} + \bar{x}^2 \\ &= \langle x^2 \rangle - 2\bar{x}^2 + \bar{x}^2 \\ &= \langle x^2 \rangle - \langle x \rangle^2\end{aligned}\quad (5.5)$$

(remember that $\bar{x} = \langle x \rangle$ is just a constant). The square root of the variance is called the *standard deviation* σ_x . To calculate the variance, we need to know the mean value, and the *second moment*

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 p(x) dx \quad . \quad (5.6)$$

It is similarly possible to define the higher-order moments $\langle x^n \rangle$.

5.1.1 Joint Distributions

Now let's consider two random variables x and y , such as the result from throwing a pair of dice, that are specified by a joint density $p(x, y)$. The expected value of a function that depends on both x and y is

$$\langle f(x, y) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) p(x, y) dx dy \quad . \quad (5.7)$$

$p(x, y)$ must be normalized, so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1 \quad . \quad (5.8)$$

It must also be normalized with respect to each variable, so that

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy \quad (5.9)$$

and

$$p(y) = \int_{-\infty}^{\infty} p(x, y) dx \quad . \quad (5.10)$$

Integrating a variable out of a joint distribution is called *marginalizing* over the variable.

For joint random variables a very important quantity is $p(x|y)$ ("the probability of x given y "). This is the probability of seeing a particular value of x if we already know the value of y , and is defined by *Bayes' rule*

$$p(x|y) = \frac{p(x, y)}{p(y)} \quad , \quad (5.11)$$

which takes the joint probability and divides out from it the known scalar probability. This is easily extended to combinations of more variables,

$$\begin{aligned} p(x, y, z) &= p(x|y, z) p(y, z) \\ &= p(x|y, z) p(y|z) p(z) \\ &= p(x, y|z) p(z) \quad , \end{aligned} \tag{5.12}$$

manipulations that will recur in later chapters.

If x and y are *independent*, $p(x, y) = p(x) p(y)$. The probability of seeing two independent events is the product of the probabilities of seeing each alone. For independent variables, the conditional distribution will then depend on just one variable, $p(x|y) = p(x)$. This provides a convenient way to remember the form of Bayes' rule, because for independent variables $p(x|y) = p(x)p(y)/p(y) = p(x)$. For *uncorrelated* variables, $\langle xy \rangle = \langle x \rangle \langle y \rangle$. Independent variables are always uncorrelated, but the converse need not be true (although it often is).

Bayes' rule has an almost mystical importance to the (frequently vocal) community of *Bayesians*, who object to any distribution that is not conditional. Equation (5.2) represents a *frequentist* viewpoint, in which probabilities are defined by observed fractions of events. If you are told that the probability of flipping a coin and seeing a head is 50%, you can perform a series of trials to check that. But, if you're told that the probability of rain tomorrow is 50%, you can't check that in the same way. The day will happen only once, so it doesn't make sense to discuss an ensemble of tomorrows. This kind of probability is really a conclusion drawn from a collection of models and prior beliefs. Since you almost always believe something, when you announce to the world $p(\text{observation})$, you are really saying $p(\text{observation}|\text{prior beliefs})$. From Bayes' rule, we know that these are related by

$$\begin{aligned} p(\text{observation}|\text{prior beliefs}) &= \frac{p(\text{observation, prior beliefs})}{p(\text{prior beliefs})} \\ &= \frac{p(\text{prior beliefs}|\text{observation}) p(\text{observation})}{p(\text{prior beliefs})} \quad . \end{aligned} \tag{5.13}$$

A term $p(\text{prior beliefs})$ is called a *prior*, and no well-equipped Bayesian would be caught without one. While this insistence can be dogmatic, it has helped guide many people to recognize and clearly state their procedure for handling advance knowledge about a system and how it gets updated based on experience. We'll have much more to say about this in Chapter 10.

To work with random variables we must understand how their distributions change when they are transformed in mathematical expressions. First, if we know a distribution $p(x)$, and are given a change of coordinates $y(x)$, then what is the distribution $p(y)$? This is easily found by remembering that probabilities are defined in terms of the value of the distribution at a point times a differential element, and equating these:

$$\begin{aligned} p(x) |dx| &= p(y) |dy| \\ \Rightarrow p(x) &= p(y) \left| \frac{dy}{dx} \right| \quad . \end{aligned} \tag{5.14}$$

In higher dimensions, the transformation of a distribution is done by multiplying it by the *Jacobian*, the determinant of the matrix of partial derivatives of the change of coordinates.

If we take $p(y) = 1$ in the unit interval and integrate both sides of equation (5.14), this relationship has a useful implication

$$\int_0^x p(x') dx' = \int_0^x 1 \frac{dy(x')}{dx} dx'$$

$$P(x) = y(x) \quad . \quad (5.15)$$

This means that if we choose the transformation $y(x)$ to be the integral $P(x)$ of a given distribution $p(x)$ (P is called the *cumulative distribution*), and pick random values of y from a uniform distribution, then the corresponding values of x will be distributed according to p .

The next question is how to combine random variables. The simplest operation is adding two independent random variables, x_1 drawn from p_1 , and x_2 drawn from p_2 , to create a new random variable $y = x_1 + x_2$. To find the probability distribution $p(y)$ for y , we must add up the probabilities for each of the different ways that x_1 and x_2 can add up to that value of y . The probability of seeing a particular pair of x_1 and x_2 is given by the product of their probabilities, and so integrating we see that

$$p(y) = \int_{-\infty}^{\infty} p_1(x) p_2(y - x) dx$$

$$= p_1(x) * p_2(x) \quad . \quad (5.16)$$

The probability distribution for the sum of two random variables is the convolution of the individual distributions. Now consider the average of N variables

$$y_N = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (5.17)$$

that are independent and identically distributed (often abbreviated as *iid*), and let's look at what happens as $N \rightarrow \infty$. The distribution of y is equal to the distribution of x convolved with itself N times, and since taking a Fourier transform turns convolution into multiplication, the Fourier transform of the distribution of y is equal to the product of the N transforms of the distribution of x . This suggests an important role for Fourier transforms in studying random processes.

5.1.2 Characteristic Functions

The Fourier transform of a probability distribution is called the *characteristic function*, and is equal to the expectation value of the complex exponential

$$\langle e^{ikx} \rangle = \int_{-\infty}^{\infty} e^{ikx} p(x) dx \quad . \quad (5.18)$$

For a multivariate distribution, the characteristic function is

$$\langle e^{i\vec{k} \cdot \vec{x}} \rangle = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{i\vec{k} \cdot \vec{x}} p(\vec{x}) d\vec{x} \quad . \quad (5.19)$$

Now let's look at the characteristic function for the deviation of y_N , the average of N iid random numbers x_1, \dots, x_N , from the average value \bar{x} :

$$\begin{aligned}
 \langle e^{ik(y_N - \bar{x})} \rangle &= \langle e^{ik(x_1 + x_2 + \dots + x_N - N\bar{x})/N} \rangle \\
 &= \langle e^{ik[(x_1 - \bar{x}) + \dots + (x_N - \bar{x})]/N} \rangle \\
 &= \langle e^{ik(x - \bar{x})/N} \rangle^N \\
 &= \left\langle 1 + \frac{ik}{N}(x - \bar{x}) - \frac{k^2}{2N^2}(x - \bar{x})^2 + \mathcal{O}\left(\frac{k^3}{N^3}\right) \right\rangle^N \\
 &= \left[1 + 0 - \frac{k^2 \sigma_x^2}{2N^2} + \mathcal{O}\left(\frac{k^3}{N^3}\right) \right]^N \\
 &\approx e^{-k^2 \sigma_x^2 / 2N} .
 \end{aligned} \tag{5.20}$$

In deriving this we have used the fact that the x_i are independent and identically distributed, and done a Taylor expansion around the average value. The last line follows because

$$\lim_{N \rightarrow \infty} \left[1 + \frac{x}{N} \right]^N = e^x \tag{5.21}$$

(which can be verified by comparing the Taylor series of both sides), and we can neglect higher-order terms in the limit $N \rightarrow \infty$. To find the probability distribution for y we now take the inverse transform

$$\begin{aligned}
 p(y_N - \bar{x}) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-k^2 \sigma_x^2 / 2N} e^{-ik(y - \bar{x})} dk \\
 &= \sqrt{\frac{N}{2\pi \sigma_x^2}} e^{-N(y_N - \bar{x})^2 / 2\sigma_x^2} .
 \end{aligned} \tag{5.22}$$

Something remarkable has happened: in the limit $N \rightarrow \infty$, the sum of N variables approaches a *Gaussian* distribution, independent of the distribution of the variables! This is called the *Central Limit Theorem*, and explains why Gaussians are so important in studying random processes. The Gaussian distribution is also called the *normal* distribution, because it is so, well, normal. Since the standard deviation is σ_x / \sqrt{N} , which vanishes as $N \rightarrow \infty$, equation (5.22) also contains the *Law of Large Numbers*: the average of a large number of random variables approaches the mean, independent of the distribution.

The characteristic function takes on an interesting form if the complex exponential is expanded in a power series:

$$\begin{aligned}
 \langle e^{i\vec{k} \cdot \vec{x}} \rangle &= \int e^{i\vec{k} \cdot \vec{x}} p(\vec{x}) d\vec{x} \\
 &= \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \dots \sum_{n_d=0}^{\infty} \frac{(ik_1)^{n_1}}{n_1!} \frac{(ik_2)^{n_2}}{n_2!} \dots \frac{(ik_N)^{n_N}}{n_N!} \langle x_1^{n_1} x_2^{n_2} \dots x_N^{n_N} \rangle .
 \end{aligned} \tag{5.23}$$

This provides a relationship between the moments of a distribution and the distribution itself (via its Fourier transform). For this reason, the characteristic function is also called the *moment generating function*. If the characteristic function is known, the moments

can be found by taking derivatives of the appropriate order and evaluating at $\vec{k} = 0$ (only one term will be nonzero).

