Notes on Random Processes

Charles Byrne (Charles_Byrne@uml.edu)
Department of Mathematical Sciences
University of Massachusetts Lowell
Lowell, MA 01854, USA

March 6, 2007

1 Random Variables as Models

When we use mathematical tools, such as differential equations, probability, or systems of linear equations, to describe a real-world situation, we say that we are employing a *mathematical model*. Such models must be sufficiently sophisticated to capture the essential features of the situation, while remaining computationally manageable. In this chapter we are interested in one particular type of mathematical model, the *random variable*.

Imagine that you are holding a baseball four feet off the ground. If you drop it, it will land on the ground directly below where you held it. The height of the ball at any time during the fall is described by the function $h(t)$ satisfying the ordinary differential equation $h''(t) = -32\frac{\text{ft}}{\text{sec}^2}$. Solving this differential equation with the initial conditions $h(0) = 4$ ft , $h'(0) = 0\frac{\text{ft}}{\text{sec}}$, we find that $h(t) = 4 - 16t^2$. Solving $h(T) = 0$ for T we find the elapsed time T until impact is $T = 0.5$ sec.. The velocity of the ball at impact is $h'(T) = -32T = -16\frac{\text{ft}}{\text{sec}}$.

Now imagine that, instead of a baseball, you are holding a feather. The feather and the baseball are both subject to the same laws of gravity, but now other aspects of the situation, which we could safely ignore in the case of the baseball, become important in the case of the feather. Like the baseball, the feather is subjected to air resistance and to whatever fluctuations in air currents may be present during its fall. Unlike the baseball, however, the effects of the air matter to the flight of the feather; in fact, they become the dominant factors. When we designed our differential-equation model for the falling baseball we performed no experiments to help us understand its behavior. We simply ignored all other aspects of the situation, and included only gravity in our mathematical model. Even the modeling of gravity was slightly simplified, in that we assumed a constant gravitational acceleration, even though Newton's Laws tell us that it increases as we approach the center of the earth. When we drop the ball and find that our model is accurate we feel no need to change it. When we drop the feather we discover immediately that a new model is needed; but what?

The first thing we observe is that the feather falls in a manner that is impossible to predict with accuracy. Dropping it once again, we notice that it behaves differently this time, landing in a different place and, perhaps, taking longer to land. How are we to model such a situation, in which repeated experiments produce different results? Can we say nothing useful about what will happen when we drop the feather the next time?

As we continue to drop the feather, we notice that, while the feather usually does not fall directly beneath the point of release, it does not fall too far away. Suppose we draw a grid of horizontal and vertical lines on the ground, dividing the ground into a pattern of squares of equal area. Now we repeatedly drop the feather and record the proportion of times the feather is (mainly) contained within each square; we also record the elapsed time. As we are about to drop the feather the next time, we may well assume that the outcome will be consistent with the behavior we have observed during the previous drops. While we cannot say for certain where the feather will fall, nor what the elapsed time will be, we feel comfortable making a *probabilistic statement* about the likelihood that the feather will land in any given square and about the elapsed time.

The squares into which the feather may land are finite, or, if we insist on creating an infinite grid, discretely infinite, while the elapsed time can be any positive real number. Let us number the squares as $n = 1, 2, 3, ...$ and let p_n be the proportion of drops that resulted in the feather landing mainly in square n. Then $p_n \geq 0$ and $\sum_{n=1}^{\infty} p_n = 1$. The sequence $p = \{p_n | n = 1, 2, ...\}$ is then a *discrete probability sequence* (dps), or a *probability sequence*, or a *discrete probability.* Now let N be the number of the square that will contain the feather on the next drop. All we can say about N is that, according to our model, the probability that N will equal n is p_n. We call N a *discrete random variable* with *probability sequence* p.

It is difficult to be more precise about what probability really means. When we say that the probability is p_n that the feather will land in square n on the next drop, where does that probability reside? Do we believe that the numbers p_n are *in the feather* somehow? Do these numbers simply describe our own ignorance, so are *in our heads*? Are they a combination of the two, in our heads as a result of our having experienced what the feather did previously? Perhaps it is best simply to view probablity as a type of mathematical model that we choose to adopt in certain situations.