Another important object is the logarithm of the characteristic function. If we choose to write this as a power series in k of the form (for the 1D case)

$$\log \langle e^{ikx} \rangle = \sum_{n=1}^{\infty} \frac{(ik)^n}{n!} C_n \quad , \quad (5.24)$$

this defines the *cumulants* C_n (note that the sum starts at $n = 1$ because $\log 1 = 0$ and so there is no constant term). The cumulants can be found by comparing this to the power series expansion of the characteristic function,

$$\exp \left(\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} C_n \right) = \sum_{n=0}^{\infty} \frac{(ik)^n}{n!} \langle x^n \rangle \quad , \quad (5.25)$$

expanding the exponential as

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots \quad , \quad (5.26)$$

and grouping terms by order of k . The cumulants have an interesting connections to Gaussianity (Problem 5.1).

5.2 STOCHASTIC PROCESSES

It is now time for time to appear in our discussion of random systems. When it does, this becomes the study of *stochastic processes*. We will look at two ways to bring in time: the evolution of probability distributions for variables correlated in time, and stochastic differential equations.

If $x(t)$ is a time-dependent random variable, its Fourier transform

$$X(\nu) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} e^{i2\pi\nu t} x(t) dt \quad (5.27)$$

is also a random variable but its *power spectral density* $S(\nu)$ is not:

$$\begin{aligned} S(\nu) &= \langle |X(\nu)|^2 \rangle = \langle X(\nu) X^*(\nu) \rangle \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} e^{i2\pi\nu t} x(t) dt \int_{-T/2}^{T/2} e^{-i2\pi\nu t'} x(t') dt' \end{aligned} \quad (5.28)$$

(where X^* is the complex conjugate of X , replacing i with $-i$). The inverse Fourier transform of the power spectral density has an interesting form,

$$\begin{aligned} &\int_{-\infty}^{\infty} S(\nu) e^{-i2\pi\nu\tau} d\nu \\ &= \int_{-\infty}^{\infty} \langle X(\nu) X^*(\nu) \rangle e^{-i2\pi\nu\tau} d\nu \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{\infty} \int_{-T/2}^{T/2} e^{i2\pi\nu t} x(t) dt \int_{-T/2}^{T/2} e^{-i2\pi\nu t'} x(t') dt' e^{-i2\pi\nu\tau} d\nu \end{aligned}$$

$$\begin{aligned}
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\infty}^{\infty} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} e^{i2\pi\nu(t-t'-\tau)} d\nu x(t)x(t') dt dt' \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} \delta(t-t'-\tau) x(t)x(t') dt dt' \\
&= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t-\tau) dt \\
&= \langle x(t)x(t-\tau) \rangle \quad , \tag{5.29}
\end{aligned}$$

found by using the Fourier transform of a *delta function*

$$\int_{-\infty}^{\infty} e^{-i2\pi\nu t} \delta(t) dt = 1 \Rightarrow \delta(t) = \int_{-\infty}^{\infty} e^{i2\pi\nu t} dt \quad , \tag{5.30}$$

where the delta function is defined by

$$\int_{-\infty}^{\infty} f(x) \delta(x-x_0) dx = f(x_0) \quad . \tag{5.31}$$

This is the *Wiener-Khinchin* theorem. It relates the spectrum of a random process to its *autocovariance function*, or, if it is normalized by the variance, the *autocorrelation function* (which features prominently in time series analysis, Chapter 16).

5.2.1 Distribution Evolution Equations

A natural way to describe a stochastic process is in terms of the probability to see a sample value x at a time t (written x_t) given a history of earlier values

$$p(x_t | x_{t_1}, x_{t_2}, \dots) \quad . \tag{5.32}$$

Given starting values for x this determines the probability distribution of the future values. If the distribution depends only on the time differences and not on the absolute time,

$$p(x_t | x_{t-\tau_1}, x_{t-\tau_2}, \dots) \quad , \tag{5.33}$$

then the process is said to be *stationary* (sometimes qualified by calling it *narrow-sense* or *strict stationarity*). A more modest definition asks only that the means and covariances of a process be independent of time, in which case the process is said to be *weak* or *wide-sense stationary*.

If the conditional distribution is limited to a finite history

$$p(x_t | x_{t-\tau_1}, x_{t-\tau_2}, \dots, x_{t-\tau_N}) \tag{5.34}$$

this is said to be an N th-order *Markov process*. If it depends on just the previous value

$$p(x_t | x_{t-\tau}) \tag{5.35}$$

it is simply called a Markov process, and if x and t are discrete variables

$$p(x_t | x_{t-1}) \tag{5.36}$$

it becomes a *Markov Chain*. As with ODEs, an N th-order Markov process for a scalar

variable can always be converted to a first-order Markov process in an N -dimensional variable.

Given the conditional distribution for one time step $p(x_t|x_{t-\tau})$ we can find the distribution two time steps ahead $p(x_{t+\tau}|x_{t-\tau})$ by adding up the probabilities for all of the ways to get from $x_{t-\tau}$ to $x_{t+\tau}$ through the intermediate value x_t . The probability for each possible path is the product of the probability to get from $x_{t-\tau}$ to x_t times the probability to go from x_t to $x_{t+\tau}$. For a discrete system this is a sum, and for a continuous system it is the integral

$$p(x_{t+\tau}|x_{t-\tau}) = \int_{-\infty}^{\infty} p(x_{t+\tau}|x_t) p(x_t|x_{t-\tau}) dx_t \quad . \quad (5.37)$$

This is called the *Chapman–Kolmogorov* equation. It can be rewritten by multiplying both sides by $p(x_{t-\tau})$ and then integrating over $x_{t-\tau}$:

$$\begin{aligned} p(x_{t+\tau}|x_{t-\tau}) p(x_{t-\tau}) &= \int_{-\infty}^{\infty} p(x_{t+\tau}|x_t) p(x_t|x_{t-\tau}) p(x_{t-\tau}) dx_t \\ p(x_{t+\tau}, x_{t-\tau}) &= \int_{-\infty}^{\infty} p(x_{t+\tau}|x_t) p(x_t, x_{t-\tau}) dx_t \\ p(x_{t+\tau}) &= \int_{-\infty}^{\infty} p(x_{t+\tau}|x_t) p(x_t) dx_t \\ &= \int_{-\infty}^{\infty} p(x_{t+\tau}, x_t) dx_t \quad . \end{aligned} \quad (5.38)$$

For a Markov chain with N states $x = (1, \dots, N)$ this becomes

$$p(x_{t+1}) = \sum_{x_t=1}^N p(x_{t+1}|x_t) p(x_t) \quad . \quad (5.39)$$

If we define an N -component vector of the state probabilities $\vec{p}_t = \{p(x_t = 1), \dots, p(x_t = N)\}$, and a matrix of transition probabilities $\mathbf{P}_{ij} = p(x_{t+1} = i | x_t = j)$, then the update for all of the states can be written as

$$\begin{aligned} \vec{p}_{t+1} &= \mathbf{P} \cdot \vec{p}_t \\ \Rightarrow \vec{p}_{t+n} &= \mathbf{P}^n \cdot \vec{p}_t \quad . \end{aligned} \quad (5.40)$$

The powers of \mathbf{P} hence determine the evolution; in particular, if it's possible to get from every state to every other state then the system is said to be *ergodic* [Reichl, 1984].

It's easy to understand what a Markov model can do. After all, equation (5.40) is a simple linear first-order finite difference equation (Section 2.5). If \mathbf{P} has eigenvectors \vec{v}_i with eigenvalues λ_i , and the starting distribution is written in terms of the eigenvectors as

$$\vec{p}_t = \sum_i \alpha_i \vec{v}_i \quad , \quad (5.41)$$

then

$$\begin{aligned} \vec{p}_{t+n} &= \mathbf{P}^n \sum_i \alpha_i \vec{v}_i \\ &= \sum_i \alpha_i \lambda_i^n \vec{v}_i \quad . \end{aligned} \quad (5.42)$$

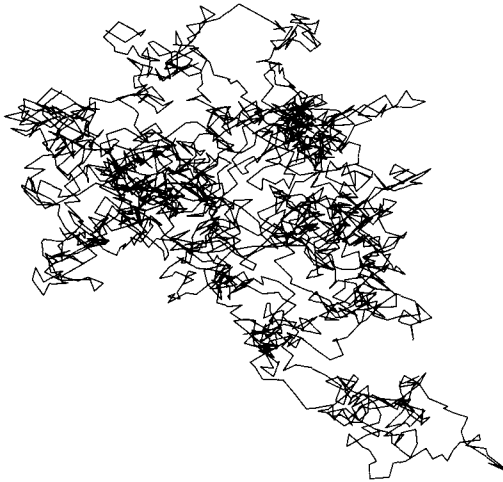


Figure 5.1. 2D Brownian motion.

One of the earliest uses of the Chapman–Kolmogorov equation was in the context of *Brownian motion*. When the Scottish botanist Robert Brown in 1827 used a microscope to look at pollen grains suspended in a solution, he saw them move in a path like the one shown in Figure 5.1. While he originally thought this was a sign of life, it was of course due to the fluctuating impacts of the solvent molecules on the pollen grain. This was a significant piece of evidence for the atomic theory of matter, and in 1905 Einstein developed a quantitative description of Brownian motion that could predict the distribution for how far a particle would travel in a given time.

For 1D Brownian motion (the generalization from 1D to higher dimensions is straightforward), let $f(x, t) dx$ be the probability to find the particle between x and $x + dx$ at time t . Because it is a probability, f must be normalized, $\int_{-\infty}^{\infty} f(x, t) dx = 1$. Brownian motion arises from the many impacts of the fluid molecules on the particle; let τ be a time that is long compared to these impacts, but short compared to the time for macroscopic motion of the particle. We can then define $p_{\tau}(x|x') dx$, which is the probability for the particle to move from x' to x in a time τ due to the fluctuating impacts on it. This can be used in the Chapman–Kolmogorov equation to write an update equation for f :

$$f(x, t + \tau) dx = dx \int_{-\infty}^{\infty} p_{\tau}(x|x') f(x', t) dx' \quad . \quad (5.43)$$

If the fluid is isotropic and homogeneous, then $p_{\tau}(x|x')$ will depend only on the position difference $x = x' + \delta$, so we can then write $p_{\tau}(x|x') = p_{\tau}(\delta)$, where by symmetry $p(\delta) = p(-\delta)$. Then the Chapman–Kolmogorov equation becomes

$$f(x, t + \tau) dx = dx \int_{-\infty}^{\infty} p_{\tau}(\delta) f(x + \delta, t) d\delta \quad . \quad (5.44)$$

Since τ has been chosen to be very small compared to the time scale for macroscopic changes in the particle position, we can expand f in a Taylor series in x and t and keep

the lowest-order terms:

$$\begin{aligned}
 & \left[f(x, t) + \frac{\partial f}{\partial t} \tau + \mathcal{O}(\tau^2) \right] dx \\
 &= dx \int_{-\infty}^{\infty} p_{\tau}(\delta) \left[f(x, t) + \frac{\partial f}{\partial x} \delta + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \delta^2 + \dots \right] d\delta \\
 &= dx f(x, t) \underbrace{\int_{-\infty}^{\infty} p_{\tau}(\delta) d\delta}_1 + dx \frac{\partial f}{\partial x} \underbrace{\int_{-\infty}^{\infty} p_{\tau}(\delta) \delta d\delta}_0 + \\
 & \quad dx \frac{1}{2} \frac{\partial^2 f}{\partial x^2} \underbrace{\int_{-\infty}^{\infty} p_{\tau}(\delta) \delta^2 d\delta}_{\langle \delta^2 \rangle} + \mathcal{O}(\delta^3) \quad . \quad (5.45)
 \end{aligned}$$

In the last line, the first integral is 1 because $p_{\tau}(\delta)$ must be normalized, the second one vanishes because it is the first moment of a symmetrical function $p_{\tau}(\delta) = p_{\tau}(-\delta)$, and the last integral is the variance. Cancelling out the $f(x, t)$ on both sides, we're left with

$$\frac{\partial f}{\partial t} = \underbrace{\frac{\langle \delta^2 \rangle}{2\tau}}_{\equiv D} \frac{\partial^2 f}{\partial x^2} \quad . \quad (5.46)$$

We've found that the evolution of the probability distribution for Brownian motion is governed by the familiar diffusion equation. In the context of stochastic processes, this is called a *Wiener process*. The diffusion equation is a particular case of the *Fokker-Plank* PDE which governs the evolution of a probability distribution for an underlying Markov process.

We obtained the diffusion equation for a system in which time and space are continuous, but for many problems they are discrete. Consider a 1D random walker that at each time step t_i can hop from a site x_n one unit to the left or right, $x_{n\pm 1}$. The change in probability $p(x_n, t_i)$ to find it at a point is then equal to the sum of the probabilities for it to hop into the point, minus the sum of probabilities to hop off the point

$$\begin{aligned}
 p(x_n, t_i) - p(x_n, t_{i-1}) &= \frac{1}{2} [p(x_{n+1}, t_{i-1}) + p(x_{n-1}, t_{i-1})] - p(x_n, t_{i-1}) \\
 p(x_n, t_i) &= \frac{1}{2} [p(x_{n+1}, t_{i-1}) + p(x_{n-1}, t_{i-1})] \quad . \quad (5.47)
 \end{aligned}$$

This can be rewritten suggestively by subtracting $p(x_n, t_{i-1})$ from both sides and rearranging:

$$\frac{p(x_n, t_i) - p(x_n, t_{i-1})}{\delta_t} = \underbrace{\frac{\delta_x^2}{2\delta_t}}_{\equiv D} \left[\frac{p(x_{n+1}, t_{i-1}) - 2p(x_n, t_{i-1}) + p(x_{n-1}, t_{i-1})}{\delta_x^2} \right] \quad (5.48)$$

These are discrete approximations of partial derivatives (Chapter 7). If we take the limit $\delta_t, \delta_x \rightarrow 0$ (keeping D finite), we find once again that

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2} \quad . \quad (5.49)$$

More generally, the time derivative of the probability to be at a state x_n is equal to the sum over states x_m of the probability to be at x_m times the rate $W_{x_m \rightarrow x_n}$ at which transitions are made from there to x_n , minus the probability to be at x_n times the rate at which transitions are made back to x_m :

$$\frac{\partial p(x_n, t)}{\partial t} = \sum_m W_{x_m \rightarrow x_n} p(x_m, t) - W_{x_n \rightarrow x_m} p(x_n, t) \quad . \quad (5.50)$$

This is called the *Master equation*. For a stationary solution the transition rate between two sites is equal in both directions, a condition called *detailed balance*.

5.2.2 Stochastic Differential Equations

An alternative analysis of Brownian motion was first done by Langevin in terms of a stochastic differential equation. A particle moving in a fluid feels a drag force, and as long as the velocity is not too great the force is given by the Stokes drag formula

$$\vec{F} = -6\pi\mu a \vec{v} \quad , \quad (5.51)$$

where μ is the viscosity of the fluid, a is the diameter of the particle, and \vec{v} is the velocity of the particle [Batchelor, 1967]. In addition to this force, we can model Brownian motion by including a fluctuating force η that is due to the molecular impacts on the particle. In terms of these forces, $\vec{F} = m\vec{a}$ for the particle becomes (in 1D):

$$m \frac{d^2 x}{dt^2} = -6\pi\mu a \frac{dx}{dt} + \eta \quad . \quad (5.52)$$

This is an example of what is now called a *Langevin equation*. Because η is a random variable, x becomes one, much like the promotion of operator types in a computer program. Therefore we cannot solve for x directly; we must instead use this differential equation to solve for observable quantities that depend on it. To do this, first recognize that

$$\frac{d(x^2)}{dt} = 2x \frac{dx}{dt} \quad (5.53)$$

and

$$\frac{d^2(x^2)}{dt^2} = 2 \left(\frac{dx}{dt} \right)^2 + 2x \frac{d^2 x}{dt^2} = 2v^2 + 2x \frac{d^2 x}{dt^2} \quad . \quad (5.54)$$

Using this, if we multiply both sides of equation (5.52) by x we can rewrite it as

$$\frac{m}{2} \frac{d^2(x^2)}{dt^2} - mv^2 = -3\pi\mu a \frac{d(x^2)}{dt} + \eta x \quad . \quad (5.55)$$

Next, let's take the time expectation value

$$\frac{m}{2} \frac{d^2 \langle x^2 \rangle}{dt^2} + 3\pi\mu a \frac{d \langle x^2 \rangle}{dt} = \underbrace{m \langle v^2 \rangle}_{kT} + \underbrace{\langle \eta x \rangle}_0 \quad . \quad (5.56)$$

In the first term on the right hand side, we've used the fact that the particle is in thermodynamic equilibrium with the fluid to apply the *Equipartition Theorem* [Gershenfeld,

1999a], which tells us that

$$\frac{1}{2}m\langle v^2 \rangle = \frac{D}{2}kT \quad , \quad (5.57)$$

where D is the dimension (1 in this case), k is Boltzmann's constant 1.38×10^{-23} (J/K), and T is the temperature (in Kelvin). The second term vanishes because the rapidly fluctuating noise term is uncorrelated with the slowly moving particle. Therefore,

$$\frac{d^2 \langle x^2 \rangle}{dt^2} + \frac{3\pi\mu a}{m} \frac{d \langle x^2 \rangle}{dt} = \frac{2kT}{m} \quad . \quad (5.58)$$

This is now an ordinary differential equation for the variance, which can easily be solved to find

$$\langle x^2 \rangle = Ae^{-6\pi\mu at/m} + \frac{kT}{3\pi\mu a} t \quad . \quad (5.59)$$

The first term is a rapidly decaying exponential transient, leaving us with

$$\langle x^2 \rangle = \frac{kT}{3\pi\mu a} t \quad . \quad (5.60)$$

This result agrees with the form of Einstein's calculation (Problem 5.3), even though we got here by a very different path. Solving more general stochastic differential equations, and justifying assumptions such as throwing out the $\langle \eta x \rangle$ term, requires extending ordinary calculus to integrals of stochastic functions. This is done by the Ito and the Stratonovich calculus [Gardiner, 1990].

5.3 RANDOM NUMBER GENERATORS

There is a frequent and apparently paradoxical need to use a computer to generate random numbers. In modeling a stochastic system it is necessary to include a source of noise, but computers are (hopefully) not noisy. One solution is to attach a computer peripheral that performs a quantum measurement (which as far as we know can be completely random), or perhaps measures the molecular fluctuations in your coffee cup (which are extremely close to random), but for most people this is not a convenient option. Instead, a more reasonable alternative is to use an algorithm that produces pseudo-random numbers that appear to be more random than can be detected by your application. There is a large difference in what is required to fool a player of a video game and a cryptographic analyst of a one-time pad. There is a corresponding broad range of choices for random number generators, based on how sensitive your problem is to the hidden order that must be present in any deterministic algorithm. While these are numerical rather than analytical methods, and so rightfully belong in the next part of this book, they are so closely connected with the rest of this chapter that it is more natural to include them here.

Why discuss random number generators when most every programming language has one built in? There are two reasons: portability and reliability. By explicitly including an algorithm for generating needed random numbers, a program will be sure to give the same answer whatever system it is run on. And built-in generators range from being much too complex for simple tasks to much too simple for complex needs.