Now let T be the elapsed time for the next feather to hit the ground. What can we say about T? Based on our prior experience, we are willing to say that, for any interval $[a, b]$ within $(0, \infty)$, the probability that T will take on a value within $[a, b]$ is the proportion of prior drops in which the elapsed time was between a and b. Then T is a *continuous random variable*, in that the values it may take on (in theory, at least) lie in a continuum. To help us calculate the probabilities associated with T we use our prior experience to specify a function $f_T(t)$, called the *probability density function* (pdf) of T, having the property that the probability that T will lie between a and b can be calculated as $\int_a^b f_T(t)dt$. Such $f_T(t)$ will have the properties $f_T(t) \geq 0$ for all positive t and $\int_0^\infty f_T(t)dt = 1$.

In the case of the falling feather we had to perform experiments to determine appropriate ps p and pdf $f_T(t)$. In practice, we often describe our random variables using a ps or pdf from a well-studied parametric family of such mathematical objects. Popular examples of such ps and pdf, such as Poisson probabilities and Gaussian pdf, are discussed early in most courses in probability theory.

It is simplest to discuss the main points of random signal processing within the context of discrete signals, so we return there now.

2 Discrete Random Signal Processing

Previously, we have encountered specific discrete functions, such as δ_k, u, e_ω, whose values at each integer n are given by an exact formula. In signal processing we must also concern ourselves with discrete functions whose values are not given by such formulas, but rather, seem to obey only probabilistic laws. We shall need such discrete functions to model noise. For example, imagine that, at each time n, a fair coin is tossed and $x(n) = 1$ if the coin shows heads, $x(n) = -1$ if the coin shows tails. We cannot determine the value of $x(n)$ from any formula; we must simply toss the coins. Given any discrete function x with values $x(n)$ that are either 1 or -1, we cannot say if x was generated by such a coin-flipping manner. In fact, any such x could have been the result of coin flips. All we can say is how likely it is that a particular x was so generated. For example, if $x(n) = 1$ for n even and $x(n) = -1$ for n odd, we feel, intuitively, that it is highly unlikely that such an x came from random coin tossing. What bothers us, of course, is that the values $x(n)$ seem so predictable;

randomness seems to require some degree of unpredictability. If we were given two such sequences, the first being the one described above, with 1 and -1 alternating, and the second exhibiting no obvious pattern, and asked to select the one generated by independent random coin tossing, we would clearly choose the second one. There is a subtle point here, however. When we say that we are "given an infinite sequence" what do we really mean? Because the issue here is not the infinite nature of the sequences, let us reformulate the discussion in terms of finite vectors of length, say, 100, with entries 1 or -1. If we are shown a print-out of two such vectors, the first with alternating 1 and -1, and the second vector exhibiting no obvious pattern, we would immediately say that it was the second one that was generated by the coin-flipping procedure, even though the two vectors are equally likely to have been so generated. The point is that we associate randomness with the absence of a pattern, more than with probability. When there is a pattern, the vector can be described in a way that is significantly shorter than simply listing its entries. Indeed, it has been suggested that a vector is random if it cannot be described in a manner shorter than simply listing its members.

2.1 The Simplest Random Sequence

We say that a sequence $x = \{x(n)\}$ is a *random sequence* or a *discrete random process* if $x(n)$ is a random variable for each integer n. A simple, yet remarkably useful, example is the random-coin-flip sequence, which we shall denote by $c = \{c(n)\}$. In this model a coin is flipped for each n and $c(n) = 1$ if the coin comes up heads, with $c(n) = -1$ if the coin comes up tails. It will be convenient to allow for the coin to be *biased*, that is, for the probabilities of heads and tails to be unequal. We denote by p the probability that heads occurs and $1 - p$ the probability of tails; the coin is called *unbiased* or *fair* if $p = 1/2$. To find the *expected value* of $c(n)$, written $E(c(n))$, we multiply each possible value of $c(n)$ by its probability and sum; that is,

$$E(c(n)) = (+1)p + (-1)(1 - p) = 2p - 1.$$

If the coin is fair then $E(c(n)) = 0$. The variance of the random variable $c(n)$, measuring its tendency to deviate from its expected value, is $var(c(n)) = E([c(n) - E(c(n))]^2)$. We have

$$var(c(n)) = [+1 - (2p - 1)]^2 p + [-1 - (2p - 1)]^2 (1 - p) = 4p - 4p^2.$$

If the coin is fair then $var(c(n)) = 1$. It is important to note that we do not change the coin at any time during the generation of the random sequence c; in particular, the p does not depend on n.

The random-coin-flip sequence c is the simplest example of a discrete random process or a random discrete function. It is important to remember that a random discrete function is not any one particular discrete function, but rather a probabilistic model chosen to allow us to talk about the probabilities associated with the values of the $x(n)$. In the next section we shall use this discrete random process to generate a wide class of discrete random processes, obtained by viewing $c = c(n)$ as the input into a linear, shift-invariant (LSI) filter.