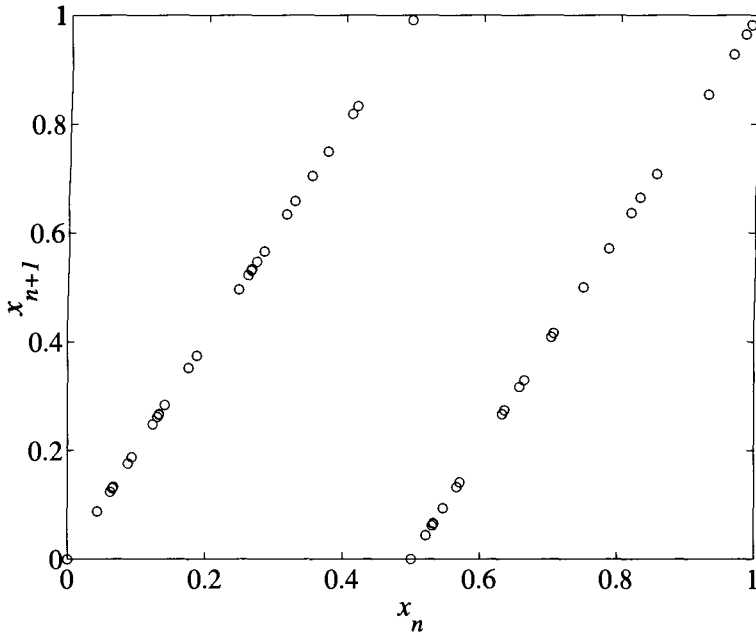


Figure 5.2. 1000 points from the map $x_{n+1} = 2x_n \pmod{1}$.

5.3.1 Linear Congruential

Linear congruential random number generators, and more sophisticated variants, are the most common technique used for producing random numbers. The simplest example is the map

$$x_{n+1} = 2x_n \pmod{1} \quad . \quad (5.61)$$

Starting with an initial value for x_0 (chosen to be between 0 and 1), this generates a series of new values. The procedure is to multiply the old value by 2, take the part that is left after dividing by 1 (i.e., the fractional part), and use this as the new value. This string of numbers is our first candidate as a random number generator. But how random are the successive values of x ? Not very; Figure 5.2 plots x_{n+1} versus x_n for a series of 1000 points. Two things are immediately apparent: the points lie on two lines rather than being uniformly distributed (as they should be for two independent random numbers), and it doesn't look like there are 1000 points in the figure. The first problem can easily be explained by looking back at the definition of the map, which shows that successive pairs of points lie on a line of slope 2, which gets wrapped around because of the *mod* operator. To understand the second problem, consider x written in a fractional binary expansion (where each digit to the right of the binary point stands for $2^{-1}, 2^{-2}, \dots$). Each iteration of this map shifts all the digits one place to the left, and throws out the digit that crosses the binary point. This means that it brings up all the bits of the starting position, and finally settles down to a fixed point at $x = 0$ when all the bits are used up.

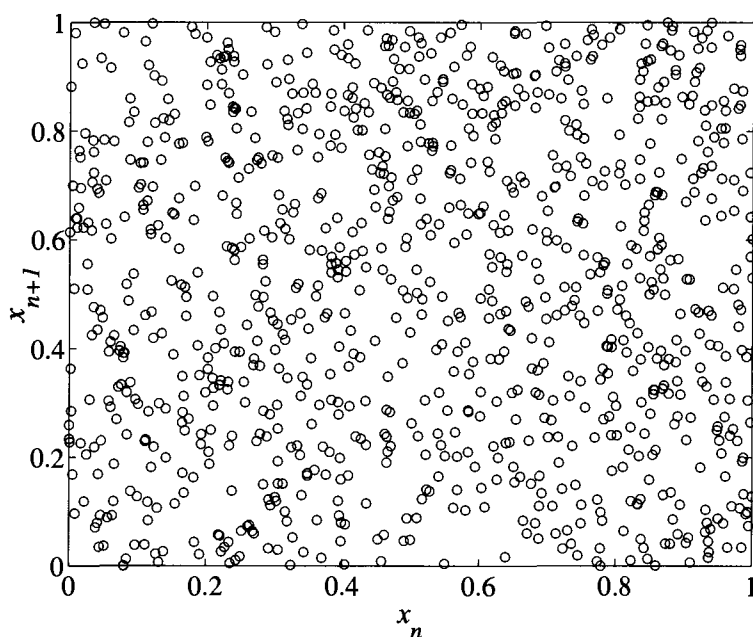


Figure 5.3. 1000 points from the map $x_{n+1} = 8121x_n + 28411 \pmod{134456}$.

This bad example can be generalized to the class of maps

$$x_{n+1} = ax_n + b \pmod{c} \quad . \quad (5.62)$$

The value of a determines the slope of the lines that the points are on and how many lines there will be ($a = 2$ gave us 2 lines). We want this to be as large as possible, so that the lines fill the space as densely as possible. Then b and c must be chosen relative to a so that the period of the map is as long as possible (it doesn't repeat after a few iterations), there are no fixed points that it can get stuck at, and the digits are otherwise as random as they can be. Choosing optimal values for a , b , and c is a surprisingly subtle problem, but good choices have been worked out as a function of the machine word size used to represent x [Knuth, 1981]. For the common case of a 32-bit integer, with the leading bit used as a sign bit, an optimal choice is

$$x_{n+1} = 8121x_n + 28411 \pmod{134456} \quad . \quad (5.63)$$

Iterating this map produces a string of integers between 1 and 134456 (or a fraction between 0 and 1 if the integer is divided by 134456), shown in Figure 5.3. This now appears to be much more random, and is adequate when there is a simple need for some numbers that “look” random.

This is still not a great generator, because there are only 134456 distinct possible values, and so in a string that long it is possible to detect the predictability. It's also easy to see here why the bits of x are not equally random: if x_n is even then x_{n+1} will be odd, and *vice versa*, so the lowest order bit of x simply oscillates at each step. Not very random. To further improve such a simple linear congruential generator, it is possible

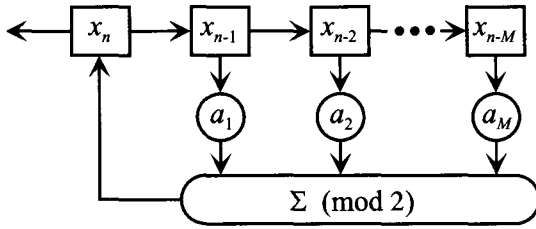


Figure 5.4. A linear feedback shift register.

to add degrees of freedom by techniques such as running multiple generators in parallel and using them to shuffle entries in a large array [Press *et al.*, 1992].

5.3.2 Linear Feedback

Linear congruential generators have the problem that all of the bits in each number are usually not equally random; *linear feedback shift registers (LFSRs)* provide a powerful alternative that can be used to generate truly psuedo-random bits. A binary linear feedback shift register is specified by a *recursion relation*

$$x_n = \sum_{i=1}^M a_i x_{n-i} \pmod{2} \quad . \quad (5.64)$$

This can be viewed as a series of registers through which the bits are shifted, with taps specified by the a_i 's that select the values to be added mod 2 (Figure 5.4).

If the taps are properly chosen, the bits that come out of the end of the shift register are as random as possible. This means that the power spectrum is flat (up to the repeat time, which is $2^M - 1$ for a register with M steps), all possible substrings of bits occur equally often, and so forth. Such a *maximal* LFSR is designed by taking the z -transform of the recursion relation, and finding the taps that make this polynomial have no smaller polynomial factors [Simon *et al.*, 1994]. Table 5.1 gives a (non-unique) choice for maximal taps for a range of register lengths. For example, for order 12 the tap values are 1, 4, 6, 12, so the recursion relation is

$$x_n = x_{n-1} + x_{n-4} + x_{n-6} + x_{n-12} \quad . \quad (5.65)$$

Because the recurrence time is exponential in the length of the register, a surprisingly modest LFSR can have an extremely long period and be hard to distinguish from random (Problem 5.2).

The most sophisticated techniques for making numbers appear random are associated with cryptography, since detectable deviations from randomness can be used to help break a code. In fact, the best way of all to improve a random number generator is to run its output through a cryptographic encoding scheme [Simmons, 1992], which pass much more sophisticated tests of randomness than the algorithms we have covered (which are child's play for a cryptanalyst). This of course does come at the expense of much more computational effort.

Table 5.1. For an LFSR $x_n = \sum_{i=1}^M a_i x_{n-i} \pmod{2}$, lag i values for which $a_i = 1$ for the given order M (all of the other a_i 's are 0).

M	i	M	i	M	i
2	1, 2	13	1, 3, 4, 13	24	1, 2, 7, 24
3	1, 3	14	1, 6, 10, 14	25	3, 25
4	1, 4	15	1, 15	26	1, 2, 6, 26
5	2, 5	16	1, 3, 12, 16	27	1, 2, 5, 27
6	1, 6	17	3, 17	28	3, 28
7	3, 7	18	7, 18	29	2, 29
8	2, 3, 4, 8	19	1, 2, 5, 19	30	1, 2, 23, 30
9	4, 9	20	3, 20	31	3, 31
10	3, 10	21	2, 21	32	1, 2, 22, 32
11	2, 11	22	1, 22	33	13, 33
12	1, 4, 6, 12	23	5, 23	34	1, 2, 27, 34

5.4 SELECTED REFERENCES

[Feller, 1968] Feller, William (1968). *An Introduction to Probability Theory and its Applications*. 3rd edn. New York, NY: Wiley.

A classis reference for probability theory.

[Gardiner, 1990] Gardiner, C.W. (1990). *Handbook of Stochastic Methods*. 2nd edn. New York, NY: Springer-Verlag.

This is a beautiful survey of techniques for working with stochastic systems.

[Knuth, 1981] Knuth, Donald E. (1981). *Semi-Numerical Algorithms*. 2nd edn. *The Art of Computer Programming*, vol. 2. Reading, MA: Addison-Wesley.

The standard starting point for questions about generating and testing random numbers.

[Press *et al.*, 1992] Press, William H., Teukolsky, Saul A., Vetterling, William T., & Flannery, Brian P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edn. New York, NY: Cambridge University Press.

Numerical Recipes has a good collection of practical routines for generating random numbers.

5.5 PROBLEMS

- (5.1) (a) Work out the first three cumulants C_1, C_2 , and C_3 .
- (b) Evaluate the first three cumulants for a Gaussian distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\bar{x})^2/2\sigma^2} \quad . \tag{5.66}$$

- (5.2) (a) For an order 4 maximal LFSR write down the bit sequence.
- (b) If an LFSR has a clock rate of 1 GHZ, how long must the register be for the time between repeats to be the age of the universe ($\sim 10^{10}$ years)?

-
- (5.3) (a) Use a Fourier transform to solve the diffusion equation (5.46) (assume that the initial condition is a normalized delta function at the origin).
- (b) What is the variance as a function of time?
- (c) How is the diffusion coefficient for Brownian motion related to the viscosity of a fluid?
- (d) Write a program (including the random number generator) to plot the position as a function of time of a random walker in 1D that at each time step has an equal probability of making a step of ± 1 . Plot an ensemble of 10 trajectories, each 1000 points long, and overlay error bars of width $3\sigma(t)$ on the plot.
- (e) What fraction of the trajectories should be contained in the error bars?

Bibliography

- [Abarbanel *et al.*, 1993] Abarbanel, H.D.I., Brown, R., Sidorowich, J.J., & Tsimring, L.S. (1993). The Analysis of Observed Chaotic Data in Physical Systems. *Reviews of Modern Physics*, 65, 1331–92.
- [Abelson *et al.*, 1997] Abelson, Harold, Mayer, Meinhard E., Sussman, Gerald J., & Wisdom, Jack. (1997). *A Computational Framework for Variational Mechanics*. Preprint.
- [Abramowitz & Stegun, 1965] Abramowitz, Milton, & Stegun, Irene A. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. New York, NY: Dover.
- [Acton, 1990] Acton, Forman S. (1990). *Numerical Methods That Work*. Washington, DC: Mathematical Association of America.
- [Aho *et al.*, 1974] Aho, Alfred V., Hopcroft, John E., & Ullman, Jeffrey D. (1974). *The Design and Analysis of Computer Algorithms*. Reading, MA: Addison-Wesley.
- [Akaike, 1979] Akaike, H. (1979). A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting. *Biometrika*, 66, 237–242.
- [Ames, 1992] Ames, William F. (1992). *Numerical Methods for Partial Differential Equations*. 3rd edn. Boston, MA: Academic Press.
- [Arfken & Weber, 1995] Arfken, George B., & Weber, Hans J. (1995). *Mathematical Methods For Physicists*. 4th edn. San Diego, CA: Academic Press.
- [Auerbach *et al.*, 1992] Auerbach, D., Grebogi, C., Ott, E., & Yorke, J.A. (1992). Controlling Chaos in High dimensional Systems. *Phys. Rev. Lett.*, 69, 3479–82.
- [Baker & Graves-Morris, 1996] Baker, George A., & Graves-Morris, Peter. (1996). *Padé Approximants*. 2nd edn. New York, NY: Cambridge University Press.
- [Balazs & Jennings, 1984] Balazs, N.L., & Jennings, B.K. (1984). Wigner's Function and Other Distribution Functions in Mock Phase Spaces. *Physics Reports*, 104, 347–91.
- [Balian, 1991] Balian, Roger. (1991). *From Microphysics to Macrophysics : Methods and Applications of Statistical Physics*. New York, NY: Springer-Verlag. Translated by D. ter Haar and J.F. Gregg, 2 volumes.
- [Barron, 1993] Barron, Andrew R. (1993). Universal Approximation Bounds for Superpositions of a Sigmoidal Function. *IEEE Transactions on Information Theory*, 39, 930–45.
- [Batchelor, 1967] Batchelor, George Keith. (1967). *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.
- [Bathe, 1996] Bathe, Klaus-Jürgen. (1996). *Finite Element Procedures*. Englewood Cliffs, NJ: Prentice-Hall.
- [Bechtel & Abrahamsen, 1991] Bechtel, William, & Abrahamsen, Adele. (1991). *Connectionism and the Mind: an Introduction to Parallel Processing in Networks*. Cambridge, MA: B. Blackwell.
- [Beck, 1990] Beck, C.(1990). Upper And Lower Bounds on the Renyi Dimensions and the Uniformity of Multifractals. *Physica D*, 41, 67–78.

- [Bell & Sejnowski, 1995] Bell, A.J., & Sejnowski, T.J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7, 1129–59.
- [Bennett, 1988] Bennett, Charles H. (1988). Notes on the History of Reversible Computation. *IBM J. Res. Dev.*, 32, 16–23.
- [Bernardo & Smith, 1994] Bernardo, Jose M., & Smith, Adrian F.M. (1994). *Bayesian Theory*. New York, NY: Wiley.
- [Besag et al., 1995] Besag, J., Green, P.J., Higdon, D., & Mengersen, K. (1995). Bayesian Computation and Stochastic Systems. *Statistical Science*, 10, 3–66.
- [Bishop, 1996] Bishop, Christopher M. (1996). *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press.
- [Boer & Uhlenbeck, 1961] Boer, J. De, & Uhlenbeck, G. (eds). (1961). *Studies in Statistical Mechanics*. Vol. 1. New York, NY: Interscience Publishers.
- [Box et al., 1994] Box, George E.P., Jenkins, Gwilym M., & Reinsel, Gregory C. (1994). *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, NJ: Prentice Hall.
- [Brenan et al., 1996] Brenan, K.E., Campbell, S.L., & Petzold, L.R. (1996). *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*. Philadelphia, PA: SIAM.
- [Brown & Hwang, 1997] Brown, Robert Grover, & Hwang, Patrick Y.C. (1997). *Introduction to Random Signals and Applied Kalman Filtering*. 3rd edn. New York, NY: Wiley.
- [Buntine, 1994] Buntine, W.L. (1994). Operations for Learning with Graphical Models. *Journal of Artificial Intelligence Research*, 2, 159–225.
- [Buntine, 1996] Buntine, W.L. (1996). A Guide to the Literature on Learning Probabilistic Networks from Data. *IEEE Transactions on Knowledge and Data Engineering*, 8, 195–210.
- [Buntine & Weigend, 1991] Buntine, Wray L., & Weigend, Andreas S. (1991). Bayesian Back-Propagation. *Complex Systems*, 5, 603–43.
- [Casdagli, 1992] Casdagli, Martin. (1992). A Dynamical Systems Approach to Modeling Input-Output Systems. *Pages 265–81 of: Casdagli, M., & Eubank, S. (eds), Nonlinear Modeling and Forecasting*. Santa Fe Institute Studies in the Sciences of Complexity. Redwood City, CA: Addison-Wesley.
- [Catlin, 1989] Catlin, Donald E. (1989). *Estimation, Control, and the Discrete Kalman Filter*. Applied Mathematical Sciences, vol. 71. New York, NY: Springer Verlag.
- [Chaitin, 1990] Chaitin, G.J. (1990). *Information, Randomness & Incompleteness*. 2nd edn. Series in Computer Science, vol. 8. Singapore: World-Scientific.
- [Chaitin, 1994] Chaitin, G.J. (1994). Randomness & Complexity in Pure Mathematics. *International Journal of Bifurcation and Chaos*, 4, 3–15.
- [Cheeseman et al., 1991] Cheeseman, P., Kanefsky, B., & Taylor, W. (1991). Where the Really Hard Problems Are. *Pages 331–37 of: Proc. of the 12th. International Joint Conference on A.I. (IJCAI-91)*. San Mateo, CA: Morgan Kaufmann.
- [Chui et al., 1994] Chui, Charles K., Montefusco, Laura, & Puccio, Luigia (eds). (1994). *Wavelets: Theory, Algorithms, and Applications*. San Diego, CA: Academic Press.
- [Coddington & Levinson, 1984] Coddington, Earl A., & Levinson, Norman. (1984). *Theory of Ordinary Differential Equations*. Malabar, FL: R.E. Krieger.
- [Collet & Eckmann, 1980] Collet, Pierre, & Eckmann, Jean-Pierre. (1980). *Iterated Maps on the Interval as Dynamical Systems*. Boston, MA: Birkhauser.
- [Comon, 1994] Comon, Pierre. (1994). Independent Component Analysis: A New Concept? *Signal Processing*, 36, 287–314.
- [Cook et al., 1989] Cook, Robert D., Malkus, David S., & Plesha, Michael E. (1989).