3 Random Discrete Functions or Discrete Random Processes

A linear, shift-invariant (LSI) operator T with impulse response function $h = \{h(k)\}$ operates on any input sequence $x = \{x(n)\}$ to produce the output sequence $y = \{y(n)\}$ according to the convolution formula

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) = \sum_{k=-\infty}^{\infty} x(k)h(n-k). \tag{3.1}$$

We learn more about the system that T represents when we select as input sinusoids at fixed frequencies. Let ω be a fixed frequency in the interval $[-\pi, \pi)$ and let $x = e_\omega$, so that $x(n) = e^{in\omega}$ for each integer n. Then Equation (3.1) shows us that the output is

$$y(n) = H(e^{i\omega})x(n),$$

where

$$H(e^{i\omega}) = \sum_{k=-\infty}^{\infty} h(k)e^{-ik\omega}. \tag{3.2}$$

This function of ω is called the *frequency-response function* of the system. We can learn even more about the system by selecting as input the sequence $x(n) = z^n$, where z is an arbitrary complex number. Then Equation (3.1) gives the output as

$$y(n) = H(z)x(n),$$

where

$$H(z) = \sum_{k=-\infty}^{\infty} h(k)z^{-k}. \tag{3.3}$$

Note that if we select $z = e^{i\omega}$ then $H(z) = H(e^{i\omega})$ as given by Equation (3.2). The function $H(z)$ of the complex variable z is the z-transform of the sequence h and also the *transfer function* of the system determined by h.

Now we take this approach one step further. Let us select as our input $x = \{x(n)\}$ the random-coin-flip sequence $c = \{c(n)\}$, with $p = 0.5$. It is important to note that such an x is not one specific discrete function, but a random model for such functions. The output $y = \{y(n)\}$ is again a random sequence, with

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)c(n-k). \tag{3.4}$$

Clearly, in order for the infinite sum to converge we would need to place restrictions on the sequence h; if $h(k)$ is zero except for finitely many values of k then we have no problem. We shall put off discussion of convergence issues and focus on statistical properties of the output random sequence y.

Let u and v be (possibly complex-valued) random variables with expected values $E(u)$ and $E(v)$, respectively. The covariance between u and v is defined to be

$$cov(u,v) = E([u - E(u)]\overline{(v - E(v))}]),$$

and the cross-correlation between u and v is

$$corr(u,v) = E(u\overline{v}).$$

It is easily shown that $cov(u,v) = corr(u,v) - E(u)\overline{E(v)}$. When $u = v$ we get $cov(u,u) = var(u)$ and $corr(u,u) = E(|u|^2)$. If $E(u) = E(v) = 0$ then $cov(u,v) = corr(u,v)$.

To illustrate, let $u = c(n)$ and $v = c(n-m)$. Then, since the coin is fair, $E(c(n)) = E(c(n-m)) = 0$ and

$$cov(c(n), c(n-m)) = corr(c(n), c(n-m)) = E(c(n)\overline{c(n-m)}).$$

Because the $c(n)$ are independent, $E(c(n)\overline{c(n-m)}) = 0$ for m not equal to 0, and $E(|c(n)|^2) = var(c(n)) = 1$. Therefore

$$cov(c(n), c(n-m)) = corr(c(n), c(n-m)) = 0, \text{for}\, m \neq 0,$$

and

$$cov(c(n), c(n)) = corr(c(n), c(n)) = 1.$$

Returning now to the output sequence $y = \{y(n)\}$ we compute the correlation $corr(y(n), y(n-m)) = E(y(n)\overline{y(n-m)})$. Using the convolution formula Equation (3.4) we find that

$$corr(y(n), y(n-m)) = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h(k)\overline{h(j)} corr(c(n-k), c(n-m-j)).$$

Since

$$corr(c(n-k), c(n-m-j)) = 0, \text{for } k \neq m+j,$$

we have

$$corr(y(n), y(n-m)) = \sum_{k=-\infty}^{\infty} h(k)\overline{h(k-m)}. \tag{3.5}$$

The expression of the right side of Equation (3.5) is the definition of the *autocorrelaton* of the sequence h, denoted $\rho_h(m)$; that is,

$$\rho_h(m) = \sum_{k=-\infty}^{\infty} h(k)\overline{h(k-m)}. \tag{3.6}$$

It is important to note that the expected value of $y(n)$ is

$$E(y(n)) = \sum_{k=-\infty}^{\infty} h(k)E(c(n-k)) = 0$$

and the correlation $corr(y(n), y(n-m))$ depends only on m; neither quantity depends on n and the sequence y is therefore called *weak-sense stationary.* Let's consider an example.