- Concepts and Applications of Finite Element Analysis*. 3rd edn. New York, NY: Wiley.
- [Cooley & Tukey, 1965] Cooley, James W., & Tukey, John W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, 19, 297–301.
- [Cover & Thomas, 1991] Cover, Thomas M., & Thomas, Joy A. (1991). *Elements of Information Theory*. New York, NY: Wiley.
- [Cybenko, 1989] Cybenko, G. (1989). Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems*, 2, 303–14.
- [Daubechies, 1988] Daubechies, I. (1988). Orthonormal Basis of Compactly Supported Wavelets. *Comm. Pure Applied Math.*, 41, 909–96.
- [Dawson *et al.*, 1994] Dawson, S., Grebogi, C., Sauer, T., & Yorke, J.A. (1994). Obstructions to Shadowing When a Lyapunov Exponent Fluctuates about Zero. *Phys. Rev. Lett.*, 73, 1927–30.
- [Dempster *et al.*, 1977] Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum Likelihood From Incomplete Data via the EM Algorithm. *J. R. Statist. Soc. B*, 39, 1–38.
- [Doolen *et al.*, 1990] Doolen, Gary D., Frisch, Uriel, Hasslacher, Brosi, Orszag, Steven, & Wolfram, Stephen (eds). (1990). *Lattice Gas Methods for Partial Differential Equations*. Santa Fe Institute Studies in the Sciences of Complexity. Reading, MA: Addison-Wesley.
- [Doyle *et al.*, 1992] Doyle, John C., Francis, Bruce A., & Tannenbaum, Allen R. (1992). *Feedback Control Theory*. New York, NY: Macmillan.
- [Dubrulle *et al.*, 1991] Dubrulle, B., Frisch, U., Hénon, M., & Rivet, J.-P. (1991). Low Viscosity Lattice Gases. *Physica D*, 47, 27–29.
- [Duda & Hart, 1973] Duda, Richard O., & Hart, Peter E. (1973). *Pattern Classification and Scene Analysis*. New York, NY: Wiley.
- [Durbin, 1960] Durbin, J. (1960). The Fitting of Time-Series Models. *Rev. Int. Inst. Statist.*, 28, 233–43.
- [Efron, 1983] Efron, Bradley. (1983). Estimating the Error Rate of a Prediction Rule: Improvements on Cross-Validation. *Journal of the American Statistical Association*, 78, 316–31.
- [Ekert & Jozsa, 1996] Ekert, Artur, & Jozsa, Richard. (1996). Quantum Computation and Shor's Factoring Algorithm. *Reviews of Modern Physics*, 68(3), 733–753.
- [Engl *et al.*, 1996] Engl, Heinz W., Hanke, Martin, & Neubauer, Andreas. (1996). *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publishers.
- [Farmer & Sidorowich, 1987] Farmer, J.D., & Sidorowich, J.J. (1987). Predicting Chaotic Time Series. *Phys. Rev. Lett.*, 59, 845–8.
- [Feller, 1968] Feller, William. (1968). *An Introduction to Probability Theory and its Applications*. 3rd edn. New York, NY: Wiley.
- [Foley *et al.*, 1990] Foley, James D., van Dam, Andries, Feiner, Steven K., & Hughes, John F. (1990). *Computer Graphics: Principles and Practice*. 2nd edn. Reading, MA: Addison-Wesley.
- [Forrest, 1993] Forrest, Stephanie. (1993). Genetic Algorithms: Principles of Natural Selection Applied to Computation. *Science*, 261, 872–78.
- [Fraser, 1989] Fraser, A.M. (1989). Information and Entropy in Strange Attractors. *IEEE Trans. Inf. Theory*, 35, 245–62.
- [Frederickson *et al.*, 1983] Frederickson, P., Kaplan, J.L., Yorke, E.D., & Yorke, J.A. (1983). The Liapunov Dimension of Strange Attractors. *J. Diff. Eq.*, 49, 185–207.
- [Fredkin & Toffoli, 1982] Fredkin, Edward, & Toffoli, Tommaso. (1982). Conservative Logic. *Int. J. Theor. Phys.*, 21, 219–53.
- [Frisch *et al.*, 1986] Frisch, E., Hasslacher, B., & Pomeau, Y. (1986). Lattice-Gas Automata for the Navier-Stokes Equation. *Physical Review Letters*, 56, 1505–8.

- [Frisch *et al.*, 1987] Frisch, Uriel, d'Humières, Dominique, Hasslacher, Brosl, Lallemand, Pierre, Pomeau, Yves, & Rivet, Jean-Pierre. (1987). Lattice Gas Hydrodynamics in Two and Three Dimensions. *Complex Systems*, 1, 649–707.
- [Fukunaga, 1990] Fukunaga, Keinosuke. (1990). *Introduction to Statistical Pattern Recognition*. 2nd edn. Boston, MA: Academic Press.
- [Gardiner, 1990] Gardiner, C.W. (1990). *Handbook of Stochastic Methods*. 2nd edn. New York, NY: Springer-Verlag.
- [Gardner, 1970] Gardner, Martin. (1970). The Fantastic Combinations of John Conway's New Solitaire Game "Life". *Scientific American*, October, 120–3. Mathematical Games.
- [Garey & Johnson, 1979] Garey, Michael R., & Johnson, David S. (1979). *Computers And Intractability: A Guide To The Theory Of NP-completeness*. San Francisco, CA: W.H. Freeman.
- [Gear, 1971] Gear, C. William. (1971). *Numerical Initial Value Problems in Ordinary Differential Equations*. Englewood Cliffs, NJ: Prentice-Hall.
- [Gershenfeld, 1989] Gershenfeld, Neil A. (1989). An Experimentalist's Introduction to the Observation of Dynamical Systems. *Pages 310–384 of*: Hao, Bai-Lin (ed), *Directions in Chaos*, vol. 2. Singapore: World Scientific.
- [Gershenfeld, 1992] Gershenfeld, Neil A. (1992). Dimension Measurement on High-Dimensional Systems. *Physics D*, 55, 135–54.
- [Gershenfeld, 1993] Gershenfeld, Neil A. (1993). Information in Dynamics. *Pages 276–80 of*: Matzke, Doug (ed), *Proceedings of the Workshop on Physics of Computation*. Piscataway, NJ: IEEE Press.
- [Gershenfeld, 1999a] Gershenfeld, Neil A. (1999a). *The Physics of Information Technology*. To be published.
- [Gershenfeld, 1999b] Gershenfeld, Neil A. (1999b). *When Things Start To Think*. New York: Henry Holt.
- [Gershenfeld & Chuang, 1997] Gershenfeld, Neil A., & Chuang, Isaac L. (1997). Bulk Spin Resonance Quantum Computation. *Science*, 275, 350–6.
- [Gershenfeld & Grinstein, 1995] Gershenfeld, Neil A., & Grinstein, Geoff. (1995). Entrainment and Communication with Dissipative Pseudorandom Dynamics. *Physical Review Letters*, 74, 5024–7.
- [Gershenfeld *et al.*, 1983] Gershenfeld, Neil A., Schadler, Edward H., & Bilaniuk, Olexa M. (1983). APL and the Numerical Solution of High-Order Linear Differential Equations. *American Journal of Physics*, 51, 743–6.
- [Gershenfeld *et al.*, 1998] Gershenfeld, Neil A., Schoner, Bernd, & Metois, Eric. (1998). *Cluster-Weighted Modeling for Time Series Prediction and Characterization*. Preprint.
- [Gersho & Gray, 1992] Gersho, Allen, & Gray, Robert M. (1992). *Vector Quantization and Signal Compression*. Boston, MA: Kluwer Academic Publishers.
- [Ghahramani & Hinton, 1998] Ghahramani, Zoubin, & Hinton, Geoffrey E. (1998). Hierarchical Nonlinear Factor Analysis and Topographic Maps. Jordan, M.I., Kearns, M.J., & Solla, S.A. (eds), *Advances in Neural Information Processing Systems*, vol. 10. Cambridge, MA: MIT Press.
- [Giona *et al.*, 1991] Giona, M., Lentini, F., & Cimagalli, V. (1991). Functional Reconstruction and Local Prediction of Chaotic Time Series. *Phys. Rev. A*, 44, 3496–502.
- [Giroso *et al.*, 1995] Giroso, Frederico, Jones, Michael, & Poggio, Tomaso. (1995). Regularization Theory and Neural Networks Architectures. *Neural Computation*, 7, 219–69.
- [Goldstein, 1980] Goldstein, Herbert. (1980). *Classical Mechanics*. 2nd edn. Reading, MA: Addison-Wesley.
- [Golub & Van Loan, 1996] Golub, G.H., & Van Loan, C.F. (1996). *Matrix Computations*. 3rd edn. Baltimore, MD: Johns Hopkins University Press.

- [Gould & Eldredge, 1977] Gould, S.J., & Eldredge, N. (1977). Punctuated Equilibrium: The Tempo and Mode of Evolution Reconsidered. *Paleobiology*, 3, 115–51.
- [Goutte, 1997] Goutte, Cyril. (1997). Note on Free Lunches and Cross-Validation. *Neural Computation*, 9, 1246–9.
- [Grassberger & Procaccia, 1983a] Grassberger, P. & Procaccia, I. (1983a). Characterization of Strange Attractors. *Phys. Rev. Lett.*, 50, 346–9.
- [Grassberger & Procaccia, 1983b] Grassberger, P., & Procaccia, I. (1983b). Measuring the Strangeness of Strange Attractors. *Physica D*, 9, 189–208.
- [Guckenheimer & Holmes, 1983] Guckenheimer, John, & Holmes, Philip. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York, NY: Springer-Verlag.
- [Guillemin & Pollack, 1974] Guillemin, Victor, & Pollack, Alan. (1974). *Differential Topology*. Englewood Cliffs, NJ: Prentice-Hall.
- [Haar, 1910] Haar, A. (1910). Zur Theorie der Orthogonalen Funktionen-systeme. *Mathematics Analysis*, 69, 331–71.
- [Hair *et al.*, 1998] Hair, Joseph, Black, William, Tatham, Ronald, & Anderson, Ralph. (1998). *Multivariate Data Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- [Hamilton, 1994] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.
- [Hardy *et al.*, 1976] Hardy, J., de Pazzis, O., & Pomeau, Y. (1976). Molecular Dynamics of a Classical Lattice Gas: Transport Properties and Time Correlation Functions. *Phys. Rev. A*, 13, 1949–61.
- [Hartigan, 1975] Hartigan, John A. (1975). *Clustering Algorithms*. New York, NY: Wiley.
- [Hartman *et al.*, 1990] Hartman, E.J., Keeler, J.D., & Kowalski, J.M. (1990). Layered Neural Networks with Gaussian Hidden Units As Universal Approximations. *Neural Computation*, 2, 210–15.
- [Hasslacher, 1987] Hasslacher, Brosi. (1987). Discrete Fluids. *Los Alamos Science*, 175–217.
- [Hertz *et al.*, 1991] Hertz, John A., Krogh, Anders S., & Palmer, Richard G. (1991). *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley.
- [Hildebrand, 1976] Hildebrand, Francis B. (1976). *Advanced Calculus for Applications*. 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.
- [Hill & Peterson, 1993] Hill, Fredrick J., & Peterson, Gerald R. (1993). *Computer Aided Logical Design with Emphasis on VLSI*. 4th edn. New York, NY: Wiley.
- [Hlawatsch & Boudreaux-Bartels, 1992] Hlawatsch, F., & Boudreaux-Bartels, G.F. (1992). Linear and Quadratic Time-Frequency Signal Representations. *IEEE Signal Processing Magazine*, 9, 21–67.
- [Honerkamp, 1994] Honerkamp, Josef. (1994). *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis*. New York, NY: VCH. Translated by Katja Lindenberg.
- [Huang, 1987] Huang, Kerson. (1987). *Statistical Mechanics*. 2nd edn. New York, NY: Wiley.
- [Hull, 1993] Hull, John. (1993). *Options, Futures, and Other Derivative Securities*. 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- [Jordan & Jacobs, 1994] Jordan, M.I., & Jacobs, R.A. (1994). Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation*, 6, 181–214.
- [Kamen, 1990] Kamen, Edward W. (1990). *Introduction to Signals and Systems*. 2nd edn. New York, NY: Macmillan.
- [Kearns & Vazirani, 1994] Kearns, Michael J., & Vazirani, Umesh V. (1994). *An Introduction to Computational Learning Theory*. Cambridge, MA: MIT Press.
- [Kempf & Frazier, 1997] Kempf, Renate, & Frazier, Chris (eds). (1997). *OpenGL*