Take $h(0) = h(1) = 0.5$ and $h(k) = 0$ otherwise. Then the system is the two-point moving-average, with

$$y(n) = 0.5x(n) + 0.5x(n-1).$$

With $x(n) = c(n)$ we have

$$y(n) = 0.5c(n) + 0.5c(n-1).$$

In the case of the random-coin-flip sequence c each $c(n)$ is unrelated to any other $c(m)$; the coin flips are independent. This is no longer the case for the $y(n)$; one effect of the filter h is to introduce correlation into the output. To illustrate, since $y(0)$ and $y(1)$ both depend, to some degree, on the value $c(0)$, they are related. Using Equation (3.6) we have

$$\rho_h(0) = h(0)h(0) + h(1)h(1) = 0.25 + 0.25 = 0.5,$$

$$\rho_h(-1) = h(0)h(1) = 0.25,$$

$$\rho_h(+1) = h(1)h(0) = 0.25,$$

and

$$\rho_h(m) = 0, \text{otherwise}.$$

So we see that $y(n)$ and $y(n-m)$ are related, for $m = -1, 0, +1$, but not otherwise.

4 Correlation Functions and Power Spectra

As we have seen, any nonrandom sequence $h = \{h(k)\}$ has its autocorrelation function defined, for each integer m, by

$$\rho_h(m) = \sum_{k=-\infty}^{\infty} h(k)\overline{h(k-m)}.$$

For a random sequence $y(n)$ that is wide-sense stationary, its correlation function is defined to be

$$\rho_y(m) = E(y(n)\overline{y(n-m)}).$$

The *power spectrum* of h is defined for ω in $[-\pi, \pi]$ by

$$S_h(\omega) = \sum_{m=-\infty}^{\infty} \rho_h(m)e^{-im\omega}.$$

It is easy to see that

$$S_h(\omega) = |H(e^{i\omega})|^2,$$

so that $S_h(\omega) \geq 0$. The power spectrum of the random sequence $y = \{y(n)\}$ is defined as

$$S_y(\omega) = \sum_{m=-\infty}^{\infty} \rho_y(m)e^{-im\omega}.$$

Although it is not immediately obvious, we also have $S_y(\omega) \geq 0$. One way to see this is to consider

$$Y(e^{i\omega}) = \sum_{n=-\infty}^{\infty} y(n)e^{-in\omega}$$

and to calculate

$$E(|Y(e^{i\omega})|^2) = \sum_{m=-\infty}^{\infty} E(y(n)\overline{y(n-m)})e^{-im\omega} = S_y(\omega).$$

Given any power spectrum $S_y(\omega)$ we can construct $H(e^{i\omega})$ by selecting an arbitrary phase angle θ and letting

$$H(e^{i\omega}) = \sqrt{S_y(\omega)}e^{i\theta}.$$

We then obtain the nonrandom sequence h associated with $H(e^{i\omega})$ using

$$h(n) = \int_{-\pi}^{\pi} H(e^{i\omega})e^{in\omega}d\omega/2\pi.$$

It follows that $\rho_h(m) = \rho_y(m)$ for each m and $S_h(\omega) = S_y(\omega)$ for each ω.

What we have discovered is that, when the input to the system is the random-coin-flip sequence c, the output sequence y has a correlation function $\rho_y(m)$ that is equal to the autocorrelation of the sequence h. As we just saw, for any weak-sense stationary random sequence y with expected value $E(y(n))$ constant and correlation function $corr(y(n), y(n-m))$ independent of n, there is a LSI system h with $\rho_h(m) = \rho_y(m)$ for each m. Therefore, any weak-sense stationary random sequence y can be viewed as the output of an LSI system, when the input is the random-coin-flip sequence $c = \{c(n)\}$.

5 Random Sinusoidal Sequences

If $A = |A|e^{i\theta}$, with amplitude $|A|$ a positive-valued random variable and phase angle θ a random variable taking values in the interval $[-\pi, \pi]$ then A is a complex-valued random variable. For a fixed frequency ω_0 we define a random sinusoidal sequence $s = \{s(n)\}$ by $s(n) = Ae^{in\omega_0}$. We assume that θ has the uniform distribution over $[-\pi, \pi]$ so that the expected value of $s(n)$ is zero. The correlation function for s is

$$\rho_s(m) = E(s(n)\overline{s(n-m)}) = E(|A|^2)e^{im\omega_0}$$

and the power spectrum of s is

$$S_s(\omega) = E(|A|^2) \sum_{m=-\infty}^{\infty} e^{im(\omega_0 - \omega)},$$

so that, by Equation (**??**), we have

$$S_s(\omega) = E(|A|^2)\delta(\omega - \omega_0).$$

We generalize this example to the case of multiple independent sinusoids. Suppose that, for $j = 1, ..., J$, we have fixed frequencies ω_j and independent complex-valued random variables A_j. We let our random sequence be defined by

$$s(n) = \sum_{j=1}^{J} A_j e^{in\omega_j}.$$

Then the correlation function for x is

$$\rho_s(m) = \sum_{j=1}^{J} E(|A_j|^2)e^{im\omega_j}$$

and the power spectrum for s is

$$S_s(\omega) = \sum_{j=1}^{J} E(|A_j|^2)\delta(\omega - \omega_j).$$

A commonly used model in signal processing is that of independent sinusoids in additive noise.

Let $q = \{q(n)\}$ be an arbitrary weak-sense stationary discrete random sequence, with correlation function $\rho_q(m)$ and power spectrum $S_q(\omega)$. We say that q is white noise if $\rho_q(m) = 0$ for m not equal to zero, or, equivalently, if the power spectrum $S_q(\omega)$ is constant over the interval $[-\pi, \pi]$. The *independent sinusoids in additive noise* model is a random sequence of the form

$$x(n) = \sum_{j=1}^{J} A_j e^{in\omega_j} + q(n).$$

The *signal power* is defined to be $\rho_s(0)$, which is the sum of the $E(|A_j|^2)$, while the noise power is $\rho_q(0)$. The *signal-to-noise ratio* (SNR) is the ratio of signal power to noise power.

It is often the case that the SNR is quite low and it is desirable to process the x to enhance this ratio. The data we have is typically finitely many values of $x(n)$, say for $n = 1, 2, ..., N$. One way to process the data is to estimate $\rho_x(m)$ for some small number of integers m around zero, using, for example, the *lag products* estimate

$$\hat{\rho}_x(m) = \frac{1}{N-m} \sum_{n=1}^{N-m} x(n)\overline{x(n-m)},$$

for $m = 0, 1, ..., M < N$ and $\hat{\rho}_x(-m) = \overline{\hat{\rho}_x(m)}$. Because $\rho_q(m) = 0$ for m not equal to zero, we will have $\hat{\rho}_x(m)$ approximating $\rho_s(m)$ for nonzero values of m, thereby reducing the effect of the noise.

The additive noise is said to be *correlated* or *non-white* if it is not the case that $\rho_x(m) = 0$ for all nonzero m. In this case the noise power spectrum is not constant, and so may be concentrated in certain regions of the interval $[-\pi, \pi]$.

6 Spread-Spectrum Communication

In this section we return to the random-coin-flip model, this time allowing the coin to be biased, that is, p need not be 0.5. Let $s = \{s(n)\}$ be a random sequence, such as $s(n) = Ae^{in\omega_0}$, with $E(s(n)) = \mu$ and correlation function $\rho_s(m)$. Define a second random sequence x by

$$x(n) = s(n)c(n).$$

The random sequence x is generated from the random signal s by randomly changing its signs. We can show that

$$E(x(n)) = \mu(2p-1)$$

and, for m not equal to zero,

$$\rho_x(m) = \rho_s(m)(2p-1)^2,$$

with $\rho_x(0) = \rho_s(0) + 4p(1-p)\mu^2$. Therefore, if $p = 1$ or $p = 0$ we get $\rho_x(m) = \rho_s(m)$ for all m, but for $p = 0.5$ we get $\rho_x(m) = 0$ for m not equal to zero. If the coin is unbiased, then the random sign changes convert the original signal s into white noise. Generally, we have

$$S_x(\omega) = (2p-1)^2 S_s(\omega) + (1-(2p-1)^2)(\mu^2 + \rho_s(0)),$$

which says that the power spectrum of x is a combination of the signal power spectrum and a white-noise power spectrum, approaching the white-noise power spectrum as p approaches 0.5. If the original signal power spectrum is concentrated within a small interval, then the effect of the random sign changes is to spread that spectrum. Once we know what the sequence c is we can recapture the original signal from $s(n) = x(n)c(n)$. The use of such a spread spectrum permits the sending of multiple narrow-band signals, without confusion, as well as protecting against any narrow-band additive interference.