- Reference Manual: The Official Reference Document to OpenGL, Version 1.1.* 2nd edn. Reading, MA: Addison Wesley.
- [Kirk *et al.*, 1993] Kirk, David B., Kerns, Douglas, Fleischer, Kurt, & Barr, Alan H. (1993). *Analog VLSI Implementation of Multi-Dimensional Gradient Descent*. Advances in Neural Information Processing Systems, vol. 5. San Mateo, CA: Morgan Kaufmann, pp. 789–96.
- [Kirkpatrick, 1984] Kirkpatrick, S. (1984). Optimization by Simulated Annealing: Quantitative Studies. *J. Stat. Phys.*, **34**, 975–86.
- [Kirkpatrick & Selman, 1994] Kirkpatrick, S., & Selman, B. (1994). Critical Behavior in the Satisfiability of Random Boolean Expressions. *Science*, **264**, 1297–301.
- [Kirkpatrick *et al.*, 1983] Kirkpatrick, S., Gelatt Jr., C.D., & Vecchi, M.P. (1983). Optimization by Simulated Annealing. *Science*, **220**, 671–80.
- [Knuth, 1981] Knuth, Donald E. (1981). *Semi-Numerical Algorithms*. 2nd edn. The Art of Computer Programming, vol. 2. Reading, MA: Addison-Wesley.
- [Kohonen, 1989] Kohonen, Teuvo. (1989). *Self-Organization and Associative Memory*. 3rd edn. New York, NY: Springer-Verlag.
- [Kolmogorov, 1957] Kolmogorov, A.N. (1957). On The Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of One Variable and Addition. *Doklady Akademii Nauk SSSR*, **114**, 953–6.
- [Landauer, 1961] Landauer, Rolf. (1961). Irreversibility and Heat Generation in the Computing Process. *IBM J. Res. Dev.*, **5**, 183–91.
- [Lawler & Wood, 1966] Lawler, E.W., & Wood, D.E. (1966). Branch-and-Bound Methods: a Survey. *Operations Research*, **14**, 699–719.
- [Leff & Rex, 1990] Leff, Harvey S. & Rex, Andrew F. (eds). (1990). *Maxwell's Demon: Entropy, Information, Computing*. Princeton, NJ: Princeton University Press.
- [Lei & Jordan, 1996] Lei, Xu, & Jordan, M.I. (1996). On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural Computation*, **8**, 129–51.
- [Levinson, 1947] Levinson, N. (1947). The Wiener RMS (Root-Mean-Square) Error Criterion in Filter Design and Prediction. *Journal of Mathematical Physics*, **25**, 261–78.
- [Lewis & Papadimitriou, 1981] Lewis, Harry R., & Papadimitriou, Christos H. (1981). *Elements of the Theory of Computation*. Englewood Cliffs, NJ: Prentice-Hall.
- [Liu & Qian, 1995] Liu, Pei-Dong, & Qian, Min. (1995). *Smooth Ergodic Theory of Random Dynamical Systems*. Lecture Notes In Mathematics, vol. 1606. New York, NY: Springer-Verlag.
- [Lloyd, 1993] Lloyd, Seth. (1993). A Potentially Realizable Quantum Computer. *Science*, **261**, 1569–1571.
- [Lorenz, 1963] Lorenz, E. (1963). Deterministic Nonperiodic Flow. *J. Atmospheric Sci.*, **20**, 130–41.
- [MacQueen, 1967] MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Pages 281–97 of: Cam, L. M. Le, & Neyman, J. (eds), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. Berkeley: University of California Press.
- [Macready & Wolpert, 1996] Macready, William G., & Wolpert, David H. (1996). What Makes an Optimization Problem Hard? *Complexity*, **1**, 40–6.
- [Mandelbrot, 1983] Mandelbrot, Benoit B. (1983). *The Fractal Geometry of Nature*. New York, NY: W.H. Freeman.
- [Marcienkiewicz, 1939] Marcienkiewicz, J. (1939). *Math. Z.*, **44**, 612.
- [Marquardt, 1963] Marquardt, Donald W. (1963). An Algorithm for the Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11**, 431–41.
- [McLachlan & Basford, 1988] McLachlan, G.J., & Basford, K.E. (1988). *Mixture Models: Inference and Applications to Clustering*. New York, NY: Marcel Dekker.

- [Merton, 1993] Merton, Robert K. (1993). *On The Shoulders Of Giants: A Shandean Postscript*. Chicago: University of Chicago Press.
- [Metropolis *et al.*, 1953] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., & Teller, E. (1953). Equation Of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087–92.
- [Mezard, 1987] Mezard, Marc. (1987). *Spin Glass Theory and Beyond*. Singapore: World Scientific.
- [Miles, 1981] Miles, J.W. (1981). The Korteweg-de Vries Equation: A Historical Essay. *Journal of Fluid Mechanics*, 106, 131–47.
- [Minsky & Papert, 1988] Minsky, Marvin, & Papert, Seymour. (1988). *Perceptrons: An Introduction to Computational Geometry*. Expanded edn. Cambridge, MA: MIT Press.
- [Nasrabadi & King, 1988] Nasrabadi, N.M., & King, R.A. (1988). Image Coding Using Vector Quantization: A review. *IEEE Trans. on Communications*, COM-36, 957–71.
- [Nelder & Mead, 1965] Nelder, J.A., & Mead, R. (1965). A Simplex Method for Function Minimization. *Computer Journal*, 7, 308–13.
- [Newland, 1994] Newland, D.E. (1994). Wavelet Analysis of Vibration, Part I: Theory. *Journal of Vibration and Acoustics*, 116, 409–16.
- [Nye, 1992] Nye, Adrian (ed). (1992). *The Definitive Guides to the X Window System*. Sebastopol, CA: O'Reilly and Associates.
- [Oppenheim & Schaffer, 1989] Oppenheim, A.V., & Schaffer, R.W. (1989). *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice Hall.
- [Packard *et al.*, 1980] Packard, N.H., Crutchfield, J.P., Farmer, J.D., & Shaw, R.S. (1980). Geometry From a Time Series. *Phys. Rev. Lett.*, 45, 712–16.
- [Parker, 1977] Parker, R.L. (1977). Understanding Inverse Theory. *Ann. Rev. Earth Planet. Sci.*, 5, 35–64.
- [Parlitz *et al.*, 1997] Parlitz, U., Junge, L., & Kocarev, L. (1997). Subharmonic Entrainment of Unstable Period Orbits and Generalized Synchronization. *Physical Review Letters*, 79, 3158–61.
- [Pearson, 1990] Pearson, Carl E. (1990). *Handbook of Applied Mathematics: Selected Results and Methods*. 2nd edn. New York, NY: Van Nostrand Reinhold.
- [Pecora *et al.*, 1997] Pecora, L.M., Carroll, T.L., Johnson, G.A., Mar, D.J., & Heagy, J.F. (1997). Fundamentals of Synchronization in Chaotic Systems, concepts, and applications. *Chaos*, 7, 520–43.
- [Powell, 1992] Powell, M.J.D. (1992). The Theory of Radial Basis Function Approximation in 1990. *Pages pp. 105–210 of: Light, Will (ed), Advances in Numerical Analysis*, vol. II. Oxford: Oxford University Press.
- [Preparata & Shamos, 1985] Preparata, Franco P. & Shamos, Michael Ian. (1985). *Computational Geometry: An Introduction*. New York, NY: Springer-Verlag.
- [Press *et al.*, 1992] Press, William H., Teukolsky, Saul A., Vetterling, William T., & Flannery, Brian P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd edn. New York, NY: Cambridge University Press.
- [Priestley, 1981] Priestley, Maurice B. (1981). *Spectral Analysis and Time Series*. New York, NY: Academic Press.
- [Priestley, 1991] Priestley, Maurice B. (1991). *Non-Linear and Non-Stationary Time Series Analysis*. San Diego, CA: Academic Press.
- [Prügel-Bennett & Shapiro, 1994] Prügel-Bennett, Adam, & Shapiro, Jonathan L. (1994). Analysis of Genetic Algorithms Using Statistical Mechanics. *Physical Review Letters*, 72, 1305–9.
- [Rabiner, 1989] Rabiner, Lawrence R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77, 257–86.

- [Reichl, 1984] Reichl, L.E. (1984). *A Modern Course in Statistical Physics*. Austin, TX: University of Texas Press.
- [Richards, 1988] Richards, Whitman (ed). (1988). *Natural Computation*. Cambridge, MA: MIT Press.
- [Richardson & Green, 1997] Richardson, Sylvia, & Green, Peter J. (1997). On Bayesian Analysis of Mixtures with an Unknown Number of Components. *J. R. Statist. Soc. B*, 59, 731–92.
- [Rissanen, 1986] Rissanen, J. (1986). Stochastic Complexity and Modelling. *Ann. Stat.*, 14, 1080–100.
- [Rothman & Zaleski, 1994] Rothman, Daniel H., & Zaleski, Stéphane. (1994). Lattice-Gas Models of Phase Separation: Interfaces, Phase Transitions, and Multiphase Flow. *Review of Modern Physics*, 66, 1417–79.
- [Rothman & Zaleski, 1997] Rothman, Daniel H., & Zaleski, Stéphane. (1997). *Lattice-Gas Cellular Automata: Simple Models of Complex Hydrodynamics*. New York, NY: Cambridge University Press.
- [Rumelhart *et al.*, 1996] Rumelhart, D. E., Durbin, R., Golden, R., & Chauvin, Y. (1996). Backpropagation: The Basic Theory. *Pages 533–66 of*: Smolensky, P., Mozer, M. C., & Rumelhart, D. E. (eds), *Mathematical Perspectives on Neural Networks*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- [Russell *et al.*, 1980] Russell, D.A., Hanson, J.D., & Ott, E. (1980). Dimension of Strange Attractors. *Physical Review Letters*, 45, 1175–8.
- [Saff & Snider, 1993] Saff, E.B., & Snider, A.D. (1993). *Fundamentals of Complex Analysis for Mathematics, Science, and Engineering*. 2nd edn. Englewood Cliffs, NJ: Prentice Hall.
- [Sauer *et al.*, 1991] Sauer, T., Yorke, J.A., & Casdagli, M. (1991). Embedology. *J. Stat. Phys.*, 65, 579–616.
- [Sauer *et al.*, 1997] Sauer, T., Grebogi, C., & Yorke, J.A. (1997). How Long do Numerical Chaotic Solutions Remain Valid? *Phys. Rev. Lett.*, 79, 59–62.
- [Scheck, 1990] Scheck, Florian. (1990). *Mechanics : From Newton's Laws to Deterministic Chaos*. New York, NY: Springer-Verlag.
- [Shaw, 1981] Shaw, R. (1981). Strange Attractors, Chaotic Behavior and Information Flow. *Z. Naturforsch. A*, 36A, 80–112.
- [Silverman, 1986] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. New York, NY: Chapman and Hall.
- [Simmons, 1992] Simmons, G.J. (ed). (1992). *Contemporary Cryptology: The Science of Information Integrity*. Piscataway, NJ: IEEE Press.
- [Simon *et al.*, 1994] Simon, M.K., Omura, J.K., Scholtz, R.A., & Levitt, B.K. (1994). *Spread Spectrum Communications Handbook*. New York, NY: McGraw-Hill.
- [Sklar, 1988] Sklar, Bernard. (1988). *Digital Communications: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice Hall.
- [Smyth *et al.*, 1997] Smyth, P., Heckerman, D., & Jordan, M.I. (1997). Probabilistic Independence Networks for Hidden Markov Probability Models. *Neural Computation*, 9, 227–69.
- [Stark *et al.*, 1997] Stark, J., Broomhead, D.S., Davies, M.E., & Huke, J. (1997). Takens Embedding Theorems for Forced and Stochastic Systems. *Nonlinear Analysis*, 30, 5303–314.
- [Stevens, 1990] Stevens, W. Richard. (1990). *UNIX Network Programming*. Englewood Cliffs, NJ: Prentice Hall.
- [Stoer & Bulirsch, 1993] Stoer, J., & Bulirsch, R. (1993). *Introduction to Numerical Analysis*. 2nd edn. New York, NY: Springer-Verlag. Translated by R. Bartels, W. Gautschi, and C. Witzgall.
- [Strang, 1986] Strang, Gilbert. (1986). *Introduction to Applied Mathematics*. Wellesley, MA: Wellesley-Cambridge Press.

- [Strang, 1988] Strang, Gilbert. (1988). *Linear Algebra and its Applications*. 3rd edn. San Diego, CA: Harcourt, Brace, Jovanovich.
- [Taft & Walden, 1990] Taft, Ed, & Walden, Jeff. (1990). *Postscript Language Reference Manual*. 2nd edn. Reading, MA: Addison-Wesley.
- [Takens, 1981] Takens, Floris. (1981). Detecting Strange Attractors in Turbulence. *Pages 366–81 of*: Rand, D.A., & Young, L.S. (eds), *Dynamical Systems and Turbulence*. Lecture Notes in Mathematics, vol. 898. New York, NY: Springer-Verlag.
- [Tanenbaum, 1988] Tanenbaum, Andrew S. (1988). *Computer Networks*. 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.
- [Taylor & Wheeler, 1992] Taylor, Edwin F., & Wheeler, John Archibald. (1992). *Spacetime Physics: Introduction to Special Relativity*. 2nd edn. New York, NY: W.H. Freeman.
- [Temam, 1988] Temam, Roger. (1988). *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*. Applied Mathematical Sciences, vol. 68. New York, NY: Springer-Verlag.
- [Theiler *et al.*, 1992] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J.D. (1992). Testing For Nonlinearity in Time Series: The Method of Surrogate Data. *Physica D*, 58, 77–94.
- [Therrien, 1989] Therrien, Charles W. (1989). *Decision, Estimation, and Classification: An Introduction to Pattern Recognition and Related Topics*. New York, NY: Wiley.
- [Toffoli & Margolus, 1991] Toffoli, Tommaso, & Margolus, Norman. (1991). *Cellular Automata Machines: A New Environment for Modeling*. Cambridge, MA: MIT Press.
- [Toffoli & Quick, 1997] Toffoli, Tommaso, & Quick, Jason. (1997). Three-Dimensional Rotations by Three Shears. *Graphical Models and Image Processing*, 59, 89–95.
- [Tong & Lim, 1980] Tong, H., & Lim, K.S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *J. Roy. Stat. Soc. B*, 42, 245–92.
- [Turing, 1936] Turing, A.M. (1936). On Computable Numbers, With An Application To The *Entscheidungsproblem*. *Proc. London Math. Soc.*, 42, 230–65.
- [Vapnik & Chervonenkis, 1971] Vapnik, V.N., & Chervonenkis, A.Y. (1971). On the Uniform Convergence of Relative Frequencies of Events To Their Probabilities. *Theory of Probability and its Applications*, 16, 264–80.
- [Weigend *et al.*, 1990] Weigend, A. S., Huberman, B. A., & Rumelhart, D. E. (1990). Predicting the Future: A Connectionist Approach. *International Journal of Neural Systems*, 1, 193–209.
- [Weigend *et al.*, 1996] Weigend, A. S., Zimmermann, H. G., & Neuneier, R. (1996). Clearing. *Pages 511–22 of*: Refenes, A.-P. N., Abu-Mostafa, Y., Moody, J., & Weigend, A. (eds), *Neural Networks in Financial Engineering (Proceedings of the Third International Conference on Neural Networks in the Capital Markets, NNCM-95)*. Singapore: World Scientific.
- [Weigend *et al.*, 1995] Weigend, A.S., Mangeas, M., & Srivastava, A.N. (1995). Nonlinear Gated Experts for Time Series: Discovering Regimes and Avoiding Overfitting. *International Journal of Neural Systems*, 6, 373–99.
- [Weigend & Gershenfeld, 1993] Weigend, Andreas S., & Gershenfeld, Neil A. (eds). (1993). *Time Series Prediction: Forecasting the Future and Understanding the Past*. Santa Fe Institute Studies in the Sciences of Complexity. Reading, MA: Addison-Wesley.
- [Wernecke, 1994] Wernecke, Josie. (1994). *The Inventor Mentor: Programming Object-Oriented 3D Graphics With Open Inventor, Release 2*. Reading, MA: Addison-Wesley.

-
- [Whitham, 1974] Whitham, Gerald B. (1974). *Linear and Nonlinear Waves*. New York, NY: Wiley-Interscience.
- [Wolaver, 1991] Wolaver, Dan H. (1991). *Phase-Locked Loop Circuit Design*. Englewood Cliffs, NJ: Prentice Hall.
- [Wolfram, 1986] Wolfram, Stephen. (1986). Cellular Automaton Fluids I: Basic Theory. *Journal of Statistical Physics*, 45, 471–526.
- [Wolpert & Macready, 1995] Wolpert, David H., & Macready, William G. (1995). *No Free-Lunch Theorems for Search*. Santa Fe Institute working paper 95-02-010.
- [Woo et al., 1997] Woo, Mason, Neider, Jackie, & Davis, Tom. (1997). *OpenGL Programming Guide: The Official Guide to Learning OpenGL, Version 1.1*. 2nd edn. Reading, MA: Addison Wesley.
- [Wyld, 1976] Wyld, Henry W. (1976). *Mathematical Methods for Physics*. Reading, MA: W.A. Benjamin.
- [Young & Gregory, 1988] Young, David M., & Gregory, Robert Todd. (1988). *A Survey of Numerical Mathematics*. New York, NY: Dover Publications. 2 volumes.
- [Yule, 1927] Yule, G. U. (1927). On a Method of Investigating Periodicities in Disturbed Series with Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions Royal Society London Ser. A*, 226, 267–98.
- [Zabusky, 1981] Zabusky, N.J. (1981). Computational Synergetics and Mathematical Innovation. *Journal of Computational Physics*, 43, 195–249.
- [Zwillinger, 1992] Zwillinger, Daniel. (1992). *Handbook of Differential Equations*. 2nd edn. New York, NY: Academic Press.