7 Stochastic Difference Equations

The ordinary first-order differential equation $y'(t) + ay(t) = f(t)$, with initial condition $y(0) = 0$, has for its solution $y(t) = e^{-at} \int_0^t e^{as} f(s) ds$. One way to look at such differential equations is to consider $f(t)$ to be the input to a system having $y(t)$ as its output. The system determines which terms will occur on the left side of the differential equation. In many applications the input $f(t)$ is viewed as random noise and the output is then a continuous-time random process. Here we want to consider the discrete analog of such differential equations.

We replace the first derivative with the first difference, $y(n+1) - y(n)$ and we replace the input with the random-coin-flip sequence $c = \{c(n)\}$, to obtain the random difference equation

$$y(n+1) - y(n) + ay(n) = c(n). \tag{7.1}$$

With $b = 1 - a$ and $0 < b < 1$ we have

$$y(n+1) - by(n) = c(n). \tag{7.2}$$

The solution is $y = \{y(n)\}$ given by

$$y(n) = b^n \sum_{k=-\infty}^{n} b^{-k} c(k). \tag{7.3}$$

Comparing this with the solution of the differential equation, we see that the term b^n plays the role of $e^{-at} = (e^{-a})^t$, so that $b = 1 - a$ is substituting for e^{-a}. The infinite sum replaces the infinite integral, with $b^{-k}c(k)$ replacing the integrand $e^{as}f(s)$.

The solution sequence y given by Equation (7.3) is a weak-sense stationary random sequence and its correlation function is

$$\rho_y(m) = b^m/(1 - b^2).$$

Since

$$b^n \sum_{k=-\infty}^{n} b^{-k} = 1 - b$$

the random sequence $(1 - b)^{-1}y(n)$ is an infinite *moving-average* random sequence formed from the random sequence c.

We can derive the solution in Equation (7.3) using z-transforms. The expression $y(n) - by(n-1)$ can be viewed as the output of a LSI system with $h(0) = 1$ and $h(1) = -b$. Then $H(z) = 1 - bz^{-1} = (z - b)/z$ and the inverse $H(z)^{-1} = z/(z - b)$ describes the inverse system. Since

$$H(z)^{-1} = z/(z - b) = 1/(1 - bz^{-1}) = 1 + bz^{-1} + b^2 z^{-2} + ...$$

the inverse system applied to input $c = \{c(n)\}$ is

$$y(n) = c(n) + bc(n-1) + b^2 c(n-2) + ... = b^n \sum_{k=-\infty}^{n} b^{-k} c(k).$$

8 Random Vectors and Correlation Matrices

In estimation and detection theory, the task is to distinguish *signal vectors* from *noise vectors*. In order to perform such a task, we need to know how signal vectors differ from noise vectors. Most frequently, what we have is statistical information. The signal vectors of interest, which we denote by $s = (s_1, ..., s_N)^T$, typically exhibit some patterns of behavior among their entries. For example, a constant signal, such as $s = (1, 1, ..., 1)^T$, has all its entries identical. A sinusoidal signal, such as $s = (1, -1, 1, -1, ..., 1, -1)^T$, exhibits a periodicity in its entries. If the signal is a vectorization of a two-dimensional image, then the patterns will be more difficult to describe, but will be there, nevertheless. In contrast, a typical noise vector, denoted $q = (q_1, ..., q_N)^T$, may have entries that are unrelated to each other, as in white noise. Of course, what is signal and what is noise depends on the context; unwanted interference in radio may be viewed as noise, even though it may be a weather report or a song.

To deal with these notions mathematically, we adopt statistical models. The entries of s and q are taken to be random variables, so that s and q are random vectors. Often we assume that the mean values, $E(s)$ and $E(q)$, are zero. Then patterns that may exist among the entries of these vectors are described in terms of *correlations*. The *noise covariance matrix*, which we denote by Q, has for its entries $Q_{mn} = E((q_m - E(q_m))\overline{(q_n - E(q_n))})$, for $m, n = 1, ..., N$. The signal covariance matrix is defined similarly. If $E(q_n) = 0$ and $E(|q_n|^2) = 1$ for each n, then Q is the *noise correlation matrix*. Such matrices Q are Hermitian and non-negative definite, that is, $x^\dagger Q x$ is non-negative, for every vector x. If Q is a positive multiple of the identity matrix, then the noise is said to be *white noise*.